Representational limitations and consequences of phonetic accommodation:

English and Hungarian speakers' imitation of word-initial voiced and voiceless stops

by

Ildikó Emese Szabó

A dissertation submitted in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

Department of Linguistics

New York University

September, 2020

Gillian Gallagher, co-chair

Laurel MacKenzie, co-chair

Acknowledgements

This work has benefited from a core personal and professional support system that spans 3 countries and 2 continents. I will try to provide an exhaustive list of everyone who contributed, but I know I will inevitably fail. All remaining omissions and errors are my own.

My committee spent hours and hours reading my dissertation. This would have warranted an enormous amount of gratitude even in the best of times. However, I would certainly like to acknowledge that the spring and summer of 2020, when most of this work happened, were hardly the best (or at least certainly not the easiest) of times. I am immensely grateful to these five brilliant women for finding the time to engage with my work in the midst of a pandemic and at a time also crucial in the struggle for racial justice in America. The extraordinary nature of their efforts did not go unnoticed, and I am deeply appreciative.

This dissertation had two co-chairs, whose contributions were both equally generous and essential. I would like to thank Gillian Gallagher, who has been advising me since the end of my first year at NYU. The discussions she inspired both in class and in meetings helped me grow as a linguist and as a phonologist. Tied with her (in all aspects really) is my other co-chair, Laurel Mackenzie. She has been instrumental in piquing my curiosity about accommodation and had a huge role in my work being as sociolinguistically grounded as it is. Both of my co-chairs also were essential in me developing a work(-life) system that's truly sustainable, the importance of which I

cannot overstate. I also cannot leave it unmentioned that Gillian and Laurel are tied in giving the most prompt and thorough feedback ever; would recommend!

I would also like to thank Lisa Davidson, who has been unbelievably generous with her feedback, making the dissertation so much richer. The discussion on voicing in this dissertation is a testimony to just how inspiring it is to work with her. Renee Blake's wealth of knowledge and perspective were indispensable in order for this project to be what it is. The sociolinguistic discussion and perspective of the dissertation shows signs of her guidance and generosity all over! Last but not least, I could not have wished for a better outside member than Susannah Levi. She helped me stay grounded and realistic when necessary (always) and challenged me where there was room for development (also always).

I would also like to thank Maria Gouskova and Juliet Stanton for both the education and their feedback, especially during the process of developing the model than ended up being Chapter 5. This work has also benefited from discussions with Lacey Wade, Chelsea Sanker, as well as attendees of ICPhS 2019, LSA 2020, and PEP Labs. In addition, I am abundantly grateful to Teresa Leung, Hannah Katz, and Seena Berg for making day-to-day life at NYU possible.

This work was also facilitated by many people in Hungary as well. I would first and foremost like to thank Katalin Mády, who provided me access to her lab while running my experiment in Hungary. In particular, I am especially grateful to Kati for helping me out remotely and on a moment's notice when my original recording device failed with participant #2 in the booth (on a day that I had 9 back-to-back appointments scheduled). This project would have also been impossible without Kinga Gárdai, whose encouragement and incredible trouble-shooting saved many a recording day. An enormous thank you goes to my old Hungarian mentors and the local linguist community, especially Péter Rebrus, Miklós Törkenczy, Beáta Gyuris and László Kálmán. Neither my professional nor my personal outlook would be the same without you.

I'd also like to give a huge thank you to all the graduate-student-giants who let me step on their shoulders for a boost from time to time. Thank you, Daniel Szeredi, for helping me make sure nothing ever got lost in cultural translation, Itamar Kastner (along with Aurica, Theo and Aidan) for making sure I landed safely in Berlin, and Daniel Duncan for all switching back and forth between work and random conversations on a moment's notice and for reminding me of music that I really should listen to more often. I owe a great many thanks to Paloma Jeretič for leading the way in scouting out the best options in everything from espressos to long-haul bus rides, Yining Nie for being a great office and room mate (simultaneously!), and Sheng-Fu Wang for both his personal and professional support and for filling our office with the coziest of music. I'd also like to thank Kimberley Baxter for some of the best and most meaningful post-colloquium conversations out there, Omar Agha and Kate Mooney for the board games and potatoes, Jailyn Pena for being my partner in hygge and apple cider donuts, and Guy Tabachnick for giving me a reason to speak Hungarian and alongside Emma Claire for the balcony garden support group.

This would not have been possible without Alicia Parrish and her unwavering support (and the couch that she and Dave, Bruce and Avi allowed me to occupy). I also would like to thank Sarah F Phillips for the coffees and chats that were a lifeline and a profile picture that will be very hard to surpass. I am also beyond grateful to Mary Robinson for being my absolute rock and my galpal through my years at NYU. I simply cannot imagine how much harder grad school is for people without someone like you.

I also owe great many thanks to Suzy Ahn, Alicia Chatten, Maddie Gilbert, Nicole Holiday, Zachary Jaggers, Rebecca Laturnus, Sean Martin, and James Whang for being my p-side community and inspiration always and forever; and Hagen Blix, Woojin Chung, Masha Esipova, Dunja Veselinović, Adina Williams, and Vera Zu for being my s-side liaisons. This list does not stop at the walls of NYU. Stony Brook is an excellent place, full of excellent people, and I loved to have you

all be part of my graduate experience, I'd especially like to thank Jon Rawski, Aniello de Santo, and Chikako Takahashi. You are excellent people and deserve all the good things.

My non-grad-school friends have been no less crucial to this endeavor, by making sure I'm a well-rounded human who spends time not even near work. While some people worked their magic in New York (thank you, Jinexa, Teagan, Alexandra, and the Group!), my friends in Budapest have been no less important. Above all, I am unspeakably grateful to Julis Pándi, who was always ready to be happy, proud, sad or angry with (and let's face it, sometimes for) me. I look to you as a point of reference in more ways than you'd think, and I truly think none of this would have been possible without you. I am also just so lucky to have people like Doró Bartha , Bálint Biczók, András Bujdosó, Emil Hollenbach, Fanni Patay, Anna Réz, Rozi Sulyok, Ádám Varga, and Anna Várhelyi in my life, who have always been there to remind me of all the things in the world that aren't linguistics.

I am a firm believer of the "home is where you travel with your laundry to" philosophy, and this dissertation came from a lot of homes for sure. Aside from my family in Hungary, I have been lucky enough to have a support system in the US too in the form of the Pasternaks (and the Sterns). Nancy, Randy, you took me in with open arms from the get go, and I am so-so grateful for absolutely everything. I would like to thank my Hungarian family: my Grandmother, (*Mama: köszönöm, hogy mindig számíthattam a szeretetedre és a támogatásodra!*) all three of my godparents (Teca, Ildi and Józsi), as well as my parents and my brother. I would like to thank my parents for doing absolutely everything to open as many doors for me as possible while closing as few as possible. Gergő, wow I'm so glad I'm not an only child. You are truly the best brother I can possibly imagine. Finally, I (and by association, this dissertation) owe a whole lot of thanks to Rob Pasternak. Bud, I am in constant awe of the way in which you simultaneously manage to ground me and lift me up. Having you in my life fills me with determination.

Abstract

There is a trade-off relationship between the intra-personal variation coming from accommodation (the process where a speaker's speech is influenced by their interlocutor's speech) and the stability provided by the speaker's phonological representations. The present work approaches this relationship from two angles. On the one hand, I investigate how a speaker's pre-existing representations limit what kinds of targets the speaker can accommodate to. On the other hand, I explore the flip side of this using computational tools: how phonological representations are altered by accommodation.

I conducted a pair of VOT experiments to learn more about how pre-existing contrasts and categories can limit accommodation. Experiment 1 was run with native speakers of English in English (an aspirating language), and Experiment 2 other with native speakers of Hungarian in Hungarian (a prevoicing language). Participants had to shadow artificially manipulated p-initial and b-initial words in their native language. The model talker was female in both studies. Her speech pattern either embodied an exaggerated version of how the participants' native language expresses the /p b/ contrast (an aspirating contrast for English-, a prevoicing contrast for Hungarian-speaking participants) or was the opposite (a prevoicing contrast for English speakers and an aspirating contrast for Hungarians). In addition, pre- and post-exposure reading data were also collected from each participant, as well as pre- and post-exposure labeling results and ratings about the model talker along 9 semantic differential scales (e.g. *organized–unorganized*, *friendly–unfriendly*).

I defined two hypotheses for how contrasts could be preserved. **Maintain contrasts** is a requirement for maintaining a distinction between any pair of contrastive sounds. This is a relatively flexible pressure, which can be satisfied if the contrast is simply shifted to a different range of an already used spectrum. This hypothesis predicts accommodation to the un-native-like realization of the /p b/ contrast (prevoicing contrast for English and aspirating contrast for Hungarian speakers). The alternative hypothesis, **Maintain categories** is a pressure to adhere to certain phonetic details of sound categories. While this requirement is not directly concerned with the preservation of contrasts, it leads to a similar effect by guaranteeing consistency in the phonetic realization of sound categories. If individual categories stay consistent, contrasts will also be preserved passively. This hypothesis predicts that English speakers will not converge with a prevoicing contrast (since the plain /p/ occupies the acoustic space where /b/ typically is in English) nor will Hungarian speakers with an aspirating contrast (for analogous reasons).

The conclusions from the two studies are quite similar, and present evidence for the second hypothesis (**Maintain categories**). In both experiments, participants accommodated to exaggerated versions of their native contrasts, but not to realizations of the contrast that were un-native-like. These results were found both in performances from the reading task and during the shadowing task itself. An inspection of the labeling data did not show any large-scale perceptual shift in the un-native-like condition either. Therefore, results from both experiments support the **Maintain categories** hypothesis, i.e. a token will only be accommodated if it is not phonetically atypical for its category.

These data also have consequences for other topics. The lack of convergence in the un-native-like conditions indicates that neither English nor Hungarian speakers perceive aspiration and prevoicing as two equally salient ends of a unified VOT spectrum. This raises the issue of when modeling VOT as a unified continuum—ranging from negative (prevoicing) to positive (long-lag or aspiration)—is justified. Moreover, this study presents evidence for likeability effects being due to

mostly *Solidarity*- and *Superiority*-related measures, while the model talker's perceived *Dynamism* is a less important component for the purposes of accommodation (at least among the examined populations). While this study found no across the board gender effects, in certain cases males showed more sensitivity to these likeability factors (with the female model talker) than females did. In addition, while the English experiment found effects of ethnicity, these could be reduced to distance effects. That is, this was likely behavior that compensated for Black / African American participants tending to have more prevoiced /b/'s in their baseline reading results.

Aside from the empirical question of how phonological representations limit accommodation, this dissertation also discusses how pre-existing representations are in turn updated and changed over the course of accommodation. This is explored through a computationally explicit model. I first outline a basic exemplar model, which is not only capable of simulating distance effects but allows for even larger-scale longitudinal changes. This model is then extended to incorporate the empirical findings of this dissertation (i.e. a mechanism implementing **Maintain categories**). This reflects a system where phonological representations are in constant interaction and are constantly changing.

Eventually, this can all be incorporated into the language change literature at large. Since broader sound changes have been theorized to be propagated as a result of small, incremental instances of accommodation, any viable theory and model for accommodation has to be compatible with sound change. I argue that **Maintain categories** is in fact compatible with different kinds of long-term sound change, and outline some topics for future research in this direction.

Contents

List of Figures

List of Tables

Chapter 1: Introduction

Accommodation is a process where a speaker's speech production is influenced by input they have just encountered from an interlocutor. This often involves the speaker making their productions more similar to those of their interlocutor, which reflects the speaker's flexibility and capacity for change. In fact, it is often assumed that accommodation forms a necessary part of the propagation of bigger sound changes themselves. However, a speaker hardly ever matches their interlocutor's productions perfectly, and they will retain many characteristics of their own speech through different contexts. This suggests that the speaker's mental representations of sound categories of the language(s) they speak (i.e. their phonological representations) constrain this process. While these representations are phonetically rich—they include episodic details of previously encountered tokens as well as some generalizations about them—and are constantly updated with new input tokens, they do guarantee some level of phonetic stability and continuity across interactions. Thus, productions coming from accommodation reflect a trade-off between variability and consistency.

This dissertation is centered around one aspect of this trade-off: how accommodation and pre-existing phonological representations influence one another. This question is approached from two angles in this work. On the one hand, I investigate how a speaker's pre-existing representations limit what kinds of targets the speaker can accommodate to (*Chapters 2–4*). On the other hand, *Chapter 5* explores the flip side of this using computational tools: how phonological representations are altered by accommodation.

While previous research has uncovered numerous ways in which accommodation can be phonetically or socially selective, little is known about how phonological representations and, specifically, pre-existing categories and contrasts can limit accommodation. I conducted a pair of experiments related to contrast preservation and VOT to learn more about this issue. These experiments also addressed other issues such as phonetic distance effects in accommodation, parallel accommodation of contrasts, the treatment of non-contrastive cues, as well as the relationships between accommodation and likeability, gender, and ethnicity.

Chapter 2 introduces the background for this question and also explores other areas of phonetics and sociolinguistics this study can contribute to. This chapter also defines two hypotheses for how contrasts could be preserved. **Maintain contrasts** is a requirement for maintaining a distinction between any pair of contrastive sounds. While it might specify the phonetic dimension and other factors, this is a relatively flexible pressure, which can be satisfied if the contrast is simply shifted to a different range of an already used spectrum. The alternative, **Maintain categories** is a pressure to adhere to certain phonetic details of sound categories. While this requirement is not directly concerned with the preservation of contrasts, it leads to a similar effect by guaranteeing consistency in the phonetic realization of sound categories. If individual categories stay consistent, contrasts will also be preserved passively. *Chapter 2* concludes by laying out the design for the pair of shadowing experiments that test these hypotheses.

In order to control for the potentially language-specific perceptibility of prevoicing and aspiration as non-native cues two experiments were run, one with an aspirating language and one with a prevoicing one, respectively. Experiment 1 was run with native speakers of English in English (*Chapter 3*) and Experiment 2 other with Native speakers of Hungarian in Hungarian (*Chapter 4*). Participants had to shadow artificially manipulated p-initial and b-initial words in their native language. The model talker was female in both studies. Her speech pattern either embodied an exaggerated version of how the participants' native language expresses the /p b/ contrast (an

aspirating contrast for English-, a prevoicing contrast for Hungarian-speaking participants) or was the opposite (a prevoicing contrast for English speakers and an aspirating contrast for Hungarians). In addition, pre- and post-exposure reading data were also collected from each participant, as well as pre- and post-exposure labeling results and ratings about the model talker along 9 semantic differential scales (e.g. *intelligent–unintelligent*, *friendly–unfriendly*).

While the native-like contrasts served as baselines for how these populations of speakers accommodate within the present experimental paradigm, the critical conditions for testing the two hypotheses were those that exposed English speakers to English words with a word-initial prevoicing contrast (plain /p/, prevoiced /b/) and exposed Hungarian speakers to Hungarian words with an aspirating contrast (aspirated /p/, plain /b/). **Maintain contrasts** predicts that participants should be able to accommodate to the un-native-like contrasts, since the /p b/ contrast in these stimuli is merely shifted, but is still maintained. However, **Maintain categories** predicts that these un-native-like contrasts will not be accommodated to, since a plain stop would be more likely to fall into a category other than the one suggested by lexical information.

For instance, in the English *Extreme Prevoicing* condition, *pooling* /ˈpuːlɪŋ/ is presented with a word-initial 15 ms VOT stop. While the lexical information suggests that it is a /p/ (because there is no English word */ˈbuːlɪŋ/), phonetically it is much more likely to be a /b/. In light of this conflict, **Maintain categories** predicts no accommodation. **Maintain contrasts**, however, predicts that English speakers could be able to shift to interpreting 15ms VOT as /p/, given that they are exposed to an interlocutor who produces /b/ with 130 ms prevoicing.

Chapter 3 presents the results of the English experiment (Experiment 1), and *Chapter 4* presents the results of the Hungarian experiment (Experiment 2). The conclusions from the two studies are quite similar. In both experiments, participants accommodated to exaggerated versions of their native contrasts, but not to realizations of the contrast that were un-native-like. These results were found both in performances from the reading task and during the shadowing task itself.

An inspection of the labeling data did not show any large-scale perceptual shift in the un-native-like condition either. Therefore, results from both experiments support the **Maintain categories** hypothesis, i.e. a token will only be accommodated if it is not phonetically atypical for its category.

These data also had consequences for other topics. The lack of convergence in the un-native-like conditions indicates that neither English nor Hungarian speakers perceive aspiration and prevoicing as two equally salient ends of a unified VOT spectrum. This raises the issue of when modeling VOT as a unified continuum—ranging from negative (prevoicing) to positive (long-lag or aspiration)—is justified. Moreover, this study presents evidence for likeability effects being due to mostly *Solidarity*- and *Superiority*-related measures, while the model talker's perceived *Dynamism* is a less important component for the purposes of accommodation (at least among the examined populations). While this study found no across the board gender effects, in certain cases males showed more sensitivity to these likeability factors (with the female model talker) than females did. In addition, while the English experiment found effects of ethnicity, these could be reduced to distance effects. That is, this was likely behavior that compensated for Black / African American participants tending to have more prevoiced /b/'s in their baseline reading results.

Aside from the empirical question of how phonological representations limit accommodation, this dissertation also discusses how pre-existing representations are in turn updated and changed over the course of accommodation. This is explored through a computationally explicit model in *Chapter 5*. In that chapter I first outline a basic model, which is not only capable of simulating distance effects but allows for even larger-scale longitudinal changes. This model is then extended to incorporate the empirical findings of this dissertation (i.e. a mechanism implementing **Maintain categories**). This reflects a system where phonological representations are in constant interaction and are constantly changing.

Finally, *Chapter 6* concludes the dissertation's main take-away, and brings together the various secondary issues it addresses. Since broader sound changes have been theorized to be

propagated as a result of small, incremental instances of accommodation, any viable theory and model for accommodation has to be compatible with sound change. Therefore, at the end of *Chapter 6*, I also argue that **Maintain categories** is in fact compatible with different kinds of long-term sound change, and outline some topics for future research in this direction.

Chapter 2: Empirical background

In this chapter I provide the necessary background that contextualizes the empirical findings of this dissertation. The empirical question this work asks is how a speaker's representations (specifically, their contrasts) limit the way they accommodate. The first few sections (*Sections 2.1-2.4*) establish this main empirical question and outline how it could be tested.

I will first summarize the literature on accommodation both being a great source of intra-speaker variation and a testimony to the flexibility of our representations and it also being limited in crucial ways at the same time (*Section 2.1*). Within this section I will also point out the gap this work occupies: representational limitations (*Section 2.1.4*). This work investigates the mechanism that drives one such limiting factor in particular, contrast preservation. Then, I will formulate two hypotheses representing two plausible mechanisms of contrast preservation, which this dissertation will test (*Section 2.2*).

I will continue by providing some background on voicing contrasts, through which this work compares the two hypotheses (*Section* 2.3). This issue will be discussed both from a cross-linguistic perspective, then with the two specific languages of interest, English and Hungarian, in mind. Subsequently, I will outline the two conditions that the two experiments in this work use to investigate the two hypotheses through the case of voicing accommodation along with the predictions each of the two hypotheses make for these specific conditions.

I will then move on to discuss the literature on other issues that this dissertation also contributes to, namely how accommodation is impacted by likeability, gender, and ethnicity (*Section 2.5*). After having explored both the main and secondary considerations for the experiment, I will present the design elements of the two experiments, which will be the same across English and Hungarian participants (*Section 2.6*). Finally, I will offer a brief summary of all the questions or issues that this dissertation could touch on (*Section 2.7*).

2.1 Accommodation is prolific but limited

Accommodation is the process whereby a speaker changes the way they speak as a reaction to their interlocutor's speech. Accommodation can take many shapes, and can affect many parts of language. I will first review some nomenclature (*Section 2.1.1*), then demonstrate that accommodation is a great source of intra-speaker variation and can surface in a number of varied ways (*Section 2.1.2*). Later, I will point out that it has its own limitations (*Section 2.1.3*). Whilte previous literature has raised attention to how accommodation is both phonetically and socially selective, the empirical half of this dissertation is concerned with the ways in which the speaker's own representation can limit their ability to accommodate to certain input from their interlocutor. *Section 2.1.4* identifies contrast preservation through which this question will be investigated.

2.1.1 Nomenclature

Accommodation has been the subject of much of the recent literature on phonetics and phonology, albeit under various names such as divergence, alignment, synchrony, entrenchment, entrainment, mimicry, imitation and accommodation. The different names for the process all have slightly different connotations in the literature. While *alignment*, *entrenchment* and *entrainment* are often used in research to emphasize the automaticity of the process (incl. Pickering and Garrod, 2004), *accommodation* is associated with communication theoretic approaches and sociolinguistics (Giles

et al., 1991). *Mimicry* is more typical in papers from the fields of cognitive science and psychology (Solanki et al., 2015), and *imitation* might reflect conscious effort (Nielsen, 2011). In this paper, I am going to use *accommodation* to describe the general process of adjusting one's speech in a reaction to someone else's speech irrespective of the direction of the adjustment. I will be using the terms *imitation* and *convergence* to describe a subset of this behavior—cases where the speaker's speech becomes more similar to the interlocutor's speech, but not necessarily implying a complete match. Similarly, *divergence* will be used to describe instances of accommodation where the speaker's speech becomes less similar to the interlocutor's over the course of time.

2.1.2 Accommodation is universal

Accommodation behavior has been observed in a broad range of contexts and phenomena. Humans are not even the only animals exhibiting imitative behavior as part of their normal communication (for a summary, see Tyack, 2008). Within humans, accommodation affects a wide array of behaviors. Humans have been described to converge with one another in terms of gestures (Dijksterhuis and Bargh, 2001) such as smiling, face rubbing and foot shaking (Chartrand and Bargh, 1999), dance (Ros et al., 2014), yawns (Yoon and Tennie, 2010; Norscia and Palagi, 2011), breathing rate (McFarland, 2001), as well as in terms of their thought patterns (Fletcher et al., 1999). This behavior seems so universal that some instances of accommodation can be observed as early as infancy (Balog and Brentari, 2008). Within linguistics, such behavior has been observed in other parts of language and speech such as morphology (Beckner et al., 2016), syntax (Bock, 1986; Estival, 1985; Xu and Reitter, 2016), and choice of lexical items (Garrod and Anderson, 1987). However, in this work I will only use accommodation and other terms to refer to phonetic accommodation.

Within phonetic accommodation, we find descriptions from a myriad of diverse languages. While English examples are by far the most common in the literature, some other examples include Belgium French (Delvaux and Soquet, 2007), Cantonese (Feldstein and Crown, 1990), Dutch

(Hanssen et al., 2007; Mitterer and Ernestus, 2008), German (Lewandowski, 2012; Schweitzer et al., 2019), Hebrew (Yaeger-Dror, 1988), Hungarian (Kontra and Gósy, 1988; Kane et al., 2011), Japanese (Welkowitz et al., 1984), Korean (Kim et al., 2011), Slovak (Reichel et al., 2018), Taiwanese Mandarin (Van den Berg, 1986), and Thai (Beebe, 1981). Several studies have also focused on bilingual participants and language contact, e.g. Beebe (1981) had Chinese-Thai bilingual 4th graders as participants, Welkowitz et al. (1984) included Japanese-American bilingual dyads in Hawaii, and Frisian-Dutch bilinguals have also been studied (Gorter, 1987; Ytsma, 1988). However, these studies are still outweighed by accommodation studies on various English dialects.

Just like how accommodation cannot be tied to a single language, it has also been observed through both perceptual and acoustic measures. In the following, I will review perceptual measures of accommodation as well as the range of phonetic dimensions affected (outside of VOT, which will be discussed in its own section later).

Accommodation as third-party perception

Accommodation is not only something scientists can measure, but something listeners perceive as well. Accordingly, multiple studies use holistic perceptual measures to assess the amount of accommodation happening in a given situation—either in conjunction with or instead of focusing on some specific phonetic dimension(s) (Babel and Bulatov, 2012; Goldinger, 1997; Gregory and Webster, 1996; Gregory et al., 1993, 1997, 2001; Kim, 2012; Pardo, 2006, 2010; Pardo et al., 2013a; Schweitzer et al., 2019).

In these studies, a set of speakers are recorded either performing a shadowing task or in a conversation with another speaker. The measure of accommodation comes from subsequent similarity assessments obtained from a different set of participants, usually in an AXB task. In an AXB task, third-party listeners assess whether the pre- or post-exposure productions of the speaker (A and B) were more similar to the stimulus that the speaker received (X). If the pre- and the post-

exposure productions will be chosen as more similar at an equal (random) rate, no accommodation has taken place (according to the listener). If listeners systematically select either the pre-exposure or the post-exposure productions as more similar to the speaker's stimulus, we have divergence or convergence on our hands, respectively.

Holistic judgments from onlookers can tell us about how accommodation is perceived, thereby reintroducing some of the social environment that accommodation normally takes place in into a controlled laboratory setting. However, since this measure introduces another factor into the situation (the listener), and since the impressions of third-party listeners do not necessarily line up with how much accommodation is happening acoustically (Babel and Bulatov, 2012), in this study I will rely on a raw acoustic measure for accommodation.

Accommodation for suprasegmental properties

Instances of accommodation are found in terms of suprasegmental prosodic and intonational patterns, such as pause duration, pause frequency, turn-taking duration, and turn-taking frequency (Gregory and Hoyt, 1982), and speech rate (Webb, 1970; Cohen Priva et al., 2017; Szabó, 2019). These dimensions indicate that while accommodation is not monolithic, it is still a holistic process rather than something that only affects a particular measure, segment, or feature (Babel and Bulatov, 2012).

Out of all the phonetic dimensions that have been investigated, fundamental frequency is by far one of the most common one. As for f_0, several papers have found evidence for accommodation (Babel and Bulatov, 2012; Goldinger, 1997; Gregory and Webster, 1996; Gregory et al., 1993, 1997, 2001; Kim, 2012; Pardo, 2006, 2010; Pardo et al., 2013a; Schweitzer et al., 2019). What is more, speakers not only accommodate for f_0, converging along this cue plays a crucial role in the social quality of conversations in general. Gregory et al. (1993) found that the conversation is rated by third-party listeners as more "smoothly going" when convergence was attested.

The present study will focus on segmental contrasts and the limitations they pose on accommodation, and therefore suprasegmental measures will not be taken here. In order to control for information that these dimensions might convey, I am not going to manipulate them in this study, and they will stay constant across conditions.

Accommodating for segmental properties

The relationship between accommodation and segmental properties has also been investigated. This study focuses on accommodation for stop VOT, which will be discussed in detail as part of *Section 2.3*. Outside of that, accommodation has been detected for vowels, measured as vowel formants (Babel, 2009; Kim, 2012; Pardo, 2006, 2010; Schweitzer et al., 2019), vowel duration (Kim, 2012; Pardo et al., 2013a), or vowel intensity (Gregory and Hoyt, 1982; Coulston et al., 2002). Convergence was found for all three measures.

Automatic alignment theory

As the previous subsections demonstrate, accommodation affects a large number of phonetic dimensions. Accordingly, researchers have attempted to explain its universality. Some argue that accommodation is a genuinely automatic reaction to hearing one's interlocutor speak, and the speaker just cannot help but converge with the stimulus provided by their interlocutor. This point is represented by Interactive Alignment / Automatic Alignment (Pickering and Garrod, 2004). Under this theory, the goal of a successful conversation is to align mental representations—i.e. to accommodate. Therefore, under Interactive Alignment theory, accommodation is not so much a facilitating factor of communication but rather inherent to its purpose. Similar ideas about the automaticity of accommodation have shown up in other works, such as Goldinger (1998) and Trudgill (2008). Other theories emphasizing the facilitating or hindering factors of accommodation will be discussed in *Section 2.1.3*.

2.1.3 The limits of accommodation

While a speaker can produce highly variable tokens through accommodation, in some core sense they still sound like themselves. This is not only due to the fact that accommodation happens in small, incremental steps, but also due to there being limiting factors. Some of these effects have been referred to as "selectivity" in accommodation (Babel, 2009). The effects that have been explored in the literature so far can be grouped into two categories: effects of phonetic sensitivity and effects of social selectivity.

Phonetic limits of accommodation

How and whether accommodation takes place depends on the phonetic target. Convergence along a given phonetic variable does not straightforwardly correlate with convergence along another—i.e. whether a given person is found to accommodate can depend on what measure we look at. This suggests that automatic accommodation phenomena can be facilitated or hindered based on the sound or phonetic dimension in question. For instance, studies like Pardo et al. (2013a) and Pardo et al. (2017) investigated multiple phonetic dimensions, and found no one-to-one correlation for any pair of them—i.e. accommodation is not a monolithic phenomenon that translates into each phonetic dimension equally. Thus, Pardo et al. warn about the generalizability of accommodation effects in general.

Moreover, accommodation does not even impact all sounds in a natural class the same way, even if we look within the same dimensions. As Babel (2012) puts it, accommodation is phonetically selective. In her study, not all vowels' formants were subject to accommodation (or at least not in the same way). Furthermore, there is evidence indicating that while convergence towards *more* prominent cues is possible, participants do not always converge with input that has *less* of the same cues than their own speech. Such a process has been described for VOT (Nielsen, 2011), where participants accommodated to /p/'s with extremely long VOT, but not to /p/'s with

shortened VOT. At the same time, for example, speech rate studies (e.g. Cohen Priva et al., 2017; Szabó, 2019) are yet to find similar limitations (e.g. that participants would willing to speed up for a fast-paced interlocutor but not willing to slow down for slow-paced ones). This suggests that if such directionality effects exist, they might be restricted to certain phonetic dimensions. These results all indicate that even though accommodation is both incredibly common, and can result in immense variation, it is not limitless. The phonetic factors discussed above can all inhibit accommodation.

Social limits of accommodation

There are also social factors hindering accommodation, and some of the most prominent examples include ideological alignment (Gregory and Hoyt, 1982) and social biases. Gregory and Hoyt (1982) conducted 5 one-on-one conversations with airmen on race relations at the air force, where they found less convergence in conversations with miscommunication and misunderstandings. They conclude that "cultural homogeneity facilitates adaptation" (Gregory and Hoyt, 1982:43).

Yu et al. (2011) show that homophobia impedes convergence with a male model talker who uses same-gender pronouns for his partner (as opposed to the control group where the partner was referred to as she/her/herself). The amount of this impediment is proportionate to the implicit homophobia of the participant, as measured by an Implicit Association Task. Similar effects were found for race and racial bias. Babel (2009, 2012) conducted a study where white native speakers of English had to shadow both a Black and a white model talker. Participants who were not shown a picture of their model talker could not reliably identify the Black speaker as Black, but when they saw him, their accommodation behavior correlated with their implicit Anti-Black bias (as assessed in an Implicit Association Task).

However, bias effects were not replicated with xenophobia. In her shadowing study using monosyllabic and disyllabic words, Kim (2012) investigated whether attitudes towards foreigners influence accommodation towards audio from a proficient non-native model talker. She found no

such relationship: the degree of convergence was independent of her participants' implicit attitude towards foreigners, it only mediated phonetic distance-based effects.

Results from likeability studies are in line with these bias effects: more convergence happens if the speaker finds the interlocutor likeable, and less convergence (or even divergence) happens in case of disliking (Natale, 1975; Babel, 2009; Bane et al., 2010; Pardo et al., 2012; Babel et al., 2014; Sonderegger, 2015; Babel et al., 2019; Schweitzer et al., 2019, inter alia). Since the present work also focuses on likeability effects, previous studies in this area will be discussed in much more detail later.

These effects make it clear that accommodation is mediated by social factors and the speaker's attitudes. It must be noted that some of the limitations are not just quantitative (i.e. that less accommodation happens in certain cases, and thus it would take longer for the speaker to match the interlocutor's values than how much time is available in the conversation), but qualitative as well. In the extreme, convergence can even be completely inhibited and surface as divergence as a result of certain attitudes. With certain speaker–interlocutor dyads, no convergence happens at all.

Social theories of accommodation

From a theoretical perspective, the capacity of social factors to mediate accommodation has long been recognized. Some theories view accommodation as a tool that speakers can (semi-)consciously use to accomplish social goals in a conversation. This line of thinking can be traced back to work not necessarily related to (and even contradicting) accommodation. In theories of identity projection (e.g. Le Page and Tabouret-Keller, 1985) speech style is a way of projecting a desired persona. Under this view, speakers do not converge with their interlocutors' speech, but with the idea of how they themselves would like to be seen instead. Therefore, in its most extreme form it directly contradicts accommodation. Eventually, this line of work was reconciled with audience design (Bell, 1984), where speech style is *chosen* by the speaker with their audience (and potentially onlookers) in mind.

The version that is compatible with audience design (and thus accommodation) is referred to as 'referee design' (Bell, 2002).

This work influenced accommodation research in two ways. The most direct extension of this is Communication Accommodation Theory (CAT: Giles et al., 1987, 1991). Under CAT, the speaker uses accommodation (converging with or diverging from the interlocutor) to modulate their social distance from the interlocutor. In a way, this hinges on the idea that acoustic or linguistic distance is correlated with perceived social distance. This is reinforced by the fact that conversations with accommodation are perceived as more smoothly going than ones without it (see Gregory et al., 1993 for a phonetic example and Chartrand and Bargh, 1999 for a behavioral one). We also find connections between accommodation and leadership role (Pardo, 2006) as well as social status, where lower status speakers, who have more incentive to accommodate to their higher-status interlocutor, do accommodate more. This idea was, to my knowledge, first expressed by Bloomfield (1933) ("[the] humble person is not imitated"; 476), and eventually experimentally tested by Gregory and Webster (1996).

It must be noted that there has also been a trend in recent studies that suggest a marriage of Interactive/Automatic Alignment Theory and CAT. More and more authors subscribe to the view that while accommodation itself is automatic, its *inhibition* can be brought about by unfavorable social percepts (e.g. Babel, 2009). This idea is supported by the fact that patients with frontal brain lesions (correlated with a loss of social inhibition) still imitate their interlocutor (Brass et al., 2003, 2005; Spengler et al., 2010). Evidence from the present work also supports the idea that social factors can mostly be observed as they inhibit rather than facilitate convergence.

These lines of research, especially on status have greatly informed the current study. To maximize the amount of convergence, I am using model talkers who are both older than the college-age participant pool. Thus, my participants will not be higher status than my model talkers, which could facilitate convergence. Even if they do not (because college-aged participants might

not perceive a professional in her 30's as having higher status than themselves), the two model talkers (the English and the Hungarian one) are the same age as each other. By keeping the model talkers' age the same across the two experiments, I try to control differences in perceived status to the extent it is possible. This is of course barring individual differences between the two model talkers and differences in the perception of status cross-culturally, which are inescapable.

2.1.4 The representational limits of accommodation

In the previous sections I demonstrated that accommodation can be selective or limited in terms of the phonetic target (the given sound or phonetic dimension) and in terms of the social factors (the speaker's attitudes towards the interlocutor). There is also some evidence suggesting that phonological rules (i.e. the set of environments triggering an alternation) are not subject to accommodation. These claims are based on studies such as Payne (1980), who observed that while children moving to a Philadelphian suburb are successful at producing the sounds involved in their new local dialect, they do not apply /æ/-tensing in the right environments. Based on such results, Trudgill (1986) concludes that accommodation is phonetic in nature, rather than phonological.

What is yet to be thoroughly investigated is how the speaker's representation of sound categories constrains their ability to accommodate to various kinds of input. I am going to call potential constraints stemming from the speakers phonological knowledge the representational limitations on accommodation. We know that speakers' representation can be very flexible,which is why accommodation can result in immense intra-personal variation in the first place, but how much of their own phonological representation do speakers bring to these interactions? Another way of looking at this question is from the perspective of the phonetic identity of the target. Can certain inputs be invalid targets of accommodation because their phonetic details mismatch the speaker's preconceptions of the token's putative phonemic category? These questions lie at the heart of the present work.

Limits regarding contrast preservation

To answer these questions, the present study compares two hypotheses about the mechanism behind contrast preservation, which is a tendency to preserve phonemic distinctions phonetically. Some of the aforementioned phonetic sensitivity effects, e.g. English speakers' lack of accommodation to shortened VOT, have been explained through contrast preservation. Nielsen (2011) finds that native English speakers do not imitate a version of /p/ with a shortened VOT. She suggests that this might be due to the resulting /p/ encroaching on English /b/'s habitual location in the acoustic space. In the following I will review the available literature on contrast preservation (outside of the context of accommodation), and then highlight points of connection.

Contrast maintenance (or contrast preservation) is one of the forces that keep languages relatively constant and stable. It is a force that prevents mergers, and maintains the individual contrasts and thus the phonological structure and system of a language. If we endorse a theory of language change via accommodation (see next subsection), then forces that prohibit language change can also be prohibiting accommodation itself. The main goal of this dissertation is to investigate through what mechanism sound systems limit phonetic accommodation. The answer to this question could inform us about mechanisms responsible for contrast preservation at large. In order to better appreciate these potential repercussions, we first need to understand the phenomenon of contrast preservation a little better.

Contrast preservation has been crucial to two areas of linguistic research. The earliest reference to such a force has been with respect to chain shifts, push chain shifts in particular. Chain shifts are a series of at least two historical sounds changes, whereby one sound ends up in an area of the phonetic space that used to be occupied by another sound. One type of chain shifts is a 'pull chain' or 'drag chain'. For example, a hypothetical pull chain is /eː/ changing to /iː/, and then another sound, e.g. /aː/ changing to /eː/. After both changes take place, what used to be /aː/ now occupies

the place that a different sound used to (/eː/, that is now /iː/). Another type of chain shifts is a 'push chain', which could arise from reordering the two changes. In the toy example above this would be /aː/ changing to /eː/ first, and thereby urging what was formerly /eː/ to change to for example /iː/ so that the contrast is maintained.

Push chains have been described as phenomena involving the "encroachment of a phoneme into the phonological space of another" and the "the second phoneme changes so that the distinction between the two is maintained" (Gordon, 2002). For over half a decade, contrast preservation has often been the reason scholars cited to explain how chain shifts happen (Martinet, 1952; Labov, 1994, inter alia) and this connection is still used to this day (Watson et al., 2000).

There is another line of work, which looks at a factor that itself influences contrast preservation, namely functional load (Gilliéron, 1918; Trubetzkoy, 1939; Martinet, 1952). The functional load hypothesis expresses the idea that the more pairs of words a given contrast distinguishes (the more "work it does"), the more likely it is to be maintained over time. This idea has been supported by statistical evidence from various languages (Kaplan, 2011; Wedel et al., 2013a,b; Szabó, to appear). In this sense, contrast preservation is also related to anti-homophony effects within the morphological paradigm. The only difference is that these accounts emphasize a preservation of difference (and thus the role of informativity) within a paradigm rather than the informational value of a segmental contrast.

While these two areas (chain shifts and functional load) have been relatively well-researched, less is understood about the short-term mechanisms behind contrast preservation on the level of the individual. It is not clear for instance, on what level of representation contrast preservation applies. Since it seems to be sensitive to contrast-specific functional load, it seems likely that contrast preservation must in some way refer to contrasts as well. However, it is unclear whether contrast preservation is only an abstract principle in speakers' grammars, generalized at the level of contrasts, or whether it also requires an adherence to the phonetic details of individual sound

categories. These two options will be in the focus of this work, and will be fleshed out in the next chapter in the form of two hypotheses. Before laying out these hypotheses, we first have to discuss a caveat: how a phenomenon found in long-term language change can be traced in short-term interactions.

Change-by-accommodation

The reader will be right to notice that accommodation is most often described as a short-term phenomenon (it will be studied here through instances of short-term exposure), whereas contrast preservation is often invoked for language change, often for chain shifts, which could happen over multiple lifetimes. However, this work is not the first linking forces of long-term language change to those of short-term accommodation. There is a body of literature that views accommodation as an essential step in language change.

This idea is related to a more general line of thinking, where "sound change is drawn from a pool of synchronous variation" (see Ohala, 1989, for summary). The more specific change-by-accommodation idea has been broadly endorsed in the literature (Trudgill, 1986, 2004, 2008; Labov, 1990; Pardo, 2006; Garrett and Johnson, 2013, among others). Under this theory, a speaker accommodates during conversations, and these incremental changes over time result in the speaker's norms shifting, and finally, this change on an individual level spreads to other members of the community, becoming a community-level change (as summarized in Sonderegger, 2012, as well). What is important here is the link between the first two steps—as Sonderegger (2012) puts it: the link between short-term individual change and long-term individual change.

The idea of whether a given speaker accommodates to their interlocutor's actual targets (e.g. Cohen Priva et al., 2017) has been questioned. Some claim that speakers converge with their *perceived idea* of their interlocutor's speech rather than the interlocutor's actual phonetic targets (Wade, 2020), while others argue that accommodation is a result of a speaker trying to project

and assert their own identity, and the way they try to do that can depend on the interlocutor, and therefore speakers adjust their speech differently in different conversations (Auer and Hinskens, 2005). However, what the target of accommodation is socially does not call into question the aspect of Change-by-accommodation (CBA) that is most relevant here, namely, that small, short-term adjustments from conversations result in a larger individual change over time.

The connection between the first two steps (short-term individual change and long-term individual change) lies at the heart of much of second dialect acquisition (e.g. Trudgill, 1981; Chambers, 1992; Munro et al., 1999; Nycz, 2013). During this process, often due to relocation, an individual is introduced to a speech community whose norms are different from the speaker's native ones. This also often coincides with the speaker's input from their native speech community drastically decreasing. In these environments, most speakers slowly adapt their new community's speech patterns, although there is great individual variation, much like in short-term accommodation itself. These processes of second dialect acquisition are not unlike what has been described for cross-dialectal accommodation (e.g. Mons Belgian French participants accommodating to the Liége dialect of Belgian French model talker, Delvaux and Soquet, 2007), and the line between the two can be hard to draw. Albeit these changes are most visible when accommodation or acquisition is cross-dialectal, this link can be presumed to be there when adjustments are smaller.

This is further reinforced by studies finding incremental medium-term accommodation. For instance, Pardo et al. (2012) find that college roommates converge to one-another over the course of a semester, especially if they maintain a good relationship. While Sonderegger (2012) points out the immense day-to-day variation, he concludes that in his corpus study on reality-show participants (from Big Brother), a close relationship can result in the two participants' speech becoming more and more similar.

A link between individual-level changes and community-changes has also been demonstrated. There are several studies pointing out that individual longitudinal changes reflect broader

community shifts (e.g. Harrington et al., 2000; Harrington, 2006, 2007; Harrington et al., 2007; Sankoff and Blondeau, 2007; MacKenzie, 2017)

Therefore, one can logically assume that some of the forces that are responsible for promoting or prohibiting sound change might also be traceable within short-term accommodation. That is, that contrast preservation does not just arise on a community level, but it is also there on the level of the individual. If long-term changes emerge from several smaller in-conversation adjustments, it can also be assumed that such forces (e.g. contrast preservation) are also present in individual interactions, blocking accommodation to certain kinds of stimuli.

While we have evidence of contrast preservation surfacing in accommodation studies (e.g. Nielsen, 2008, which will be discussed later), the exact mechanism behind it is not yet understood. In the following section I will offer two hypotheses for what types of restrictions protect contrasts amid synchronic variation.

2.2 Two hypotheses

As mentioned before, at the crux of this dissertation lies the question of how much a speaker's pre-established representations limit them in accommodation—i.e. predetermine what types of targets they can and cannot accommodate to. In this section, I will describe two hypotheses: the speaker either only brings with them a vague notion of which sounds are contrastive in their language or, alternatively, these contrastive sounds are also restricted in terms of what exact phonetic properties they can and cannot have (potentially in addition to the vaguer restriction). Both types of restrictions limit what type of input a speaker can accommodate to, and they could both result in contrasts being maintained.

2.2.1 Maintain contrasts

The first hypothesis is that in speakers' grammar there is only an abstract pressure that mandates that contrasts that preexist in the language must be maintained. I am going to refer to this hypothesis as **Maintain contrasts**. This imperative on what realizations of a contrast are permissible is quite abstract, and therefore also quite permissive. For the sake of specificity, I will assume that this permissiveness is limited to one crucial aspect: where exactly the category boundary lies. In order to do that, restriction needs to be specific along four lines: restrictions are specific to phonetic dimension, relative values, magnitude of difference and environment. In reality, some of these specific criteria might not exist, and thus **Maintain contrasts** is even more abstract and permissive.

In the following, I am going to define Maintain contrasts to be as strict as possible, to illustrate that it still allows for great flexibility that is not possible under its alternative. For instance, an English speaker's grammar might involve a constraint stating that the center of the /p/ and /b/ categories must be distinguished by VOT word-initially in a way that /p/ must always have a longer VOT than /b/ by at least a 30 ms difference between means (30 ms is chosen somewhat arbitrarily here). Specifying VOT mandates an adherence to a *phonetic dimension (or dimensions)*. The fact that it is specified that /p/ should have the longer VOT of the two represents specificity on the level of *relative values*. The inclusion that the difference should be at least 30 ms big, is *magnitudinal* specificity. Finally, this restriction is also specific to an *environment*, i.e. this constraint only applies word-initially, and a similar but different restriction protects the contrast elsewhere, potentially through a different set of cues. Along similar lines, word-finally a similar, but different restriction might mandate that /p/ and /b/ must be distinguished via a difference in the stop's duration and the duration of the preceding vowel.

Under this maximally strict definition, the word-initial restriction of **Maintain contrasts** for the /p b/ contrast is satisfied as long as all of the following four criteria are met: First, /p/ and /b/

are distinguished from one another word-initially. Second, that distinction is (at least) along VOT. Third, this distinction is such that /p/ has to have a higher value (here, a longer VOT) than /b/ along this dimension. Finally, the difference has to be on average 30 ms or more.

If one of these criteria are not met, the contrast is unsatisfactory for **Maintain contrasts**. For instance, it can be violated if /p b/ are merged word-initially (environmental violation)[1], not distinguished by VOT but intensity instead (phonetic dimension violation), a plain /p/ is contrasted with an aspirated /b/ (relative value violation) or the difference is too small (a violation of magnitude). While a **Maintain contrasts**-type restriction can be violated in a number of ways, crucially, it is tolerant of the contrast being shifted to a different range of the given phonetic dimension(s). For instance, **Maintain contrasts** should be satisfied if /p/ and /b/ are contrasted by VOT, but in the negative range (i.e. the speaker has a plain /p/ is contrasted with a prevoiced /b/).

If a given realization of a contrast violates **Maintain contrasts** (i.e. violates a **Maintain contrasts**-type restriction), then it is an illicit realization for that contrast. If the given realization of a contrast is illicit, then a listener will not recognize it as a valid realization of the /p b/ contrast. If lexical information is present, as it often is, then an illicit realization of a contrast does not result in communication failure in the discourse, but the speaker might be viewed less favorably.

To say what an illicit realization means in terms of representations requires us to make further stipulations. I am going to assume that while anomalous tokens can be processed as part of a given category (mainly via lexical information), such a token cannot influence production. In terms of accommodation, it means that tokens in violation of **Maintain contrasts** will not

[1]It is possible that Maintain contrasts is violated in some environments, such as in the case of word-final devoicing. Some argue that variation and neutralization is more common in these environments precisely because it is hard to perceive, for instance, a voicing distinction in these positions. Therefore, Maintain contrasts might also be more permissive of violations in environments where the distinction (and thus compliance with Maintain contrasts) is harder to perceive to begin with.

form a valid target and thus will not be accommodated to. A distribution that satisfies **Maintain contrasts** will not be hindered (by **Maintain contrasts** at least) and all else being equal, it will be a valid target. As such, it will be subject to effects of the already discovered mediating factors (e.g. sympathy, ideological bias etc.) as any other target. This explains why an extremely long VOT is imitated for /p/, while a shortened VOT is not (Nielsen, 2011). A lengthening the VOT of /p/ does not threaten the contrast in any way, while if the VOT is shortened past a certain limit, /p/ and /b/ will not be distinguished by *enough* VOT.

Such **Maintain contrasts**-type restrictions can be made gradient to account for effects of functional load mentioned in *Section 2.1.3*. Since this present work uses only one contrast, cross-contrast restrictions will not be necessary. Therefore, gradience of restriction is omitted from further discussion in this work, while the option is available for future work.

As a result, **Maintain contrasts** types of restrictions are largely flexible. It is satisfied as long as tokens of a sound pair form a bimodal distribution along the expected phonetic dimensions (with maintaining an appropriate distance or non-overlap between the categories). This flexibility straightforwardly allows for (and explains) phenomena like chain shifts, where contrasts are maintained, but one or both members of the contrast are relocated in the acoustic space.

Such flexibility is also seen, for instance, in L1 acquisition, where infants show sensitivity to bimodal distributions, (Maye et al., 2002). This has also been demonstrated with adults, who are more likely to say that two sounds are different if they previously heard them as part of a bimodal rather than a unimodal distribution (Maye and Gerken, 2000). It must be noted, however, that the adult study was carried out with a distinction that was not present in the participants' native language, prevoiced and short-lag alveolar stops. Contrasting prevoiced and short-lag stops will be relevant for this study as well. These behavioral results have also been corroborated in the modeling literature. Mixture of Gaussian models can effectively learn sound categories based on distributional information. Two categories are learned most efficiently when tokens form a bimodal

distribution, with the distributions only allowing for little overlap (e.g. Mielke, 2005; Vallabha et al., 2007; McMurray et al., 2009).

Another type of evidence for flexible representations comes from the attunement literature. Attunement is the process during which a listener adjusts their perception to the speaker they are listening to. In this phenomenon we find that perception is surprisingly invariable. For instance, the F1 and F2 of /i/ are very different when it is produced by a child than when it is produced by an adult male. In spite of that, people have no problem recognizing both as /i/ and correct for the vast differences between acoustic properties, probably with the help of other sounds (i.e. contrasts) in the speaker's sound system.

All in all, there is evidence suggesting that representations are somewhat flexible in terms of what realizations are tolerated. Knowing that there is a contrast to look out for along with distributional information can alleviate a lot of difficulties, especially in perception and sound-identification.

2.2.2 Maintain categories

As we have seen, **Maintain contrasts** restrictions are specific to a given contrast and make reference to oppositions in the language. Thus, the way contrasts are maintained is only through a direct pressure, which explicitly mandates the maintenance of contrasts. There is an alternate possibility, that contrast preservation is also facilitated by a strict adherence to the phonetic realization of each category. If each sound category is realized relatively consistently (i.e. has the same phonetic specifications over time), preexisting contrasts are also necessarily maintained in the system without anything in the grammar explicitly mandating that. I will call the hypothesis under which the grammar imposes such a restriction on categories **Maintain categories**.

Under **Maintain categories**, sound categories have to stay consistent over time, and therefore the speaker cannot endorse tokens that based on their phonetic form would be categorized as another

segment. To be precise, a token of a category cannot be of such a phonetic form that it would be more likely to belong to a different category than its intended one.**Maintain categories** is therefore a stricter requirement than Maintain contrasts. In addition to the aforementioned four ways Maintain contrasts limits realizations of categories, Maintain categories also provides a more concrete set of expectations about the phonetic details of a representative of each category.

Those expectations are in terms of categorization (tokens cannot be more phonetically typical of a category other than their intended one), and can therefore be half-bounded. This means that some restrictions under the hypothesis only provide a mandatory minimum (lower bound) or maximum value (upper bound) for certain phonetic dimensions and restrict the distribution on one side only. For instance, a restriction consistent with **Maintain categories** would be one that mandates that an English word-initial /p/ must have a VOT of at least 60 ms on average (again, 60 ms is somewhat arbitrary here).

While **Maintain categories** is less permissive in terms of alternate realizations of a contrast than **Maintain contrasts** is, changes that do not involve adjustments in categorization are still allowed. A **Maintain categories** restriction for /p/, for instance, accepts distributions where /p/'s VOT is *longer* than it is in English ususally, but does not permit a deviation from the English sound categories downwards (towards shorter VOT). This can explain why native English speakers converge with /p/ targets with extremely long VOT (e.g. Shockley et al., 2004; Nielsen, 2011), whereas /p/'s with a shortened VOT make an invalid target. This is because the interval of acceptable /p/ VOT's is left-bounded, and /p/ tokens with a shorter lag than 60 ms cannot be admitted because they are threatening the boundaries of the /p/ category.

The difference between the two hypotheses can also be expressed in terms of how much of their own sound categories a listener brings to a conversation. Under **Maintain contrasts**, this can be just the basics. The listener brings a knowledge of oppositions in the sound system of the given language, including what cues are used to distinguish members of each contrast in each

position and the direction and required magnitude of difference, but in every other sense they are willing to adjust to however the speaker realizes the contrasts in terms of phonetic detail. Under **Maintain categories**, a listener brings a more concrete set of expectations. They seek to use their own categorization algorithms on the input they get from their interlocutor. The speaker's categories have to conform to the listener's idea of what a certain sound category looks like in order for the listener to be able to accommodate them.

It is important to note, that **Maintain categories** in and of itself does not predict that the process of *categorization* will not change. These predictions only limit whether a perceived token can then be used by the listener as a target for their own production during imitation. Listeners can and do adjust to their speakers. None of these predictions necessarily apply for non-lexical imitation (e.g. imitation of syllables or nonce-words) either.

Since **Maintain categories** does not make direct reference to oppositions in a sound system, it seems at first as though it cannot account for the gradient effect functional load has on contrast preservation, since functional load is a contrast-specific measure. It would in theory be possible to incorporate gradience for each restriction, which could speak to the importance of the contrast the given restriction protects. For instance, the relative importance of a word-initial /p/'s VOT not going below 60 ms might be different from the importance of a word-initial /t/'s VOT not going below 70 ms. However, the predictions made by **Maintain categories** alone and those made by a combination of the two hypotheses do not differ for the experiments outlined in this work. Therefore, the main question will be whether the restrictions responsible for contrast preservation reference phonetic details of a given category or not—i.e. whether there is any evidence for **Maintain categories**. Whether it works alone or in combination with the more abstract **Maintain contrasts** restriction is outside of the scope of present research.

The evidence for a more phonetically explicit category-level pressure like **Maintain categories** comes from the fact that while distributional information is sufficient for learning two *new*

categories, we also have evidence for pre-existing categories impacting production. Examples for this come from studies where while certain categories are imitated, some of the fine-grained distinctions are lost in the process. Studies from both English and Dutch show that, for instance, while the imitation of prevoicing is possible in both languages, fine-grained differences in the amount of prevoicing are not and are collapsed to a speaker-specific value.

This line of work is reinforced by the Perceptual Magnet Theory. This theory, whose consequences have been demonstrated for both adults and 6-month-old infants (Kuhl, 1991; Kuhl and Iverson, 1995), argues that discrimination is worse between tokens which are near particularly "good" (typical) instances of their category than it is between comparable tokens which are further away from "good" realizations . This indicates that while representations can be quite flexible, as long as distributional and lexical information is available, categories that are already formed do leave a mark on the way adults and even infants interact with subsequent input. Moreover, Perceptual Magnet effects also suggest that "goodness" of a token for its intended category is an especially important part of that, which supports **Maintain categories**.

2.3 Voicing contrasts and accommodation

In this dissertation I am going to investigate whether there is more support for **Maintain Contrasts** or **Maintain Categories** as a force limiting accommodation, and I am going to do it through VOT manipulation in voicing contrasts. The reason VOT was selected as a case study is twofold. First, as we will see the word-initial voiced-voiceless contrast (e.g. /p b/) can be expressed in two different ways cross-linguistically, and these two ways arguably use the same cue (VOT), but in different ways. Therefore, "sliding" the voicing contrast along VOT is an appropriate test case for seeing how flexible the representations of speakers are. In the first half of this section I will provide a brief background on voicing contrasts cross-linguistically (*Section 2.3.1*), and present the literature that is available on how this contrast manifests in the two languages this study focuses on, namely

English (*Section* 2.3.2) and Hungarian (*Section* 2.3.3). The second reason for choosing VOT is that, as we will see, there has been extensive work on VOT accommodation (at least in English), which allows us to situate the findings in a broader context. This context will be reviewed later on in this section for English and Hungarian separately.

2.3.1 Voicing contrasts crosslinguistically

Most of the world's languages have multiple homorganic stops, distinguished by laryngeal cues. A lot of these systems have a two-way voicing contrast, which is most commonly described as a voiced vs. voiceless opposition. Word-initially, this contrast is most often expressed in one of two ways using the relative timing of voice onset and the burst of the closure (Lisker and Abramson, 1964). The first type contrasts a plain stop (short-lag, voicing starts shortly after burst) with an aspirated one (long-lag, voicing starts long after burst). The languages, which use this contrast (e.g. German, Cantonese, Mandarin, and English) are called aspirating languages.[2] The other type of two-way voicing contrast is a prevoicing or "true voicing" contrast, where voiced stops are prevoiced (have voicing during closure), and voiceless stops are plain (short-lag). Prevoicing languages include Dutch, Polish, Spanish, Tamil, and Hungarian.

It must be noted that most phonetic work has made a crucial assumption: namely, that prevoicing and aspiration occupy two opposite (but more or less equivalent) extremes of the same phonetic spectrum. This spectrum is referred to as Voice Onset Time or VOT (e.g. Lisker and Abramson, 1964; Abramson and Whalen, 2017; Kharlamov, 2018; Kim et al., 2018; Seyfarth and

[2]Aspirating languages can differ in terms of the amount of voicing lag on "long-lag" (aspirated) stops, which could complicate this typology (Docherty, 1992; Cho and Ladefoged, 1999). However, this is not a crucial distinction for the purposes of the present work.

Garellek, 2018; Cho et al., 2019). Under this view, prevoicing is simply "negative VOT".[3] Since in case of prevoicing, voicing precedes the burst of the closure itself, the onset of voicing is in the negative range, if the (start of the) burst is taken to be the 0 ms time point. If we subscribe to this view, we could say that both types of two-way voicing contrasts (aspirating and prevoicing) utilize the same cue (VOT) to distinguish between voiced and voiceless stops.

However, this view could be contested from the point of view of production (Davidson, 2016, 2017). Articulatorily speaking, prevoicing and aspiration require different mechanisms. Prevoicing is a cue that requires constant flow of air during closure (in order to maintain the vibration of the vocal folds), and therefore the pressure in the supra-glottal tract and cavities must be lower than the sub-glottal pressure. Simultaneously, this air flow itself raises the air pressure in the oral cavity, since in stops the oral cavity is closed, and the velum is raised, blocking off the nasal cavity and thus air cannot escape. Therefore the potential duration of phonation during closure is aerodynamically restricted. In order to circumvent this, languages which have post-pausal phonation in voiced stops (i.e. prevoicing languages) must develop various strategies to sustain voicing (for examples, see Ohala, 2011; Solé, 2018, among others). This is in contrast with producing a long-lag stop, whose duration is not aerodynamically limited. Some claim that the relative difficulty of initiating and maintaining voicing after a pause is responsible for prevoiced–plain voicing contrasts and voiced stops as such being more marked cross-linguistically (Ohala, 1983). The putative difficulty of maintaining long instances prevoicing could impact participants' ability to converge with prevoiced stimuli in the present study above a certain point, which we would not see when participants have to converge with shorter cues. This can be explored in more detail in this study.

[3]Though there is some recent work which uses the term VOT to only refer to aspiration (positive VOT Grácz, 2011), and yet others reserve the term "negative VOT" for cases if and only if voicing during closure lasts for more than 50% of the total closure duration (Abramson and Whalen, 2017).

Since prevoicing and aspiration are very different cues acoustically as well (prevoicing is low-frequency periodic sound, aspiration is characteristically aperiodic noise), it is plausible that prevoicing and aspiration are not equally perceptually salient either. Research suggest that the distinction between short-lag and long-lag VOT is more psycho-acoustically natural than the short-lag vs. prevoiced distinction. One line of such research relies on VOT discrimination tasks carried with non-human animals. For instance, chinchillas not only seem to have a very English-like VOT boundary, but their VOT boundary shows English-like place effects as well (Kuhl and Miller, 1978). Similar studies have been carried out with rhesus monkeys, budgerigars, and Japanese quails (see Rojczyk, 2011, for a review). Another line of work investigates human language acquisition. Infants up until 6 months of age distinguish between 3 categories along VOT, and draw the two boundaries around -30 ms and +30 ms of VOT (Lasky et al., 1975). While aspirating languages have a boundary that closely resembles one of these (the positive boundary at +30 ms), neither of these boundaries is identical to the VOT boundary at 0 ms that is often cited for prevoicing languages. Therefore, infants learning prevoicing languages cannot take advantage of their natural predispositions, must suppress the two "natural" boundaries, and learn a third language-specific one. This lead some to conclude that a contrast distinguishing between short-lag and long-lag stops is more psychoacoustically natural than a contrast between a short-lag stop and a prevoiced stop (Serniclaes, 2005, inter alia).

Such an asymmetry could conceivably lead to prevoicing and aspiration not being equally easy to adopt and abandon for adult speakers. While this study does not directly assess the relative ease of *detecting* prevoicing and aspiration from a nonnative perspective, it could contribute to the issue of how easy they are (relative to each other) to pivot to as a contrastive cue for voicing. While keeping the controversies in mind, for the sake of convenience, I will use VOT to refer to a putative single continuum ranging from prevoicing ("negative VOT") through aspiration (positive VOT),

unless stated otherwise. At the same time, the dissertation itself will also discuss the tenability of this single continuum in light of the results of the two experiments.

The phonetic phenomenon of a plain stop (a short lag stop with e.g. 15 ms VOT) could be part of both aspirating and prevoicing systems. As Lisker and Abramson note, short-lag stops in English, which are interpreted as a token of a voiced category, are phonetically "not unlike" short-lag stops in other languages, where they represent a voiceless sound (Lisker and Abramson, 1967:7). However, they play a different role in the two types of voicing contrasts. In aspirating languages, they contrast with an aspirated stop, e.g. in English a plain labial stop is phonemically voiced, a /b/, and contrasts with an aspirated stop, a voiceless /p/. In prevoicing languages, a phonetically similar plain stop is phonemically voiceless and contrasts with a voiced stop. For instance, in Hungarian a plain labial stop is categorized as a voiceless /p/, contrasting with a prevoiced /b/.

The two languages this study focuses use these two kinds of voicing contrasts. English is an aspirating language, while Hungarian is a prevoicing language. In the following I will review the literature available on both of these languages. Since the experiments in this study involve labial stops word initially in a read environment, I will pay special interest to data on /p/ /b/, in reading tasks, and word-initially whenever possible.

2.3.2 The voicing contrast of English(es)

Since one of the languages in this study is English, in this section I am going to review available literature on English VOT and prevoicing in more detail. I will first discuss English VOT production along with the variation we observe across Englishes. Then I will give a brief review of the interaction of VOT and prosodic prominence—i.e. how word- and syllable-initial VOT varies based on the stress and accent. Then I will discuss the perceptual distinction between English voiced and voiceless stops. Finally, I will discuss English VOT accommodation, which has mainly focused on voiceless stops, and finally evidence of English speakers being exposed to other types of

voicing contrasts in a laboratory environment. Since most of this research is on English, the results of studies from these last two subsections are especially informative to the present experiments.

English VOT / prevoicing production

English is an aspirating language, and word-initially the voicing contrast in "mainstream" (prestigious) dialects in the US is mostly expressed as plain vs. aspirated. Even the earliest accounts pointed that voiced stops in American English show a bimodal distribution: they are a mixture of short-lag (mode around 0 ms) and prevoiced stops (mode around −100 ms Lisker and Abramson, 1967). Prevoicing is regularly found in postvocalic environments (Westbury and Keating, 1986), and some studies found that over half of word-initial post-pausal voiced stops had some prevoicing (e.g. Smith, 1978; Flege, 1982), which is the context we will focus on in this dissertation. However, contradictory evidence surfaced since then (Keating, 1984; Ball and Rahilly, 2014; Carford, 2001; Cruttenden, 2013; Davenport and Hannahs, 2010), which might indicate a difference in methodology, a change in progress, or both. Most recently only a quarter (Davidson, 2016) or as little as 10% (Davidson, 2017) of phrase-initial voiced stops were found to have at least partial phonation. Among word-initial voiceless stops in English, /p/ has the shortest VOT, but it is still within the usually understood range of long-lag (over 40 ms). Studies of read speech typically find VOT values for /p/ between 60 ms and 110 ms (e.g. Allen et al., 2003), but the range can be broader depending on the study (e.g. Chodroff and Wilson, 2017, found a range of 46–139 ms).

There are also some other dialects of English that tend to have more prevoicing in voiced stops. For instance, more prevoicing can be seen in some regional dialects spoken in the American South (Jacewicz et al., 2009), African American Language (AAL, Ryalls et al., 1997, 2004) Standard Southern British English (Docherty, 1992), and Shetlandian English (Scobbie, 2006). These dialects also tend to have less aspiration in voiceless stops, which could suggest that the two cues (prevoicing and aspiration) are in some sense tied to one-another, possibly even occupy two

ends of the same VOT continuum. However, the relationship between the two cues could be more complex as well. For instance, it could be possible that this is language- or system-dependent. It is possible that while speakers of some aspirating languages might not perceive prevoicing to be part of the VOT continuum, speakers of some prevoicing languages do view aspiration and prevoicing as two extremes of the same cue continuum. I know of no study that systematically compared the treatment of prevoicing and aspiration as a non-native cue to a voicing contrast, but the two experiments carried out in this work could be able to uncover such putative differences.

The voicing contrasts of Englishes have also changed over the years, which indicates a certain amount of flexibility in the phonetic detail of their realization. These changes paint an asymmetric picture. While multiple dialects of English have been shown to abandon prevoicing in phrase-initial position and rely more heavily on aspiration as the main cue of the voicing contrast, there are very few instances of the opposite. A change has been documented for Australian English (Millasseau et al., 2019), Scottish English (Masuya, 1997; Docherty et al., 2011), and Scots (Johnston, 1997; Stuart-Smith et al., 2015) among others, where these dialects are using less and less prevoicing on voiced stops (and more aspiration on voiceless stops). The only known example of the opposite is a trend reversal in teenage Glaswegian Scottish English speakers, who are returning to the more prevoiced voiced stops seen in their (great-)grandparents' generation (Stuart-Smith et al., 2015; Sonderegger et al., 2020). While social effects like prestige definitely play a role in these changes, the fact that more dialects are changing *away from* prevoicing rather than towards it might indicate an asymmetry between prevoicing and aspirating systems.

English VOT /prevoicing and prosodic prominence

VOT not only serves the purpose of distinguishing between /p/ and /b/ in English, but it can also carry emphasis, as it is influenced by both lexical stress and phrasal accent. Since this work manipulates VOT directly, we need to understand what supra-segmental meanings certain VOT

values might carry. The relationship between VOT and prominence is quite clear for voiceless stops. English voiceless stops /p t k/ have longer VOT in stressed syllables (Gimson, 1962; Lisker and Abramson, 1967, and on), and in phrase accented syllables (Cole et al., 2003). The relationship between VOT and prosodic prominence is less clear for voiced stops. Earlier studies could not pin down the exact effect of prominence on voiced stops (Lisker and Abramson, 1967), and some later studies found that in read "Radio News speech" found that the effect of prominence is dependent on the given voiced stop. Cole et al. (2003) found that a stressed /ɡ/ had a shorter lag, but stressed /b/ and /d/ had longer lag (more positive VOT) than their unstressed counterparts.

Yet other studies found no difference between stressed and unstressed English voiced stops in terms of percentage of closure voiced (Jacewicz et al., 2009) nor in terms of prevoicing duration (Simonet et al., 2014). These two studies are of particular interest, because their findings only held for a subset of their participants. Jacewicz et al. (2009) found that the effect of prominence on voiced stops is dialect-dependent. In their study they compared the percentage of voicing during closure of postvocalic, word-final voiced stops read out in the form of a minimal-pair word list by speakers from North Carolina and Wisconsin. North Carolinians produced voiced stops with largely voiced closures, and the proportion of voicing during closure did not change between low-, mid- and high-emphasis contexts. Since high-emphasis contexts also tend to lengthen closure duration, the fact that the percentage of voicing during closure remained relatively stable means that participants in this group likely even lengthened the raw duration of voicing during closure in high-emphasis contexts. At the same time, Wisconsinites had at most 67% voicing during closure, and this number decreased with the lengthening of the closure duration itself, indicating that the duration of voicing was consistent or decreased (it just took up less and less of the closure duration). This means that the Wisconsinite group's stops were "plainer" (or at least no more prevoiced) when emphasized.

In Simonet et al. (2014), they measured the VOT of lexically stressed and unstressed word-initial stops, and compared productions from two groups: monolingual English speakers

and bilingual speakers who were also "proficient" in Spanish. The monolingual speakers mostly produced plain /d/'s (21.9 ms in stressed vs. 26.3 ms in unstressed position), and while there was "at most [...] a marginal trend" towards shorter lag stops in stressed position, this was not significant. At the same time, bilingual speakers, who produced more prevoiced /d/ tokens to begin with, produced most of this prevoicing in *unstressed* positions, whereas their stressed productions were significantly plainer. This pattern is somewhat similar to the pattern Jacewicz et al. (2009) found for Wisconsinites, where unemphasized productions could be somewhat prevoiced, but emphasized productions were closer to the "canonical" plain voiced stop values of English.

What we can take away from these studies is that while the VOT of English voiceless stops is always lengthened by prosodic prominence, the effect of prominence on voiced stops varies based on the speaker's baseline. For speakers of certain (usually less prevoiced) dialects, prominence has a subtractive effect: the distribution of voiced stop productions seems to become more concentrated in the short-lag range when stressed or accented. For other speakers with more voiced baseline productions, prominence either does not change or even lengthens instances of prevoicing (depending on whether we measure voicing as raw VOT or percentage of voicing during closure).

English VOT /prevoicing perception

VOT plays an important role in not just the production, but the perception of the voicing contrast as well. While the labial /p b/ boundary has received little attention, we have some information on the boundary for voiced and voiceless alveolar stops (/t d/). In Mack's (1989) study of English monolinguals, who at the time were students at Brown University, she found that speakers categorically label a 0 ms VOT stop a voiced /d/, but at 30 ms VOT they almost exclusively label it a voiceless /t/, with a transition at the 10 ms and 20 ms steps, which are somewhat ambiguous. She also conducted a 20 ms difference detection task, where participants heard two steps of the

continuum 20 ms apart (e.g. a 50 ms VOT stop and a 70 ms VOT stop) and had to say if they thought the two tokens were different. The biggest chance of detecting the difference was when participants had to compare 10 ms VOT with 30 ms VOT and when comparing 0 ms VOT with 20 ms VOT. There was also a third (albeit smaller) peak when 20 ms VOT and 40 ms VOT were compared. Participants were only able to distinguish between 10 ms VOT and -10 ms VOT stimuli and between 0 ms and -20 ms VOT stimuli at chance. The results suggest that 20 ms VOT is indeed ambiguous for most speakers, and that the category boundary must be somewhere in that vicinity. More recent studies had similar results (Takahashi, 2020), indicating that the boundary has not subject to much change for American English-speakers. Other studies with a different methodology (a labeling task) found the boundary to be somewhat higher: a given individual's voiced-voiceless boundary was somewhere between 28 ms and 44 ms (Keating et al., 1981).

Here we must mention that fundamental frequency in the following vowel also cues voicing, but as a secondary cue (Lisker and Abramson, 1964; Shultz et al., 2012, inter alia). However, it co-varies with Voice Onset Time for most speakers (Chodroff and Wilson, 2017) and it affects perception mostly under adverse listening conditions. Under normal listening conditions, native English speakers rely solely on VOT, and only use information on onset f_0 mostly under noisy circumstances (Winn et al., 2013), when aspiration is especially hard to hear. In this dissertation f_0 will not be manipulated, but maintained as a constant across comparable tokens—i.e. multiple versions of the same word will always have the same f_0 information, and will only vary in terms of the duration of aspiration and prevoicing. This is largely based on the fact that listening conditions will be carefully controlled and therefore participants will likely not rely on it as much.

English VOT / prevoicing accommodation

Aside from maybe fundamental frequency, adjustments in Voice Onset Time have been the most commonly investigated acoustic measures for accommodation, and most of this research has been

done on English. As this is the measure the experiments in this dissertation also use, I am now going to describe the available literature on VOT accommodation in more detail.

Several studies have found that English speakers lengthen the VOT of their voiceless stops when they are exposed to stops with long VOT. This has been demonstrated with extremely long manipulated VOT's, e.g. around twice as long on average than the participants' baseline in Shockley et al., 2004, but also replicated with long VOT, derived from instructing the model talker to lengthen their VOT (but not manipulating the stimuli any further (Tobin, 2013, 2015 had a mean of 110.45 ms).

Shockley et al. (2004) used a shadowing task, where after a baseline reading task, participants listen to audio stimuli of individual words and have to "identify the word by saying it out loud" (following Goldinger, 1998). They found three things. First, participants converged with the model talker's naturally occurring VOT, which was longer than the participants' baseline. Second, they also converged with the model talkers when their VOT was artificially lengthened (doubled). Third, artificially extended VOT elicited more accommodation from the participants (stronger effects of perceived convergence from the listeners) than the model talker's naturally occurring VOT values, which suggests that the effects were not due to a task effect—e.g. participants producing longer VOT's in a shadowing task than in a reading task irrespective of the stimuli.

While Shockley et al. (2004) used an AXB task to measure accommodation, it has been found that the frequency with which third-party listeners detect accommodation in a shadowing task correlates with the amount of VOT accommodation for the given participant (Schertz et al., 2019). This correlation goes away when the VOT difference is neutralized, which indicates that the original correlation was not due to listeners picking up on another phonetic factor whose accommodation also happened to correlate with VOT accommodation. This not only suggests that VOT accommodation is important for the perception of accommodation overall but also indicates that the results from Shockley et al.'s AXB measure are also applicable to VOT duration.

Nielsen and colleagues also conducted experiments with lengthened VOT exposure, but measured convergence as a change (increase) in VOT from pre- to post-exposure. In these studies, participants read out a list of words both before and after being exposed to manipulated stimuli (words with either lengthened or shortened VOT). Exposure, however, was somewhat different than in Shockley et al. (2004). Exposure in these studies involved only a listening task, where participants did not need to produce anything while listening to words from the model talker.

In these exposure-words Nielsen and colleagues manually manipulated the VOT of a word-initial /p/ to be at least 100 ms or were shortened by exactly 40 ms, which resulted in a mean VOT of 30.36 ms,SD=8.95 ms. She found that participants, who were exposed to the lengthened listening stimuli produced higher post-exposure read values than their baselines recorded before exposure were (Nielsen, 2008, 2011, replicated by Mielke et al., 2013). In their study the reading and the listening sets did not match, which also allowed them to demonstrate that participants generalize the exposure patterns to novel words and also to a novel segment from the same natural class (/k/).

In the shortened VOT condition (only in Nielsen, 2008, 2011), there was no effect of exposure (i.e. pre- and post-exposure reading values did not differ). This suggests that this VOT accommodation might be unidirectional—i.e. speakers are only willing to *lengthen* but not shorten the VOT of their voiceless stops as a reaction to stimuli. She argues that the lack of convergence towards a shorter VOT might be due to the fact that shortening the VOT of an aspirated stop brings tokens closer to the categorical boundary between /p/ and /b/, thereby neutralizing or at least meddling with a phonologically meaningful distinction. Indeed, her participants were never exposed to /b/ words, manipulated or otherwise, which means participants had no way of knowing if the /p b/ contrast was maintained by the model talker.

While Nielsen (2008, 2011) used 30 fillers in her reading task, they were outnumbered by her 120 target words (100 /p/-initial and 20 /k/-initial words in order to test generalizability). Since

the filler-to-target ratio was quite skewed towards /p/-initial filler words, the inclusion of fillers likely did not mask the focus of the experiment (word-initial /p/'s).

While Nielsen's studies find no effect of a shortened /p/ VOT, other, albeit somewhat different, studies do find effects of shortened VOT. For instance, in a priming study, Levi (2015) finds that participants are able to prime themselves by exploiting the natural VOT ranking in English ($VOT_{/p/} < VOT_{/t/} < VOT_{/k/}$). When participants had to read a /t/-word after a /p/-word (*pan*), they produced shorter VOT's than when a /t/-word followed a /k/-word (*keen*) or a neutral prime (e.g. *main*). Other participants had to repeat words after a model talker, whose /p/ and /k/ were manipulated to show an unnatural pattern (/p/ had a longer VOT than /k/). They also produced their /t/ with shorter VOT when the /t/-word followed a /k/-word with an unnaturally short VOT as compared to the /t/ word following a lengthened /p/ word or a neutral prime. In a continuum imitation task, native English speaking participants converge with short-VOT (prevoiced) stimuli produced by a native speaker of Thai (Olmstead et al., 2013). In his semester-long language pedagogical study Nagle (2019) also finds evidence of native English speakers learning to imitate shorter VOT voiceless stops (but only in imitation and not in a picture naming task).

These studies indicate that accommodation to shorter VOT (moving closer to a categorical boundary) is possible, but it is questionable how well these findings can be extrapolated to speakers accommodating to a truly plain vs. prevoiced contrast. For instance, the study of Levi (2015) demonstrates shortened VOT, but the shortened /k/ primes are still in the aspirated range (even if their VOT is short for a /k/). Olmstead et al. (2013) and Nagle (2019) demonstrate convergence to either prevoiced stimuli or a prevoicing contrast, but their stimuli are outside of a lexical environment—i.e. they use nonce syllables or words of a foreign language. There fore, these studies suggest that English speakers can prevoice and learn a *new* prevoicing contrast, but they can only be seen as suggestive of accommodation to shorter VOT being possible in an English lexical environment.

Since they *are* indicative of VOT shortening being possible in the process of imitations, a more direct investigation of Nielsen's null results is in order. It is possible that Nielsen's null results were at least in part due to the lack of /b/-words during exposure. Without /b/-initial words, participants had no way of knowing whether the contrast was in fact preserved even under the peculiar speech patterns of the model talker. The current study expands on Nielsen's work by including /b/-initial words as well, thereby demonstrating to participants that the /p b/ contrast is not neutralized in the model talker's speech.

Much of the design is borrowed from Shockley et al. (2004). I will also use a shadowing design, but expanding on Nielsen's idea of comparing the effects of exposure to lengthened vs. shortened /p/ VOT, I will not only include /p/-initial words, but /b/-initial words as well in order to demonstrate that the contrast is preserved in each case. The use of a mixture of /p/-words and /b/-words exposes participants to not only a certain realization of a sound category but to a certain realization of the /p b/ contrast. In this study, I will not use fillers, following Shockley et al.'s design, partially as a trade-off to keep the duration of the experiment in a feasible range (30–40 minutes). The exact conditions of the experiment will be described in the next section (*Section 2.4*) in detail.

I will also slightly alter the traditional shadowing design (used in Goldinger, 1998; Shockley et al., 2004; Babel et al., 2014, and others). Borrowing from Nielsen and colleagues, I will add a post-exposure reading block on top of the pre-exposure reading and the shadowing task. A pre-exposure reading task can be much more reliably compared with a post-exposure reading task, where the task is the same, than with a shadowing task, where results can easily be impacted by task effects. By a meta-comparison of *patterns* in accommodation behavior during the reading tasks with *patterns* during the shadowing task (rather than raw values), we can not only control for task effects within the tasks but isolate these potential task effects as well. Lastly, while Shockley et al. (2004) used an AXB task with third-party listeners to assess convergence, like Nielsen (2008) and others, I use VOT duration in order to measure convergence acoustically.

In the next subsection I will review studies which, rather than exposing participants to a single sound category, expose participants to a voicing contrast. This is of particular interest, this study will expose some participants to foreign-sounding contrasts, which might come with complications.

Laboratory exposure of English speakers to different voicing systems

The experiments carried out in the dissertation involve some of the participants having to shadow a model talker with a prevoicing voicing contrast, which is quite unlike the typical plain vs. aspirated contrast in English. In this subsection I will review previous results on English speakers being exposed to challenging types of voicing contrasts outside of accommodation studies.

One particular challenge that participants might face with my stimuli is hearing prevoicing to begin with, since if the participant cannot tell prevoiced stops apart from plain ones, they might conclude that the prevoicing model talker does not maintain the voicing contrast at all. However, previous research suggests that English speakers can be trained to distinguish between plain and prevoiced stops, and can not only hear prevoicing as a cue, but to some extent they can even produce it. Even though prevoicing is a sub-phonemic cue in English—i.e. the presence or absence of prevoicing in and of itself does not distinguish between different phonemes—it has been shown that monolingual native speakers of English can be trained to distinguish between plain and prevoiced stops (Baese-Berk, 2010a,b). Moreover, when perceptual training is supplemented with production training, English speakers are also able to produce the distinction themselves (Baese-Berk, 2019).[4]

[4]While production training seems to interfere with the success of perceptual learning (Baese-Berk, 2010a), this effect goes away when the speaker participants in three training sessions (Baese-Berk, 2019) instead of two (Baese-Berk, 2010a). This effect of mixed-modality interference is not specific to perceptual and production trainings, it has

Some research on L2 pronunciation shows that native English speakers learning Spanish are able to lower their VOT of voiceless stops in order to produce a more plain (and thus more native Spanish-like) target (Hutchinson and Dmitrieva, 2019; Schuhmann and Huffman, 2019), but they find differing results for voiced stops. Hutchinson and Dmitrieva (2019) found that when taking a semester of Spanish, English speakers can learn to imitate a prevoiced—short lag contrast successfully. However, the success rate of English speakers producing prevoicing word-initially in a foreign language is tied to their tendency to prevoice in English word-initially (Hutchinson and Dmitrieva, 2019). Schuhmann and Huffman (2019), on the other hand found no significant effect of lower VOT (more prevoicing), and they argue that this suggests an articulatory difficulty of producing (post-pausal) prevoicing.

However, these studies were using either only nonce-words (or syllables) or words from a foreign language, and therefore they demonstrate the learnability of a *new* contrast. Having to adjust one's pre-existing English contrast to a new type of pattern can pose a different challenge altogether. While there are no studies investigating this with a plain vs. prevoiced contrast, studies with ambiguous stimuli indicate that lexical information facilitates rather than hinders perceptual learning. Kraljic and Samuel (2006) show that when lexical information is available, English speakers can be trained to perceive a smaller contrast between /t/ and an alveolar stop with ambiguous voicing (/?td/) as well as between /d/ and the ambiguous stop (/?td/) in a word-medial position. Neither of these happen when the training items are nonce-words.

Kraljic and Samuel (2006) exposed monolingual native speakers to multi-syllabic words which contained either a single /t/ or a /d/ in word-medial position. They created an ambiguous stops (/?td/) from a "mixture" of /t/ and /d/ in terms of amplitude of the burst, VOT and closure duration. In one condition, all /t/'s were replaced with the ambiguous /?td/ (?t condition), and in the other, all

also been described for articulatory and visual feedback interfering with each others' success rate in speech therapy (Lin et al., 2019).

/d/'s were replaced with /ʔtd/ (ʔd condition). Participants then had to complete a lexical decision task with the stimuli of their condition (as well as with other filler words and nonce-words). In two control conditions /ʔtd/ also replaced /t/'s and /d/'s respectively, but in nonce-words. Participants successfully adapted to the modified contrasts in the test conditions, where lexical information was available—participants had an accuracy rate of 98.1% in the ʔt condition and 84.6% in the ʔd condition. No perceptual learning occurred in the control conditions with nonce-words.

Curiously enough, participants even generalized this pattern to /p/'s and /b/'s, which they were not exposed to in the training phase, even though the data contained multiple /k/'s (e.g. *crocodile* /ˈkrɒkədaɪl/, *kingdom* /ˈkɪŋdəm/) that were never manipulated. In a subsequent labeling task, ambiguous stimuli were categorized by the two test conditions differently, whereas the two control groups did not differ. This study indicates that English speakers can perceptually adjust to changes in the VOT values of stops (at least when the adjustments stop at an ambiguous level). It is yet to be tested if this perceptual learning also translates into imitation of said stimuli, but it is certainly a promising result for the purposes of the current study. Moreover, it also shows the importance of lexical information. Therefore, in the current experiments I will only use words which do not form part of a /p b/ minimal pair.

2.3.3 The voicing contrast of Hungarian

Compared to English, the phonetic parameters of the Hungarian voicing contrast have been documented to a much lesser extent. In the following few subsections I am going to summarize what is available with respect to VOT production and VOT's relationship to prosodic prominence. Since I could not find information on the perception of VOT or prevoicing production, I will review literature on other prevoicing languages and extrapolate from them. Finally, I will discuss findings from the only VOT accommodation study from Hungarian (impressionistic descriptions from hours of a sociological interview).

Hungarian VOT / prevoicing production

In the earliest available description on Hungarian VOT production, Lisker and Abramson (1964) describe Hungarian word-initial /p/'s to have 0–10 ms VOT values, and /b/'s to have 65—125 ms of prevoicing (VOT of −125 to −65ms). This has largely been corroborated by more recent word-initial and word-medial measurements (Gósy, 2001; Gósy and Ringen, 2009; Gráczi, 2011). Gósy (2001) finds that the average VOT of /p/'s in a word list reading task is 24.64 ms (range of 13.2–34.8 ms). While all /p/'s in the word list were word-medial, she finds similar read values for CV syllables in isolation (24.05 ms mean, range of 14.0–38.4 ms).

Later studies found somewhat shorter VOT's. In Gósy and Ringen (2009), word-initial /p/'s had 9.7 ms of VOT (30.8 ms wide range) and /b/'s had −94.6 ms of VOT on average, and while the range for the latter was very wide (178.4 ms), all recorded tokens were prevoiced. Participants in this study were instructed to speak "carefully". The fact that their /p/'s became "plainer" under this condition echoes Jacewicz et al.'s finding that Wisconsinites, who mostly produced plain but on occasion also prevoiced /b/'s, produced exclusively plain /b/'s when asked to emphasize words. Gósy and Ringen were the only ones who looked for potential gender differences, but they found no systematic difference between males and females' VOT production.

The most recent study (Gráczi, 2011) found shorter cues overall (less aspiration, less prevoicing), albeit her words were embedded in a carrier sentence (they were still sentence-initial). She measured a range of −40 to −125 ms for prevoiced /b/'s (median around −75 ms). Only 4 of them stopped phonation during closure or did not prevoice at all, these tokens had a VOT of about 13 ms. In comparison, /p/'s had a VOT of 15±7 ms.

Hungarian VOT / prevoicing and prosodic prominence

In contrast to English, much less information is available on the effects of prosodic prominence on the realization of the voicing contrast in Hungarian stops. However, based on data from other

prevoicing languages, such as Spanish and Dutch, we have reason to suspect that VOT (and especially prevoicing) might vary with emphasis. In their study mentioned in the English section, Simonet et al. (2014) also recorded word-initial stressed and unstressed stops from monolingual Spanish (in Spanish) speakers as well as Spanish productions from their bilingual speakers. While in English they found that speakers lengthened the amount of aspiration on their /t/'s and bilingual speakers even produced more prevoicing in their /d/'s (and monolinguals did not change their /d/'s), they found a different pattern for Spanish. In Spanish, prominence affected /d/'s rather than /t/'s: Spanish /d/'s had longer prevoicing when lexically stressed. This effect was the same for monolingual and bilingual participants.

These findings are in line with previous research on Spanish, referenced in Simonet et al. (2014). A similar effect of lexical stress and phrasal accent was found for intervocalic stops in Dutch (Cho and McQueen, 2005). Dutch /d/'s have more prevoicing when stressed or accented, but unlike in Spanish where /t/ is unaffected by stress, the VOT of Dutch /t/'s shortens when in a prominent syllable. The difference between the two studies' results for /t/ could be idiosyncratic or could be an effect of environment: the Spanish study that found no change looked at word-initial stops, whereas the Dutch study that found a shortening of short-lag VOT looked at intervocalic stops.

From these studies we can presume that Hungarian might behave similarly—i.e. that Hungarian voiced stops are more prevoiced when prosodically prominent, whereas plain voiceless stops are more concentrated and have a shorter lag VOT in similar positions. In conjunction with the English data on prominence reviewed in the previous subsection, research suggests that emphasis has an effect on VOT, in that it exaggerates the typical VOT properties of the given sound category. As such, this effect is dependent on what the category typically looks like. Prosodic prominence makes prevoiced categories more prevoiced, aspirated categories more aspirated and plain stops "plainer" (shorter lag).

Hungarian VOT / prevoicing perception

In terms of perception, there is again little to know about the Hungarian perceptual boundary, but we can extrapolate from results on other, similarly prevoicing languages. For instance, results from Keating et al. (1981) suggest that the boundary for Polish word-initial [ta-] and [da-] syllables is somewhere around 0 ms, specifically, an AX same-different task indicated that their participants' median boundaries were between -4 ms and 6 ms across the multiple experiments they report on. However, when they compared Polish listeners with English listeners in a labeling task, Polish listeners showed much more interpersonal variation in the location of their voiced-voiceless boundary than English speakers did. This was seen from a higher standard deviation and a wider range of boundaries—on a continuum from -100 to 50 ms VOT, estimates for any given Polish participant's boundary ranged from -20 ms to 29 ms (depending on the participant), whereas all English-speaking participants' boundaries were between 28 ms and 44 ms. These results are likely to be indicative of where the Hungarian VOT boundary might be as well, and therefore we can expect to see a less clear boundary for Hungarian speakers as a group than for English speakers, in a much lower range as well.

Hungarian VOT / prevoicing accommodation

The only mention of VOT in a Hungarian accommodation study was in Kontra and Gósy (1988). A series of interviews were recorded with a Hungarian-born man emigrating to the US 30 years beforehand. Based on impressionistic assessment of the recordings, Kontra and Gósy found that initially the speaker was aspirating his word-initial voiceless stops, especially in contexts where he was having issues with lexical retrieval. While this effect gradually decreased over the course of the conversation, it provides evidence of a shift in someone's cuing of a natural class over time. The speaker adapted aspiration as a cue to voiceless stops. It also demonstrates that aspiration is a learnable cue for native speakers of a voicing language (at least over the course of 30 years).

Unfortunately, the authors do not describe whether this was concurrent with less prevoicing in voiced stops. Therefore we cannot know if the speaker's contrast was shifted to a higher range of VOT (whether the speaker had plain voiced stops contrasting with aspirated voiceless stops) or whether the phonetic distance between the two contrasting natural classes was increased as a result of more prevoicing (prevoiced voiced stops contrasting with aspirated voiceless stops). What happens to the other half of the contrast when one half changes is something that is explicitly addressed through the present study.

2.4 Conditions and predictions

I carry out two experiments involving VOT accommodation to investigate the question of through what mechanism contrasts are maintained, i.e. how one's representations limit what types of input they can accommodate to. The design of the experiments builds on Nielsen's (2008, 2011) and Shockley et al.'s (2004) work, but the design has been modified in ways to test the two hypotheses outlined in *Section 2.2*, namely: **Maintain contrasts** and **Maintain categories**. **Maintain contrasts** is a more abstract pressure, which mandates the maintenance of a contrast, but does not restrict the exact realization of the contrastive categories, as long as a distinction is maintained along the same dimension(s). By contrast, **Maintain categories** is a more concrete pressure that mandates the adherence to the phonetic detail of each sound category.

The experiments comprise a shadowing task and a pre- and post-exposure reading task to measure immediate and short-term accommodation, respectively. These tasks are accompanied by a labeling task at the beginning and one at the end of the study to measure any potential shifts of perceptual boundaries that the exposure to shadowing stimuli might cause. This experiment is carried out in two languages: English and Hungarian. Aside from the linguistic background of the participants and the words used as stimuli, the two experiments are identical. The English experiment has self-identified monolingual English speakers as participants, listening to minimally

altered English words produced by a native English speaker. In the Hungarian experiment self-identified monolingual Hungarian native speakers listen to artificially manipulated Hungarian words produced by a native Hungarian speaker.

The way these experiments test the hypotheses discussed in *Section 2.2* is through exposing participants to /p/- and /b/-initial word stimuli with different VOT properties during the shadowing task. In both the English and the Hungarian experiment, the group of participants is split into two conditions that only differ in the voicing contrast they hear during shadowing. In one condition, participants shadow a model talker who uses an exaggerated version of an aspirating contrast for word-initial /p/'s and /b/'s (*Extreme Aspirating* condition). This is realized as a contrast of 130 ms VOT /p/'s and 15 ms VOT /b/'s—i.e. aspirated /p/ vs. plain /b/. In the other condition, participants shadow a model talker who expresses the /p b/ contrast by exaggerated prevoicing (*Extreme Prevoicing* condition). This is realized as a contrast of 15 ms VOT /p/'s and −130 ms VOT /b/'s—i.e. plain /p/ vs. prevoiced /b/.[5]

These two conditions are of course not equally "foreign" to the participants, and the English and Hungarian participants also differ in which distinction is more native-like to them (compared to their mother tongue, the language of the experiment). The following subsections will discuss how the interpretation of the same two conditions differs in the context of the two languages.

The two conditions in the context of the English experiment

Figure 2.1 shows how the stimuli of the two conditions compare to habitual English values in the English experiment. The gray ovals reflect that in English, the VOT of a word-initial /p/ is usually in the 46–139 ms range, and the VOT of a word-initial /b/ is 11–20 ms (Chodroff and Wilson, 2017). The black boxes represent the model talker's production values—*Extreme Prevoicing* (*Extr.*

[5]In the end, in the Hungarian experiment's *Extreme Aspirating* condition a 130 ms /p/ was contrasted with a 5 ms /b/, for reasons that will be discussed in more detail in *Chapter 4*.

Prev.) on top, *Extreme Aspirating* (*Extr. Asp.*) below. It can be seen that the *Extr. Asp.* stimuli is much closer acoustically to what the /p/ /b/ distinction in English usually looks like. The /b/ target is actually is within the typical range, and while the /p/ target is outside of the typical /p/ range, it is *more* aspirated rather than less, thereby exaggerating the contrast rather than threatening it. At the same time, the *Extr. Prev.* stimuli are very unlike English is typically. Compared to habitual VOT values, the /p/ target is shifted down, and it is actually within the range of typical English /b/. However, since the /b/ target is shifted even further down well into the prevoiced range, the contrast is preserved, and if anything the size of the difference is increased.

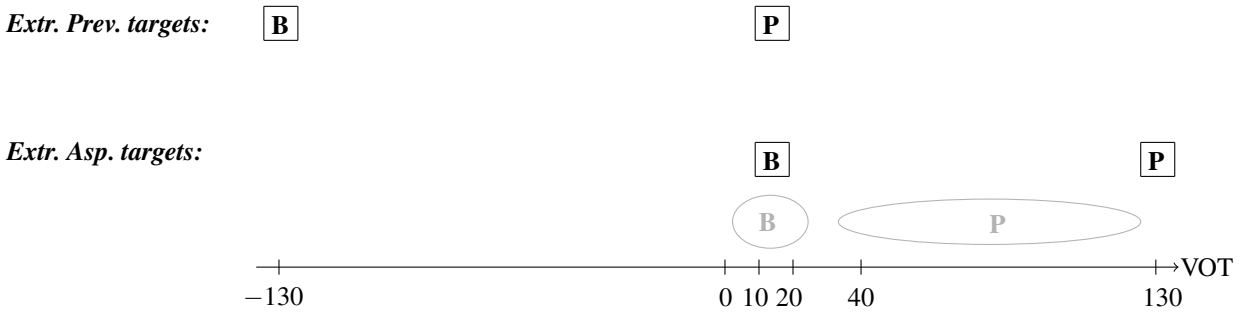

Figure 2.1: Conditions in the English experiment

With this setup, the two hypotheses make largely the same predictions. The predictions of the **Maintain contrasts** hypothesis are presented visually in *Figure 2.2* and then the predictions of **Maintain categories** are in *Figure 2.3*. First, I will demonstrate how the two make almost the same predictions, and then I will point out the locus of the difference, which will be the most crucial point for this study. The "✓" and "✗" signs above the target values in the black squares mark whether the given hypotheses predicts accommodation to the given target.

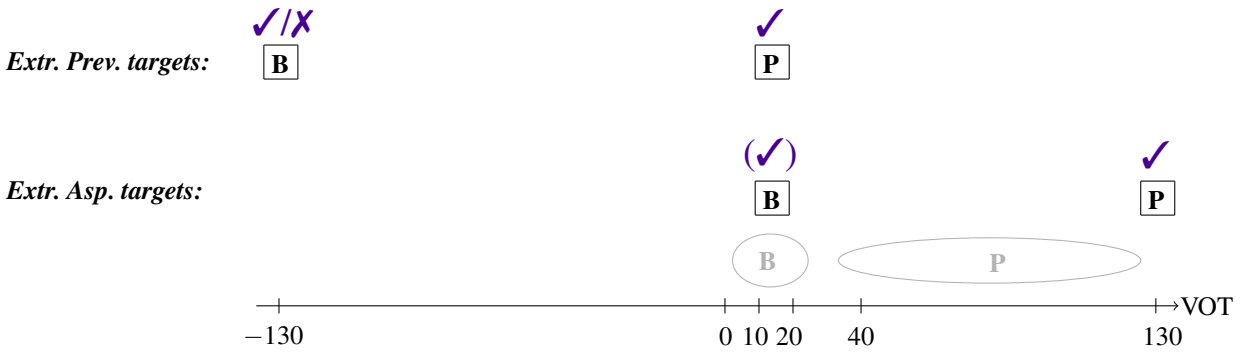

Figure 2.2: Predictions of the maintain contrasts hypothesis for English
Gray: typical English values; ✓: hypothesis predicts convergence; ✗: hypothesis predicts no convergence

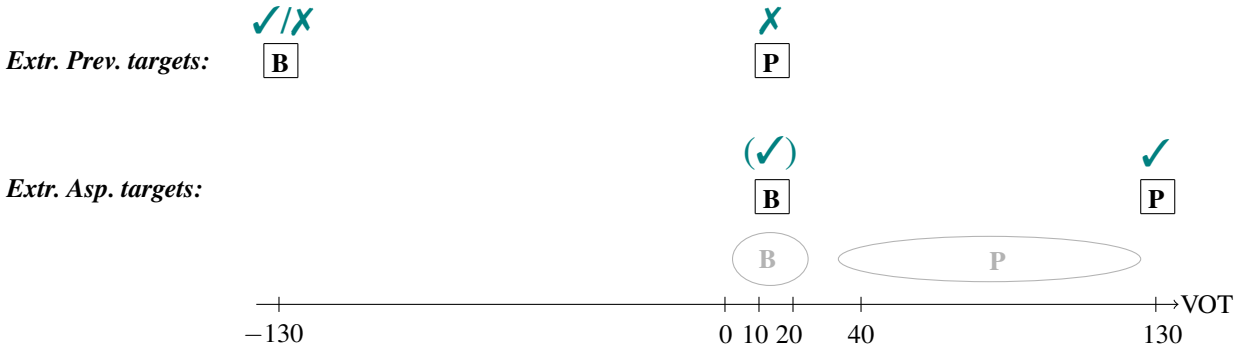

Figure 2.3: Predictions of the maintain categories hypothesis for English
Gray: typical English values; ✓: hypothesis predicts convergence; ✗: hypothesis predicts no convergence

In terms of the *Extr. Asp.* stimuli, the predictions do not differ. Both hypotheses predict that participants will be able to accommodate to an extremely aspirated /p/ (130 ms VOT target; marked by "✓") as well as to a plain /b/ (15 ms VOT target). The reason why this is marked by a "(✓)" rather than with a "✓" is because for most English speakers, a plain 15 ms VOT is a target that they might be matching to begin with, in which case no convergence is necessary or possible. Under **Maintain contrasts**, the two targets (extremely aspirated /p/, plain /b/) are a valid realization of the contrast. this is expected because the contrast is still maintained (with the same phonetic dimension, relative ranking of segments, magnitude and in the same environment)—in fact, the size of the distinction is even bigger than it typically is in English. **Maintain categories**-type restrictions also find both targets valid. In case of the plain /b/, that is somewhat trivial: it being an existing realization of a

sound category of the language, it must be a valid target for that same category's realization. While the aspirated /p/ deviates from typical English values, crucially, it deviates in a direction in which its distribution is not limited—i.e. towards longer VOT's.

The case of *Extr. Prev.* is of more interest, since, as discussed before, it is less like English, and therefore accommodating to these stimuli can be a greater challenge for English native speakers. Neither of the two hypotheses make any predictions about whether native English speakers will in actuality accommodate to a prevoiced /b/. From the perspective of **Maintain contrasts**, the contrast is adequately maintained, and therefore there is nothing wrong with a prevoiced /b/, which is a member of this realization of the contrast. From the perspective of **Maintain categories** restrictions, an extremely prevoiced /b/ is a perfectly adequate target as well, because the category of /b/'s only has an upper bound for VOT—i.e. more prevoicing is permitted. The reason that this is marked by "✓/✗" rather than by just "✓" is because the hypotheses only predict that a prevoiced /b/ will be a *valid* target. Whether it is accommodated to also hinges on whether it is also an *implementable* target. It might be the case that English speakers for reasons outside of representation and phonology, could be unable to consistently produce word-initial prevoicing in an imitational context.

Now we are left with the only case in which the two hypotheses make differing predictions: the plain /p/'s in *Extr. Prev.* **Maintain contrasts** predicts that this target is perfectly valid. The opposition of a prevoiced /b/ and a plain /p/ is a satisfactory way of realizing the word-initial /p b/ contrast in English. Even though the prevoiced range is not a part of the VOT spectrum English usually uses, the disctinction is along VOT, /p/ has a longer (more positive) VOT than /b/ does, and by a substantial amount (here, there is a difference of 145 ms between the two). Therefore, **Maintain contrasts** predicts that English speakers will accommodate to a plain /p/.

As opposed to **Maintain contrasts**, **Maintain categories** predicts no accommodation to a plain /p/. This is largely based on the fact that 15 ms is an unsatisfactorily short VOT value for an

English /p/ to have. Since this hypothesis is not directly concerned with contrasts, it does little good that even while having a plain /p/, the model talker still clearly maintains a distinction (a bimodal distribution) between her /p/'s and her /b/'s. In the model talker's speech plain /p/'s do not actually encroach on her own /b/ category, which is prevoiced. However, since it encroaches on the *listeners'* /b/ category, the plain /p/ is an invalid target.

All in all, in the English experiment the greatest point of interest will be whether accommodation can be observed with respect to the plain /p/'s in *Extr. Prev.* If accommodation is observed in this case, it can be seen as evidence for **Maintain contrasts**—i.e. that a speaker's representations are flexible enough to be able to shift a contrast to a new range of the phonetic dimensions the contrast is originally defined along. If we see no accommodation, that is either evidence supporting **Maintain categories**—i.e. that there is a part of a speaker's phonology that mandates adherence to the phonetic details of a category—or it could indicate something about prevoicing as a cue. A lack of convergence to plain /p/'s could indicate that prevoicing is not perceived as an equal part of the VOT spectrum (ranging from prevoiced through short-lag to long-lag), at least not by speakers of languages with a two-way voicing contrast.

The two conditions in the context of the Hungarian experiment

Which of the two conditions seems more natural is different in the case of Hungarian participants *Figure 2.4*. In Hungarian, the voicing contrast is expressed by the presence or absence of prevoicing: /b/'s are heavily prevoiced (-125 to -60 ms VOT Lisker and Abramson, 1964; Gósy, 2001; Gósy and Ringen, 2009; Gráczi, 2011), while /p/'s are plain (13.2-34.8 ms Gósy, 2001). The way the /p b/ contrast is expressed in *Extr. Prev.* is much closer to these values. The /b/ target has 130 ms of prevoicing, which is still beyond what Hungarian native speakers themselves produce, and the plain /p/ with its 15 ms VOT is well within the plain range. On the other hand, the *Extr. Asp.* stimuli challenge typical Hungarian stop categories. While the plain /b/ in the stimuli (5 ms VOT) would

be most likely categorized as a typical /p/ without much uncertainty, in the experiment it clearly appears in words that start with /b/, so its identity is reinforced lexically. Even if the model talker shifts /b/'s VOT "up" (compared to typical values), /p/ is also shifted up in return. The /p/ target in *Extr. Asp.* has 130 ms VOT, which maintains the distinction, albeit in a different range of the VOT spectrum.

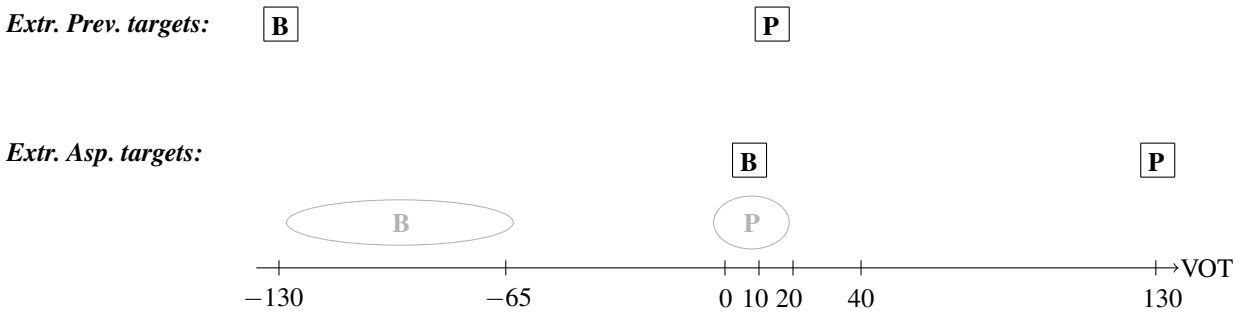

Figure 2.4: Conditions in the Hungarian experiment

Just like in the English case, the two hypotheses make almost identical predictions except for one case. The predictions of **Maintain contrasts** can be seen in *Figure 2.5* on top, and the predictions of **Maintain categories** are in *Figure 2.6* below. Participants are expected to accommodate to stimuli in the *Extreme Prevoicing* condition by both hypotheses—albeit, in the case of plain /p/'s they might already match the target 15 ms VOT to begin with. The more interesting condition here is the *Extreme Aspirating* condition, since this is the one that requires more representational flexibility to accommodate to. Similarly to how whether English-speaking participants accommodate to a prevoiced /b/ is a not necessarily just a matter of valid targets, the two hypotheses cannot make concrete predictions about whether Hungarian native speakers will accommodate to an aspirated /p/. While representationally these targets would be fine from the perspective of both hypotheses, whether they will be accommodated to in a measurable way also depends on whether such realizations are feasible, and thus the two hypotheses remain agnostic about whether accommodation will be seen.

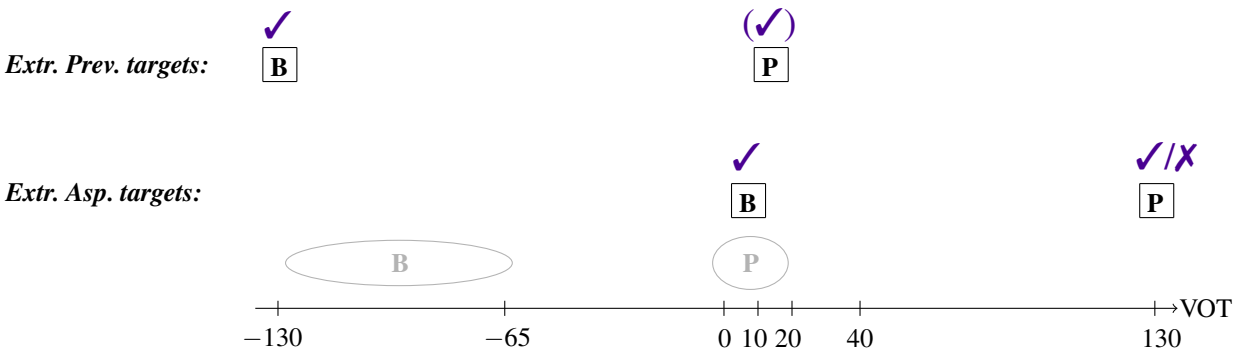

Figure 2.5: Predictions of the maintain contrasts hypothesis for Hungarian
Gray: typical Hungarian values; ✓: hypothesis predicts convergence; ✗: hypothesis predicts no convergence

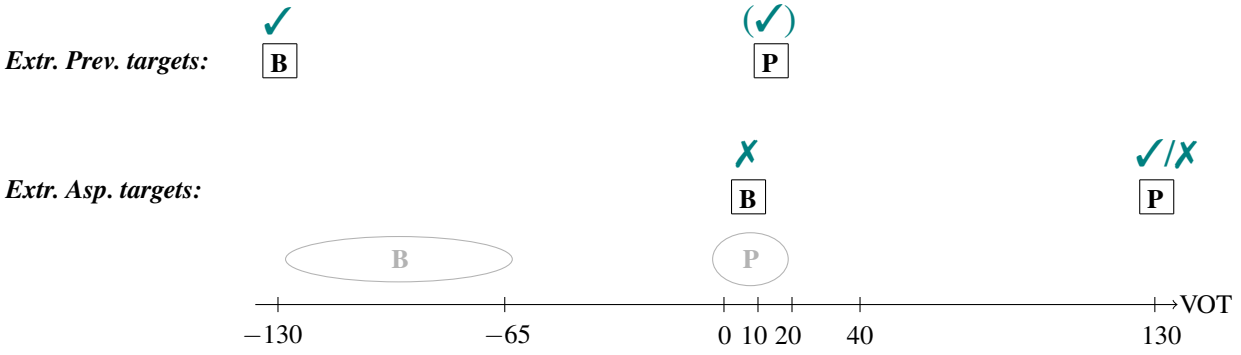

Figure 2.6: Predictions of the maintain categories hypothesis for Hungarian
Gray: typical Hungarian values; ✓: hypothesis predicts convergence; ✗: hypothesis predicts no convergence

However, the two hypotheses make different predictions about the plain /b/'s in *Extr. Asp.* From the perspective of **Maintain contrasts**, a plain /b/ is a valid target, because the word-initial opposition of a plain /b/ and a very aspirated /p/ mean a sufficient distinction between members of the /p b/ contrast. The contrast is still maintained along VOT, /p/ has a longer (less negative) VOT than /b/, and the distinction is also substantial in terms of size. Since a short lag is also implementable from an articulatory perspective, **Maintain contrasts** predicts accommodation. At

the same time **Maintain categories** predicts no accommodation, since a plain stop is not a valid realization of the category /b/ in Hungarian. The main reason for that is because of the VOT being too close to 0, such a token would be categorized as a token of /p/ rather than of /b/. Since such a /b/ token is an invalid target, **Maintain categories** predicts that no accommodation will happen to a plain /b/ in Hungarian.

Therefore, depending on whether we see accommodation to plain /b/'s in Hungarian can help us decide between the two hypotheses. If we see accommodation, it is support for the **Maintain contrasts** hypotheses—i.e. representations can tolerate shifts in the realization of a contrast as long as the contrast itself is not threatened. If, however, we do not see accommodation, it is either a reflection of participants adhering to the phonetic details of their native categories (i.e. the **Maintain categories** hypotheses) or it tells us something about aspiration. Such an outcome could possibly mean that aspiration is not a salient enough part of the VOT spectrum for Hungarian speakers to tolerate a shifting of their stop categories into this region.

This situation parallels the interpretations of a possible outcome where native English speakers do not imitate a plain /p/ (in *Extr. Prev.*). There, we said that that could be either seen as evidence supporting **Maintain categories**, or as something indicative of the fact that English speakers might not view prevoicing as a salient enough part of the VOT spectrum. It is precisely because of this ambiguity (between **Maintain categories** and the relative salience of cues as alternatives to short-lag VOT) that a comparison of the English and Hungarian experiments is essential. If we see accommodation to either a plain /p/ in English or to a plain /b/ in Hungarian, but not both, then the lack of accommodation in the other case must be reflective of something specific to the given cue or contrast, since if it was something universal about phonologies like **Maintain categories**, then we would see its effects cross-linguistically.

If we do not see convergence to these targets in either case, then we have a somewhat trickier situation on our hands. Logically, then it must either be due to **Maintain categories** (an adherence to

a certain amount of phonetic detail when producing an instance of a category) or a reflection of both phonetic biases at once. This would mean that native speakers of *both* languages that we examined show a certain amount of resistance towards phonetically radically different implementations of their sound categories (i.e. English speakers do not want to abandon aspiration for the sake of prevoicing and Hungarian speakers do not want to abandon prevoicing for the sake of aspiration). At one point we must ask ourselves the question to what extent the adherence to phonetic detail (as expressed by **Maintain categories**), and a cross-linguistically holding preference for certain cues are different. I argue, that these two interpretations, while not necessarily equal, are not that different from one another. In case we see no accommodation at all in either of the two relevant types of stimuli, we can interpret that as a sign of an adherence to phonetic detail, that is, support for the **Maintain categories** hypothesis.

2.5 Other, related issues

In the previous four sections I provided some basic background on accommodation, established the main question of this dissertation, and formulated two hypotheses for a possible answer. I also defined voicing contrasts (in aspirating vs. prevoicing languages) as the case through which we can best test these hypotheses. After reviewing the literature on voicing contrasts cross-linguistically and specifically for English and Hungarian (the two languages the hypotheses will be tested on), I described what the two conditions in the experiment will exactly look like, and what predictions the two hypotheses make for these conditions in the two testing languages respectively. Aside from its main focus, this dissertation will also allow me to address some secondary issues, namely whether accommodation behavior for /p/ predicts accommodation behavior for /b/ as well as the relationship between accommodation and extra-linguistic factors like gender, ethnicity, and likeability. The available literature on these topics will be reviewed here.

2.5.1 Linguistic issues around accommodation

Aside from the issue of how representations limit accommodation, there is one more issue that these accommodation experiments will allow me to address, but whose background I have not reviewed yet. In these experiments participants will be exposed to one of two versions of a contrast. This allows us to observe whether or not a participant's accommodation behavior for one segment (e.g. /b/) correlates with the accommodation behavior along another (/p/). That is, do the productions of contrastive voiced and voiceless stops co-vary?

There is some evidence from previous literature that sounds which form a natural class co-vary. Vowel studies measuring accommodation as a change in formant productions found mixed results. While some, like Babel (2009), find no correlation between the accommodation of different vowels, others do. Sanker (2020) exposed her participants to tokens of /ɛ/ with manipulated F1 (lowered or raised). Participants not only adjusted their /ɛ/ productions to converge with the model talker's tokens, but they also generalized this to another mid vowel (/ʌ/). Interestingly, this did not carry over from front mid /ɛ/ to front vowels of other heights—neither low /æ/ nor high /ɪ/. This suggests that generalizations do not apply to sounds that are *contrastive* along a given manipulated phonetic dimension, only to sounds that are *similar* along that given phonetic dimension.

At the same time, we might expect that the generalizability of F1 accommodation (and the limits of this generalizability) does not necessarily carry over to the VOT of stops. Unlike for vowel formants, VOT accommodation seems to more clearly generalize to other voiceless stops. When exposing participants to artificially lengthened VOT's on /p/, Nielsen (2008, 2011) found that they not only lengthened the VOT of their own /p/ productions but the VOT's of their /t/ and /k/ as well.[6] However, pairs of contrastive voiced—voiceless stops are yet to be tested. While the fact

[6]Mielke et al. (2013) found that generalizing from segment to natural class was only exhibited by neuro-typical participants, and speakers with Autism Spectrum Disorder only showed a lengthening on the segmental level (for /p/).

that there seems to be a cross-dialectal inverse correlation between prevoicing and aspirating in English—dialects of English that prevoice voiced stops more tend to aspirate voiceless stops less —this does not necessarily indicate that these categories "move together" synchronicly as well. It might be possible that these patterns are a result of two steps in Englishes (cue-enhancement and then cue shortening in some order). The present study is an interesting case of testing the cross-contrast generalizability of accommodation. What is more, such generalizations will be facilitated in this study, because the two sounds (/p/ and /b/) are simultaneously "shifted" in the less native-like conditions (*Extreme Prevoicing* for English, and *Extreme Aspirating* for Hungarian).

2.5.2 Extra-linguistic variables

In this subsection, I will move on to other extra-linguistic variables which will also be involved in this experiment. I will review the available literature on accommodation and its interaction with gender, ethnicity, and likeability, and discuss where the present work could add to pre-existing research. However, we must also keep in mind that accommodation along different phonetic dimensions does not necessarily correlate. This means that we need to be careful when inferring from socially stratified accommodation patterns as measured one way (e.g. as AXB listener judgements) to social stratification in accommodation when measured differently (e.g. as change in vowel formants) (Pardo et al., 2013a, 2017).

2.5.2.1 Gender and accommodation

The gender of both the speaker and the model talker / interlocutor have been shown to have an effect on the accommodation patterns that can be observed in the interaction, even in shadowing tasks. To my knowledge all studies used only cis-gendered binary people both as participants and as model talkers (in shadowing tasks). The following effects might play out differently for trans and non-binary participants, but at the moment not much is known about that. In these studies

the gender of model talkers is either signaled by pronouns or it is assumed that the cis-gendered model talker was gendered correctly by all participants. These studies were not designed to and thus cannot differentiate potential effects of gender identity from effects of perceived gender.

It has been established that both males and females converge with one another along some dimension(s) (Bilous and Krauss, 1988). This means that while a 'male dominance hypothesis' (Thorne and Henley, 1975) could be true in terms of number of interruptions or how much each participant speaks during the conversation, it is not the case that males simply dominate conversations and females converge with them.

When accommodation is measured as third-party listener judgments, results are somewhat mixed on who accommodates more to who. When assessed via AXB judgements, females were found to converge more often by third-party listeners (Namy et al., 2002; Pardo, 2006, 2010; Babel et al., 2014). In a study that also involved AI voices, third-party listeners did not find that males converged with either the human female voice or the female AI at all (Cohn et al., 2019). Giles et al. (1991) attributed it to "greater affiliative strategy" on the part of females, and Nygaard and Queen (2000) argue that females might be more attentive and pick up more information—based on a voice identification task, where females demonstrated a more frequent use of indexical features. However, AXB tasks also introduce the third-party listener, the one judging whether accommodation happens, as a variable. Namy et al. (2002) found that convergence was more often detected by female listeners, which could be consistent with the idea of females being more sensitive to socially indexed information. At the same time, females might be *detected* to converge more often than males even when males converge more than females in terms of raw acoustic measures, like f_0 (Babel and Bulatov, 2012).

Aside from discussions on who accommodates more in general, a difference in accommodation trajectories was also noticed. In a shadowing task of multiple repetitions (as measured by formant values), women with higher formant values accommodate gradually, while men did so

immediately, from the first repetition on (Babel, 2009). Convergence here was measured by formant values and the model talker was a male, who naturally had lower formant values than the females in the study (i.e. there is a confound between habitual F1 and F2 values and gender). Therefore, women had "further to go" if they wanted to match the model talker's targets. Thus, the difference in accommodation trajectories could also be explained as an effect of phonetic distance: converging to a further target happened gradually but accommodation to a nearer target happened immediately.

Some interactions between gender and other social variables have also been documented. While the more attractive female shadowers find a male model talker, the more they converge with them, this patterns is flipped for males (Babel, 2009). Males diverge from the male model talker when they find him attractive. Other studies simply found that females showed a larger effect of attractiveness on accommodation than males did (Babel et al., 2014).

There is also an interaction between gender and leadership. Though the data is highly complex, in a map task when one participant has to request information from the other, receiver-oriented convergence was more common in female-female dyads—i.e. the information giver converged more to the receiver—while male-male dyads not only exhibited more convergence overall, but convergence was also typically giver-oriented—the one who requested the information converged with the information giver (Pardo, 2006). These patterns were not replicated when one of the parties was given explicit (but secret) instruction to imitate the other participant without being noticed. When the receiver was instructed to converge with the giver, both parties (irrespective of gender) ended up converging. While in M-M dyads the receiver converged more (as instructed), in F-F dyads, the giver converged more, which was consistent with Pardo (2006), but went against the instruction. When the giver was instructed to converge with the information receiver, the only convergence was detected in M-M dyads. While this was the strategy observed in F-F dyads in Pardo (2006), apparently explicitly instructing the giver to converge with the receiver resulted in interference, and thus no convergence happened.

When we look at VOT specifically, we find results with competing interpretations. This study could help disambiguate between them. Nielsen (2008) finds that more males than females lengthen their VOT of voiceless stops when shadowing artificially lengthened VOT stimuli coming from a male model talker. She mentions two possible explanations for this. First, it could be that the $\sim$100 ms VOT was still too short to converge with for females, who had longer VOT's even pre-exposure. This way, even if a female's post-exposure productions were closer to the target than a male's were, statistically speaking, she might have not been detected to converge overall, because of her baseline having been so close to the target to begin with. Second, since the model talker was male, it is possible that male participants converged more because of same-sex convergence effects (Giles et al., 1991). There is also a third alternative: that VOT might be a dimension along which males converge more. This is to be tested in the current study. An experiment was proposed to choose between these possible explanations (Graff et al., 2009), but the final results from this study were never published, and because of the multiple interacting design variables, preliminary results from individual data were hard to interpret.

This will be tested in the current study by not only exposing participants to an even longer VOT values (130 ms vs. the at least 100 ms in Nielsen, 2008) but also by using a female model talker rather than Nielsen's male model talker. If the first explanation is true, and 100 ms was simply too close to the females' baseline, then in this study we expect to see no difference between males and females in terms of how much they accommodate. If we see that female participants, who have longer habitual VOT's than males, converge more to the female model talker, we must infer that Nielsen's results were due to same-sex convergence. If we see males accommodate more, then maybe VOT is a dimension along which males converge more than females.

2.5.2.2 Ethnicity and accommodation

As mentioned before, Babel (2009, 2012) found that the more Anti-Black bias a white native speaker of English exhibits in an Implicit Association Task, the less likely they are to converge with the Black model talker—in fact they might even diverge.

Aside from the speakers' perception of the model talker (and their associated biases), the speakers' own ethnic identity might also modulate their degree of accommodation. In the experiments in this work the model talkers are white, but participants in the English dataset are ethnically diverse. Most research regarding accommodation has focused on bi-dialectal speakers of African American Language (AAL) and how their speech changes when talking to Black vs. white interlocutors. There is ample evidence of African Americans code-switching and adjusting their speech depending on the interlocutor's perceived identity and the formality of the situation (e.g., Labov, 1972; Garner and Rubin, 1986; Fasold et al., 1987; Lacy, 2004; Wolfram, 2007; Rahman, 2008; Scanlon and Wassink, 2010). It has also been noted that this code-switching, when done in the right settings, is also perceived favorably by Black listeners (Koch et al., 2001).

However, little is known about how or whether the speaker's ethnicity influences their shadowing behavior with a white model talker. It should also be added that while shadowing a white model talker in a laboratory might indeed elicit code-switching from a speaker of AAL, their baseline will be recorded in a reading task (also in a laboratory setting) conducted by a (different) white experimenter. This means that participants' experimental baseline might already reflect adjustments to a formal setting and thus might differ greatly from their habitual speech patterns in informal settings. This difference might be especially great for speakers for whom these adjustments involve dialectal code switching as well (e.g. speakers of AAL).

2.5.2.3 Likeability and accommodation

The relationship between participants (or between the speaker and the model talker in a shadowing task) has been shown to influence how much accommodation happens. Over the course of a longer time, closeness of the relationship correlated with the amount of accommodation observed for college students (Pardo et al., 2012) as well as for reality show participants (Bane et al., 2010; Sonderegger, 2015). As mentioned before, ideological closeness also resulted in more convergence even in shorter conversations (Gregory and Hoyt, 1982).

Immediate impressions of one's interlocutor (or model talker) can also facilitate or inhibit accommodation. As mentioned before, attractiveness can correlate (or inversely correlate) with accommodation behavior, which is further mediated by gender (Babel, 2009; Babel et al., 2014). Moreover, sympathy and liking have also been demonstrated to have a similar effect: the more likeable the participant found their interlocutor, the more likely they were to converge with them and the more they disliked them, the more likely they were to diverge (Natale, 1975; Babel, 2010; Schweitzer et al., 2019).

It has been found that visual images facilitate accommodation. Gregory et al. (1997, 2001) looked at what happens when f_0 is eliminated from the speech signal with a high-pass filter. In their experiment participants heard their interlocutor through audio, which was either filtered or not, and had a chance to see them as well via CCTV. Even though fundamental frequency could have been recoverable from other, secondary information (e.g. the periodicity of glottal pulses and other harmonics), participants exhibited less f_0 convergence than controls whose input was not filtered. Moreover, participants in the high-pass filter condition looked up less at the CCTV screen showing their interlocutor. This not only suggests that the presence of low frequency energy (and f_0) adds to the social quality of the conversation, but that the amount of convergence is correlated with how often their participants looked at their interlocutor.

Visual information of the model talker is also important in less social tasks, such as shadowing studies, where the participants repeat words after pre-recorded audio. Babel (2009) finds more accommodation in her Social Condition, where participant saw a still image of the model talker while they were exposed to their voice as opposed to the Asocial Condition, where no image accompanied the audio. This is standard design in shadowing studies, and the experiments in this dissertation will also follow suit in order to elicit as much accommodation as possible.

My design for both studies will involve manipulation of stimuli to the point where they are unnatural sounding. This has been demonstrated to have an effect on perceptual shifts and "pleasantness" ratings. Participants prefer a Pleasant Control guise (with varied intonation patterns and English-like sound categories) over a Pleasant one with an unnatural Shift (artificially implemented back vowel lowering; Babel et al., 2019). However, participants still showed a perceptual shift in the Shifted case (especially if the guise was also Unpleasant—monotonous intonation with creak). Since convergence is mediated by likeability, we might expect less convergence from participants who are exposed to an unfamiliar/unnatural pattern and find it off-putting. However, the findings of Babel et al. (2019) indicate that the information is still processed and adjusted to in perception. Thus, we can still expect a change in their perception (in the labeling task), even if we do not see it in their production (in the reading and shadowing tasks).

While the importance of likeability has been established, most studies used very general scales to measure it. Other non-accommodation studies in the field have established a few clusters of personal features that can contribute to how likeable a person is perceived to be. Carranza and Ryan distinguish between two categories, *status* and *solidarity* (Carranza and Ryan, 1975; Ryan and Carranza, 1975). In their social evaluation study Zahn and Hopper (1985) studied 56 personality traits and found that ratings for some of their traits heavily correlate with one another. Based on covariance, they established three groups: *attractiveness*, *superiority*, and *dynamism*, where *attractiveness* largely overlaps with Carranza and Ryan's *solidarity* measure. Bauman (2013)

distilled the the 56 features into 9, forming 3 groups each. *Solidarity* is made up of *friendliness*, *honesty*, and *politeness*. *Superiority* has components of *organizedness*, *intelligence*, and *class*. Finally, *dynamism* comprises *talkativeness*, *confidence*, and *energy-level/lazyness*. These 9 scales represent relatively easy to judge features (as opposed to the congregate labels like *solidarity*), which participants could have an easier time using. This study relies on the likeability measures of Bauman (2013).

2.6 Outline of methods

In the first part of this section I have set up the main question that this dissertation asks (on what level do our representations limit accommodation) and defined two competing hypotheses for how this might manifest for contrasts. I then singled out VOT contrasts as a case study, and provided a background on both VOT in the two test languages (English and Hungarian) and VOT accommodation in the two languages. This allowed us to establish the two conditions that the English and Hungarian experiment will each be broken down into and to explore what specific predictions the two hypotheses make for each of these conditions, considering the specific phonetic properties of word-initial voicing contrasts in stops in both languages. After the main question was established, I also reviewed literature on phenomena (generalizability of accommodation and the interaction of extralinguistic factors with accommodation), which might also influence the results of these experiments or to which issues these experiments could contribute.

In this section I lay out the basic design of the two experiments for this work—parts that the English and the Hungarian experiment have in common. The experimental design includes a variety of tasks. Both experiments are made up of eight stages, which are presented in *Table 2.1*. Before each stage, the task is briefly described to the participants in writing via Psychopy, which the participant has to self-advance from. It is also emphasized that they can take short breaks at

these points. The design takes about 40 minutes to complete. During the task, short (one-sentence) prompts remind them of the current task in each trial.

	Instruction	**Stimuli**	**Example**
Rating	Rate speaker for properties	semantic differential scales (10)	shy–talkative
PRE-Labeling	Select the word you hear	word, audio on a VOT continuum (11*10)	*binning* / *pinning* with 45 ms VOT
Familiarization	Read the word silently	written word (40*1)	*basin*
PRE-Read	Read the word out loud	written word (40*2)	*poser*
Shadowing	Repeat the word you hear	word, audio, 1 of 2 conditions (30*6)	*pollen* with 15 ms VOT
POST-Read	Read the word out loud	written word (40*2)	*buzzer*
POST-Labeling	Select the word you hear	word (audio on a VOT continuum (11*10)	*binning* / *pinning* with 75 ms VOT
Questionnaire	Fill in the questionnaire	—	Age:

Table 2.1: Procedure of the experiment

The first stage is a rating task, where participants have to listen to a text on shopping for mattresses as read out by the model talker. During this task they also have to rate the model talker on nine 1-to-9 semantic differential scales for various attributes. In semantic differential scale tasks (Osgood et al., 1957; Osgood, 1964) participants have to rate the stimulus based on where it falls between two extremes—e.g. how *shy* vs. *talkative* the person who is talking might be. These scales are contrasted with Likert-scales which allow for the measurement of reactions on a single scale—e.g. how much a participant agrees with a given statement. Participants are informed of the model talker's gender, either by pronouns (e.g. *her*) or by choice of lexical item (*nő* 'woman').

Following Bauman (2013), the 9 trait pairs in this experiment are evenly selected from the 3 categories established by Zahn and Hopper (1985). *Table 2.2* contains the extremes on the 9 semantic differential scales in the order in which they are printed on the rating sheet—sometimes the more positive end of the spectrum is on the left, sometimes it is on the right. I modify Bauman's scales at one point. Instead of her low class—high class, I use low status—high status, because the Hungarian equivalent of class (*osztály*) does not trivially go with the adjectives high and low, and alternatives, such as *working class* are strongly associated with the communist regime. Finally, participants also rate how attractive they find the model talker personally on a Likert scale of 1-to-9 (from *not at all* to *very*). Participants will be asked this question again at the end as part of the *Sociolinguistic questionnaire*.

Solidarity	**Superiority**	**Dynamism**
friendly — unfriendly	organized — disorganized	shy — talkative
dishonest — honest	lower status — upper status	unsure — confident
rude — polite	intelligent — unintelligent	energetic — lazy

Table 2.2: Semantic differential scales

At the second stage, participants complete a labeling task with words on a VOT continuum. The continuum has 11 steps, each 15 ms apart from −60 ms to +90 ms VOT. In English, this task involves words synthesized on a continuum between *binning* /ˈbɪnɪŋ/ and *pinning* /ˈpɪnɪŋ/. In Hungarian, the endpoints of the continuum are *boros* /ˈboros/ 'wine-like, wine-related' and *poros* /ˈporoʃ/ 'dusty'. The trials start with a blank gray screen. After 500 ms, the instructions appear, and another 500 ms later (1,000 ms into the trial) one of the 11 items on the continuum is played. Participants have to indicate which word they think they heard, using F and J on the keyboard. As mentioned above, the two options were different for participants of the English and the Hungarian experiment. The delay is instigated in order to avoid interference from the sound of the key being pushed. Each of the 11 items are presented to every participant in 10 randomized blocks (110 decisions/participant). This establishes their baseline VOT boundary, their boundary prior to exposure (*PRE-Labeling*).

Participants are then familiarized with the words used in the shadowing and reading tasks. The total set of 40 words from their native language is presented to the participants one by one. In order to control for any potential frequency effect (Goldinger, 1998; Goldinger and Azuma, 2004; Nielsen, 2011), all words were chosen to be low-frequency. No fillers are included. The exact list of words for each language and their frequencies will be shown later in their respective chapters (*Chapter 3* for English and *Chapter 4* for Hungarian). In each trial, the instructions (*"Read this word out silently in your head:"*) and the word appear on the screen simultaneously. The word is displayed for 2,000 ms, and the instruction is displayed for 3,000 ms (a second longer than the word). When the instruction disappears, a new trial starts with a new word. Thus, words are presented every 3 seconds. Participants are asked to first read the words without saying them out loud in order to reduce hyperarticulation effects in their baseline productions (Goldinger and Azuma, 2004; Nielsen, 2011). Next, participants are shown the 40 words one-by-one again in the same manner, but this time the text instructs them to say the words out loud. Similarly to *Familiarization*, a new

word is presented every 3 seconds. The entire list is presented twice in a randomized order each time, which yields two pre-exposure baseline productions per word per participant (*PRE-Read*).

After this, the experiment contains 6 blocks of *Shadowing* (Goldinger, 1998), where participants hear recordings of some of the words from the reading task and are asked to identify and repeat the English/ Hungarian word they hear. The words are a semi-randomly selected subset of 30 words selected from the 40 words they were previously familiarized with. Each participant hears 8 monomorphemic b-words, 7 polymorphemic b-words, 8 monomorphemic p-words, and 7 polymorphemic p-words—a different set of 10 items are withheld from each participant. First in each trial the instruction appears (*"Listen to our speaker say an English word and repeat the word she said."*) along with an image of the model talker. The image is displayed while the audio stimulus was playing in order to facilitate accommodation (Babel, 2009). It stays on screen for 1,300 ms. The audio stimuli starts playing 100 ms into the trial. 1,500 ms into the trial, once the audio stimuli is done, the words *"Repeat the word"* are displayed in the middle of the screen instead of the image of the model talker. After the text is on for another 1,500 ms (3 seconds into the trial), the screen goes gray for 1,000 ms to indicate a split between trials. A new audio stimulus is thus presented every 4,000 ms. These productions provide the *Shadowing* data, 6 recordings per participant for each of the 30 semi-randomly chosen words. Subsequently, the reading out loud task is repeated to record post-exposure values and measure non-immediate effects of exposure—*POST-Read*. This is done with the full list of 40 words, i.e. withheld items are included. Subsequently, the labeling task is also repeated *POST-Labeling* with the same continuum and methods as in *PRE-Labeling*. While the *POST-Read* values reflect the amount of accommodation slightly after exposure, *POST-Labeling* results test whether participants categorize differently after being exposed to the audio stimuli. Changes in categorization might still happen even when the participant did not accommodate in production (see Babel et al., 2019).

Lastly, participants fill out a sociolinguistic questionnaire, which records their age, gender identity and linguistic history, including their native language, place of birth, languages they are or were regularly exposed to, what environment, and for how long. These questions are all free-form. Participants have to assess again how attractive they found the model talker on a 1-to-9 scale just like at the beginning, in order to see if the length and nature of the tasks themselves changed their opinion. They can also provide their reasons for their assessment at the end as well as general comments on the study.

2.7 Research questions of the experiments

Now that we have an explicit design, we should recap the questions this dissertation could contribute to. This study can contribute to a number of issues. The main question for this study is what mechanism drives contrast maintenance, more specifically whether phonetic details of individual sound categories restrict the process of accommodation. The two hypotheses I have discussed in the introduction are **Maintain contrasts** and **Maintain categories**. **Maintain contrasts** is an abstract pressure to maintain a distinction between each pair of contrastive sounds along some potentially predefined dimension(s). This pressure can be satisfied if the contrast is moved to a different part of the same phonetic dimension or spectrum. **Maintain categories**, however, expresses a pressure for speakers to adhere to the phonetic specifications of each category in their native language. Unlike **Maintain contrasts**, **Maintain categories** involves more phonetic detail, and is harder to satisfy. Moving a contrast to a different part of the same phonetic dimension might not necessarily meet this requirement. In the two experiments in this study, English- and Hungarian-speaking participants will be exposed to either an extreme version of a prevoicing contrast (prevoiced /b/, plain /p/) or an extreme version of an aspirating contrast (plain vs. aspirated) as a baseline. While the former will be challenging for English speakers' representations, the latter will challenge Hungarian speakers more.

The two hypotheses make different predictions. For the extreme prevoicing contrast, **Maintain contrasts** predicts that this contrast should be subject to imitation even by English speakers, because the contrast is maintained in a different part of the VOT spectrum. Similarly, Hungarian speakers should not have any problem accommodating to a plain /b/ (in the extreme aspirating contrast). At the same time, **Maintain categories** finds that such realizations are not satisfactory manifestations of the English and the Hungarian /p b/ contrast, respectively, because they do not adhere to the phonetic specifications of the participants' original categories. Therefore, **Maintain categories** predicts no accommodation for either a plain /p/ (in *Extr. Prev.*) in the English experiment or a plain /b/ (in *Extr. Asp.*) in the Hungarian experiment.

This study can also be informative about how accommodation generalizes on a segmental level. Several studies have found speakers generalizing across a natural class (e.g. from /ɛ/ to all mid vowels, and from /p/ to /p t k/). At the same time, there is little research on contrasts "moving" together, specifically whether the amount accommodated for one member of the contrast correlates with the amount of accommodation found for the other member. This study could contribute to the issue by comparing individual patterns for /p/ accommodation and /b/ accommodation and seeing if there is correlation between the two.

These experiments could also address task effects. Comparing participant behavior during the shadowing task and the reading task could accomplish this in two ways. First, if there is convergence in *Extr. Prev.* shadowing, does converging to un-nativelike stimuli persist after exposure, i.e. does it show up in the reading task too? Second, are there people who show a pattern while exposed to the speaker (converge or diverge), but which goes away post-exposure. To put it differently, are there *strictly immediate convergers*, who show no change in reading, but adjust their speech to the model talker while directly exposed to her? We can compare participant behavior not only across reading and shadowing, but also across labeling and reading or shadowing. Thereby we can investigate if changes in production (convergence or divergence) are reflected in a change in where the

category boundary is drawn. The flipside could also be investigated: whether a change in labeling productions correlates with a certain accommodation profile during reading and/or shadowing or whether an adjustment of categorical boundaries occurs even in participants who do not change their productions.

This study could also uncover effects of articulatory fatigue. It could accomplish that by looking at whether we see particular patterns co-occurring with imitation of longer articulatory gestures (long VOT or long prevoicing) that we do not see when participants are imitating shorter gestures. For instance, do participants maintain similar amounts of convergence for long aspiration and prevoicing throughout the shadowing task? If we see less accommodation in later repetitions, it could be indicative of some sort of fatigue associated with realizing longer cues for a sustained period of time. If we see such subsiding accommodation, then we must address what it means for the theory that language change can happen via accommodation. In order to maintain this theory, we must suppose either that these difficulties can be overcome with time, or in the less likely case that change in the direction of articulatorily *more* difficult categories is not possible (at least via accommodation).

Aside from linguistic variables, this study could also address the relationship of accommodation and four different extralinguistic variables: the gender and ethnicity of the speaker and the liking and attraction they experience towards the model talker. It is important to note that these effects might apply differently in different cultures and communities and through them in different languages. In terms of gender, there are many points where we could observe gender effects, but a particularly important aspect would be to clarify a finding of Nielsen (2008, 2011). Nielsen found that males converged more with an artificially lengthened VOT than females did. She proposes that this could either be because the at least 100 ms VOT was still not far enough from females' habitual productions to elicit detectable convergence or because of a bias for same-sex accommodation giving males a boost. A third possibility is that VOT is a phonetic dimension along which males

show more convergence. While this study is not designed to systematically test the nuances of this effect (especially the effect of inter-speaker baseline VOT variation across genders), results could be informative nonetheless. In the experiments in this dissertation the extremely aspirated target will have even longer VOT than in Nielsen's work (130 ms VOT) and the model talker will be female. If males converge more, then her results were due to a gender bias (i.e. males converge more for VOT than females do). If males and females converge the same amount then Nielsen's (2008) results were due to a ceiling effect in female productions and her model talker having been too close to the female participants' baseline in the study. If females end up converging more in this experiment, then (in conjunction with Nielsen's results) we have evidence for same-sex convergence.

Moreover, since the study is also recording self-identified ethnicity information from the participants, there will also be an opportunity to examine and discuss any effects of ethnicity (with a white model talker).

Furthermore, this study could reveal more about the exact role various facets of likeability and attraction play in accommodation. While several studies have found that participants converge more to an interlocutor that they find likeable, this was often measured with a single scale ("How likeable do you find this person?"). In this study, I will take a more nuanced approach and try to tease apart three facets of likeability, namely: `Solidarity` (friendliness, honesty, and politeness), `Superiority` (intelligence, social status and organizedness), and `Dynamism` (talkativeness, self confidence, and energy level). In addition to these three measures, I will also include attraction as a measure, since likeability and attraction could be closely related, and since most studies were done without attraction or even sexual orientation data, these two could be hard to tease apart.

Chapter 3: The English experiment

A study was run with native English-speaking participants in order to investigate the questions outlined in the previous chapter. This section will first present the study's methods (participants and stimuli) and recap its procedure (*Section 3.1*). The results will be broken down into three sections based on tasks that each participant completed (*Section 3.3*: reading, *Section 3.4*: shadowing, and *Section 3.5*: labeling), and the rating results will be incorporated as independent variables of these sections. Each of these section will start with an introductory section. This will contain an overview of results, a discussion of statistical methods used for that task and also a reminder of all the questions that can be answered with data from the given task. Then, each section will discuss the *Extreme Aspirating* condition first, and then the *Extreme Prevoicing* condition. Finally, sections will conclude with a summary. After each task has been discussed, I will return to the main questions once more and provide an interim summary of how close the English data got us to an answer.

3.1 Methods

In this section I will describe the methods of the English experiment. First, I will review the participants (*Section 3.1.1*). Then, I will discuss the materials for all four tasks (rating, labeling, reading and shadowing, *Section 3.1.2*). Finally, I will summarize the procedure, which was described in more detail in *Section 2.6*.

3.1.1 The Model talker and the Participants

The model talker was a 36 year-old cis-female. She is a phonetically trained native speaker of English, without any hearing disorder or speech impediment. She also provided a recent picture for the study. A total of 45 participants were recorded for the experiment, 41 of which are analyzed here. 2 participants had to be excluded because of previously unreported experience with another language, 1 because of equipment failure and 1 because of the quality of recording and the participant's lack of compliance with the task. Participants were identified by a unique, self-generated code both on their paper-based sheets and in Psychopy. The number of participants by condition and gender are shown in *Table 3.1*. 22 participants identified as female, 19 as male, no other gender identities were reported. The age of female participants was 18–24 years old (mean: 19.68), and males were aged 18–27 (mean: 21.05). Participants were all enrolled in college at the time of the experiment.

	Extr. aspiration	**Extr. prevoicing**	**Total**
Female	11	11	**22**
Male	10	9	**19**
Total	**21**	**20**	

Table 3.1: Participants' breakdown by condition and gender

All participants' only native language was English and none of them reported having been diagnosed with a hearing disorder. 2 male participants reported their parents being a native speaker of another language (Korean and Mandarin), who were assigned to the *Extreme Prevoicing* condition to avoid potential confounds. Other assignments happened at random. All participants were American, except for 1 British male and a Canadian female, who were both randomly assigned in the *Extreme Prevoicing* condition.

The ethnic make up of the two groups sorted into the *Extr. Asp.* and the *Extr. Prev.* conditions was quite similar (*Table 3.2*), but there were some differences. While both groups had 3 people with South-East Asian ethnicities, in *Extr. Asp.* these were 1 person identifying as *Asian*, 1 as *Indian*, 1 as *Malaysian Filippino*, in the *Extr. Prev.* 1 person identified as *Asian*, 1 as *Korean* and 1 as *Chinese*. The Middle-Easterners in *Extr. Prev.* identified as *Middle Eastern* and *white/Middle Eastern*, and the 2 Hispanic-LatinX people in *Extr. Asp.* identified as *Hispanic* and as *Puerto Rican*. The Mixed population varied the most in their descriptions. In *Extr. Asp.*, it was made up of people who self-identified as *Mixed* or *Persian-Panamaian*, and in the *Extr. Prev.* the group was made up of people with *"white, black"*, *Asian/Hispanic*, *South Asian/Middle Eastern*, *Mixed* and *"Afro-Latina and white"* and *Mixed (Indian, Irish, German)* identities. There were also two people speaking non-American Englishes: there was a Canadian participant (F22 Extr. Prev., *Asian/Hispanic*), and a British participant (M04, Extr. Prev., *Black/African-American*).

3.1.2 Materials

Audio stimuli for the entire experiment (rating, labeling, and shadowing) were recorded with the same model talker. She was a native speaker of English and English only, identifies as female, and is 36 years old. She is phonetically trained and speaks with a region-neutral North American accent.

3.1.2.1 Rating

In the rating task, participants listened to a short audio recording of the model talker discussing aspects to consider when shopping for a mattress. The text was an English translation of an article from a Hungarian lifestyle blog (see Appendix, *Figure A.1*). It was chosen because of its neutral topic and tone, as well as compatibility with both English and Hungarian. The speaker was recorded reading the text 3 times at a medium pace, and the most natural-sounding production was used. Recordings were sampled at 44.1 kHz, with a KayPentax CSL Model 4500 in a sound booth.

3.1.2.2 Labeling stimuli

In the labeling tasks, participants were exposed to a word created on a continuum between *binning* /ˈbɪnɪŋ/ and *pinning* /ˈpɪnɪŋ/. Items were spliced together in Praat to form an 11-step continuum with 15 ms steps ranging from −60 ms VOT (prevoicing) to +90 ms VOT (aspiration). Each step was spliced from the model talker's reading of the two words in a word list 6 times each—along with the words described in *Section 3.1.2.3*. Recordings were sampled at 44.1 kHz, with a TASCAM DR-40 PCM recorder in a sound booth. Each token was spliced together from the following 3 or 4 parts. *Figure 3.1* demonstrates the stimulus with no aspiration and 30 ms of prevoicing.

Every token started with 120 ms silence. Then, for prevoiced tokens, a 15, 30, 45 or 60 ms long instance of prevoicing was cut (between the yellow and blue lines). These bits were obtained from the same naturally produced token. In order to make sure the eventual stimuli sounded natural, a naturally produced single instance of prevoicing was used. Since prevoicing is extremely rare in word- and utterance-initial voiced stops in English, there were very few natural productions to choose from. In the end, bits of prevoicing were cut out from one instance of prevoicing that had a total duration of 172 ms, which had a consistent amplitude for 130 ms. This was followed by a burst with about 0 ms aspiration ("∼0 ms long burst", between the blue and green lines), obtained from a token of *pinning* /ˈpɪnɪŋ/. In order to control for any effects of f_0, the "ending" was obtained from a *binning* /ˈbɪnɪŋ/ token after f0 was stable, and F1, F2 and F3 were present (right of the green line).

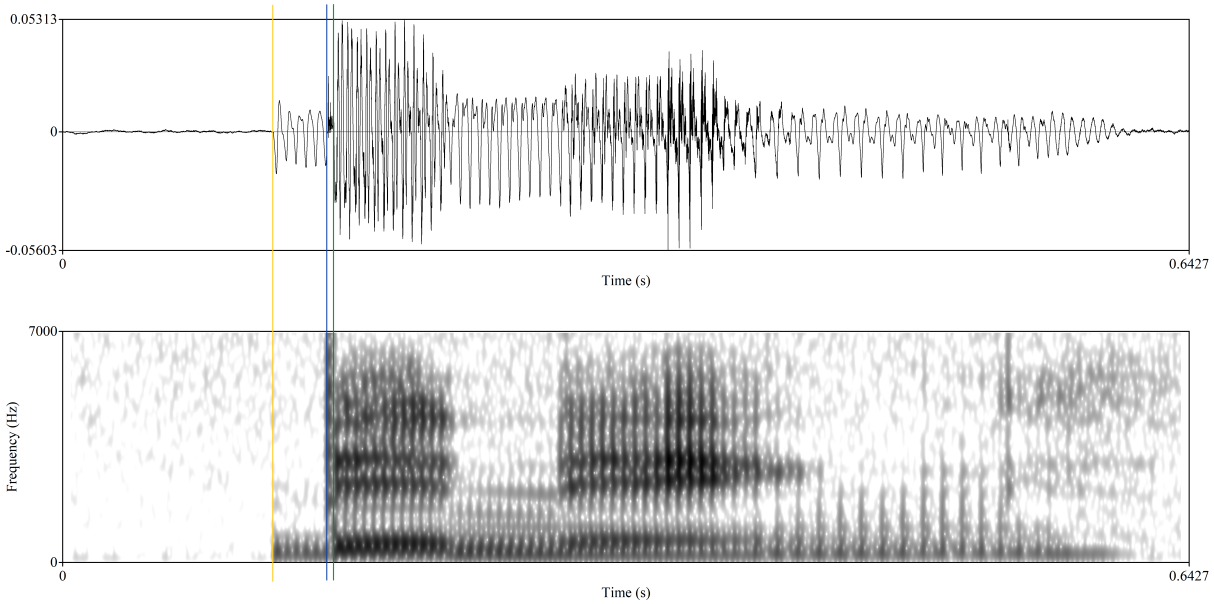

Figure 3.1: Waveform and spectrogram of binning/pinning stimulus with 30 ms of prevoicing and no aspiration

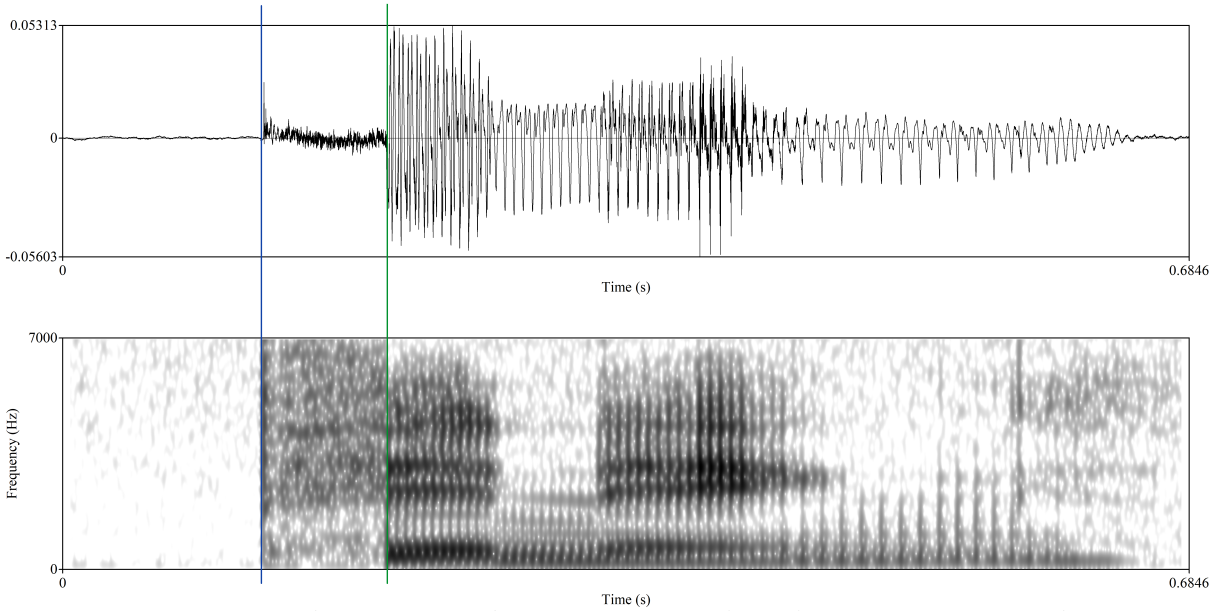

Figure 3.2: Waveform and spectrogram of binning/pinning stimulus with 75 ms of aspiration and no prevoicing

The 0 ms VOT step was made up of the 120 ms silence, the 0 ms burst, and the *-inning* ending. The aspirated tokens (with more than 0 ms VOT, e.g. *Figure 3.2*) were made up of 120 ms silence, and a burst with some aspiration (0, 15, 30, 45, 60, 75 or 90 ms; between the blue and green

lines), which was a longer single except from the same *pinning* /ˈpɪnɪŋ/ token that the 0 ms burst came from. Finally, the ending (right of the line) was spliced to the end to create the stimulus. The burst itself, as well as the ending (-*inning*) were identical in all 11 created items.

Ethnicity	Extr. Asp. (21)	Extr. Prev. (20)
white or Caucasian	9	7
Black or African American	5	3
South-East Asian	3	3
Hispanic or LatinX	2	0
Middle-Eastern	0	2
Mixed	2	5

Table 3.2: Ethnic break-down by condition

3.1.2.3 Reading and Shadowing stimuli

During the reading and shadowing tasks participants had to read and repeat words, respectively. All words have SVC(CC)VC(C) structure—the word-initial consonant (S) is a bilabial stop, and the other consonants are non-stops (fricatives or sonorants). Bracketed consonants are optional. None of the words used in these tasks are a part of voiced-voiceless minimal pairs. The full list contained 40 words, displayed along with their frequency information in *Tables 3.3–3.4*. The list is made up of 20 p-initial and 20 b-initial disyllabic words, containing both mono- and polymorphemic words (10–10 of each). All the words are of relatively low frequency, which have been argued to show more accommodation than high-frequency words, at least for female shadowers (for an overview see Pardo et al., 2017). In the reading task every participant encountered each word.

Word	IPA	Frequency	Word	IPA	Frequency
baffle	ˈbæfəl	8 (0.16)	basher	ˈbæʃɚ	16 (0.31)
banish	ˈbænɪʃ	69 (1.35)	beaming	ˈbimɪŋ	42 (0.82)
basil	ˈbeɪzəl	69 (1.35)	boiler	ˈboɪlɚ	144 (2.82)
basin	ˈbeɪsən	96 (1.88)	boomers	ˈbumɚz	6 (0.12)
beaver	ˈbivɚ	246 (4.82)	boyish	ˈboɪɪʃ	31 (0.61)
bellow	ˈbɛlo	23 (0.45)	bullies	ˈbʊliz	89 (1.75)
bison	ˈbaɪsən	17 (0.33)	bullish	ˈbʊlɪʃ	1 (0.02)
bowel	ˈbaʊəl	183 (3.59)	bunnies	ˈbʌniz	92 (1.80)
bushel	bʊʃəl	32 (0.63)	burner	ˈbɝnɚ	93 (1.82)
burrow	ˈbɝo	28 (0.55)	buzzer	ˈbʌzɚ	201 (3.94)

Table 3.3: English b-words, word frequency in SUBTLEXus 1.0
Corpus of 51M (in brackets frequency/million); Left: monomorphemic, right: polymorphemic

Word	IPA	Frequency	Word	IPA	Frequency
panel	ˈpænəl	372 (7.29)	painless	ˈpeɪnləs	99 (1.94)
panther	pænθəˠ	131 (2.57)	pausing	ˈpɔzɪŋ	16 (0.31)
parlor	ˈpɑɹləˠ	303 (5.94)	paving	ˈpeɪvɪŋ	22 (0.43)
pelvis	ˈpɛlvɪs	103 (2.02)	peeler	ˈpiləˠ	15 (0.29)
pension	ˈpɛnʃən	247 (4.84)	peeving	ˈpivɪŋ	1 (0.02)
perish	ˈpɛɹɪʃ	132 (2.59)	ponies	ˈponiz	142 (2.78)
pollen	ˈpɑlən	62 (1.22)	pooling	ˈpulɪŋ	24 (0.43)
possum	ˈpɑsəm	105 (2.06)	poser	ˈpozəˠ	17 (0.33)
puffin	ˈpʌfɪn	8 (0.16)	pursing	ˈpɝsɪŋ	6 (0.12)
puma	ˈpumə	7 (0.14)	pusher	ˈpʊʃəˠ	64 (1.25)

Table 3.4: English p-words, word frequency in SUBTLEXus 1.0
Corpus of 51M (in brackets frequency/million); Left: monomorphemic, right: polymorphemic

For the audio shadowing stimuli, the model talker was recorded reading out all 40 words 6 times in a self-paced word reading task along with the labelling words (*binning* and *pinning*). Each of the 6 lists contained the words in a different randomized order. Recordings were sampled at 44.1 kHz, with a TASCAM DR-40 PCM recorder in a sound booth. Two conditions were created with the original recordings (*Table 3.3*). In the *Extreme Aspirating* condition (*Extr. Asp.*) participants heard p-words with a lot of aspiration (130 ms VOT), and b-words with a plain /b/ (15 ms VOT). In the *Extreme Prevoicing* condition (*Extr. Prev.*) /p/'s were plain (15 ms), and /b/'s were extremely prevoiced (130 ms prevoicing: −130 ms VOT). The VOT for plain stops falls within the range of VOT for plain stops both in English and in Hungarian, while aspirated stops and prevoiced stops are exaggerated versions of English voiceless and Hungarian voiced stops, respectively. [1]

[1]Standard values will be discussed in the lit review chapter

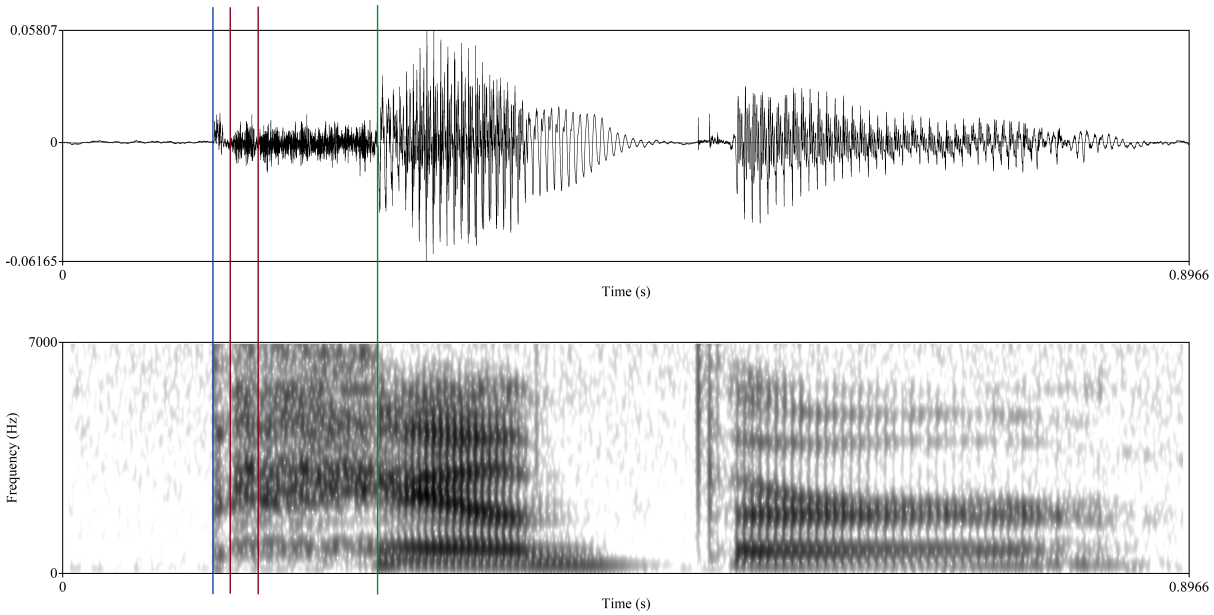

Figure 3.4: Waveform and spectrogram of audio stimulus panther with no prevoicing and 130 ms aspiration

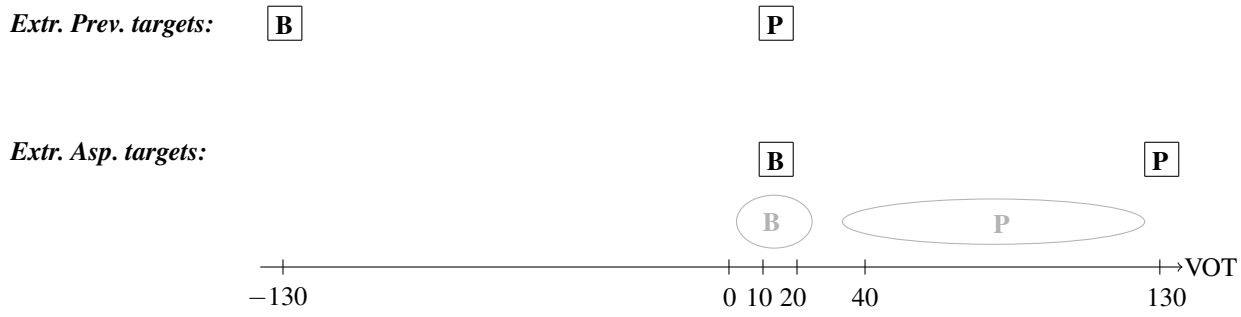

Figure 3.3: Conditions in the English experiment

In the *Extreme Aspirating* condition voiceless stops were aspirated (130 ms VOT), and voiced stops were plain (15 ms VOT), and in the *Extreme Prevoicing* condition voiceless stops were plain (15 ms VOT) and voiced stops were prevoiced (−130 ms VOT). In the *Extreme Aspirating* condition, the aspirated voiceless stops started with 120 ms silence, then for each word a token was selected with longer aspiration, which was then split into two (burst-bit and long-bit). In *Figure 3.4*, the burst-bit starts with the burst (blue line) and lasts until the first red line. The long bit is the longest part of aspiration (in this case 97 ms), between the second red line and the green line.

The "burst-bit" was usually 15 ms (between the blue line and the first red line). In between the two bits (between the two red lines), I spliced a subpart of the long-bit, which completed the combined duration of burst and aspiration to 130 ms (in this case 18 ms). In some cases the same aspiration was used for different words, but the aspiration always came from a token where the following vowel had the same quality and rhoticity. For instance, *pelvis* /ˈpɛlvɪs/, and *pension* /ˈpɛnʃən/ has the same burst and aspiration, but the *perish* /ˈpɛɹɪʃ/ stimuli was created with a different aspiration, extracted from an instance of *perish*. The silence, burst, and aspiration were combined with a word ending. The plain voiced stops in the *Extreme Aspirating* condition also start with a 120 ms lead silence, followed by a 15 ms burst from an instance of *panel* /ˈpænəl/, which was chosen because the burst was sufficiently loud and the first 15 ms of the burst and release did not reflect the following vowel. This happens to be the "burst bit" (between the blue line and the first red line) in *Figure 3.4*. These two parts were then spliced with the word ending.

In the *Extreme Prevoicing* condition, voiceless stops are plain (15 ms VOT). These stops started with 120 ms silence, followed by a 15 ms burst (the 15 ms "burst bit" that was used in the aspirated version), and then an ending, which was the same between the aspirated and the plain version of the word. Prevoiced voiced stops (-130 ms VOT), like all items, also started with 120 ms silence. It was concatenated with 130 ms continuous prevoicing from a production of *beanie* /ˈbini/ (total duration of prevoicing: 172 ms, consistent amplitude for 130 ms), followed by the 15 ms burst used for all voiced stops in the *Extreme Aspirating* condition, and then the word ending, which was the same ending that was used for the given word in the *Extreme Aspirating* condition.

3.1.3 Procedure

Each participant completed the same set of tasks in the same order with a laptop and Audio-technica ATH-ANC7b headphones in a laboratory environment (*Table 2.1*). The entire session was recorded in a sound booth with a TASCAM DR-40 PCM recorder and an audio-technica AT8531

lapel microphone clipped on the participant's clothing, sampled at 44.1 kHz. The experiment was presented to participants through Psychopy 3.0 (Peirce, 2007), and in the form of a paper based Rating sheet and Sociolinguistic questionnaire. In Psychopy, the trials were presented with white letters on a gray background.

	Instruction	**Stimuli**	**Example**
Rating	Rate speaker for properties	semantic differential scales (10)	shy–talkative
PRE-Labeling	Select the word you hear	word, audio on a VOT continuum (11*10)	*binning* / *pinning* with 45 ms VOT
Familiarization	Read the word silently	written word (40*1)	*basin*
PRE-Read	Read the word out loud	written word (40*2)	*poser*
Shadowing	Repeat the word you hear	word, audio, 1 of 2 conditions (30*6)	*pollen* with 15 ms VOT
POST-Read	Read the word out loud	written word (40*2)	*buzzer*
POST-Labeling	Select the word you hear	word (audio on a VOT continuum (11*10)	*binning* / *pinning* with 75 ms VOT
Questionnaire	Fill in the questionnaire	—	Age:

Table 3.5: Procedure of the experiment

The experiment was carried out as described in *Section 2.6*, it is also recapped in *Table 3.5*. First, participants completed a *Rating* task, where they listened to the model talker read out a passage about various things to consider when shopping for a mattress and subsequently rated her on 9 semantic differential scales, which almost exactly followed Bauman's (2013; see *Table 3.6*). Participants also rated on a 1-to-9 Likert scale how attractive they found the model talker personally (from *not at all* to *very*). During the passage, the model talker's image was shown on the screen.

Solidarity	**Superiority**	**Dynamism**
friendly — unfriendly	organized — disorganized	shy — talkative
dishonest — honest	lower status — upper status	unsure — confident
rude — polite	intelligent — unintelligent	energetic — lazy

Table 3.6: Semantic differential scales

Then participants completed a labeling task (*PRE-Labeling*) with items corresponding to 11 equidistant steps of a VOT continuum, described in *Section 3.1.2.2*. Each of the 11 items were presented to every participant 10 times in randomized blocks (110 decisions/participant).

After labeling, participants were familiarized with the words used in the shadowing and reading tasks, where the total set of 40 stimuli words (from *Tables 3.3–3.4*) was presented individually. Participants were asked to first read the words without saying them out loud (*Familiarization*). They were then shown the 40 words one-by-one again in the same manner, but this time they also had to say the words out loud. This was repeated twice, which yielded 80 baseline productions (*PRE-Read*) from each participant.

After this, the experiment contained 6 blocks of *Shadowing* (Goldinger, 1998), where participants heard recordings of words fitting one of the two conditions (either from the *Extreme Aspirating* or the *Extreme Prevoicing* condition from *Figure 3.3*), and participants were asked to identify and repeat the English words they hear. While the audio was playing, a picture of the model talker was showing on the screen, which was replaced with a short instruction ("Repeat") after the audio was done. The words in the task were a semi-randomly selected subset of 30 words

selected from the 40 words in *Tables 3.3–3.4*. Each participant heard 8 monomorphemic b-words, 7 polymorphemic b-words, 8 monomorphemic p-words, and 7 polymorphemic p-words—a different set of 10 items were withheld from each participant. These productions provided the *Shadowing* data, 6 recordings per participant for each of the 30 semi-randomly chosen words.

Then the reading out loud task was repeated twice per word every word (even withheld ones, i.e. 2×40=80 tokens per participant). These provided *POST-Read* values that can be compared with the *PRE-Read* values to measure non-immediate effects of accommodation. The labeling task was also repeated *POST-Labeling*. These data can be compared with the *PRE-Labeling* data to detect any changes in categorization.

Lastly, participants filled out a sociolinguistic questionnaire, recording their age, gender identity and linguistic history, including their native language, place of birth, languages they are or were regularly exposed to, what environment, and for how long, all in a free-form answer. Participants also gave another assessment of how attractive they found the model talker personally on a 1-to-9 Likert scale. Participants were encouraged to provide their reasons for this assessment and to leave any general comments about the study they wanted to.

3.2 Data processing and analysis

Labeling data were obtained automatically from Psychopy. Because of the nature of the task, all 9,020 observations collected from non-excluded participants could be kept in the analysis. The reading and shadowing data were hand-annotated by me ($\sim$75%) and a trained undergraduate research assistant ($\sim$25%). I double-checked about half of these for consistency. I was responsible for making all decisions regarding the exclusion of tokens or the annotation of atypical realizations.

In the reading task 3,282 /p/ tokens and 3,284 /b/ tokens were recorded from non-excluded participants. In *Extr. Asp.* 1 /p/ and 0 /b/ tokens had to be excluded, and in *Extr. Prev.* 8 /p/ tokens and 5 /b/ tokens had to be excluded. These occurred mostly because of the participant skipping

the word (not reading out anything when prompted) or yawning. As a result the final datasets were made up of 1,681 /p/'s and 1,680 /b/'s in *Extr. Asp.* and 1,592 /p/'s and 1,599 /b/'s in *Extr. Prev*.

In the shadowing task, 3,693 instances of /p/ and 3,692 instances of /b/ were recorded. Of these, 3 tokens of /p/ and 15 tokens of /b/ had to be excluded in *Extr. Asp.*—mostly because of the participant yawning through or skipping words. In comparison, in *Extr. Prev.* 106 tokens of /p/ and 11 tokens of /b/ had to be excluded. Most of the /p/ tokens had to be excluded because in response to hearing a p-word with a plain /p/ (15 ms VOT) they produced the wrong word that was not on the list (e.g. *fooling* or *booling* instead of *pooling*). These tokens were most often accompanied with incredulous, questioning intonation, indicating hesitation or uncertainty about the lexicality of a given form. The intonational pattern suggests that participants were surprised by encountering a lexical item that was unfamiliar (e.g. *booling*) or unexpected within the context (because *fooling* was not on the list of words they encountered in the reading task, that they were told they were hearing words from during shadowing). Thus, these tokens should be considered tokens of /f/ and /b/ rather than of /p/ and were excluded from the /p/ dataset. Because the b-counterpart of the p-initial stimuli were always nonce words (the words were chosen to *not* form a minimal pair with the other labial stop e.g. #/ˈbulŋ/) these tokens were not included in the /b/ dataset either. Even though these tokens could not be used in the analysis, the frequency of such "errors" indicates that the plain [p] tokens were often misperceived as a /b/ by native English speakers.

The 9 ratings collected on the model talker's likeability were transformed into 3 measures. For every participant, likeability data was averaged into three measures. Ratings along the *friendly–unfriendly*, *dishonest–honest*, and *rude–polite* scales were averaged and called ***Solidarity***. *Organized–disorganized*, *lower status–higher status*, and *intelligent–unintelligent* ratings were averaged as ***Superiority***. The average of *shy–talkative*, *unsure–confident*, and *energetic–lazy* provided the ***Dynamism*** measure.

Attraction ratings will be excluded from subsequent analysis because of ambiguity of interpretation. The question was worded as 'How attractive do you find her personally?' with the intention to learn how strong of an attraction the participant actually experiences toward the model talker. However, there was an alternative interpretation of 'How likely would you be to describe her as "attractive"?'—i.e. how conventionally "attractive" do you think she is. Based on the responses to the open ended question ('Do you have a guess as to why [you rated her X]?') some participants interpreted the question one way and some the other. For instance, someone rated her 7 on the attraction scale, but in his verbal response he said "I'm not attracted to women, but would never call her ugly.", which indicates that the rating of 7 is a result of interpreting the question the second way. Since the open-ended question was not always answered or the answer provided did not disambiguate between which interpretation the participant followed, the data cannot be used to investigate the relationship of accommodation and attractiveness.

3.3 Reading results

In this section I will discuss the reading results. First, I will provide a preview of the results, discuss the statistical methods that will be used, and then give a reminder of what to focus on within the results, which aspects of the data can answer which of our questions laid out at the beginning (*Section 3.3.1*). After that the *Extreme Aspirating* condition and the *Extreme Prevoicing* condition will be discussed in more detail (*Section 3.3.2* and *Section 3.3.3*, respectively). The section will end with a brief summary of the reading results.

3.3.1 Overview and statistical methods

The experiment with native English speakers found that participants converged with exaggerated productions of English voiceless labial stops (very aspirated /p/'s). Moreover, they also converged with plain voiced stops (/b/'s with 15 ms VOT) to the extent that their productions at the beginning of the task do not already match them. At the same time, barely any convergence was observed in the *Extreme Prevoicing* condition. Neither the plain /p/ (15 ms VOT) nor the very prevoiced /b/ (130 ms prevoicing) of the model talker were imitated in the reading task.

Participants often produced the category /b/ with a bimodal distribution both before and after exposure—i.e. as a combination of plain [b]'s and [b]'s with substantial prevoicing, and participants rarely produced [b]'s with short prevoicing. Changes in /b/ productions (whether in the direction of more or less prevoicing) were largely categorical: participants changed the proportion of plain vs. prevoiced /b/'s. However, some participants changed their /b/'s in a gradient way—i.e. they shifted the range of prevoicing of their prevoiced tokens). *Figure 3.5* shows a short overview of what effect `Exposure` had (if any) in the two conditions by segment (without any measures of likeability for now).

Extralinguistic factors had limited effects on accommodation in this experiment. Likeability effects could not be established with this population in either condition. Individual variation was seen across accommodation behaviors—participants with similar starting distributions reacted to the stimuli in different ways, which could not be explained by the likeability data. In terms of `Gender`, no effects could be established. Male and female participants showed no statistically significant differences in their accommodation behavior. They were equivalent in terms of amount converged or how often they converged or diverged to the model talker. Effects of racial background could only be found to a small extent. Participants in the *Extr. Asp.* condition who identified as black or African American tended to have more prevoiced /b/'s before exposure, and when exposed

to plain /b/'s they tended to converge more. This might be a result of simply having more room to demonstrate convergence rather than an increased predisposition to accommodation in this situation.

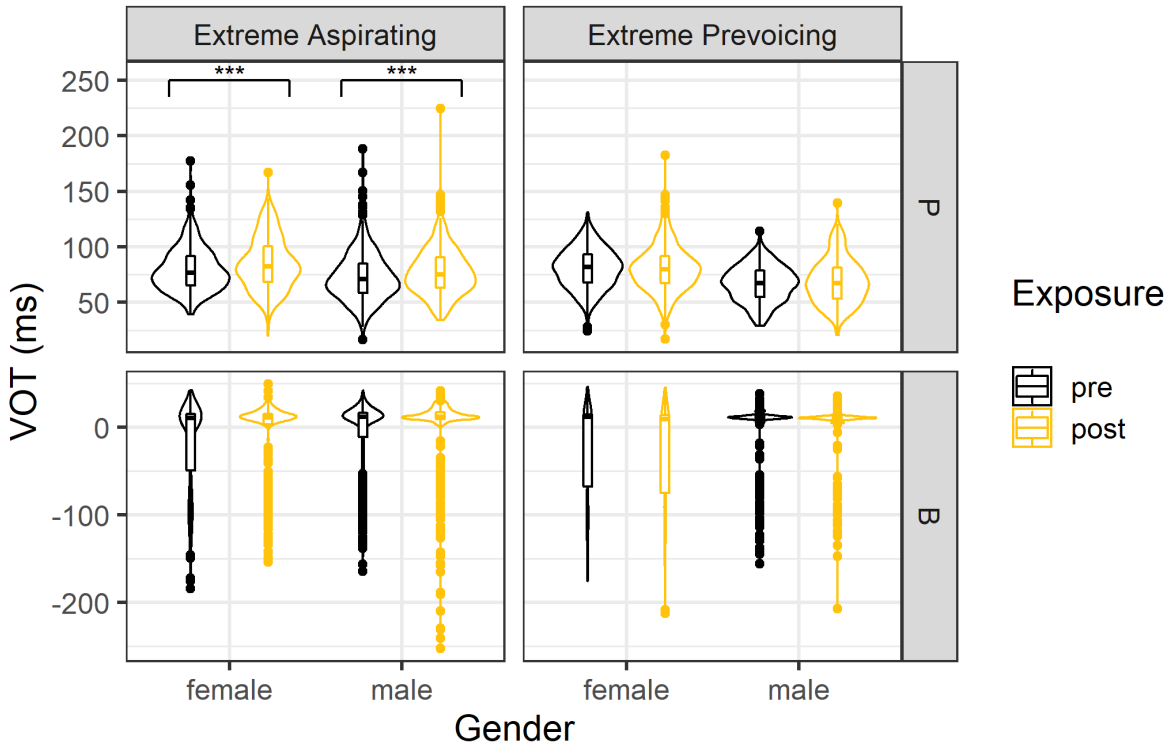

Figure 3.5: Effect of `Exposure` *in the reading data from English-speaking participants*
Note the different VOT axes for /p/ and /b/

Descriptive statistics

In the *Extreme Aspirating* condition there were a total of 1,681 read /p/ tokens and 1,680 read /b/ tokens, while the *Extreme Prevoicing* condition had 1,592 /p/'s and 1,599 /b/'s. This section provides brief descriptive statistics of the participants' baseline reading productions, as assessed by the PRE-Reading task. Since these were recorded prior to exposure to the modified stimuli, these productions reflect natural English productions most closely (with the significant constraints of word-list reading tasks).

Participants' pre-exposure /p/ productions had a mean of 75.588 ms (long lag VOT), and a median of 73.834 ms, which indicates that the distribution was unimodal and quite close to a normal distribution. Standard deviation was 19.932 ms. This dataset was analysed in a linear mixed effect model (with `lmer` in R 3.6.1) where the dependent variable was VOT (in ms), and the independent variables were `Gender` and `Condition`, with a random by-word intercept. No by-participant intercept was added in order not to absorb any potential `Gender` effects.

The model indicates that males in both conditions had a lower VOT than females, especially in *Extr. Prev.* where males had even shorter aspiration than males in *Extr. Asp.*. However, as *Figure 3.6* indicates, the difference between males was likely due to the long VOT productions of M01. When his tokens are excluded from the model, we no longer see effects differentiating between the two conditions, and the difference between male and female participants also goes away when corrected for the number of tests run.

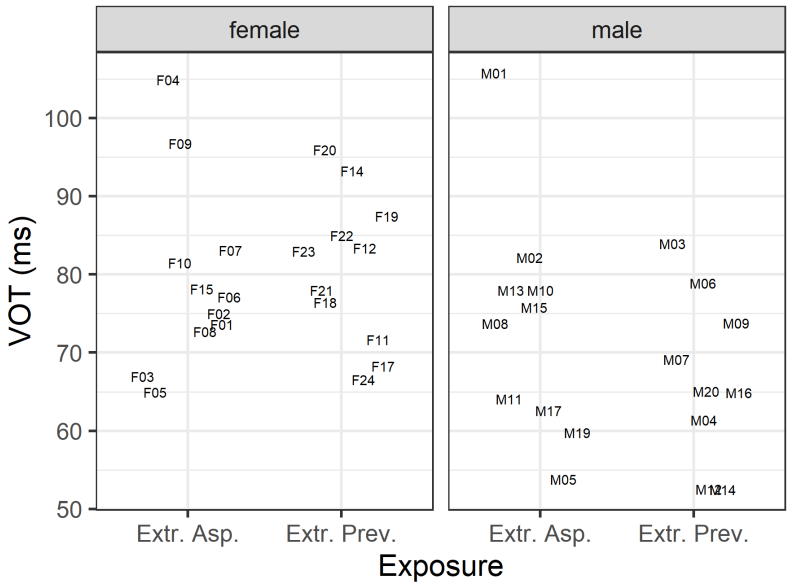

Figure 3.6: Pre-exposure /p/ reading data from English-speaking participants
By-participant means by gender

According to the model, males produced /p/'s with 9.751 ms shorter VOT's than females in both condition, but this effect goes away once correcting for the number of tests run. Visual

inspection also supports a systematic difference between males and females (*Figures 3.6–3.7*). Females across the two conditions did not differ significantly (p=0.7803), nor did males (p=0.5451).

	Estimate	Std. Error	Pr(> |t|)	
(Intercept)	79.515	3.496	<0.0001	***
Gender [male]	−9.751	4.787	0.0491	.
Condition [Extr. Prev.]	1.276	4.541	0.7803	
Gender [male] × Condition [Extr. Prev.]	−4.136	6.770	0.5451	

Table 3.7: LMER model of reading /p/ tokens' VOT (in ms) before exposure without the outlier M01; Threshold for significance (adjusted with Bonferroni correction): p<0.0125

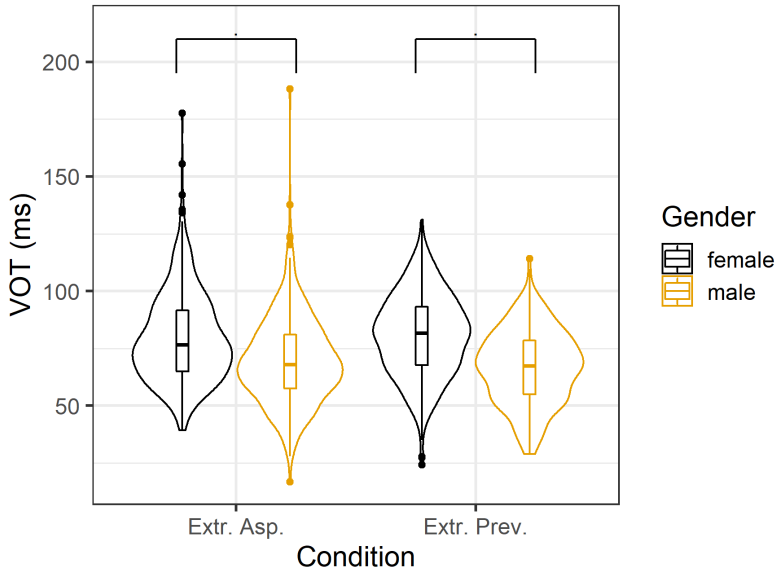

Figure 3.7: Pre-exposure /p/ reading data from English-speaking participants Without the outlier, M01

The /b/ tokens are shown in *Figure 3.9*. Productions of /b/ had a mean of -10.795 ms (slightly prevoiced), but values similar to the mean were not often attested. The median was 11.457 ms (short lag VOT), which suggests that the mean being slightly prevoiced was a result of a mixture of plain and prevoiced tokens. The standard deviation is quite large too (47.679 ms), which similarly suggests

a bimodal distribution. *Figure 3.8* shows each participant's pre-exposure mean by condition and gender.

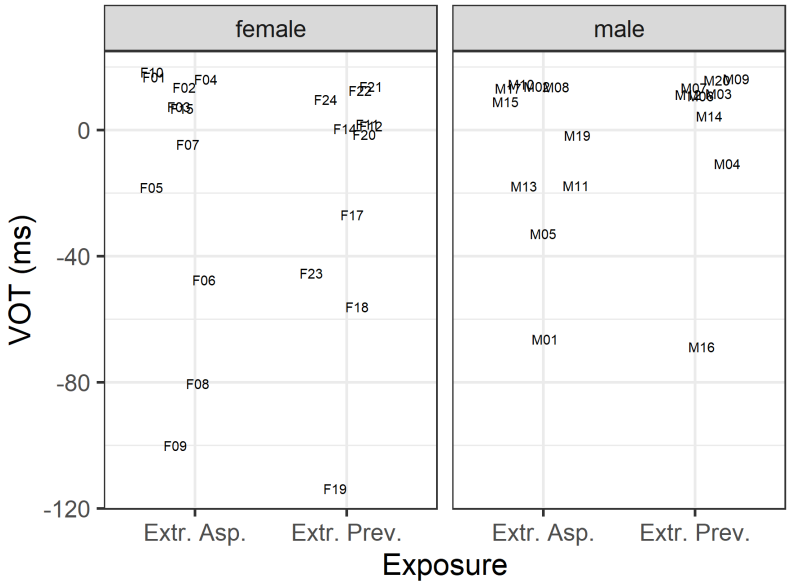

Figure 3.8: Pre-exposure /b/ reading data from English-speaking participants
By-participant means by gender

Across the two conditions there were a total of 391 prevoiced /b/ tokens and 1249 plain ones, and looking at their properties separately reveals a truly bimodal distribution. The plain /b/'s were quite concentrated (mean: 13.993 ms, median: 12.735 ms, SD: 5.751 ms), while most of the prevoiced tokens had a substantial amount of prevoicing, but with a lot of variation (mean: -89.980 ms, median: -91.746 ms, SD: 34.551 ms).

Statistical analysis of the entire pre-exposure /b/ dataset (again, without a by-participant random intercept) shows only a `Gender` difference in this case. Males had less prevoicing (β=8.370, p=0.0103) in both conditions, and the difference between males in *Extr. Asp.* and males in *Extr. Prev.* was only significant when not correcting for the number of tests run (p=0.0229). The pre-exposure difference between males in the two conditions will be revisited in the section discussing

the *Extreme Prevoicing* reading data specifically *Section 3.3.3*. *Figure 3.9* shows these results visually.

	Estimate	Std. Error	Pr(> |t|)	
(Intercept)	−15.567	2.267	<0.0001	***
Gender [male]	8.370	3.257	0.0103	*
Condition [Extr. Prev.]	−2.956	3.178	0.3524	
Gender [male] × Condition [Extr. Prev.]	10.639	4.672	0.0229	.

Table 3.8: LMER model of reading /b/ tokens' VOT (in ms) before exposure; Threshold for significance (adjusted with Bonferroni correction): p<0.0125

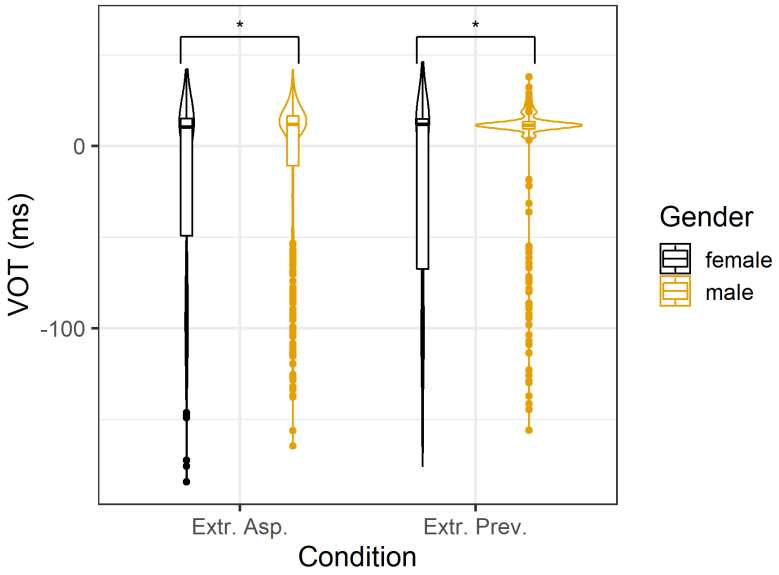

Figure 3.9: Pre-exposure /b/ reading data from English-speaking participants

Statistical analysis

The reading data were then analyzed in four separate linear mixed-effect regression models (one per segment-condition combination: *Extr. Asp.* /p/; *Extr. Asp.* /b/; *Extr. Prev.* /p/; *Extr. Prev.* /b/) using `lme4` in R. The dependent variable was VOT produced by the participant, and the independent

variables were `Gender` (*female* or *male*) and `Exposure` (whether the token was produced before or after the shadowing task, i.e. *pre* or *post* exposure). By-participant and by-word random intercepts were added. Not all the words participants read out were also featured in the Exposure/Shadowing phase. Adding a factor of whether the given word was included during shadowing did not improve the model's fit and was therefore omitted from future analyses (χ^2 test, p=0.7290 for the /p/ dataset and p=0.1658 for the /b/ dataset), which replicated previous findings (e.g. Nielsen, 2008) of accommodation generalizing over lexical items.

Foci of attention

So far in this section I described the statistical models that will be used to analyze the reading data, and gave a brief overview about the nature of the PRE-Read data set. This was done in order to provide some context about the baseline productions of the participants, which were all comparable to previous research. In this subsection, I will offer a reminder of what specific questions we are seeking to answer with the reading data. These will be discussed in three groups: the main question, linguistic questions, and extralinguistic (socio-linguistic) questions.

First, the main question is whether participants can shift their productions towards a type of contrast that is unlike that of their native language. To answer this, we need to look at *Extr. Prev.* /p/'s, and whether exposure to a plain-prevoiced contrast can make participants shift their /p/'s towards a more short-lag realization. Examining the *Extr. Prev.* /p/'s can help us disambiguate between the **Maintain categories** hypothesis, which predicts no convergence, and the **Maintain contrasts** hypothesis, which does.

Second, the reading data can help answer some questions related to linguistic variables and methodology. For instance, do members of a voicing contrast move together? We have evidence of participants generalizing from a segment to a natural class sharing the given cue (e.g. exposure to /p/'s affects productions of not only /p/ but /t k/ as well), but do people generalize to the use of

an entire phonetic dimension, not just the cue? This is especially relevant in *Extr. Prev.*, because shifting /p/ towards less aspiration means moving their /p/ closer to /b/ (potentially have /p/ encroach on /b/'s acoustic space). Do participants who converge with a plain /p/ also move their /b/, does accommodation behavior along /p/ predict accommodation behavior along /b/? Furthermore, do accommodation effects persist? While this requires a comparison of shadowing and reading data, the reading data can show whether there are accommodation effects that outlast direct exposure.

Third, we can use this dataset to glean a better understanding of what effects certain extralinguistic variables have on accommodation. Is there a gendered component to convergence? Do we see a difference in male and female accommodation behaviors for *Extr. Asp.* /p/'s specifically? The setup here is similar to Nielsen's (2008) study, except the target is even further from the participants' normal productions (130 ms VOT), thereby giving even females room to demonstrate convergence. Moreover, data was also collected on ethnicity, which allows us to find any ethnicity-specific patterns in the dataset. Since the ethnic background of participants was so diverse, they will be discussed in a qualitative way along with holistic individual patterns rather than with statistical models. In addition, we can see what components of likeability (if any out of `Solidarity, Superiority,` and `Dynamism`) influence accommodation to a shadowed model talker most strongly. Separate models will be dedicated to this issue by condition and segment. As discussed above, attraction data will not be included because of the ambiguity inherent in the answers.

3.3.2 The Extreme Aspirating condition

Accommodation of /p/'s

The model on the *Extr. Asp.* /p/ reading data indicates convergence (*Table 3.9*): exposure to voiceless stops with an extremely long VOT (130 ms) resulted in significantly longer VOT in those stops after exposure (β=5.602 $p<0.0001$). While there was a trend of males having shorter VOT both before and after exposure, neither the main effect nor the interaction of `Gender` were significant in this

model (p=0.274 and p=0.487, respectively). While starting differences between male and female participants could have been absorbed by a by-participant random intercept, the lack of interaction between Gender and Exposure indicates that male and female participants accommodated in comparable ways (they all converged).

	Estimate	Std. Error	Pr(> |t|)	
(Intercept)	79.519	4.077	<0.0001	***
Gender [male]	−6.154	5.467	0.274	
Exposure [post]	5.602	1.158	<0.0001	***
Gender [male] × Exposure [post]	−1.168	1.680	0.487	

Table 3.9: LMER model of reading /p/ tokens' VOT (in ms) in the Extr. Asp. condition; Threshold for significance (adjusted with Bonferroni correction): p<0.0125

On an individual level, almost every speaker got closer to the target 130 ms VOT (i.e. lengthened their VOT), and the effect was not demonstrably enhanced or dampened by any of the three likeability measures that were examined (ratings that were averaged into one rating for Superiority, Solidarity, and Dynamism each per person). Rather, convergence was an overall tendency that almost every participant in this condition exhibited (*Figure 3.10*).

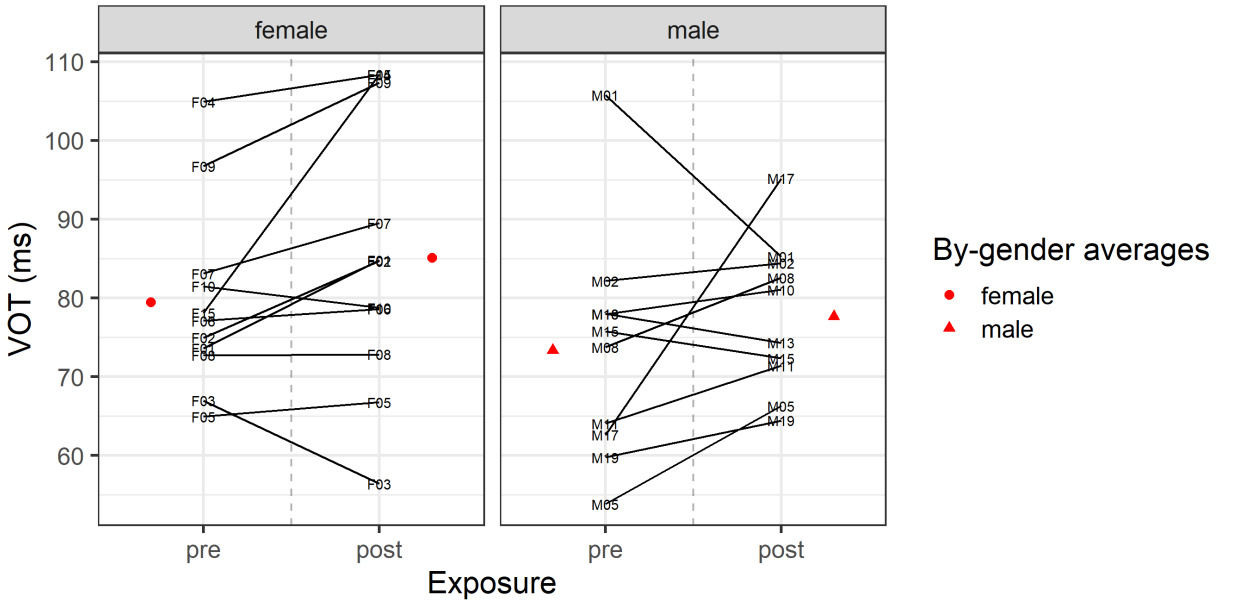

Figure 3.10: Change in mean VOT of /p/ in Extr. Asp. with by-gender averages

Since combining the different likeability measures in one model often resulted in a failure to converge, three separate linear mixed-effect regression models were created, combining `Condition`, `Gender`, `Exposure`, and one of the three likeability measures (`Superiority`, `Solidarity` or `Dynamism`) as independent variables. The dependent variable was duration of VOT (in ms; positive for aspiration, negative for prevoicing), and by-person and by-word random intercepts were also added. While there did seem to be an effect of `Superiority` (*Table A.1* in the Appendix), this seemed to be a result of over-fitting for the behavior of two participants (*Figure A.3* in the Appendix). `Solidarity` and `Dynamism` did not improve the model's fit. Their statistical tables and figures are in the Appendix in full detail.

Not much of a connection could be found between ethnicity and either read /p/ VOT's or accommodation of VOT in the *Extr. Asp.* condition (*Figure 3.11*). While white males had longer VOT's than males of other ethnicities, this observations is only based on 3 white male speakers, and therefore could easily be accidental.

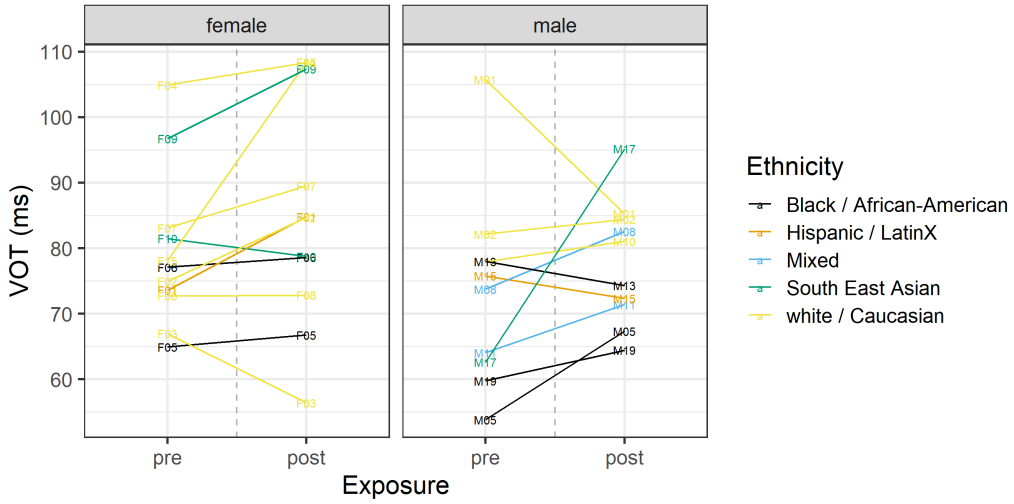

Figure 3.11: Reading performance for /p/'s in Extr. Asp. by ethnicity
Model talker's VOT: 130 ms

Accommodation of /b/'s

At first glance, there does not seem to be any accommodation for /b/ in the *Extr. Asp.* reading task (target: 15 ms VOT). When the entire dataset is tested, we see no effects, including no effect of `Exposure` (p=0.0409, which does not meet the Bonferroni corrected significance level of α=0.0125). However, the 15 ms target VOT is quite typical for an English [b], and indeed, a lot of participants were around the eventual target to even in their pre-exposure reading task at the start of the experiment (*Figure 3.12*). These participants would not be able to demonstrate convergence (they could not possibly go any closer), and thus might mask the behavior of those who did have room to converge. *Figure 3.12* reveals a trend, where participants whose mean was near the target diverged or did not change their productions, and those whose pre-exposure mean was further away from the target (/b/ with 15 ms VOT) mostly converged to it. This trend was not universal, with some prevoicers not changing their productions (e.g. F05, F07, F08—all females), and some even diverged (e.g. M11, M13, M17—all males).

Therefore, one possible reason for the lack of statistically significant convergence so far could be that the participants who were already matching the target /b/ (a plain unaspirated stop)

even before exposure muddled any effect coming from the rest of the group. To limit the "on target" participants from creating a ceiling/floor effect, the dataset was restricted to the 13 out of 21 participants whose pre-exposure mean was at least 5 ms away from the target (i.e. participants whose pre-exposure mean B was 10–20 ms). Even though participants were selected to be simply *further away* from the target in either direction, everyone's mean in this dataset was *below* the target rather than above it—i.e. no one had a mean production for B over 20 ms. The highest mean in this group was 8.93 ms.

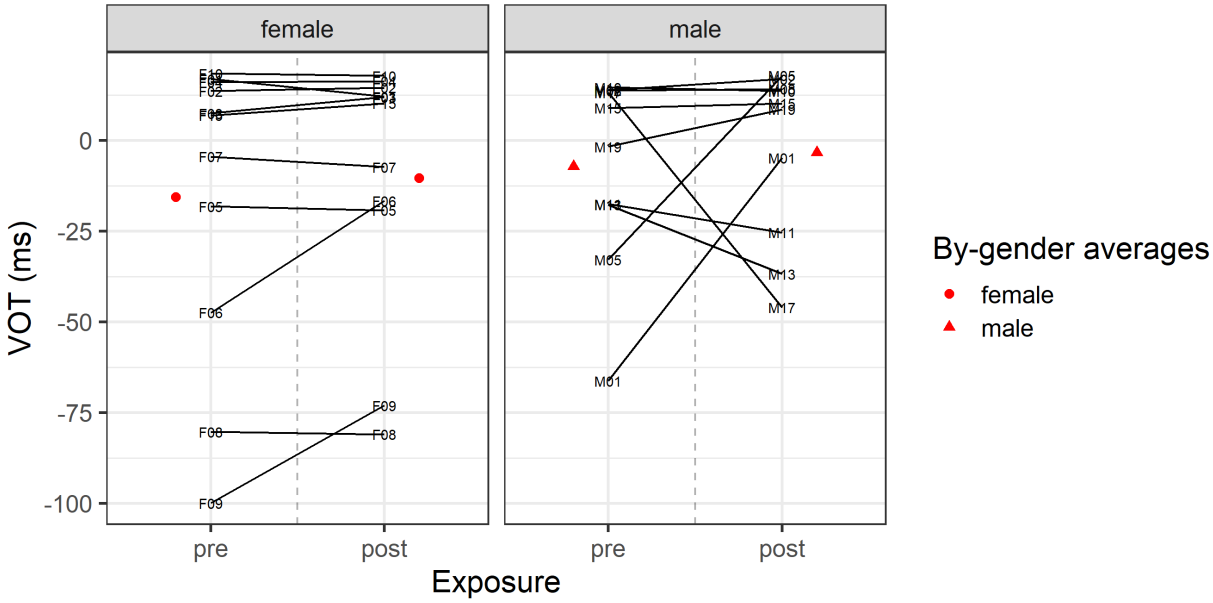

Figure 3.12: Change in mean VOT of /b/ in Extr. Asp. with by-gender averages

When the model was run on this dataset (N=1,039 from 13 out of 21 participants; *Table 3.10*) significant convergence was found, suggesting that while some participants were "on target" the whole time, those who had room (participants who produced at least some prevoiced tokens) converged to a plain /b/, which is more common in English. It is also interesting to note, that the model could not establish a baseline Intercept, which indicates that the distribution of B productions was not unimodal—indeed /b/'s formed a bimodal distribution with two peaks around plain stops

and substantially prevoiced stops (*Figure 3.13*). There does not seem to be any gender-grading of accommodation patterns in this case either (p=0.1411).[2]

	Estimate	Std. Error	Pr(> |t|)	
(Intercept)	−33.740	12.191	0.0175	.
Gender [male]	12.518	17.872	0.4977	
Exposure [post]	8.697	3.455	0.0120	*
Gender [male] × Exposure [post]	7.488	5.084	0.1411	

Table 3.10: LMER model of reading B tokens' VOT (ms) in Extr. Asp., participants who "had room" to converge; Threshold for significance (adjusted with Bonferroni correction): p<0.0125

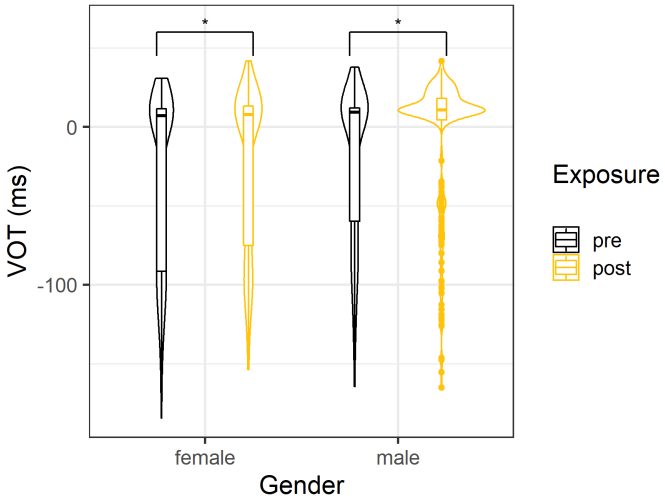

Figure 3.13: VOT of read /b/'s in the Extr. Asp. condition, before and after exposure by gender Restricted to participants who had room to accommodate

Just like in the case of *Extr. Asp.* /p/'s, the effect of `Exposure` to plain /b/ does not seem to be mediated by any of the three likeability measures. These analyses are only based on the 13/21 participants (1,039 tokens) whose pre-exposure mean was further away from the target than 5 ms. While some effects can be found, those seem to be due to a confound of the ratings the participant

[2]From this point on, plots summarizing all four models of reading data will feature a plot with just this subset of participants for *Extr. Asp.* /b/'s rather than the entire *Extr. Asp.* /b/ dataset in *Figure 3.5*.

gave to the model talker and their original phonetic distance from the model talker's productions. The statistical tables and figures illustrating the data can be found in the Appendix (*Tables A.4–A.6* and *Figures A.6–A.10*). Since in these cases there are potential confounds between a participant's baseline VOT and their accommodation behavior, some of these plots show by-participant raw values rather than just the amount of change in VOT productions.

Unlike for /p/, where no connection was found between ethnicity and VOT, /b/ VOT's were contingent on ethnicity. Most prevoiced mean VOT's for /b/ came from people of color *Figure 3.14*. This was especially true of Black / African-American participants, all of whom produced enough prevoiced tokens to bring their pre-exposure mean VOT into decidedly the negative VOT range. This corroborates Ryalls et al.'s (1997) results, where Black participants had longer prevoicing on their voiced stops than their white participants did in a word reading task. This finding was based on a sample of ten 20–30 year-old participants per ethnic group (N=20 in total), but not found in older speaker's productions (Ryalls et al., 2004). Since the population used in this study is college-aged, these results are consistent with both Ryalls et al. (1997) and Ryalls et al. (2004).

Moreover, some of the biggest instances of convergence also came from Black / African-American participants. This is understandable, since by producing a lot of prevoiced /b/'s as a baseline (in the PRE-Read), they had more of an opportunity to demonstrate convergence (drop prevoicing) than other participants who were meeting the 15 ms VOT target to begin with. At the same time, not all Black / African-American participants converged (see M13).

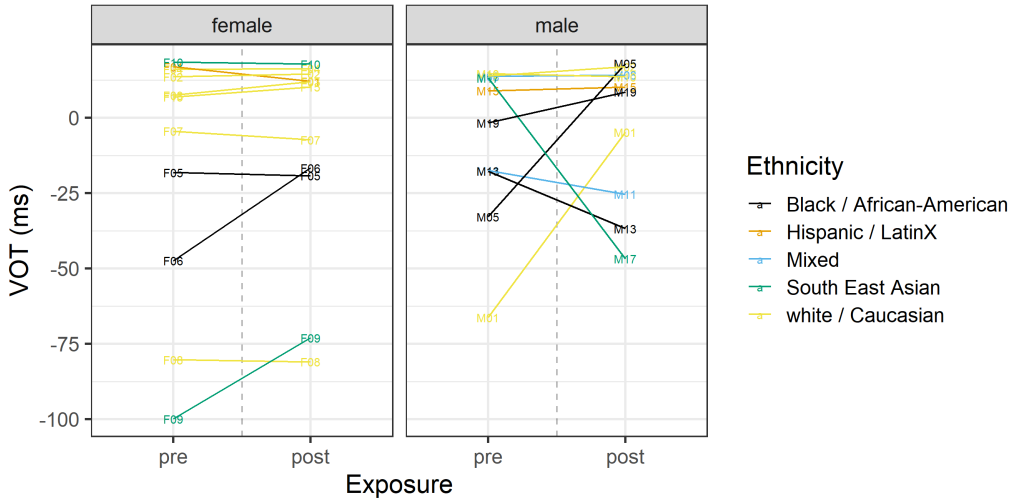

Figure 3.14: Reading performance for /b/'s in Extr. Asp. by ethnicity
Model talker's VOT: 15 ms

Patterns in the treatment of the /p b/ contrast in the Extreme Aspirating reading data

In the previous sections we saw that most participants converged with the extremely aspirated /p/'s and plain /b/'s of the model talker (130 ms VOT and 15 ms VOT, respectively). In this section, we will compare /p/ and /b/ accommodation and see whether the treatment of these two sounds matched up on an individual level.

Accommodation observed in /p/-words was slightly different from that observed in /b/-words. Overall, more speakers changed their /p/ productions than did their /b/ productions, but the handful of people who changed their /b/'s made bigger changes than were observable for /p/. *Figure 3.15* shows how a participant's /p/ accommodation behavior compared to their /b/ accommodation behavior. The x axis shows if the participants shifted the mean VOT of their /b/ productions in a positive or negative directions. If their /b/'s became less prevoiced, that is a positive change (e.g. a change in means from -60 ms to -45 ms would translate into a positive change of 15 ms) and an increase in prevoicing is a negative value. The y axis shows how their /p/ productions changed. More aspiration means a positive value on the y axis (e.g. a change from 55 ms to 75 ms is a positive change of 20 ms), and a decrease in aspiration is negative on the y axis.

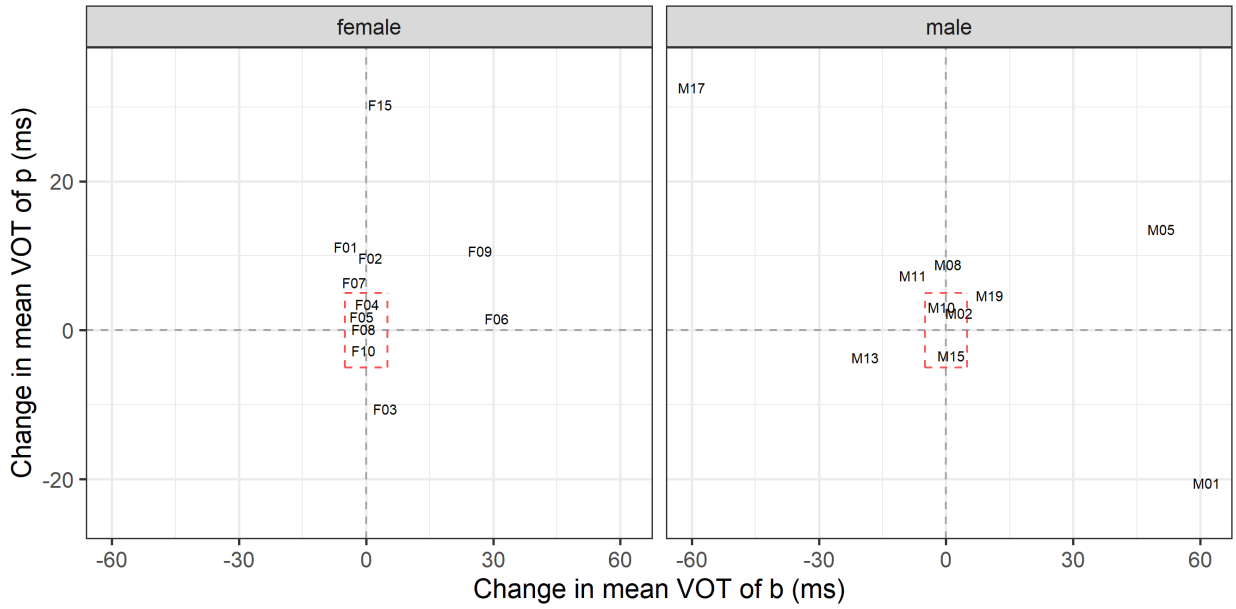

Figure 3.15: Change in mean VOT of /p/ and /b/ in Extr. Asp. per person;
The red rectangle shows 5 ms change of means in either direction for reference

In *Figure 3.15*, most datapoints have a non-zero value on the y axis, but the biggest excursions from zero are along the x axis. This essentially means that aspiration as a cue was frequently adjusted in smaller increments, while prevoicing as a cue was less often changing, but when it did, it involved larger changes.

If a participant's accommodation behavior for the two phonemes was correlated then they would be in the top right or the bottom left sections of *Figure 3.15*—i.e. either they changed both their /p/ and /b/ in a positive direction (top right; less prevoicing/more aspiration for /b/, more aspiration for /p/), which in this condition means convergence, or changed both sounds negatively (bottom left; more prevoicing/less aspiration for /b/, less aspiration for /p/), which in *Extr. Asp.* means divergence.

Participants do not all fall into these two quadrants. A participant's accommodation behavior for the two phonemes was not necessarily correlated. Even among the 13 participants whose /b/

baseline was at least 5 ms from the target, no correlation was found between /p/ accommodation and /b/ accommodation (Pearson's r, p=0.5605), and participants followed different trajectories.

All in all, in the *Extreme Aspirating* condition we saw that participants converged with the model talker's extremely aspirated (130 ms VOT) [p]'s as well as her [b]'s (15 ms VOT) – provided they were not hitting the target even without exposure. Changes to the realization of /p/'s were more common, but also on a smaller scale than changes of /b/'s, which while rarer were also bigger changes. On an individual level these observations translated into various strategies, attesting the individual variation so often described in the literature. The most common patterns were however accommodation and /p/ accommodation (for participants whose /b/ productions matched the target). While some participants did move their /p/ and /b/ categories closer to one another, the categories never overlapped. This is not surprising, since this condition is an exaggerated version of everyday English, in which /p/ and /b/ form a neat bimodal distribution word-initially. This study could not establish a robust relationship between accommodation and any of the three likeability measures. While no ethnicity-related effects were seen for /p/, most prevoicers for /b/ were people of color. This was especially true for Black / African-American participants, more of whom converged compared to white / Caucasian participants.

3.3.3 The Extreme Prevoicing condition

While participants in *Extr. Asp.* largely exhibited convergence towards the target (unless they were on target to begin with), in *Extr. Prev.* participants were less likely to adopt the model talker's /p b/ contrast, which was less like English (/p/:15 ms VOT, /b/: 130 ms of prevoicing). Overall, no significant effects were found in the baseline analysis (without likeability ratings) of the dataset for either /p/ or /b/, but there was a relationship between `Solidarity` ratings and /p/ accommodation.

Accommodation of /p/'s

The model run on the *Extr. Prev.* /p/ dataset from the reading task (*Table 3.11*) found no convergence or divergence for either gender (p=0.962 for the *female* baseline, and p=0.179 for an interaction for *males*). There was an almost significant effect of `Gender` indicating that males had on average 13.887 ms shorter VOT's than females both before and after exposure, but this effect goes away after correcting for the number of tests run (p=0.014 $>$ α=0.0125).

	Estimate	Std. Error	Pr($>$ |t|)	
`(Intercept)`	80.792	3.677	$<$0.0001	***
`Gender [male]`	$-$13.887	5.122	0.014	.
`Exposure [post]`	0.047	0.983	0.962	
`Gender [male]` $\times$ `Exposure [post]`	1.974	1.469	0.179	

Table 3.11: LMER model of reading P tokens' VOT (in ms) in Extr. Prev.; Threshold for significance (adjusted with Bonferroni correction): p<0.0125

These results are shown in *Figure 3.16* for all participants in *Extr. Prev.* combined and individually in *Figure 3.17*. There does not seem to be an overall tendency: some participants lengthen their VOT, some shorten it, some stay the same. These patterns do not correlate with the participants' starting VOT ('pre' column), and indivdual variation is exhibited by how participants starting from similar values can adopt different productions after exposure (consider F14 vs. F20, or M16 vs. M20). While *Figure 3.16* indicates that the median did not change for either gender group, in *Figure 3.17* we can see that if anything, the male group-level mean got further from the target (red triangles), though as the model indicates, this divergence is not significant either.

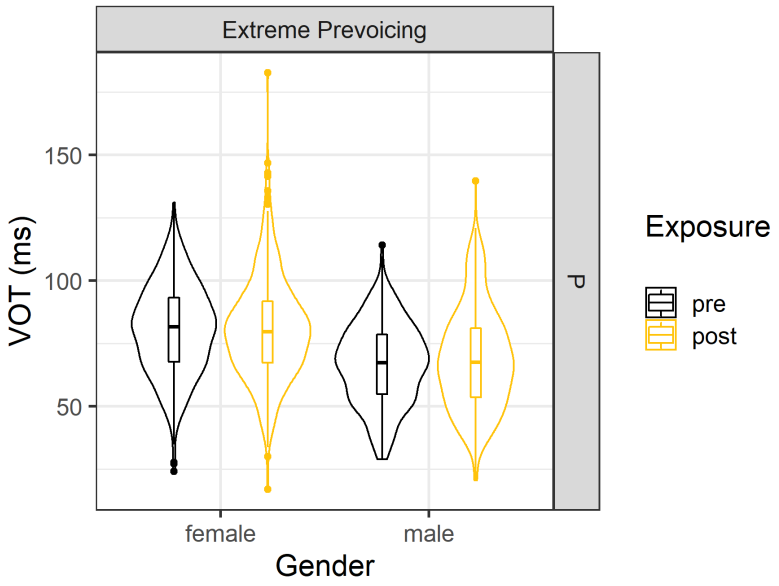

Figure 3.16: Effect of `Exposure` *in the reading data in Extr. Prev. /p/'s*

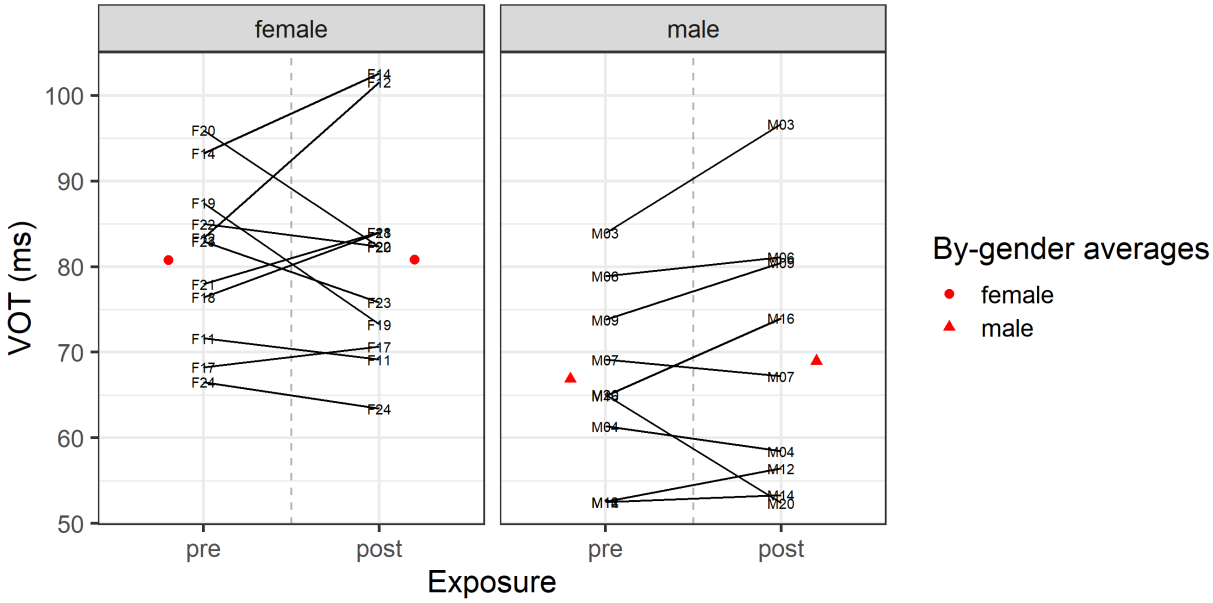

Figure 3.17: Change in mean VOT of /p/ in Extr. Prev. with by-gender averages

There were four speakers who substantially shortened their VOT of /p/ after exposure (*Table 3.12*). However, when compared with the people in *Extr. Asp.* there were neither more of them in *Extr. Prev.* who shortened their /p/ VOT's post-exposure, nor did the magnitude of the change

surpass that in *Extr. Asp.* In *Extr. Asp.* this was likely the result of fatigue or hyperarticulated (and thus unnaturalistically aspirated) baselines, as we will see later. However, if the changes in *Extr. Asp.* are results of a task effect, then effects observed in *Extr. Prev.* can hardly be claimed to be a result of convergence unless further evidence is presented.

Participant	**Condition**	**Difference**	**Pre-exposure**	**Post-exposure**
M01	Extr. Asp.	−20.6 ms	105.8 ms	85.2 ms
F03		−10.6 ms	67 ms	56.4 ms
F19	Extr. Prev.	−14.1 ms	87.4 ms	73.3 ms
F20		−13.7 ms	96 ms	82.3 ms
M20		−12.6 ms	65 ms	52.4 ms
F23		−7.2 ms	83 ms	75.8 ms

Table 3.12: Participants who shortened their mean /p/ VOT by more than 5 ms after exposure

Post-hoc analyses suggest that accommodation might have been contingent on likeability, specifically `Solidarity`, in this condition. `Solidarity` improved the model's fit (AIC of 13,160 vs. AIC of 13,185 in the baseline) without a trade-off in model complexity (BIC 13,219 vs, BIC 13,222; p>0.0001). The model with `Solidarity` (*Table A.7* in the Appendix) indicates an effect that was hidden in the baseline model. Participants giving the model talker a low `Solidarity` rating (rated her on the unfriendly, dishonest or rude ends of the respective spectra) tended to diverge (the baseline `Exposure` effect goes in the direction of more aspiration; β=23.554, p=0.0003), but this changes into convergence for participants who rate her higher (higher than 6.16 to be precise; `Exposure [post]` $\times$ `Solidarity`: β=−3.823, p=0.0003). Thus, someone rating the model talker high for `Solidarity`-related measures also produces shorter VOT values (i.e. converged with the model talker's 15 ms VOT /p/'s).

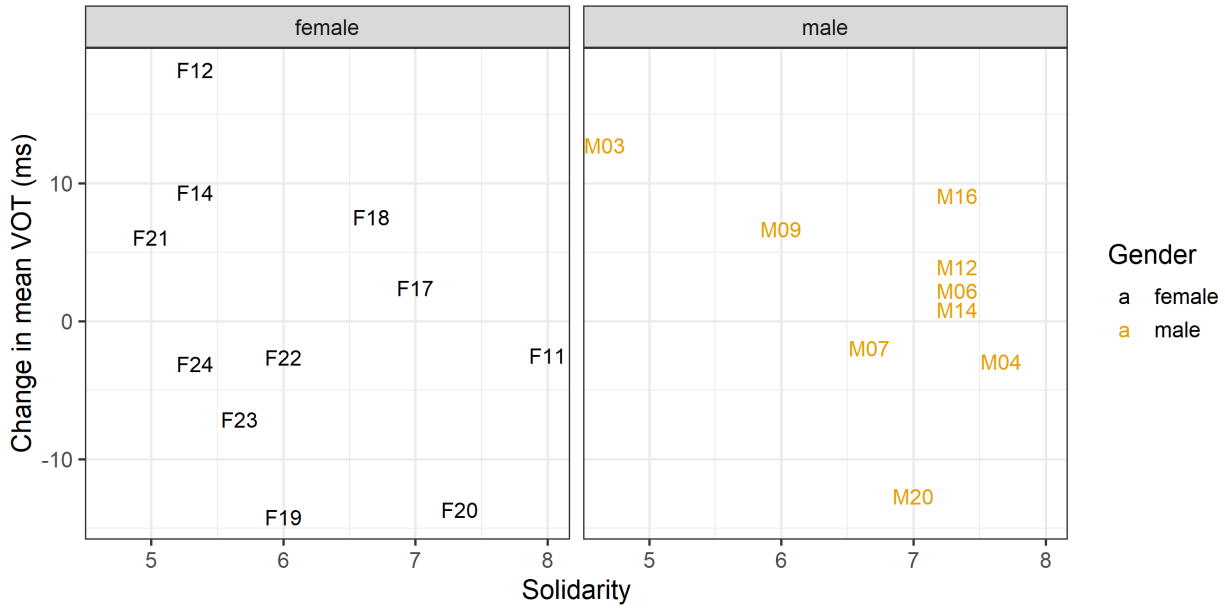

Figure 3.18: Change in mean /p/ VOT in Extr. Prev. in the reading task by gender and Solidarity rating

While there is undoubtedly a substantial amount of individual variation, visual inspection does confirm some correlation (*Figure 3.18*). This plot shows the relationship between the `Solidariy` ratings and /p/ accommodation. Each participant is represented as a single text label along the x and y axes, where the x axis represents the `Solidarity` rating they gave to the model talker, and the y axis represents what direction the participant changed their /p/ VOT after exposure to the audio stimuli. Participants who lengthened their VOT, whose VOT changed in a positive direction (i.e. convergers) are upwards on the y axis, and divergers (who shortened their /p/ VOT) are lower down the y axis, as they adjusted their VOT in a negative direction. For example, F12 is female (left half of the plot), and she gave the model talker a `Superiority` rating of 5.33, and after exposure her mean VOT was 18.2 ms longer than her pre-exposure average VOT—i.e. she diverged from the model talker.

There are also some reasons pointing in the direction that unlike in the *Extr. Asp.* dataset, these effects are not a result of overfitting for the dataset. First, `Solidarity` ratings have a wide and quite even distribution. Second, there is no confound between pre-exposure values and ratings

(e.g. those who rated her high on `Solidarity` happening to produce longer VOT's before exposure and thus potentially providing a hyperarticulated baseline; effect of `Solidarity` on VOT before exposure p=0.2269, also see *Figure 3.19*).

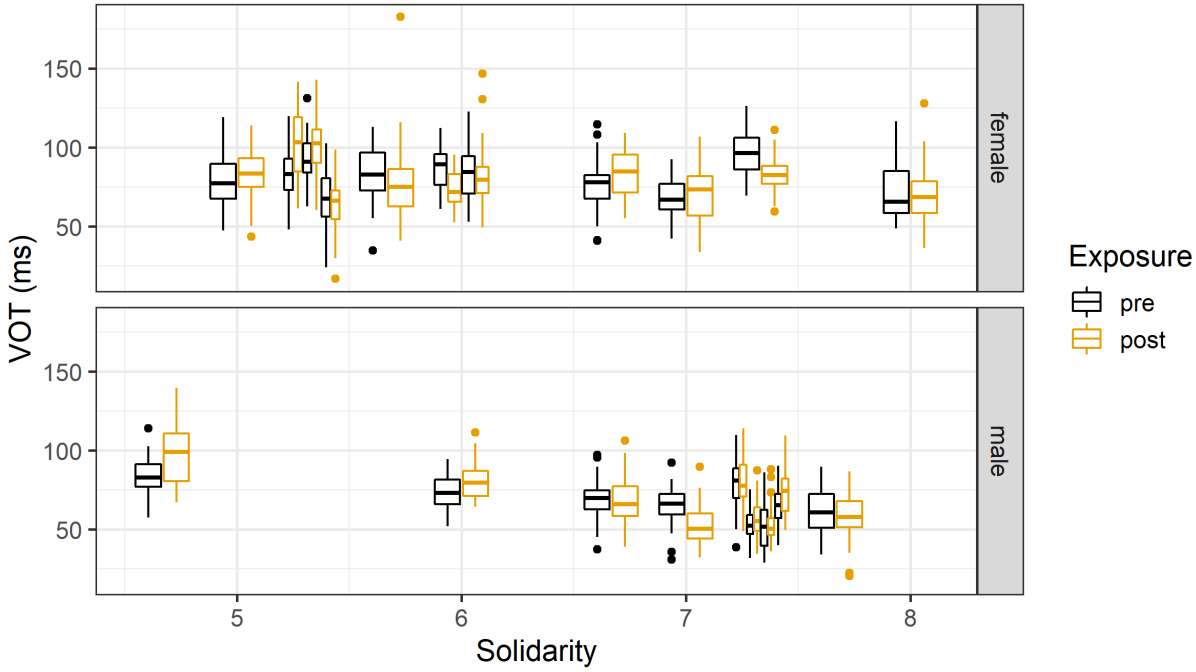

Figure 3.19: Change in mean /p/ VOT in Extr. Prev. in the reading task by gender and Solidarity rating

While some participants might have converged with plain /p/'s, about half of females and most males diverged, which tended to go with low ratings along `Solidarity`-related social scales (*friendliness*, *honesty*, and *politeness*). This increase in the amount of aspiration observed from participants who rated her low can most soundly be explained as divergence fueled by the participants' disliking of the model talker. Thus, the effect of *Solidarity* found in the data might be more of a connection between disliking and divergence, rather than between liking and convergence.

While there does seem to be a correlation between VOT's produced by speakers and `Solidarity` ratings (participants rating the model talker high on the scales corresponding to `Solidarity` converged with her, and those who rated her low tended to diverge from her in the post-exposure reading task) no such connection was found between accommodation and `Superiority`

or `Dynamism`. The statistical tables and figures can be found in the Appendix (*Tables A.8–A.9* and *Figures A.11–A.12*) No link was found between ethnicity and either VOT productions themselves or VOT accommodation behavior (*Figure 3.20*).

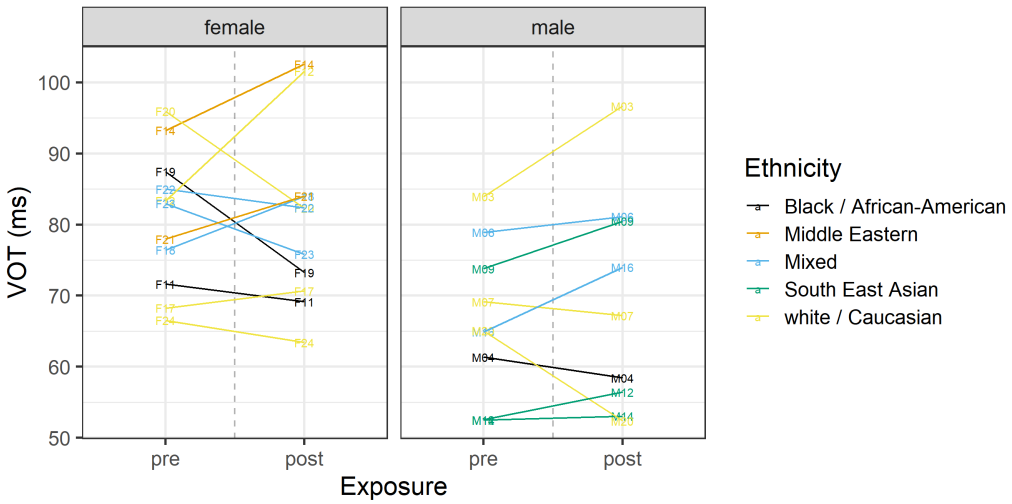

Figure 3.20: Reading performance for /p/'s in Extr. Prev. by ethnicity
Model talker's VOT: 15 ms

Accommodation of /b/'s

No socially independent accommodation can be found in the /b/ reading dataset either. *Figure 3.21* shows individual changes in mean VOT by gender. Most participants produced almost exclusively or exclusively plain /b/'s to begin with and those of them who change (e.g. M14) change towards more plain productions (i.e. diverge from the prevoiced target). Those who demonstrated an ability to produce substantial prevoicing (either in duration or in frequency) do not show a consistent pattern either. Some of them converge with the target (e.g. M04 and F14 do), but just as many of them diverge from it (i.e. produce less prevoiced tokens, like F19 and M16).

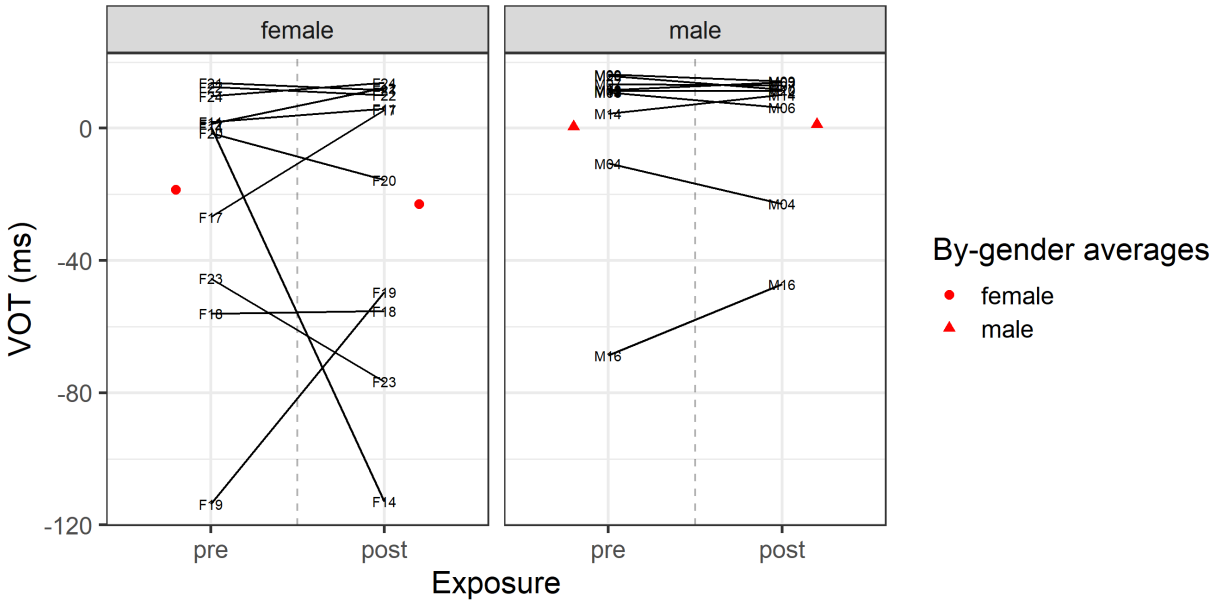

Figure 3.21: Change in mean VOT of /b/ in Extr. Prev. with by-gender averages

For reference, in *Extr. Asp.* (left side of *Figure 3.22*), there was a visible concentration of tokens around the 15 ms VOT /b/ target. In *Extr. Prev.* (right-hand side of *Figure 3.22*; −130 ms VOT /b/ target) only females show a weak trend of skewing towards negative VOT values (i.e. toward prevoicing) and the group-level median moves towards a shorter (albeit still positive) VOT value.

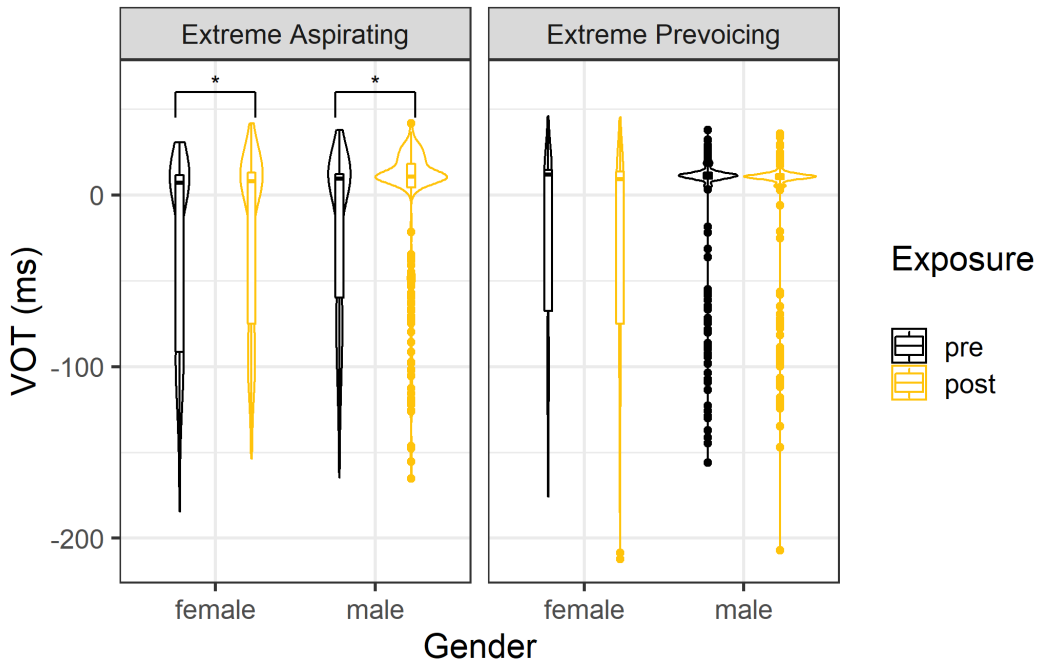

Figure 3.22: VOT of read /b/'s in both conditions before and after exposure by gender Restricted to participants who had room to accommodate

This effect of `Exposure`, especially for females was not statistically significant (*Table 3.13*), but it was trending negative (`Exposure`: β=−4.305, p=0.0913). Since this downward trend in VOT is most strongly exhibited by females, a post-hoc model was run just on the female participants, but again, it was not statistically significant (p=0.1550). It must also be noted that this model, similarly to the one run on *Extr. Asp.* /b/'s, does not even find a significant intercept, which indicates that /b/ productions showed a bimodal distribution (they were a mixture of plain and prevoiced tokens).

	Estimate	Std. Error	Pr(> |t|)
`(Intercept)`	−18.513	9.446	0.0650
`Gender [male]`	19.008	14.045	0.1921
`Exposure [post]`	−4.305	2.548	0.0913
`Gender [male]` × `Exposure [post]`	4.996	3.795	0.1883

Table 3.13: Linear mixed-effect regression model of reading B tokens' VOT (in ms) in the Extr. Prev. condition; Threshold for significance (adjusted with Bonferroni correction): p<0.0167

Males in *Extr. Prev.* (far right side of *Figure 3.22*) do not change their behavior much either, but they appear to be different from the other three groups from the start. Most of their /b/ productions are plain (short-lag positive VOT tokens) with only a handful of outlying prevoiced tokens. This indicates that males in the *Extr. Prev.* produced their voiced stops mostly plain with little prevoicing to begin with (over 75% of their tokens had short lag), which suggests that males randomly assigned to the *Extr. Prev.* condition were less able or willing to prevoice to begin with than males assigned to the *Extr. Asp.* condition.

Crucially, their pre-exposure productions do not resemble those of the males in *Extr. Asp.* The males in *Extr. Prev.* prevoiced less than males in *Extr. Asp.* (44/360 prevoiced tokens, cf. 102/400 in *Extr. Asp.*). A χ^2-test (with Yates' correction) confirmed that this difference is significant (χ^2=20.6761, p<0.0001). Thus, part of the reason why at least males exhibited no convergence could be because of the set of males assigned to the *Extr. Prev.* condition was particularly unable or unwilling to prevoice.

Unlike for the /p/ data, in the /b/ data there are no effects that would indicate accommodation being conditional on one of the likeability factors. The statistical tables and figures can be found in the Appendix for more details (*Tables A.10–A.12* and *Figures A.13–A.15*).

In terms of ethnicity, we see a similar tendency for *Extr. Prev.* /b/ as we did in the *Extr. Asp.* /b/ data. White participants in this condition were less likely to prevoice than people of color (*Figure 3.23*). No other generalization can be formed about accommodation. Most participants either did not change or diverged from the target (prevoiced less), and some participants converged with the target (prevoiced more). This was more or less independent of ethnicity: accommodation behavior had more to do with baseline values than ethnicity. While ethnicity could be correlated with baseline values, there is no evidence for it also being a factor in how much people accommodate.

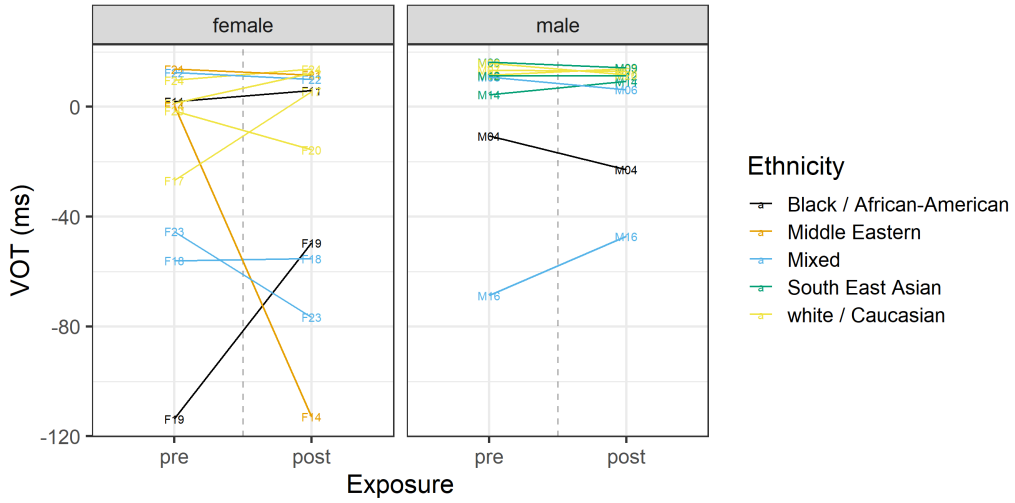

Figure 3.23: Reading performance for /b/'s in Extr. Prev. by ethnicity
Model talker's VOT: −130 ms

Patterns in the treatment of the /p b/ contrast in the Extreme Prevoicing reading data

The observations we saw in *Extr. Asp.* about differences between /p/ and /b/ accommodation are also borne out here but to a lesser extent. Changes in /p/ were much more common than changes in /b/, and as we saw earlier males in this condition were especially reluctant to change their /b/'s. While changes in /b/ production were less frequent, they were also bigger changes (a more drastic shift in VOT). These observations reinforce that the VOT of voiceless stops is something English-speakers adjust more than prevoicing or lack thereof on voiced stops. Just like in *Extr. Asp.*, accommodation behavior for /p/'s and /b/'s was not correlated (Pearson's r, p=0.4490). However, in *Extr. Prev.* there was far less convergence than in *Extr. Asp.* overall—to the point where the base models (without extra-linguistic predictors) did not even find convergence in *Extr. Prev*.

The target values that participants were exposed to in *Extr. Prev.* could have urged participants to move their voiced and voiceless bilabial stop categories closer to each other by decreasing the amount of aspiration on their /p/'s, but not prevoicing their word-initial /b/'s (that much). While some participants did follow this strategy, they were few and far inbetween, and even for them, categories sometimes touched, but did not overlap.

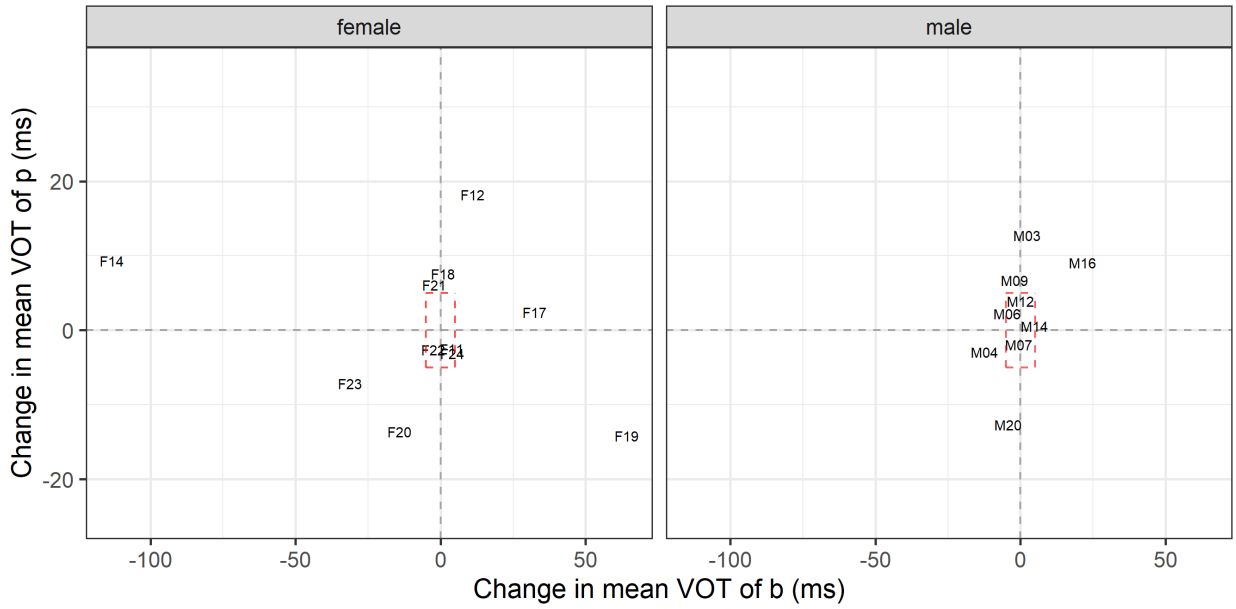

Figure 3.24: Change in mean VOT of /p/ and /b/ in Extr. Prev. per person; The red rectangle shows 5 ms change of means in either direction for reference

3.3.4 Summary

This section described the results observed in the two reading tasks, one before the shadowing task was completed (pre-exposure baseline) and one afterwards (post exposure). Participants were assigned to listen to either stimuli showing a /p/—/b/ contrast relying on *Extreme Aspiration* (/p/ is aspirated, 130 ms VOT; /b/ is plain, 15 ms VOT) or stimuli showing an *Extreme Prevoicing* /p/—/b/ contrast (/p/ is plain, 15 ms VOT; /b/ is prevoiced, 130 ms prevoicing).

We saw convergence with the target in the *Extr. Asp.* condition in the reading data. This was observable all over the /p/-dataset (participants converged to the model talker's 130 ms VOT /p/'s), and present in the subset of the /b/ dataset, which consisted of participants who were not already hitting the target (15 ms) before exposure. These changes, however did not straightforwardly correlate with any of the three likeability measures that this study looks at (`Superiority`, `Solidarity`, and `Dynamism`). Some participants showed considerably less aspiration on their /p/'s

after being exposed to highly aspirated tokens (M01, M15, F03). However, these participants were either hyperarticulating pre-exposure (thereby producing unnaturalistically long VOT's as baselines) or it could be argued that they hypoarticulated post-exposure as a task effect because of the duration of the experiment.

In *Extr. Prev.* there was no statistically significant convergence or divergence for either /p/ or /b/. However, for /p/, there seemed to be an effect of `Solidarity`. Participants who rated the model talker low on solidarity (more unfriendly, dishonest or rude) tended to diverge (they produced longer VOT's β=29.348, p=0.0007), while those who rated her high (more friendly, honest or polite) tended to converge (they produced shorter VOT's β=$-$5.647, p$<$0.0001). Such an effect of `Solidarity` was not found in the *Extr. Asp.* dataset. A possible explanation for the lack of `Solidarity` effect there is that since the stimuli were more English-like, maybe convergence was more automatic, and social factors were only drawn upon when the stimuli were odd or un-English-like in some sense.

However, the magnitude of the convergence (decrease in aspiration) in *Extr. Prev.* was comparable to the amount of decrease in aspiration in *Extr. Asp.*, which I have argued to likely be a result of task effect, and therefore this effect has to be treated with a grain of salt. In the case of /b/'s it could be established that at least the randomly assigned males in *Extr. Prev.* prevoiced significantly less even before exposure than randomly assigned males in *Extr. Asp.* did, which might have impacted their ability to accommodate to stimuli relying on a prevoicing contrast. They might not have perceived the contrast accurately to begin with or were not able to produce sustained stretches of prevoicing consistently.

The divergence of participants who rated the model talker low on `Solidarity`-related measures is more likely to indicate socially-mitigated accommodation, but this manifested in only a handful of participants actually converging with the stimuli, thereby potentially adjusting their categories to the model talker. Therefore, we do not have strong evidence for categories being flexible in terms of their phonetic detail. As it currently stands the English dataset supports the

Maintain categories hypothesis, which mandates an adherence to phonetic properties of categories rather than an abstract pressure to maintain contrastive categories as distinct. By maintaining the phonetic specifications of each category, contrasts are also maintained, albeit in an indirect way.

Gender effects have been sporadically seen to interact with various likeability factors. However, most of these effects are probably a result of either extreme values coming from a few participants or a confound between baseline VOT values and a given likeability score assigned on a small set of speakers (6 males in the *Extr. Asp.* condition). Models without likeability predictors could not establish a gender effect. This suggests that effects observed by Nielsen (2008) were likely due to a 100 ms target being too close to the females' baseline production in her study, and thus less convergence was attested. This notion of "too close" will receive more attention once the Hungarian reading data is evaluated. Ethnicity did affect /b/ productions, but not /p/ productions: participants who prevoiced at any point of the reading task (in either condition) were more likely to be people of color than white / Caucasian. However, accommodation behavior itself had much more to do with baseline values than ethnicity, and no link between ethnicity and accommodation could be established in the reading task.

In both conditions, accommodation to the model talker's /p/ was slightly more frequent but had a smaller magnitude, whereas accommodation observed in B-words was less common, but had a larger magnitude than for P-words. This was less striking in *Extreme Prevoicing* where there was no overall effect that would indicate systematic convergence. In other words, aspiration was more frequently adjusted in smaller increments, while participants changed the amount or frequency of prevoicing less often, but when they did, they made larger changes.

This could be related to the fact that aspiration is a contrastive cue for English stops (distinguishing voiced and voiceless stops), and thus movement is more limited—i.e. decreasing aspiration beyond a certain point can affect categorical perception. At the same time, since aspiration is also used for prosodic reasons (i.e. it accompanies intonational patterns), it is maybe more

available to English speakers articulatorily. In contrast, prevoicing is a truly optional cue in English, and as such, what amount of prevoicing a speaker expresses carries very little meaning in English (if any). Thus English speakers have more freedom in terms of the prevoicing range, but also have less experience with this cue. Therefore, native speakers of English might have less practice in manipulating the amount of prevoicing they produce in a fine-grained way. The lack of experience hearing prevoicing might amount to perceptual difficulties as well, which makes manipulating this cue even more difficult for native English speakers.

3.4 Shadowing results

In this section, we will look at the shadowing dataset. The first part of this section will offer a brief overview and some basic raw facts about the data (*Section 3.4.1*). In this part, I also will describe what statistical methods will be used to approach the data, as well as remind the reader which ones of our initial questions can be answered by the shadowing data and how. Then I will discuss first the *Extreme Aspirating* then the *Extreme Prevoicing* condition separately (*Section 3.4.2–3.4.3*, respectively). The section will end with a summary of the main findings.

3.4.1 Overview and statistical methods

Convergence was attested in the *Extreme Aspirating* condition for both /p/'s and /b/'s—towards the middle and the end of the shadowing task, in particular. Since participants largely converged in the reading task, there were no instances of participants only being affected by the model talker while directly exposed to her. Accommodation to /p/ was largely gradual (top left facet of *Figure 3.25*). During the shadowing task both instantaneous and gradual convergence trajectories were observed for /b/ accommodation. While the former was mostly attested by females, the latter was more typical of males (bottom left facet of *Figure 3.25*). Males' /b/'s were more prevoiced during this section than during their reading task, even though male productions used less prominent cues (less

aspiration and less prevoicing) overall. (They were also more prevoiced than *Extr. Prev.* men, but as we have seen men in *Extr. Prev.* were especially unlikely to prevoice to begin with). Thus, men had a different reaction to the *Extr. Asp.* stimuli from women, which can possibly indicate that they interpreted the stimuli in this condition as overall "emphasized". However, this is more likely to be a dialectal effect than a gender effect—dialects with more prevoicing could have been better represented among male speakers than female speakers.

In the *Extr. Prev.* condition of the shadowing task, little change was observed for /p/, and these values were quite similar to the values attained from the reading tasks. This can be seen from the flat trajectories in the top right facet of *Figure 3.25*. Participants started prevoicing their /b/'s more when shadowing the model talker, but this was restricted to the middle repetitions indicating a potential articulatory or realizational difficulty with a prevoiced /b/. It is also worth noting that this effect was limited to productions during the task that involved direct exposure, and did not carry over to the next task (reading). Such behavior was mostly observed among females on a group level, even though as we will see, many males exhibited similar behavior (see bottom right facet of *Figure 3.25*). Likeability effects could not be established on data from the *Extreme Aspirating* condition, but accommodation tended to correlate with higher `Superiority` ratings, for both /p/ and /b/ in males and for /b/ in females in *Extreme Prevoicing*.

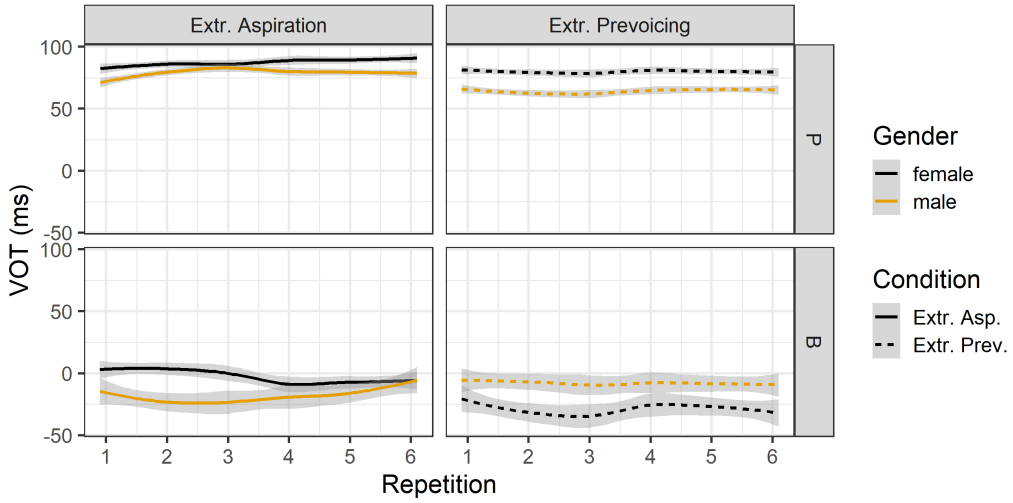

Figure 3.25: Smoothed results of the English shadowing data by condition and segment

Descriptive statistics

Each participant was presented with a random selection of 30 words out of the 40 that they saw in the reading task. There were 6 repetitions of each of their word from each participant. The *Extreme Aspirating* dataset consisted of 1,887 /p/'s and 1,875 /b/'s (with the exclusion of 3 /p/'s and 15 /b/'s). The excluded tokens had to be excluded because the participant skipped them, or because the token could not be segmented because of noise or hypoarticulation. In contrast, the *Extreme Prevoicing* condition had 1,697 /p/'s and 1791 /b/'s. While this condition had one fewer participant (i.e. 90 fewer recordings of /p/ and /b/ each), much more tokens had to be excluded in *Extr. Prev.* than in *Extr. Asp.* Most of the excluded tokens were /p/'s (106 /p/'s vs. 11 /b/'s). The reasons and consequences for this were discussed in *Section 3.2* above.

Statistical analysis

Data were separated into four groups by condition and segment (*Extr. Asp.* /p/'s; *Extr. Asp.* /b/'s; *Extr. Prev.* /p/'s; *Extr. Prev. /b/'s*) and analyzed in separate linear mixed-effect regression models. These models differed depending on whether likeability measures were included. In the baseline

models, the independent variables were `Gender`, `Repetition` (1–6) with their interaction included, and the dependent variable was VOT in ms. These models also featured a by-participant and by-word random intercept.

Just like with the reading data, each of the three likeability predictors were run separately with all four sections of the dataset, resulting in 12 models in total. In the models investigating likeability, the independent variables were `Gender`, `Repetition` and one of the three likeability measures (i.e. either `Solidarity`, or `Superiority` or `Dynamism`) as well as their two- and three-way interactions. These models also had a by-word random intercept. However, they did not have a by-participant random intercept, for two reasons. One, since we only have one `Solidarity`, `Superiority` and `Dynamism` rating per participant, a random intercept could easily be chosen to incorporate any effect of these variables, and thus absorb any potential likeability effect. Two, the baseline for shadowing productions (Repetition 1) were not independent of the model talker and their likeability. While in the reading task the pre-exposure reading production, which the model used as a baseline, could be considered relatively unaffected by the participant's opinion of the model talker, each shadowing production (even the first one) followed visual and auditory exposure to the model talker. As such, likeability effects could show up even in the first repetition, and if these effects are consistent throughout repetitions, they cannot be detected. For these reasons a by-participant random intercept was omitted from the likeability models—but as mentioned above, it was included in the baseline models.

Foci of attention

This section is a reminder of what question this task aims to answer and how those questions will be tested, broken down into three groups. First, the main focus will be whether contrasts are maintained with an adherence to phonetic detail or whether a bimodal distribution between contrasting segments is enough for speakers to interpret and adopt the input as target. If it is the

former (**Maintain categories**), then we will not see accommodation in the case of a plain /p/ (in *Extr. Prev.*), because a plain /p/ encroaches on the usual phonetic profile of an English /b/, which is mostly plain. If the latter hypothesis is true (**Maintain contrasts**) we will see accommodation in case of a plain /p/, because in *Extr. Prev.* it is still presented as a part of a clear contrast between a prevoiced /b/ and a plain /p/.

Second, this study can also contribute to answering other linguistic and methodological questions. Comparing accommodation behavior for /p/ and /b/ could be informative of whether accommodation to one category goes along with accommodation to the other. Furthermore, analysis of the shadowing data forms a crucial part of seeing task effects. Were there participants exhibiting convergence or divergence during shadowing, who did not show a difference in their reading productions? In particular, is it possible that participants did accommodate to the *Extr. Prev.* stimuli, but such effects went away once they were not directly exposed to the un-English-like stimuli? Moreover, does initiation of longer articulatory gestures (long VOT or long prevoicing) have the same trajectory as imitation of shorter gestures?

Third, this study can contribute to some discussion on extra-linguistic factors as well. Can we see evidence of VOT accommodation being gendered? Specifically, do we see a difference for *Extr. Asp.* /p/ similar to what Nielsen (2011) found? Moreover, since participants were asked to self-identify by ethnicity, we can also look at whether there are any connections between the ethnic background of participants and their accommodation behavior. As the ethnic background of participants was very diverse, each ethnic group has too few participants in it. Therefore the use of statistical methods would not be justified, and these data will be discussed qualitatively rather than through quantitative methods. Lastly, can we isolate certain components of likeability that are responsible for (VOT) accommodation? We will divide likeability into three components (`Solidarity`, `Superiority` and `Dynamism`) to address that.

	Estimate	Std. Error	Pr(> |t|)	
(Intercept)	82.669	4.848	<0.0001	***
Gender [male]	−10.574	6.638	0.1252	
Rep 2	5.586	1.977	0.0048	.
Rep 3	3.162	1.974	0.1094	
Rep 4	6.618	1.974	0.0008	***
Rep 5	6.395	1.974	0.0012	*
Rep 6	8.346	1.974	<0.0001	***
Gender [male] × Rep 2	1.202	2.863	0.6747	
Gender [male] × Rep 3	7.334	2.861	0.0104	.
Gender [male] × Rep 4	0.709	2.861	0.8043	
Gender [male] × Rep 5	1.577	2.861	0.5816	
Gender [male] × Rep 6	−2.012	2.866	0.4828	

Table 3.14: LMER model of shadowing /p/ productions in the Extreme Aspirating condition (target: 130 ms); Threshold for significance (adjusted with Bonferroni correction): p<0.0042

3.4.2 The Extreme Aspirating condition

Accommodation of /p/'s

The model on the *Extr. Asp.* /p/ data (*Table 3.14*) indicates that the convergence found in reading happened gradually throughout the shadowing task. Participants produced longer VOT's in Repetition 4, 5, and 6 than in the baseline Repetition 1 (by 6.618 ms, 6.395 ms, and 8.346 ms, respectively). Repetition 2 also trended towards convergence, but this effect was no longer significant, once corrected for the number of tests run (p=0.0048). This trajectory was gender-independent (none of the p-values of Gender effects reached significance).

Figure 3.26 shows this visually. The bolded lines are a product of locally estimated scatterplot smoothing (loess) over the dataset, and the gray ribbon around it is its confidence interval. Each thin gray line connects observations of a given participant's by-repetition means. For instance one of the gray lines in the male (bottom) plot is M01's Rep 1 mean is connected with his Rep 2, Rep 3, Rep 4, Rep 5, and Rep 6 means, another represents M02's progression through repetitions, etc. As we saw convergence in the reading data, this convergence persists even when the participants are not directly exposed to the stimuli from the model talker.

While the highest average VOT's consistently came from females, so did the lowest. Female VOT's tended to concentrate around a somewhat higher region of VOT than males' did, but this was not significant (as we saw in *Table 3.14*, p=0.1252). On an individual level, there are some participants who stay stagnant, but most converge gradually throughout the task. Except for one female, no one reaches a mean of 130 ms, which is where the target was (F04).

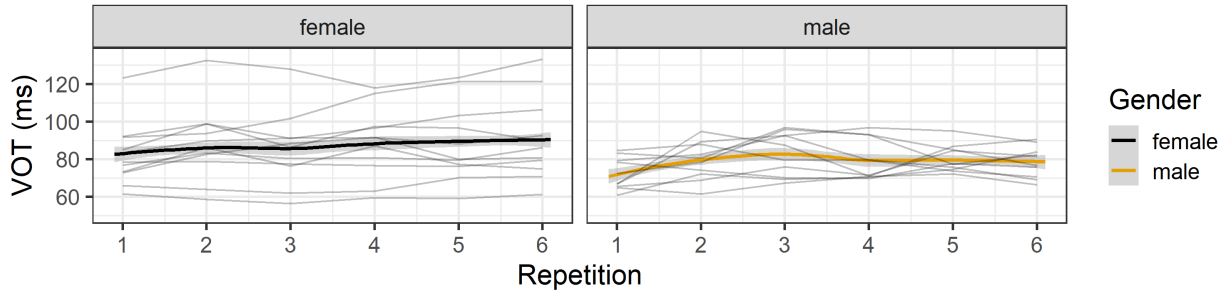

Figure 3.26: Smoothed results of the Extr. Asp. /p/ shadowing data

No extra-linguistic factor could be tied to accommodation. In terms of likeability, none of the three measures (out of `Solidarity`, `Superiority`, and `Dynamism`) improve the model's fit compared to a model that has none of these factors, but includes a by-participant random intercept (p=1). The statistical tables can be found in *Tables A.13–A.18*, and the figures of plotted results in *Figures A.16–A.18* in the Appendix. Ethnicity was not a good predictor of shadowing behavior

either, participants showed more or less a flat trajectory, and members of no ethnic groups produced consistently higher or lower VOT's than others (*Figure A.19* in the Appendix).

Accommodation of /b/'s

In the *Extreme Aspirating* condition's /b/ data initially it seems like all effects go away when correcting for the number of tests run (*Table A.19* in the appendix). However, once the dataset is restricted to those who were at least 5 ms away from the target in their mean pre-read production, like we did for the reading data, we find that convergence was only observed in males.[3]

Then the same analysis was also conducted on the 1,166 /b/ tokens from the 13 out of the total 21 participants whose PRE-Read mean was at least 5 ms away from the 15 ms VOT target. *Table 3.15* shows that the model does not find an intercept (p=0.6905), which as discussed before, indicates that the /b/'s did not have a unimodal distribution, but instead were split between plain and substantially prevoiced. Moreover, males in *Extr. Asp.* were closer to the 15 ms target /b/ in Repetitions 4 and 6 (by 29.368 ms and 35.035 ms, respectively) than the females (or they themselves) were initially (baseline: females in Rep 1), but did not differ from them in the other repetitions.

[3]The lack of effect in the model run on the whole dataset *Table A.19* is not just due to a single outlier (M17). This participant is a male who closely matched the target in their pre-read productions, so his curve is only shown in the top plot in *Figure 3.27*. In Rep 1 he had the highest mean in the entire dataset, then rapidly decreased his VOT to a very prevoiced range for Reps 3, 4, and 5. When only his data were excluded from the model, the same lack of effects still persisted.

	Estimate	Std. Error	Pr(> |t|)	
(Intercept)	−4.161	10.291	0.6905	
Gender [male]	−32.330	14.679	0.0417	.
Rep 2	−0.917	6.846	0.8935	
Rep 3	−5.823	6.846	0.3952	
Rep 4	−17.451	6.846	0.0109	.
Rep 5	−10.481	6.846	0.1261	
Rep 6	−13.233	6.863	0.0541	
Gender [male] × Rep 2	−11.998	10.092	0.2348	
Gender [male] × Rep 3	5.983	10.092	0.5534	
Gender [male] × Rep 4	29.368	10.092	0.0037	**
Gender [male] × Rep 5	21.832	10.124	0.0313	.
Gender [male] × Rep 6	35.035	10.104	0.0005	***

Table 3.15: LMER model of shadowing /b/ productions in the Extreme Aspirating condition (target: 15 ms); Participants who "had room" to accommodate
Threshold for significance (adjusted with Bonferroni correction): p<0.0042

This statistical test paints a more complicated picture. To gain a better understanding of what is happening, we need to take a look at the data visually (*Figure 3.27*). We can see here, that males tended to use prevoiced VOT ranges more than females did, and this is especially true in the dataset restricted to those who "had room". The females were mostly around the plain /b/ target (15 ms VOT) throughout the entirity of the shadowing task.

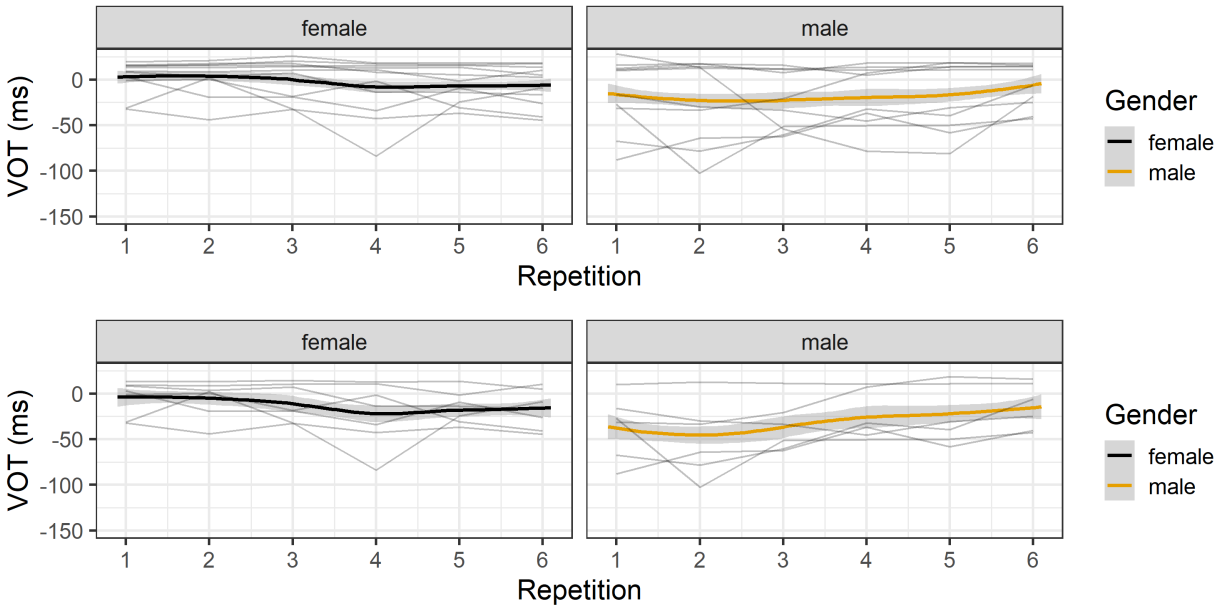

Figure 3.27: Smoothed results of the Extr. Asp. /b/ shadowing data (all participants on top, those who "had room" on the bottom)

One could think that the difference between males and females is a result of different starting points: that males are simply more prone to prevoicing their /b/'s compared to females, and thus have "a longer way to go" to match the target. However, this is refuted once we contextualize the shadowing data with productions from the two reading tasks. *Figure 3.28* shows all the PRE-Read task's tokens, shadowing tokens, and POST-Read tokens, collapsed into a single by-segment, -gender, and -condition mean each. For example, in the top left plot the line in black shows means of the female /p/ productions in *Extr. Asp.*. The first datapoint is the mean of all PRE-Read /p/'s from these participants, the middle one is the mean of all their /p/'s in the shadowing task (averaged over all the repetitions), and the third point is the mean of all their POST-Read /p/'s. The yellow line in that same plot shows the same for /p/'s from the males in *Extr. Asp.*

While tokens from the reading and shadowing productions should not be directly compared with one another, as any possible results might be more indicative of the completely different nature of the two tasks than anything else, we can see that males overall tended to have less extreme productions than females. Males overall had less aspiration on their /p/'s and less prevoicing on

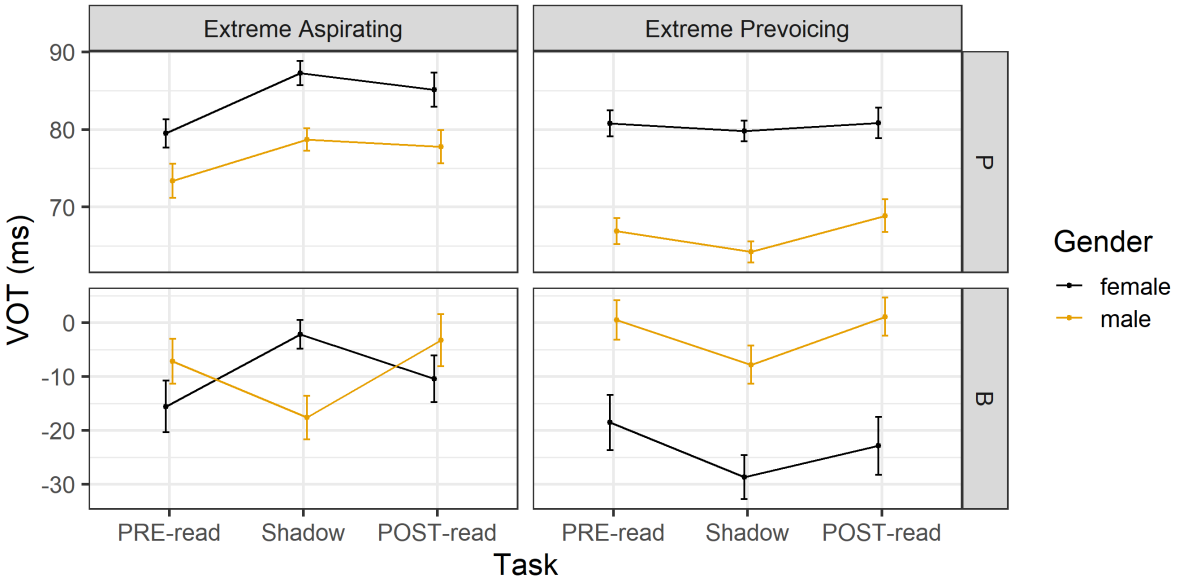

Figure 3.28: All productions from all English-speaking participants averaged by task: PRE-Read (2 reps), Shadowing (6 reps) and POST-Read (2 reps)

their /b/'s than females did, across all tasks and conditions (remember though that males in *Extr. Prev.* were especially unlikely to prevoice their /b/'s). This tendency is flipped in one case only, for /b/'s in the *Extr. Asp.* shadowing task (middle datapoints in the bottom left plot). Here, males tended to prevoice more (diverging from the 15 ms VOT target value), while females produced largely plain tokens on average (converging with the plain /b/ target).

This indicates that males and females had a somewhat different reaction to the stimuli in *Extr. Asp.* (a plain /b/ and a very aspirated /p/). Females reacted to stimuli in this condition by trying to match both of them, thus converging. At the same time, males made the cues of both segments (prevoicing and aspiration) more prominent, which in this study is interpreted as convergence to the aspirated /p/ and diverging from the plain /b/. It must be noted that this strategy was only attested in the shadowing task, and only by males who were far enough from the target in their PRE-Read, i.e. males that tended to prevoice at least some of the time, enough so to distance their mean from the 15 ms target by at least 5 ms. Within this group almost all males exhibited this pattern.

This strategy widens the distinction between /p/ and /b/, and suggests a general approach of enunciating the words more clearly. It could be a result of interpreting the extreme aspiration on /p/'s as emphasis and generalizing to the entire context (i.e. the task). As we have seen in *Chapter 2*, dialects of English differ in how stress and accent manifest on utterance-initial voiced stops—dialects, which have less prevoicing, prosodically prominent stops will be even "plainer", whereas in dialect, which have more prevoicing habitually do not shorten (maybe even lengthen) prevoicing. The reason why we might have seen a gender-divide in this respect could be a confound with ethnicity. In the *Extr. Asp.* condition most females were white (6/11) and from the Mid-Atlantic and Northeast (8/11), while the males were more diverse in terms of both ethnicity and place of birth. This asymmetry makes it more likely that more males than females spoke dialects that prevoice more in prominent environments.

If this was indeed a manifestation of emphasis, one of three things must to be true about these prevoicers in *Extr. Asp.* First, they could have ignored the exact phonetic specification of plain /b/'s and just generalized from the perceptually salient extreme aspiration on /p/'s. Second, they could have interpreted the plain /b/'s to be the model talker's enunciated performance (which could be plausible if the model talker is one of those people who categorically do *not* prevoice utterance-initially). Third, it is possible that these participants did not perceive the prevoicing accurately. This third option is somewhat unlikely, as there is indication from previous literature that English speakers are able to perceive the difference between prevoiced and plain stops, albeit in non-lexical environments.

Later in the shadowing task, these same participants gradually reduce the amount of prevoicing they produce, which could either indicate articulatory fatigue or a decrease in interest in or motivation regarding the task. This is captured by the statistically significant difference between Repetition 1 and Reps 4 and 6 (as it was shown in *Table 3.15*).

In *Section 3.3.2* we saw that participants either matched the target to begin with or got closer to it during the reading task. Therefore we did not have an opportunity to observe any convergence that appears only *during* exposure, and does not continue afterwards. In fact, we initially see divergence in male productions during shadowing, but since we do not see `Gender` effects in the reading data, this likely has turned into convergence in the reading task.

As opposed to the /p/ production where ethnicity did not seem to be an influencing factor, *Extr. Asp.* /b/ productions varied with ethnicity (*Figure 3.29*). Most participants who produced any prevoicing—i.e. diverged from the 15 ms VOT target—were people of color, namely South East Asian, Black / African-American or Mixed participants (the person categorized as Mixed self-identified as "Mixed" as well). While there were some white / Caucasian participants who prevoiced as well, most of them produced mostly or exclusively plain /b/'s. Similarly to the findings in the reading dataset, this can also reflect dialectal differences, similarly to the reading data. For instance, the Black / African-American participants' prevoicing is not surprising, since AAL tends to have prevoicing on voiced stops, unlike most dialects spoken by white people in the North-East US (Ryalls et al., 1997). The two Hispanic / LatinX participants did not prevoice at all. None

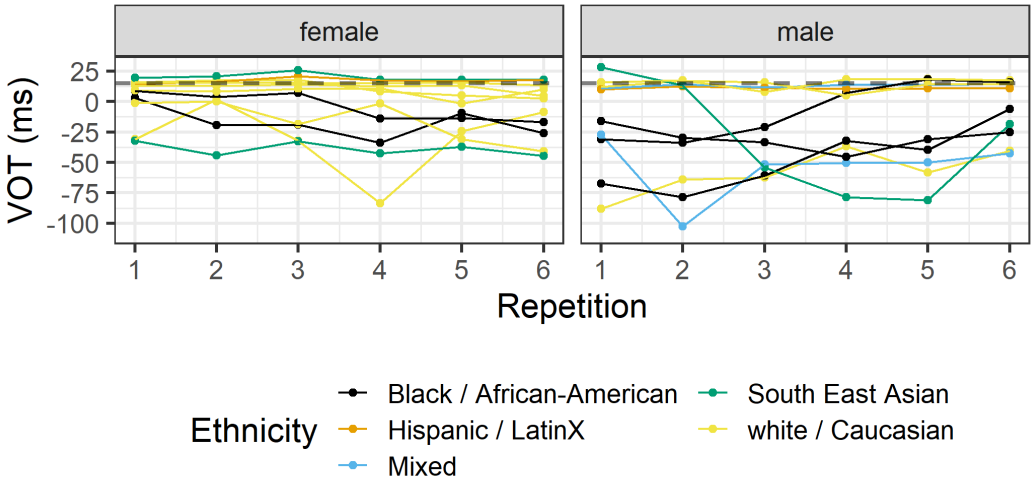

Figure 3.29: Shadowing performance for /b/'s in Extr. Asp. by ethnicity
Gray dashed line indicates the model talker's VOT (15 ms)

of three models run with the likeability measures (`Solidarity`, `Superiority`, and `Dynamism`) find significant effects (see *Tables A.20–A.25* and *Figures A.20–A.22*in the Appendix) nor do the measures improve on the baseline model without them (p=1).

Patterns in the treatment of the /p b/ contrast in the Extreme Aspirating shadowing data

On a group-level the /p/ and /b/ categories seem adjacent but with limited overlap. This is shown in *Figure 3.30*, which plots all tokens from the shadowing task from all participants in *Extr. Asp.* The categories only touch for a few participants, and on an individual level we saw no mergers and even overlaps between the two categories were minimal and were only observed in two participants (*Figure 3.31*). Participants with close or touching categories were more commonly attested in male than in female productions, especially, but not exclusively those who did not produce prevoicing on their /b/'s. The two categories being closer for males than females is hardly unexpected, since, as mentioned before, males used less prominent cues (less aspiration and less prevoicing) to distinguish their bilabial stops from one another.

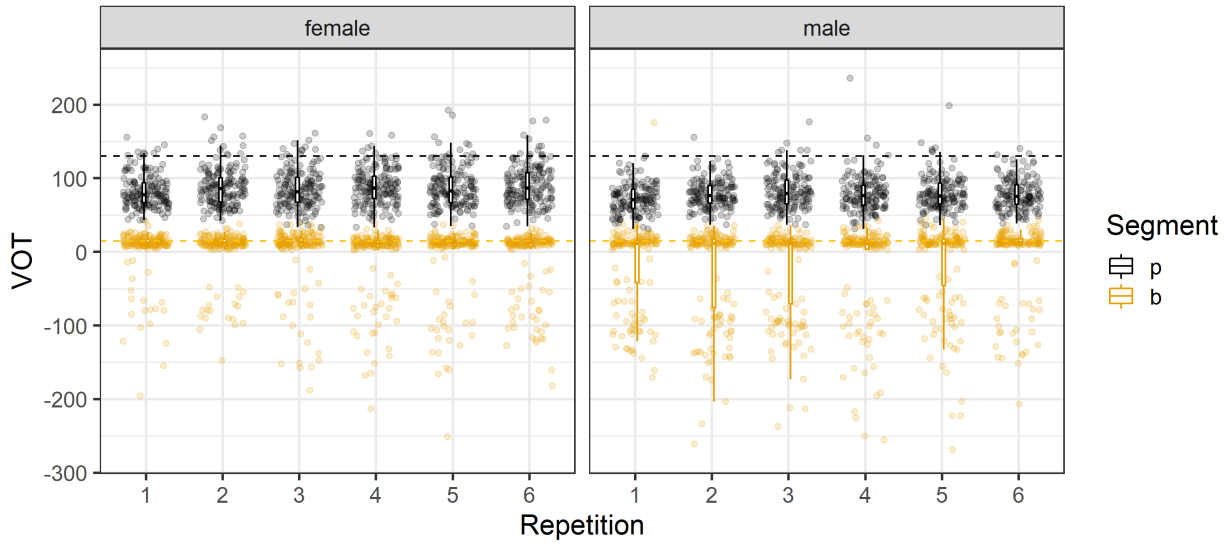

Figure 3.30: All participants' shadowing productions in Extr. Asp.

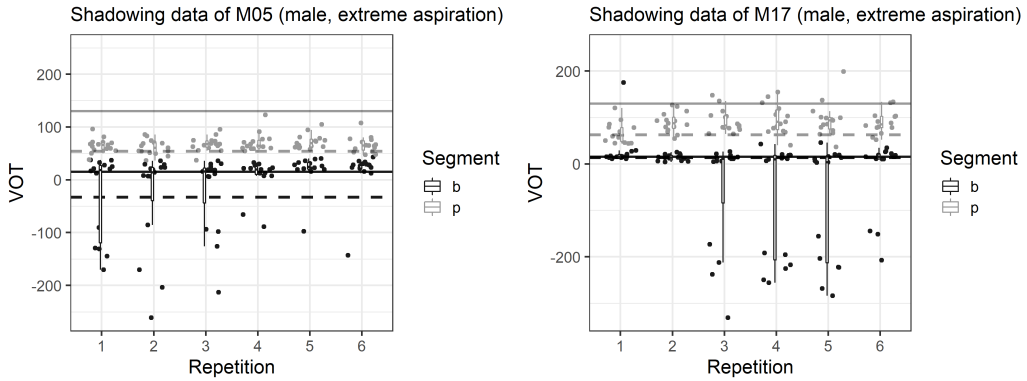

Figure 3.31: The two participants whose /p/ and /b/ productions overlapped in shadowing
Solid lines of respective colors show the target, and dashed lines show the participant's PRE-Read average

At the same time, there were also participants whose categories became more distinct from one another with each repetition. *Figure 3.32* shows one of the clearest examples of this. F15 (left) gradually increases aspiration on her /p/ while staying "on target" with her /b/ (target: 15 ms). We also see examples of the distancing patterns, where the participant simultaneously increases aspiration on their /p/ and prevoicing on their /b/ (M13, right). As discussed before, this strategy was mostly attested in males. While these figures show examples of a clear and monotonic trajectory, convergence (or less commonly divergence) was non-monotonic for most participants, which is indicative of the immense amounts of variation that is inherent to shadowing.

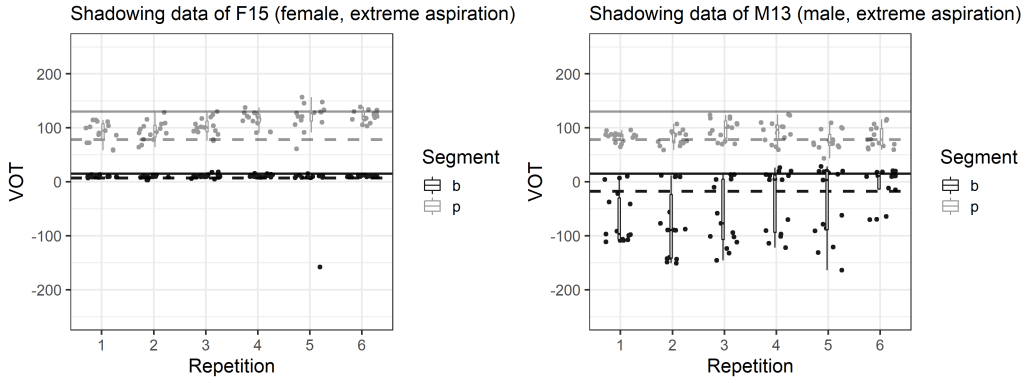

Figure 3.32: Examples of /p/ and /b/ productions becoming more distinct in shadowing
Solid lines of respective colors show the target, and dashed lines show the participant's PRE-Read average

All in all, we saw a mixture of flat trajectories and steadily rising values for VOT for *Extr. Asp.* /p/'s, where the target was a stop with 130 ms VOT. In conjunction with the decisive effect of `Exposure` in the reading data (participants lengthened their VOT), the flat trajectories can be interpreted as instantaneous convergence, where participants hit the ceiling of their VOT production in the very first repetition (at least their relative ceiling attainable in this task with these stimuli), while monotonic rise in VOT values indicates gradual convergence.

As for the /b/ data, females and males showed different patterns. While females uniformly converged with the 15 ms plain /b/ stimuli (either all at once or gradually), most males fell into two camps. They either matched it in their PRE-Read already and then stayed with it through the shadowing task, or they did not match it in the PRE-Read task, and then diverged from it even stronger in shadowing (i.e. prevoiced more and more), which could be interpreted as adding emphasis or enunciation. Aside from these effects, no strong likeability effects could be established.

3.4.3 The Extreme Prevoicing condition

Accommodation of /p/'s

In the *Extreme Prevoicing* /p/ dataset, the model finds only the intercept, and no other significant values. While men in this condition tended to have less aspirated voiceless stops throughout the shadowing task, this is not a significant difference, once Bonferroni correction is applied (p=0.0123). This could be a result of including by-participant random intercepts, which could also absorb some of the variation coming from gender differences. None of the repetitions significantly differed from Repetition 1 (the baseline).

	Estimate	Std. Error	Pr(> |t|)	
(Intercept)	81.253	4.169	<0.0001	***
Gender [male]	−15.965	5.840	0.0123	.
Rep 2	−1.416	1.715	0.4093	
Rep 3	−2.717	1.710	0.1123	
Rep 4	−0.329	1.710	0.8473	
Rep 5	−2.062	1.713	0.2288	
Rep 6	−1.684	1.721	0.3280	
Gender [male] × Rep 2	−1.040	2.575	0.6864	
Gender [male] × Rep 3	−0.085	2.563	0.9737	
Gender [male] × Rep 4	0.389	2.557	0.8792	
Gender [male] × Rep 5	1.531	2.562	0.5503	
Gender [male] × Rep 6	1.678	2.577	0.5150	

Table 3.16: LMER model of shadowing /p/ productions in the Extreme Prevoicing condition (target: 15 ms); Threshold for significance (adjusted with Bonferroni correction): p<0.0042

The trajectories of participants are in fact flat over the 6 repetitions, both as a group and individually (*Figure 3.33*). Again, in this plot the thicker lines are a result of locally estimated scatterplot smoothing over all data (females in *Extr. Prev.* in black, males in yellow) and the thinner gray lines each represent an individual's trajectory (in terms of by-repetition means).

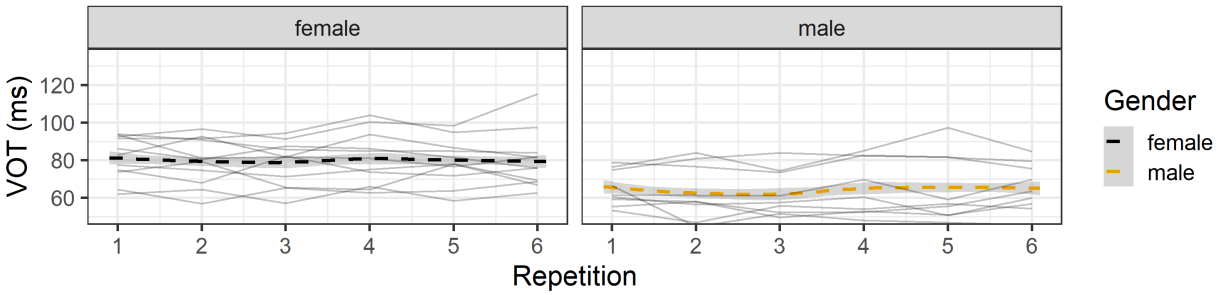

Figure 3.33: Smoothed results of the Extr. Prev. /p/ shadowing data

While such flat trajectories could be a result of instantaneous convergence or divergence (as we've seen for the *Extr. Asp.* /p/ and /b/ data), this is unlikely to be the case here since participants with varying starting VOT's all produced flat trajectories. If the flat trajectories were indicative of instantaneous convergence, then we would see some participant hovering closer to the 15 ms target in a relatively flat way and some starting further away, but gradually approaching it. Since we do not see such a distribution in the data, results are likely reflective of no accommodation and thus we do not find any strictly immediate convergers, who converged with the 15 ms plain /p/ stimuli during shadowing but abandoned that behavior once they were no longer exposed to the model talker.

As for likeability effects, `Solidarity` and `Dynamism` ratings did not explain the interpersonal variation we see. Neither do they add to the baseline model (p=1), nor does visual inspection suggest an effect of these two factors on accommodation. The tables from the statistical analysis with these two factors can be found in *Tables A.26–A.27*, *Tables A.30–A.31*, and *Figures A.23–A.24* in the Appendix.

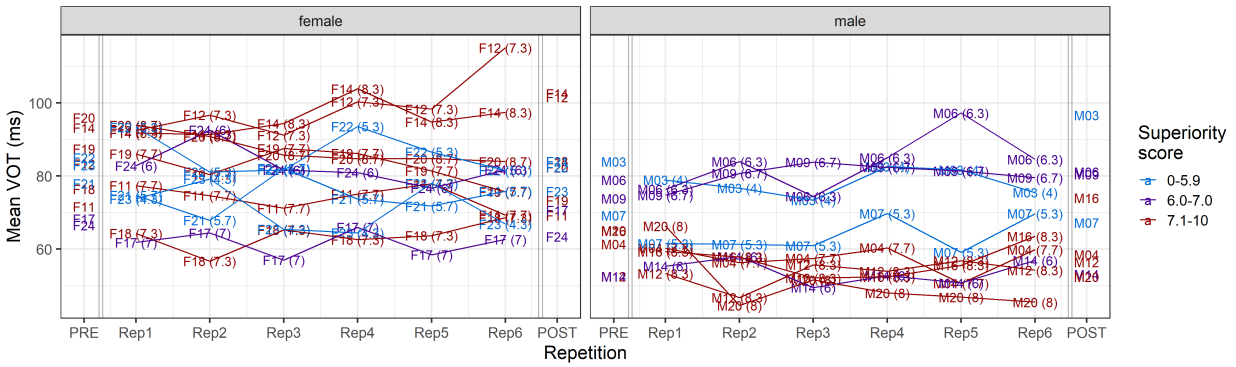

Figure 3.34: Participants' /p/ shadowing trajectories by `Superiority` *ratings in Extr. Prev.*

However, there did seem to be a pattern involving males' `Superiority` rating (see *Figure 3.34* as well as *Tables A.28–A.29* in the Appendix). Males who rated her higher for `Superiority`-related measures (more organized, intelligent, and of a higher status) produced some of the shortest VOT's in the whole dataset—on average −6.00 ms shorter than the 64.797 ms intercept (p=0.0002).

This plot shows the participants' individual trajectories, colored by their ratings given. The colors are assigned in a binned fashon: blue for average ratings below 6.0, purple for ratings between 6.0 and 7.0, and red for ratings above 7.0. These values were chosen semi-arbitrarily in a way that is intuitive and also allows for the easiest telling apart of individual lines with the fewest colors possible. The pattern itself surfaces as a more-or less binary divide among males rather than fine-grained likeability grading. Participants rating the model talker 7 or higher produced shorter mean VOT's and those below produced longer lags (even M14 with his short VOT's but his `Superiority` rating of 6 seems to be somewhat of an exception). Aside from this exception, the pattern seems to hold. This pattern is completely absent from the female productions. Similarly to the /p/ productions in *Extr. Asp.*, /p/ productions in *Extr. Prev.* did not seem to have any correlation to ethnicity, and neither do accommodation trajectories. The individual trajectories by ethnicity are shown in *Figure A.25* in the Appendix.

Accommodation of /b/'s

The situation was somewhat different for /b/'s in the *Extreme Prevoicing* condition. When the model was first run on all participants in this dataset, no effects were found (*Table A.32* in the appendix). Productions from Repetition 3 initially appeared to be more prevoiced, but this did not prove significant in the end (β=$-$12.9, p=0.0044).

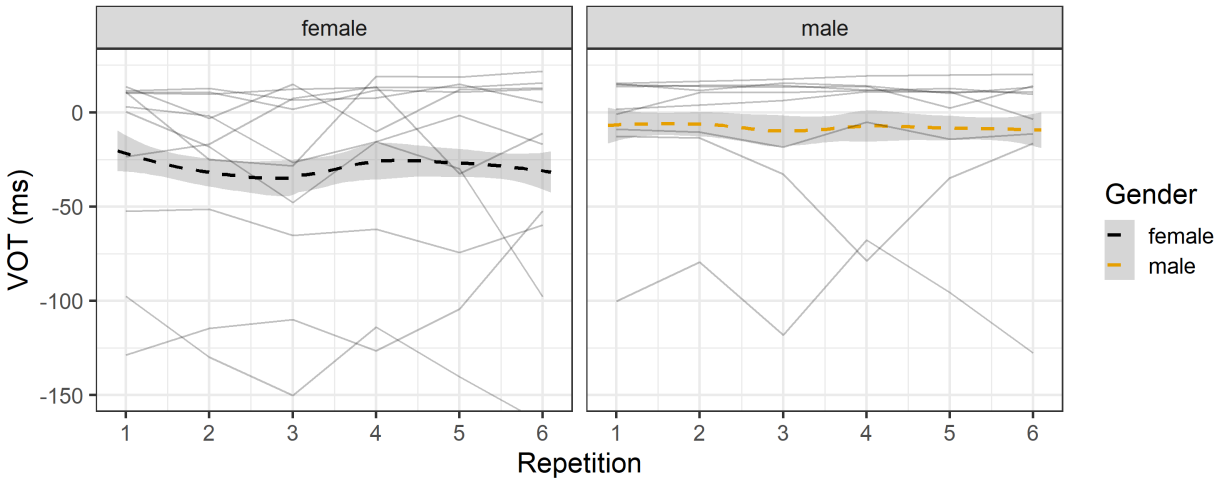

Figure 3.35: The use of prevoicing in Extr. Prev.

However, females had visibly more change in their VOT's than males did (*Figure 3.35*). As established earlier, males in *Extr. Prev.* were especially unlikely (probably unwilling and/or unable) to prevoice even before any exposure. While some females had flatter trajectories with short-lag means during shadowing, almost all males followed this pattern. Therefore a separate post-hoc model was run on females only.

	Estimate	Std. Error	Pr(> |t|)	
(Intercept)	−22.116	15.526	0.1820	
Rep 2	−7.767	4.869	0.1110	
Rep 3	−12.934	4.854	0.0078	*
Rep 4	−3.323	4.861	0.4944	
Rep 5	−6.491	4.861	0.1821	
Rep 6	−8.352	4.862	0.0861	

Table 3.17: LMER model of shadowing /b/'s from females in the Extreme Prevoicing condition (target: −130 ms); Threshold for significance (adjusted with Bonferroni correction): p<0.0083

Within the female dataset Repetition 3 was found to have more prevoicing than the baseline (Repetition 1), while the other repetitions did not differ from Rep 1 (*Table 3.17*, Rep 3: β=−12.934 ms, p=0.0078). It is apparent from the individual trajectories that the situation is somewhat more muddled than simply a steep and universal dip at Rep 3 and then immediate recovery.

Most female participants attempt more prevoicing in one of the later repetitions than they had initially, but most instances of it can be seen in Repetition 3. It is also the case that after a few repetitions like that, they do not sustain this level of prevoicing throughout the experiment (unlike, for example, the trajectories we saw for *Extr. Asp.* /p/ and /b/ convergence, where convergence was monotonic). Even though the trajectories are often more complex, than a monotonic decrease of VOT plus an inflection point, followed by monotonic increase, I will refer to these trajectories as V-shape for short. Such trajectories could be indicative of the participant attempting to prevoice, but either failing to maintain (long) prevoicing consistently or losing interest or motivation to do so. V-shapes are also exhibited by males who prevoice.

There were quite a few strictly immediate convergers in this condition—participants actually got closer to the −130 ms VOT target /b/ during shadowing, but only for a little bit, then reverted to more plain productions, and showed no effect of exposure in their POST-Read productions. There participants largely followed a V-shaped trajectory. *Figure 3.36* shows some of these trajectories.

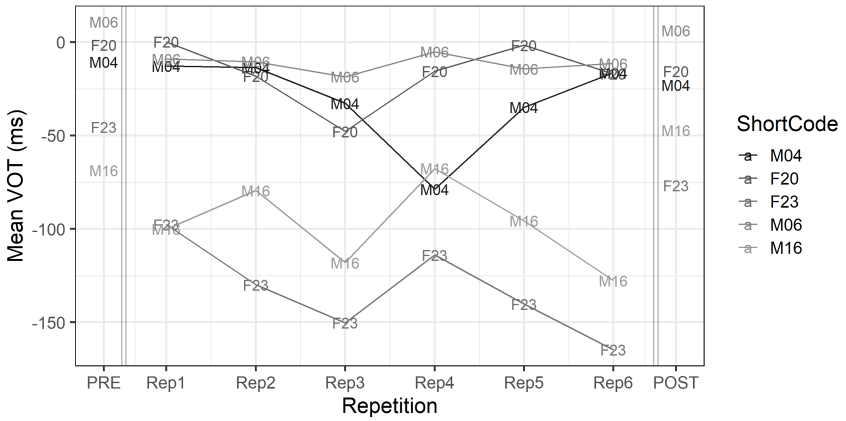

Figure 3.36: Strictly immediate convergers to Extr. Prev. /b/

This prevoicing pattern attested in *Extr. Prev.* (repeated in *Figure 3.38* can be compared with the use of prevoicing in *Extr. Asp.* /b/'s (repeated in *Figure 3.37*), where even though their stimuli had plain /b/'s, some males also produced prevoicing. In both conditions, prevoicing tended to decrease in the second half of the shadowing task, but the conditions differed in how quickly

prevoicing appeared in the first place. While in *Extr. Prev.* prevoicing tended to appear gradually (and then disappear throughout later repetitions), caused a V-shape, males in *Extr. Asp.* produced a lot of prevoicing initially (instantaneous accommodation), which then disappeared. The fact that these cues disappeared over the course of the task is most likely a testimony to how hard it is to produce consistent word-initially prevoicing for English speakers, especially throughout such a long task.

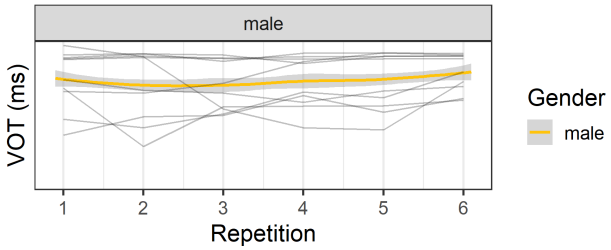

Figure 3.37: The use of prevoicing in Extr. Asp. (males only)

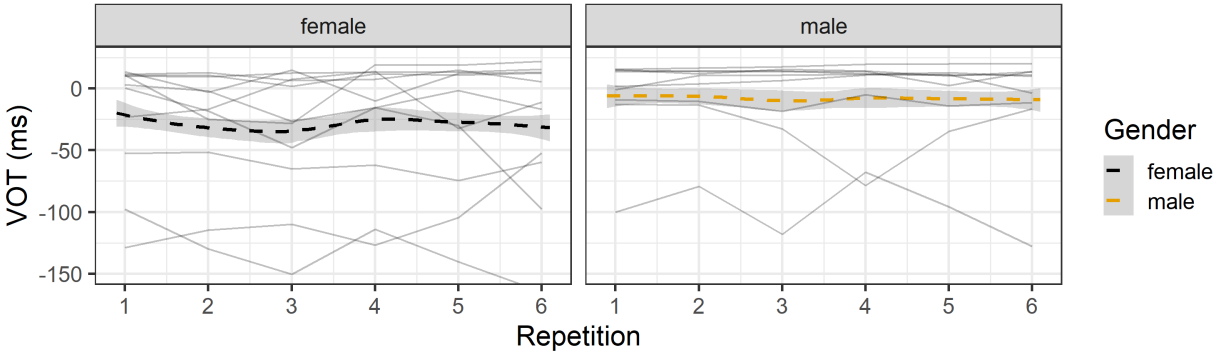

Figure 3.38: The use of prevoicing in Extr. Prev.

The fact that some *Extr. Asp.* males started to produce extreme prevoicing from the first repetition on instead of gradually appearing by Repetion 3 might indicate that the *Extr. Asp.* stimuli were easier to incorporate than the *Extr. Prev.* stimuli. The prevoicing reaction to the plain /b/ stimuli in *Extr. Asp.* was more or less organic. The /p/ and /b/ in *Extr. Asp.* were quite similar to what the categories sound like in everyday English, and therefore it was presumably easier to identify, comprehend and incorporate. In contrast, in *Extr. Prev.*, participants might have needed

some time to "wrap their heads around" and adapt to the un-English-like stimuli. This strangeness mostly came from the word-initial /p/'s being plain (15 ms VOT), which could be seen from the 106 /p/ tokens that had to be excluded from the shadowing analysis, because participants often assumed that they heard a b-initial nonceword. The confusion around /p/'s could have also hindered adaptation to /b/ stimuli because the contrast (prevoiced vs. plain stop) was unfamiliar to most English speakers.

In terms of likeability, `Solidarity` and `Dynamism` could not be linked to accommodation. They did not improve the model compared to the baseline that had no likeability features, but had a by-participant intercept (p=1, see *Tables A.33–A.34*, *Tables A.40–A.41*, and *Figures A.26–A.27* in the Appendix).

However, accommodation behavior showed sensitivity to `Superiority` ratings. Data are shown in *Figure 3.39*. Males who rated the model talker low on scales related to `Superiority` had significantly higher (less prevoiced) values than females throughout the shadowing task (`Gender`: β=115.585, p=0.0008, *Tables A.35–A.36*). There was an interaction between `Gender` and `Superiority`, but this went away after Bonferroni correction was applied. Since this interaction that would moderate the effect of `Gender` is no longer significant, we are left with an unlikely prediction: without the interaction, results just indicate that males had 115.585 ms less prevoicing than females on average.

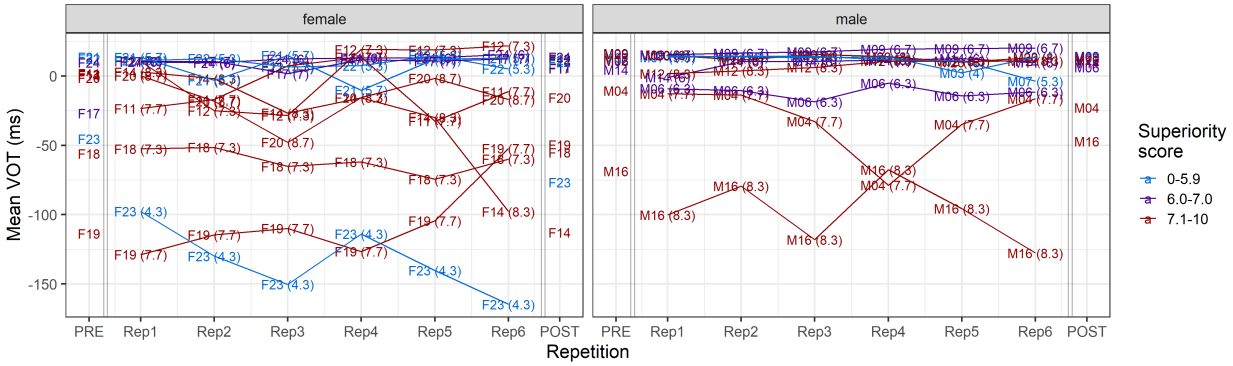

Figure 3.39: Participants' /b/ shadowing trajectories by `Superiority` *ratings in Extr. Prev.*

Since this is an unlikely fit for the data, in order to better understand this, males and females were analysed in two separate post-hoc model. In the female model (*Table A.37*) no effects were found. However, looking at *Figure 3.39*, we can see that F23's behavior is somewhat of an outlier. Once she is removed (*Table A.38*), we can see a reverse correlation between VOT and `Superiority`: females who rated her higher on `Superiority` produced more prevoiced tokens than those who rated her low—by an estimated -13.017 ms for per score point on the rating scale, p=0.0019). This is measured from the unattested 77.964 ms baseline intercept (for a hypothetical 0 `Superiority` score). In practice this is capturing the fact that the only prevoiced mean values for any repetition came from females who rated the model talker high for `Superiority`-related traits.[4]

Statistically speaking, males also exhibited a similar pattern (*Table A.39*). VOT is decreased by 11.426 ms per `Superiority` point ($p<0.0001$) compared to the hypothetical, unattested 69.740 ms baseline intercept (p=0.0006).

As far as ethnicity is concerned, prevoicing tended to more often come from people of color than from participants who self-identified as white / Caucasian (*Figure 3.40*). There were 3 participants (2 females and 1 male) with Mixed ethnic identities who produced means with more than 100 ms prevoicing. They self-identified as "South East Asian, Middle Eastern" (F23) and "Afro-Latina and white" (F18) for the two females, and as "white, black" on the male side (M16). This is consistent with findings from the reading data as well as the /b/ productions in the *Extreme Aspirating* condition.

[4]The one exception is F23, who rated her 4.3. Even so, in the reading data she was the clearest case of convergence for both /p/ and /b/.

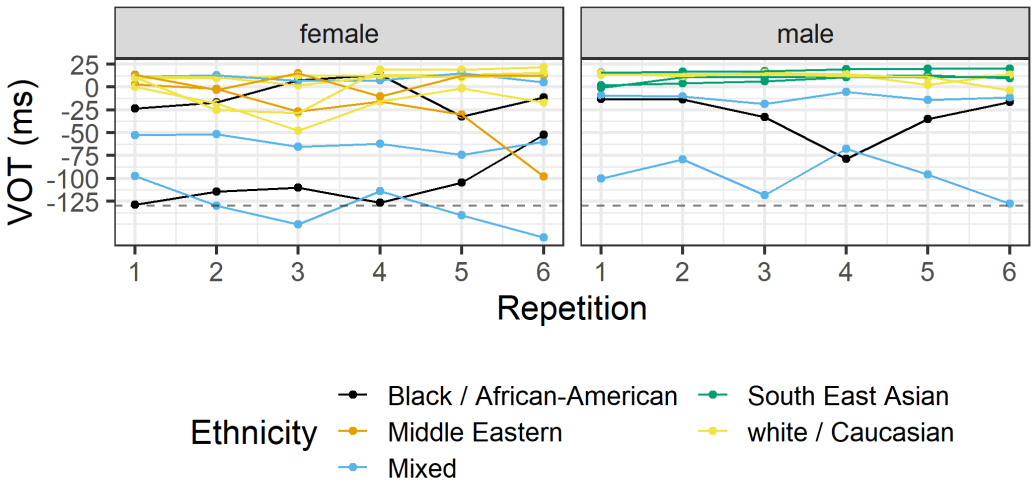

Figure 3.40: Shadowing performance for /b/'s in Extr. Prev. by ethnicity
Gray dashed line indicates the model talker's VOT (−130 ms)

To sum up, in *Extr. Prev.* we saw fewer instances of convergence to the target /b/ (either all at once or gradually) than in *Extr. Asp.* This is consistent with the findings of the reading data, where there was significant convergence on a group level in *Extr. Asp.*, but only very few individual instances of convergence were found in *Extr. Prev.* Males were especially unlikely to prevoice (see the discussion on their PRE-Read data in *Section 3.3.3*) and most of their productions changed very little during the shadowing task. The females and the few men whose VOT changed exhibited a V-shaped pattern, where they attempted to prevoice midway through the task, but then abandoned it for the rest of shadowing (some of them did not even show any sign of convergence in their POST-Read). This pattern shared some features with what males did in *Extr. Asp.*, where they started with a lot of prevoicing, but also lost it by the end of the task, which could indicate articulatory difficulty or decrease in motivation or attention. The crucial difference was that while some males in *Extr. Asp.* started prevoicing immediately (in Rep 1), males and females in *Extr. Prev.* only started later in the task. This could be a result of the *Extr. Prev.* stimuli being more confusing or perceptually and representationally challenging. An effect of `Superiority` can be seen among females in their accommodation to /p/ and among males for both /p/ and /b/. The other two likeability measures

could not be linked to accommodation. Ethnicity was correlated with prevoicing in /b/ productions: people of color (particularly Black / African-American participants) tended to prevoice more.

Patterns in the treatment of the /p b/ contrast in the Extreme Prevoicing shadowing data

We see a minimal overlap of the /p b/ categories in *Extr. Prev.* (all partificpants pooled together in *Figure 3.41*), which is similar to what we observed in *Extr. Asp.*. The plot of all *Extr. Asp.* tokens is repeated as *Figure 3.42* to facilitate the following comparisons between the two conditions. Since the /p/ category in *Extr. Prev.* (*Figure 3.41* on top) is not pulled upwards (towards more aspiration), the maxmimum of the range of /p/ productions is lower here than it was in *Extr. Asp.*—i.e. the most aspirated tokens are not as aspirated as they were when participants were exposed to 130 ms VOT /p/ tokens. At the same time, the minimum (least aspirated tokens) does not seem to be much lower. Thus, in *Extr. Prev.* we see a narrower range in /p/ production, with a slight downward shift in concentration.

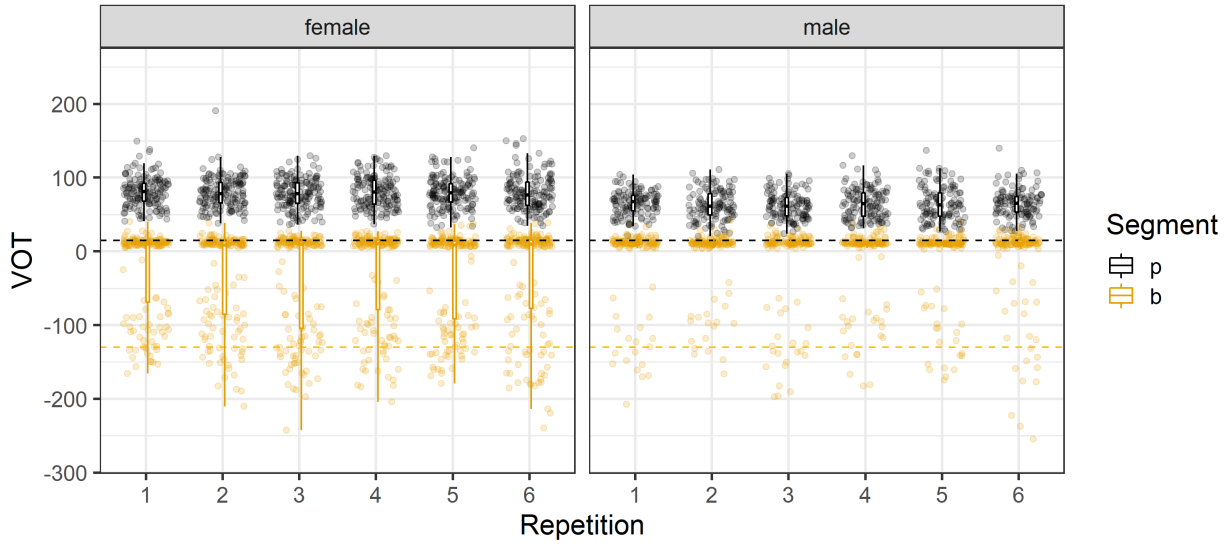

Figure 3.41: All participants' shadowing productions in Extr. Prev

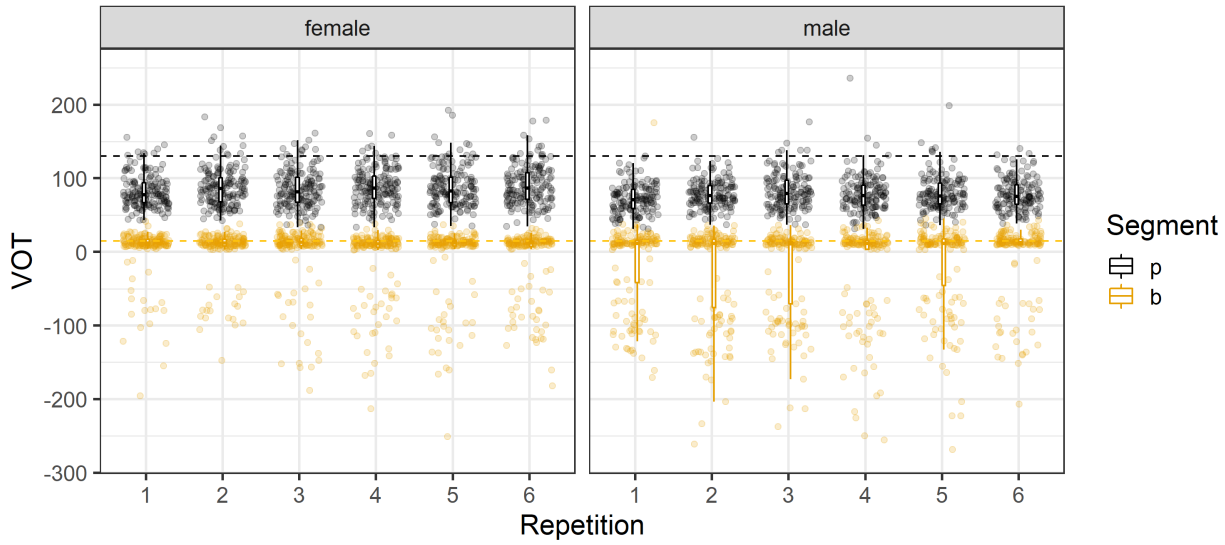

Figure 3.42: All participants' shadowing productions in Extr. Asp.

As for /b/'s, males and females in the two conditions compare differently. While we see more prevoicing in females' production in *Extr. Prev.* (*Figure 3.41*, top) than we did in *Extr. Asp.* (*Figure 3.42*, bottom), there is still a great concentration of /b/'s in the short-lag range, which, in combination with the relatively consistent lower bound of /p/'s still means a relatively clear distinction between /p/'s and /b/'s for females. As a reminder, males in *Extr. Asp.* reacted to the combination of extremely aspirated /p/'s and plain /b/'s by not only aspirating their /p/'s more but also by prevoicing their /b/'s, thereby accentuating the contrast. At the same time, in *Extr. Prev.* we see less predisposition to prevoicing /b/'s in males to begin with, and similarly, their pooled shadowing productions are hardly ever prevoiced. As a result of lack of prevoicing, lower baseline VOT for /p/ than for females, and a lack of pull towards higher ranges of aspiration that was present in *Extr. Asp.*, males' productions in *Extr. Prev.* shows some overlap between /p/ and /b/.

However, just like in *Extr. Asp.*, the minimal overlap on a group level translated to limited overlap on an individual level, as few individuals had in fact touching or overlapping voiced and voiceless categories. There were participants, whose /p/ and /b/ were close to one another but did

not end up overlapping (*Figure 3.43*). There participants exhibited a clear distinction between /p/ and /b/.

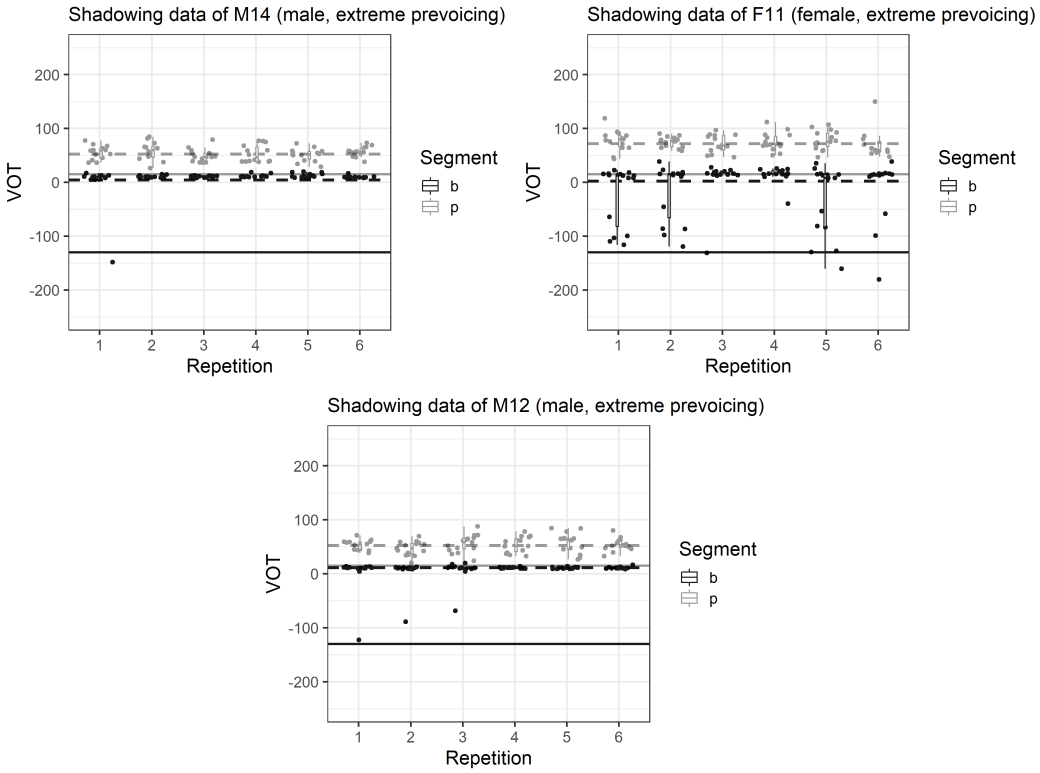

Figure 3.43: Participants whose /p/ and /b/ productions got closer to each other during shadowing
Solid lines of respective colors show the target, and dashed lines show the participant's PRE-Read average

There were two participants, for whom the two categories overlapped, but this was also marginal (*Figure 3.44*). The overlap involved /p/ tokens that were outliers of their category, and thus comparatively few tokens were produced in this "gray area". However, these participants' productions provide evidence of voiceless tokens produced with less aspiration than 40 ms. It is worth noting, that not even in the case of these participants do we necessarily have to stipulate that a shift in their /p b/ boundary occurred as a result of exposure.

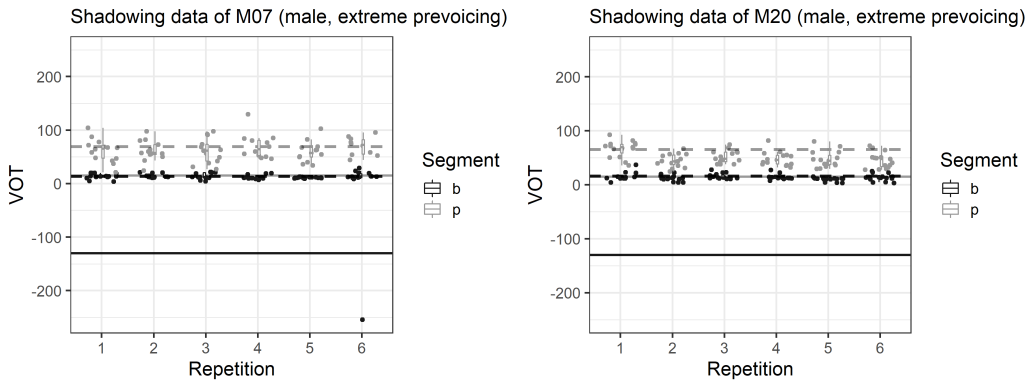

Figure 3.44: The two participants whose /p/ and /b/ productions overlapped in shadowing
Solid lines of respective colors show the target, and dashed lines show the participant's PRE-Read average

There were a handful of participants who ended up distancing their categories from one another during the shadowing task (*Figure 3.45*) either by only moving their /b/'s towards prevoicing (M16 on the left) or both of their categories further from one another (F12 on the right). Distancing in this case can only be observed in terms of shifting the concentration of tokens towards prevoicing, as no participant shifted their short-lag /b/ tokens towards even shorter lags or abandoned plain tokens entirely.

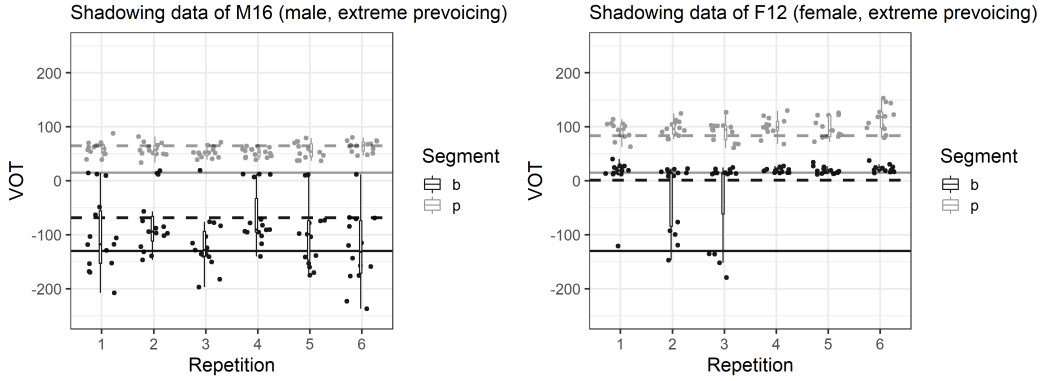

Figure 3.45: Participants distancing their /p/ and /b/ productions during shadowing
Solid lines of respective colors show the target, and dashed lines show the participant's PRE-Read average

There was one participant who shifted her categories in parallel (*Figure 3.46*). F23 shifted both her /p/ and /b/ downward in tandem, maintaining a distinction between the two. However,

aside from her, there was no other evidence of the voiced and voiceless stop categories moving in tandem.

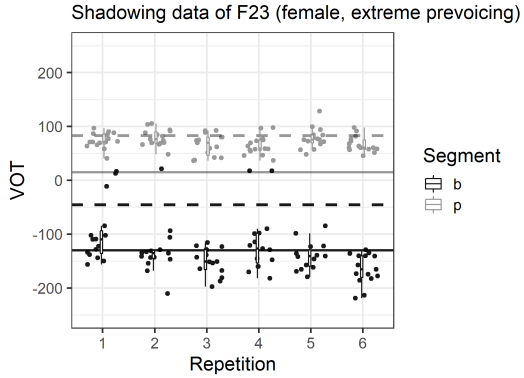

Figure 3.46: The participant shifting her /p/ and /b/ in parallel during shadowing
Solid lines of respective colors show the target, and dashed lines show the participant's PRE-Read average

There was also a pattern of a V-shape for prevoicing of /b/'s attested by a few individuals (e.g. *Figure 3.47*). F18 started prevoicing more towards the middle of the task, but then abandoned it. As discussed before, such behavior could be an indicator of a decrease in either ability or motivation to maintain consistently long prevoicing throughout the task. At the same time it also indicates that the participant had to "warm themselves up to" the stimuli, which is in contrast with some males in *Extr. Asp.*, who attempted sustained prevoicing from the first repetition on (but then also abandoned it).

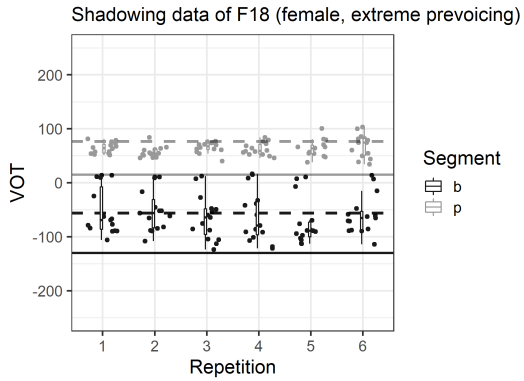

Figure 3.47: A participant showing a V-shape in their /b/ productions during shadowing
Solid lines of respective colors show the target, and dashed lines show the participant's PRE-Read average

As we have seen throughout this section, individual patterns are often non-monotonic and blurred by repetition-to-repetition variation. There is also high levels of individual variation in terms of reactions to the stimuli, which allows no pattern or patterns to emerge as dominant—unlike in *Extr. Asp.* where most participants followed either flat and converging trajectories for /p/ (and for /b/ for females) and flat or diverging for /b/ for males. Shadowing performances in *Extr. Prev.* also echo the findings of the reading task in the same condition, where no dominant pattern(s) of either convergence or divergence emerged.

All in all, while some participants did seem to endorse the un-English-like stimuli and converged, which indicates that it is possible to perform convergence to cross-categorical tokens, followers of this pattern were in a minority, and thus further research is required to establish this phenomenon. At the same time, we did find encouraging results of `Superiority`-based likeability effects, which points to a promising direction for potential further research.

3.4.4 Summary

During the shadowing task, there was a big difference between the behavior of participants in the *Extr. Asp.* and those in the *Extr. Prev.* conditions. In *Extr. Asp.*, participants imitated the 130 ms VOT /p/ target gradually. Repetitions 4, 5, and 6 all had longer VOT's than the baseline, Rep 1 (by 6.618 ms, 6.395 ms, and 8.346 ms, respectively). While females used a wider range of VOT values and tended to produce longer VOT's than males, this difference was not significant. Since a lot of participants already matched the 15 ms VOT /b/ target in their PRE-Read, the investigation of the /b/ dataset was restricted to participants who did not, and thus had room to demonstrate convergence. In this dataset, females did not show changes *within* the task, but were roughly matching the target, whereas males actually diverged from it. Compared to Rep 1, they produced longer VOT's in Rep 4 by 29.368 ms and in Rep 6 by 35.035 ms. When we looked at individual trajectories this turned out

to be a result of males prevoicing substantially in the baseline Repetition (Rep 1) and then stopped gradually.

However, this could be an effect of dialectal background rather than being a gender-effect, since while females in *Extr. Asp.* were mostly white and from the East-coast or California, males were both ethnically and geographically more diverse. People of color in this study prevoiced their /b/'s more overall than white / Caucasian participants did and different dialectal backgrounds or baselines could indeed lead to differences when realizing prosodic prominence (emphasis). Speakers of dialects that prevoice more, tend to exhibit more prevoicing in emphasized contexts, while speakers of dialects that typically do not prevoice post-pausal voiced stops produce few prevoiced tokens (if any) when emphasizing the given word. The effect we see here could thus be the result of all participants in *Extr. Asp.* potentially interpreting the plain /b/ and aspirated /p/ stimuli as being emphasised, and producing emphasized tokens in response to them. These emphasized tokens could have then looked different based on the given participant's dialectal background (prevoiced for "prevoicers" and plain for "less prevoicers". Because of the gender-dialect confound this could then show up as an apparent gender effect. Likeability effects were not found in *Extr. Asp.*

While convergence was the most common general pattern in *Extr. Asp.*, in *Extr. Prev.*, participants as a group were not found to converge with the model talker's 15 ms VOT /p/'s or −130 ms VOT /b/'s. However, in the /p/ dataset, we saw convergence contingent on `Superiority`. The higher participants rated the model talker for these measures, the shorter lag VOT they produced on their /p/), but this was only exhibited by males. A similar pattern was found for /b/, but not just for males, but females as well (once an outlier female, F23, was excluded). Participants prevoiced more if they rated the model talker high on `Superiority`-related measures—participants prevoicing at all were also ones who rated the model talker 7.3 or higher. Females also exhibited a more general, socially independent effect as well: they tended to prevoice more in Rep 3 than in the baseline

Rep 1. This indicates a V-shape pattern, where participants started to converge with the target (i.e. prevoice) toward the middle of the task, but this prevoicing disappeared toward the end of the task. Just like in *Extr. Asp.*, people of color (especially Black / African-American participants) prevoiced the most often in *Extr. Prev.* as well.

3.5 Labeling results

Participants completed a labeling task twice: once at the start and then at the end of the experiment. The PRE-Labeling task was carried out at the beginning (after the rating task, but before the PRE-Read and the shadowing task), and the POST-Labeling was toward the end of the experiment (following the shadowing and POST-Read tasks, right before the questionnaire). In both tasks, participants were exposed to words synthesized on a continuum between *binning* /ˈbɪnɪŋ/ and *pinning* /ˈpɪnɪŋ/, and participants had to choose which word they heard. The VOT continuum involved 11 steps from -60 ms to 90 ms, in 15 ms increments. The *Extr. Asp.* data consists of 2310 pre-exposure and 2310 post-exposure responses, and the *Extr. Prev.* data consists of 2200 pre-exposure and 2200 post-exposure responses.

In general, we can see that participants drew the boundary somewhere between 15 ms and 30 ms before exposure. *Figure 3.48* represents the labeling data as a proportion of /b/ responses. The circles (for *Extr. Asp.*) and the squares (for *Extr. Prev.*) show what proportion of the responses was *binning* at any given step—i.e. how often participants thought the stimuli at a given VOT step started with a /b/. The smoothed curves are fitted to the raw data of /p/ and /b/ responses, not just the proportions represented by the circles and squares. As a result, the curve might not always fit the line that could be drawn by connecting the symbols.

Irrespective of condition or whether the response was recorded before or after exposure, stimuli starting with a 15 ms VOT stop were overwhelmingly categorized as /b/-initial (*binning*), whereas 30 ms VOT were almost exclusively labeled as /p/-initial (*pinning*). The curve has been

smoothed with binomial smoothing. In the first half of the curve (at negative VOT's) the probability is around 1, which means that participants almost always chose *binning*. With longer lag stimuli, the probability is around 0, i.e. participants almost never chose *binning* (they almost always chose *pinning*). The dots (for *Extr. Asp.*) and squares (for *Extr. Prev.*) show the proportion of responses that were '*binning*' at any given step across the given group.

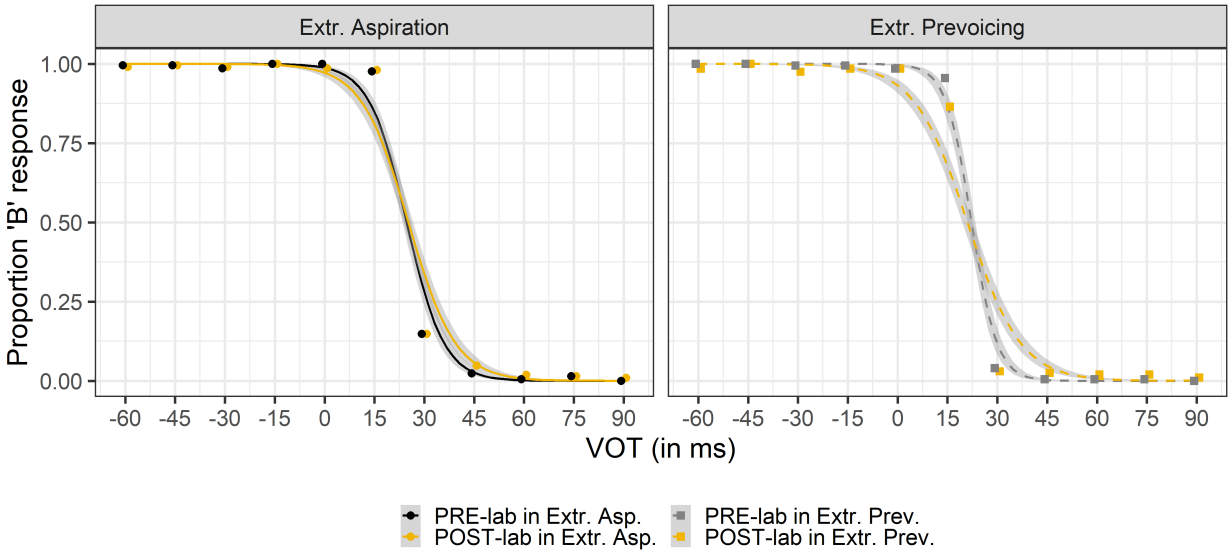

Figure 3.48: Labeling performance by condition

While the stimuli in *Extr. Asp.* had no demonstrable effect on labeling performance (left-hand plot in *Figure 3.48*), we see a change from PRE-Labeling to POST-Labeling happened at the 15 ms step in *Extr. Prev.* (right-hand plot). On average, participants in this condition (as a group) were less likely to call a stop with 15 ms VOT a /b/ after being exposed to short-lag /p/'s and long-lag /b/'s than they were beforehand. Before exposure 95.5% of the time participants labeled the stimulus with the 15 ms VOT stop as *binning* (191 out of 200 tokens), but this dropped to 86.5% after exposure (173 out of 200 tokens). This difference is significant (p=0.0143, with R's `lsmeans` package, the p-value is Tukey-adjusted). The change goes in the expected direction: before exposure they almost unanimously judged a 15 ms VOT stop as a /b/ (typical for English), after being exposed

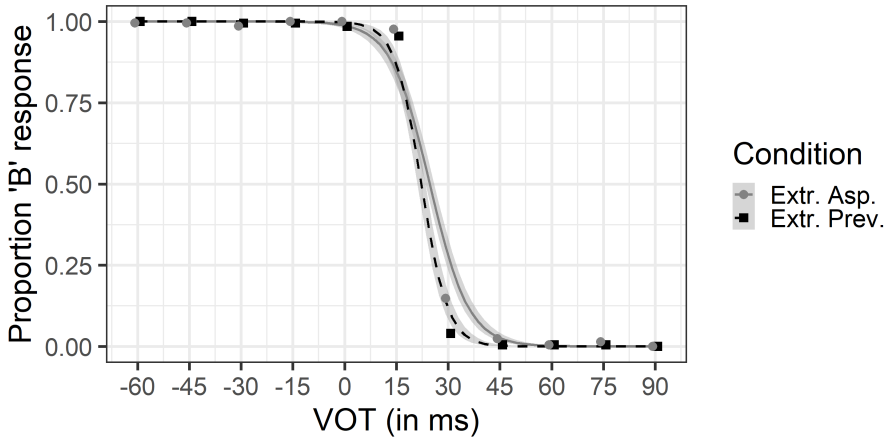

Figure 3.49: PRE-Labeling performance by condition

to someone who produced 15 ms VOT /p/'s and very prevoiced /b/'s participants had to revisit their English-based categories and were somewhat less likely to judge a 15 ms VOT stop as a /b/.

The change in *Extr. Prev.* is not a result of a sampling error or a confound. The pre-exposure response-rate to the 15 ms VOT token was comparable between *Extr. Asp.* and *Extr. Prev.* participants *Figure 3.49*, the difference only showed up after being exposed to their respective stimuli. While participants in *Extr. Asp.* and *Extr. Prev.* responded *binning* to 15 ms VOT stimuli at slightly different rates—97.5% (205 out of 210) and 95.5% (191 out of 200), respectively—this difference was not significant (p=0.6510). [5]

On an individual level, the change in *Extr. Prev.* emerges from two tendencies from a total of 8 participants. This group is ethnically diverse (they self identified as: "South East Asian, Middle Eastern", "African-American", "Black/African-American", "Asian/Hispanic", "Chinese", "Asian", "white", "Afro-Latina and white"), therefore no link can be established between perceptual shifts and ethnicity in this task. First, 5 participants (out of the total 20 in this condition) categorized one fewer 15 ms token as *binning* after exposure than they did beforehand.

[5]While there is a pre-exposure difference between participants in the two conditions for 30 ms stimuli (p=0.0029), neither groups changed from these values after exposure (p=1 for *Extr. Asp.* and p=0.9486 for *Extr. Prev.*).

Most of these people went from labeling all 10 15 ms VOT stimuli as *binning* pre-exposure but only 9 afterwards. One person (M09) went down from 8 to 7. They are shown in *Figure 3.50*. It is important to note that while most of them labeled 10 out of 10 tokens as /b/ to begin with, this was not necessarily the case: F23 labeled only a minority of 15 ms VOT tokens (3/10) as /b/, and M09 labeled them as /b/ 8 out of 10 times. It must be noted that they were the only two participants labeling less than 9 tokens of 15 ms VOT stimuli as /b/ among all participants (including both conditions). All considering, the change in the labeling performance of these five people is small enough to be explicable by errors (e.g. due to them accidentally pushing the wrong button).

The other tendency involved a larger shift. There were three participants, whose treatment of the 15 ms VOT stimulus changed by more than one judgement. They are shown in *Figure 3.51*. It is likely that the change in these three participants' labeling performance reflects a real perceptual shift of their category boundaries. They all started from labeling 10 out of 10 15 ms VOT tokens as *binning*. While one of them (F11) only shifted to 8 out of 10 afterwards, the other two (F24 and M14) demonstrated a larger shift. They went from labeling stimuli with 15 ms VOT categorically

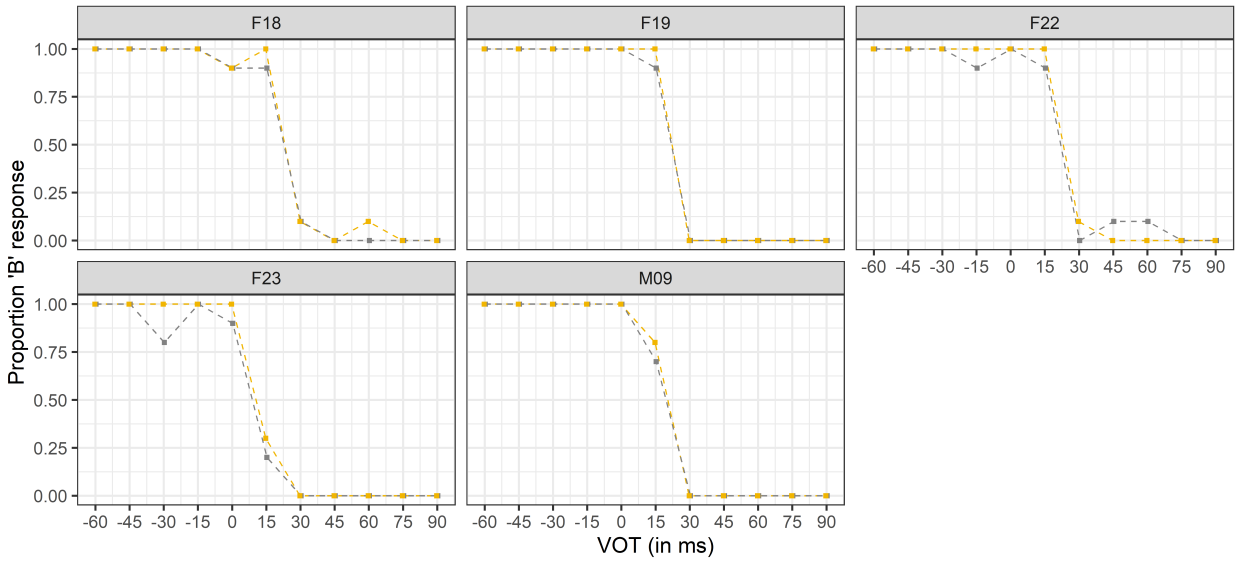

Figure 3.50: Participants in Extr. Prev. who labeled one fewer 15 ms VOT token as /b/ post-exposure
Gray is pre-exposure, yellow is post-exposure

as /b/-initial, to labeling it as such only half the time or even less. However, it must be noted that not even in case of these bigger changes did the participants completely reassign the category label of any VOT step (i.e. no participant went from categorically recognising certain step as /b/, but categorically recognising it as /p/ after exposure or vice versa), changes only introduced ambiguity at previously unambiguous steps.

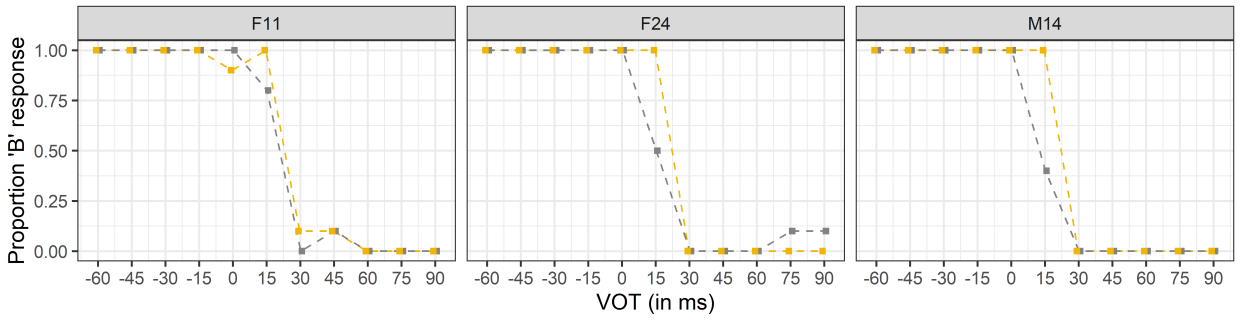

Figure 3.51: Participants in Extr. Prev. who labeled many fewer 15 ms VOT tokens as /b/ post-exposure
Gray is pre-exposure, yellow is post-exposure

While connections based on three participants are somewhat spurious, we can see some commonalities in their shadowing productions. The perceptual shift occurring for these participants was reflected in their production during the shadowing task, and we can see common tendencies. They all converged to a plain /p/ during the shadowing task. Since they did not change their /b/ productions, their /p/ and /b/ got closer to each other. There was no change in any of their reading productions.

However, this connection only went one way: while the perceptual shift could be linked to a certain production profile, no part of their production predicted the perceptual shift per se. In the rest of this section I will demonstrate that neither producing short-lag /p/'s nor a closeness of /p/ and /b/ was a reliable predictor of perceptual changes. First, the perceptual shift we did see for F11, F24, and M14 is not a direct result of having a short(er)-lag /p/, even though /p/ encroaching on /b/ might otherwise challenge participants to adjust their perceptual boundary. This can be seen from the fact that participants with the shortest VOT's for /p/ (e.g. F17, F18, M12, or M20) did not show

a perceptual shift. Second, the perceptual shift did not correlate with /p/ having such short lag that /p/ and /b/ overlap either. The three participants showing a perceptual shift did not have any actual overlap between their /p/ and /b/ in shadowing. At the same time, there were two participants who did have a marginal amount of overlap between their /p/ and /b/ (M07 and M20, *Figure 3.44*), but there was no indication of even a minor perceptual shift in these people's labeling data.

At the same time, it is not surprising that perceptual shift went together with /p/ convergence (to the extent it did) and not with /b/ convergence. Convergence to a prevoiced /b/ involves an exaggeration of a cue, and if occurring without a change in /p/, it moves the contrasting categories further from each other. Therefore it does not necessitate a perceptual shift. Indeed, perceptual shift was not correlated with more prevoicing in /b/ productions in either direction. The /b/'s of the three participants showing bigger perceptual shifts did not seem to move, and the participants who seemed to converge with the model talker's prevoiced /b/'s, did not show any evidence of their perceptual boundaries shifting between the PRE-Labeling and POST-Labeling tasks. For instance, M16 prevoiced his /b/'s so much, he often even overshot the target, but his labeling performance did not change. Cases like M16's show that there were participants who perceived the prevoiced /b/'s accurately enough to produce more prevoicing themselves, but did not apply this information to pin down the model talker's categories more closely. They did not use it to enforce a strict restriction that her /b/'s must be prevoiced, and any non-prevoiced stop must be a /p/ for her.

Similarly to the prevoiced /b/ targets in *Extr. Prev.*, converging with the short-lag /b/ target or the long-lag /p/ target in *Extr. Asp.* did not require an adjustment of English-like category boundaries. Thus, the lack of effect of exposure in the *Extr. Asp.* labeling data is not surprising either.

3.6 Interim discussion

In this section, I will summarize the results of the English experiment, and take a look at how they inform each of the questions set out at the beginning. In the end, I will also discuss how results from native English speakers affect things going into the Hungarian experiment: what aspects of the Hungarian native speakers' performance will be of particular interest given the results from the English experiment.

3.6.1 Summary of results

The English experiment included a reading, shadowing, and labeling component. In the reading task, participants converged with the 130 ms VOT /p/ target of *Extr. Asp.* as well as with the 15 ms VOT /b/ target, albeit the latter was obscured by a ceiling effect (participants who matched the target even without exposure could not get any closer to it). Much less change was seen in *Extr. Prev.*, where stimuli were a plain /p/ (15 ms VOT) and a prevoiced /b/ (-130 ms VOT). While no statistically significant convergence was found for either /p/ or /b/ at large, there was a correlation between `Solidarity` and POST-Read /p/ productions. This mostly surfaced in the form of participants, who rated the model talker to be ruder, unfriendlier, and less honest tended to diverge (aspirate their /p/'s) more than those who did not.

The shadowing task paints a little more complicated picture, but the main tendencies were similar. In *Extr. Asp.*, /p/ targets were imitated gradually by most participants. As for /b/, the two genders showed different tendencies. While females converged with the plain /b/ target, males who did not match the target to begin with diverged from it, but only initially. Thus the profile of these males ended up being as follows. In addition to converging with the /p/ target (aspirating more), they initially also prevoiced their /b/'s more, but over the course of several repetitions, the initial prevoicing tended to disappear. This pattern was likely the result of dialectal differences between males and females: the subset of participants who identified as male were from a more diverse

(and perhaps more likely to prevoice) background than the mostly white participant group (mostly from the Mid-Atlantic and the Northeast), who identified as females. For speakers of dialects that use prevoicing more often, emphasis could have involved more prevoicing, whereas speakers of dialects that do not tend to have that much prevoicing tend to prevoice *less* in prominent positions. Therefore, it is possible that all participants simply expressed more emphasis on their /p/'s and /b/'s, but executed that differently depending on their dialectal background.

This resulted in some participants diverging from the model talker's plain /b/ targets (by prevoicing more) when in fact they converged with the expressed prominence (in response to emphasized stimuli, they produced emphasized tokens). This suggests that the target of accommodation might not always be the acoustic signal (the raw phonetic data itself) but the pattern therein. Similar results were found by Nielsen and Scarborough (2019). When forced to choose, their participants converged with the nasalization pattern of the model talker even at the cost of diverging from the model talker's raw values in terms of Nielsen and Scarborough's nasalization metric (A1–P0).

In *Extr. Prev.*, participants as a whole were not found to converge with the model talker's plain /p/ and prevoiced /b/ productions during the shadowing task. However, there were some more fine-grained patterns. In terms of /p/, males who rated the model talker higher on `Superiority` produced tokens with shorter lags (approached the target more closely) than those who rated her lower. As for /b/, females as a group had significantly more prevoicing in Repetition 3 than before or after. This means that prevoicing trajectories often showed a characteristic V-shape: participants started prevoicing more and more, but then stopped, resulting in a prevoicing peak (a minimum along the VOT axis) towards the middle of the task. An effect of `Superiority`, similar to the effect found for /p/, was found for /b/, but in this case it applied to all participants, not just males. Participants who gave the model talker high `Superiority`-related ratings ended up prevoicing more. In fact, almost all participants who prevoiced rated the model talker 7.3 or above. It should be noted that instances of less aspiration for /p/ and more prevoicing for /b/ than in the reading task

cannot necessarily be interpreted as convergence. Since shadowing and reading productions cannot be directly compared in a meaningful way, only instances where there is a non-flat trajectory (e.g. the V-shapes for prevoicing) can be considered convergence.

The only change in labeling performance happened in the `Extr. Prev.` condition: participants were less likely to rate a token with 15 ms VOT as /b/ after being exposed to the condition's 15 ms /p/ and −130 ms /b/ stimuli. This tendency was observed in the production of 8 participants, with 3 participants contributing to the bulk of the change. Two of these went from always labeling 15 ms VOT tokens as /b/ to only doing so about half of the time. The three major contributors converged with the model talker's /p/ in shadowing (and not with her /b/). This relationship between perception and production was only one-sided: no specific production profile could be isolated to predict the change in labeling performance. No link was found between labeling performance and read productions.

3.6.2 Mechanisms of contrast maintenance

In *Chapter 2*, I defined two hypotheses for what mechanisms are responsible for contrast maintenance. The first relies on **Maintain contrasts**, an abstract pressure to keep categories that are contrastive separate in production. Under this hypothesis, there are no restrictions on what a member of a category might look like, as long as the category is different from other contrastive sounds in the language. The dimension(s) along which the sounds have to differ might potentially be specified as well as which of the two contrastive sounds have higher values along these dimensions. Aside from that, this hypothesis presumes that the imperative to maintain a distinction between members of a contrast is largely abstract and flexible in terms of phonetic realization. Therefore under this hypothesis, a speaker may be able to change the phonetic space of a category within a dimension so long as contrasts are maintained by changing all categories that also use that dimension—e.g.

/p/'s VOT can be shifted downward but only as long as it does not clash with /b/—either because the change is small or because /b/ is also shifted further down.

The other hypothesis relies on **Maintain categories**, which requires each individual sound category to stay phonetically consistent within the speaker's productions. Thus it imposes phonetic restrictions on tokens from these categories. Thus the categories stay relatively static over time, which also results in contrasts being maintained, albeit in an indirect way.

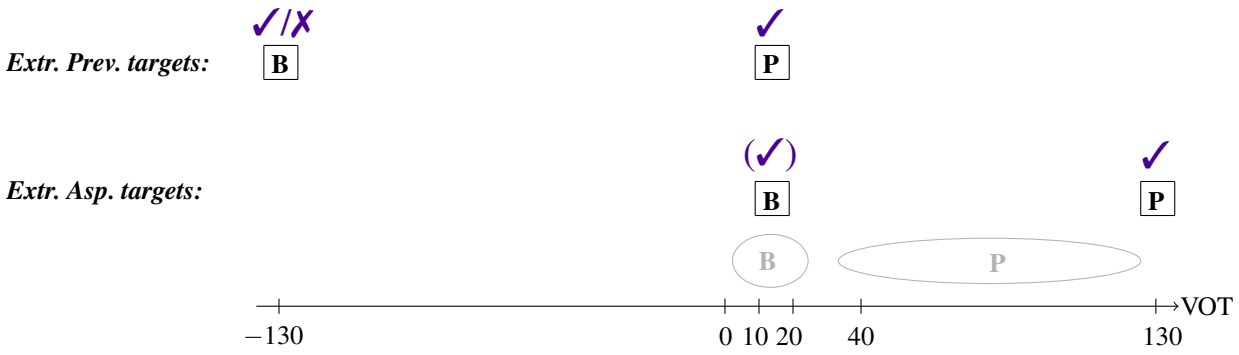

Figure 3.52: Predictions of the maintain contrasts hypothesis for English
Gray: typical English values; ✓: hypothesis predicts convergence; ✗: hypothesis predicts no convergence

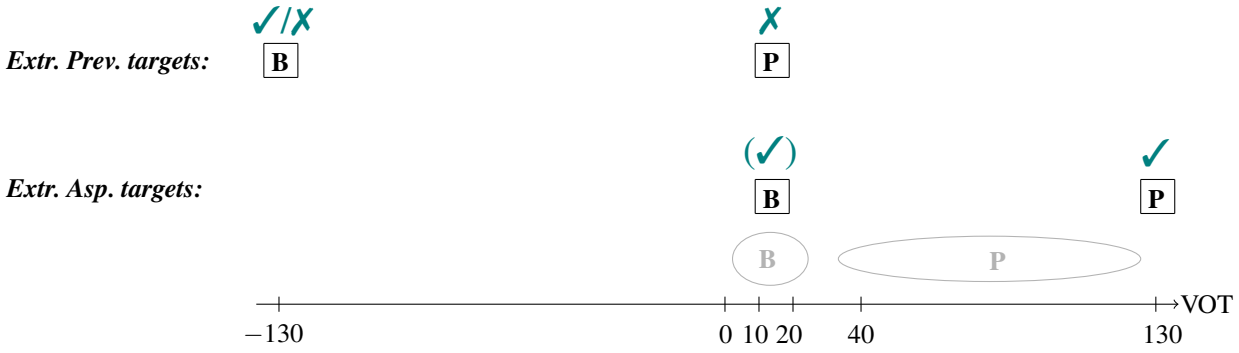

Figure 3.53: Predictions of the maintain categories hypothesis for English
Gray: typical English values; ✓: hypothesis predicts convergence; ✗: hypothesis predicts no convergence

As we saw in *Section 2.4*, these two hypotheses have slightly different predictions for the English experiment (*Figure 3.52 & 3.53*), which I will re-summarize here. The two hypotheses make largely the same predictions for 3 out of 4 stimuli. They both predict that extremely aspirated stops

are a valid target for a /p/, and thus (notwithstanding likeability effects) participants will converge with *Extr. Asp.* /p/'s—marked with '✓'). The plain /b/ stimuli in *Extr. Asp.* (15 ms VOT) do not require any adjustment compared to the values that most English speakers habitually produce, and thus participants will end up being similar to the target, but they likely had been to begin with—as denoted by '(✓)'. The two hypotheses do not make differing predictions for *Extr. Prev.* /b/'s (very prevoiced with −130 ms VOT) either, specifically neither can predict any behavior in particular. Prevoicing is a cue that English does not use systematically, and therefore it is questionable whether participants end up imitating it. This is not for representational reasons, but for reasons of lack of articulatory practice with prevoicing. As a result, we might see accommodation, no accommodation or even interpersonal variation. The non-committalness of the two hypotheses is signalled by '✓/✗'.

The two hypotheses only differ in their prediction for whether participants accommodate to /p/'s in *Extr. Prev.* In this condition, participants were exposed to plain /p/'s (15 ms VOT) and prevoiced /b/'s (−130 ms VOT). **Maintain contrasts** predicts no problem in participants endorsing the 15 ms VOT stop as a target for /p/, because /p/ and /b/ are still clearly distinct from one another ('✓'). On the other hand, **Maintain categories** predicts that participants will not accommodate to such stimuli, because 15 ms is an atypical VOT value for a /p/ in English, in fact, a more prototypical token of /b/, a contrastive category. Therefore it is not a valid target for /p/, and the speaker cannot incorporate these exemplars into their representation, which necessarily means that the speaker cannot accommodate to it ('✗').

In the English dataset we find convergence with *Extr. Asp.* /p/ (130 ms VOT) and among participants who did not match the plain /b/ in *Extr. Asp.* (15 ms VOT) to begin with we see convergence for the voiced stop too. These patterns are present in both the reading and the shadowing data and show no likeability effect. The *Extr. Prev.* dataset is more complicated. Without social variables we do not see any convergence or divergence for plain /p/'s (15 ms VOT), but there does seem to be an effect of `Solidarity` in the /p/ reading data. However, this leads to actual

convergence in very few cases in the dataset. For the prevoiced /b/'s (−130 ms) in *Extr. Prev.* we see no convergence in the reading data ('**R:✗**'), but we see some, especially V-shaped convergence in the shadowing task ('**S:✓**'). We also find an effect of `Superiority` in the shadowing data, but mostly in the /b/ dataset. Therefore, the English dataset supports the **Maintain categories** hypothesis over the more abstract and flexible *Maintain contrasts* hypothesis.

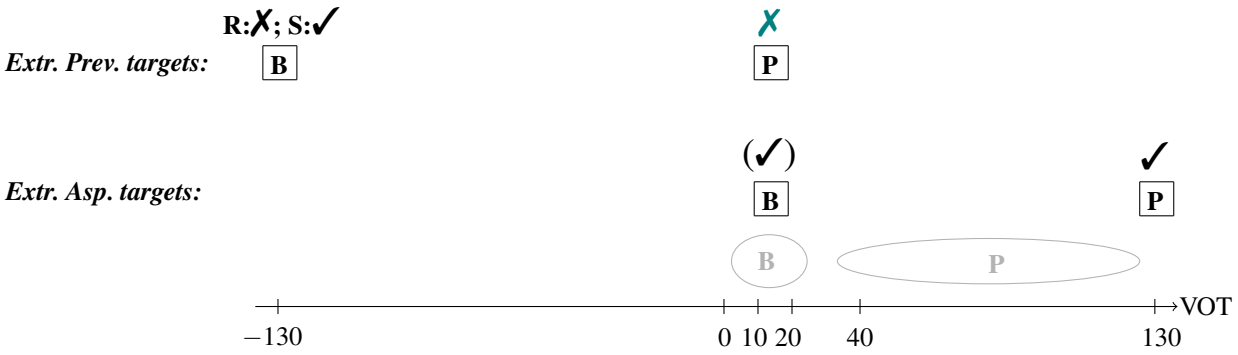

Figure 3.54: Participants' behavior in the English dataset
Gray: typical English values; ✓: accommodation; ✗: no accommodation

In order to differentiate between the two hypotheses, we need to focus on the /p/ data in *Extr. Prev.*, where neither the reading nor the shadowing task found overwhelming evidence for convergence. This could be for two reasons. First, this could be because participants did not endorse the plain /p/ tokens they heard as valid targets. If this is the case, the null result from the study supports the **Maintain categories** hypothesis. Alternatively, this could be a result of prevoicing not being a perceptually adequate cue for English speakers. This could be especially plausible because we found a large group of (especially male) participants, who did not imitate the *Extr. Prev.* prevoiced /b/ tokens, which might also suggest difficulties. It is important to note that if participants only had production issues with prevoicing /b/'s (and their representations were flexible enough), then it would have only affected /b/ productions and participants should have been able to produce shorter lag on /p/'s.

A perceptual difficulty could either surface with not perceiving prevoicing at all, or just not perceiving it as part of the VOT spectrum, and both result in not perceiving a sufficient distinction between /p/ and /b/ and thus either could lead to the results we found in the *Extr. Prev.* /p/ dataset. The former case (not hearing prevoicing at all) is somewhat less likely, because some studies have had success with getting monolingual native English speakers to imitate prevoicing when instructed. Beach et al. (2001) found that participants imitated prevoicing in a task where they had to shadow pairs of contrasting syllables consisting of a bilabial stop and /a/ (e.g. /pa/ — /ba/; /p^{h}a/ — /pa/), and Olmstead et al. (2013) found small but significant amounts of prevoicing when English speakers were shadowing prevoiced tokens on a /pa/ /ba/ continuum. Their results indicate that under certain circumstances (when contrasted with other stops and without lexical information) native English speakers can, in fact, perceive prevoicing or a lack thereof well enough to imitate it. It must be noted that it is uncertain, if this sensitivity is something English speakers also similarly perceive in the context of words in their native language (rather than syllables from a speaker of a foreign language). The Olmstead et al. (2013) thus provides weak evidence for English speakers having been able hear the prevoicing itself. Therefore any perceptual issue around prevoicing is less likely to be related to English speakers not hearing prevoicing (i.e. literally not being able to distinguish prevoicing from silence as such).

However, hearing a cue does not necessarily mean that one also perceives it as a part of a particular spectrum. In this case, perceptual issues could stem from participants not identifying prevoicing and aspiration as two ends of the same spectrum (VOT), but seeing them as two different kinds of cues entirely. If English speakers assume that a voicing contrast must be along VOT, but do not recognize prevoicing as simply a negative VOT value, then a contrast between a plain /p/ and a prevoiced /b/ does not even satisfy the looser requirement posed by **Maintain contrasts**, let alone **Maintain categories**—i.e. that the two sounds must stay distinctive along VOT.

Another way of expressing this is that maybe the phonology of English speakers defines a voicing contrast between stops as one of aspiration rather than VOT—i.e. "how long after the burst did the following vowel's voicing start". For instance, the heavily prevoiced /b/ stimuli in *Extr. Prev.* has 0 ms of aspiration, and the *Extr. Prev.* plain /p/ has 15 ms aspiration. This difference of 0 ms vs. 15 ms is quite small, and participants could conceivably think that the model talker did not maintain the /p b/ contrast in her speech. That is a violation even under a more abstract contrast preservation principles (like **Maintain contrasts**), and thus the model talker's targets are not valid, especially the plain /p/, which is very different from English aspirated /p/.

It is important to note that participants might still hear the prevoicing on /b/, but do not consider it to be on the same spectrum. To give a rather extreme example, English speakers do not consider aspiration to be on the same scale as intensity either. While English speakers are perfectly capable of hearing (and converging in terms of) intensity, an intensity-based /p b/ contrast would be dissatisfactory for English speakers as well. Combined with the fact that English speakers might not be very good at hearing and producing prevoicing reliably and accurately (because it is not a contrastive cue in English), this perception issue could also lead to the lack of convergence we saw in *Extr. Prev.* in the English experiment.

Thus, we are left with two competing interpretations for why participants did not converge towards the plain /p/ in the *Extr. Prev.* stimulus. First, an adherence to phonetic detail of sound categories (**Maintain categories**) could be part of the participants' grammar, and the plain /p/ stimulus was violating that principle by encroaching too much on the typical location of /b/. Second, English speakers might not perceive prevoicing as the negative end of the VOT continuum, but as its own cue itself, which resulted in an insufficient distinction between /p/ and /b/ in terms of the timing of the stop burst and the start of the vowel. The Hungarian dataset could help disambiguate between these two interpretations. How exactly will be discussed, later, in *Section 3.6.11*.

3.6.3 Categories moving together

The Engish experiment did not find much correlation between the treatment of /p/ and /b/ on an individual level. In *Extr. Asp.* there was across the board convergence with /p/, and /b/ was more limited by some participants simply being so close to the target to begin with that they could not demonstrate convergence by matching the target after exposure. In *Extr. Prev.*, participants did not show much convergence in either the reading or the shadowing task, and exhibited various strategies in whether they moved their VOT downward or upward for the two sounds and correlation between the two could not be established.

3.6.4 Task effects

Participants' productions differed in the reading task and in the shadowing task. The shadowing task in general, elicited values that were closer to the target (*Figure 3.55*), but this is likely due to the difference in task than anything else, and thus, cannot be directly compared. This was also observed on an individual level: most (but not all) participants producing tokens closer to the target during shadowing than in the reading tasks.

However, there were also cases where accommodation was only temporary. Participants exhibiting a V-shape trajectory during shadowing converged to the model talker for a short time, but then abandoned it in subsequent repetitions within the same task (and did not show any exposure effect in the POST-Read). This was mostly observed while participants were shadowing the prevoiced /b/ target in *Extr. Prev.* Such participants could be qualified as strictly immediate convergers, because while their reading productions suggest that the exposure had no effect on them, they did in fact show exposure effects while directly exposed to the stimuli.

The reason why this pattern was especially prolific in *Extr. Prev.* and for /b/'s at that could stem from multiple things. This behavior could be a result of articulatory challenges around prevoicing. Alternatively, it could be indicative of participants exhibiting more convergence while

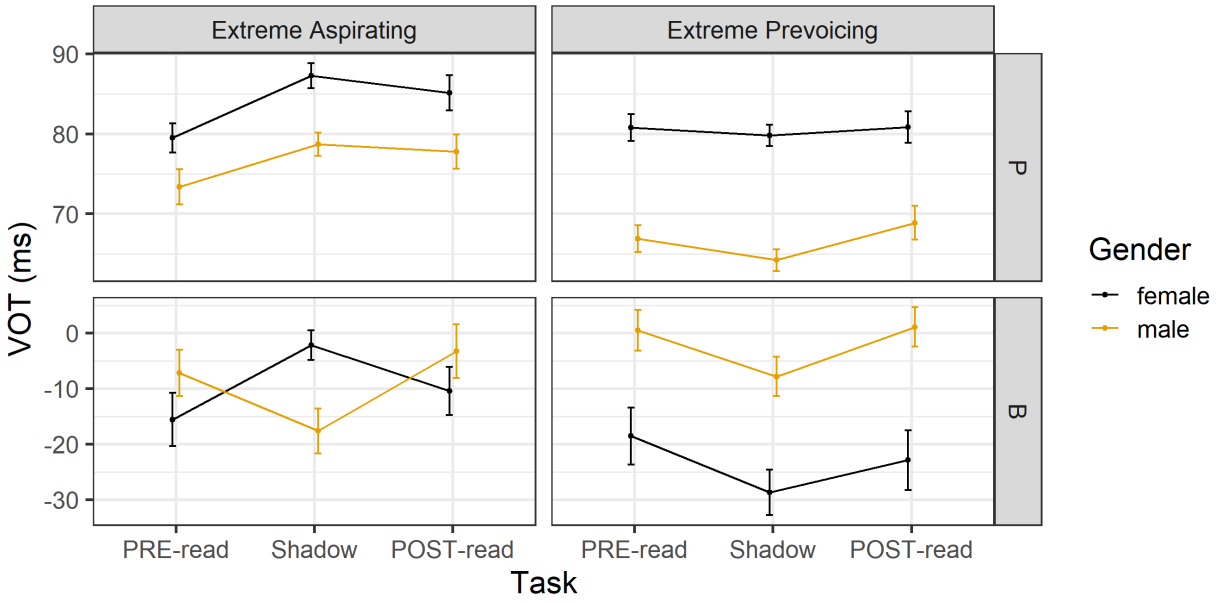

Figure 3.55: All productions from all English-speaking participants averaged by task: PRE-Read (2 reps), Shadowing (6 reps) and POST-Read (2 reps)

directly exposed to the model talker, but not once she is "gone", especially since in *Extr. Prev.* the model talker's productions were peculiar or un-English-like. In real life such behavior would be unsurprising: approximating atypical productions could lead to social gains (face saving) in a face-to-face context, but there would be no social benefits in maintaining adjustments to atypical productions of a one-off interlocutor.

3.6.5 Shift of boundaries

Only small changes were found in the categorization of short-lag VOT stops and only in the *Extr. Prev.* condition. In the *Extr. Asp.* condition no changes were seen. In *Extr. Prev.*, participants were slightly but significantly less likely to label stimuli consisting of 15 ms VOT stop + ɪmŋ as a *binning* (rather than *pinning*) after being exposed to /b/'s with 130 ms prevoicing and /p/'s with 15 ms aspiration from the same talker (95.5% /b/ labels pre-exposure and 86.5% afterwards). This effect was based on changes in 8 (out of 20) participants' performance, with the bulk of the

change coming from 3 people. While the 3 participants who showed bigger changes in labeling did converge to the model talker's /p/ productions, this did not carry over to the other 5 participants whose labeling performances only reflected a little change. There were also participants, whose production changed without their labeling production being affected by it. Therefore we could only establish a weak one-way relationship: if one's performance for 15 ms VOT labeling stimuli underwent a bigger change, the participant also demonstrated some convergence to the 15 ms VOT /p/ stimuli, but a change in labeling performance was not necessary for producing shorter-lag /p/'s post-exposure.

3.6.6 Articulatory fatigue

English-speaking participants produced both aspiration (long lag) and prevoicing (albeit to a different extent), but they showed different imitation trajectories for the two cues during shadowing. While aspiration was imitated either "all at once" (to the same extent from the first repetition on) or gradually, but in any case was most intense in later repetitions, the use of prevoicing tended to cease after a few repetitions—i.e. it peaked early on in the task.

A decreasing amount of prevoicing later on in shadowing was found in two cases: participants (typically females) converging with the prevoiced /b/ stimuli in *Extr. Prev.* and males in *Extr. Asp.*, who for some reason reacted to the aspirated /p/'s and plain /b/'s in the stimuli by enhancing both cues: producing more aspiration on /p/'s and more prevoicing on /b/'s. There was a difference between these two groups. Convergers in *Extr. Prev.* showed a more V-like pattern (they increased the amount of prevoicing up to a point, then decreased it, resulting in a prevoicing peak in the middle repetitions), while the males in *Extr. Asp.* tended to prevoice a lot early on and gradually lost it as the task progressed. This difference might have to do with the fact that the *Extr. Prev* stimuli were unlike typical English productions (prevoiced /b/ and plain /p/), and thus participants might have needed some time to adjust to them.

The fact that these decreasing patterns were only found for prevoicing, but not for aspiration indicates a difference between the two cues. Maybe participants found prevoicing hard to produce or it at least its production required more effort, and participants lost interest in maintaining that effort as the task continued.

3.6.7 Gender

Male and female participants behaved largely similarly, aside from a few notable exceptions. First, while females converged with both the aspirated /p/ and the plain /b/ target in *Extr. Asp.* shadowing, males only converged with the aspirated /p/ and diverged from the plain /b/, and prevoiced their /b/'s instead. This was likely because males interpreted the very long lag in /p/ (130 ms aspiration) as a general cue for emphasis, as it is often used in English, and generalized that information to the rest of the task, emphasizing (and thus prevoicing) /b/'s in the process. In order to do this, they crucially had to disregard that /b/ tokens in the stimuli were not prevoiced, but had short lag (15 ms) VOT values. However, this is most likely a confound between dialects (i.e. ethnicity and place of birth) and gender rather than a result of males not hearing prevoicing as well as females, since there is no indication in the literature for monolingual English-speaking males' and females' differing in their abilities to perceive prevoicing.

Second, males showed wider effects of `Superiority` in shadowing *Extr. Prev.* stimuli (plain /p/ and prevoiced /b/) than females did (right-hand side and left-hand side of *Figure 3.56*, respectively). While an effect of `Superiority` was traceable in /b/ productions among both males and females (higher `Superiority` ratings cooccurred with more prevoicing), only males showed a similar effect for /p/ (less aspiration, i.e. convergence, going together with higher ratings).

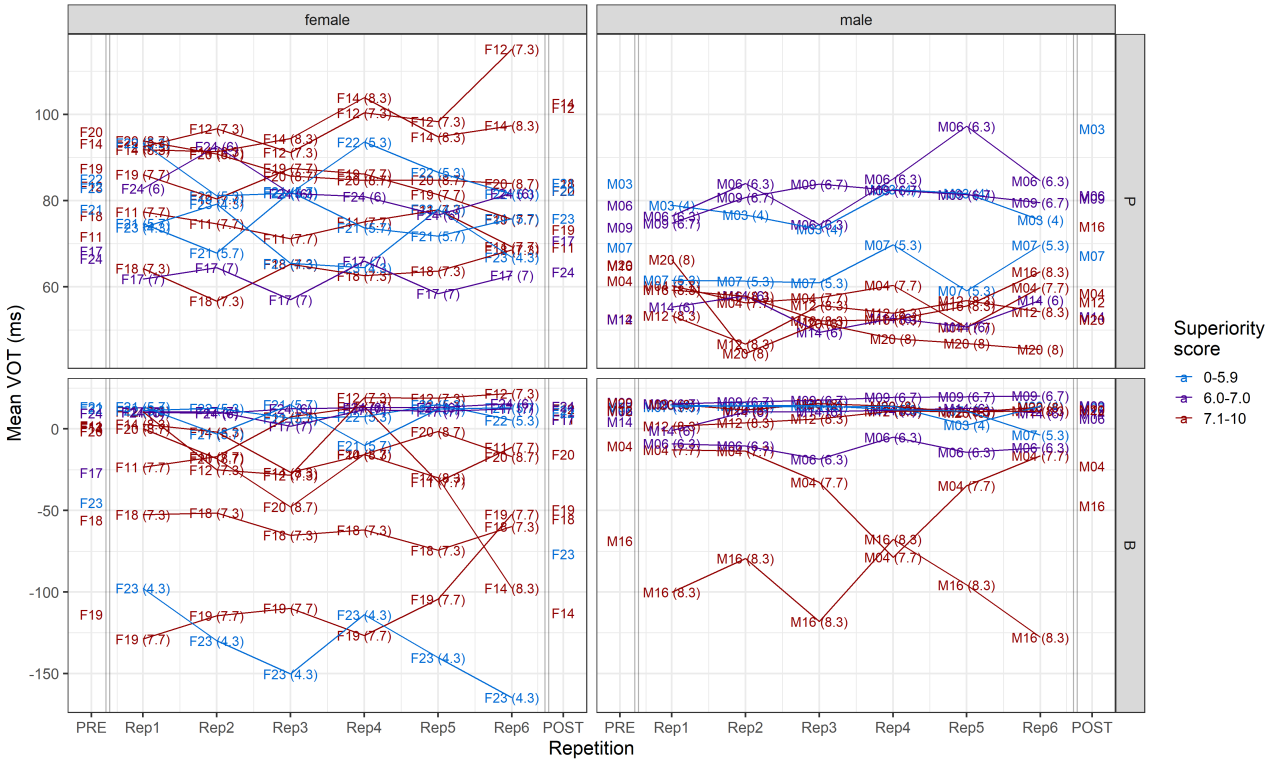

Figure 3.56: Participants' /p/ and /b/ shadowing trajectories by `Superiority` *ratings in Extr. Prev.*

This could be a social effect, in which females were more likely to imitate the model talker's un-English-like cues no matter how high they rated her for *Superiority*-related features, while males only endorsed these target values if they perceived the model talker to be sufficiently intelligent, organized, and of high status. At the same time, there is also a possibility that this effect is due to overfitting, as it relies on few participants' productions.

In general, males tended to use less extreme cues (less aspiration on /p/'s and less prevoicing on /b/'s) during the two reading tasks than females. Aside from the notable example of males in *Extr. Asp.* , who reacted to plain /b/ and aspirated /p/ stimuli by not only aspirating their /p/'s but prevoicing their /b/'s more as well, this was true in shadowing as well. Moreover, most participants whose pre-exposure baseline /b/'s were almost exclusively plain were male, to the point where males in the *Extr. Prev.* dataset prevoiced significantly less before exposure than any other subgroup of

participants. While other studies found that males aspirate less than females, it might be interesting to see if they are also less likely to use the completely optional cue of prevoicing on utterance-initial voiced stops.

The behavior of males and females in shadowing the very aspirated /p/ in *Extr. Asp.* was of particular interest because of Nielsen's (2008; 2011) finding that more males converged with the extremely aspirated /p/'s in her study than females did. This study found no difference between male and female shadowing trajectories. Both males and females overwhelmingly converged with the 130 ms VOT target /p/.

The crucial difference from Nielsen's study is in how extreme the aspiration in the stimuli was. While Nielsen exposed her participants to /p/'s with at least (but usually around) 100 ms VOT, in this study, /p/'s had uniformly 130 ms VOT. What could have lead to fewer females converging in her study that 100 ms was maybe too low of a target for females to demonstrate convergence. While few participants produced a mean of at least 100 ms VOT on /p/'s before exposure, such productions were not uncommon either. Participants producing these values had less room to demonstrate convergence, or perhaps did not even perceive a difference between the model talker and themselves in terms of VOT. By making the target /p/'s have 130 ms of VOT this study allowed females more room to converge, and indeed they did. Thus Nielsen's results were most likely due to a ceiling effect rather than gender-grading in VOT accommodation.

3.6.8 Ethnicity

People of color, especially Black / African-American participants, were more likely to prevoice their /b/'s than white / Caucasian participants did. This finding is replicating previous results of others, e.g. Ryalls et al. (1997), who found that Black participants prevoiced more in a reading task than white participants did. This study found similar results in both the reading and the shadowing task, which suggests that this has more to do with a difference in baseline productions rather than a

different aptitude or willingness to converge. This is reinforced by the fact that we saw comparable trends in both conditions, even though prevoicing was divergent from the 15 ms VOT /b/ target of *Extr. Asp.*, but convergent with the −130 ms VOT target of *Extr. Prev.* These effects might have been intertwined with `Gender`—as females in *Extr. Asp.* were less diverse than males were. There were no ethnicity-related tendencies in the labeling data (the participants who contributed to the group-level change in *Extr. Prev.* were an ethnically diverse group), which is consistent with the idea that accommodation itself was not contingent on ethnicity.

3.6.9 Likeability

Instead of using one monolithic likeability rating, this study examined three different facets of likeability (each coming together from 3 semantic differential scales) in order to find out if any components of likeability contribute more to likeability effects than others. These three facets were `Solidarity` (which was the average of ratings from the *friendly–unfriendly*, *honest–dishonest*, and *rude–polite* scales), `Superiority` (from the *organized–disorganized*, *lower–upper status*, and *intelligent–unintelligent* scales), and `Dynamism` (coming from *shy–talkative*, *unsure–confident*, and *energetic–lazy* ratings). Out of the three, `Dynamism` had no demonstrable effect on accommodation.

There was some indication of `Solidarity` playing a role. The reading task in *Extr. Prev.* indicated more imitation of plain /p/'s when the participant previously rated the model talker high for `Solidarity`-related measures. *Figure 3.57* below shows the relationship between `Solidarity` and how the participants' VOT in the POST-Read compared to their PRE-Read baseline. The correlation in general was weak and could be best described as likeability-conditioned divergence. This means that the correlation manifested most often in the form of diverging from the model talker when (strongly) "disliking" her, rather than as converging when "liking" her.

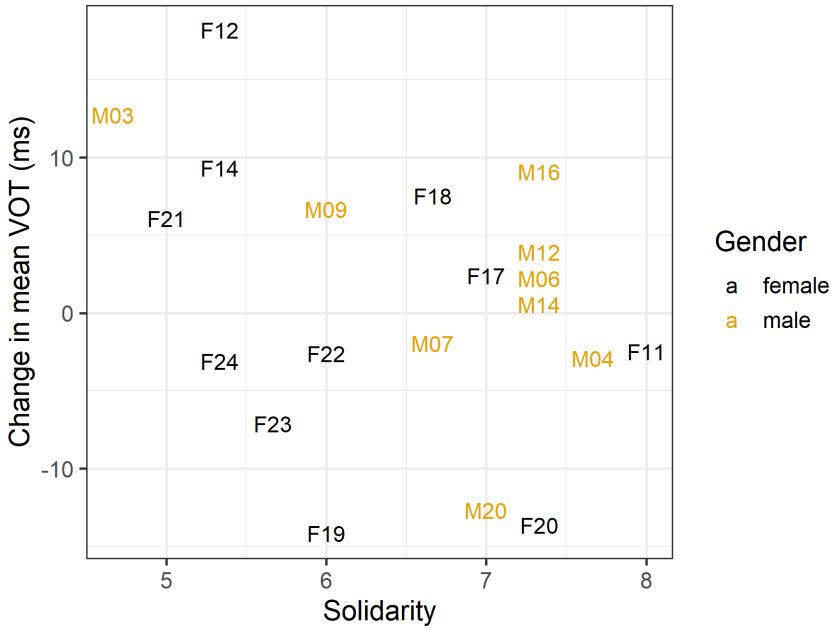

Figure 3.57: Change in mean /p/ VOT in Extr. Prev. in the reading task by gender and `Solidarity` *rating*

Somewhat puzzlingly, in the shadowing section it was not `Solidarity` that proved to be important, but `Superiority`. Since most trajectories were flat and we cannot directly compare shadowing productions with read ones, we cannot necessarily interpret any behavior as convergence or divergence. However, there was a correlation between the values produced and `Superiority` ratings. For /p/, it was only observable in males: males who rated her high `Superiority`-related measures produced shorter lag VOT's (i.e. tokens closer to the target 15 ms VOT) than males who rated her low. A similar behavior was observed for both females and males in the /b/ dataset: participants rating her high produced more prevoicing (tokens closer to the target, which had 130 ms prevoicing) than those who rated her low (see *Figure 3.56* in the previous section).

When hearing the phrase "likeable", `Solidarity`-related measures are the first to come to mind (friendliness, politeness, and honesty). While it has been shown that features involved in the measures which are here called `Superiority` and `Dynamism` also contribute to likeability, they are not the most direct synonyms. Thus, it seems counter-intuitive that in the *Extr. Prev.*

shadowing task it was `Superiority` that showed the strongest correlation with VOT productions. This could stem from the nature of the task itself, as well as from the fact that it was observed in *Extr. Prev.* The stimuli in *Extr. Prev.* with its prevoiced /b/'s and especially its plain /p/'s was not what English typically sounds like. These productions might in and of themselves seem odd or suspicious. Thus, in order to converge with the stimuli they needed to trust the model talker's productions enough to find them valid targets. They needed to "suspend their disbelief" and rely on the model talker's (English) competence. How competent the participants viewed the model talker to be is closely tied to the social features that were compiled into `Superiority` (status, intelligence, and organizedness). The reason for not seeing such effects in *Extr. Asp.* could be that those stimuli were much less "suspicious" than the *Extr. Prev.* stimuli, and thus required less of a leap of faith from the participants.

3.6.10 Attractiveness

While there is evidence for participants treating the question regarding attractiveness in different ways, it was not always clear which approach a given participant followed. Therefore, the information gathered on attractiveness could not be evaluated in this study.

3.6.11 In anticipation of the Hungarian data

The Hungarian experiment can help contextualize the English results in multiple ways. First, the fact that there was no group-level statistically significant convergence in *Extr. Prev.* could be indicative of two different things. One, it can stem from participants' reluctance to endorse such "un-English-like" stimuli, and therefore it can serve as evidence supporting the **Maintain categories** hypothesis over the **Maintain contrasts** hypothesis. Two, it can be a result of prevoicing not being treated as an equivalent part of the VOT spectrum by English speakers, who are only used to it as an optional cue and might even have perceptual difficulties with it.

Running the same experiment with native speakers in Hungarian can help disambiguate. If the former interpretation (adherence to nativelike phonetic specifications of categories) is behind the English results, we would not see convergence in Hungarian for *Extr. Asp.* /b/'s, which with their short lag VOT would encroach on the acoustic space of a typical Hungarian /p/. If, on the other hand English results were due to some perceptual peculiarity of prevoicing, then Hungarians will not have a problem in this case. Since the Hungarian experiment involves a challenge with perceiving aspiration accurately (and its distinguishability from plain stops) rather than prevoicing, perceptual issues regarding prevoicing should not influence the outcome.

In a similar vein, the Hungarian experiment can help to interpret the directional preference found in English: more aspiration was imitated more than less aspiration was. This effect can be either idiosyncratic or systematic on two different levels. An idiosyncratic pattern would mean that no similar effect is found in Hungarian: either we find convergence to all kinds of cues or not. However, this effect can be systematic in that it can tell us something about aspiration—that increased aspiration is in a sense universally preferable over reduced aspiration. In this case we expect the same asymmetry to surface in Hungarian: aspirated /p/'s will be imitated more than plain ones. It can also be systematic in terms of the contrast system: more of a cue that the language already uses is in some sense preferable over less of it. In this case, we expect a similar pattern to be replicated in Hungarian, but not for /p/'s and aspiration, but /b/'s and prevoicing—i.e. more prevoicing will be more commonly imitated than less prevoicing. These two are not mutually exclusive.

We also need to continue to pay attention to any gender-specific patterns as well as any likeability effects that might surface. In terms of likeability, `Superiority` seems to have had the most convincing effect on accommodation (aside from the somewhat spurious `Solidarity` effect in the reading task). This effect was observed during shadowing stimuli in *Extr. Prev.*, where targets were more out of the ordinary than in *Extr. Asp.* Similar effects of `Superiority` could be expected

in the *Extr. Asp.* condition of the Hungarian experiment, since a plain—aspirated voicing contrast would be highly unusual for monolingual Hungarian speakers.

Chapter 4: The Hungarian experiment

This section will be about the Hungarian counterpart of the English experiment presented in *Chapter 3*. First, I am going to describe the difference in methods (participant pool and stimuli), and briefly review the procedure of the experiment as a reminder (*Section 4.1*). Then I am going to present the results of the experiment broken down by task: first, reading (*Section 4.3*), then shadowing (*Section 4.4*), and finally, labeling (*Section 4.5*). The data recorded in the likeability rating task at the start of the experiment will be incorporated into these sections as a predictor of the recorded phonetic data.

Each of the result sections will start with a brief overview and the discussion of statistical methods used. Then the data will be analyzed by condition. The order of the conditions will be different from the English experiment. While in the English chapter I started with the *Extreme Aspirating* condition, which was more like typical English productions, in Hungarian each section will look at the *Extreme Prevoicing* condition first, since it is more similar to typical Hungarian productions. After *Extr. Prev.*, each section will then present results from *Extr. Asp.*. Finally, each section will have a short summary of the given tasks' results. The chapter will conclude with a discussion section, where I revisit the issues this experiment was meant to address, and interpret the results in relation to these questions.

4.1 Methods

In this section I am going to describe the methods of the Hungarian experiment. Since it is very similar to the English study, I will first focus on the sources of difference: different participants and stimuli. I will start by describing the model talker and the participants in *Section 4.1.1*, followed by a description of the materials used in the study (*Section 4.1.2*). Finally, the section will end with a short reminder of the procedure of the experiment (*Section 4.1.3*).

4.1.1 The Model talker and the participants

The model talker of the Hungarian experiment was a 29 years old cis-female. She is a phonetically trained native monolingual Hungarian speaker. She has no speech impediments or hearing disorder. She was photographed in a well-lit room for the purpose of the experiment.

Participants were identified by a unique, self-generated 6-digit code both on their paper sheets and in Psychopy. For the purposes of plotting and analysis, these participant codes were translated into a code made up of the letter F or M ('female' or 'male') and a number. In total, 48 participants were recorded for this experiment. 7 of them had to be excluded because of either recording issues, task issues, or the participant's non-compliance with the task. Out of the remaining 41 participants, 21 self-identified as male and 20 as female. No other identities were represented among the participants (see *Table 4.1* for a by-condition by-gender breakdown). The female participants were ages 18–24 (mean: 20.1), and the males were ages 18–25 (mean: 20.9). The experiment was advertised on college campuses, and while all participants were college-age, not all of them were necessarily college students.

	Extr. aspiration	Extr. prevoicing	Total
Female	10	10	**20**
Male	10	11	**21**
Total	**20**	**21**	

Table 4.1: Participants' breakdown by condition and gender

All participants self-identified as monolingual native speakers of Hungarian. In addition, none of them lived abroad, went to a bilingual school, had a foreign language taught to them by a native speaker or had advanced skills in any foreign language (nothing above B2 on the Common European Framework of Reference for Languages). All participants reported "Hungarian" as their ethnicity.

4.1.2 Materials

Rating

During the rating task, participants listened to a short audio recording of the model talker discussing what to consider when shopping for a mattress. The text was a simplified version of an article from a Hungarian lifestyle blog (see Appendix, *Figure A.2*). It was chosen because of its neutral topic and tone, as well as compatibility with both English and Hungarian. The speaker was recorded reading the text 9 times at a medium pace, and the most natural-sounding production was used. The recordings were made in a sound booth with a DPA 4066 omni-directional head-mounted microphone, through an Analogue-to-Digital (AD) cable. The recordings were sampled at at 44.1 kHz.

Labeling

The stimuli for the labeling data were created similarly to the English experiment's labeling stimuli. The stimuli comprised 11 equidistant steps of a VOT continuum, where one extreme was the word *boros* /ˈboroʃ/ 'wine-related' and the other was *poros* /ˈporoʃ/ 'dusty'. The tokens differed in how much VOT the initial stops had, which ranged from prevoicing (-60 ms VOT) to aspirated (90 ms VOT). The steps had -60, -45, -30, -15, 0, 15, 30, 45, 60, 75 or 90 ms VOT. Each stimulus was spliced from parts from actual word recordings of the model talker. The words *boros* and *poros* were included in the word list that the model talker was asked to read out when recording for stimuli for the reading and shadowing tasks.

Each audio file started with a 120 ms silence. For prevoiced tokens, this was followed by either 15 ms, 30 ms, 45 ms, or 60 ms of prevoicing (between the yellow and blue line in *Figure 4.1*). These bits of prevoicing all came from a single production of the model talker (*béllel* /ˈbeːlːɛl/ 'with guts' with 160 ms of naturally occurring prevoicing). This was followed by a burst (from a token of *pofon* /ˈpofon/ 'slap on the face (N)'). This burst (between the blue and green lines in *Figure 4.1*) was the same for all items (prevoiced and aspirated alike). Then this was followed by an "ending" (*-oros* /ˈoroʃ/, right of the green line in *Figure 4.1*) from an instance of *poros* /ˈporoʃ/. In order to control for any cues to voicing in the f_0, this ending was obtained after f_0 was stable, and F1, F2, and F3 were all present. The 0 ms VOT token was made similarly, only without the prevoicing bit.

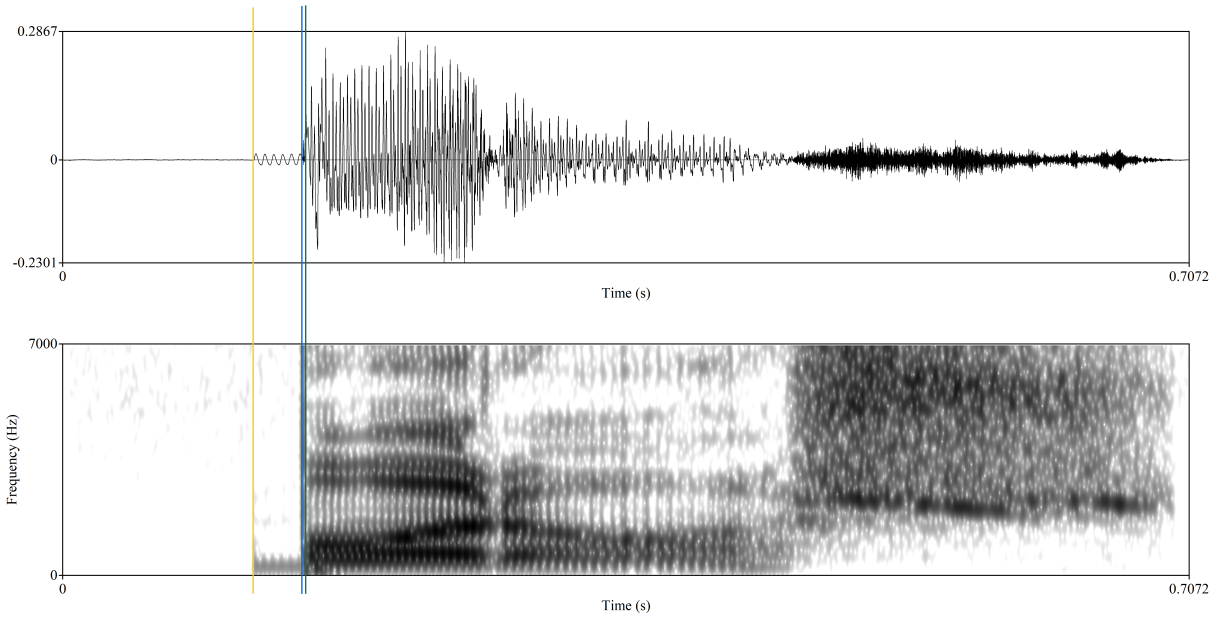

Figure 4.1: Waveform and spectrogram of audio stimulus boros/poros with 30 ms of prevoicing and 0 ms aspiration

The aspirated tokens were made in a similar way—an example of such a token is shown in *Figure 4.2*. Like the prevoiced tokens, they also started with 120 ms silence. They of course did not have a bit of prevoicing, so the silence was followed by the aforementioned almost 0 ms burst from a token of *pofon* /ˈpofon/ ‘slap on the face (N)’—this is between the blue and the red line. The tokens then continued with a piece of aspiration of the appropriate duration (15, 30, 45, 60, 75, or 90 ms—between the red and green line). These were all a contiguous bit, obtained from a token of *poros* /ˈporoʃ/ ‘dusty’. This token was recorded with an instruction to aspirate a lot, and as a result it had 187 ms of VOT in total. Finally, each stimulus ended with the same “ending” (*-oros* /ˈoroʃ/) that was used for the prevoiced tokens (right of the green line).

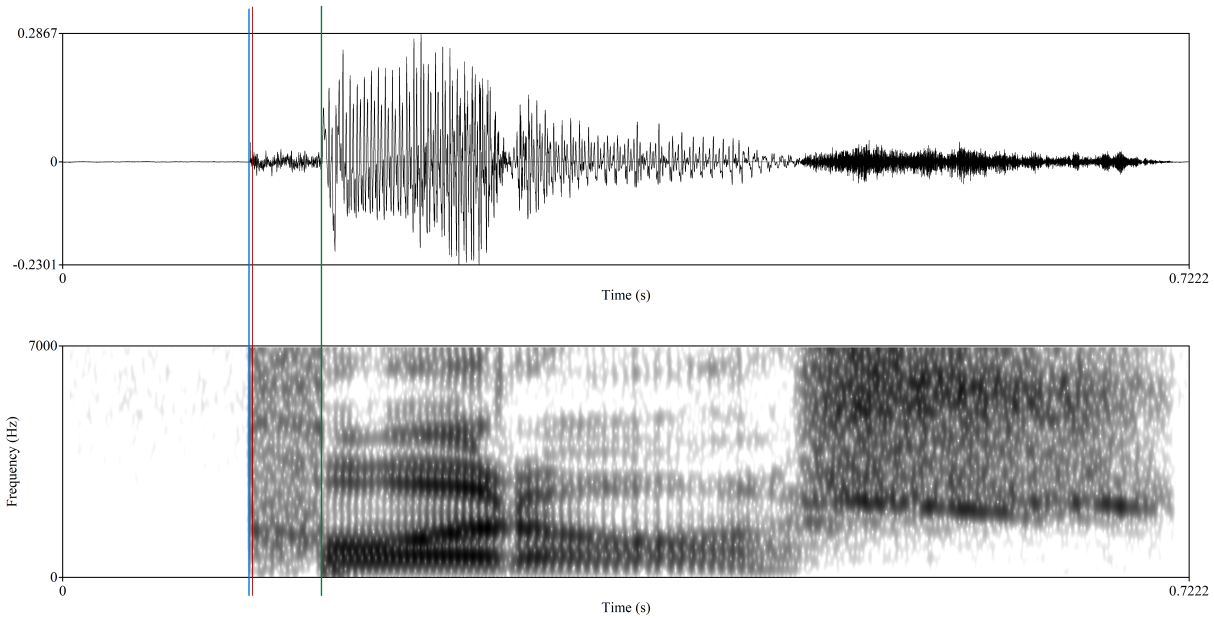

Figure 4.2: Waveform and spectrogram of audio stimulus boros/poros with 75 ms of aspiration and no prevoicing

Reading and shadowing

Just like in the English experiment, the shadowing data involved a subset of words that were used in the reading tasks. In the reading tasks, participants had to read the 40 items twice each, and in the shadowing task they had to repeat a semi-random subset of 30 of them 6 times each. All words followed a SVC(CC)VC(C) structure, where S is a bilabial stop, V is a vowel, and C is a consonant that is not a stop. The full list of the 20 p-initial and 20 b-initial words are in *Tables 4.2 & 4.3*, respectively. The left-hand side of both tables contains monomorphemic words, and the right-hand side contains polymorphemic ones. Frequencies are from the Hungarian Webcorpus (Halácsy et al., 2004). These were the words participants saw in the reading task (and during Familiarization).

Word	IPA	Frequency	Word	IPA	Frequency
páfrány ‘fern’	ˈpaːfraːɲ	291 (0.49)	pálmán ‘on palm tree’	ˈpaːlmaːn	5 (0.01)
panel ‘block (N)’	ˈpɑnɛl	2572 (4.37)	párol ‘steam (V)’	ˈpaːrol	20 (0.03)
parázs ‘ember’	ˈpɑraːʒ	1006 (1.71)	párnán ‘on pillow’	ˈpaːrnaːn	284 (0.48)
parfüm ‘perfume’	ˈpɑrfym	643 (1.09)	páván ‘on peacock’	ˈpaːvaːn	8 (0.01)
peron ‘platform’	ˈpɛron	626 (1.06)	perel ‘sue’	ˈpɛrɛl	224 (0.38)
persely ‘piggy bank’	ˈpɛrʃɛj	225 (0.38)	pofon ‘slap (N)’	ˈpofon	2009 (3.41)
pimasz ‘cheeky’	ˈpimɑs	740 (1.26)	pólón ‘on T-shirt’	ˈpoːloːn	178 (0.30)
pollen ‘pollen’	ˈpolːɛn	509 (0.86)	pózol ‘pose (V)’	ˈpoːzol	93 (0.16)
póráz ‘leash’	ˈpoːraːz	789 (1.34)	puszil ‘kiss (V)’	ˈpusil	32 (0.05)
pulzus ‘pulse (N)’	ˈpulzuʃ	544 (0.92)	pürés ‘mushy’	ˈpyreːʃ	2 (0.01)

Table 4.2: Hungarian p-words, frequency from Hungarian Webcorpus
Corpus size: 589M; Word’s frequency per million in brackets
Left: monomorphemic, right: polymorphemic

Word	IPA	Frequency	Word	IPA	Frequency
bálvány ‘idol’	ˈbaːlvaːɲ	686 (1.16)	bányász ‘miner’	ˈbaːɲaːs	1956 (3.32)
balzsam ‘balm’	ˈbɑlʒɑm	287 (0.49)	bénán ‘lame (Adv.)’	ˈbeːnaːn	249 (0.42)
bámul ‘stare’	ˈbaːmul	1106 (1.88)	béllel ‘with bowel’	ˈbeːlːɛl	13 (0.02)
banán ‘banana’	ˈbɑnaːn	1751 (2.97)	bomlás ‘disintegration’	ˈbomlaːʃ	740 (1.26)
bazár ‘bazaar’	ˈbɑzaːr	530 (0.90)	borsón ‘on pea’	ˈborʃoːn	17 (0.03)
bivaly ‘bison’	ˈbivɑj	687 (1.17)	borzas ‘scruffy’	ˈborzɑʃ	474 (0.80)
bizarr ‘bizarre’	ˈbizɑrː	2679 (4.55)	bosszús ‘angry’	ˈbosːuːʃ	394 (0.67)
bónusz ‘bonus’	ˈboːnus	952 (1.62)	búsul ‘be dismayed’	ˈbuːʃul	148 (0.25)
bölény ‘buffalo’	ˈbøleːɲ	328 (0.56)	bűvész ‘magician’	ˈbyːveːs	566 (0.96)
búvár ‘diver’	ˈbuːvaːr	1717 (2.92)	bűvöl ‘charm (V)’	ˈbyːvøl	72 (0.12)

Table 4.3: Hungarian b-words, word frequency from Hungarian Webcorpus
Corpus size: 589M; Word's frequency per million in brackets
Left: monomorphemic, right: polymorphemic

To create the audio stimuli for the shadowing task, the model talker was recorded reading these words out loud 6 times (along with *boros* /ˈboroʃ/ ‘wine-related’ and *poros* /ˈporoʃ/ ‘dusty’ for the labeling task). The order of words was randomized within each of the 6 lists. For 3 of the recordings she was encouraged to read as naturally as possible, and the other 3 were meant to elicit aspiration. The phonetically trained model talker was aware of this intent. Since long aspiration is not typical on Hungarian voiceless stops, she was recommended to try and imitate an American accent. Recordings were made in a sound booth with a DPA 4066 omni-directional head-mounted microphone, through an AD cable, with a sampling rate of 44.1 kHz.

Two different sets of shadowing stimuli were created for the two conditions of the experiment (*Figure 4.3*). The *Extreme Prevoicing* condition mimicked a prevoicing system but with exaggerated

prevoicing (plain /p/ with 15 ms of VOT and prevoiced /b/ with −130 ms of VOT). As the figure shows, this is very similar to what Hungarian typically looks like. The *Extreme Aspirating* condition is more similar to a typical aspirating contrast, where an aspirated /p/ (130 ms VOT) is contrasted with a plain /b/. The plain /b/ in the Hungarian experiment had to deviate from the original plan of having 15 ms VOT (as much as the plain /p/ had in the *Extr. Prev.* condition), because the model talker hardly produced any natural /b/'s with this much aspiration. Since splicing together very short tokens can sound artificial ("clicky"), the plain /b/'s in *Extr. Asp.* had 5 ms aspiration each.

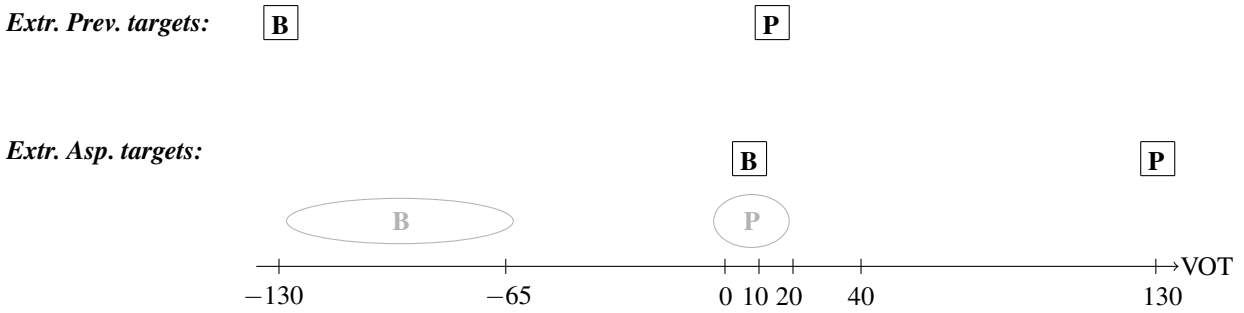

Figure 4.3: Conditions in the Hungarian experiment

The /p/ words were generated in two versions: with 130 ms VOT for *Extr. Asp.* and with 15 ms VOT for *Extr. Prev.* An example of a 130 ms VOT stimulus is shown in *Figure 4.4*. The stimuli were generated from 120 ms silence, followed by a less than 1 ms burst (between the blue and the first red line in the figure), then aspiration (between the blue and green lines), and finally, an ending (right of the green line). In this example for *pollen* /ˈpolːɛn/ 'pollen', the "burst bit" came from a token of *pofon* /ˈpofon/ 'slap in the face (N)'. The same burst was used for groups of words where a /p/ was followed by the same vowel—e.g. *pollen* /ˈpolːɛn/ 'pollen' and *pofon* /ˈpofon/ 'slap in the face (N)' had the same burst bit, but the burst in *puszil* /ˈpusil/ 'kisses (V)' came from a different word (*pulzus* /ˈpulzuʃ/ 'pulse (N)'). The reason for using the same burst for different lexical items was borne out of necessity: the burst from some words did not easily yield itself to splicing, and reusing some of the bursts allowed us to avoid mechanic-sounding stimuli.

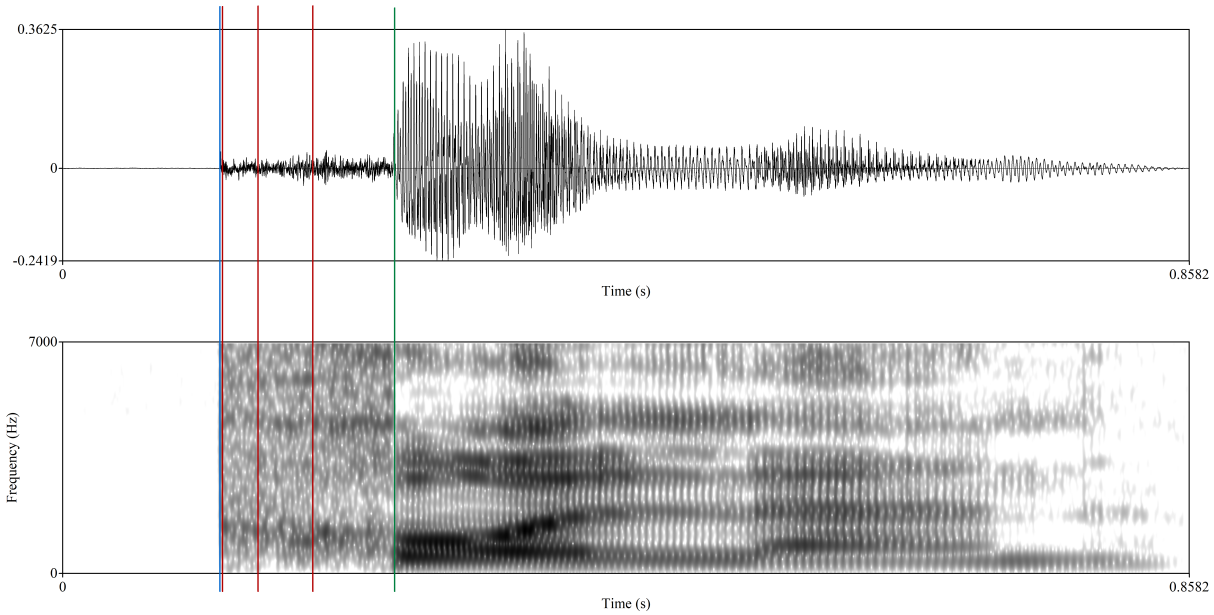

*Figure 4.4: Waveform and spectrogram of Hungarian audio stimulus /*polːɛn*/ with no prevoicing and 130 ms aspiration*

After the burst, 130 ms of aspiration was put together from multiple (2 or 3) pieces. The exact pieces that were used for the 130 ms VOT *pollen* /ˈpolːɛn/ 'pollen' stimulus are divided by red lines in *Figure 4.4*. In order to make the resulting stimulus as natural as possible, the pieces all began and end at a 0-crossing. Moreover, the pieces used for a given word always came from the same token, but just like with the burst, the word whose aspiration was taken might not be the same as the stimulus. For instance, the 130 ms aspiration for *pollen* /ˈpolːɛn/ 'pollen' came from a different word, *póráz* /ˈporaːz/ 'leash', but all the parts that made up the eventual stimulus (*pollen*) came from the same token of *póráz*. Finally, an "ending" (i.e. the rest of the word after the stop) was spliced to the stimuli—right of the green line in *Figure 4.4*. This was obtained after F1, F2 and F3 were all present. The 15 ms plain /p/-initial words for *Extr. Prev.* only differed in that instead of the multiple pieces of aspiration, they had only one, 15 ms long bit of aspiration.

The /b/ words also had two versions: one with 5 ms aspiration for *Extr. Asp.* and one with 130 ms prevoicing (and no aspiration) for *Extr. Prev.* They were all made up of 120 ms silence,

optional prevoicing (for *Extr. Prev.* tokens), a short burst, optional 15 ms aspiration (for *Extr. Asp. tokens*), and then an ending, which was obtained after F1, F2, and F3 were all present. An example is shown below in *Figure **??*** on the example of *balzsam* /ˈbɒlʒɒm/ ‘balm’. The prevoicing was added as a contiguous piece (between the yellow and the blue line), derived from a token of *béllel* /ˈbeːlːɛl/ ‘with guts’, which had 160 ms prevoicing naturally. We can see the burst between the blue and the green line in *Figure **??***, and the ending is the part right of the green line.

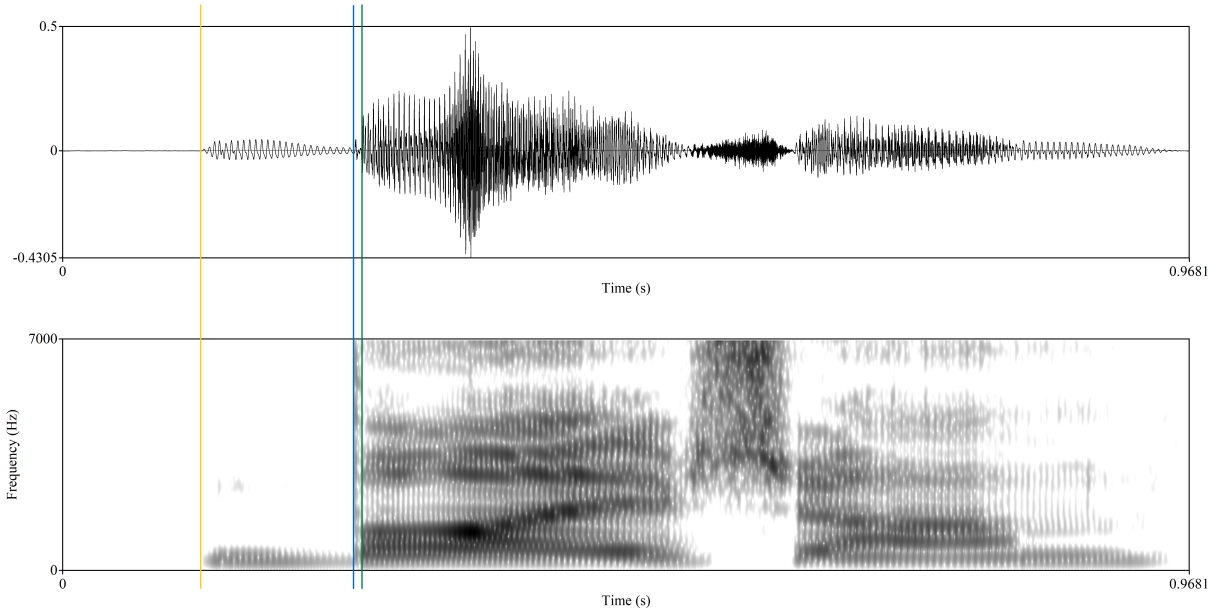

Figure 4.5: nah

4.1.3 Procedure

Participants completed the tasks in a sound-attenuated booth in the same order with Audio-technica ATH-ANC7b headphones. Because of technical issues, 29 participants were recorded with a TASCAM DR-40 PCM recorder and an audio-technica AT8531 lapel microphone clipped on the participant’s clothing, sampled at 44.1 kHz. 9 participants were recorded with the same microphone but from the laptop itself through an external sound card (M-Audio Mobile Pre attached with a USB). 3 participants were recorded with a DPA 4066 omni-directional head-mounted microphone,

through an Analogue-to-Digital (AD) cable. The experiment was presented via Psychopy 3 (Peirce, 2007), and except for the paper-based Rating task and the Sociolinguistic questionnaire, tasks had to be completed through Psychopy with a laptop. The procedure was identical to that of the English experiment, and it is summarized in *Table 4.4*. The experiment took about 40 minutes to complete.

	Instruction	**Stimuli**	**Example**
Rating	Rate speaker for properties	semantic differential scales (10)	shy–talkative
PRE-Labeling	Select the word you hear	word, audio on a VOT continuum (11*10)	*binning* / *pinning* with 45 ms VOT
Familiarization	Read the word silently	written word (40*1)	*basin*
PRE-Read	Read the word out loud	written word (40*2)	*poser*
Shadowing	Repeat the word you hear	word, audio, 1 of 2 conditions (30*6)	*pollen* with 15 ms VOT
POST-Read	Read the word out loud	written word (40*2)	*buzzer*
POST-Labeling	Select the word you hear	word (audio on a VOT continuum (11*10)	*binning* / *pinning* with 75 ms VOT
Questionnaire	Fill in the questionnaire	—	Age:

Table 4.4: Procedure of the experiment

Participants first completed a *Rating* task, where they listened to the model talker reading out a text on mattress shopping (full text in the *Appendix*, *Figure A.2*) and rated her on 9 semantic differential scales for certain features as well as on a 1-to-9 Likert-scale on how attractive they found her personally. This was completed on paper. A picture of the model talker was displayed during the reading passage. The full list of features grouped by the eventual measures they were converted into (*Superiority*, *Solidarity*, and *Dynamism*) are in *Table 4.5*. Participants were told to pay attention to the fact that some of the scales had the "positive" value on the left-hand side while for other scales it's on the right.

Solidarity	**Superiority**	**Dynamism**
friendly — unfriendly	organized — disorganized	shy — talkative
dishonest — honest	lower status — upper status	unsure — confident
rude — polite	intelligent — unintelligent	energetic — lazy

Table 4.5: Semantic differential scales

After that, participants completed the *PRE-Labeling* task, where they had to rate 11 steps on a word-based VOT continuum from −60 ms to 90 ms VOT. After hearing a stimulus, they had to indicate whether they heard the word *boros* /ˈboroʃ/ 'wine-related' or *poros* /ˈporoʃ/ 'dusty' by pressing either 'F' or 'J' on the keyboard, respectively. This was repeated in 10 blocks, yielding 110 observations/participant, which formed the pre-exposure baseline for the labeling data.

Then participants were exposed to the 40 words from *Tables 4.2–4.3* one-by-one in written form. In the first iteration they were asked to only read them silently (*Familiarization*). Then they saw each item in two more blocks, and had to read them out loud (*PRE-Read*). These 80 recorded tokens/participant formed the pre-exposure baseline for the Reading data.

Then followed 6 blocks of *Shadowing* task, where participants were exposed to a semi-random 30-word subset of the 40 words (balanced for initial segment and morphemic complexity),

which they had to "identify by saying it out loud". While the audio was playing, the screen showed a picture of the model talker. Participants were made aware of the fact that all of these words will be words that they already saw in the reading task. A random half of the participants completed the experiment with the *Extreme Aspirating* stimuli (aspirated /p/, plain /b/), and the other half with the *Extreme Prevoicing* stimuli (plain /p/, prevoiced /b/).

Afterwards, participants completed the reading task again (*POST-Read*) in two blocks, and 10 more blocks of the labeling task (*POST-Labeling*). These data were compared with the respective pre-exposure baseline values.

Finally, they completed a sociolinguistic questionnaire (on paper), which asked about their age, gender identity, place of birth, native language, and their competence and exposure to other languages, all in a free-form answer. Participants were asked to provide another assessment of their attraction towards the model talker on a 1-to-9 Likert scale. Participants were also able to say why they gave that rating and they could also leave any general comments about the study.

4.2 Data processing and analysis

The labeling dataset consists of 9,020 responses (*boros* or *poros* responses), which was automatically extracted from the output of Psychopy. Words were automatically segmented for the reading and shadowing data, then each word was manually adjusted and annotated by the author. The reading data comprises 3,280 /p/ and 3,280 /b/. I had to exclude 4 /p/ and 3 /b/ tokens in *Extr. Asp.*, and 4 /p/ and 5 /b/ tokens in *Extr. Prev.* These tokens were excluded because the participant skipped them or because they could not be segmented because of creakiness, yawning or noisy recording quality. This resulted in the final dataset: the *Extr. Asp.* had 1,596 /p/'s and 1,597 /b/'s, and the *Extr. Prev.* data had 1,676 /p/'s and 1,675 /b/'s.

In the Hungarian shadowing task, many fewer tokens were excluded than in the English shadowing task. In the English experiment, when participants in *Extr. Prev.* had to shadow a plain

/p/, they often produced the word with an English [b], which always resulted in nonce-words. For instance, when they heard the model talker say *pooling*, they said #[ˈbulŋ]. Such tokens were often excluded from the analysis. I argued that rather than just an instance of shadowing, these tokens must be a realization of a different category (/b/), because they were paired with a questioning or incredulous intonation pattern, which reflects that the participants themselves thought they were hearing and pronouncing a nonce-word, i.e. a /b/-initial word. Since they aimed for a /b/, these productions do not say anything about their representations of /p/, and must be excluded from the /p/ dataset.

While the Hungarian participants produced some uncharacteristic values, there was no evidence that these tokens were not from the shadowed category—i.e. there is no evidence that plain (non-prevoiced) /b/'s in *Extr. Asp.* were indeed shadowed as a completely different segment, a /p/. We have two reasons for this. First, none of their productions could be excluded based on intonation. Second, as we will see in *Section 4.3.1*, while realizations were mostly as expected from the literature, there were a surprising number of atypical tokens in the PRE-Read task. In total, 3,690 /p/ and 3,690 /b/ productions were recorded. 8 /b/'s in *Extr. Asp.* and 4 /p/'s in *Extr. Prev.* had to be excluded. This resulted in an *Extr. Asp.* dataset of 1,800 /p/'s and 1,792 /b/'s and an *Extr. Prev.* dataset of 1,886 /p/'s and 1,890 /b/'s.

9 ratings on the model talker's likeability were collected from each participant. On a by-participants level these were averaged into three measures. The average of ratings along the *friendly — unfriendly*, *dishonest — honest*, and *rude — polite* is called ***Solidarity***. *Superiority* is the average of the *organized — disorganized*, *lower status — higher status*, and *intelligent — unintelligent* ratings. The ratings of the *shy — talkative*, *unsure — confident*, and *energetic — lazy* spectra were averaged as the ***Dynamism*** measure. The attraction data will be excluded from the Hungarian dataset as well because of the ambiguous interpretations of the answer.

4.3 Reading results

In this section I am going to present results from the Hungarian reading tasks (*PRE-Read* and *POST-Read*). First, I will provide an overview section with some basic information about the participants' baselines and the statistical methods used (*Section 4.3.1*). Then I will go on to discuss the results of the *Extreme Prevoicing* condition (*Section 4.3.2*), where stimuli were most similar to the typical way the /p b/ contrast manifests in Hungarian. Then I will do the same for the *Extreme Aspirating* dataset, which required more of an adjustment from Hungarian native speakers (*Section 4.3.3*).

4.3.1 Overview and statistical methods

In the following, I will share some basic details about the *PRE-Read* data, in order to give the reader an idea of what the baseline reading productions in Hungarian looked like to begin with. Then I will discuss the statistical methods and models and also highlight some key points of interest in it by recapping the questions we will focus on.

The experiment found that Hungarian speakers elongated the prevoicing in their read /b/'s after being exposed to extremely prevoiced /b/ tokens in a shadowing task (/b/'s with 130 ms prevoicing). They similarly converged with the plain /p/'s—i.e. they lessened the amount of aspiration in their /p/'s. These can be seen on the right-hand side of *Figure 4.6*. When a different set of participants were exposed to a plain /b/ (5 ms VOT) and an aspirated /p/ (130 ms VOT) in the *Extreme Aspirating* condition, no convergence was found (left-hand side of *Figure 4.6*).

As can be seen, Hungarian /b/ productions were bimodal: most tokens were prevoiced, but there were a few plain [b] tokens as well. Accommodation behavior most often entailed the participant changing the ratio of the two, but sometimes we saw adjustments in terms of the ranges of VOT values—e.g. sometimes participants changed how much prevoicing there was on their prevoiced tokens.

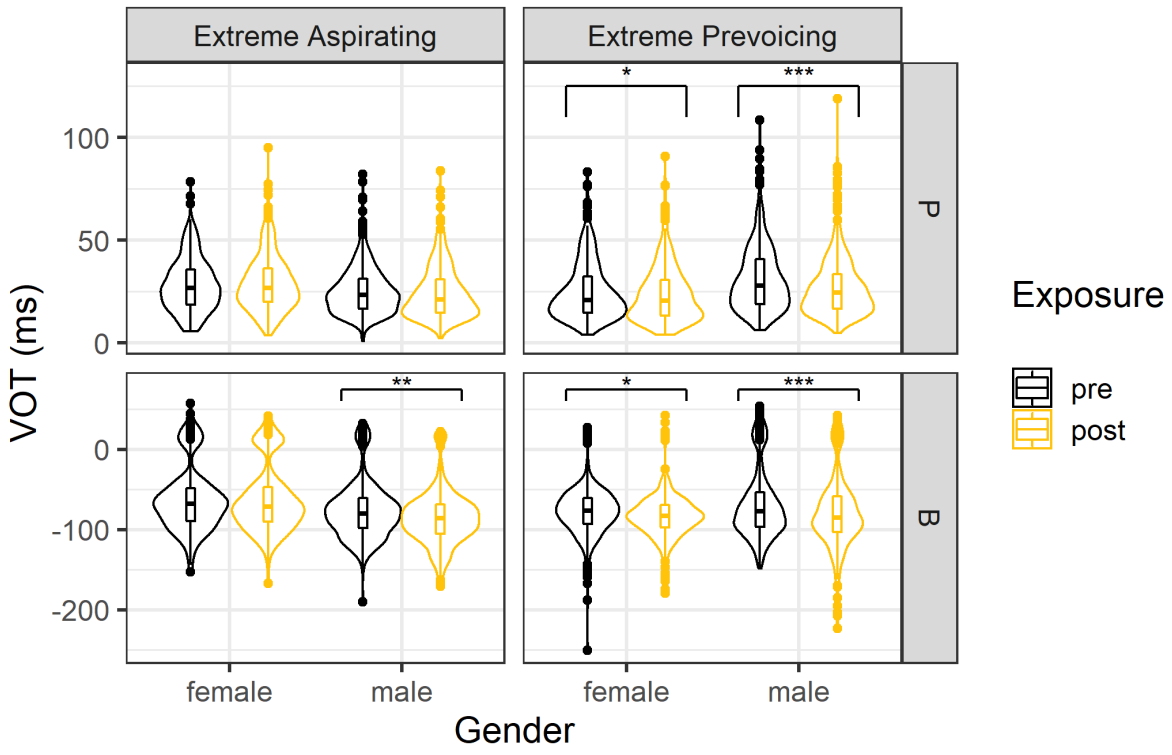

Figure 4.6: Effect of `Exposure` *in the reading data from Hungarian-speaking participants Note the different VOT axes for /p/ and /b/*

In terms of likeability, there were two notable effects. First, there seemed to be a correlation between convergence for /b/'s in *Extr. Prev.* and `Superiority` ratings, but only for males. Second, in *Extr. Asp.*, we saw that `Solidarity` correlated with accommodation behavior for /b/'s, but it mostly mediated the amount of divergence, and resulted in few instances of convergence. Participants who rated the model talker low tended to diverge more than those rating her higher, but some females who rated the model talker high even converged with her. Thus, this effect finds a link between disliking and divergence rather than liking and convergence.

Descriptive statistics

The *Extreme Aspirating* dataset was made up of 1,596 /p/ tokens and 1,597 /b/ tokens, while the *Extreme Prevoicing* dataset consisted of 1,676 /p/'s and 1,675 /b/'s. Before diving into the analysis of the full dataset, in this section I will describe the half of this that was collected before any exposure to the shadowing stimuli happened. The purpose of this is both to give a general idea about what the pre-exposure baselines of our participants looked like and to see if participants in the two conditions differ in any accidental but significant ways.

The /p/ productions had on average 27.663 ms VOT (median: 24.821 ms, standard deviation: 14.356 ms). *Figure 4.7* shows individual means by gender and condition, and *Figure 4.8* shows the distribution of the data pooled by gender and condition. The mean and median are perfectly in line with previous findings (e.g. mean of 24.64 ms for word-initial /p/ in Gósy, 2001). Unsurprisingly, this is much shorter than the baseline VOT was in the English dataset, but it must also be noted that the standard deviation is barely smaller (cf. sd=19.932 ms for English), which means that the Hungarian productions were not much more concentrated. This is somewhat surprising after the ranges found in the Hungarian literature were much narrower than those found for English (e.g. 13.2–34.8 ms in Gósy, 2001 for Hungarian vs. 46–139 ms in Chodroff and Wilson, 2017 for English).

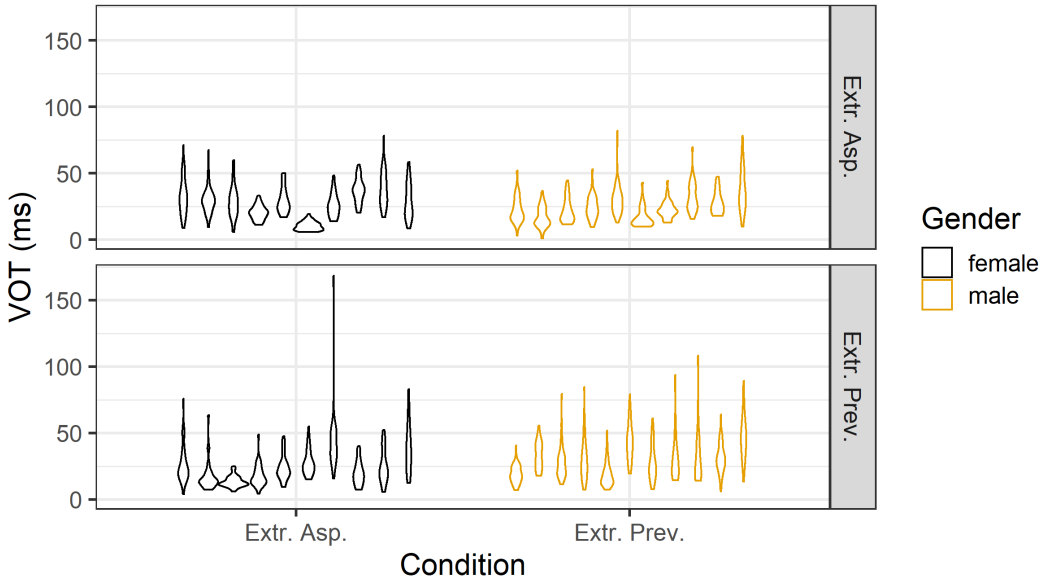

Figure 4.7: Individual pre-exposure means in the Hungarian /p/ reading data

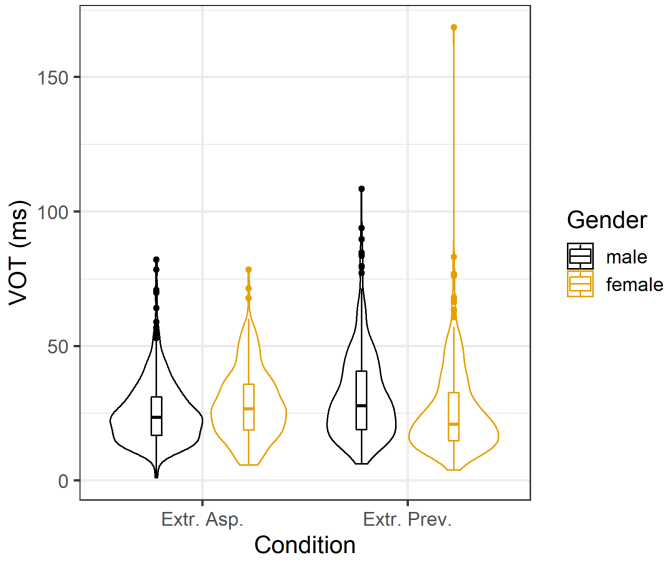

Figure 4.8: Pre-exposure /p/ reading data from Hungarian participants

While in English we found that males had shorter /p/ VOT's on average than females, Hungarian males and females do not differ significantly in terms of their mean values *Figures 4.7 & 4.8*. This is reinforced by a statistical model where neither `Gender`, nor `Condition`, nor the interaction of the two were significant (p=0.4360, p=0.4400, and p=0.1010, respectively; *Table 4.6*).

	Estimate	Std. Error	Pr(> |t|)	
(Intercept)	28.302	3.021	<0.0001	***
Gender [male]	−2.886	3.664	0.4360	
Condition [Extr. Prev.]	−2.860	3.664	0.4400	
Gender [male] × Condition [Extr. Prev.]	8.604	5.123	0.1010	

Table 4.6: LMER model of Hungarian reading /p/ tokens' VOT (in ms) before exposure; Threshold for significance (adjusted with Bonferroni correction): p<0.0125

As for /b/ tokens, the mean VOT was −70.636 ms, the median value was −75.283 ms median with a standard deviation of 38.611 ms. As can be seen in *Figure 4.9* individually and in *Figure 4.10* on a group-level, these means and medians came from a bimodal distribution—the category /b/ comprises a mixture of plain and prevoiced tokens. This was true for English as well, but while the English distribution heavily skewed towards plain tokens, Hungarian /b/'s are more often prevoiced than plain. 1,477 out of the total 1,633 tokens were prevoiced (cf. 391 / 1,640 prevoiced /b/'s in English). On average, these /b/'s had −80.203 ms VOT, (−78,936 ms median, 26.035 ms standard deviation). If anything, these are slightly less prevoiced than the 391 prevoiced /b/ tokens in the English PRE-Read task (−89.980 ms mean, −91.746 ms median, 34.551 ms standard deviation). This supports the observation first noted by Lisker and Abramson (1967) that prevoiced realizations of voiced stops in English are very similar to prevoiced stops in prevoicing languages (i.e. languages where prevoicing is a contrastive cue). The standard deviation in Hungarian was slightly smaller (26.035 ms) than it was in English (34.551 ms), which indicates that prevoiced tokens in Hungarian had a slightly more concentrated distribution. Only 156 out of 1,633 tokens were realized plain in the Hungarian PRE-Read task (cf. 1,249 / 1,640 in English). Interestingly, the Hungarian plain /b/'s VOT was slightly longer than that of English plain /b/'s (Hungarian 19.941 ms mean, 19.022 ms median, 10.626 ms SD; English: 13.993 ms mean, 12.735 ms median, 5.751 ms SD). This means that there was a slight overlap between Hungarian /p/'s and /b/'s VOT values to begin with.

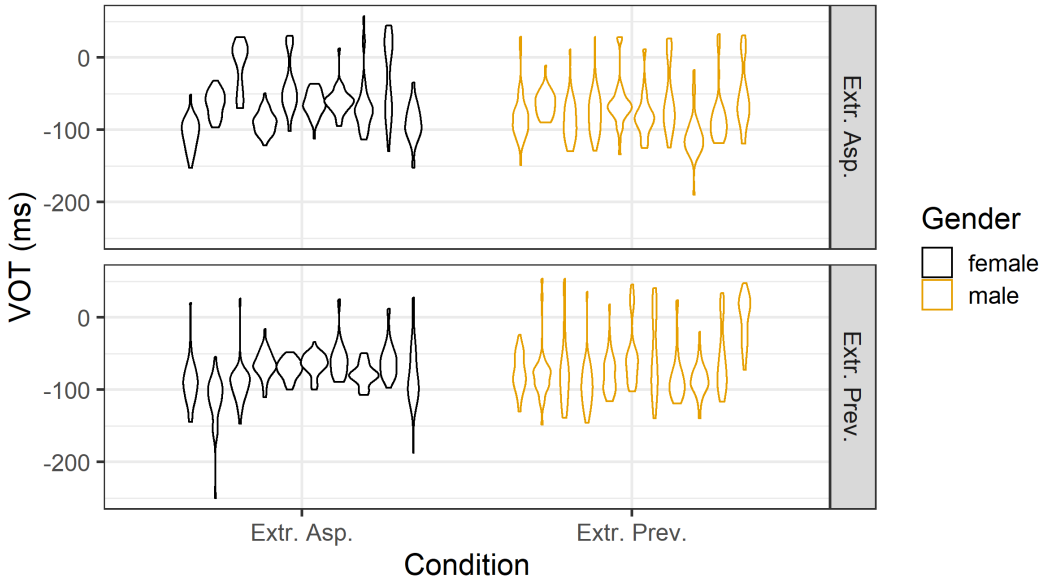

Figure 4.9: Individual pre-exposure means in the Hungarian /b/ reading data

Statistically speaking VOT was not affected by either `Gender` or `Condition`, or their interaction (p=0.1930, p=0.1360, p=0.1090, respectively; *Table 4.7*). The lack of significance of the `Condition` predictor (and its gendered interaction) are reassuring as they indicate that the random split of participants into the two conditions was indeed random. It must also be noted that where in the English data, statistical models often struggled to find an intercept for /b/'s (because of the bimodal distribution of the data), this model on Hungarian pre-exposure /b/'s has no problem finding one. This indicates that the Hungarian dataset is less bimodal, which also shows from the split between prevoiced and plain stops. While Hungarian stops skew towards prevoiced realizations and English stops skew towards plain ones, the *extent* to which Hungarian /b/'s are skewed in their respective direction is even more (1,477 vs. 156) than how skewed English /b/'s were (391 vs. 1,249).

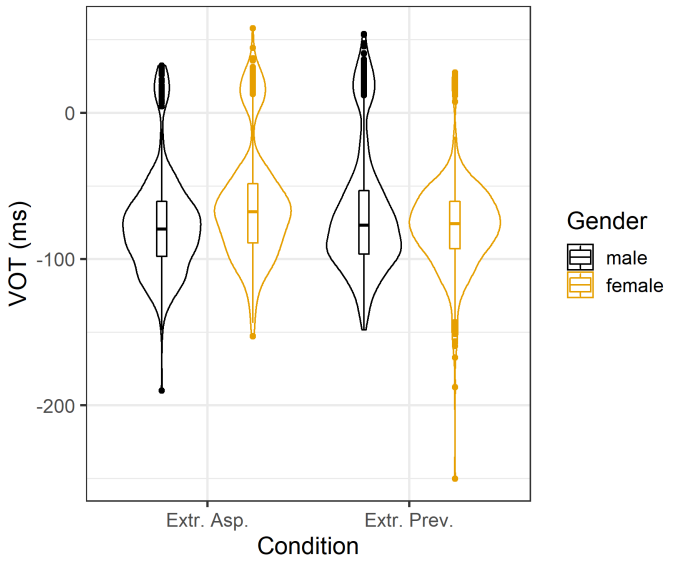

Figure 4.10: Pre-exposure /b/ reading data from Hungarian participants

	Estimate	Std. Error	Pr(> |t|)	
`(Intercept)`	−61.772	7.417	<0.0001	***
`Gender [male]`	−13.581	10.242	0.1930	
`Condition [Extr. Prev.]`	−15.608	10.241	0.1360	
`Gender [male]` × `Condition [Extr. Prev.]`	23.517	14.319	0.1090	

Table 4.7: LMER model of Hungarian reading /b/ tokens' VOT (in ms) before exposure; Threshold for significance (adjusted with Bonferroni correction): p<0.0125

Statistical analysis

Data was analyzed in four separate linear mixed-effect regression models by condition and segment: there is a model for *Extr. Prev.* /p/'s, *Extr. Prev.* /b/'s, *Extr. Asp.* /p/'s, and *Extr. Asp.* /b/'s. This was done in R (R Core Team, 2013) using `lme4`. The dependent variable was VOT duration with the independent variables of `Gender` (*male* or *female*) and `Exposure` (*pre-* or *post*-exposure) and the interaction of these two was also tested. Not all words from the Reading task were also included in Shadowing, but whether a word was included or not did not improve the model's fit, and was omitted

from the analysis. This was tested with χ^2 tests, which showed that there was no improvement in either the model for /p/'s (p=0.5148) or the model for /b/'s (p=0.6471). By-participant and by-word random intercepts were also added to the models.

Because of convergence issues, likeability measures were tested in three separate models each in each of the 4 data subsets. In these models, aside from the independent variables discussed before, one of the three likeability measures was also added. This resulted in one model with `Solidarity`, one with `Superiority`, and one with `Dynamism` for *Extr. Prev.* /p/'s, *Extr. Prev.* /b/'s, *Extr. Asp.* /p/'s, and *Extr. Asp.* /b/'s each. These models will be discussed with the appropriate subset of data.

Foci of attention

Just like the English experiment, the Hungarian dataset also allows for investigating a number of questions. The main focus will be to test the **Maintain categories** and the **Maintain contrasts** hypotheses. The first mandates an adherence to exact phonetic detail of each category, whereas the latter requires the maintenance of contrasts, but is more flexible in terms of where along the typical phonetic dimensions the categories are themselves. The English dataset found no evidence of such flexibility, which can be interpreted as either weak evidence for **Maintain categories** (adherence to phonetic detail) or a sign of prevoicing being a sub-optimal cue for native English speakers. In the Hungarian dataset we will have to inspect how plain /b/'s in *Extr. Asp.* are shadowed, as in order to converge with a plain /b/, participants have to be flexible enough to shift both of their categories upwards on the VOT continuum, and potentially readjust their categorization of certain types of tokens. If we see convergence with plain /b/'s in *Extr. Asp.*, that is evidence for **Maintain contrasts**—and means that the English results were stemming from prevoicing being a hard or less useful cue for English speakers. If, on the other hand we see no convergence in this case,

we can infer that an adherence to phonetic detail plays a role in contrast maintenance (**Maintain categories**).

There are further linguistic and methodological issues that this dataset can address. First, while there was no evidence of members of a voicing contrast moving together in the English dataset, in this experiment we could test this question again, for another (type of) language. This could be especially interesting in the Hungarian *Extr. Asp.* case. If we see people shift their /b/ productions' VOT upward (towards less prevoicing), do these same speakers also shift their /p/ upwards (towards more aspiration), thereby maintaining the same size of gap between their two categories? Second, the Reading dataset can capture a very specific kind of accommodation: one that outlives the exposure. Therefore, any convergence we see in this task must be something that remains around even after the model talker is no longer "present".

Lastly, this dataset can also tell us something about the role of extra-linguistic factors in accommodation, namely `Gender` and facets of Likeability (`Solidarity`, `Superiority`, and `Dynamism`). Since extra-linguistic factors may be interpreted and are present in different ways, cross-culturally, our Hungarian population might show different effects than the English-speaking one did.

4.3.2 The Extreme Prevoicing condition

In the *Extreme Prevoicing* condition, the stimuli that participants were exposed to during shadowing was similar to how the /p b/ contrast is expressed in their native language (Hungarian). Tokens of /p/ were plain (15 ms VOT), and tokens of /b/ were prevoiced (-130 ms VOT).

Accommodation of /p/'s

Exposure to a plain /p/ resulted in a trend of shorter VOT's in the Reading task compared to the pre-exposure read baseline (*Table 4.8*). This trend is almost significant after the threshold for

significance is adjusted with Bonferroni correction (β=−1.762, p=0.0170 > α=0.0125). There is also a trending interaction suggesting that this tendency was even stronger among males (β=−1.748, p=0.0867). The dataset is illustrated in *Figure 4.11*.

	Estimate	Std. Error	Pr(> |t|)	
(Intercept)	25.443	3.246	<0.0001	***
Gender [male]	5.717	3.875	0.1560	
Exposure [post]	−1.762	0.737	0.0170	.
Gender [male] × Exposure [post]	−1.748	1.020	0.0867	

Table 4.8: LMER model of reading /p/ tokens' VOT (in ms) in the Extr. Prev. condition; Threshold for significance (adjusted with Bonferroni correction): p<0.0125

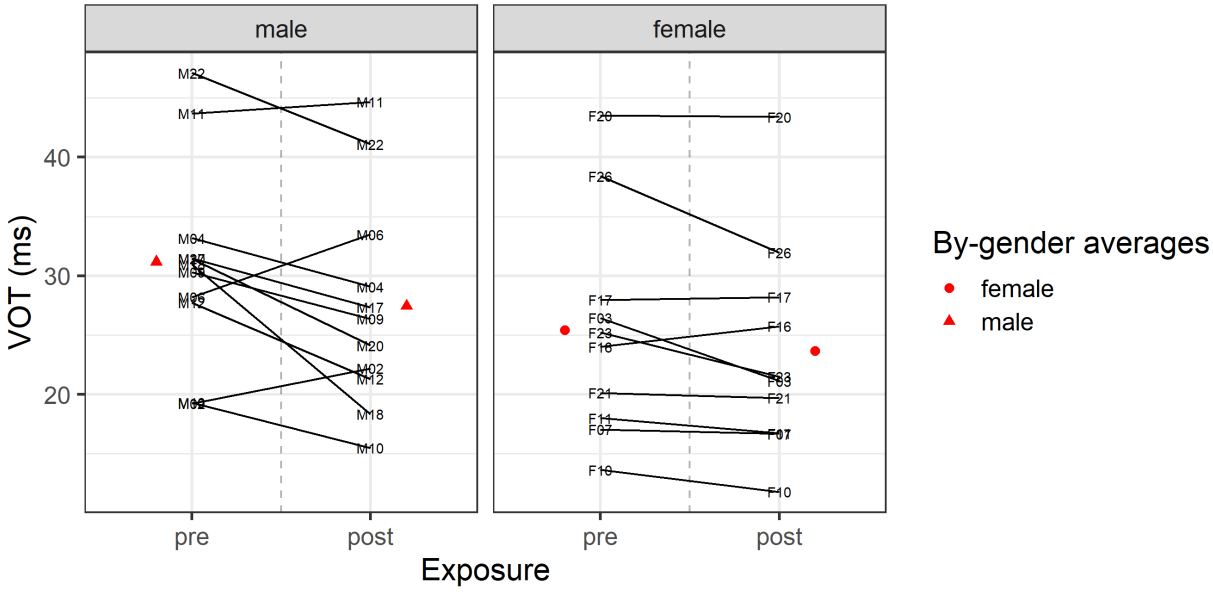

Figure 4.11: Change in mean VOT of Hungarian /p/'s in Extr. Prev. with by-gender averages

Because of these trends, the dataset was split into two (males and females), and the data was modeled with two LMER models where VOT duration (in ms) was predicted by `Exposure` (pre- vs. post) with a by-participant and by-word random intercept. In both of these models, there was a significant effect of exposure (β=−1.7616, p=0.0118 for females; β=−3.508, p <0.0001 for

males). As can be seen from both these effect sizes and the individualized data in (*Figure 4.11*), males indeed reduced their VOT by more than females did. This could potentially be related to the fact that males' means tended to be higher to begin with: 7/10 females' mean VOT was between 15 ms and 30 ms, whereas only 4/11 males'). Perhaps, because of this, some of the females with a pre-exposure mean in this range might have felt that they were "close" enough to the target. This could be similar to how convergence in the English *Extr. Asp.* condition for /b/ was masked by participants whose baselines matched the 15 ms VOT plain /b/ target to begin with.

Likeability data were also recorded in the form of a `Superiority`, a `Solidarity` and a `Dynamism` rating about the model talker from each participant. These were tested as separate predictors, but there does not seem to be a correlation between accommodation and any of these three measures. The results from all three models can be found in the Appendix (*Tables A.42–A.44*). `Solidarity` and `Dynamism` do not improve the model's fit, and the effect of `Superiority` that we do see statistically seems to be overfitting for a few extreme participants (*Figures A.28–A.30* in the Appendix, respectively).

Accommodation of /b/'s

Just like for /p/'s we also see a weak trend of convergence to prevoiced /b/'s in the Hungarian dataset (`Exposure`: β=$-$4.442, p=0.0449) and a trend indicating that it might be gendered (`Exposure` $\times$ `Gender`: β=$-$5.447, p=0.0754). When testing males and females separately, we find convergence for both groups, but especially for males (β=$-$4.442, p=0.0142, α=0.025 for females; β=$-$9.882, p$<$0.0001 for males).

	Estimate	Std. Error	Pr(> |t|)	
(Intercept)	−77.401	7.279	<0.0001	***
Gender [male]	9.963	9.851	0.3239	
Exposure [post]	−4.442	2.213	0.0449	.
Gender [male] × Exposure [post]	−5.447	3.062	0.0754	

Table 4.9: LMER model of reading /b/ tokens' VOT (in ms) in the Extr. Prev. condition; Threshold for significance (adjusted with Bonferroni correction): p<0.0125

We can see (*Figure 4.12*) that these effects belie the variety of behaviors participants attested. Unlike for /p/'s there is no immediately apparent group of participants who could be responsible for the male-female difference. This is certainly less convergence than we saw from English speakers when they were exposed to a native-like but extreme cue (*Extr. Asp.* /p/'s with 130 ms VOT). One reason for this could be that as part of a wide range, Hungarian participants often already produced a few tokens in the vicinity of the 130 ms prevoiced target even pre-exposure, while in English no /p/ was produced with a VOT of 130 ms or more pre-exposure, and the range of VOT for pre-exposure /p/ productions was narrower. This could have resulted in less convergence, similar to Nielsen's (2008, 2011) observation on how females, whose VOT was closer to the 100 ms target VOT she used, converged less than males did, whose VOT was further away on average. Another reason could be that prevoicing as a cue does not lend itself to fine-grained adjustments as much as aspiration does. A lack of fine-grained control over prevoicing might also be supported by the fact that there are even participants whose *mean* VOT is below the target (namely, M09 and F07).

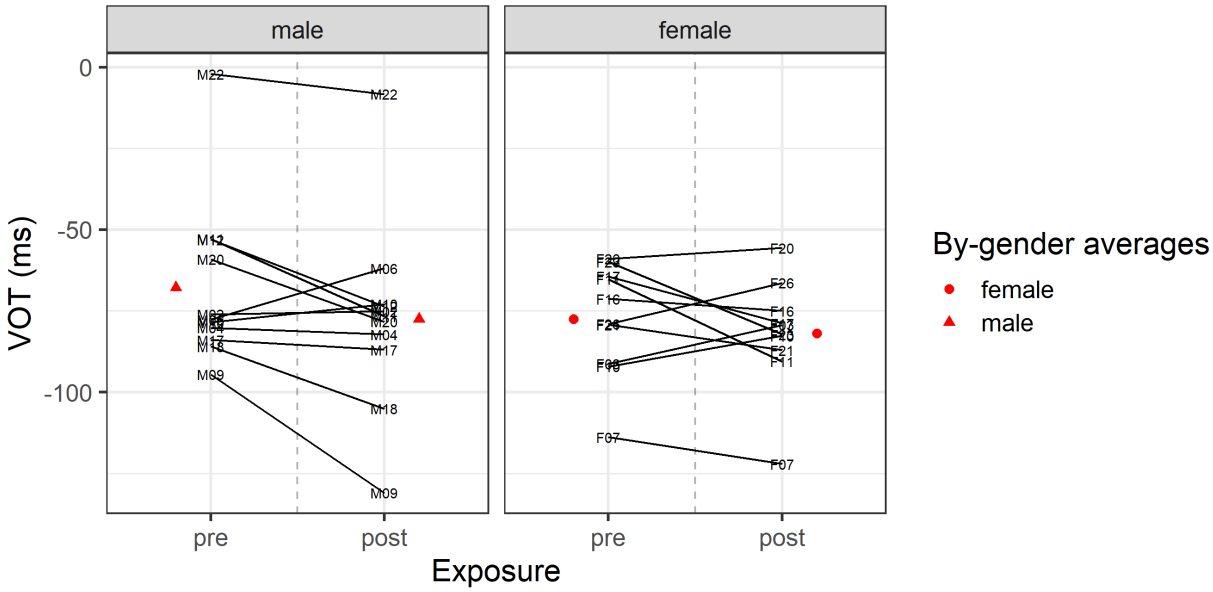

Figure 4.12: Change in mean VOT of Hungarian /b/'s in Extr. Prev. with by-gender averages

Out of the three likeability measures, neither `Solidarity` nor `Dynamism` can be credibly linked to accommodation (see *Tables A.46 & A.47* and *Figures A.31 & A.32* in the Appendix). `Superiority` shows more promise (see *Table A.45* in the Appendix). `Superiority` explains the /b/ data by establishing a correlation between males' VOT and the rating they gave to the model talker (but does not find any effects for females): the higher they rated her, the more likely they were to converge with her very prevoiced /b/'s (`Gender` $\times$ `Exposure:` β=58.602, p=0.0008; `Gender` $\times$ `Exposure` $\times$ `Superiority:` β=$-$9.300, p=0.0004).

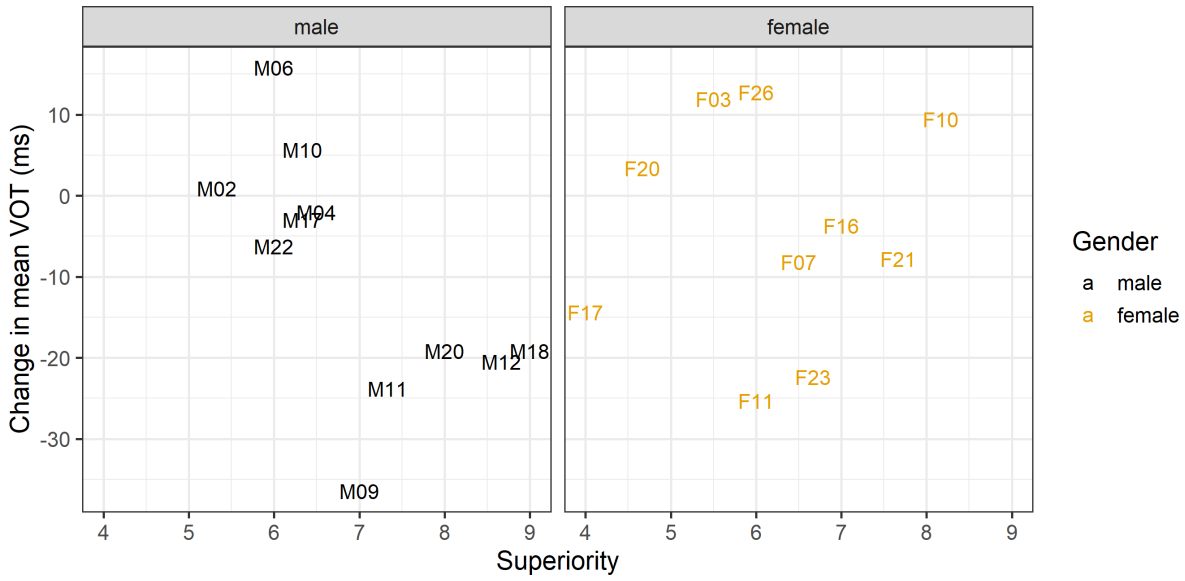

Figure 4.13: Change in mean /b/ VOT in Extr. Prev. in the Hungarian reading task by gender and Superiority rating

While this effect benefits from extreme productions like those of M06, M10 and M09, there also seems to be more of a binary split, where male participants form two clusters based on their /b/'s and `Superiority` ratings (cf. participants giving ratings 7 and up vs. below 7; *Figure 4.13*). This correlation does not arise as a result of a confound with baseline VOT values. *Figure 4.14* shows no relationship that would indicate such a confound—e.g. the participants who diverged or did not converge were not systematically closer to the target −130 ms VOT. This indicates that the statistically determined correlation is not a result of over-fitting to some idiosyncracies of the dataset.

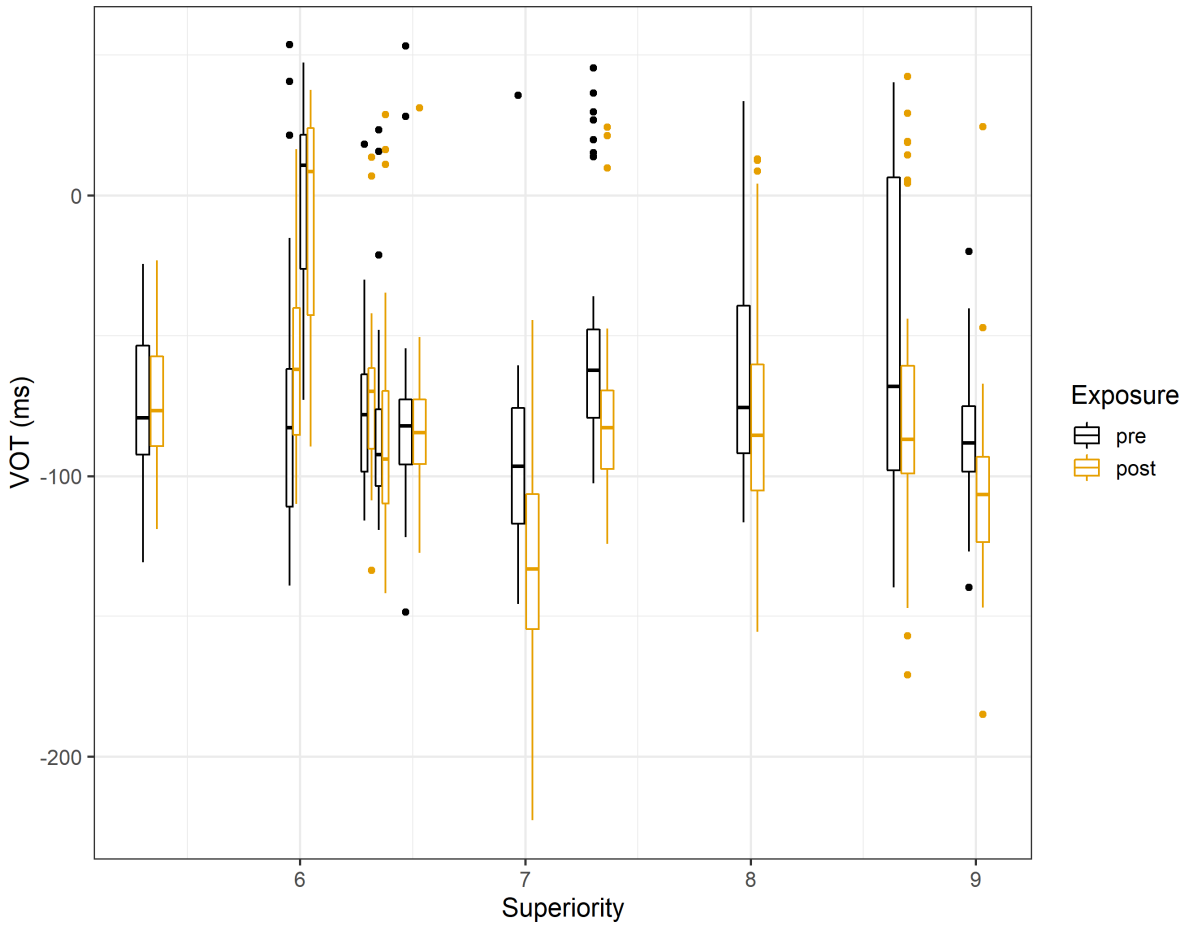

Figure 4.14: Change in /b/ VOT in Extr. Prev. in the Hungarian reading task by gender and Superiority rating

Patterns in the treatment of the /p b/ contrast in the Extreme Prevoicing reading data

Now that we have discussed the treatment of /p/ and the treatment of /b/ separately, it is time to turn our attention to how the contrast itself was treated by individuals. Tokens of /b/ were most often prevoiced in this dataset, which also means that the /p/ and /b/ categories were largely distinct and overlaps were marginal. At the same time, the production of a handful of plain /b/ tokens was fairly common (e.g. *Figure 4.15*), even if plain /b/'s were always outweighed by prevoiced /b/ tokens. These prevoiced tokens tended to have longer stretches of prevoicing in them, which meant that the distribution of /b/'s was bimodal.

Figure 4.15: Reading patterns of M17 in the Hungarian dataset

In terms of *changes* to the contrast as such, we saw that participants converged to both a plain /p/ (15 ms VOT) and a heavily prevoiced /b/ (−130 VOT). *Figure 4.16* shows that these behaviors were not entirely independent from one-another (males on the left, females on the right). In this figure, the y-axis represents how much the participant's mean VOT of /p/ changed from pre- to post-exposure. Since most participants are below 0 on this axis, we can conclude that most

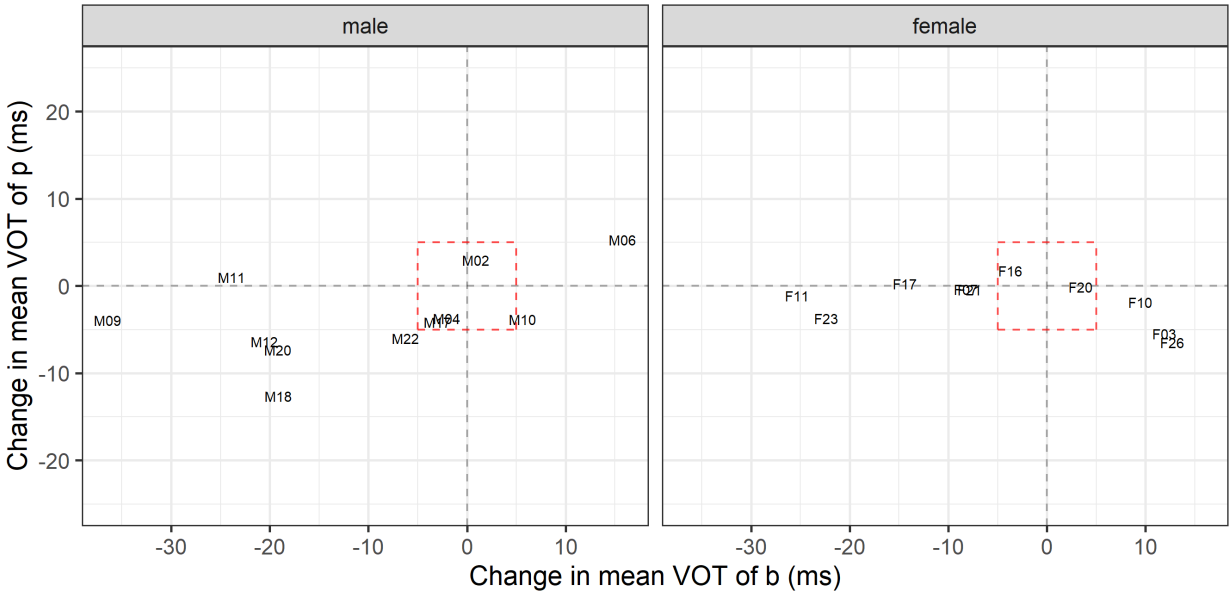

Figure 4.16: Change in mean VOT of /p/ and /b/ in Extr. Prev. per person in the Hungarian data; The red rectangle shows 5 ms change of means in either direction for reference

participants *shortened* their /p/'s VOT after being exposed to 15 ms VOT stimuli, which was in fact somewhat below the pre-exposure average. The x-axis meanwhile represents the change in mean VOT of /b/'s for every participant. Most participants are left of the gray dashed vertical line indicating 0, which means that most participants decreased the VOT of their /b/'s as well.

Two things are apparent from this plot. First, changes in /b/ were both more frequent and bigger than changes in /p/ productions. Some changes along the x-axis are bigger than 20 ms, while there is only one person (M18), whose /p/ VOT changes by more than even 10 ms. Second, behavior for the two categories seems to be correlated, especially for males—i.e. those who reduced their /b/'s VOT (i.e. converged), tended to do the same with their /p/'s VOT. This can be seen from how the most populated quadrant is the bottom left one (people who converged for both /p/ and /b/). However, this relationship was not significant (Pearson's r, p=0.2844), which could be because this tendency had its exceptions. Most notably, F10, F03, and F26 converged with the model talker's plain /p/ (aspirated less, i.e. changed their VOT in a negative direction), but diverged from her extremely prevoiced /b/ (prevoiced less, i.e. adjusted their VOT in a positive direction). Such a pattern could be explained by hypoarticulation or a loss of former hyperarticulation in the POST-Read task.

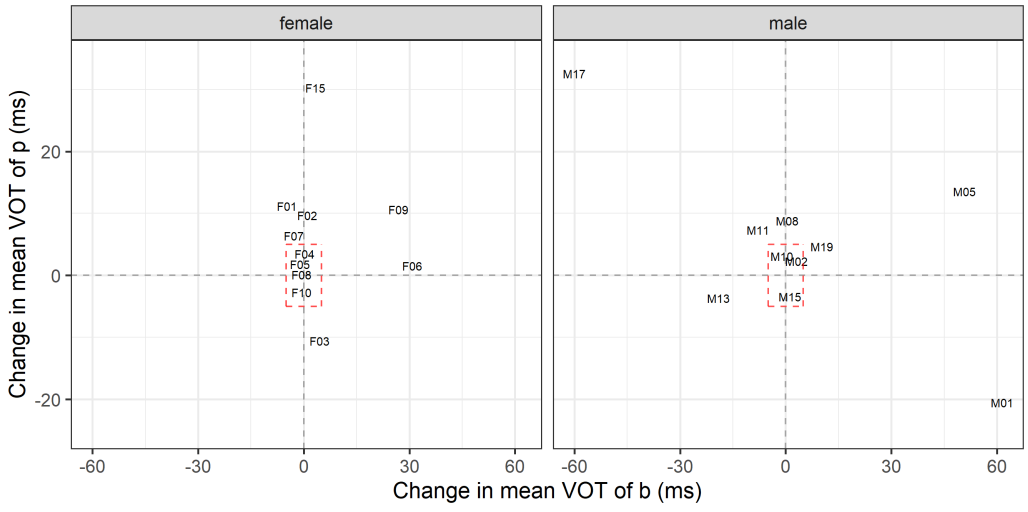

Figure 4.17: Change in mean VOT of /p/ and /b/ per person in the English Extr. Asp. data; The red rectangle shows 5 ms change of means in either direction for reference

These trends are even more apparent when compared with the behavior exhibited by English speakers in *Extr. Asp.* — the condition that resembled their VOT values the most. This is repeated in *Figure 4.17*.Many more Hungarian speakers changed their /b/ productions than the English speakers in either condition, while English speakers changed their /p/ productions more often than Hungarian speakers did. This indicates that English speakers were more able to adjust aspiration than Hungarian speakers. Moreover, while Hungarian speakers might not have had perfect control over the amount of prevoicing they used, as indicated by participants regularly overshooting the target by as much as 50 ms or more, prevoicing still lent itself to more variation in Hungarian. At the same time, changes in /b/ productions were nowhere near as drastic in Hungarian as they were in English, which indicates that the English mostly plain /b/'s allowed for far more room to demonstrate accommodation when a participant did converge with a prevoiced token.

4.3.3 The Extreme Aspirating condition

In the *Extreme Aspirating* condition participants were exposed to an extreme version of an aspirating contrast during shadowing. During the shadowing task, all the /p/'s participants heard were aspirated (had a VOT of 130 ms), and /b/'s were plain with no prevoicing (5 ms VOT).

Accommodation of /p/'s

The baseline model indicates no significant effect of exposure to the aspirated /p/'s (*Table 4.10*). The interaction between `Gender` and `Exposure` indicates a trend of divergence (shorter VOT's), but that effect is not significant under the Bonferroni-adjusted threshold (α=0.0125). This is not significant when only the male dataset is tested, but only barely so (p=0.0533).

	Estimate	Std. Error	Pr(> |t|)	
(Intercept)	28.303	2.713	<0.0001	***
Gender [male]	−2.882	3.301	0.3937	
Exposure [post]	0.861	0.626	0.1696	
Gender [male] × Exposure [post]	−1.981	0.885	0.0254	.

Table 4.10: LMER model of reading /p/ tokens' VOT (in ms) in the Extr. Asp. condition; Threshold for significance (adjusted with Bonferroni correction): p<0.0125

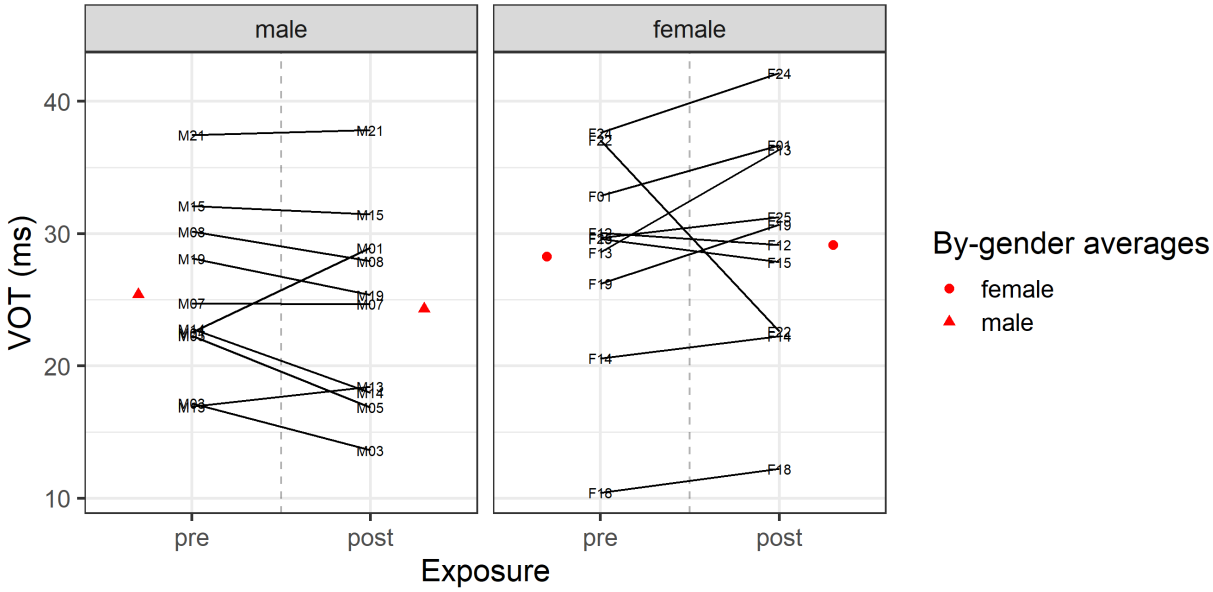

Figure 4.18: Change in mean VOT of Hungarian /p/'s in Extr. Asp. with by-gender averages

On an individual level (*Figure 4.18*), while there are some people (mostly females like F24) who converged—i.e. their /p/'s were most aspirated after exposure than beforehand—there are several people who diverged, and most made no or little changes to their /p/ productions. None of the three likeability measures seemed to mediate accommodation behavior. The full statistical tables are in *Tables A.48–A.50* and plots in *Figures A.33–A.35* in the Appendix.

Accommodation of /b/'s

No convergence was found in the /b/ dataset either (*Table 4.11*). In fact, males diverged from the model talker's plain /b/'s to a significant degree (i.e. they prevoiced more; `Gender [male]` × `Exposure [post]`: β=−8.891, p=0.0041).

	Estimate	Std. Error	Pr(> |t|)	
`(Intercept)`	−61.953	7.456	<0.0001	***
`Gender [male]`	−13.547	10.320	0.2051	
`Exposure [post]`	−0.243	2.184	0.9113	
`Gender [male]` × `Exposure [post]`	−8.891	3.090	0.0041	**

Table 4.11: LMER model of reading /b/ tokens' VOT (in ms) in the Extr. Asp. condition; Threshold for significance (adjusted with Bonferroni correction): p<0.0125

On an individual level, no males converged substantially (had much less prevoicing), many stayed roughly the same and even more diverged. For females, the lack of change on a group-level is a result of varying individual strategies. For instance, some females, like F13 and F19, converged

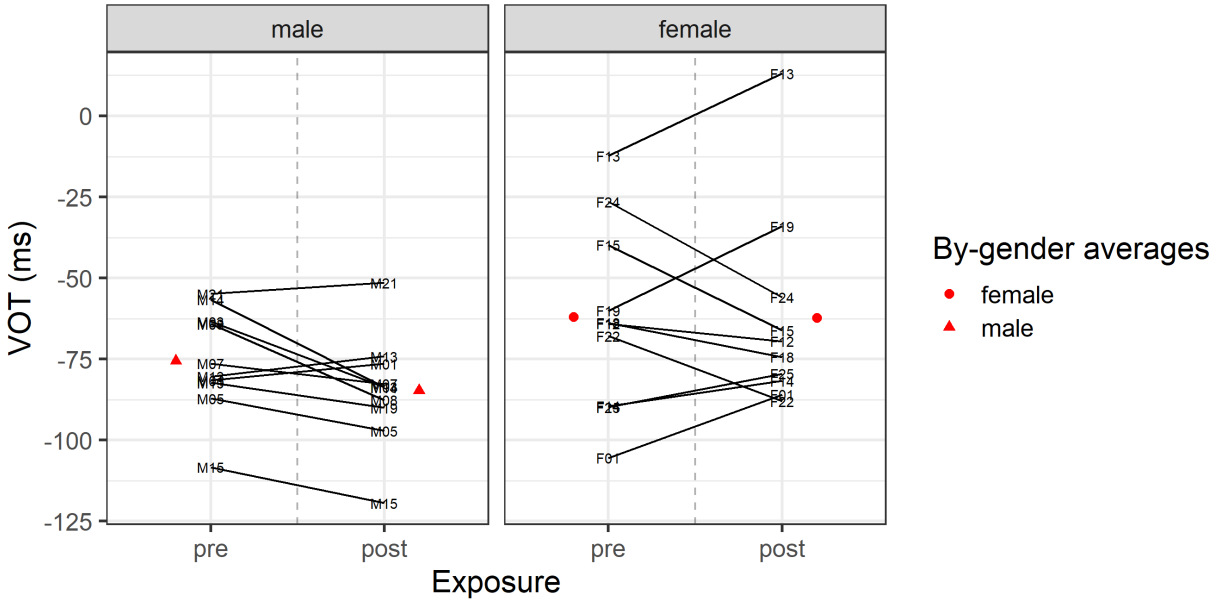

Figure 4.19: Change in mean VOT of Hungarian /b/'s in Extr. Asp. with by-gender averages

by a lot (prevoiced a lot less), but some others, like F24 and F15, diverged (prevoiced more) to a great extent. As a result of these strategies being roughly equally attested, on a group-level females seemed to not change at all. This is different from what we saw for /p/'s in this condition, where no change on a group-level was a result of only small or no changes on an individual level. Since the individual changes were much bigger for /b/, but balanced each other on a group level, it is probable that /b/ accommodation behavior was correlated with some of the likeability measures.

Indeed, we see effects for all three variables (statistical models in *Tables A.51–A.53*), but most clearly for `Solidarity` (*Figure 4.20* below). Adding `Solidarity` improves the explanatory power of the model. We find that while participants tended to diverge when they rated the participant low (`Exposure [post]`: β=−53.002, p<0.0001), this gradually gave way to convergence with higher `Solidarity` ratings (`Exposure [post]` × `Solidarity`: β=7.893, p<0.0001)—the model predicts convergence with ratings of at least 6.71. This pattern is found independently of `Gender` (`Gender [male]` × `Exposure [post]` × `Solidarity`: p=0.0931 > α=0.00625).

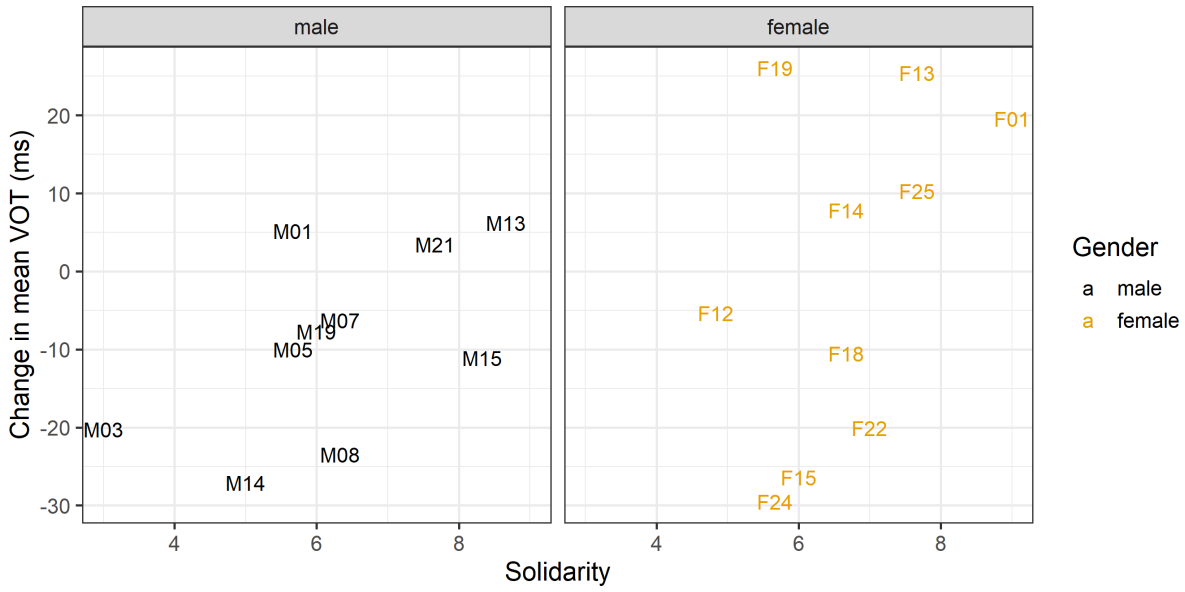

Figure 4.20: Change in mean /b/ VOT in Extr. Asp. in the Hungarian reading task by gender and Solidarity rating

In practice, most males ended up diverging, but the amount to which they diverged from the target was mediated by how they rated the model talker on `Solidarity`-related measures. For females, this effect showed up as a more-or-less exceptionless binary split between participants who rated the model talker lower (still mostly between 5 and 7), and those who rated her higher and converged with her. This subgroup of females is the only evidence of convergence to a plain /b/ (5 ms VOT), and thus should be further inspected for their behavior during Shadowing.

Similar, but less clear effects were seen for `Superiority` and `Dynamism` as well (*Figures A.36–A.37* in the Appendix), but these correlations are a little more spurious than the effect of `Solidarity` presented above.

Patterns in the treatment of the /p b/ contrast in the Extreme Aspirating reading data

We have seen above what group-level tendencies could be observed in accommodation to /p/ and /b/ separately. In this subsection, I am going to discuss how these tendencies came together and what we can say about the category representations of the Hungarian participants.

While overlaps between /p/ and /b/ were not attested among English-speaking participants, some Hungarian participants' /p/ and /b/ categories did overlap. In *Extr. Asp.* every participant who produced plain /b/'s at all still had a bimodal distribution of /b/'s (i.e. short negative VOT was not attested). Both before- and after exposure, all participants produced mainly prevoiced /b/'s, except for one participant, F13, who seemed to converge with the plain /b/ target. Her post-exposure productions were almost-exclusively made up of plain /b/'s (see *Figure 4.21*). In general, /p/ and /b/ tokens were often distinct from each other in terms of VOT, and in most cases there were at most a handful of plain /b/ tokens that were in the typical /p/ range. However, there were some notable exceptions, such as F13 and F24 (right-hand side of *Figure 4.23* below). This is in contrast with how neatly /p/'s and /b/'s were distinguished in the English dataset.

Figure 4.21: Reading patterns of F13 in the Hungarian dataset

In terms of holistic, contrast-level patterns, one could assume that /p/ accommodation and /b/ accommodation were completely independent. We saw significant `Solidarity`-mediated convergence to and divergence from the model talker's plain /b/ targets in *Extr. Asp.*, while no patterns of either convergence or divergence were observed for aspirated /p/ targets. At first blush, this suggests that there must not have been any relationship between the treatment of /p/ and /b/ on an individual level either.

However, this is not what the data shows. There was a moderately positive correlation between the change in means for /p/ and the change in means for /b/ (Pearson's r=0.6227, p=0.0034). This is also reflected in the figure below (*Figure* 4.22). While the amount of changes is vastly different along the two axes (changes in /b/ productions were much larger than changes in /p/ productions), the direction of most participants' changes along one axis matched the other. That is, if someone converged for one segment, they tended to do the same for the other (no matter how big or small the change was).

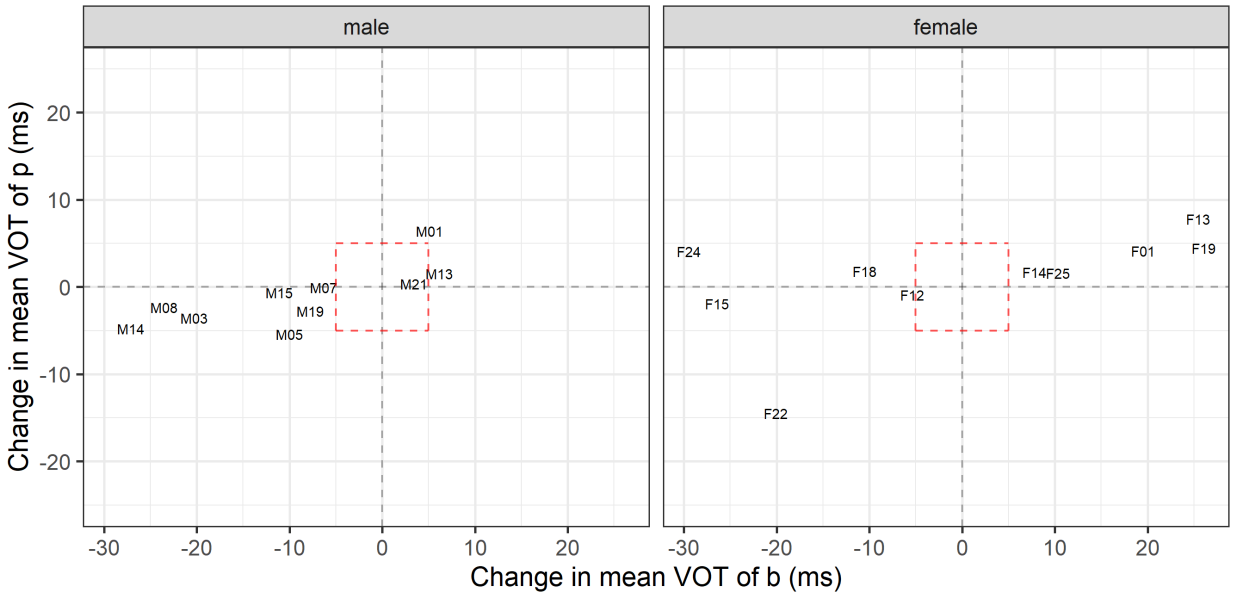

Figure 4.22: Change in mean VOT of /p/ and /b/ in Extr. Asp. per person in the Hungarian data; The red rectangle shows 5 ms change of means in either direction for reference

The only exceptions from this were F18 and F24, who aspirated their /p/'s more (positive along the y axis), but also prevoiced their /b/'s more (negative along the x axis). *Figure 4.23* shows that interestingly, they differed in how aspirated their /p/'s were to begin with: F24 tended to produce more aspirated /p/'s than F18, both before and after exposure. In case of F18, the shift in the amount of prevoicing is also accompanied by a widening of the prevoicing range mostly upwards—i.e. in

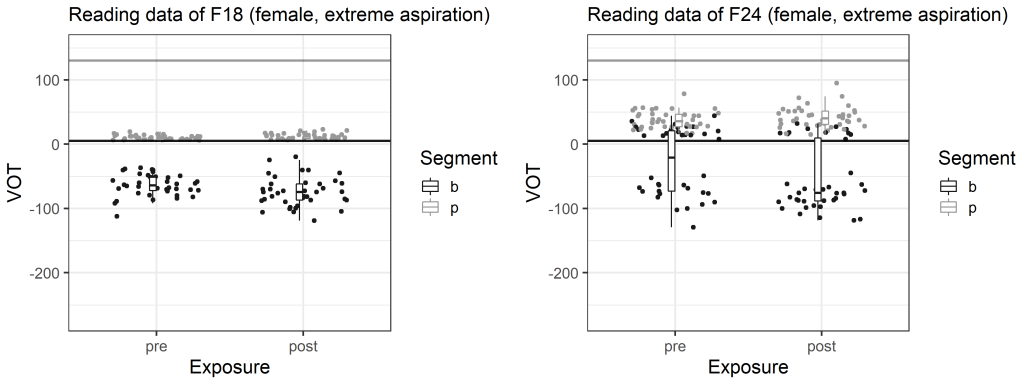

Figure 4.23: Reading patterns of F18 and F24 in the Hungarian dataset

the post-exposure reading task she produced tokens that were less prevoiced than her productions in the PRE-Read, while her /p/ productions barely changed. For F24, we can see a clear indication of more aspiration on /p/'s, while for /b/'s the VOT ranges used for plain and prevoiced /b/'s barely changed, merely the proportion of plain vs. prevoiced /b/'s shifted.

It can also be seen from *Figure 4.22* that changes along the x axis (changes in /b/) were far larger in their amount than changes along the y axis were (changes in /p/). This was also the case in the Hungarian *Extr. Prev.* data. To some extent this was true in the English dataset as well (those who changed their /b/'s made larger changes than those changing their /p/'s). There was one crucial difference though: while in the English dataset, the biggest changes were indeed observed in /b/ productions, changes in /p/'s were far more common. In the Hungarian dataset, /b/'s were not only changed the most but the most often as well.

This suggests a different relationship to prevoicing and aspiration as cues in the two languages: while aspiration might lend itself to more adjustments for English speakers (and doing the same for prevoicing might be a more difficult challenge), Hungarian speakers might find adjustments to prevoicing easier. This is, however, further contextualized by the fact that in the *Extr. Prev.* some Hungarian speakers substantially overshot the target amount of prevoicing in the post-exposure reading task. This suggests that while prevoicing might be easier to manipulate for Hungarian speakers, it might not be easy to manipulate *precisely*.

4.3.4 Summary

In this section we looked at the Reading results from Hungarian participants. While the reading task itself was the same across the two conditions, participants were exposed to different stimuli during a shadowing task that intervened between the PRE-Read and POST-Read task. In the *Extreme Prevoicing* condition participants were exposed to a prevoicing /p b/ contrast (plain /p/'s with 15 ms aspiration and prevoiced /b/'s with 130 ms prevoicing, i.e. -130 ms VOT), which was quite similar

to the way Hungarian normally expresses a /p b/ contrast. In the *Extreme Aspirating* condition, participants were exposed to an aspirating /p b/ contrast (aspirated /p/'s with 130 ms VOT and plain /b/'s with 5 ms VOT).

In the *Extreme Prevoicing* condition, most participants decreased the amount of aspiration on their /p/'s and prevoiced their /b/'s more (converged with the model talker on both sounds). The /p/ pattern was exhibited by males and females to a different extent (β=$-$3.508, p <0.0001 for males; β=$-$1.7616, p=0.0118 for females). This difference could be because males in this condition tended to have longer VOT's in the baseline PRE-Read and thus had more room to demonstrate convergence. Males and females also differed in the magnitude of /b/ accommodation (β=$-$9.882, p<0.0001 for males; β=$-$4.442, p=0.0142, α=0.025 for females). This time there was no systemic difference in pre-exposure reading values that would explain the gender difference. While we found similar gender-patterns for both segments (males converging to a larger extent than females), no statistically significant correlation was found between a given participant's treatment of /p/ and /b/ (Pearson's r, p=0.2844).

No link could be established between the three likeability measures and /p/ accommodation, but there did seem to be an effect of `Superiority` and /b/ accommodation, but for males only. That is, males who rated the model talker low on `Superiority`-related scales (i.e. *intelligent—unintelligent*, *organized—disorganized*, and *higher status—lower status*) tended to produce similar or even less prevoiced tokens after being exposed to the model talker's highly prevoiced tokens, while males who rated her higher on these scales produced more prevoicing (converged) after exposure.

In the *Extreme Aspirating* condition, there was no convergence in the baseline models for either /p/ or /b/. While females tended to either stay the same or converge with the model talker's aspirated /p/'s (i.e. have more aspiration post-exposure), most males if anything, got further from the target (aspirated less, i.e. diverged). However, the amount of change on average was so close

to 0, that no statistically robust patterns were found for /p/. As for /b/, males actually diverged from the plain /b/ target (they prevoiced more; β=$-$8.891, p=0.0041, α=0.0125), while females did not change. However, the apparent lack of change for females was a result of a mixture of some participants converging with and others clearly diverging from the target.

While likeability measures found no further patterns in the /p/ data, the male and female patterns for /b/ came together to form a uniform likeability pattern. All three measures correlated with accommodation behavior to some extent, but in the end it was `Solidarity` that proved the most useful for explaining the pattern. Males mostly rated the model talker low on these scales and diverged from her (i.e. prevoiced more), the amount of divergence was mostly proportional to the ratings they gave, and some males who rated the model talker positively on `Solidarity`-related scales even converged. Females tended to give her higher ratings to begin with, and 5 out of 10 females adjusted their /b/ productions towards less prevoicing (i.e. convergence). The other 5 females tended to rate the model talker lower and also diverged from her (prevoiced more), resulting in the apparent non-effect on a group-level. Even though the effects we found in the /b/ data were not detectable in the /p/ data, there was a moderately strong correlation between /p/ and /b/ accommodation in *Extr. Asp.* (Pearson's r=0.6227, p=0.0034). This could be because the same tendencies were there in the /p/ data as well, they were just so small in magnitude to detect over noise.

In both conditions /b/ accommodation was not only larger in magnitude compared to /p/ accommodation, but it was also more common. This is different from the English data, where changes in /b/ production were bigger (just like in Hungarian), but they were also rarer than the smaller but more frequent changes in /p/ production. The fact that both languages' speakers showed the biggest changes in terms of their /b/ productions could stem from the fact that both languages use the same range of prevoicing (albeit prevoicing is less often attested in English, but when it is, it is in a similar range to Hungarian prevoicing). The difference between English and Hungarian speakers

in which segment (and thus which cue) they adjusted could be a result of a difference in how often the speakers adjust these respective cues in their everyday speech. Since Hungarian word-initial voiced stops have prevoicing more often than their counterparts in English, Hungarian speakers might be able to more reliably produce prevoicing and therefore might be more accustomed to the variation and manipulation of this cue. It must be noted, that this does not necessarily translate into an ability to adjust the duration of prevoicing in a fine-grained way, which was seen from examples where some participants not only converged with a prevoiced /b/, but regularly overshot—after exposure, they produced even more prevoicing than they heard the model talker do.

While there were differences between the Hungarian and the English-speaking participants, in terms of sensitivity to certain cues, the overall, general patterns were largely similar. In both languages, participants converged with stimuli that were similar to their respective native language (*Extr. Asp.* for English speakers and *Extr. Prev.* for Hungarian speakers), albeit there was a small difference. English speakers' convergence with their native-like /b/'s (plain /b/'s in *Extr. Asp.*) was independent of all three likeability ratings (and only absent if their baseline productions matched the target to begin with). However, while convergence was detected in the Hungarian *Extr. Prev.* for /b/ on a group level even without the addition of likeability predictors, it was enhanced by high ratings on `Superiority`-related scales.

Convergence to the other ("un-native-like") condition was limited in both experiments. The likeability effect for Hungarian *Extr. Asp.* /b/'s is even consistent with the `Solidarity` effect found in the English data, where participants' accommodation behavior regarding the "un-English-like" *Extr. Prev.* stimuli was contingent on the `Solidarity` rating the given participant gave to the model talker. Just like here, in the English data it also manifested in the form of varying degrees of divergence, with only a few cases of convergence, suggesting a link between *dis*liking and divergence, rather than a link between liking and convergence.

This indicates that the lack of convergence in the English experiment (in *Extr. Prev.*) was not simply due to a general suboptimality of prevoicing as a cue, but that speakers of any language adhere to the phonetic specifications of their sound categories to some extent (i.e. they support the **Maintain categories** hypothesis over **Maintain contrasts**). However, we first need to take the data from the Shadowing task to account to confirm this.

4.4 Shadowing results

In this section I am going to present results from the shadowing data in the Hungarian experiment. First, I will provide an overview of the results as well as some basic information about the dataset, the statistical models that were used and remind the reader of what the questions at issue are and what parts of the dataset could help answer them (*Section 4.4.1*). Then I will present the results in more detail, broken down by condition. Just like in the case of the reading data, I will start with the *Extr. Prev.* condition, where stimuli were closer to the typical way the word-initial /p b/ contrast is realized in Hungarian (*Section 4.4.2*). Then, I will move on to the *Extr. Asp.* condition (*Section 4.4.3*). The section concludes with a summary of findings.

4.4.1 Overview and statistical methods

In this experiment, trajectories were a lot flatter than in the English experiment, and no difference was found between repetitions in either condition for either segment (*Figure 4.24*). The interpretation of these flat trajectories (whether it was instantaneous convergence or a lack of exposure effect) was largely done through a comparison of the two conditions.

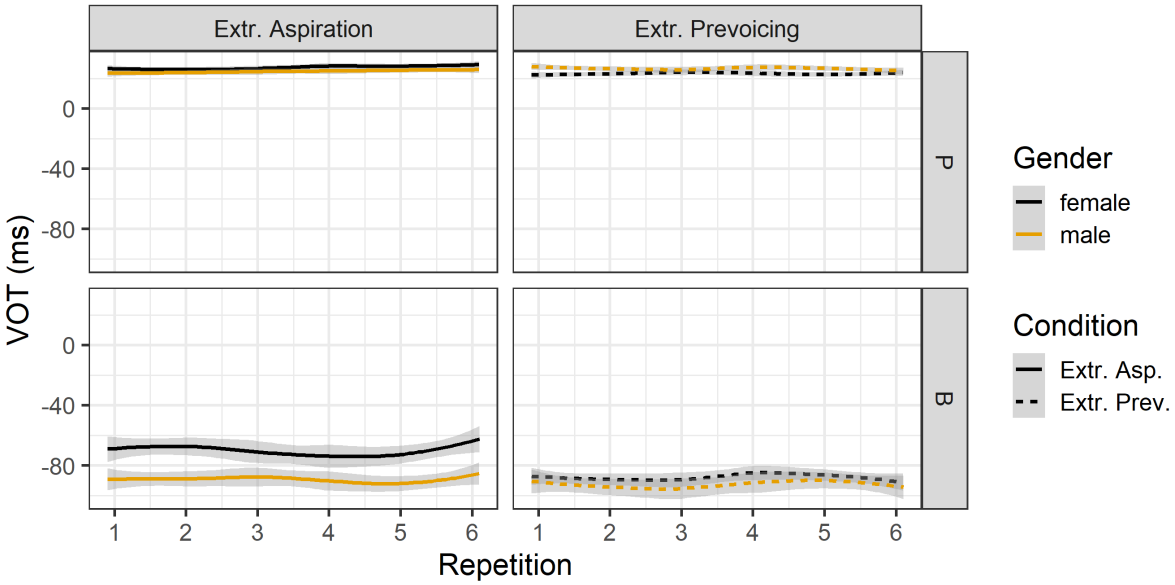

Figure 4.24: Smoothed results of the Hungarian shadowing data by condition and segment

I argue that the flat trajectories in *Extr. Prev.* are signs of convergence for both /p/ and /b/. Since this convergence happens to the same extent from the first repetition on, the repetition-by-repetition curves are flat. Participants often match (or on rare occasions even overshot) the target provided by the model talker in the *Extr. Prev.* shadowing task. In contrast, in the *Extr. Asp.* shadowing task there are no signs of group-level convergence, even though some individuals do produce values closer to the target, particularly for the plain /b/. This is in line with the strictly immediate convergence we saw in the English data in *Extr. Prev.* (the "un-English-like" condition). In the Hungarian *Extr. Asp.* shadowing data, some of the participants who showed no accommodation in reading do produce more prevoiced tokens during the shadowing section. Some signs of partial convergence (A-shaped trajectories) are also observed for /p/'s, but these are outnumbered by the vast majority of flat trajectories.

In terms of likeability, no effects can be established in either of the two conditions. This is in contrast with the English shadowing results from the "un-English-like" *Extr. Prev.* condition,

where we found a relationship between likeability (in terms of `Superiority`) and accommodation behavior for both /p/ and /b/.

Descriptive statistics

During the shadowing task, each participant heard a semi-randomly chosen set of 30 words of the 40 they all had to read out in the reading task. The participant encountered each of their 30 words 6 times throughout the task. The *Extreme Prevoicing* dataset consisted of 1,886 /p/'s and 1,890 /b/'s. 4 /p/ tokens were excluded because the participants either skipped the given item (3), or a creak made segmentation impossible (1). The *Extreme Aspirating* dataset was somewhat smaller, because of there only being 20 participants in *Extr. Asp.* compared to the 21 in *Extr. Prev.*, which meant 90 fewer recordings. The dataset was made up of 1,800 /p/'s and 1,792 /b/'s. 8 /b/ tokens had to be excluded (7 from the same speaker) because of yawns or realizations without a clear burst—i.e. the stop became nasalized or spirantized. This is in contrast with the English dataset, where in the more "unnatural" condition (in that case, *Extr. Prev.*) over 100 /p/ tokens had to be excluded based on participants' incredulous intonation patterns during the token, which indicated that they interpreted the plain /p/'s as tokens of /b/ and thus heard and repeated the p-initial words as b-initial nonce-words. In the Hungarian dataset, participants exhibited no signs of interpreting a plain /b/ as a /p/ even though a typical /b/ would normally be prevoiced in Hungarian.

Statistical analysis

Data were separated into four subsets by segment and condition: *Extr. Prev.* /p/ dataset, *Extr. Prev.* /b/ dataset, *Extr. Asp.* /p/ dataset, and *Extr. Asp.* /b/ dataset. These subsets were analyzed separately in linear mixed-effect regression models. The dependent variable was VOT (in ms), the independent variables were `Gender` and `Repetition` (1–6) with their interaction. By-participant and by-word random intercepts were added. These were considered the four "baseline" models.

In order to test likeability factors, three further models were run for each subset of the data (12 models in total), each including one of the likeability measures (`Solidarity`, `Superiority`, or `Dynamism`). The three likeability measures had to be separated into three models because adding the three measures resulted in models that did not converge. On top of the given likeability measure, these models also had `Gender` and `Repetition` (1–6) as independent variables, just like the baseline models. All interactions were included. Each model had a by-word random intercept but no by-participant random intercept (just like in the English dataset).

Foci of attention

The Hungarian shadowing dataset addresses the same questions as the English shadowing dataset did, but from the perspective of a prevoicing language, thereby hopefully eliminating language-specific and cue-specific effects. In this subsection I am going to recap what these questions are and in what part of the data we can find answers to each of them.

The main focus of this study is whether a pressure to adhere to the exact phonetic properties of sound categories plays a role in contrast maintenance as observed through an accommodation study. If it is (**Maintain categories**), then we expect to see no accommodation to plain /b/ stimuli (in *Extr. Asp.*), since a plain stop deviates so much from the typical realization of a word-initial /b/ that it would not even be categorized as a /b/ (without access to lexical information). Alternatively, contrasts can also be maintained via a more abstract and flexible mechanism (**Maintain contrasts**), which is satisfied as long as a bimodal distribution and sufficient distance is maintained between the contrastive segments without a reference to the exact phonetic characteristics of each category. If this is the case, plain /b/'s in *Extr. Asp.* should be a valid target of accommodation, since plain /b/'s in *Extr. Asp.* are clearly distinct from aspirated /p/'s.

Second, this study will also allow us to address other linguistic and methodological issues. One of these is whether segments which are both manipulated along the same phonetic dimension are

accommodated to together. This can be answered through comparing the amount of accommodation (convergence or divergence) for /p/ and /b/ on an individual level. Moreover, the data can tell us about whether there are any differences in how participants accommodate to longer gestures compared to shorter ones. In the English datasets we saw some effects of fatigue (during shadowing an increased amount of prevoicing was often only present until about halfway through the task and then gradually disappeared). In the Hungarian dataset we might see similar trajectories for aspiration, the cue that Hungarian speakers have less practice with. Furthermore, the shadowing data can be compared with the reading data to inform us about task effects. Such a comparison could reveal whether there were participants who were strictly immediate convergers or divergers—i.e. who converged to or diverged from the model talker, but only while exposed to them (during shadowing), and their post-exposure reading production did not reflect such a change.

Third, we can investigate the contribution of extra-linguistic factors as well, namely gender and likeability. While I found no gender effects in the English data (and argued that previous results from Nielsen, 2011 were due to phonetic distance and not gender), it is possible that we see some effects in Hungarian. This would corroborate the notion that social variables are not universal, but might surface differently across cultures. The main question around likeability is what components of this measure are responsible for effects seen in previous studies. To test this, likeability is divided into three components: Solidarity, Superiority and Dynamism. Unlike the English data, the Hungarian data cannot address the issue of ethnicity, as no ethnic variation was captured in the participant pool.

4.4.2 The Extreme Prevoicing condition

Accommodation of /p/'s

In this condition, /p/'s had 15 ms aspiration in the stimuli. When the entire dataset was tested, no effects were found (*Table 4.12*).[1] This means that no repetitions were significantly different from the baseline (`Repetition 1`). This is in contrast with the English dataset, where in a comparable situation, those participants who did not meet the plain /b/ target in *Extr. Asp.* to begin with did converge to the target in `Rep 4` and `Rep 6`. No effects or interactions of gender were significant either.

The fact that repetitions were not different from one another could be interpreted in two ways. It could either indicate that participants' productions were not influenced by the shadowing stimuli at all or that they were accommodating to the stimuli to the same extent throughout the entire task—and thus the repetitions did not differ from *each other*. I argue that the latter is more likely.

[1]Since 15 ms is within the typical range of word-initial VOT values for /p/ in Hungarian, the tests were repeated with a subset of the data—those whose pre-exposure read mean /p/ was either below 10 ms VOT or above 20 ms. This subset contained 1,257 tokens from 14 (out of 21) participants. No effects were found in this case either (*Table A.54* in the Appendix). Since the dataset from those who "had room" to demonstrate convergence was not significantly different from the entire *Extr. Prev.* dataset, I will use the full dataset in the following

	Estimate	Std. Error	Pr(> |t|)	
(Intercept)	22.101	2.999	<0.00001	***
Gender [male]	5.003	3.669	0.1867	
Rep 2	0.106	1.021	0.9174	
Rep 3	1.772	1.021	0.0829	
Rep 4	1.131	1.021	0.2684	
Rep 5	−0.321	1.021	0.7531	
Rep 6	1.477	1.025	0.1498	
Gender [male] × Rep 2	−0.885	1.412	0.5311	
Gender [male] × Rep 3	−3.198	1.413	0.0238	.
Gender [male] × Rep 4	−1.053	1.412	0.4561	
Gender [male] × Rep 5	−0.866	1.412	0.5399	
Gender [male] × Rep 6	−3.752	1.415	0.0081	.

Table 4.12: LMER model of shadowing /p/ productions in the Hungarian Extr. Prev. condition (target: 15 ms); Threshold for significance (adjusted with Bonferroni correction): p<0.0042

When we look at the data (*Figure 4.25*) we can see that indeed, the trajectories are flat. The by-repetition mean productions of the participants (traced by the light gray lines) are often in the vicinity of the model talker's VOT values (15 ms) even though there are some exceptions (participants with mean VOT over 30 ms).

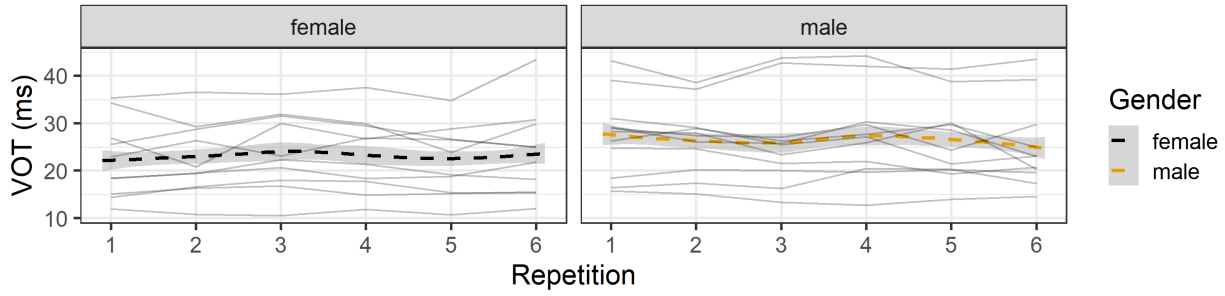

Figure 4.25: Smoothed results of the Hungarian Extr. Prev. /p/ shadowing data (all participants)

These observations are easier to interpret when put in the context of the other tasks (PRE- and POST-Read) and the other condition (*Extr. Asp.*) A plot of cross-task comparisons is included in *Figure 4.26*. In this plot we see /p/ tokens from PRE-Read, Shadowing and POST-Read collapsed into one set each. The left-hand side shows observations from *Extr. Asp.*, the right-hand side shows observations from *Extr. Prev.*

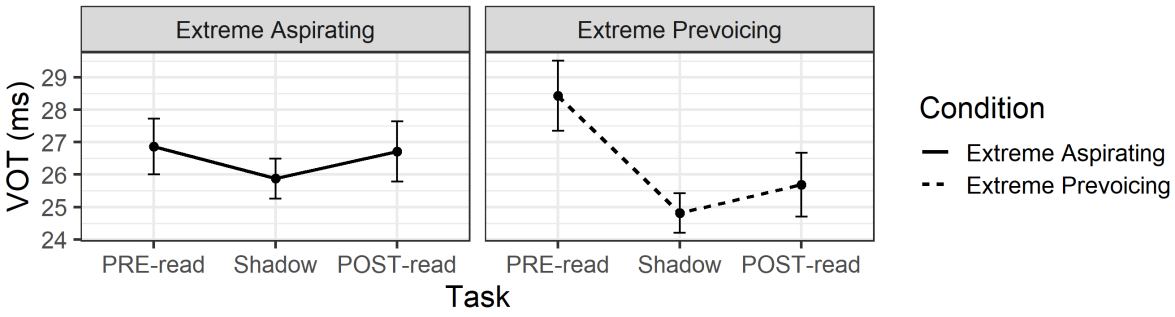

Figure 4.26: Cross-task averages of /p/ VOT by condition in the Hungarian experiments
Brackets represent the confidence interval of the estimate

We must be cautious when interpreting this figure. Since cross-task comparisons are also subject to task-effects, participants being closer to the target during shadowing compared to PRE-Read might not necessarily indicate convergence. Because of that, I restrict quantitative comparisons to meta-comparisons of trajectories across conditions—i.e. I compare the *relationship* between read and shadowed values between the two conditions rather than the shadowed and read data themselves.

While the PRE-Read values in *Extr. Prev.*, which were used as a baseline, were higher than those in *Extr. Asp.*, this was not significant (`Condition`: p=0.5274), and thus we can say that the

	Estimate	Std. Error	Pr(> |t|)	
`(Intercept)`	27.070	2.293	<0.00001	***
`Condition [Extr. Prev.]`	1.578	2.475	0.5274	
`Task [Shadowing]`	−0.984	0.402	0.0145	.
`Condition [Extr. Prev.]` × `Task [Shadow]`	−2.905	0.562	<0.00001	***

Table 4.13: LMER model of /p/ productions in the Hungarian recorded during PRE-Read and Shadowing; Threshold for significance (adjusted with Bonferroni correction): p<0.0125

two groups' baselines are comparable. This can be seen from the LMER model below (*Table 4.13*, with by-word and by-participant random intercepts).

What is more, the *Extr. Prev.* group's shadowing productions have significantly shorter VOT, i.e. closer to the target (β=$-$2.905, p<0.00001). This in and of itself is a cross-task comparison and as I pointed it out before could also indicate task effects and not just convergence. However, this effect was only present in the *Extr. Prev.* dataset, and nothing comparable was found in *Extr. Asp.* (`Task`, p=0.0145 $>$ α=0.0125). This means that while participants in the two conditions had comparable baseline productions, they reacted to the stimuli differently, since any task effect should have affected participants in the two conditions similarly. This difference between the two conditions makes it likely that the decreased VOT we see in *Extr. Prev.* Shadowing is likely convergence and not just a task effect. Within that, it is a case of instantaneous convergence, since the same amount of accommodation was present from `Repetition 1` on.

In terms of likeability effects, neither `Solidarity`, `Superiority` or `Dynamism` could be linked to accommodation. While there were significant effects in all three models, these all seem to be the results of overfitting or confounds. The `Superiority` effect for males does not so much capture accommodation behavior, but take advantage of a confound between males' PRE-Read baselines values and the ratings they gave. The `Superiority` effect looks like it is a result of over-fitting the model to a few participants' productions, and the `Dynamism` effect is a results of a confound between ratings given by females and their baselines. These plots and models are shown in full in the Appendix (*Figures A.38–A.40* and *Tables A.55–A.60*).

Accommodation of /b/'s

Stimuli tokens of /b/ in *Extr. Prev.* all had 130 ms prevoicing. Just like in the case of /p/ data, no effects were found in the /b/ data either (*Table 4.14*). Again, this means that no repetition differed

	Estimate	Std. Error	Pr(> |t|)	
(Intercept)	−87.681	7.340	<0.00001	***
Gender [male]	−3.123	9.945	0.7560	
Rep 2	−0.439	3.998	0.9130	
Rep 3	−1.768	3.998	0.6580	
Rep 4	3.289	3.998	0.4110	
Rep 5	0.455	3.998	0.9090	
Rep 6	−2.501	3.998	0.5320	
Gender [male] × Rep 2	−3.227	5.524	0.5590	
Gender [male] × Rep 3	−2.769	5.524	0.6160	
Gender [male] × Rep 4	−3.777	5.524	0.4940	
Gender [male] × Rep 5	−0.311	5.524	0.9550	
Gender [male] × Rep 6	0.017	5.524	0.9980	

Table 4.14: LMER model of shadowing /b/ productions in the Hungarian Extr. Prev. condition (target: −130 ms); Threshold for significance (adjusted with Bonferroni correction): p<0.0042

significantly from the baseline Rep 1. However, we can compare these productions to the *Extr. Asp.* dataset, where participants were exposed to plain /b/'s (5 ms VOT).

In the *Extr. Prev.* dataset 1,794 /b/ tokens were prevoiced, which is 94.9% of the total 1,890 /b/ tokens in this condition. This is significantly more than the 88.8% ratio we see in *Extr. Asp.* (χ^2 test with Yates' correction: χ^2=45.1966, p $<$ 0.00001). This difference is not due to some confound—e.g. that the participants sorted into *Extr. Prev.* tended to produce more prevoiced tokens to begin with. In the PRE-Read there was no significant difference between the two conditions neither in terms of ratio of prevoiced tokens (χ^2 test with Yates' correction: p=0.0838) nor in the amount of prevoicing (Condition: p=0.1360, see *Table 4.7* in *Section 4.3.1* on the pre-exposure reading productions). Indeed, visually, there are a lot of participants producing /b/'s with around

130 ms of prevoicing (or more), which suggests instantaneous convergence rather than no effect of exposure.

The fact that most people's trajectories were flat through the task (*Figure 4.27*) is important for another reason. In the English dataset we found that V-shaped trajectories were characteristic of the /b/ data, where participants prevoiced more and more until about halfway through the task, but then this prevoicing also gradually disappeared by the end of the task, often without a trace. These V-shaped trajectories are nowhere to be found, which indicates that in the case of English speakers it was in fact of a sign of some sort of fatigue. This is also supported by the fact that there are some participants in the Hungarian data whose *mean* productions are consistently below the −130 ms VOT target. This means that at least some people must have consistently produced /b/'s that were more prevoiced than the heavily prevoiced target.

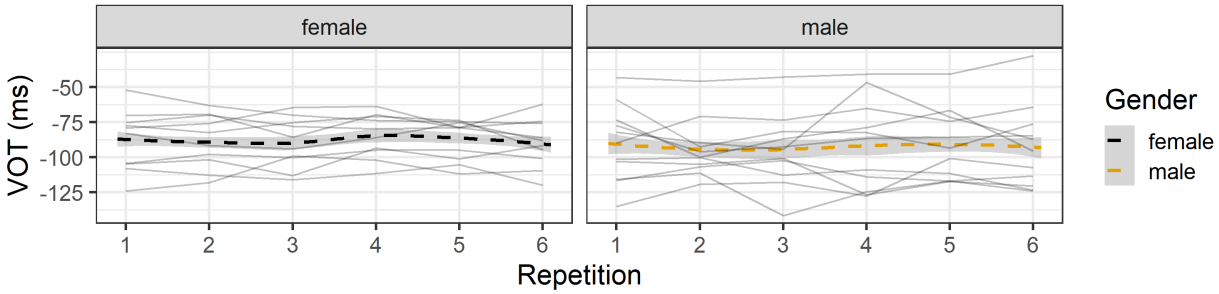

Figure 4.27: Smoothed results of the Hungarian Extr. Prev. /b/ shadowing data (all participants)

None of the three likeability effects could be linked to accommodation behavior for *Extr. Prev.* /b/'s. This is in line with the English data, where in the more English-like condition (in that case *Extr. Asp.*) we also found convergence across the board. The plots and statistical models can be found in *Figures A.41–A.43* and *Tables A.61–A.66* in the Appendix.

Patterns in the treatment of the /p b/ contrast in the Extreme Prevoicing shadowing data

After discussing both the /p/ dataset and the /b/ dataset separately, in this subsection I am going to put the two together and see how the contrast as a whole was affected during the shadowing task.

First, I am going to start by looking at where the two categories were with respect to one another (i.e. degree of overlap) and then see how the contrast changed over time (i.e. if the two sounds got closer to one another or further apart) on an individual level.

Tokens of /p/ were almost all below 50 ms VOT, and were especially concentrated around 0 (marked by the black line on *Figure 4.28*). As mentioned before, /b/ tokens not only reached the target (−130 ms VOT), but sometimes surpassed it too, especially for men. Just like in the Reading data, overlaps between the /p/ and /b/ categories were attested in the Shadowing data as well. The extent of this was limited. This can be seen on a group level in *Figure 4.28* from how few yellow dots (/b/'s) are in the short-lag region. This was also true on an individual level: while there were participants who produced plain /b/'s in the VOT range where /b/ normally was, there were few of them with only a handful of such tokens each (e.g. see *Figure 4.29*). Even for these participants, plain /p/'s tended to have a wider range (higher maximum VOT) than the handful of plain /b/ tokens did.

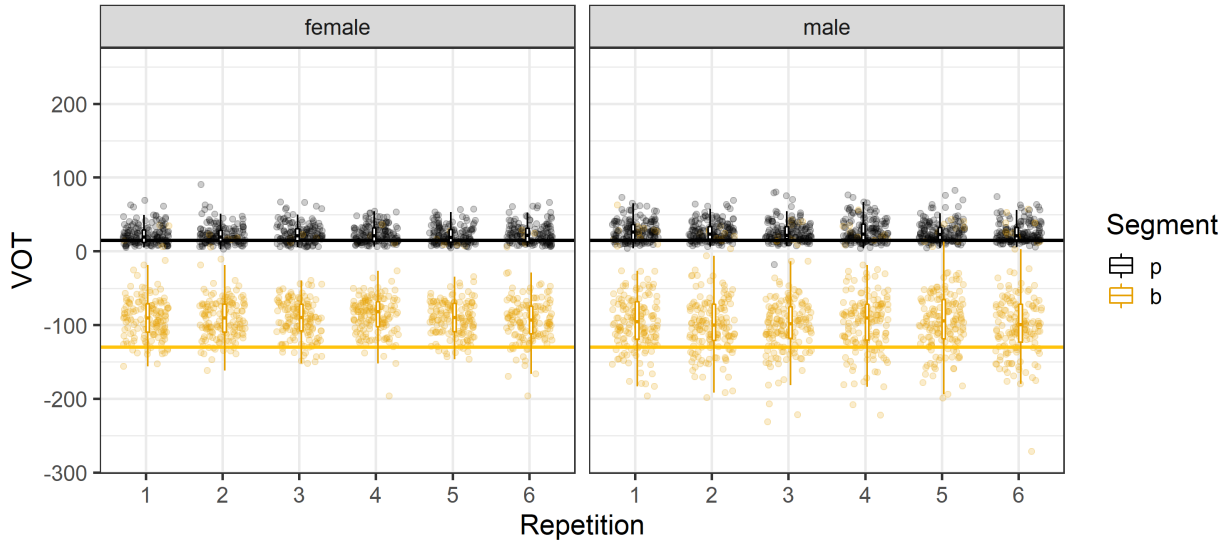

Figure 4.28: All participants' shadowing productions in Hungarian Extr. Prev

In terms of trajectories, since the accommodation trajectories were largely flat for both /b/ and /p/, the categories neither moved closer to or or became more distant from one another on an individual level either.

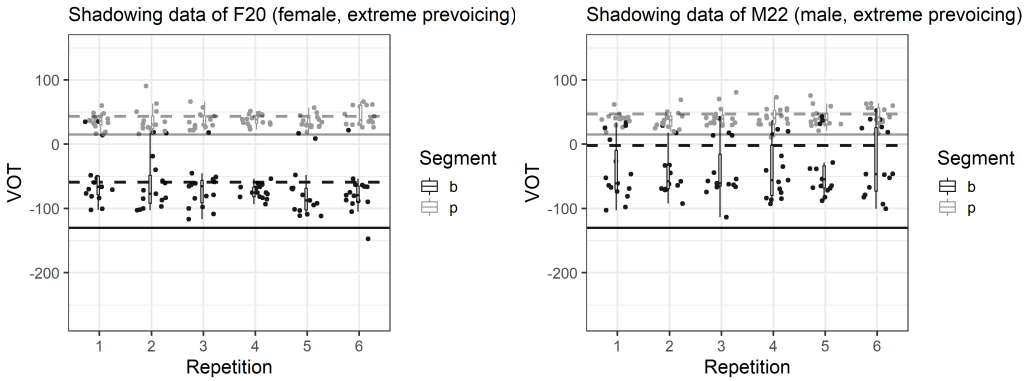

Figure 4.29: Shadowing patterns of F20 and M22 in the Hungarian dataset
Solid lines are target values, dashed lines are the participant's individual baseline from PRE-Read

4.4.3 The Extreme Aspirating condition

Accommodation of /p/'s

In *Extr. Asp.*, the /p/'s in the stimuli had 130 ms VOT, i.e. they were heavily aspirated. When a model was run on the tokens that participants uttered as a response to these stimuli (*Table 4.15*), no effects were found. There was a trend towards longer VOT in `Repetition 6` (compared to the baseline `Rep 1`), but after applying Bonferroni correction to compensate for the number of tests that were run, this was no longer significant (p=0.0080 $>$ α=0.0042).

	Estimate	Std. Error	Pr(> |t|)	
(Intercept)	26.339	2.911	<0.00001	***
Gender [male]	−2.762	3.598	0.4512	
Rep 2	−0.426	1.078	0.6925	
Rep 3	−0.053	1.077	0.9611	
Rep 4	1.931	1.077	0.0733	
Rep 5	0.810	1.077	0.4523	
Rep 6	2.862	1.077	0.0080	.
Gender [male] × Rep 2	0.900	1.524	0.5549	
Gender [male] × Rep 3	1.315	1.524	0.3881	
Gender [male] × Rep 4	−0.270	1.524	0.8595	
Gender [male] × Rep 5	1.212	1.524	0.4264	
Gender [male] × Rep 6	−0.419	1.524	0.7832	

Table 4.15: LMER model of shadowing /p/ productions in the Hungarian Extr. Asp. condition (target: 130 ms); Threshold for significance (adjusted with Bonferroni correction): p<0.0042

Indeed, the trajectories were almost all flat on an individual level (individual trajectories are shown by the faint gray lines in *Figure 4.30*). While the flat trajectories in *Extr. Prev.* were a sign of instantaneous convergence—the same amount of convergence happening from Rep 1 on—in *Extr. Asp.* it is more likely to be a sign of no convergence happening at all. This interpretation is supported by the fact that while the VOT values recorded during shadowing in *Extr. Prev.* were much closer to the target plain /p/ than those recorded during the PRE-Read from the same participants, in *Extr. Asp.* /p/'s did not change much at all from the PRE-Read to Shadowing.

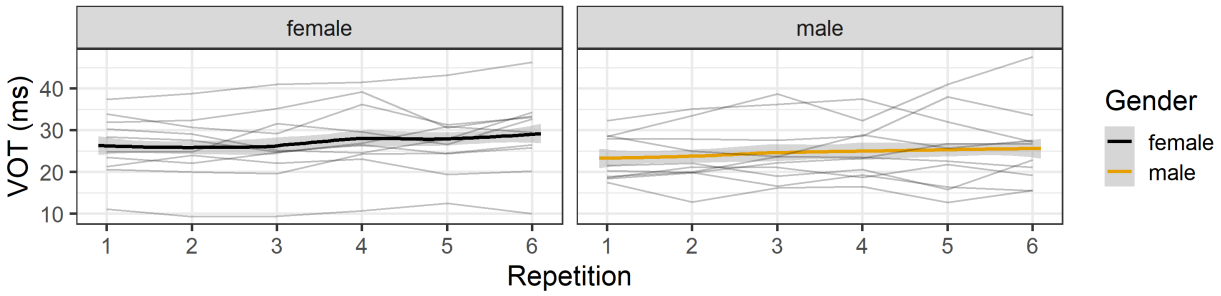

Figure 4.30: Smoothed results of the Hungarian Extr. Asp. /p/ shadowing data

Aside from flat trajectories, there were also some (but not many) A-shaped curves with VOT until about the middle of the task, then decreasing. This could could indicate (articulatory or mental) fatigue and is similar to the instances of V-shapes found for prevoicing in the English dataset (where some participants prevoiced more up until a point during shadowing, but then this prevoicing also disappeared by the end of the task). The magnitude of increase was at most around 5–10 ms, which participants should have been able to maintain throughout the task from an articulatory perspective. Therefore, the gradual disappearence of this aspiration is more likely due to mental fatigue (decreased attention to the task) than an articulatory one.

In terms of likeability effects, `Superiority` and `Dynamism` did not affect accommodation. While there was an effect of `Superiority` and one of `Solidarity` in the dataset, these were likely due to overfitting for a few females' and males' productions, respectively. The plots and statistical models can be found in the Appendix (*Figures A.44–A.46* and *Tables A.67–A.72*).

Accommodation of /b/'s

There was very little happening in the /b/ dataset, where stimuli were 5 ms VOT /b/'s. None of the tested variables affected VOT productions significantly (see *Table A.73* in the Appendix). Trajectories were flat, and tokens were mostly prevoiced. There was no sign of the V-shaped trajectories that could be observed in the English dataset in cases when participants did attempt substantial amounts of prevoicing. While all but one participant's means were in the prevoiced

(negative) region, the VOT values are still somewhat higher (i.e. the tokens are less prevoiced) than the values we saw in *Extr. Prev.*

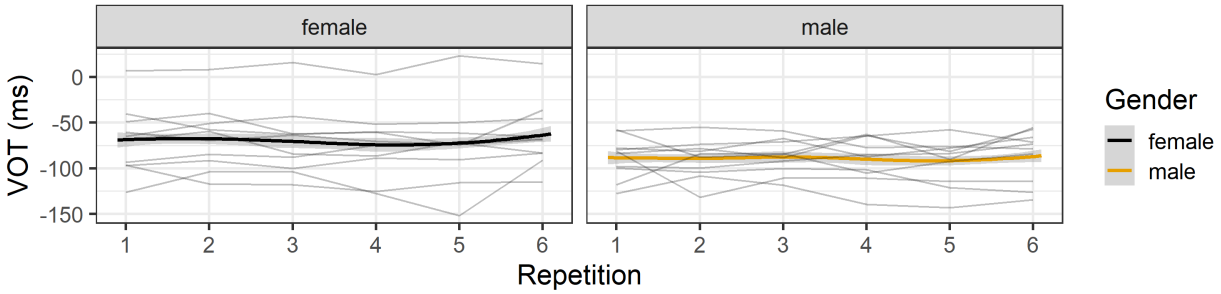

Figure 4.31: Smoothed results of the Hungarian Extr. Asp. /b/ shadowing data

In terms of likeability, no relationship could be established between accommodation and either `Solidarity`, `Superiority`, or `Dynamism`. While effects were found in the statistical models for all of these variables for one gender, these effects were in an unexpected direction (likeability correlated with *divergence*) and were likely due to over-fitting for a few participants. The plots and model estimates can be found in the Appendix (*Figures A.47–A.49* and *Tables A.74–A.79*).

Patterns in the treatment of the /p b/ contrast in the Extreme Aspirating shadowing data

Unlike in the *Extr. Prev.* condition, participants in the *Extr. Asp.* condition did not seem to be affected by the stimuli—they neither converged nor diverged with it. As can be seen from *Figure 4.32*, there was a large amount of overlap between /p/ and /b/, even more than in *Extr. Prev.*, where the stimuli likely have pulled the two categories further apart.

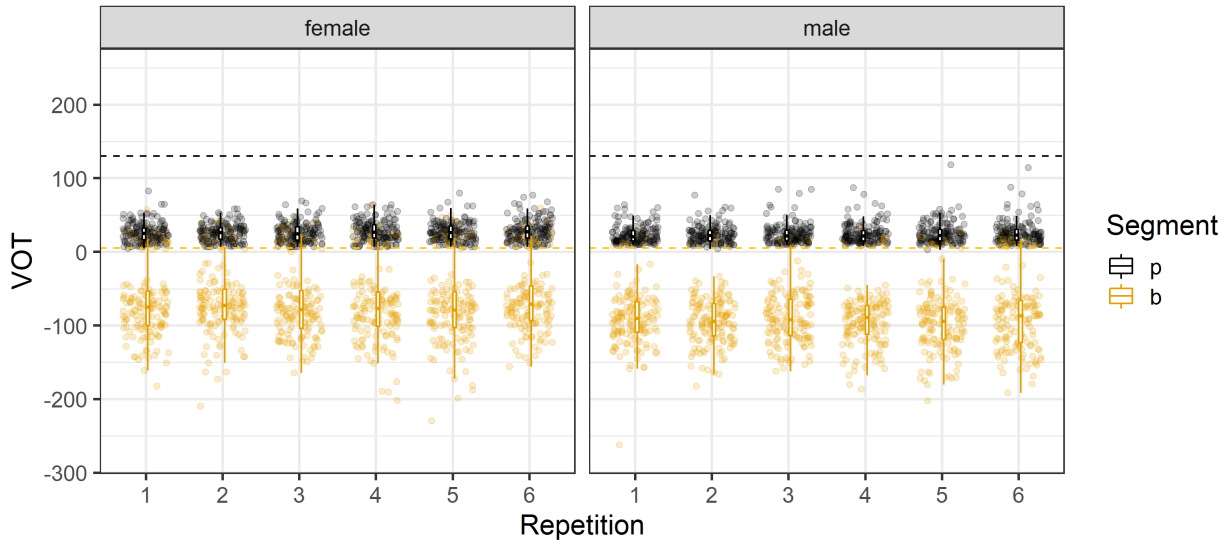

Figure 4.32: All participants' shadowing productions in Hungarian Extr. Asp

These overlaps were not just present on a group level, but could be observed in the productions of many individuals as well. Some of the most extreme examples are shown in *Figure 4.33*. Some of the participants with the most overlap in the shadowing productions hsad less overlap in their reading productions. This suggests that some of the overlaps might have indicated strictly immediate convergence (i.e. convergence that was only present while the participant was exposed to the model talker and could not be found during the POST-Read afterwards). Thus, even though on a group level participants did not accommodate to the stimuli, it appears that certain individuals did. This is all in line with what we saw for the English dataset in *Extr. Prev.* (no convergence on a group level, and a handful of participants converging with the model talker, but only while exposed to her). Interestingly enough, strictly immediate convergence is restricted to unnativelike /b/'s in both languages.

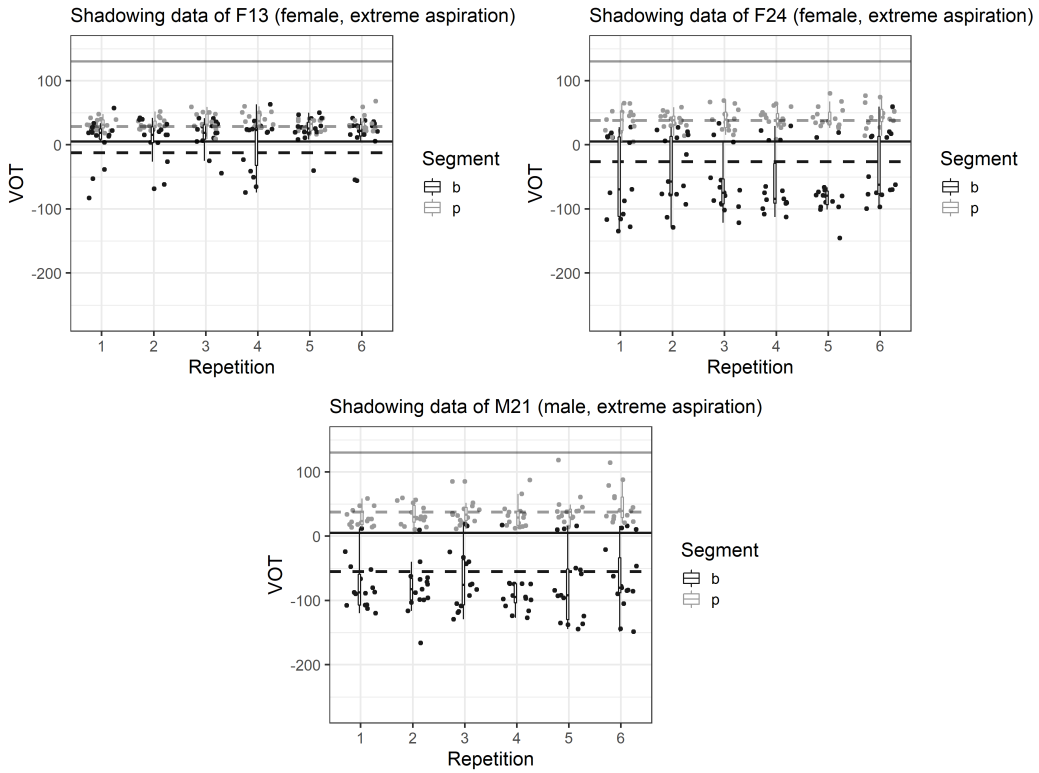

Figure 4.33: Shadowing patterns of F13, F24, and M21 in the Hungarian dataset
Solid lines are target values, dashed lines are the participant's individual baseline from PRE-Read

There was some overlap between /p/ and /b/ tokens. While some plain /b/'s encroached on the territory of short-lag /p/'s, plain /p/'s tended to have longer VOT than the /b/ tokens that were realized as plain stops. However, the extent of this overlap was not only greater here compared to read productions, but there were a lot more of these individual-level overlaps in the *Extr. Asp.* shadowing data compared to the *Extr. Prev.* condition. This could further indicate that the flat trajectories in *Extr. Prev.* were indeed cases of instantaneous convergence, since the categories were "pulled apart" by the stimuli (15 ms VOT /p/ and −130 ms VOT /b/).

Since trajectories were flat, no patterns of /p/ and /b/ either "moving together" or moving further apart could be established.

4.4.4 Summary

The results from the two conditions were quite different, just like in the English dataset, but unlike in the English dataset, there were similarly flat trajectories in both. In *Extr. Prev.*, there was convergence for both /p/ and /b/. It appeared in `Repetition 1` and remained consistent for the rest of the task, resulting in a flat trajectory. The reason for believing that it was still convergence was because it was close to the target and different from the PRE-Read in a way that was not observed in *Extr. Prev.*, thus was unlikely to be a result of task effects. There was some degree of overlap (more than in either of the English conditions). This was a result of some /b/ tokens being plain (rather than prevoiced). These tokens were the minority, however (5.1%), thus /p b/ overlaps were relatively few. Neither likeability effects nor gender effects were found in *Extr. Prev.*

In *Extr. Asp.* no group-level convergence was seen for for either /p/ or /b/. Certain individuals exhibited strictly immediate convergence with the model talker's plain /b/'s, but there were only few of them. No V-shaped trajectories were seen for /b/'s (unlike in the English data), which suggests that the Hungarian participants were able and willing to maintain the large degree of prevoicing in their productions. A small amount of A-shaped trajectories were seen in the /p/ dataset (some participants gradually increased the amount of aspiration on their /p/'s, but this gradually disappeared during the second half of the task). Since the peak mean VOT was still below 40 ms, this was likely not due to articulatory difficulties with maintaining aspiration, but potentially due to attention fatigue instead. In *Extr. Asp.* there was even more of an overlap between /p/'s and /b/'s due to plain /b/'s than in the *Extr. Prev.* shadowing task, but the ratio of plain /b/'s was still relatively low (11.2%). Just like in *Extr. Prev.*, no effects of any of the three likeability measures were found in *Extr. Asp.* either.

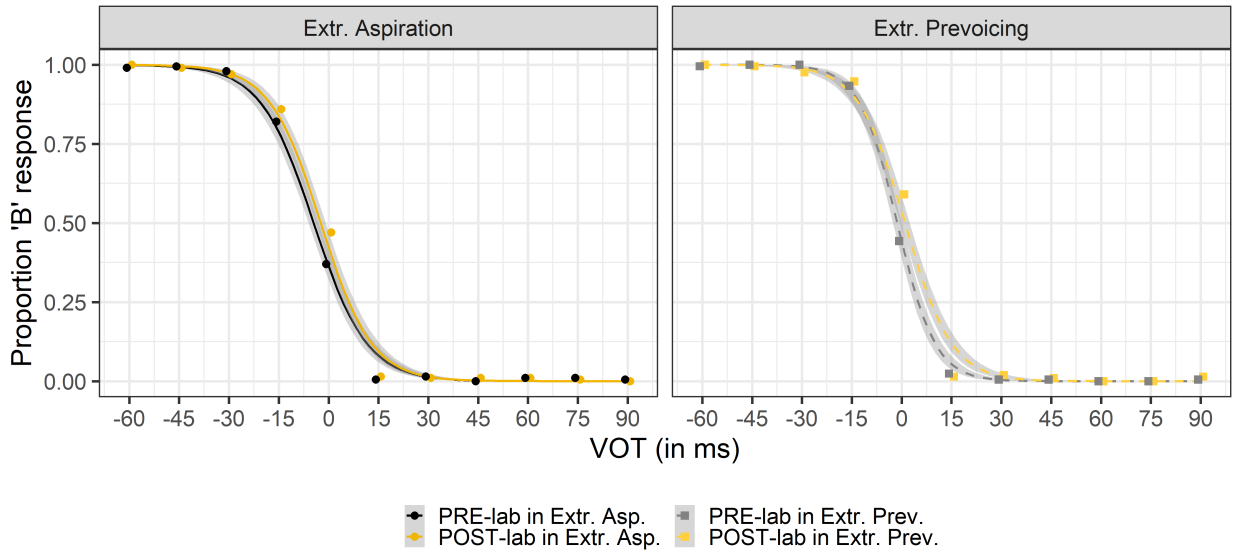

Figure 4.34: Labeling performance by condition in the Hungarian experiment

4.5 Labeling results

Participants also carried out a labeling task, both before exposure (after the rating task, before PRE-Read, i.e. PRE-Labeling) and after exposure (after POST-Read, before the questionnaire, i.e. POST-Label). During this task they had to rate tokens representing 11 steps of a VOT continuum and had to decide whether the word they heard was *boros* /ˈboroʃ/ 'wine-related' or *poros* /ˈporoʃ/ 'dusty'. The continuum went from +90 ms VOT (aspirated) to −60 ms VOT (prevoiced) and steps were 15 ms apart. Each participant judged each step of the continuum 10 times, resulting in a total of 110 responses. The *Extr. Prev.* dataset consists of 2,310 PRE-Labeling and 2,310 POST-Labeling judgements, and the *Extr. Asp.* dataset consists of 2,200 PRE-Labeling and 2,200 POST-Labeling judgements.

Figure 4.34 shows the data from both conditions as a proportion of /b/ (*boros*) responses. The circles (*Extr. Asp.*) and squares (*Extr. Prev.*) represent what proportion of responses to the given stimuli was *boros*. Since the smoothed curves were fitted to the raw dataset of /p/ and /b/ responses and not the proportions, the curves do not always touch all the symbols.

In both conditions, participants' /p b/ boundary is around 0 ms, which can be seen from all four curves having their inflection point around 0 ms (the 0 ms step is ambiguous, the proportion of /b/ judgements is around or slightly below half). This was not only the case on a group level, but also individually—most participants labeled the 0 ms stimuli some of the time as /b/, some of the time as /p/. This is different from the English data, where the inflection point of the curves did not coincide with any of the steps themselves (most 15 ms stimuli were labeled as *binning*, and most 30 ms stimuli were labeled as *pinning*). This indicates that the 0 ms step was almost exactly on the /p b/ boundary and was almost maximally ambiguous for most participants.

There was no demonstrable effect of exposure in the *Extr. Asp.* condition (left-hand side of the *Figure 4.34*). The biggest difference between pre-exposure and post-exposure behavior was seen at the 0 ms step (37% /b/ in PRE-Labeling and 47% /b/ in the POST-Labeling). Even though this pattern was attested in the behavior of 13 out of 20 participants, the changes in each case were so small (one decision's worth of difference) that this was not significant (p=0.1797 with `lsmeans`). This indicates that the *Extr. Asp.* stimuli did not compel participants to adjust their categories in perception. Accordingly, we saw no change in the "un-Hungarian-like" condition (*Extr. Asp.*) in production either. Thus, the lack of exposure effect in the labeling task are consistent with a lack of exposure effects in production.

These results can be contrasted with the English data, where participants in *Extr. Prev.* (the "un-English-like" condition) changed their labeling behavior as a result of exposure: they were less likely to rate 15 ms VOT stops as /b/ after exposure than they were beforehand. This change tended to go with a change in production, mostly convergence to /b/'s during shadowing. This happened in *Extr. Prev.*, which was the condition less like the English-speakers' typical productions.

However, a significant difference was found at the 0 ms step in the Hungarian *Extr. Prev.* condition (right-hand plot). Before exposure, participants labeled on average 44.29% of the 0 ms stimuli as *boros* /ˈboros/ 'wine-related'. This number went up: participants labeled 59.05% of these

stimuli as *boros* after being exposed to the *Extr. Prev.* stimuli. On an individual level, this change came together from an overall tendency exhibited by most participants in *Extr. Prev.* 13 out of 21 participants gave more /b/ responses to the 0 ms VOT stimuli than beforehand (on average 1–3 of their decisions differed). The change in labeling could not necessarily be matched with any pattern in production changes, especially since most participants in *Extr. Prev.* changed their productions.

This difference between pre-exposure and post-exposure behavior seems inexplicable and counterintuitive at first. *Extr. Prev.* was the condition that is most similar to Hungarian (the shadowing stimuli had 15 ms VOT /p/'s and −130 ms VOT /b/'s), and therefore changes to the categories or the categorization were not required to accommodate. If the *Extr. Prev.* stimuli differed from the participants' habitual speech it was for /b/: the model talker's /b/'s had more prevoicing than Hungarian typically has. However, according to the labeling results, participants became more willing to call a plain (0 ms VOT) stimulus a /b/ after being exposed to stimuli with more prevoicing than what their own speech has—i.e. they became more permissive with what they label as a b-word. At first blush, this seems like an unexpected development.

However, the answer lies in not the /b/'s but the /p/'s. The change in labeling makes more sense if instead of interpreting it as participants becoming more permissive or willing to label 0 ms VOT stimuli as a b-word, we should interpret this change as participants becoming more reluctant to label it a p-word. This is an understandable outcome of a study that repeatedly exposed participants to only a single value of /p/ VOT during shadowing. This single value of /p/ (15 ms) was higher than 0 ms, thus participants could have inferred that the model talker's /p/'s have 15 ms VOT, and whatever has a VOT below that is not a /p/. While in this light the change becomes explicable, it is still interesting why participants seemed to focus on /p/ VOT's rather than /b/ VOT's, which could be the subject of future research.

4.6 Interim discussion

In this section, I am going to summarize the results of the Hungarian experiment, followed by a brief discussion on how this dataset informs the questions outlined at the beginning. These results will be briefly compared with the English results whenever relevant.

4.6.1 Summary of results

Just like the English experiment, the Hungarian experiment also collected three types of phonetic data: production data from a reading and a shadowing task as well as perceptual information from a labeling task. In the reading task, participants converged with the more Hungarian-like stimuli in *Extr. Prev.* (15 ms VOT /p/ and −130 ms VOT /b/). Males demonstrated bigger changes for both segments: they shortened their /p/ VOT more and increased the amount of prevoicing on /b/'s by more than females did. For /p/'s this could be explained by males tending to have longer VOT baselines. Even though it was non-significant, there was a trend for males to produce /p/'s with more VOT in the PRE-Read, and thus they had more room to demonstrate convergence. However, no such relationship was found for /b/'s. In addition, males' productions seemed to have been influenced by an effect of `Superiority` (males who rated the model talker higher converged more than those who rated her low). While the gender effects were present in both the /p/ and /b/, no correlation was found between the treatment of the two categories on an individual level.

In the *Extr. Asp.* condition, where participants were exposed to plain /b/'s (5 ms VOT) and aspirated /p/'s (130 ms), no effect of exposure was found for either segment (for either gender). However, there seemed to be a relationship between the `Solidarity`-related ratings a participant gave and the decrease in their /b/'s prevoicing. Participants who rated the model talker high tended to converge with her (prevoice less) and those who rated her low tended to diverge (prevoice more). This effect was observed by most males and more than half of females diverging, which indicates a link between disliking and divergence (rather than one between liking and convergence).

A combination of these two tendencies gave the impression of no change for /b/'s in *Extr. Asp.* In this condition the change a given participant made to their /p/'s correlated with the change they made in terms of their /b/'s.

In the shadowing data, participants' trajectories were flat in both conditions (i.e. the mean production values of a given participant did not change much from repetition to repetition). In the *Extr. Prev.* condition this could be interpreted as convergence, taking place from the first repetition on. This convergence was not affected by gender either. The resulting productions of /p/ and /b/ overlapped to some extent (due to plain /b/'s), which was more than what we saw in either of the English conditions, but less than what we saw in the other Hungarian condition (*Extr. Asp.*).

In the Hungarian *Extr. Asp.* condition no group-level convergence was found for either /p/ or /b/ with or without likeability variables.There were a few participants whose read productions did not indicate any change but produced shadowed tokens much closer to the model talker's target. Since this was not true of participants across the board (most participants' shadowed productions differed little from their pre- and post-exposure read productions), this is more likely to be a sign of strictly immediate convergence than a task effect (a systematic difference between how participants produce words in the context of the two tasks). In *Extr. Asp.* for /p/'s we did see some (though not many) instances of A-shaped curves, which indicate that some participants started to prevoice more and more until a certain point, when they abandoned this trajectory and reverted to their starting values. This trajectory is essentially the aspirating equivalent of the V-shape we saw for prevoicing in the English data. However, it must be noted that the"aspirated" values that resulted in the A-shaped curves were still often in the short-lag region (below 40 ms VOT).

In the labeling task we only saw a difference in *Extr. Prev.* Participants who were exposed to the plain (15 ms VOT) /p/ and prevoiced (-130 ms VOT) /b/ stimuli of *Extr. Prev.* ended up being more reluctant to call a 0 ms VOT stop a /p/. This could be a result of the 15 ms VOT /p/ stimuli setting too "high" of a standard for /p/'s (i.e. that they must be distinctly short-lag). In the *Extr. Asp.*

condition, however, no changes were found, which indicates no perceptual adjustments were made to attune to the model talker's productions. While this differs from the English results found in their respective "atypical" condition (*Extr. Prev.* in the English case), it was in line with the fact that in the Hungarian *Extr. Asp.* condition no group-level adjustments were found in production either.

4.6.2 Mechanisms of contrast maintenance

In *Chapter 2*, I formulated two hypotheses for how contrasts might be maintained and thus how potential targets of accommodation can be limited by the speaker's pre-existing representations. The first of these is **Maintain contrasts**, which is a pressure to maintain distinctive pairs of segments as distinctive. This pressure is satisfied as long as the two segments in question form a bimodal distribution—perhaps along some predefined phonetic dimension(s)—and does not necessarily restrict the exact phonetic parameters of either of the two categories. The alternative is **Maintain categories**, which acts perhaps in addition to Maintain contrasts. Maintain categories mandates an adherence to the phonetic details of a given category. More specifically, it mandates that a category cannot be realized in a way that the speaker themselves would not recognize (categorize) the resulting tokens as tokens of their intended category.

As we saw earlier (in *Section 2.4*), these two hypotheses make slightly different predictions for which targets will and will not be accommodated to in the Hungarian experiment. These predictions are repeated in *Figures 4.35 & 4.36* below. The two hypotheses make the same predictions for 3 out of the 4 types of stimuli. Their predictions overlap for both of the *Extr. Prev.* stimuli, and predict convergence ('✓') or a matching of the targets by default without even converging ('(✓)'). Maintain contrasts finds these targets satisfactory, since the prevoiced /b/ and the plain /p/ are significantly differently from one another, in fact, the contrast is maintained in a very similar way to how it is typically done in Hungarian. Maintain categories is also satisfied, since both the /b/'s with 130 ms prevoicing and the plain /p/'s (15 ms VOT) are quite typical exemplars of their respective

categories. While the 130 ms prevoicing might be longer than the usual amount of prevoicing in word-initial Hungarian /b/'s, it is an exaggerated version of a characteristic cue, which still does not challenge categorization.

The predictions of the two hypotheses is the same for the aspirated /p/'s in *Extr. Asp.* as well. Both hypotheses are somewhat agnostic as to whether convergence happens ('✓/✗'), since long aspiration is a cue that Hungarian rarely uses (if ever), and thus participants might not be able to replicate it consistently throughout the experiment. However, on a representational level, both pressures find a /p/ with 130 ms VOT to be a perfectly valid target. For Maintain contrasts, this comes from the fact that it is sufficiently distinct from the plain /b/ that the model talker contrasts it with. From the perspective of Maintain categories, the aspirated /p/ is a valid target, since, even though such long aspiration would be unusual for a /p/, the exaggeration of positive VOT does not interfere with it being categorized as a /p/.

The two hypotheses differ, however, in their predictions about whether the plain /b/ target in *Extr. Asp.* will be accommodated to by Hungarian speakers. Maintain contrasts predicts accommodation, since while the plain stop with 5 ms VOT is an atypical realization for a Hungarian /b/, it is contrasted with an aspirated /p/ in the model talker's speech, which means the contrast is preserved ('✓' in the top figure). However, Maintain categories is not satisfied, since the /b/ with 5 ms VOT is encroaching on the canonical place of /p/ in the representations of Hungarian speakers (illustrated by the '✗' in the bottom figure).

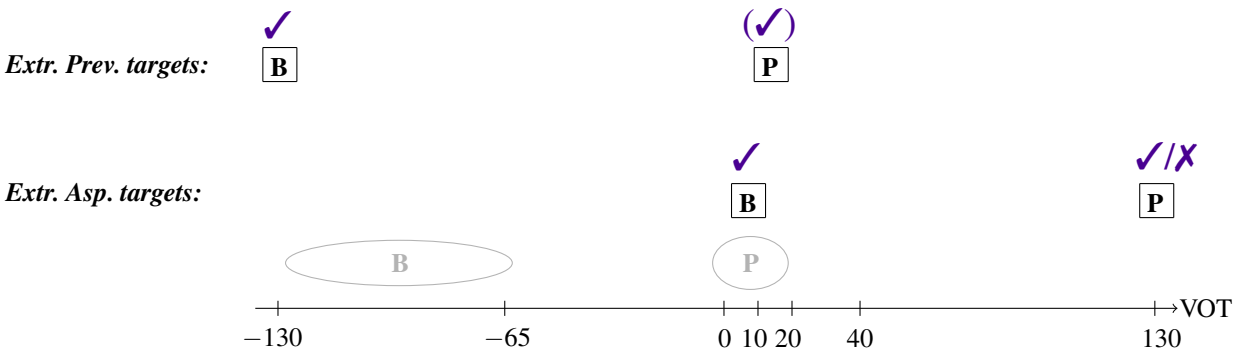

Figure 4.35: Predictions of the maintain contrasts hypothesis for Hungarian
Gray: typical Hungarian values; ✓: hypothesis predicts convergence; ✗: hypothesis predicts no convergence

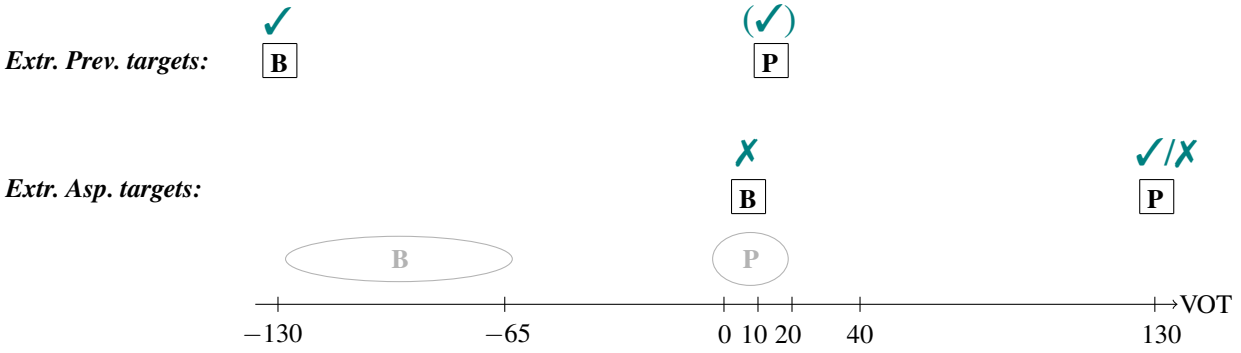

Figure 4.36: Predictions of the maintain categories hypothesis for Hungarian
Gray: typical Hungarian values; ✓: hypothesis predicts convergence; ✗: hypothesis predicts no convergence

The Hungarian results are in line with the predictions of **Maintain categories** rather than with those of Maintain contrasts. While we see convergence for both /p/ and /b/ in *Extr. Prev.*, no convergence was observed for either /p/ or /b/ in *Extr. Asp.* (only some socially mediated divergence). The picture here is much clearer than in the English study: while participants endorsed the *Extr. Prev.* targets, they did not do so with the *Extr. Asp.* targets.

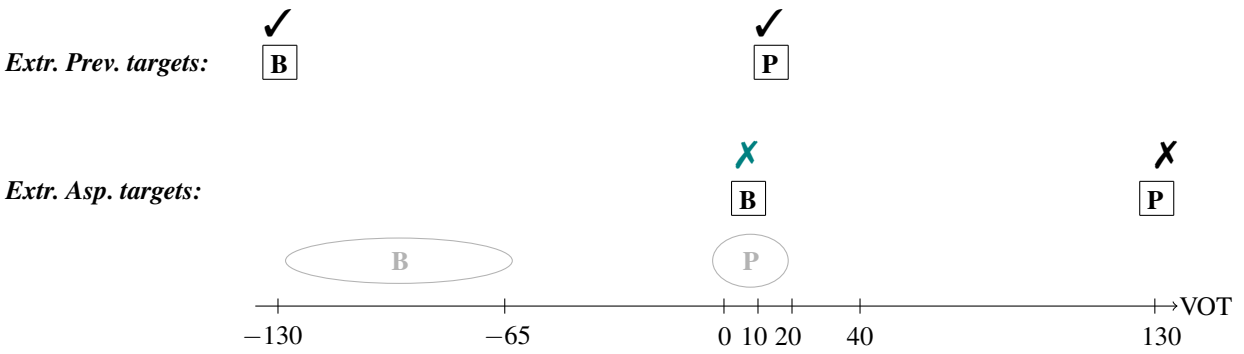

Figure 4.37: Participants' behavior in the Hungarian dataset
Gray: typical English values; ✓: accommodation; ✗: no accommodation

The lack of convergence to *Extr. Asp.* stimuli itself could indicate one of two things. First, it could be reflective of the limitations **Maintain categories** poses on what a valid target is (i.e. it could be considered evidence in favor of the Maintain categories hypothesis). Second, it could be reflective of Hungarian speakers not considering aspiration as a cue to be on the same continuum as prevoicing and short-lag VOT are. Under this second explanation, Hungarian speakers might not have seen a short-lag vs. long-lag contrast as a valid voicing contrast even under Maintain contrasts (since Maintain contrasts might require a distinction along a continuum that runs from prevoicing to no prevoicing).

One must remember that we were left with a similar outcome in the English data as well. Based on the English results we concluded that the lack of convergence in English *Extr. Prev.* is either the result of **Maintain categories** or of English speakers not viewing prevoicing as part of the VOT continuum. Since based on previous results (e.g. Olmstead et al., 2013) we can assume that *hearing* these non-native cues in and of itself is not an issue, we are left with two possibilities. First, it could be the case that **Maintain categories** restricts the phonetic properties of exemplars of a category, and a plain stop is not an adequate target for either an English /p/, or a Hungarian /b/, because the plain stop encroaches on the representations of other segments in both of these

languages. Second, it could indicate that languages view the use of non-native cues (or ranges of cues) as an insufficient way of distinguishing a contrast.

While these two alternatives might be distinct in some respect, they are not all that different from one another in terms of the main question: whether the speaker's pre-existing phonetic knowledge can limit accommodation. The Hungarian and English results suggest that it does: speakers do not imitate certain realizations of categories that are too different from their own preconceived representations of these categories.

4.6.3 Categories moving together

In the Hungarian experiment we see some evidence of accommodation to /p/ being in correlation to that of /b/. This evidence comes from the reading data, and specifically from *Extr. Asp.*, where not much convergence took place and treatment of /b/ was contingent on `Solidarity`. I claimed that this `Solidarity` effect was more divergence stemming from disliking than convergence enhanced by a positive attitude towards the model talker.

The fact that correlation between the treatment of /p/ and /b/ was not found in *Extr. Prev.*, could be explained by either two things. The correlation could exist, but was muddled by the fact that a given participant's productions were likely not equally far away from the model talker's /p/ and her /b/. It is possible that every participant in *Extr. Prev.* converged with the model talker as much as they could, but for instance, some participants' baseline /p/ was much closer to the model talker's /p/ than their baseline /b/ was to the model talker's /b/, while others had their /b/'s more closely aligned with the model talker's. Even if they all converged with the model talker maximally, the extent to which they changed their /p/'s and /b/'s would not be correlated. This would have been especially likely if the model talker's targets were within the range of the participants' baseline productions, which they happened to be in this case. Alternatively, maybe it is only *divergence* of /p/ and /b/ that are related to one another, and that is why it only shows up in the condition where

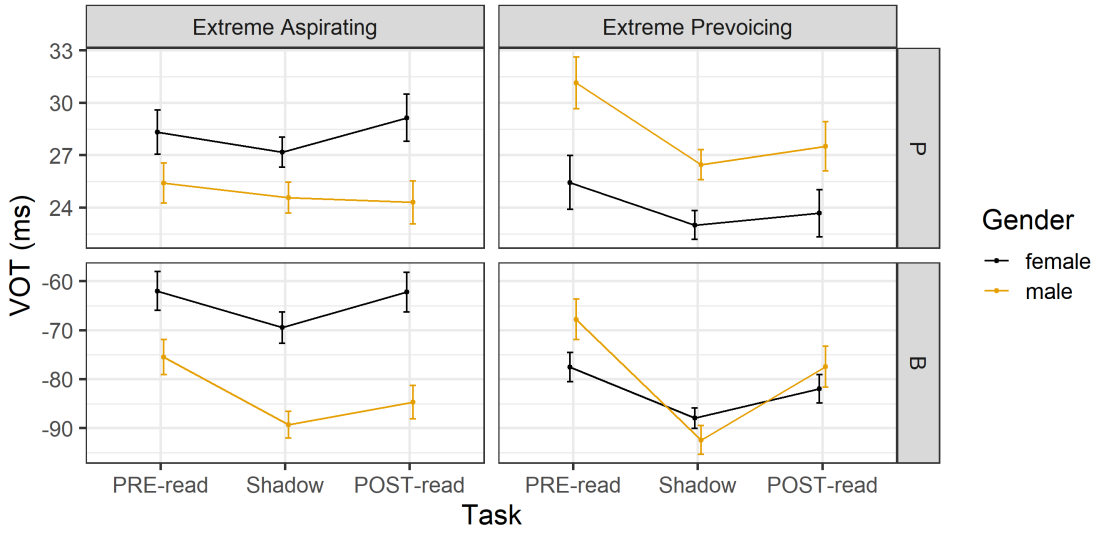

Figure 4.38: All productions from all Hungarian participants averaged by task: PRE-Read (2 reps), Shadowing (6 reps) and POST-read (2 reps)

we find divergence encouraged by disliking. This one study does not provide enough evidence to decide, and this issue must be investigated further.

4.6.4 Task effects

In general, in *Extr. Prev.*, participants' productions during shadowing were often closer to the target than their productions during reading (right-hand side of *Figure 4.38*). However, this was not true in *Extr. Asp.* (right-hand side of *Figure 4.38*). This was true on an individual level too: most participants' shadowed productions in *Extr. Prev.* were closer to the targets than their read productions, but only few participants exhibited the same patterns in *Extr. Asp.* To put it differently, there were very few instances of strictly immediate convergence in *Extr. Asp.*—convergence that could be observed during the exposure task (shadowing) but not once the participant is not directly exposed to the stimuli from the model talker.

This is different from the English dataset, where even in the "un-English-like" condition *Extr. Prev.*, participants' productions were in general closer to the target during shadowing than

during reading (at least for /b/). This could be further support for the English-speaking participants having converged with a prevoiced /b/.

It is worth noting that shadowing trajectories were much flatter in Hungarian than what we saw in English. This meant that people either did not converge at all (*Extr. Asp.*) or converged all at once (*Extr. Prev.*) The lack of convergence suggests that the Hungarian speakers were maybe less inclined to endorse productions that they were uncertain about. The instantaneous convergence rather than gradual convergence indicates that maybe the targets that did induce convergence were in such a range for Hungarian speakers that converging to them did not require that much exposure.

4.6.5 Shift of boundaries

When Hungarian participants were exposed to a contrast unlike their native one (*Extr. Asp.* condition), no measurable shift occurred in perception, mirroring the lack of change in production. This is in contrast with the English dataset, where although exposure to an unnativelike contrast (English *Extr. Prev.* condition) only induced production changes in one sound (/b/) and only during shadowing, there was a measurable perceptual shift on a group-level (even though clearly spear-headed by a handful of participants).

At the same time, there did seem to be an adjustment of boundaries among the Hungarian participants, but in reaction to the more nativelike contrast (*Extr. Prev.* condition). Since this condition should not have required perceptual adjustments from the participants, this might indicate that the 15 ms VOT for /p/, while within the range of typical productions, was not completely "nativelike".

4.6.6 Articulatory fatigue

Whereas in the English data there were multiple examples of V-shaped trajectories, where certain participants started to prevoice more, only to gradually prevoice less and less in the second half

of the task, the counterpart of that was rarely observed in the Hungarian data. The counterpart of the V-shape would be a trajectory, where participants start aspirating more, but gradually revert to their baseline in subsequent repetitions (A-shape). The maximum amount of aspiration for the few participants in the Hungarian dataset who did that tended to remain under 40 ms. Their maxima being so low makes it unlikely that it was articulatory fatigue, and an interpretation of mental or attention fatigue (or boredom) is much more likely.

4.6.7 Gender

Males and females behaved comparably throughout the experiment, and even when not, most exceptions could be explained by distance effects or differences in likeability ratings. For instance, while males seemed to converge more (in amount, not in frequency) than females in the *Extr. Prev.* reading data, males' /p/ baselines also tended to be further from the target at the start, so they needed to compensate more. However, not all effects could be explained that way. For instance, males also tended to converge more in the *Extr. Prev.* /b/ data, even though their baselines were not significantly further away from the target.[2]

Moreover, there seemed to be some gender-effects with respect to likeability that were only observed among males. Namely, for males, there seemed to be a `Superiority` effect in *Extr. Prev.* for /b/'s (-130 ms VOT), and during shadowing, males showed an effect of `Solidarity` with respect to *Extr. Asp.* /p/'s (130 ms VOT), which was nowhere to be found in the female data. This could be explained by multiple things. First, it is possible that males accommodate more in terms of VOT. This would be compatible with Nielsen's (2008) finding that more males in her study converged with her male model talker ($\sim$100 ms /p/ targets) than females did. However, this

[2]Even though it was not significant, *Figure 4.38* suggests a gender difference in the baselines (PRE-Read), which would mean that the same distance effect we see for /p/'s was also the reason for males converging more than females for /b/'s as well (i.e. males simply made up for having to start from further).

explanation is unlikely, since the English experiment in this study has provided evidence in favor of Nielsen's finding being a result of a distance effect (males' VOT's were further away from the model talker, so more of them needed to adjust their VOT production in order to match or get close to the model talker's target). Second, it could be the case that males might be susceptible to likeability-based effects. It is worth noting that likeability-based effects in this study have mostly been observed through divergence correlated with disliking. This would be an interesting finding, since previous studies (on English) have either indicated that females might be more sensitive to socially indexed information than males or found no gender-based difference. Since female model talkers are less commonly used and since we know especially little about how and whether gender interacts with accommodation in Hungarian, the findings of the Hungarian experiment cannot be readily contextualized, and should be investigated further.

4.6.8 Ethnicity

Since the Hungarian participant pool was ethnically homogeneous, ethnicity-related effects could not be observed in the dataset.

4.6.9 Likeability

There were two effects of likeability in the Hungarian dataset, and they were only observed during the reading task. First, in *Extr. Prev.*, where the targets were a plain /p/ and a prevoiced /b/, males diverged from the model talker's prevoiced /b/ if they rated her low on `Superiority`. This the more surprising effect of the two, and its robustness should be tested in further studies.

The other effect was in the *Extr. Asp.* condition. This one is easier to explain, since this was the condition that sounded *unlike* Hungarian typically does, so whether a participant was willing to endorse these targets could have conceivably been contingent on their attitude towards the model talker. Participants tended to diverge more from the model talker's plain /b/'s during the reading

task if they rated her low on Solidarity-related scales. This effect was quite similar to the effect in the English reading data, where participants in *Extr. Prev.* who rated the English-speaking model talker low on Solidarity-related scales tended to diverge from her plain /p/ (*Figure 3.18*). Both the English *Extr. Prev.* effect and the Hungarian *Extr. Asp.* effect are an active divergence from representationally anomalous stimuli in the case of disliking.

Based on the two experiments together, we can say that the component of likeability that was likely responsible for previously recorded effects on accommodation was either Solidarity (*friendly–unfriendly*, *honest–dishonest*, and *polite–rude*) or Superiority (*organized–unorganized*, *high status–low status*, and *intelligent—unintelligent*) or a combination of the two, but Dynamism-related characteristics (*talkative–shy*, *confident–unsure*, and *energetic–lazy*) likely had little to do with previous results.

The effects of likeability were quite messy in this study, which could have been partially exacerbated by the fact that the two ends of the scales were often flipped (following Bauman, 2013), which could have led some participants to misuse certain scales. Further research (especially with a larger set of model talkers) is necessary to corroborate these results.

4.6.10 Attractiveness

While data was recorded on attractiveness as well, both before and after the task, Hungarian responses were also subject to the same interpretational ambiguity that we saw in the English data. Therefore, these data could not be analyzed.

Chapter 5: Modeling contrast preservation in accommodation

In the previous two chapters I have investigated how phonological representations affect accommodation. Results from the experiments in *Chapters 3 & 4* indicate that the phonetic properties of a token relative to its category can limit accommodation to it. In this chapter, I am going to look at the flip-side of this question, namely, how representations themselves are impacted in the course of accommodation. I am first going to discuss what insights we can glean from previous models of accommodation, speech perception, and production (*Section 5.1*). I am going to outline a computational model that can simulate accommodation and which will serve as our basic model (*Section 5.2*). Finally, I am going to discuss what modifications are necessary to this basic model in light of the experimental results and demonstrate its abilities with simulations (*Section 5.3*).

5.1 Background

While there has been ample experimental research on accommodation itself, much less attention has been devoted to modeling it. While formally explicit models of accommodation are exceedingly rare in the literature, two key features of models have been identified through experimental results: episodic memory and activation. In the following, I am going to review these two concepts. I am also going to discuss the concept of repertoires, and why I will eventually digress from them. Finally, I

am going to talk about the concept of resonance, which makes some interesting predictions for the experiments in this study and thus could prove useful during the computational implementation of accommodation.

5.1.1 Requirements for models

It has been argued that in order to model accommodation processes, we need to assume an episodic or exemplar representation of phonological categories. This has been the approach taken by all accommodation modeling work I am aware of (e.g. Johnson, 1997; Kaschak and Glenberg, 2004; Wagner et al., 2006; Tobin, 2015; Tobin et al., 2017; Wang, 2019). An exemplar model (Bybee, 2001; Pierrehumbert, 2001, 2006) has two key advantages. First, accommodation itself is a testimony to speakers being sensitive to the phonetic detail of their interlocutor's speech, which means that the input must be represented in a similarly detailed way. Such detailed representations are one of the main tenets of exemplar theory: memory traces of input tokens contain all of the tokens' phonetic details as well as social information about the interlocutor (to the extent that the speaker's memory allows it). Moreover, exemplar models are *dynamic*, i.e. they can be continuously updated with new information. This is necessary for modeling accommodation, a process that involves the speaker reacting to stimuli and changing their production as a result of it.

Another necessary component for accommodation models is activation (e.g. Chartrand and Bargh, 1999; Johnson, 1997; Goldinger and Azuma, 2003, 2004; Babel, 2009; Tobin, 2015; Tobin et al., 2017). Activation is a neurological concept, which comes from the (visual and auditory) perception literature. When an input token is perceived, it activates parts of the pre-existing representation that are most similar to it. In the perception literature the resulting activation pattern is used for the categorization of this input token, but it can also be useful for modeling how the speaker produces tokens during accommodation (see e.g. Tobin, 2015). Activation has been a key part of the formally explicit model of the perception-production loop by Roon and Gafos (2013). Their

model is designed to explain reaction time effects in priming studies through tracing activation peaks regarding the movement of articulators in quite a detailed way (e.g. various areas of the tongue).

5.1.2 Production repertoires in accommodation

It has also been argued that over the course of accommodation speakers only produce tokens from their "repertoire" (ranges of production they are already proficient in). In her word-shadowing experiment, Babel (2009) observes two different trajectories of formant accommodation to a male model talker. Female participants, whose formants were further away from the male model talker even after normalization, approximated the model talker more and more, indicating that accommodation happened incrementally throughout several trials. At the same time, male participants accommodated after the first block of exposure and then stayed at that level of accommodation for the rest of the experiment—i.e. their productions did not get any more similar to the model talker in subsequent trials. She argues that the reason for the differing patterns is the difference between the production repertoires of male and female speakers. While male speakers might have more production targets similar to the male model talker in their repertoires, which are thus more easily activated, female participants likely have fewer, and increased activation throughout trials helps to "highlight" these target exemplars (Babel, 2009:134).

However, assuming that input tokens the speaker perceives and outputs that they produce are stored separately can have some undesirable consequences when it comes to modeling long-term sound change. Babel (2009) theorizes that more permanent sound change happens via tokens from a different word affecting the production of the sound in other, non-matching environments. However, assuming that new productions are always chosen from the speaker's repertoire set means that the speaker will never produce any truly novel tokens for the given word throughout their lifetime.

This is in conflict with the idea that accommodation is in fact a crucial step in large scale language change (also see *Section 2.1.4*).

Moreover, this could rule out certain instances of second dialect acquisition (see review in Siegel, 2010) as well as longitudinal changes within one's idiolect (Harrington et al., 2000; Harrington, 2006, 2007; Harrington et al., 2007; Sankoff and Blondeau, 2007; MacKenzie, 2017). Later in the modeling chapter (in *Section 5.2*) I am going to show that the same results can be achieved even if all tokens are stored together, but the speaker's own productions simply vastly outnumber other tokens. This allows us to leave the option for long-term change open.

5.1.3 Resonance and typicality

The findings from the experiments in this dissertation indicate that accommodation is impeded when lexical information about a given stimulus conflicts with its phonetic profile. For instance, English speakers did not converge to stimuli in *Extr. Asp.*, such as *pooling* /ˈpuːlɪŋ/ with 15 ms VOT on the initial stop. Lexical information here suggests the initial stop is a /p/ (since */ˈbuːlɪŋ/ is not a word of English), but its phonetic profile (short-lag VOT) suggests that it is a /b/. This incongruity between lexical and phonetic information has been predicted to interfere with categorization, via the concept of resonance.

Resonance was introduced by Adaptive Resonance Theory (ART) (Grossberg, 1980, 1999, 2003), which was primarily developed for categorization during auditory and visual perception and relies heavily on neurological insights. ART claims that expectations and pre-learned stereotypes and concepts can influence speech perception (Grossberg, 2003). This has been corroborated by the importance of language-specific cues in speech perception (Wagner et al., 2006).

This is modeled through the concept of resonance in visual and audio object recognition. Resonance happens when bottom-up information (i.e. information about the object itself) matches top-down information (i.e. expectations about the object based on circumstantial information). In

case there is no match the token is ignored. ART predicts that stimuli whose lexical and phonetic information are in conflict will not cause a perceptual shift, nor will they be accommodated to. If we see no accommodation in cases where there is a mismatch between lexical and phonetic information (i.e. *Extr. Prev.* in English and *Extr. Asp.* in Hungarian), the concept of resonance can be useful over the course of modeling as well.

5.2 The basic model

In this section I am going to describe the basic model through which I am going to illustrate one possibility for how **Maintain categories** could be implemented. This model is an exemplar model. The exemplar framework (Johnson, 1997; Bybee, 1999; Pierrehumbert, 2001) is particularly suited to model accommodation phenomena, because of its ability to provide memory-based models, where previous memories of the speaker are activated by perceiving the speech of their interlocutor (for a syntactic example, see Kaschak and Glenberg, 2004). In line with previous work, I will also assume that whenever a perceived token is added to the representation, it *activates* a number of similar tokens that are already part of the speaker's representation (Goldinger and Azuma, 2003; Grossberg, 2003; Wagner et al., 2006).

5.2.1 The algorithm

The model's workings can be broken down into discrete steps following the pseudo-code in Algorithm 1 below. First, we need to set up the *Speaker set* and *Interlocutor Set*, which are essentially the phonological representations of the speaker and the interlocutor, respectively. The Speaker set and the Interlocutor set are set up and populated based on separate distributional specifications for each relevant phonetic dimension (Lines 1–4). At this point, these sets are the implementation of the speaker and model talker's representations of the same phonological category. These sets contain

every token the speaker or interlocutor has encountered, respectively (regardless of whether it was produced by them or perceived by them).

In its current form tokens in a given *Speaker Set* can be interpreted on various levels: as instances of a stop in the same word (e.g. tokens of /k/ in the word /ˈkin/), instances of the same sound category (e.g. tokens of /k/), or either of these restricted by speaker (e.g. various productions of /k/ in /ˈkin/ from speaker X). While the model itself is agnostic to this decision, I will choose to use the *Speaker Set* to reflect an entire sound category, because there is evidence that speakers exhibit phoneme- and feature-level generalizations in accommodation (among others, see McQueen et al., 2006; Nielsen, 2008; Babel, 2009).[1]

After the two sets are initiated, a resting activation pattern is generated in the Speaker Set which reflects that the speaker is more susceptible to some types of production than others (Lines 5–6). An initial token, t_0, is created. This token has the median values of the Speaker Set for every dimension, representing an "average" token from the speaker. This token activates some or all tokens in the speaker's own representation (i.e. the Speaker Set). There are multiple ways of implementing this, which will be described in detail in *Section 5.2.3*. In the simulation in *Section 5.2.4* this is implemented as an increase in the activation level of the n number of tokens in the Speaker set closest to the input token t_0. This increase is proportionate to the distance between the given token in the Speaker set and the input token. Since t_0 was chosen to be the median value, this means that at the start of the interaction, these two steps will result in the activation of the speaker's most commonly perceived tokens, which can be considered roughly equivalent to resting activation. The *Speaker Set* and the *Interlocutor Set* at this point represent the mental state of the speaker and the interlocutor at the start of the interaction—before anything has been uttered.

[1]These tokens can come from semantically and morphologically unrelated words, and while the model does not encode semantic or morphological similarity, these could be added to the model through an n-dimensional representation for semantic and morphological representation that is somewhat separate from the phonological representations.

Algorithm 1 The algorithm of the basic model

1: SET UP the *Speaker Set* (S_{Sp})

2: POPULATE the *Speaker Set* with l number of tokens (t), based on the D_S distribution, defined along d dimensions (*Speaker's initial state*) with an activation level

▷ *For example: if dimension d is duration, t is of form (duration, activation level), such as t_1 = {duration: 1.05, activation level: 0})*

3: SET UP *Interlocutor Set* (S_{Int})

4: POPULATE the *Interlocutor Set* based on elements of D_I (*Interlocutor Distribution*)

5: BASELINE PRODUCTION Create a token t_0 with the median values of the *Speaker Set*

6: RESTING ACTIVATION Increase the activation level of some or all tokens in the *Speaker Set* based on the token's similarity to t_0 (defined by activation function A)

7: **for** i iterations **do**:

8: CHOOSE RANDOM TOKEN from S_{Int}, $t_{l+(i*2)-1}$[2]

9: ACTIVATE(int) Increase the activation level of some or all tokens in the *Speaker Set* based on the token's similarity to $t_{l+(i*2)-1}$ (defined by activation function A)

10: INCORPORATE(int) $t_{l+(i*2)-1}$ to the *Speaker Set* (S_{Sp})

11: PRODUCE a new speaker token ($t_{l+(i*2)}$) that's the weighted average of the activated tokens

12: ACTIVATE(sp) Increase the activation level of some or all tokens in the *Speaker Set* based on the token's similarity to speaker token $t_{l+(i*2)-1}$ (same mechanism as in Line 9)

13: INCORPORATE(sp) ($t_{l+(i*2)}$) to the *Speaker Set* (S_{Sp})

14: DEACTIVATE some tokens in the *Speaker Set* (S_{Sp})

15: **end for**

From here on, the model proceeds in iterations (Lines 7–15) and goes through the same set of steps in each iteration. First, it chooses a token from the *Interlocutor Set* (Line 8), which represents

the interlocutor uttering a token. In a natural interaction this token could be truly randomly chosen; in the experiments in the previous chapter this was fixed to be a certain value each time (e.g. the target /p/ in *Extr. Asp.* had 130 ms VOT each time). This token is then perceived by the speaker, which is reflected by the token activating the exemplars in the speaker's representation which are most similar to it (Line 9). This is done in the exact same way as earlier in the case of Resting activation (Line 6): it increases the activation level of the closest n number of tokens by an amount that is proportional to the given token's distance from $t_{l+(i*2)-1}$. More details on alternatives will be presented in *Section 5.2.4*. Following this step, the speaker commits this token to memory: the recently uttered interlocutor token is placed in the *Speaker Set* (Line 10).

The order of these two steps (9 & 10) is crucial, because if the token is incorporated before the activation step, then because the distance between the new token and itself is 0 in any case, it will give itself extremely high activation levels. This would predict that any speaker can accommodate to any perceived token, even extreme ones, nearly perfectly upon the first hearing, which is not what previous studies suggest.

Then the speaker produces a token in return: a new token is created based on the weighted average of the activated tokens, with randomized noise (Line 11). The speaker's new token also activates some of the representation and becomes part of their representation: this token is also added to the *Speaker set* (Lines 12 and 13). The final step in every iteration is deactivation. This step could be optional and variants of this function are also discussed in Section 5.2.3. In the simulation below it will be implemented as each token's activation level being slightly decreased by the same amount.

5.2.2 Parameters

The model outlined above has several parameters that can be adjusted independently. First, l is the number of tokens initially populating the *Speaker Set* and *Interlocutor Set*. The more tokens there

are, the more systematic variability we can see in the placement of the tokens themselves. Also, in this model, the bigger the *Speaker Set* is, in particular, the more "inertia" the speaker's productions will have. This is because at the end of the day, shadowed productions are a weighted average of activated tokens. Because of the resting activation step before the "interaction" starts, many of these typical tokens will already be highly activated. Thus, the speaker's more typical productions will outweigh the outliers in the production process—until a critical mass of any given kind of outlier is collected. Thus, the more exemplars the speaker has, the smaller the effect of extreme input tokens will be—i.e. the speaker will not be steered away from their usual productions that easily.

This is corroborated by frequency effects observed in the accommodation literature, where words of which the speaker has more exemplars (i.e. high frequency words) are less susceptible to accommodation than ones that the speaker rarely uses and encounters (i.e. low frequency words; Goldinger, 1998 and on). Another form of evidence for this can be found in the second dialect acquisition literature, which suggests that young children pick up new dialects more easily than adults.

The phonetic dimensions themselves can also be externally specified. This means that the model is not only capable of simulating VOT accommodation, but also of simulating accommodation along any phonetic dimension. The model does not limit the number of dimensions either, since the model operates with Euclidean distance, which can be calculated between any two n-dimensional points. Therefore, the number of dimensions can also be set arbitrarily, i.e. one dimension for VOT and two dimensions (e.g. F1 and F2) for vowel formants or even more. However, one must be cautious when using multiple dimensions, and make sure that these dimensions are related in some sense. Recent research indicates that accommodation behavior along various dimensions is not necessarily correlated (e.g. Pardo et al., 2013a, 2017), which would mean that, for instance, predictions based on a *Speaker Set* defined along VOT and intensity might not be very meaningful.

The distribution of each set has to be described with a mean and a standard deviation, which can differ between dimensions and sets. For instance, the speaker set might have a mean of 10 ms and a standard deviation of 0.8 for VOT, while we can choose an interlocutor with a mean VOT of 65 ms and a standard deviation of 2. This property allows for simulations of interactions where the speaker and interlocutor have very different habitual productions. Standard deviation describes how "spread out" the distribution is. By increasing the standard deviation of a set, we get distributions where productions are more different from one another. A speaker with a bigger standard deviation of encountered tokens can be better at accommodation, because a random new perceived token will have a bigger chance of falling near some exemplar the speaker already encountered or produced. This prediction is not surprising, considering findings from speech attunement. Laturnus (2017) finds lifetime experience listening to non-native speech positively correlates with comprehension accuracy of previously unheard accented talkers.

While this prediction is made on the level of the individual—i.e. contrasting individuals with a more diverse set of tokens of a given category with individuals with a less varied exemplar pool—it is worth considering if it also applies to categories. That is, whether categories with more variation are more easily accommodated (if speakers encounter this variation to the same extent). It is plausible that such varied categories might be more often accommodated to and therefore might even be more likely to change over time. However, since the theory presented here is only concerned with the propagation of change and not its actuation, this prediction is dependent on whether change is actuated to equivalently often in various categories, which lies outside of the scope of this work to determine.

It is also an interesting question whether this generalizes to categories with allophonic variation (whether categories with allophonic variation are especially susceptible to change over time), as allophonic variation provides a constant source of variation of disparate forms. This is contingent on how important of a role phonological context plays in the representation of sound

categories. While at one extreme some argue that exemplar representations are word-specific (in which case allophones, which are necessarily encountered in different environments, would not be able to influence each other's production), others claim that forms from other contexts can influence the way a new instance of a category is produced. This work remains agnostic on this divide. However, if the previous two stipulations hold (change is actuated at a roughly equal rate across categories and tokens of a category are stored and considered together irrespective of their potentially different phonological environments) then this model predicts that categories with more varied realizations (or categories with allophonic variation) are more likely to be accommodated to and more likely to change over time than categories whose realization is more consistent.

The number of iterations can also be independently chosen, reflecting how long the interaction was. More exposure (more iterations, a longer interaction) will result in more accommodation in the model up to a certain point. Interlocutor tokens encountered in previous iterations are likely going to fall close to future tokens from the same interlocutor, thereby forming a stronger and stronger basis for target productions approximating the interlocutor's speech. This is in line with the finding that the amount of exposure correlates with the amount of accommodation in case there is substantial difference between the speaker's and the interlocutor's speech (Goldinger, 1998).

5.2.3 Activation and deactivation functions

Finally, there are also multiple options for how activation and deactivation might happen. These options can be combined freely with one another, and they do not yield drastically different results for the purposes of this work. A function can also be included to simulate the forgetting of tokens by randomly deleting a pre-specified number of tokens in each iteration, but this function is not used during the course of this project.

The first activation function is quite simple: it activates the n tokens that are closest to the newly perceived token, where n is a number chosen by the user. The activation level of these n

tokens are incremented by a fixed amount, which can also be chosen freely. The number of tokens activated in each iteration (n) should be chosen to be smaller than l (the number of tokens the *Speaker Set* has at the start). If this criterion is not met (n is greater or equal to l), all tokens will be activated by every input token. As a result, the activation pattern will be the same (all tokens will be equally activated) irrespective of the phonetic properties of the input token. Thus, if $n \geq l$, we are essentially left with an activation function that does not implement activation.

The second option is to increase the activation of a given token proportionately to its distance from the new input token. During this process, each token's activation level is incremented by the sigmoid of the reciprocal of its distance from the new token (activation $= \sigma$ (1/distance)).[3] This results in closer tokens having a larger boost in their activation levels while the activation level of far away tokens only increases a little.

The third function is a hybrid of the previous two: the activation level of the closest n tokens increases by an amount that is the sigmoid of the reciprocal of its distance from the new token (activation $= \sigma$ (1/distance)). On the one hand, this approach reflects the intuition that an input will activate tokens that are similar to it more so than ones that are less similar. On the other hand, by only activating the closest n tokens it presents a cut-off point: beyond a certain amount of dissimilarity, activation will not happen. This third variant of the activation function is used in all of the simulations in this chapter.

It must be noted that none of these three activation functions interact with the size of the *Speaker Set* (l) aside from the consideration that l should be larger than n (the number of tokens that are activated in each iteration). However, the way n itself is chosen can drastically alter the results of the simulations. The choice of n can have similar effects on the model's tendency to accommodate as the choice of l, but through a different mechanism. As I discussed at the beginning

[3]The sigmoid function is a monotonic function that maps a negative x to a value between -1 and 0 and a positive x to a number between 0 and 1. It shrinks distances between numbers of very different magnitudes.

of *Section 5.2.2*, the bigger l is, the more robust the speaker's production patterns are—i.e. the less swayed they are by extreme input tokens. This is true of n as well. The smaller n is, the more impact the input tokens can have on the speaker's productions—the more likely the speaker is to accommodate. While the sum of increases in activation level itself is not limited (i.e. there is no fixed amount of activation that needs to be doled out between the n tokens), the smaller n is, the fewer tokens get activated, and at the end of the day, it is only tokens with *some* amount of activation that can sway productions.

This effect becomes more prominent if we consider later iterations, when the speaker has encountered multiple similar tokens from the same atypical interlocutor. Since these are extremes, initially they are alone in a sparse area of the Speaker Set (that is what makes them atypical). However, as the speaker encounters more and more tokens from the same interlocutor, this previously sparse area of their phonetic space will slowly become populated with tokens from this one interlocutor. While these atypical tokens will still likely be vastly outnumbered by the speaker's own tokens, if it is chosen to be a relatively small number, e.g. 20, then at the 21$^{\text{st}}$ iteration, the n number of tokens closest to the input will exclusively be from this one atypical interlocutor. Since these tokens will be much more similar to the input (as they come from the same interlocutor) than anything the model talker so far has produced, the amount by which their activation levels will be incremented will be high as well. Therefore, within a few iterations the speaker will match the model talker's productions. Since such behavior is rarely attested (everyone's productions show a certain level of robustness and consistency, even across conversations) n should be chosen to be relatively high (compared to the number of iterations run with the model).

The two deactivation functions correspond to a fixed and a flexible approach. They both decrease the activation level of every token by the same amount, but that amount is what differentiates them. The fixed deactivation function decreases the activation level of all tokens by a *pre-specified* amount in every iteration. The flexible function does not need a pre-specified amount. In every

iteration it decreases the activation level of all tokens by the lowest non-zero activation level in the set—i.e. by the minimal non-zero amount of activation a token has in the set.

5.2.4 Accounting for previous results and extensions

This section reviews how the model outlined above can simulate effects and phenomena that we are already aware of from previous literature. The phenomena it deals with are distance effects, inter-speaker variability and social factors in accommodation.

Distance effects

This model is a single-pool model, i.e. all tokens that the speaker encounters are stored in a single pool of exemplars and there is no set of distinguished tokens in the speaker's representation that would correspond to their "repertoire". Previous studies have suggested that speakers only use their repertoire when accommodating (e.g. Babel, 2010). I am going to call these models dual-pool models, where stored exemplars are separated into two pools: the speaker's repertoire and a pool with more peripheral tokens from their interlocutors.[4] I have argued in *Chapter 2* that such a model would predict that participants' productions will not change through interactions. This would be in conflict with the Change by Accommodation theory, as it does not allow for participants changing their productions through interactions (i.e. second dialect acquisition or the propagation of language change). Because of these concerns, the model adopted here will be a *single-pool model*.

The main reason for Babel to suggest a dual-pool model is distance effects. There has been evidence suggesting that participants whose baseline is closer to the target show different

[4]Potentially tokens could be further separated by interlocutor (into as many sets as the speaker had through their lifetime), making such models multi-pool models. However, the key distinction for the purposes of this work is between single-pool models and all others, i.e. the differences between dual- and multi-pool models is not relevant. Therefore, I will disregard multi-pool models in the following.

accommodation behavior than those being further away. Most recently, Hosseini-Kivanani et al. (2019) argued that some of the differences that seem to be socially structured might be due to participants tending to diverge when they are close to the target to begin with. Moreover, larger effect sizes of convergence can be observed, if the baseline of the participant is further from the target. That is, if a given group's baselines are systematically closer to the target provided by the model talker, that group will have less of an opportunity to demonstrate convergence—there is less distance for them to cover. This has also been found in Szabó (2019), where there initially seemed to be a relationship between the chronological age of the speaker and speech rate accommodation. While older speakers had a slower speech rate initially, they were found to converge more with faster interlocutors than younger speakers were, who spoke faster to begin with. Thus, the chronological age effect was found to be a result of older speakers compensating for their slower habitual speech rate.

Babel also found something similar in her work. In her study of vowel formants with a male model talker she found that male participants, whose F1 and F2 were closer to the model talker's to begin with, converged to the model talker "all at once" (i.e. they converged with the model talker in the first repetition and then maintained those productions throughout the rest of the task, but did not get even closer). At the same time, females, whose baselines were further away from the model talker, converged with him gradually. A similar effect has also been borne out in the English experiment, in the shadowing task of the *Extr. Asp.* condition. Females tended to converge with the model talker's plain /b/ all at once, while males converged more gradually.

Because of distance effects, Babel (2009) suggests that a speaker's productions during accommodation must be limited to their *repertoires*. I will call a model like that a dual-pool model because the speaker's representation contains two different pools of tokens (the full set of tokens they encounter and the tokens they can reliably produce, i.e. their repertoire). The model

presented above is a single-pool model, i.e. there is no set of distinguished tokens in the speaker's representation that would correspond to their "repertoire".

What differentiates the single-pool model from a dual-pool model is Line 10 of the algorithm—adding the perceived token to the *Speaker set*. In a single-pool model, *Speaker set* constitutes the entirety of a speaker's representation, and since memory traces of a perceived token can influence representations, the interlocutor's token is added to it automatically. In contrast, in a dual-pool model, the *Speaker set* is a distinguished part of the representation, but probably not the only part since presumably perceived tokens are also stored somewhere. In a dual pool model the *Speaker set* is the part responsible for establishing acoustic targets for new productions (i.e. "production repertoire"), and only produced tokens are added to it, since simply hearing a token will not mean it is now part of the speaker's repertoire. In this model, perceived tokens could still activate some of the tokens that are phonetically similar to them, and thus influence production via the activation pattern they trigger (see Line 11), but they aren't added to the *Speaker set* directly.

I argue that not only do dual-pool models limit accommodation in a way that is not compatible with the Change-via-accommodation theory, but that other models that do not distinguish a specific repertoire can also account for distance effects. In a dual-pool model repeated exposure to a certain kind of token cannot add up over time. Thus, a dual-pool model predicts that participants' productions will not change through even repeated interactions. This would be in conflict with the theory of Change-via-accommodation, as it does not allow for participants changing their productions through interactions (i.e. second dialect acquisition or the propagation of language change).

Simulating distance effects

In the following I briefly demonstrate the results of a short simulation. Four models were generated, a "female" single-pool model, a "male" single-pool model, a "female" dual-pool model, and a

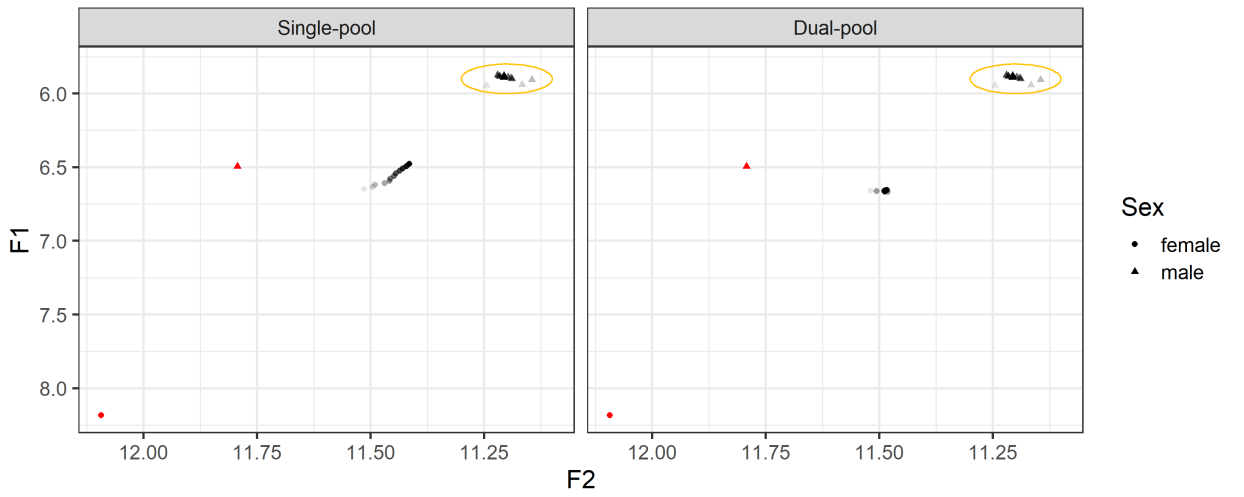

Figure 5.1: Comparing F1 and F2 accommodation simulations with a single-pool (left) and a dual-pool model (right); Red: initial production, Black: subsequent productions; Yellow oval: Model talker's distribution

"male" dual pool model, with 150,000 front low vowel tokens. The four models were exposed to 60 iterations of a randomly chosen input token. The male and female distributions were based on personal estimates of the pre-exposure reading values of participants from Babel (2009), and the input tokens were generated based on the normalized F1 and F2 distribution of the experiment's model talker.[5] *Figure 5.1* shows the results. The model talker's productions were chosen from within the yellow oval. Shapes in the plot represent "gender" (dots for the "female" models, triangles for the "male" ones), and the opacity of the shapes represents how late during the simulation the token was uttered (i.e. fainter symbols represent earlier iterations, darker ones represent later ones). The original mean of both the female and the male are plotted in red of the respective shape. Subsequent productions are plotted in black. For the sake of visual clarity, only every fifth production was plotted.

The simulations with the single-pool models (left-hand side of the plot) match the observations from Babel (2009). For the "female" single-pool model, accommodation proceeds slowly

[5]Male: speaker F1 mean=6.4993, standard deviation=0.5, F2 mean=11.8038, sd=0.5; female speaker: F1 mean=8.1909, sd=0.5, F2 mean=12.0984, sd=0.5; model talker F1 mean=5.9014, sd=0.1, F2 mean=11.2013 sd=0.1.

but surely throughout all iterations. It produces constantly decreasing values for both F1 and F2 (getting closer and closer to the interlocutor's values), with the last value still not matching the interlocutor's average (last, 60. "female" production: F1=6.4768, F2=11.4151). The "male" single-pool model reaches the interlocutor token's average faster than the "female" model. The "male" model produces F1$\sim$5.9, F2$\sim$11.2 tokens from the second iteration on, with the exception of two further F1 $>$ 5.95 values. This demonstrates that while the accommodation in case of a "male" model is nearly instantaneous, for the "female" model it takes a longer time, just like the experimental results indicated.

For the dual-pool model, after a certain number of iterations the algorithm cannot approach the interlocutor target any more. This is an instance of the limitation of the repertoire—the model cannot produce entirely new variants it never produced before. The "male" model actually starts diverging from the interlocutor after having met the mean of the "interlocutor". This is because there is a certain amount of variation in the Interlocutor set, and the model likely encountered more peripheral tokens in later iterations. Its productions are very similar to the "single-pool" male model. The "female" dual-pool model barely shows any graduality in its accommodation. It does not get as close to the model talker as the single-pool model does (it finishes at F1=6.6539, F2=11.4837 on the sixtieth iteration), and it does most of the accommodation instantaneously (early on). Interestingly enough since the dual-pool model not only seems to hit a wall, but hits it relatively early on in the simulation, it seems like if anything, the single-pool model simulates results from Babel (2009) better than the dual-pool model does with its repertoire.

The single-pool model can thus be equipotent to a dual-pool model in terms of distance effects (if not better). All-at-once accommodation is expected in cases where the input from the interlocutor is close to the speaker's baseline. The speaker's typical productions form the densest area of their representation. This is a stipulation that is based on two arguments. First, the speaker perceives their own production the most, which in and of itself allows for some consistency. Second,

each speaker's production is a product of their environment, therefore the peers and community the speaker acquired language in must also resemble their own productions, and these likely outnumber or outweigh production from others.

When the newly perceived token falls close to the speaker's own productions, it will fall into this densely populated area of the perception-production pool. Thus, it will activate a fairly large number of exemplars that are all highly similar to it. Since the speaker's next production will be the weighted average of their activated tokens, the new production will also be quite similar to the perceived token. Therefore, it necessarily follows that accommodation to close-to-baseline input will be instantaneous.

Likewise, gradual accommodation patterns also follow from this mechanism. When the newly perceived token differs from the speaker's normal production, the new token will fall in a more sparsely populated area of the perception-production pool, which rarely is activated, whereas the denser parts of the pool will contain tokens similar to the speaker's production, which are activated most often. Therefore the weighted average of the activated tokens will fall somewhere in-between the new token and the most frequently activated tokens, which resemble the speaker's production. With every round of exposure to similar stimulus, new tokens are gained that increase activation in the sparsely populated area of the pool where the first new token fell. Thus the average, which provides the next production target, will move closer and closer to the new tokens.

Inter-speaker variability

We can also simulate the natural inter-speaker variability in terms of how much speakers accommodate (Chartrand and Bargh, 1999). A given speaker will accommodate less if the new input's influence is outweighed by the speakers' own tokens—i.e. tokens in the densest area of the representation, which can be implemented in two ways, even without tagging the speaker's own tokens in a special way.

The first option is to give the speaker's tokens higher activation levels either manually when the Speaker Set is populated or by making the model "produce" some typical tokens at the start.[6] The former is a less preferred option, since the process of populating the speaker set is a random sample from a distribution, and as such, it also generates some outliers, which will have the same starting activation as tokens which are more typical (closer to the mean). The latter implementation involves high activation of tokens in the densest area (the "middle") of the speaker's representation. This can be induced by producing one or more tokens at the beginning of the simulations before the simulated interaction between speaker and interlocutor starts. Without outside stimulus, the speakers' productions reflect the distribution of their representation, and thus these tokens will further activate tokens in denser areas of the representation when they are incorporated. These two approaches essentially are the same, they increase the average activation level of the speaker's tokens before the interaction starts.

The second option involves the number of activated tokens per iteration (n). As described before, the number of tokens activated in each round (n) can also make the model more or less likely to accommodate. The smaller n is, the more rapidly the model starts approximating the inputs it receives, which was discussed in *Section 5.2.3*.

Neither of these two solutions (activating tokens before the interaction, or increasing n) can *prevent* accommodation from happening, they merely stall it. If enough iterations of exposure happen, models will slowly start to converge with the input. However, convergence can be slowed down or postponed by even thousands of instances of exposure. Eventually, we would see signs of accommodation even in these cases, but since the studies investigating accommodation (like

[6]It must be noted that human participants (coming from other interactions) would not need to actually produce any tokens before interaction. We could assume that humans would have some resting activation in the densest areas of their representation. This step of "production" before the interaction only serves to compensate for the fact that our computationally generated model does not come with this feature.

shadowing experiments) contain a few hundred instances of exposure at most, stalled accommodation and no accommodation are indistinguishable within these experiments. Thus, tweaking these parameters can be used to simulate the inter-personal differences we see in experiments with the caveat that eventual convergence is predicted in each case. While this prediction is hard to falsify, evidence for it could be gathered in second-language acquisition studies, where researchers have an opportunity to observe both short- and long-term accommodation phenomena.

Social factors

While social dimensions are not integrated into the basic model, basic exemplar theory was always envisioned to encode social tagging on each exemplar (Bybee, 2001; Pierrehumbert, 2001, 2006). Under this view each exemplar is tagged for social variables (such as age, gender, sex and race of the person who the token comes from). In theory this could be done in a categorical way, as key-value pairs in the dictionary that each token is represented as in the *Speaker Set* and *Interlocutor Set*. However, it might be better to represent each of these variables on a scale (of 0 to 1, for instance), which can be used to compute the difference or the relationship between the speaker and interlocutor.

As an extension of Giles et al.'s (1991) Communication Accommodation Theory (CAT), one could assume that the degree of accommodation is fundamentally determined by the relation the speaker seeks with the interlocutor. If the speaker wishes to have a closer relationship with the interlocutor, then they will accommodate more than if they were not seeking a closer relationship with them. I theorize that all relevant social percepts of the speaker about the interlocutor translate into an attitude measure. This measure could be interpreted as "(dis)liking" on the part of the speaker towards the interlocutor, but can potentially also incorporate more complex attitudes such as social biases. This attitude could be expressed on a numeric scale (between 0 and 1), which

then could be used as a constant multiplier in the chosen activation function (see Section 5.2.3) whenever perceived tokens from the given interlocutor activate some of the speaker's tokens.

If this number for the given speaker-interlocutor pair is 1 (the speaker has a maximally favorable attitude towards the interlocutor), the model carries on as it would without the inclusion of attitudes, and accommodation happens by default. At the other extreme, if this measure is at 0, no accommodation happens, since the interlocutor's tokens do not activate any part of the speaker's representation. This also reflects the view that convergence is automatic, but it can be suppressed by disliking or negative attitudes about the interlocutor. In this sense the effects of likeability and attitude can only be inhibitory, and never facilitating (since with the maximum attitude of 1, we're still only at where we would have been with no likeability effects).

It is worthwhile to note that this solution does not resolve the issue of when divergence happens, since not even a negative attitude score would activate tokens that are *further away* from the input. The implementation of divergence is left for future work.

5.3 Incorporating Maintain categories

In this section I am going to describe how the algorithm described in *Section 5.2* can be adjusted in order to explain the outcome of the experiment. The experimental results indicate that speakers are sensitive to the phonetic detail of input tokens, specifically sensitive to whether the phonetic profile of an input token matches the speaker's pre-conceived idea of the token's category. Participants were seen to converge with stimuli where the /p b/ contrast was realized as an exaggerated version of how it is in their native language (i.e. English participants converged with a 130 ms VOT /p/ and a plain /b/ and Hungarian participants converged with a plain /p/ and a /b/ with 130 ms prevoicing). At the same time, English participants did not imitate a prevoicing contrast nor did Hungarians imitate an aspirating /p b/ contrast. These results indicate that the phonetic properties of a token—i.e. how typical of a realization it is for its category—also matter in the course of accommodation, which

supports the **Maintain categories** hypothesis. In the following, I propose that these results can be modeled by adding a filter to the basic model described in *Section 5.2*.

5.3.1 The new algorithm

Maintain categories can be implemented by a filter on top of a categorization algorithm: input tokens will have a different effect on the speaker depending on how typical they are for their category. Such a filter would disallow atypical tokens from activating pre-existing tokens in the speaker's mental representations thereby preventing these tokens from having an effect on the speaker's future productions. This filter would need to tap into a classification function that determines the phonemic category of a token based on its phonetic properties. A classification algorithm was not part of the basic model, but can be independently motivated to be a part of speakers' perception and production processes.

This section is restricted to implementing **Maintain categories**, which works with the output of a classifier, but is agnostic to how that classifier works. The filter representing Maintain categories can itself be implemented in two ways, and in the following I will go through these two options. Both options affect the Activation step (in Line 9 and 10 in the original model, Algorithm 1) based on whether the token's *phonetic category* matches its *lexical category*. The first approach is a categorical one, while the second approach is a gradient one. The rest of the model will remain the same.

The first option is to implement a pass/fail structure for activation, shown in Algorithm 2. The token is classified twice: once based on lexical information, where it gets a label c_l (lexical category), and once based on its phonetic properties, where it gets a label c_p (phonetic category). These are shown in Lines 9–10. If the two labels match (Line 11), i.e. if the phonetic information is congruent with the lexical information, then the activation step happens as described earlier in *Section 5.2* (Line 12). If, however, based on phonetic information the token is more likely to be

a member of a different category than what the lexical information suggests, then the activation step is skipped. Irrespective of whether the activation step happened, the model carries on: it incorporates the new input token into the speaker's representation, produces a new output token based on the tokens with the highest levels of activation and stores the new output token in the speaker's representation as well (Line 14 and on).

With this algorithm, typical tokens (or at least tokens that are not anomalous from a classification point of view) influence the speaker's response and thus can induce convergence, whereas atypical tokens that disrupt classification do not. For instance, this is what would happen with the plain /p/ stimuli in the English *Extr. Prev.* condition in the word *pooling* /ˈpuːlŋ/. The lexical information (the absence of */ˈbuːlŋ/) suggests that the stop is a /p/, but the 15 ms VOT suggests that it is a /b/. Since it receives mismatching labels during the two classification processes, the activation process is skipped, and thus the token will have no effect on subsequent productions. At the same time, a token that is extreme in the other direction might not disrupt classification and thus induces convergence. For example, a /p/ token in the word *pooling* /ˈpuːlŋ/ with 130 ms aspiration will get recognized as a /p/ based on both lexical and phonetic information. Since the two labels match (both are '/p/'), the activation step happens as normal. This token activates the tokens closest to it, and since the next production is always a weighted average of activated tokens (weighted based on activation level), the token the speaker produces next will be swayed in the direction of the input token—i.e. convergence will happen.

An alternative would be to implement **Maintain categories** in a gradient way, which is shown in Algorithm 3. With this implementation the model compares lexical information about the phonemic identity of sounds with their phonetic profile, and penalizes incongruities. This is very similar to the concept of resonance in Adaptive Resonance Theory, which was discussed in *Section 5.1*.

Algorithm 2 Implementing Maintain categories categorically

1: SET UP the *Speaker Set* (S_{Sp})

2: POPULATE the *Speaker Set* with l number of tokens (t), based on the D_S distribution, defined along d dimensions (*Speaker's initial state*) with an activation level

3: SET UP *Interlocutor Set* (S_{Int})

4: POPULATE the *Interlocutor Set* based on elements of D_I (*Interlocutor Distribution*)

5: BASELINE PRODUCTION Create a token t_0 with the median values of the Speaker Set

6: RESTING ACTIVATION Increase the activation level of some or all tokens in the *Speaker Set* based on the token's similarity to t_0 (defined by activation function A) *Speaker Set*

7: **for** i iterations **do**:

8: CHOOSE RANDOM TOKEN from S_{Int}, $t_{l+(i*2)-1}$

9: CLASSIFY(Lexical) the token: $t_{l+(i*2)-1}$ is given a label c_l based on lexical information

10: CLASSIFY(Phonetic) the token: $t_{l+(i*2)-1}$ is given a label c_p based on its phonetic properties

11: **if** c_p == c_l **then**

12: ACTIVATE(int): Increase the activation level of some or all tokens in the *Speaker Set* based on the token's similarity to $t_{l+(i*2)-1}$ (defined by activation function A)

13: **end if**

14: INCORPORATE(int) $t_{l+(i*2)-1}$ to the *Speaker Set* (S_{Sp})

15: PRODUCE a new speaker token ($t_{l+(i*2)}$) that's the weighted average of the activated tokens

16: ACTIVATE(sp): Increase the activation level of some or all tokens in the *Speaker Set* based on the token's similarity to the speaker's new production ($t_{l+(i*2)}$, same mechanism as Line 12)

17: INCORPORATE(sp) ($t_{l+(i*2)}$) to the *Speaker Set* (S_{Sp})

18: DEACTIVATE some tokens in the *Speaker Set* (S_{Sp})

19: **end for**

In such a model the less typical a token is, the less of an impact it has on the phonetic properties of the speaker's next production. This option seems more appealing because gradient coefficients allow for modeling the variability we saw in participants' reactions to the same stimuli. In this scenario, the speaker assigns a category label to the input token, just like in the previous solution (Line 9). Then, however they do not simply assign another, phonetic label to the token, but assess the probability of the input token belonging to category c_l. In this model, this number (m) is the Bayesian probability of the token belonging to the category that it was assigned based on lexical information given its phonetic profile.[7,8] These probabilities were computed anew in every iteration, incorporating the updates from previous iterations. This means that the probability of a given token belonging to category c_l given its phonetic information changed over time—i.e. the standards of typicality were constantly shifting.

This number (m for match) will be between 0 and 1. For an even more drastic picture, this number could be optionally rounded up and down near the edges (e.g. probabilities below 0.2 could be rounded down to 0 and probabilities above 0.8 can be rounded up to 1). Such a rounding threshold can introduce further variation between participants—some will be more dismissive of tokens (and disallow even mildly atypical tokens from affecting their representations), while others will be more permissive (and be willing to converge with a wider range of productions).

[7]In this model, conditional probabilities were estimated with Kernel's probability density function in Python.

[8]Instead of Bayesian probabilities, logistic regression or machine learning algorithms like support-vector machines would presumably also have been able to yield a suitable m.

Algorithm 3 Implementing Maintain categories gradiently

1: SET UP the *Speaker Set* (S_{Sp})

2: POPULATE the *Speaker Set* with l number of tokens (t), based on the D_S distribution, defined along d dimensions (*Speaker's initial state*) with an activation level

3: SET UP *Interlocutor Set* (S_{Int})

4: POPULATE the *Interlocutor Set* based on elements of D_I (*Interlocutor Distribution*)

5: BASELINE PRODUCTION Create a token t_0 with the median values of the Speaker Set

6: RESTING ACTIVATION Increase the activation level of some or all tokens in the *Speaker Set* based on the token's similarity to t_0 (defined by activation function A) *Speaker Set*

7: **for** i iterations **do**:

8: CHOOSE RANDOM TOKEN from S_{Int}, $t_{l+(i*2)-1}$

9: CLASSIFY(Lexical) the token: $t_{l+(i*2)-1}$ is given a label c_l based on lexical information

10: CLASSIFY(Phonetic) the token: $t_{l+(i*2)-1}$ is given a probability of belonging to c_l

11: ACTIVATE(int): Increase the activation level of some or all tokens in the *Speaker Set* based on the token's similarity to $t_{l+(i*2)-1}$ (defined by activation function A, **multiplied by coefficient** m **reflecting** $t_{l+(i*2)-1}$**'s match with** c_l)

12: INCORPORATE(sp) $t_{l+(i*2)-1}$ to the *Speaker Set* (S_{Sp})

13: PRODUCE a new speaker token ($t_{l+(i*2)}$) that's the weighted average of the activated tokens

14: ACTIVATE(sp): Increase the activation level of some or all tokens in the *Speaker Set* based on the token's similarity to the speaker's new production ($t_{l+(i*2)}$, same mechanism as Line 11)

15: INCORPORATE(int) ($t_{l+(i*2)}$) to the *Speaker Set* (S_{Sp})

16: DEACTIVATE some tokens in the *Speaker Set* (S_{Sp})

17: **end for**

Unlike in the categorical option, the Activation step (Line 11) always happens in the gradient option, even if the token is a particularly bad match for their lexical category. However, the influence an input token is allowed to have on the speaker's representation is in some way proportional to how good of a fit it is for its lexical category. If a certain token is a good fit (m is 1), then it will be allowed to have maximal impact, and activate tokens in the speaker's representation just like input tokens in the basic model in *Section 5.2*. From this perspective, a token that is atypical, but in a way that does not interfere with categorization will be deemed a good match for its lexical category. For instance, an initial stop in *pooling* /ˈpuːlɪŋ/ with 130 ms VOT will be a good fit for its category, because its probability to be categorized as something else (e.g. a /b/) is very very small.

If, however, the input token is phonetically a bad match for its lexical category, its potential to activate other tokens will be minimized. Therefore, even if it is not the first time the speaker encounters such an anomalous token, the previous atypical tokens will not be activated to a degree that is sufficient for convergence. If we assume that some of the speaker's own (typical) tokens started out activated before the input token was uttered, we will be left with a weighted average that is so heavily skewed by the speaker's own typical productions that the new input token will not be able to influence the speaker's next production. This cycle could be perpetuated without the speaker's productions being swayed measurably over time. if each production that the speaker utters in response is relatively typical, thereby reinforcing the activation levels of the speaker's typical tokens.

Notice that in both of these alternatives the token is eventually added to the *Speaker set*, unlike in a dual-pool model. While this step could also be omitted in case of a mismatch between the token's lexical category and phonetic category (as such incongruous tokens are "ignored" in Adaptive Resonance Theory), leaving this step in was a conscious decision that I made for two reasons. The first one is a smaller, more formal point. I believe that the gradient approach is a closer fit for data we have so far, and the incorporation of a token into a set cannot be implemented in

a gradient way. The second reason is more substantial and relates to the potential for individual change. If mismatching tokens are not incorporated into the speaker's representation (i.e. they are not committed to memory), then such tokens will not only be unable to affect the speaker's productions within the interaction, but they cannot sway their production over a longer span of time (over perhaps years' of interactions) either.

I believe that models of accommodation must allow for smaller changes accumulating into larger changes over time, even if those larger changes involve altering the phonetic classification of certain types of tokens. However, whether or not atypical or anomalous tokens are incorporated into the speaker's representation is independent from the way **Maintain categories** is modeled in this work, since they do not make a difference for short-span interactions. If new information surfaces about the capacity of long-term interactions to change representations, the model could easily be adjusted by omitting the incorporation step.

5.3.2 Simulations

In this section I am going to describe some basic simulations that were run with this model. Since the two experiments had largely similar results, these simulations will only focus on the case of English. Specifically, it will model participants' behavior regarding the two different /p/'s (130 ms VOT in *Extr. Asp.* and 15 ms VOT in *Extr. Prev.*), which were the sounds that posed representational difficulties to the participants (in *Extr. Prev.*). While /b/ tokens will be included in the Speaker set, because they are necessary for the categorization process, the model only considers /p/ input. Accommodation to especially the prevoiced /b/'s (in *Extr. Prev.*) can also depend on the participants' articulatory aptitude and ability to manipulate prevoicing as a cue, which can introduce further complications in modeling the results, which are outside of the scope of this work.

Four basic simulations have been run with the gradient Maintain categories model in a two-by-two design (by condition and "speaker"). Each model started out with 100,000 /p/'s and

100,000 /b/'s, which were generated with a random seed based on real participants' productions from the PRE-Read (see below). These models than "shadowed" 20 /p/ productions that rather than being randomly chosen were either all aspirated (130 ms VOT) or had short-lag VOT (15 ms VOT), to match the experimental conditions (*Extr. Asp.* and *Extr. Prev.*, respectively). The exposure tokens (20/model) were fewer than what was used in the experiment (90/participant), since participants also likely had many more than 100,000 exemplars of /p/ in their representations.

The starting distributions of the models were always based on an actual participant from the English experiment. Out of the four models, two represented an English speaker who does not prevoice their /b/'s at all and two represented an English speaker who prevoices more than half of their /b/'s. The means and standard deviations were based on the pre-exposure reading /p/ and /b/ data of F02, and F08, respectively. Their read productions are shown in *Figure 5.2* below. In total, there were four models: a model of a non-prevoicer in *Extr. Asp.*, a non-prevoicer in *Extr. Prev.*, a prevoicer in *Extr. Asp.*, and a prevoicer in *Extr. Prev.*

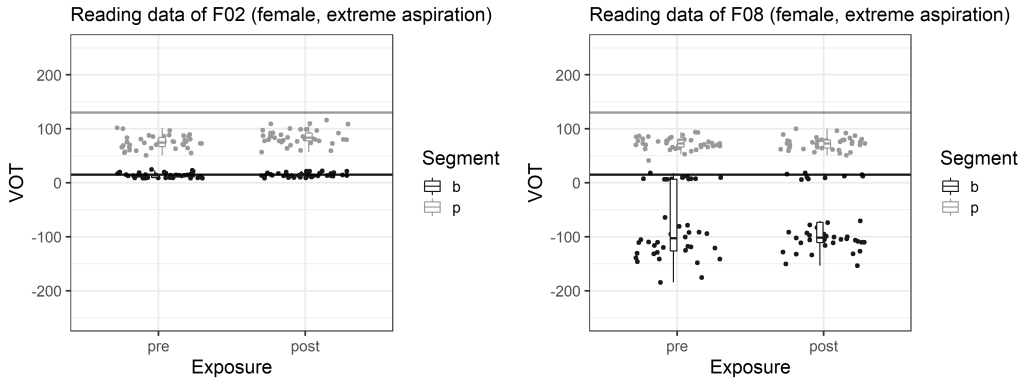

Figure 5.2: The reading productions of F02 and F08 in the English experiment

Figure 5.3 shows the results of the /p/ accommodation models by condition (reflected by color) and by "speaker" (reflected by shape). Both speakers show the same trajectories for both conditions. In the *Extr. Asp.* condition, both participants converge with the 130 ms VOT input token, but do not quite reach it. This is in line with what we saw in the English *Extr. Asp.* condition's data.

The model also correctly predicts the lack of convergence in *Extr. Prev.* (towards the 15 ms VOT /p/).

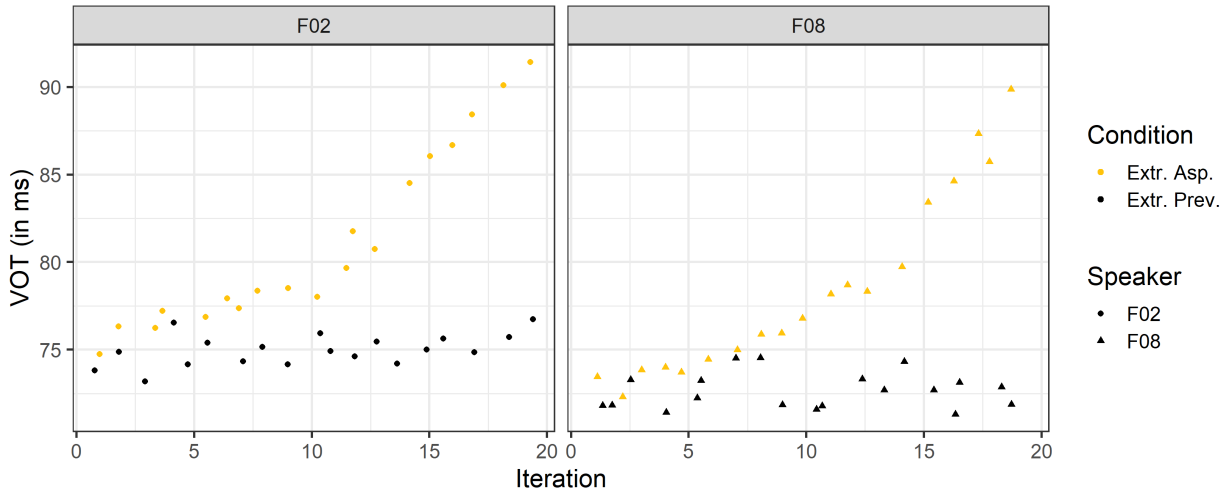

Figure 5.3: Four simulations with the gradient Maintain categories algorithm (/p/ productions only)

These results are independent of what the model talker's /b/'s look like, the only thing that matters is that the 15 ms VOT /p/ is within the range of a typical /b/ and therefore it is a fairly atypical /p/. The model is insensitive to the categories in the *input*, all that matters is the categories that the speaker has (and had before exposure). This is exactly what the results of Nielsen (2008) and this dissertation jointly suggest: irrespective of whether English-speaking participants are shown that the model talker preserved the /p b/ contrast elsewhere (see this work) or not (Nielsen, 2008), they will not converge with a plain /p/, as it does not make a valid target.

5.3.3 Implementing Maintain categories and likeability measures

Lastly, I would like to point out that Maintain categories and likeability are treated quite similarly in this model. As I pointed out in *Section 5.2.4*, likeability could be incorporated as a coefficient which describes the desired relationship or attitude of the speaker towards the interlocutor. The coefficient is between 0 and 1, and it is applied to the activation function, either allowing it to proceed normally (at a value of 1) or dampening its effect (at values closer to 0). This is quite similar to how the

coefficient m alters the workings of the activation function in the gradient Maintain categories model (in Algorithm 3).

I argue that this similarity between implementations is not necessarily a problem, since likeability and the typicality of input tend to influence each other as well. In a matched-guise perceptual study, Babel et al. (2019) found that participants rate the model talker as less "pleasant" if their speech contains an unnatural shift. This indicates that phonetic typicality can influence likeability.[9] Even though Babel et al. (2019) found that perceptual shifts happened even as a reaction to unpleasant guises, it might be the case that this would not have led to convergence in production. In this hypothetical case, it might have been difficult to tease apart how much of the lessened or no convergence is due to the unnaturalness of stimuli itself, and how much is induced by the *the perceived unpleasantness* of the guise in and of itself (which could have in turn been also influenced by their unnatural sound patterns). At the extreme one could assume that every time when participants (do not) react to unnatural stimuli, it is only because of likeabilty and attitudes, and the reason why other factors such as phonetic typicality can have an effect on convergence is because they can contribute to the attitude that the speaker forms of their interlocutor (e.g. less typical speakers are perceived less positively). Such a theory could paint likeability and social attitudes as the main gatekeeper and source of stability of language, and it could lead to even contrast preservation being an epiphenomenon of likeability.

This is certainly an interesting topic for future research, but the present work is unfortunately not in a position to address this, since likeability ratings were given *before* the shadowing task, and 15 minutes of exposure to the unnatural speech patterns might have changed the participants' views

[9]Although there is less evidence on this, one might also assume that likeability also affects how harshly the speaker reacts to atypical stimuli: perhaps a speaker might be more like to tolerate atypical productions from a likeable interlocutor.

on the model talker. However, I believe that implementing likeability and atypicality of tokens by the same tools computationally is not a problem with, but rather an advantage of the present model.

5.4 Conclusions

This dissertation examines the two-way relationship between phonological representations and accommodation. The previous three chapters concluded that there is a pressure for categories to be phonetically consistent (**Maintain categories**). In this system, the way pre-existing representations limit accommodation is by providing a basis for standards of typicality that the new input tokens can be measured up against. This chapter discussed how accommodation affects phonological representations in turn.

In this chapter I offered a computationally explicit model of accommodation that implements an adherence to the phonetic details of categories (**Maintain categories**) as a co-efficient of the activation function derived from the Bayesian probability of the token's lexical category given its phonetic properties. I demonstrated that this model is capable of simulating the experimental findings of this work—i.e. that participants converge with extreme tokens, unless those tokens are typical realizations *for another category* or at least are categorically ambiguous, in which case, no accommodation was found in the experiments. The models followed these trajectories.

Within the model, accommodation changes representations in every instance of exposure. The input is not only added to the representation that the speaker stores of the given category, but it also changes the activation pattern of the representation, which directly affects the next production(s). In this exemplar model the standards of typicality imposed on input tokens are dynamically evaluated (i.e. they are updated from iteration to iteration rather than being only established once at the beginning of the simulation). This means that new tokens (and accommodation) themselves can also affect what counts as a typical realization of a category over time. This allows for an opportunity of incremental long-term change over time.

Chapter 6: Conclusions

This dissertation focused on the relationship between phonological representations and accommodation. This relationship is mutual: phonological representations restrict accommodation, and accommodation in turn alters phonological representations over time. The first half of this issue was approached empirically through contrasting mechanisms for contrast preservation by comparing the treatment of VOT contrasts in aspirating and prevoicing languages (*Chapters 2–4*). *Chapter 5* offered a computationally explicit model of both accommodation as such and of the particular effects found in the empirical half of this study.

Chapters 3 & 4 presented results from two experiments. In both experiments participants had to complete a six-repetition shadowing task, sandwiched between pre- and post-exposure labeling and reading tasks involving p-initial and b-initial words. Participants also had to rate their model talker beforehand along 9 different semantic differential scales for likeability. Data on the participants' ethnic and gender identity were also collected. Only two genders were represented among participants: male and female. Experiment 1 (*Chapter 3*) was conducted with native speakers of English (an aspirating language), an English-speaker white female model talker, and English words as stimuli, while Experiment 2 (*Chapter 4* was conducted with native speakers of Hungarian (a prevoicing language), a Hungarian white female model talker, and Hungarian words.

Both experiments had two conditions, which only differed in the audio stimuli that were used in the shadowing task. In the *Extreme Aspirating* condition 130 ms VOT (aspirated) word-

initial /p/'s were contrasted with 15 ms VOT (plain) word-initial /b/'s. In the *Extreme Prevoicing* condition 15 ms VOT (plain) word-initial /p/'s were contrasted with −130 ms VOT (prevoiced) word-initial /b/'s. While the former reflects an exaggerated version of the naturally occurring /p b/ contrast in English and is more of a departure from Hungarian, the latter is the opposite—*Extreme Prevoicing* is quite un-English-like, but more natural for Hungarian speakers.

These experiments were set up to test two competing hypotheses for how contrasts are preserved in language. The first hypothesis (**Maintain contrasts**) is a pressure to maintain contrasting sounds as distinct from one another, but it is quite abstract, as it does not limit the phonetic realization of either category itself. **Maintain categories** is a more concrete imperative to adhere to the phonetic details of any given category, i.e. that no token of a given category should be so atypical that it is no longer likely to belong to its intended category. While Maintain contrasts predicts that participants should be able to accommodate to both conditions equally, Maintain categories predicts that accommodation should only be possible when the labels assigned to the stimuli based on lexical and phonetic information do not conflict—i.e. no accommodation in the English *Extreme Prevoicing* condition or the Hungarian *Extreme Aspirating* condition. In the end, experimental data supported the predictions of **Maintain categories**.

6.1 Summary of results

6.1.1 Experiment 1: English

While there was a lot of individual variation in both experiments, in Experiment 1 (the English experiment), there were different group-level effects in the two conditions. While participants in the *Extreme Aspirating* condition (*Extr. Asp.*) accommodated to both the /p/ and /b/ stimuli, no group level change was recorded in the *Extreme Prevoicing* condition (*Extr. Prev.*).

While no perceptual change was observed in the labeling task in *Extr. Asp.*, in production participants converged with both the aspirated /p/ and the plain /b/ stimuli in terms of their shadowed as well as their read productions. At the same time, there was no significant correlation between individual participants' treatment of /p/ and their treatment of /b/. In the case of /b/, convergence was modulated by two tendencies. First, convergence to the plain /b/ was muddled by many participants already matching the 15 ms VOT target in their pre-exposure reading productions (their baseline). Since these participants could not get closer to the target, they did not have any room to exhibit accommodation either. Second, during the shadowing task, a group of male participants diverged from the stimuli (i.e. reacted to a plain /b/ by prevoicing their own /b/'s compared to their) baseline. These participants tended to be either People of Color or from areas of the US where word-initial prevoicing is more common in the regional dialect. This increase in prevoicing was the strongest in the first (or first few) repetitions, and gradually disappeared throughout the task. The fact that most participants converged to the target while others even diverged from it could be explained as all participants emphasizing their /b/'s (presumably as a generalization from the very salient aspiration on /p/'s), but because of a difference in dialectal backgrounds, participants executed this emphasis differently. If the participant spoke a non-prevoicing dialect (or at least one where prevoicing is infrequent), then their emphasized /b/'s lost even the little prevoicing they might have had in the PRE-Read, and if they spoke a dialect more prone to prevoicing, emphasis on their /b/'s manifested as prevoicing.

In terms of extralinguistic factors, no effects of likeability were observed in this condition. With respect to ethnicity, participants of color tended to have more prevoicing on their /b/'s in the PRE-Read, while white participants most often had plain /b/'s. This went hand in hand with the fact that the biggest cases of convergence were seen from participants who were People of Color, i.e. those who had room to converge in the first place. This indicates that a participant's ethnicity only affected their baseline, but not necessarily their ability or willingness to accommodate to

the white model talker. With respect to gender, there were two tendential differences between males and females. The first is the one already described before (rather than converging, a subset of males prevoiced their /b/'s more during shadowing). The second was also observed during shadowing, where females were more likely to converge to the aspirated /p/ at once (i.e. in the first repetition, and then maintaining those productions throughout the rest of the task), whereas males tended to approach the target gradually (incrementally throughout the task). I argue that this was a distance effect—i.e. males' baseline /p/ productions were more different from the model talker's than females' were, and the gap between their productions took more time to close.

In the *Extreme Prevoicing* condition (*Extr. Prev.*), there was much less change in production, even though participants were significantly less likely to label a 15 ms VOT stimulus as /b/ after exposure to the prevoicing contrast in *Extr. Prev.* than they were at the start of the experiment. This effect was mostly carried by a few participants, whose shadowing productions (atypically for this condition) also reflected change.

In terms of production, participants as a group exhibited no convergence to either the plain /p/, nor the prevoiced /b/ target. On an individual level, /p/ accommodation and /b/ accommodation were still not correlated. While on a group level no effects were seen, there were some participants who partially converged with the prevoiced /b/ during shadowing. This tended to take a V-shape (the VOT of /b/ decreased gradually in the first few repetitions, hit its minimum towards the middle of the task, and then subsequently gradually disappeared). This was only statistically significant for females. I argued that this effect was restricted to females because the males who were randomly sorted into *Extr. Prev.* were especially unlikely to prevoice, which was confirmed by a comparison between the baselines of men in *Extr. Asp.* and *Extr. Prev.*

In terms of extra-linguistic factors, likeability effects were common in *Extr. Prev.* In the reading dataset, there was a correlation between `Solidarity` ratings and /p/ accommodation—participants who rated the model talker low on `Solidarity`-related measures (*friendly — un-*

friendly, *dishonest — honest*, and *rude — polite*) were more likely to diverge from her (aspirate their /p/'s more) than those who did not. Moreover, there was a similar `Superiority` effect in the shadowing dataset—participants who rated the model talker low at the beginning of the experiment, tended to diverge from her. For females this was only observed for /b/'s, but males exhibited this pattern for both /b/'s and /p/'s. In terms of ethnicity, there were no notable effects in this condition.

6.1.2 Experiment 2: Hungarian

In Experiment 2 Hungarian participants completed the same task with Hungarian words and a Hungarian female model talker. In this experiment all participants identified as "Hungarian" in terms of ethnicity, therefore ethnicity effects could not be observed. In Experiment 2, it was *Extr. Prev.* that was more similar to the participants' native language, Hungarian. Even so, exposure had an effect on participants' labeling performance in this condition. After exposure to the plain /p/ (15 ms VOT) and prevoiced /b/ (-130 ms VOT), participants were more reluctant to label a word-initial stop with 0 ms VOT as a /p/. Rather than being spearheaded by a handful of participants, this pattern was exhibited by most people in *Extr. Prev.* While this effect could be explained by participants potentially learning to judge 0 ms VOT as too short for /p/ (since the exposure /p/'s had 15 ms VOT), this is still a surprising effect, since changes in perception (i.e. attunement) were not necessary in order to correctly process the shadowing stimuli in this condition.

In the *Extr. Prev.* production tasks, we saw very similar results to what was observed in the English *Extr. Asp.* condition. Hungarian participants converged with both the plain /p/ and the prevoiced /b/ in the *Extr. Prev.* stimuli, both in the reading and the shadowing data, while /p/ accommodation and /b/ accommodation were not correlated on an individual level. Both on a group and on an individual level, there was some overlap between VOT values of /p/'s and /b/'s, which was not attested in the English data. In the shadowing task convergence was almost exclusively observable in the form of instantaneous convergence (i.e. flat VOT trajectories across

the six repetitions). I argued that it was still convergence based on a comparison with productions from *Extr. Asp.*

In terms of extra-linguistic factors, there was an effect of `Superiority`, but only for /b/ and only in the male dataset. Males in *Extr. Prev.* converged with the model talker more if they rated her favorably on scales related to `Superiority` (*organized — disorganized*, *lower status — higher status*, and *intelligent — unintelligent*). This effect is somewhat surprising, since in Experiment 1 likeability effects were only observed in the less "native-like" condition. Aside from the genderedness of this likeabilty effect, there was another gender effect in *Extr. Prev.*, also in the reading dataset. While both genders converged with the model talker's plain /p/ (15 ms VOT), the amount of change in males' productions was greater than in females' productions. However, this can be a phonetic distance effect, since the baseline of males was further away from the target than females' were. During the shadowing task no effects of either likeability or gender were recorded.

The stimuli in the *Extr. Asp.* condition modeled an aspirating contrast, dissimilar to how the /p b/ contrast is typically realized in Hungarian. In this condition, no convergence was found on a group level for either /p/ or /b/ in either the reading or the shadowing task and no perceptual change could be traced in the labeling task either. While the trajectories were largely flat during the shadowing task (no change from repetition to repetition), there were some instances of A-shaped trajectories. In these cases, participants partially converged with the aspirated /p/ stimuli (even though they typically stayed below 50 ms VOT in their productions), but this convergence gradually disappeared in the second half of the task.

Upon closer inspection of the data, there was also one effect of likeability in the reading data. An effect of `Solidarity` could be established both for males and females for /b/—participants who rated the model talker low for the relevant measures also tended to diverge to a proportionate extent. In the case of males, this resulted in significant group-level divergence. While this effect only concerned /b/ productions, and could not be found for /p/'s, /p/ and /b/ accommodation were

correlated on an individual level. This indicates that there might have been some conditioned divergence in the /p/ dataset as well, albeit to a smaller magnitude. In this condition there was even more of an overlap between /p/ and /b/ productions in shadowing and POST-Read productions than in *Extr. Asp.*

6.2 Conclusions and implications

The findings from this dissertation contribute to various topics. First, I am going to discuss the relevance of this dissertation's findings with respect to the interaction of phonological representations and accommodation (*Section 6.2.1*). Then, I am going to discuss the phonetic conclusions and implications of this work, mostly drawing on the experimental results (*Section 6.2.3*). Then, I am going to move on to the sociolinguistic contributions of this work, namely findings on likeability, gender, and ethnicity (*Section 6.2.3*). Finally, I am going to conclude this section by discussing some broader implications on the accommodation literature in general (*Section 6.2.4*) including phonetic distance effects, parallel accommodation to contrasting categories as well as how the present work is compatible with sound change phenomena.

6.2.1 Representational limitations and consequences of accommodation

The main conclusion of these results is that the speaker's phonological representations do place limits on what targets are possible to accommodate to. Participants in both experiments accommodated to extreme realizations of their native contrasts (English *Extr. Asp.* and Hungarian *Extr. Prev.*), where the /p b/ contrast was encoded by the same cues as it typically is, but these cues were exaggerated. The /p b/ contrast was clearly maintained in the atypical conditions as well (English *Extr. Prev.* and Hungarian *Extr. Asp.*), but in an atypical way for the given language.

This indicates that contrast preservation is not only present during accommodation as a flexible requirement to maintain contrasting sounds as phonetically distinct (i.e. **Maintain contrasts**),

but how that distinction is made also matters. In order for convergence to happen, categories in the input also need to be phonetically consistent with how they are realized in the speaker's speech (**Maintain categories**). That is, while in perception speakers are able to attune to interlocutors' speech patterns and have a remarkable ability to infer from bimodal distributions, they also bring a significant amount of the phonetic detail of their own categories to the table, which can prevent them from accommodating to certain kinds of bimodal stimuli. The experiments in this work show that over a short amount of exposure (e.g. a shadowing task), these preconceptions about the phonetic properties of categories cannot be overwritten. This might be possible over longer spans of exposure, for that see *Section 6.2.4.3*.

These results can also be modeled in an exemplar model, utilizing activation as well as the concept of resonance. I have given a computationally explicit implementation of such a model in *Chapter 5*. In this model, I have successfully implemented **Maintain categories** as the influence of the Bayesian probability of the given input token belonging to its lexically-determined (top-down) category given its phonetic properties (bottom-up information). The model predicts that if the token's phonetic properties suggest that it is unlikely to belong to the category that lexical information suggests, then accommodation is significantly dampened or even completely blocked. However, this does not limit participants from accommodating to stimuli that are atypical but in a way that does not impede categorization (e.g. extremely aspirated English /p/'s). This prediction lines up with the experimental results summarized in *Section 6.1*.

This model also demonstrates that in addition to the speaker's pre-existing representations limiting what types of input they can accommodate to, accommodation also affects representations. Input tokens not only activate certain parts of the speaker's representation, resulting in unique activation patterns which influence the subsequent production(s), they also get added to the speaker's representation. By being added to the set of exemplars that the speaker stores, they also (very slightly) influence the categorization algorithm, and thus the Bayesian probability for belonging to

a given category will be updated. This means that every input token has the potential to ever-so-slightly shift the phonetic standards for belonging to a certain category, which in turn results in an ever-so-slightly altered processing of future tokens. Thus, not only does accommodation affect phonological representations, but by altering them, previous instances of accommodation have the capacity to affect the *limitations* of these representations on future intances of accommodation. The longer-term consequences of this will be discussed in *Section 6.2.4.3*.

It is informative to consider these results in the context of previous research. First and foremost, the results from this dissertation contextualize Nielsen's work (2008; 2011). In her shadowing experiment participants were exposed to either /p/-words with over 100 ms of aspiration or /p/ words whose natural aspiration was decreased by 40 ms. She found that participants accommodated to the former (increased the amount of aspiration on their /p/'s) but not to the latter (did not decrease their amount of aspiration). The findings of this present work disambiguate Nielsen's results by establishing that her effects hold even if participants also receive evidence of the distinction being maintained (i.e. plain /p/'s are contrasted with prevoiced /b/'s in the stimuli). This shows that the speakers' typical phonetic categorization is enforced on typical and atypical stimuli alike, in a way that is not directly related to contrast preservation. However, if the phonetic consistency of individual sound categories is maintained, contrasts are also being preserved indirectly as a consequence.

While this research found no accommodation to plain /p/'s, there is evidence of English speakers imitating voiceless stops with shortened VOT. For instance, Levi (2015) finds that exposure to artificially shortened word-initial /k/'s primes her participants' /t/ productions, and these participants produce their /t/'s with shorter VOT word-initially than participants in the baseline condition do. While this seems in contradiction with the *Extr. Prev.* English results (where no accommodation to a plain /p/ was found), the two findings are not necessarily contradictory. A crucial difference is in the exact length of aspiration in the shortened VOT stimuli in the two studies. The

VOT values for the shortened /k/ primes in Levi (2015) came from naturally produced /p/ stops, and resulted in values that while indeed short for a /k/ are still within the aspirated range (over 40 ms). As I have argued, the compatibility of bottom-up and top-down information is essential for accommodation, and in Levi's stimuli the phonetic properties of her shortened /k/'s were presumably still in accordance with the information that participants received from lexical information (the word-initial /k/ appeared in the word *keen* /ˈkiːn/ and */ˈgiːn)/ is not a word of English). This is in contrast with the 15 ms VOT values used in the present experiments for plain /p/.

Another type of evidence for English speakers imitating voiceless stops with shortened VOT comes from continuum-imitation tasks and other tasks where participants have to imitate words from a prevoicing language (Olmstead et al., 2013; Nagle, 2019, among others). When participants (are often explicitly asked to) imitate syllables on a VOT continuum or foreign words from a prevoicing language, they are largely successful in producing both plain /p/'s and prevoiced /b/'s. While this does show that English speakers are able to differentiate even between plain and prevoiced word-initial stops, which is only a subphonemic distinction in English, in these instances participants do not have access to lexical information. In the case of syllable continua, we can confidently say that the participant has no access to lexical information, and the stop does not receive a top-down (lexical) label. However, it can also be argued that the phonemic identity of sounds in foreign words is not available (or not to the same extent) as of sounds in one's native language(s), especially in the case of learners that are at early stages of their studies of the given language. While there might be some orthographic information available, its influence on phonemic identity is questionable. For instance, while the /p/ in Hungarian *pír* /ˈpiːr/ 'blush' is spelled with the same letter as the /p/ in English *peer* /ˈpiːɚ/, even though it is phonetically more similar to the /b/ in English *beer* /ˈbiːɚ/, it is not clear that these sounds are represented together by English-speaking learners of Hungarian. While the mental representations of L1's and L2's and the cross-linguistic phonemic identity of sounds goes beyond the scope of this work, we can at least say that it is

questionable whether the stops in these such stimuli receive a top-down (lexical) label in the same way that stops in English words do. If that is the case, the issue of discrepancy between lexical and phonetic information cannot arise in the first place, and thus accommodation is not limited by phonological representations either.

6.2.2 Phonetic implications

This dissertation compared the treatment of un-native-like voicing contrasts in the context of both aspirating and prevoicing languages. The fact that speakers of neither type of languages converged with the contrasts from the other is particularly significant. Besides the implication that in order for a token to be imitated it needs to meet certain phonetic criteria specific to its lexically established category, it also makes broader implications for aspiration and prevoicing as cues. Aspiration and prevoicing are commonly treated as opposite ends of the VOT continuum (Lisker and Abramson, 1964; Abramson and Whalen, 2017; Cho et al., 2019, *inter alia*). While languages with a three-way voicing opposition often distinguish between aspirated (long-lag), plain (short-lag) and prevoiced stops, the two cues are articulatorily and acoustically quite different, which has led some to caution against this approach for languages with a binary voicing contrast.

The present work shows that speakers adhere to certain phonetic ranges for their categories, and they do not treat their respective subphonemic cue to voicing in the same way as they do their contrastive one. In both of the languages in the experiment one of these cues is contrastive—aspiration (long-lag) in English and prevoicing in Hungarian—and its presence or absence alone can distinguish between voiced and voiceless sounds. The other cue is subphonemic—the presence or absence of prevoicing in English or aspiration in Hungarian does not distinguish contrastive sounds. When comparing the treatment of *Extr. Asp.* /p/ and *Extr. Prev.* /b/ in Experiments 1 and 2, we can see that English and Hungarian accommodate to extreme aspiration and extreme prevoicing differently, in a language-specific way. This suggests that to speakers of a given binary

voicing language (irrespective of whether it is an aspirating or a prevoicing language) aspiration and prevoicing do not occupy two equally salient ends of the same spectrum.

One way this differentiation is exhibited is through appearance of V-shaped and A-shaped trajectories in the English and Hungarian data, respectively. In the English dataset, we found V-shapes for accommodation to prevoiced /b/'s, where some participants partially converged with the prevoiced stimuli, but then gradually lost it. This gradual decrease of prevoicing was observed both in the *Extr. Prev.* dataset, and interestingly in *Extr. Asp.* for a subset of male speakers, where I argued that the initial divergence from the plain /b/ target (the appearance of prevoicing itself) was a sign of emphasis. This shows that for some reason even the English speakers that did prevoice at some point did not maintain it throughout.

The opposite of this (A-shaped patterns in Hungarian) was somewhat less common, but it was observed in the *Extr. Asp.* condition where some participants initially started to converge with the aspirated /p/ target, but then gradually lost this amount of aspiration over the second half of the task. The fact that this pattern was less common could have been a result of potential social stigmatization of aspirated voiceless stops in Hungarian. If such a stigma exists, it must mean that aspiration is at least somewhat perceptually salient to Hungarian speakers. This could potentially raise some interesting questions about aspiration's relative salience in Hungarian compared to, say, the perceptual salience of prevoicing in English.

The two patterns both reflect partial accommodation to a subphonemic cue that was gradually abandoned later in the task. In the case of English the peak amount of prevoicing was sometimes substantial (around 75 ms prevoicing), therefore its loss could either reflect articulatory fatigue (participants not being practiced enough to maintain such amounts of prevoicing for a sustained time period) or attention fatigue (participants either lost interest or changed their minds). However, in Hungarian peak amounts of aspiration were often only around 40 ms, which makes the articulatory fatigue interpretation less likely. Such trajectories were completely absent when participants

imitated extreme cues that are used contrastively in their native languages—i.e. there were no A-shapes when English speakers imitated extremely aspirated /p/'s, nor V-shapes when Hungarian speakers imitated extremely prevoiced /b/'s. This could reflect that speakers of languages with a binary voicing contrast are simply less accustomed to producing certain cues for a sustained period of time.

Another form of evidence for a difference in treating extreme aspiration and extreme prevoicing can be seen in how *often* a given cue was adjusted or enhanced. For instance, English participants (in both conditions) were more likely to adjust their /p/ productions as a result of exposure than their /b/'s. However, the opposite was true for Hungarian speakers.[1] While in theory both English- and Hungarian-speaking participants had room on the VOT continuum to distance their /p/'s and /b/'s, English speakers more often utilized the positive half of this spectrum and moved towards more aspiration, which Hungarian speakers were more likely to increase the amount of prevoicing on their /b/'s. In both languages the most common adjustments affected cues whose presence was contrastive (compared to plain stops, which were seldom changed in order to converge with the target by adapting nonnative cues). This suggests that the two cues lent themselves to manipulation to differing degrees by the speakers of the two languages.

In the end, the present data suggests that from the perspective of a specific speaker of most languages with a binary voicing distinction, prevoicing and aspiration are not equally salient ends of the same spectrum. It was contrastive cues (rather than subphonemic ones) that lent themselves more to modifications, perhaps indicating a greater mastery or confidence with these cues. At the

[1] At the same time it must be noted that in both languages changes in /b/ productions were bigger than changes in /p/ productions, which suggests that speakers have less fine-grained control over prevoicing as a cue than over aspiration. This is corroborated by the fact that while not even English speakers surpassed the 130 ms VOT /p/ target in *Extr. Asp.*, Hungarian speakers (and even one English speaker) regularly produced /b/'s that overshot the target of *Extr. Prev.* (with 130 ms of prevoicing).

same time, subphonemic cues, whose presence is optional, were less often and less consistently adjusted—i.e. if participants adopted them as a way of converging, they were most often not maintained throughout the shadowing task, but disappeared by the end of it.

The findings of this dissertation suggest that aspiration and prevoicing can be considered as equally available and salient ends of a (VOT) spectrum only in certain languages (e.g. languages with a three-way voicing contrast), and VOT might not be a universally present spectrum. However, this study leaves much for future research. First, it remains to be seen whether the results of Experiments 1 and 2 can be generalized to other aspirating and prevoicing languages, respectively. As does the issue of whether there are cross-linguistic tendencies regarding aspiration and prevoicing—i.e. whether these cues are equally "unwieldy" to speakers whose native language does not use them contrastively.

6.2.3 Sociolinguistic implications

In the following section I am going to discuss the findings of this work with respect to three relevant extra-linguistic variables: likeability, gender, and ethnicity.

6.2.3.1 Likeability

Previous research on accommodation identified likeability effects. The more a participant liked their interlocutor or model talker, the more likely they were to converge with them, while the participant disliking their interlocutor made them more likely to diverge from the interlocutor (Natale, 1975; Babel, 2010; Schweitzer et al., 2019). This is also compatible with Communication Accommodation Theory (Giles et al., 1991, and on), which argues that accommodation is a tool for the speaker to adjust social distance between the interlocutor and themselves by adjusting the phonetic distance between them. If the speaker views the interlocutor more positively, they will have a stronger motivation to close the social gap between them. However, previous studies

mostly measured the speaker's attitudes towards their interlocutor in particular on a single scale of likeability. At the same time, research in other areas of sociolinguistics has described likeability as a complex feature, which can be broken down into different components (Carranza and Ryan, 1975; Ryan and Carranza, 1975; Zahn and Hopper, 1985). In this work, I aimed to tease apart the contributions of three different facets of likeability: *Solidarity*, *Superiority*, and *Dynamism* measured by three ratings each that the participant gave at the beginning of the study based on a picture and a read passage by the model talker.

The results of the two experiments indicate that *Solidarity* and *Superiority* might have been the driving components of these likeability effects, and *Dynamism* had no demonstrable effect on accommodation behaviors. These effects showed up both in Experiment 1 and 2 and in both tasks, albeit with varying consistency. The two *Solidarity* effects show up consistently across the two experiments, which suggests that these effects might be robust, and the *Superiority* effects are either culturally specific or a result of noise in the data.

Effects of *Solidarity* could be seen in the English *Extr. Prev.* condition for /p/ as well as the Hungarian *Extr. Asp.* condition for /b/'s (but only for males), but only during the reading task in both experiments. These are both cases when a plain stop in the input was encroaching on the typical phonetic space of another category in the speaker's pre-existing representation. These effects indicate that when it came to representationally challenging inputs, it mattered more how likeable the participant thought the model talker was. Both *Solidarity* effects mostly manifested in divergence (conditioned by disliking), rather than convergence conditioned by a positive attitude. This indicates that those who disliked the model talker were more likely to distance themselves from her atypical productions, but a favorable opinion of the model talker did not drive participants to adopt the atypical way her /p b/ contrasts were realized. Given that most accommodation studies have been conducted on English, these results are of great importance. This pair of *Solidarity* effects

indicate that while the assessment of likeability can be culturally specific, (at least some) likeability effects might be shared across (at least some European and North-American) cultures.

The two *Superiority* effects lined up less nicely across the two experiments. First, there was an effect of *Superiority* in the English shadowing dataset in *Extr. Prev.*, which could be observed for the prevoiced /b/ among females, and for both the plain /p/ and the prevoiced /b/ among males. This effect divided participants into two somewhat monolithic groups with little within-group variation (high-raters with closer-to-target productions and low-raters further away from the target). This could indicate that Superiority is factored in as a binary "pass-fail" filter—the participant either viewed the model talker as someone with authority or they did not. Again, this effect was observed in the less English-like condition (*Extr. Asp.*), which means that the participants' perception of the model talker was only factored in when the stimuli were atypical. The Hungarian *Superiority* effect was the only one that was observed in the more native-like condition.

Second, there was an effect of *Superiority* in the Hungarian *Extr. Prev.* condition for /b/ among males. Males who rated the model talker high on the relevant scales converged with her prevoiced /b/, while low-raters did not. Interestingly, this is the only effect that lines up with how likeability effects were previously observed in the literature (as convergence conditioned by a favorable perception of the model talker), whereas the rest of the effects exhibited in this study were a relationship between *disliking* and divergence. The difference between previous results and the present study could stem from the nature of the stimuli. While previous studies mostly operated with stimuli that resembled spontaneous productions or involved a real-life interlocution between speakers and interlocutors, there could have been interactions between the naturalness of the stimuli and likeability effects that remained hidden. For instance, it could be the case that for certain kinds of unnaturalistic stimuli, the spectrum of behaviors is restricted from divergence to no change, whereas participants' behavior tends to converge or not change when exposed to

naturalistic stimuli. This is consistent with the idea that convergence is automatic, and it is lack of accommodation or divergence that is socially mediated.

We must also speak of why likeability effects were mostly unseen in the more nativelike conditions (with the exception of the *Superiority* effect on /b/ accommodation in the Hungarian *Extr. Prev.* reading data). This might be most likely a result of noise. First, the two ends of the semantic differential scales were often flipped (4 out of 9 scales had the "positive" feature on the left, the other 5 had the "negative" feature on the left)—following e.g. Bauman (2013) in order to avoid certain task effects. Likely as a result of that, for certain participants the ratings responsible for a given measure did not always correlate with one another. For instance, one participant in the English dataset rated the model talker to be particularly unintelligent (1 on the unintelligent — intelligent scale), whereas she rated her to be fairly high status and organized (low status — high status: 8; unorganized — organized: 8), two features that also contribute to *Superiority*, and whose perception is supposed to correlate with perceived intelligence. While it is possible that this participant simply found the model talker to be organized, and of high status, but unintelligent, it is far more likely that the participant did not notice that some of these scales were "flipped". Another source of noise could have been that the likeability data was recorded at the very beginning of the study, and it is possible that 30–40 minutes of exposure to the model talker's voice in a fairly repetitive task might have shifted their views over the task. Because of these reasons, further evidence is needed to reinforce these effects, especially in a cross-linguistic context.

6.2.3.2 Gender

The effect of the speaker's gender has also been of particular interest to accommodation research, but recently more and more evidence has surfaced indicating that any putative gender effects must be specific to a given phonetic dimension, as accommodation along different phonetic variables is not necessarily correlated (Pardo et al., 2013b, 2017). In this study, true gender effects—i.e. ones that

could not be reduced to phonetic distance effects—were only present in the form of gender-specific likeability effects. The gender asymmetry in likeability effects largely surfaced as certain effects being specific to males only. While males in these environments diverged from the model talker if they perceived her unfavorably, females tended to either not change their productions or converge. Thus, these effects were a result of males' accommodation behavior being more sensitive to their *negative* opinions of the model talker, and their disliking of the model talker was more likely to show up as phonetic divergence than female participants' disliking was. This could potentially be related to males feeling less pressure to converge or at least to not diverge from the model talker. However, this theory would need to be tested against a more robust dataset, potentially with multiple model talkers of varying genders.

These results are somewhat surprising in the context of previous literature, which emphasizes the need for further testing. First, while females might sometimes be perceived to converge more than males by third-party listeners, at least in terms of fundamental frequency, males have been found to accommodate more to their interlocutors (Babel and Bulatov, 2012). Since the present study was also based on acoustic rather than perceptual measures of accommodation, the fact that males' accommodation patterns were potentially more susceptible to inhibitory effects of disliking is unexpected. Moreover, some researchers argue that females are more attentive to social information and are quicker to pick up on it (Nygaard and Queen, 2000). While this could indicate that they are also more *sensitive* to social factors as well, the present work suggests that better perception does not necessarily lead to greater sensitivity.

This work also contextualizes a gender effect that Nielsen (2008) found for VOT accommodation. In her study, males were more likely to converge with the model talker's 100 ms or longer VOT than females were. The fact that these effects were not replicated in the present study when the model talker's target /p/'s had even longer VOT (130 ms in *Extr. Asp.*) suggests that Nielsen's results were simply a distance effect. The most likely reason for her observing fewer females converging is

likely that the 100 ms VOT might still have been too short (too close to the females' baseline VOT) to elicit larger changes. The present study with its 130 ms VOT /p/ targets allowed even females, who tend to have longer VOT's on voiceless stops in English, to demonstrate convergence rather than matching (or being close to) the target right from the start.

There was another gender effect that could be explained as a phonetic distance effect. While both males and females converged with the aspirated /p/ targets in the English *Extr. Asp.* condition's shadowing task, males and females exhibited different trajectories in doing so. Females' productions were more or less the same across the six repetitions (indicating that they converged immediately with the model talker, and then maintained that closeness throughout the task). At the same time, male participants approached the target more gradually, getting closer and closer to it throughout the task. This could be explained as a distance effect stemming from the fact that in English females tend to have longer VOT's on voiceless stops than males. As a result, the /p/ representation of females in this study was quite close to the model talker's 130 ms VOT target /p/, and females did not need to change their productions by much in order to match the target. Since they only had to make a relatively small adjustment, they were able to do it all at once, whereas males had a longer way to go, and covered that distance gradually. A similar effect was found in Babel (2009) for vowel formants and modeled in *Chapter 5* of this work.

To summarize, the only gender effects in VOT accommodation that could not be reduced to phonetic distance effects were males' higher sensitivity to their own negative opinions of the model talker. However, these effects need to be tested further. Since the experiments in this dissertation (and to my knowledge all previous research on the relationship of gender and accommodation) were restricted to (cis-gendered) male and female participants, further research on other genders would provide a more complete picture. Moreover, participants in this study were explicitly informed of the model talker's gender, but in the future it would be important to tease apart effects of the interlocutor's gender and their *perceived* gender on the speaker's productions as well.

6.2.3.3 Ethnicity

This study also recorded information on how participants identified ethnically. To my knowledge, this is the first accommodation study to do so. While all the participants that have been reached by the Hungarian study identified the same way ("Hungarian"), the set of English-speaking participants was much more diverse. There were some effects of ethnicity recorded in the dataset, all in the /b/ dataset. In both conditions of Experiment 1, people of color (especially Black / African-American participants) had more prevoiced /b/ tokens than their white peers in the reading as well as in the shadowing data. This is in line with the results of Ryalls et al. (1997), who found that in a reading task Black participants prevoiced their voiced stops more often than white participants did.

In the *Extr. Asp.* reading dataset the biggest changes in /b/ productions also came from Black / African American participants, which is most likely a result of their baselines being more prevoiced. That is, by having more prevoicing initially, Black / African American participants simply had more of an opportunity to demonstrate convergence to the plain /b/ stimuli (15 ms VOT), because they were further away from this target at the start. This means that the present study found no ethnicity-based difference in participants' willingness or ability to accommodate to a single white female model talker.

This work, however, is only a start in studying the effects of ethnicity on accommodation. First, Experiments 1 and 2 only used a single white / Hungarian female model talker, respectively, and like in the case of gender, using a more ethnically diverse set of model talkers could uncover further effects. For example, it is possible that participants could exhibit different behaviors when interacting with model talkers or interlocutors of different ethnic backgrounds, which could be especially interesting in the case of multi-dialectal participants, who might exhibit dialectal code switching phenomena when accommodating. Similarly, since certain ethnic backgrounds have historically been associated with power and status, and power and status have been demonstrated

to have an effect on accommodation (e.g. Gregory and Webster, 1996), a larger and ethnically more diverse dataset can allow complex effects to surface, related to the speaker's ethnicity, the interlocutor's ethnicity, as well as interactions of the two. Moreover, since ethnically-based power structures can be culturally specific, ethnically diverse studies based in other cultures and languages could be especially informative.

6.2.4 Broader implications for accommodation

In this section I am going to briefly review a few less central insights and contributions of this work, relating to phonetic distance effects, parallel accommodation of contrasts as well as different types of long-term language change. In the last of these I am going to discuss how long-term language change phenomena like second dialect acquisition and chain shifts are still compatible with a pressure for phonetic consistency like *Maintain categories*.

6.2.4.1 Phonetic distance effects

This work provided further examples of phonetic distance effects—effects where different groups of participants exhibit different behaviors, which could mostly be derived from their differences in baseline values. Such effects have previously been described for age-grading in speech rate accommodation (where older speakers accommodate more to compensate for their initially slower speech rate; Szabó, 2019) and for vowel formant accommodation (where males, whose baseline was closer to the male model talker's productions, converged immediately, while females, who had a longer way to go, did so gradually; Babel, 2009).

In this dissertation, similar effects were observed for VOT accommodation. English-speaking females, who tend to produce voiceless stops with longer VOT than males, converged immediately to the aspirated /p/ in English *Extr. Asp.* during the shadowing task, while males did so gradually. Furthermore, in Experiment 2's reading task, males converged more to *Extr. Prev.'s*

plain /p/ target than females did, but their baselines were also more aspirated to begin with, and therefore they had more room to demonstrate convergence. Moreover, ethnicity effects (e.g. Black / African-American participants exhibiting the biggest changes in the English *Extr. Asp.* /b/ dataset) could also be reduced to a difference in baselines.

Distance effects showed up in both experiments, which indicates that such effects are universal rather than culturally or linguistically specific. However, there could also be cross-linguistic differences. Distance effects often involve one set of participants matching the target (and thus stopping to accommodate) sooner than others. The threshold for when participants "match the target" can potentially vary from language to language. When English-speaking participants were converging with plain /b/'s (*Extr. Asp.*, 15 ms VOT), participants who were further away than about 5 ms from that target almost exceptionlessly changed their productions to match the target, and as a result, got quite close to it in the POST Read. When Hungarian participants were exposed to their language's plain stop (/p/ in *Extr. Prev.*, 15 ms VOT), most participants within 10-20 ms from the target did not converge with the target. This suggests that English speakers probably had stricter "standards" for how closely a plain stop needed to be matched than Hungarian speakers did. This was presumably because of language-specific distributions and slightly different perceptual magnet effects in English and Hungarian. In sum, distance effects seem to be observable in multiple languages, albeit perceptual and distributional differences between languages can make their exact details language-specific.

In this study, certain social effects ended up more easily explained as distance effects stemming from systematically different baselines of various social groups. This means that without checking baseline differences, phonetic distance effects can potentially be misattributed to social factors, such as gender and ethnicity. Since accommodation is commonly measured as a difference in differences (i.e. how the difference between the speaker's production and their interlocutor's production changed over time), there is a chance that some of the social effects described in the

literature can be potentially explained by phonetic distance effects. An important line of future work could be teasing apart cases where accommodation is truly conditioned by a social effect (where certain social conditions or circumstances result in some participants being less willing or able to accommodate than others) and cases where phonetic distance effects can potentially play a part in (if not be the entire source of) effects that originally seemed to be conditioned by certain facets of the speaker's identity.

6.2.4.2 (Somewhat) Parallel accommodation of contrasts

Previous studies found evidence of members of natural classes being accommodated to in similar ways. These studies only exposed participants to one member of the natural class, but saw convergence for all members. For instance, exposure to /p/'s with long VOT results in longer VOT's for voiceless stops as a whole (Nielsen, 2008), and speakers generalize raised or lowered F1 of /ɛ/ to other mid vowels (Sanker, 2020).

The present study was somewhat different from these experiments, because it did not test the generalizability of certain productional shifts, but instead exposed participants to contrasts where one or both members of the contrast were shifted. While there was some indication of correlation between the treatment of /p/ and /b/, this was limited to only one condition of one of the experiments (in Hungarian *Extr. Asp.* read productions). In the end, this study found nowhere near as robust of an effect for parallel treatment of contrasts as previous studies found for parallel treatment of members of natural classes. However, these stimuli were not specifically designed to test parallel movement, and targeted experiments could be more informative.

There are some reasons not to expect /p/ and /b/ to move in perfect tandem in other conditions. First, the stimuli that were more native-like in the two experiments (*Extr. Asp.* in English, *Extr. Prev.* in Hungarian) were not equally close to everyday productions. For instance, in *Extr. Asp.* the plain /b/ (with 15 ms VOT) was much closer to a typical English /b/, than the aspirated /p/ was

(with 130 ms VOT). Therefore, even if participants maximally tried to match both, they might have reached the plain /b/ sooner (even immediately) than the extremely aspirated /p/, which could lead to the two categories not changing in tandem. The same is true for the Hungarian *Extr. Prev.* condition. In the English *Extr. Prev.* condition accommodation to the prevoiced /b/ target (-130 ms VOT) could have been impeded by articulatory difficulties (or difficulties with consistently producing highly prevoiced tokens over the course of the entire experiment). This meant that on top of the differing levels of reluctance of imitating a plain /p/, the success of imitating a prevoiced /b/ also varied from person to person, making correlation unlikely.

The predictions of the model presented in *Chapter 5* line up with this. While we might expect to see matching *directions* of changes with these experiments' stimuli (i.e. if the speaker converges with one sound, they will with the other as well), changes of matching *extents* would be unexpected from a modeling context as well. Since the two categories (/p/ and /b/) have their own distinct distributions, it would be exceedingly unlikely that stimuli were systematically equally far away from both members of the contrast for all (or even most) participants.

6.2.4.3 Accommodation and language change

In this section I am going to discuss topics of future research that could relate accommodation (and **Maintain categories**) to long-term language change. Since **Maintain categories** in particular is a force for phonetic stability and consistency, at first blush it could seem like it would not be compatible with sound change phenomena. For instance, acquiring a second dialect often requires the speaker to remap their phonetic categories to new areas of the phonetic space as a result of constant exposure to a new dialect, sometimes resulting in productions that could be atypical for the speaker's original dialect.

For example, Canadian English and Inland Northern American English both distinguish between the words *trap* and *lot*. In Canadian English *trap* is realized with a "backed/ æ/", i.e. as

/ˈtrap/ and *lot* is realized as /ˈlɒt/. At the same time, in Inland Northern American English, *trap* is realized as /ˈtrɪəp/, but the vowel in *lot* is fronted—i.e. /ˈlat/. This means that the Canadian TRAP-vowel overlaps with the Inland Northern American LOT-vowel, which can both be realized as a low central vowel /a/. If a speaker of Canadian English moves to Chicago, their process of acquiring the local dialect would require a drastic remapping of their vowel categories, much like what was required of participants in the English *Extreme Prevoicing* condition. However, in the present study no accommodation was found from English speakers towards a prevoicing contrast, which was explained by the constraint of **Maintain categories**. Therefore, it might seem at first as if **Maintain categories** was incompatible with the acquisition a a dialect that is very different from the speaker's native dialect(s).

However, in this section I will explain how a force like Maintain categories can still be compatible with more drastic instances of sound change—second dialect acquisition and chain shifts, respectively. Reconciling these phenomena with accommodation can provide topics for further research, especially from a modeling standpoint. In this section, I only attempt to offer explanations for how these results are compatible with the *propagation* stage of language change, and not touch on issues of the actuation problem—i.e. how language change is initiated and how some initial innovations build up enough systemic momentum to spread further while others revert and disappear.

The phenomenon that can be most straight-forwardly explained with this model is second dialect acquisition. As briefly mentioned before, a new token from the interlocutor being added to the speaker's representation (however typical or atypical) updates the probabilities associated with certain categories and phonetic properties. This capacity for updates is exactly what enables exemplar models to represent change in the individual productions. As a result of the updates, each token has the potential to shift the phonetic categorization algorithm that represents **Maintain categories** (modeled by a Bayesian probability of a token belonging to the category it was assigned

to based on lexical information, considering its phonetic properties). This means that even identical tokens can elicit a different reaction (and activation pattern) depending on when they are perceived—i.e. whether the given token is the first atypical token of its kind or if it is just one in a long line of similar (albeit atypical) tokens.

This is true even within shorter interactions, but effects are especially noticeable over the course of longer (or more) interactions. If a speaker is for example exposed to a new dialect (through a move or a change in the communities they interact with) they will start collecting more and more tokens from speakers of this new dialect. Some of these tokens might be initially atypical for the speaker, but over time these tokens will increase in number, and eventually shift the speaker's notion of typicality. As a result, the speaker will be influenced by (and could accommodate to) even variants that they initially had considered atypical.

By contrast, a phenomenon that could be tricky to integrate into the theory of accommodation outlined in this dissertation is chain shifts. While contrast preservation has often been cited as a motivation for chain shifts throughout history, chain shifts seem to be more compatible with a **Maintain contrasts** type restriction at first consideration. Maintain contrasts restrictions are satisfied by the contrast being maintained in a different part of phonetic space, while **Maintain categories** seem quite "all or nothing"—i.e. the fact that the other category is shifted as well does not make an atypical and unlikely token any more acceptable. However, I will briefly demonstrate that a mandated preservation of categories (Maintain categories) is indeed compatible with chain shifts.

The propagation of chain shifts can also be modeled with the series of small gradual changes to the speaker's representations mentioned above. Among chain shifts, pull chains and push chains are distinguished. A pull chain, where the initial change is a sound (sound A) moving *away from* other contrastive sounds, then another sound (sound B) moves into sound A's former space, is the easier one to explain. This should not be a problem for the present model, since just like in the case

of the simulation of English-speakers converging with extremely aspirated /p/'s in *Section 5.3.2*, the model presented here allows for categories shifting away from each other (to spaces not already occupied by other sounds).

For instance, let us assume the pull-shift interpretation of a small part of the Great Vowel Shift. During this historical vowel shift both /iː/ changed to /əɪ/, which later also changed to /aɪ/, in words like *time* /ˈtaɪm/ and /eː/ changed to /iː/ in words like *meet* /ˈmiːt/. The pull chain interpretation assumes that the former happened first. That is, /iː/ was more and more often realized as [əɪ] rather than as [iː] (i.e. in its initial spot). This is compatible with **Maintain categories**, because /iː/ is moving away from other sounds rather than being realized overlapping with another sound category. In fact, [iː] is seldom used at this stage, as neither *time* nor *meet* is realized with an [iː] particularly often. This creates a gap in the phonetic space (in the area of front high tense vowels), opening up room for the next step of the change, when /eː/ moves in to occupy this space. Provided that our mid-change hypothetical historical speaker is constantly exposed to interlocutors who have this second change, they should also be able to imitate it, since the realization of [iː] is not a common realization of any other category—the vowel in *time*, for example, is already realized in a different part of the phonetic space (as [ˈtəɪm]).

Push chains might be somewhat more problematic from a modeling point of view, since the movement of sound B into sound A's initial space is not preceded by sound A first moving out of it. While implementing the present model to push chains requires further work, it is important to remember that the model does not necessarily preclude sound categories from moving towards each other. It only limits convergence towards tokens that are atypical realizations of a category in a certain way such that they would be unlikely to belong to their lexical category (and might be more likely to be a realization of another category instead). With that in mind, it is conceivable that changes involving sounds slowly moving closer to each other (like push chains or mergers) can also be modeled with the tools presented in this work.

The main finding of this work is that there is a two-way interaction between phonological representations and accommodation. This manifests both as phonological representations limiting accommodation by requiring an adherence to certain phonetic properties of sound categories, and also as accommodation in turn potentially shifting these categories as well as the phonetic standards they impose on valid targets of accommodation. In this section I offered some theoretical speculations on how this theory of accommodation can be incorporated into a broader view of language change and pointed to areas of future research where the computational model from this work could be implemented for certain types of language change. However, these are only theoretical speculations, which must be backed up by either historical or longitudinal data, and further simulations are certainly needed. Moreover, while any theory of accommodation must be able to incorporate long-term language changes in some way, I only argued that the *propagation* of these changes is compatible with the presented model (since propagation must involve interaction, and interactions have the potential for accommodation). The issue of how such changes are *initiated* is a separate, much more difficult problem that is definitely worth pursuing, but lies beyond the scope of this work.

Appendix

Rating stimuli

According to experts, doing research online about mattresses is useful, but it is even more important to try the product in person before buying it. The most important aspect is that the mattress supports your body well, so that you don't sink into it too far, but that it is also not too hard. When choosing the perfect mattress, you should consider sleeping habits, body shape and body weight. Lie on the mattress, turn around, sit on it, and if you feel comfortable in every position, you have found the right one. As a mattress is not a cheap product, it is worthwhile to find out in advance what the store policy is if there is a problem later down the line. Bigger stores can replace a mattress even after several weeks or give full refunds. In the days after a purchase it is therefore advisable to thoroughly test the mattress to find out if you have made a good decision as soon as possible.

Figure A.1: Rating text for English participants: Shopping for a mattress

A szakértők szerint hasznos az interneten utánanézni a matracoknak, de még ennél is fontosabb, hogy a vásárlás előtt élőben kipróbáájuk a terméket. A legfontosabb szempont, hogy a matrac jól alátámassza a testünket, vagyis ne süppedjünk bele, de ne is legyen túl kemény. Az ideális matrac kiválasztásánál érdemes figyelembe venni az alvási szokásokat, a testalkatot és a testsúlyt is. Feküdjünk a matracra, forgolódjunk rajta, üljünk rá, és ha minden helyzetben kényelmesnek érezzük, akkor megtaláltuk az igazit. Mivel nem olcsó termékről van szó, nem árt előre kideríteni, hogy milyen lehetőséget kínál a bolt, ha később derülne fény valamilyen problémára. A nagyobb üzletek akár több hét után is kicserélik vagy visszaváltják a náluk vásárolt matracot. A vásárlás utáni napokban ezért érdemes alaposan letesztelni a terméket, hogy minél előbb kiderüljön, jól döntöttünk-e.

Figure A.2: Rating text for Hungarian participants: Matracvásárlás

English reading data: The Extreme Aspirating condition

	Estimate	Std. Error	Pr(> |t|)	
(Intercept)	92.909	13.968	<0.0001	***
Gender [male]	−3.057	34.274	0.9299	
Exposure [post]	2.328	4.110	0.5711	
Superiority	−2.325	2.314	0.3284	
Gender [male] × Exposure [post]	−38.651	10.153	0.0001	***
Gender [male] × Sup.	−0.019	4.987	0.9970	
Exposure [post] × Sup.	0.568	0.685	0.4067	
Gender [male] × Exposure [post] × Sup.	5.227	1.477	0.0004	***

Table A.1: LMER model of English read /p/ tokens' VOT in Extr. Asp.;
Threshold for significance (adjusted with Bonferroni correction): p<0.00625

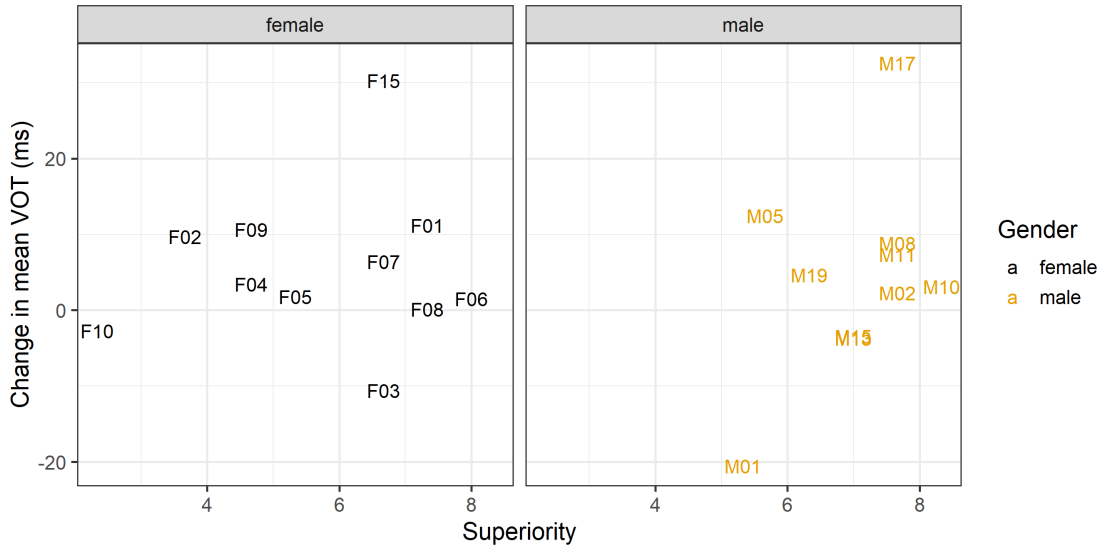

Figure A.3: Change in mean /p/ VOT in Extr. Asp. in the reading task by gender and Superiority rating

	Estimate	Std. Error	Pr(> |t|)	
(Intercept)	98.021	17.348	<0.0001	***
Gender [male]	−8.825	26.308	0.7412	
Exposure [post]	−5.794	5.186	0.2640	
Solidarity	−2.963	2.698	0.2867	
Gender [male] × Exposure [post]	7.749	7.900	0.3268	
Gender [male] × Sol.	0.450	4.097	0.9137	
Exposure [post] × Sol.	1.825	0.809	0.0243	.
Gender [male] × Exposure [post] × Sol.	−1.431	1.230	0.2448	

Table A.2: LMER model of English read /p/ tokens' VOT in Extr. Asp.;
Threshold for significance (adjusted with Bonferroni correction): p<0.00625

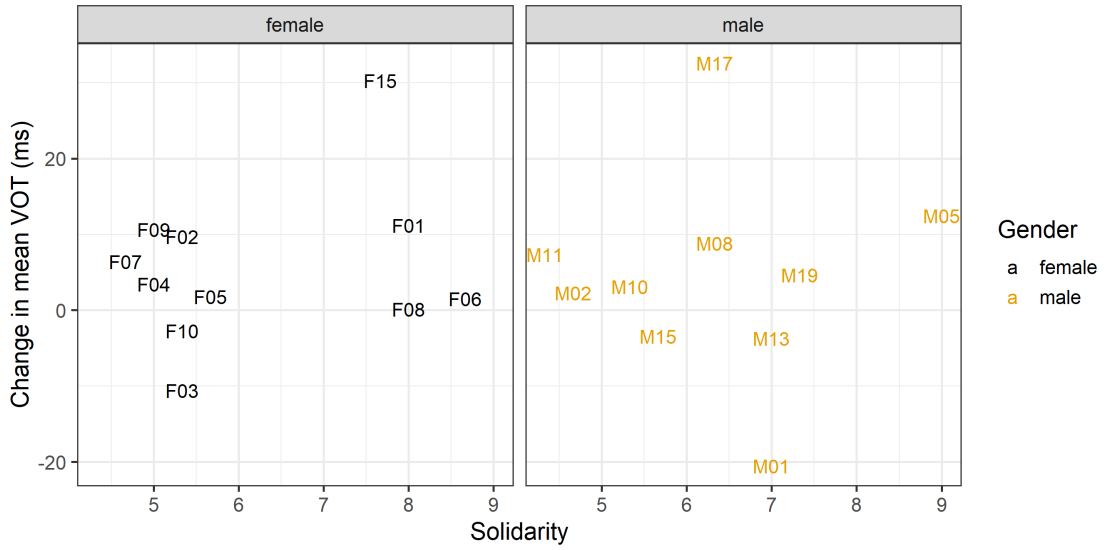

Figure A.4: Change in mean /p/ VOT in Extr. Asp. in the reading task by gender and Solidarity rating

	Estimate	Std. Error	Pr(> |t|)	
(Intercept)	94.646	13.881	<0.0001	***
Gender [male]	−50.520	47.257	0.2993	
Exposure [post]	−4.945	4.105	0.2285	
Dynamism	−2.868	2.510	0.2683	
Gender [male] × Exposure [post]	20.569	14.066	0.1438	
Gender [male] × Dyn.	7.715	7.872	0.3402	
Exposure [post] × Dyn.	1.999	0.747	0.0075	.
Gender [male] × Exposure [post] × Dyn.	−3.854	2.343	0.1002	

Table A.3: LMER model of English read /p/ tokens' VOT in Extr. Asp.;
Threshold for significance (adjusted with Bonferroni correction): p<0.00625

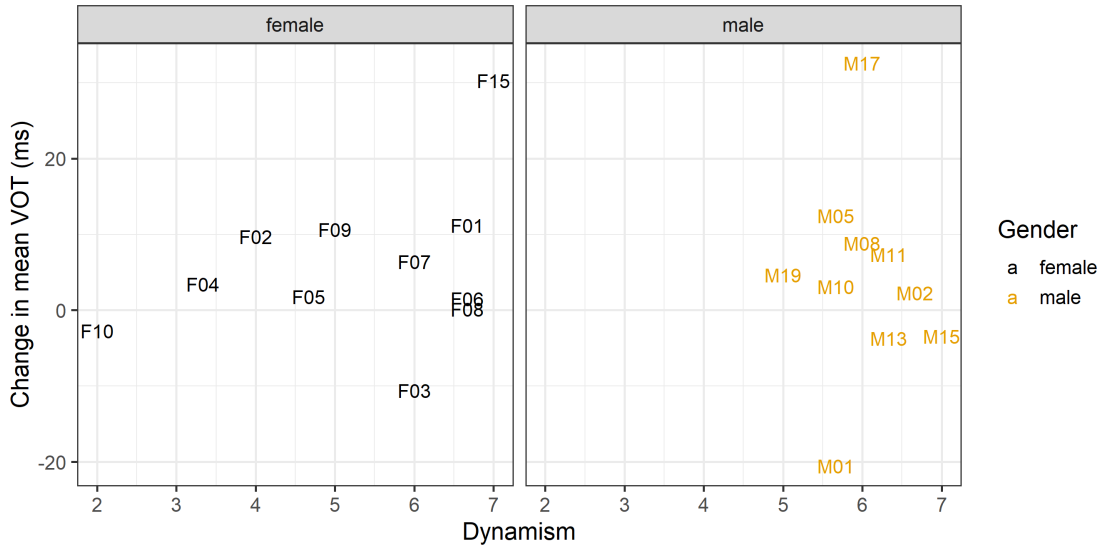

Figure A.5: Change in mean /p/ VOT in Extr. Asp. in the reading task by gender and Dynamism rating

	Estimate	Std. Error	Pr(> |t|)	
(Intercept)	−85.872	81.502	0.3190	
Gender [male]	−67.484	140.146	0.6410	
Exposure [post]	8.845	20.971	0.6730	
Superiority	8.050	12.421	0.5330	
Gender [male] × Exposure [post]	228.045	36.064	<0.0001	***
Gender [male] × Sup.	12.278	21.383	0.5790	
Exposure [post] × Sup.	−0.023	3.196	0.9940	
Gender [male] × Exposure [post] × Sup.	−33.931	5.503	<0.0001	***

Table A.4: LMER model of English read /b/ tokens' VOT in Extr. Asp.;
Threshold for significance (adjusted with Bonferroni correction): p<0.00625

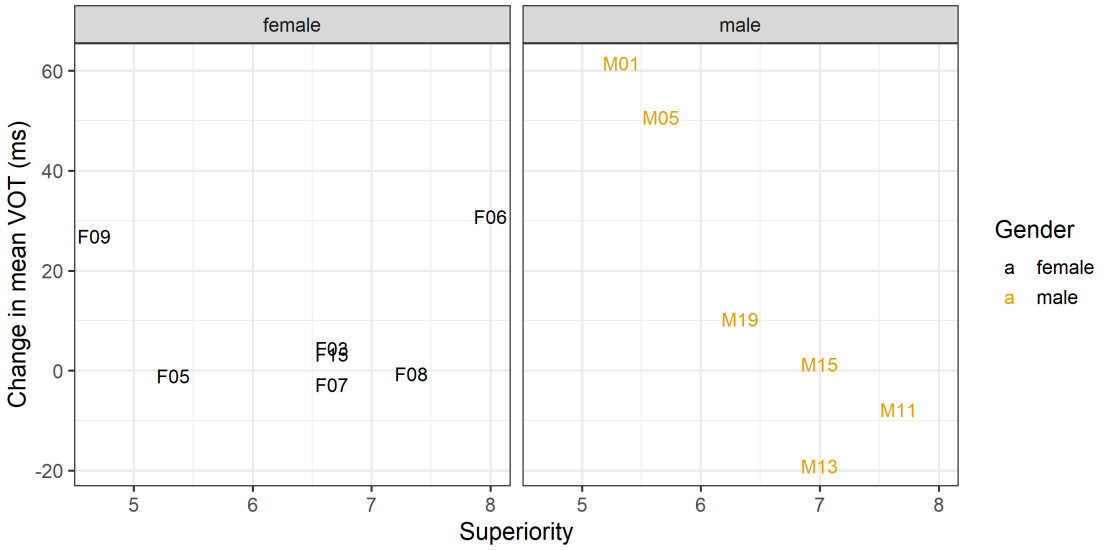

Figure A.6: Change in mean /b/ VOT in Extr. Asp. in the reading task by gender and Superiority rating

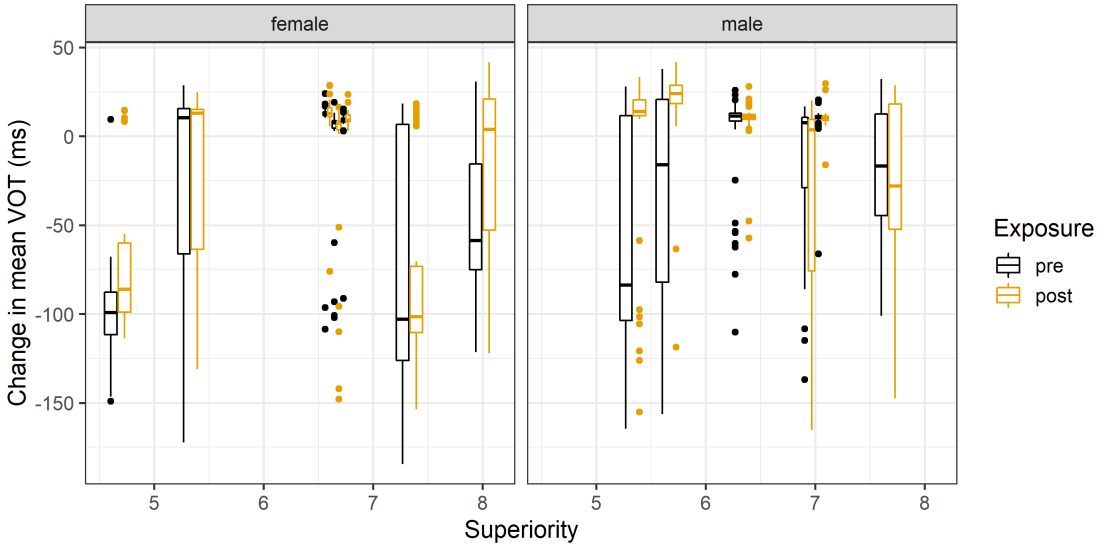

Figure A.7: /b/ VOT datapoints in Extr. Asp. in the reading task by gender and Superiority rating

	Estimate	Std. Error	Pr(> |t|)	
(Intercept)	−6.484	58.206	0.9137	
Gender [male]	21.107	89.748	0.8192	
Exposure [post]	−5.217	14.957	0.7273	
Solidarity	−4.240	8.813	0.6416	
Gender [male] × Exposure [post]	−59.783	23.067	0.0097	
Gender [male] × Sol.	−1.093	13.282	0.9362	
Exposure [post] × Sol.	2.164	2.265	0.3395	
Gender [male] × Exposure [post] × Sol.	9.913	3.414	0.0038	**

Table A.5: LMER model of English read /b/ tokens' VOT in Extr. Asp.; Threshold for significance (adjusted with Bonferroni correction): p<0.00625

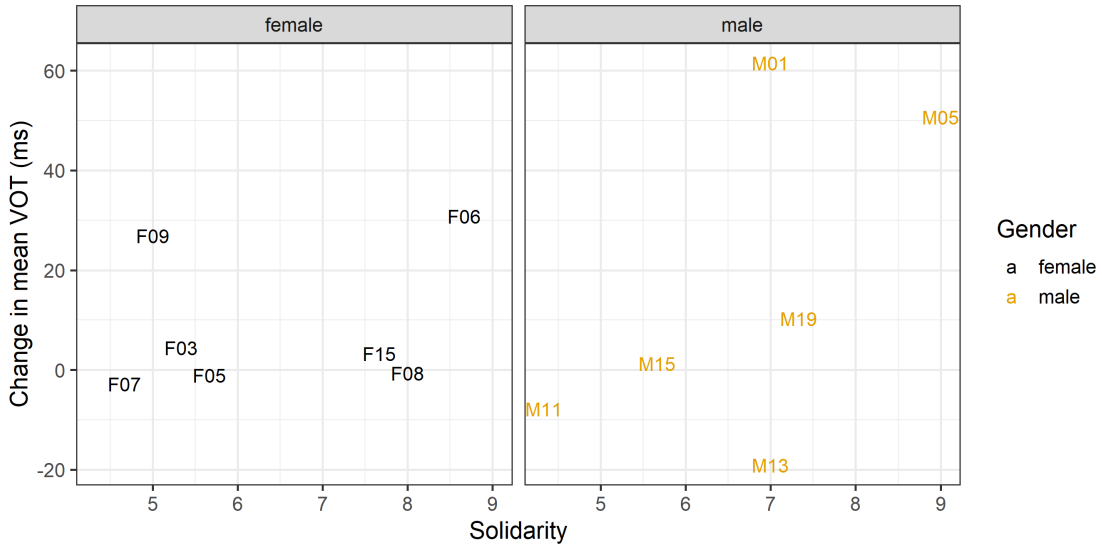

Figure A.8: Change in mean /b/ VOT in Extr. Asp. in the reading task by gender and Solidarity rating

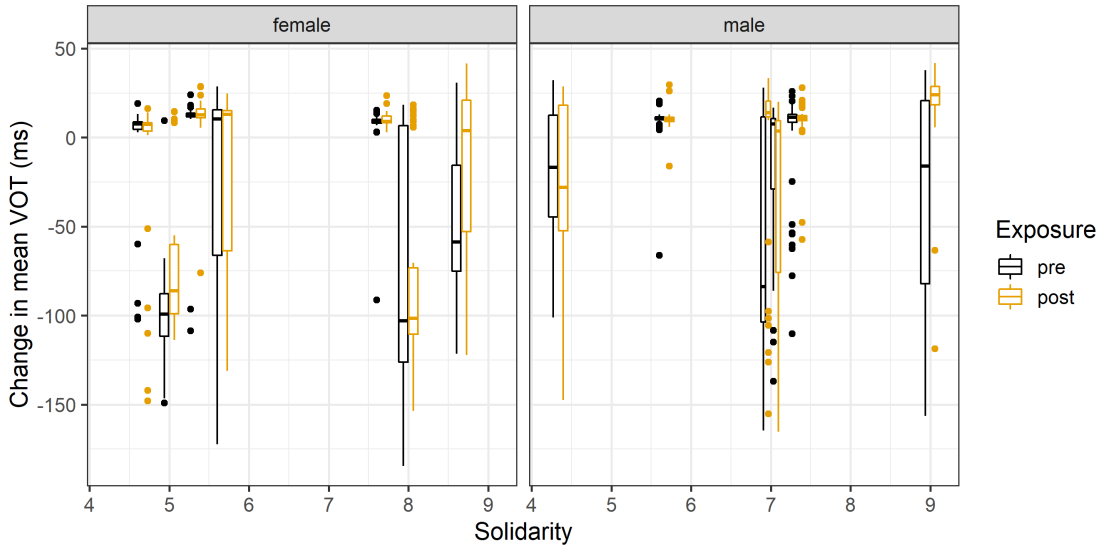

Figure A.9: Change in mean /b/ VOT in Extr. Asp. in the reading task by gender and Solidarity rating

	Estimate	Std. Error	Pr(> |t|)	
(Intercept)	−92.732	97.416	0.3652	
Gender [male]	−6.998	165.757	0.9672	
Exposure [post]	11.372	25.447	0.6551	
Dynamism	9.832	16.087	0.5557	
Gender [male] × Exposure [post]	**140.910**	**43.301**	**0.0012**	**
Gender [male] × Dyn.	3.253	27.438	0.9082	
Exposure [post] × Dyn.	−0.446	4.202	0.9155	
Gender [male] × Exposure [post] × Dyn.	**−22.237**	**7.168**	**0.0020**	**

Table A.6: LMER model of English read /b/ tokens' VOT in Extr. Asp.;
Threshold for significance (adjusted with Bonferroni correction): p<0.00625

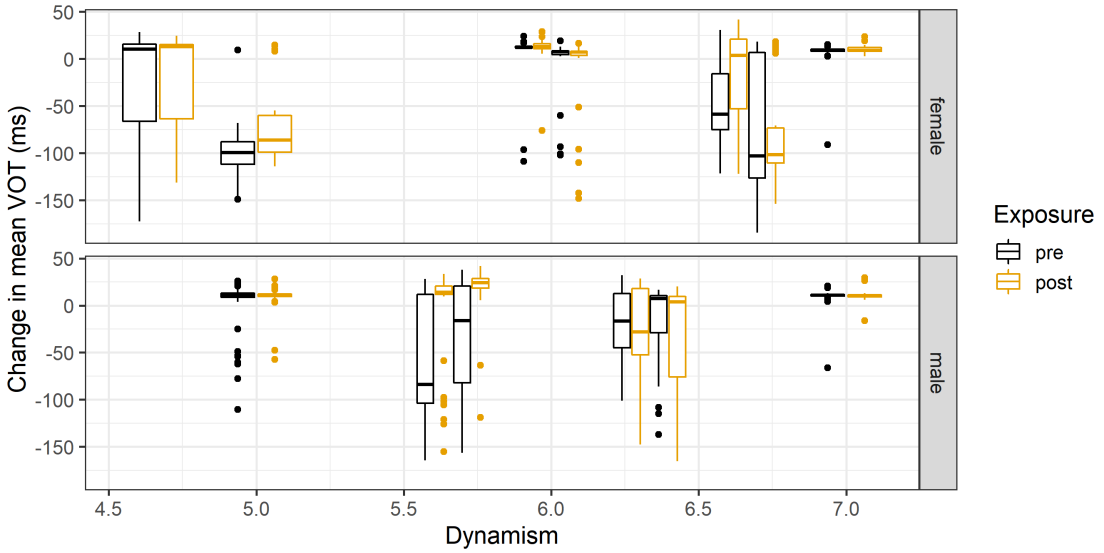

Figure A.10: Change in mean /p/ VOT in Extr. Asp. in the reading task by gender and Dynamism rating

English reading data: The Extreme Prevoicing condition

	Estimate	Std. Error	Pr(> |t|)	
(Intercept)	88.580	19.756	0.0003	***
Gender [male]	33.056	31.848	0.3139	
Exposure [post]	23.554	6.570	0.0003	***
Solidarity	−1.266	3.169	0.6945	
Gender [male] × Exposure [post]	6.407	10.605	0.5458	
Gender [male] × Sol.	−6.765	4.826	0.1790	
Exposure [post] × Sol.	−3.823	1.057	0.0003	***
Gender [male] × Exposure [post] × Sol.	−0.276	1.607	0.8635	

Table A.7: LMER model of English read /p/ tokens' VOT in Extr. Prev.;
Threshold for significance (adjusted with Bonferroni correction): p<0.00625

	Estimate	Std. Error	Pr(> |t|)	
(Intercept)	63.806	16.952	0.0015	**
Gender [male]	32.782	23.878	0.1878	
Exposure [post]	−2.014	5.337	0.7059	
Superiority	2.480	2.426	0.3210	
Gender [male] × Exposure [post]	16.639	7.560	0.0279	.
Gender [male] × Sup.	−6.884	3.448	0.0623	
Exposure [post] × Sup.	0.301	0.766	0.6944	
Gender [male] × Exposure [post] × Sup.	−2.167	1.091	0.0471	.

Table A.8: LMER model of English read /p/ tokens' VOT in Extr. Prev..; Threshold for significance (adjusted with Bonferroni correction): p<0.00625

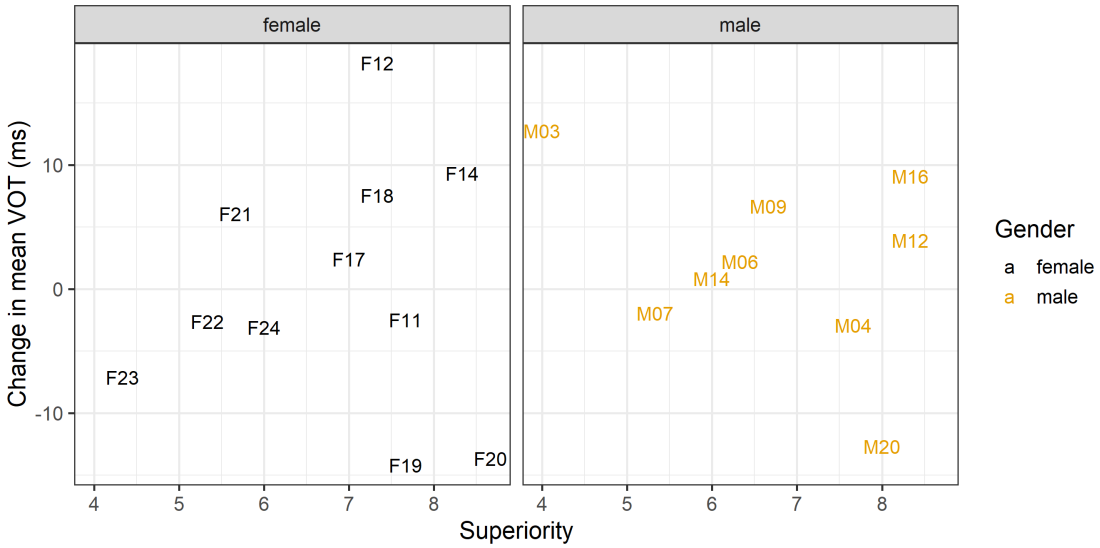

Figure A.11: Change in mean /p/ VOT in Extr. Prev. in the reading task by gender and Superiority rating

	Estimate	Std. Error	Pr(> |t|)	
(Intercept)	77.618	17.113	0.0003	***
Gender [male]	18.491	22.549	0.4237	
Exposure [post]	14.956	5.247	0.0044	**
Dynamism	0.543	2.866	0.8521	
Gender [male] × Exposure [post]	−8.910	6.934	0.1990	
Gender [male] × Dyn.	−5.264	3.684	0.1713	
Exposure [post] × Dyn.	−2.551	0.882	0.0039	**
Gender [male] × Exposure [post] × Dyn.	1.900	1.133	0.0936	

Table A.9: LMER model of English read /p/ tokens' VOT in Extr. Prev.; Threshold for significance (adjusted with Bonferroni correction): p<0.00625

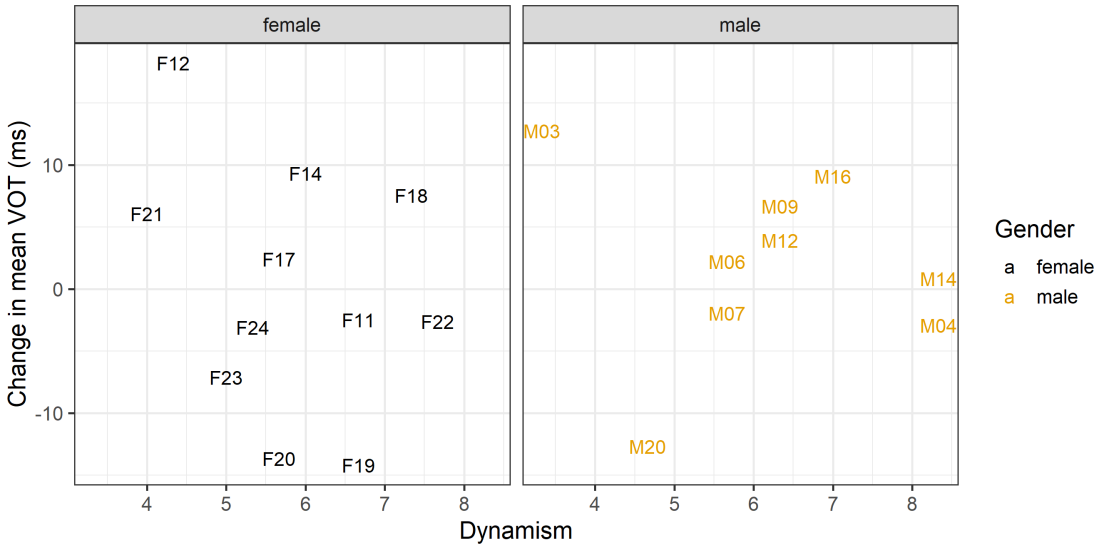

Figure A.12: Change in mean /p/ VOT in Extr. Prev. in the reading task by gender and Dynamism rating

	Estimate	Std. Error	Pr(> |t|)
(Intercept)	1.668	52.114	0.975
Gender [male]	52.294	73.620	0.487
Exposure [post]	13.330	13.834	0.335
Superiority	−2.948	7.479	0.698
Gender [male] × Exposure [post]	−16.597	19.549	0.396
Gender [male] × Sup.	−4.985	10.629	0.645
Exposure [post] × Sup.	−2.576	1.986	0.195
Gender [male] × Exposure [post] × Sup.	3.149	2.823	0.265

Table A.10: LMER model of reading /b/ tokens' VOT in Extr. Prev.; Threshold for significance (adjusted with Bonferroni correction): p<0.0071

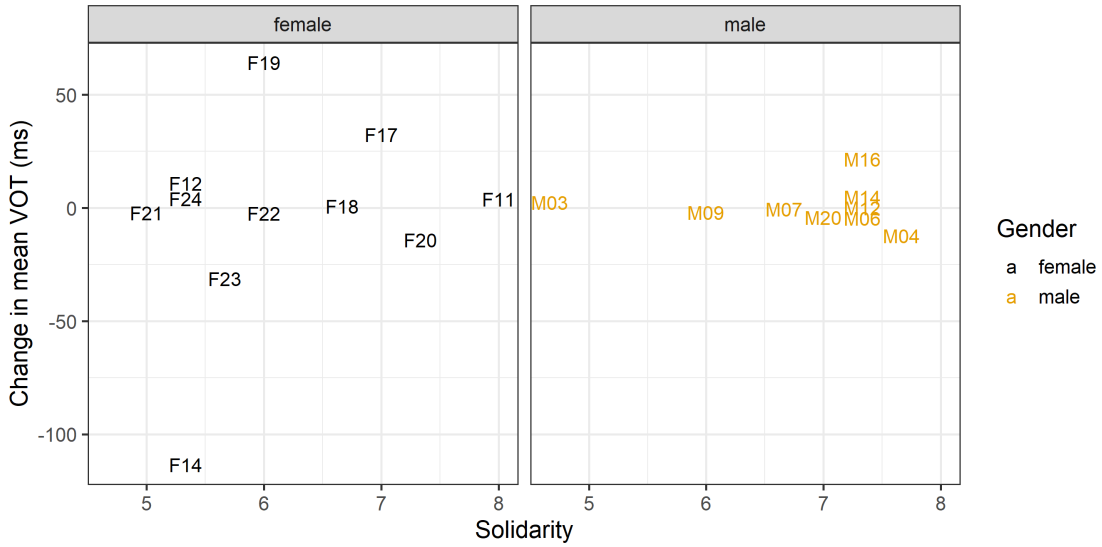

Figure A.13: Change in mean /b/ VOT in Extr. Prev. in the reading task by gender and Superiority rating

	Estimate	Std. Error	Pr(> |t|)	
(Intercept)	6.055	65.744	0.9277	
Gender [male]	59.408	106.214	0.5834	
Exposure [post]	−74.540	17.049	<0.0001	***
Solidarity	−3.996	10.568	0.7102	
Gender [male] × Exposure [post]	78.275	27.535	0.0045	**
Gender [male] × Sol.	−5.539	16.094	0.7350	
Exposure [post] × Sol.	11.416	2.740	<0.0001	***
Gender [male] × Exposure [post] × Sol.	−11.876	4.172	0.0045	**

Table A.11: LMER model of English read /b/ tokens' VOT in Extr. Prev.; Threshold for significance (adjusted with Bonferroni correction): p<0.00625

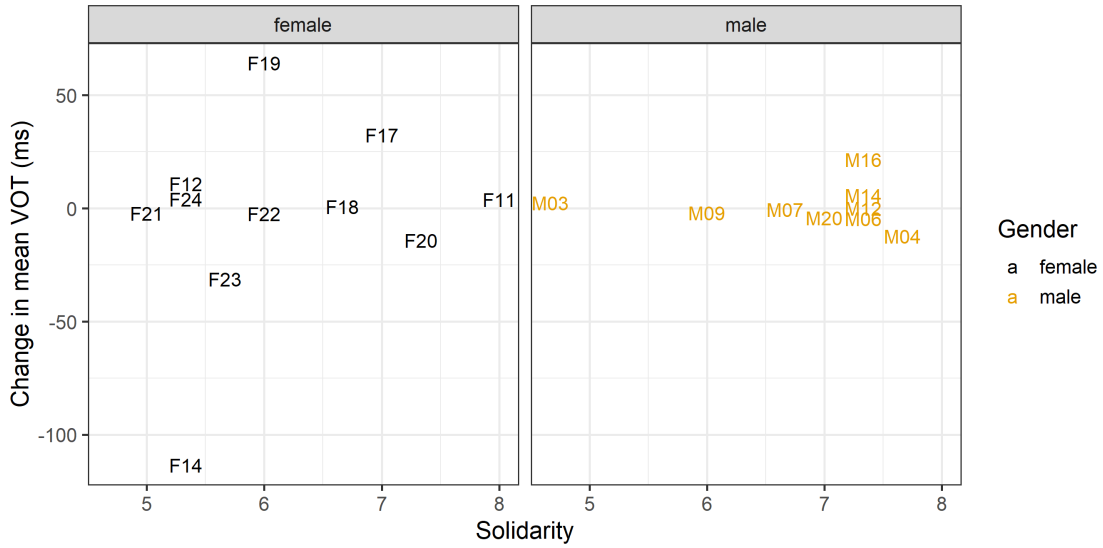

Figure A.14: Change in mean /b/ VOT in Extr. Prev. in the reading task by gender and Solidarity rating

	Estimate	Std. Error	Pr(> |t|)
(Intercept)	41.783	50.441	0.4192
Gender [male]	−1.408	66.655	0.9834
Exposure [post]	−23.400	13.649	0.0867
Dynamism	−10.311	8.472	0.2406
Gender [male] × Exposure [post]	22.954	18.019	0.2029
Gender [male] × Dyn.	3.862	10.889	0.7273
Exposure [post] × Dyn.	3.263	2.292	0.1546
Gender [male] × Exposure [post] × Dyn.	−3.095	2.943	0.2931

Table A.12: LMER model of English read /b/ tokens' VOT in Extr. Prev.; Threshold for significance (adjusted with Bonferroni correction): p<0.00625

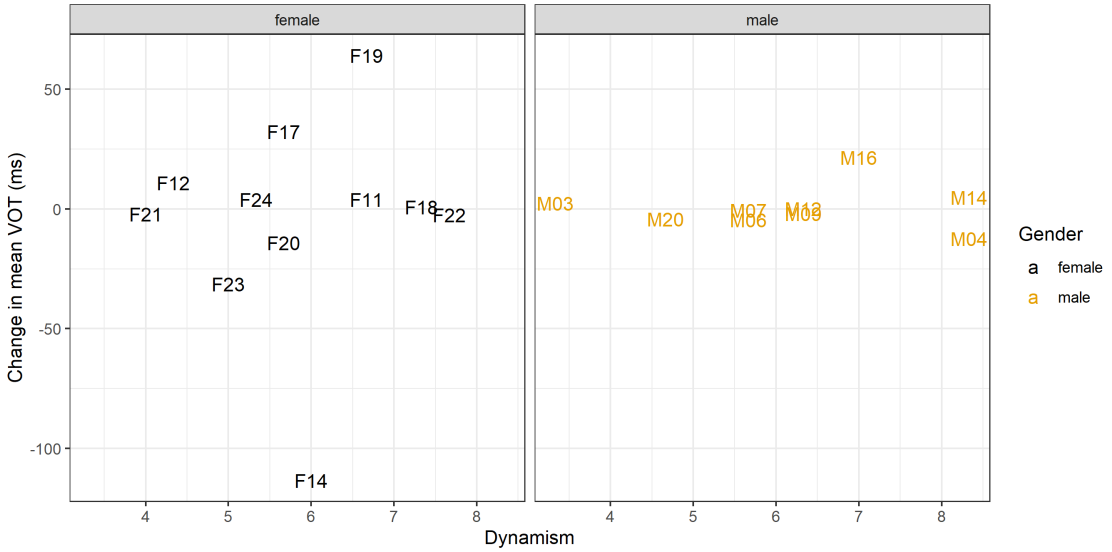

Figure A.15: Change in mean read /b/ VOT's in the English Extr. Prev. by gender and Dynamism rating

English shadowing data: The Extreme Aspirating condition

	Estimate	Std. Error	Pr(> |t|)	
(Intercept)	80.834	7.994	<0.0001	***
Solidarity	0.299	1.224	0.807	
Rep 2	5.718	11.088	0.6061	
Rep 3	2.729	11.076	0.8054	
Rep 4	−2.588	11.076	0.8153	
Rep 5	−5.594	11.076	0.6136	
Rep 6	3.996	11.076	0.7183	
Gender [male]	−7.074	11.942	0.5537	
Solidarity × Rep 2	−0.013	1.731	0.9939	
Solidarity × Rep 3	0.069	1.730	0.968	
Solidarity × Rep 4	1.475	1.730	0.3939	
Solidarity × Rep 5	1.921	1.730	0.267	
Solidarity × Rep 6	0.697	1.730	0.687	
Solidarity × Gender [male]	−0.560	1.860	0.7635	
Rep 2 × Gender [male]	25.569	16.871	0.1298	
Rep 3 × Gender [male]	35.606	16.863	0.0349	.
Rep 4 × Gender [male]	16.435	16.863	0.3299	
Rep 5 × Gender [male]	32.149	16.863	0.0567	
Rep 6 × Gender [male]	19.250	16.878	0.2542	

Table A.13: LMER model of English shadowed /p/ tokens' VOT in Extr. Asp. Part 1/2; Threshold for significance (adjusted with Bonferroni correction): p<0.0020

	Estimate	Std. Error	Pr(> |t|)
Solidarity × Rep 2 × Gender [male]	−3.876	2.627	0.1403
Solidarity × Rep 3 × Gender [male]	−4.488	2.626	0.0876
Solidarity × Rep 4 × Gender [male]	−2.510	2.626	0.3394
Solidarity × Rep 5 × Gender [male]	−4.870	2.626	0.0638
Solidarity × Rep 6 × Gender [male]	−3.389	2.627	0.1973

Table A.14: LMER model of English shadowed /p/ tokens' VOT in Extr. Asp.
Part 2/2; Threshold for significance (adjusted with Bonferroni correction): p<0.0020

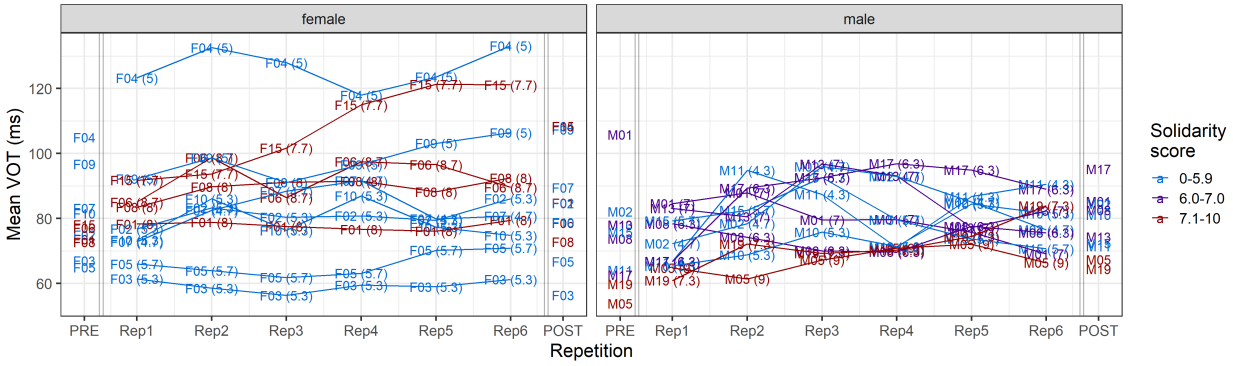

Figure A.16: Participants' shadowing trajectories by Solidarity rating in the English Extr. Asp.

	Estimate	Std. Error	Pr(> |t|)	
(Intercept)	87.977	6.449	<0.0001	***
Superiority	−0.914	1.043	0.3812	
Rep 2	10.193	8.910	0.2528	
Rep 3	0.858	8.841	0.9227	
Rep 4	2.674	8.841	0.7624	
Rep 5	1.282	8.841	0.8847	
Rep 6	7.053	8.841	0.4251	
Gender [male]	−3.647	15.443	0.8133	
Superiority × Rep 2	−0.785	1.484	0.5968	
Superiority × Rep 3	0.400	1.474	0.7860	
Superiority × Rep 4	0.685	1.474	0.6421	
Superiority × Rep 5	0.888	1.474	0.5469	
Superiority × Rep 6	0.225	1.474	0.8788	
Superiority × Gender [male]	−0.818	2.247	0.7161	
Rep 2 × Gender [male]	−27.121	21.857	0.2148	
Rep 3 × Gender [male]	−32.519	21.829	0.1365	
Rep 4 × Gender [male]	−20.240	21.829	0.3539	
Rep 5 × Gender [male]	−41.750	21.829	0.0560	
Rep 6 × Gender [male]	−49.687	21.888	0.0233	.

Table A.15: LMER model of English shadowed /p/ tokens' VOT in Extr. Asp. Part 1/2; Threshold for significance (adjusted with Bonferroni correction): p<0.0020

	Estimate	Std. Error	Pr(> |t|)	
Superiority × Rep 2 × Gender [male]	4.157	3.181	0.1914	
Superiority × Rep 3 × Gender [male]	5.594	3.176	0.0784	
Superiority × Rep 4 × Gender [male]	2.854	3.176	0.3690	
Superiority × Rep 5 × Gender [male]	5.999	3.176	0.0591	
Superiority × Rep 6 × Gender [male]	6.739	3.187	0.0346	.

Table A.16: LMER model of English shadowed /p/ tokens' VOT in Extr. Asp.
Part 2/2; Threshold for significance (adjusted with Bonferroni correction): p<0.0020

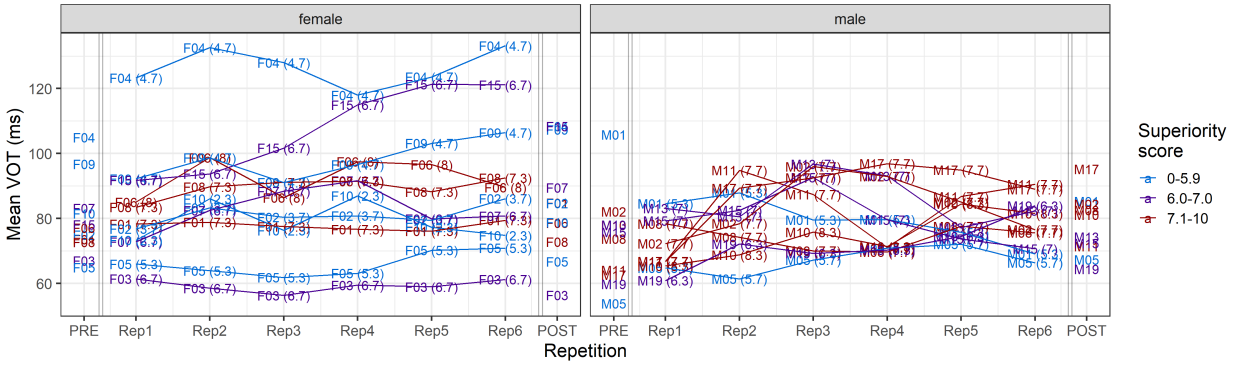

Figure A.17: Participants' shadowing trajectories by Superiority rating in the English Extr. Asp.

	Estimate	Std. Error	Pr(> |t|)	
(Intercept)	91.990	6.378	<0.0001	***
Dynamism	−1.758	1.124	0.1179	
Rep 2	12.261	8.798	0.1636	
Rep 3	0.081	8.725	0.9926	
Rep 4	−1.258	8.725	0.8854	
Rep 5	−3.866	8.725	0.6577	
Rep 6	1.551	8.725	0.8589	
Gender [male]	−66.228	21.158	0.0018	**
Dynamism × Rep 2	−1.245	1.599	0.4364	
Dynamism × Rep 3	0.584	1.588	0.7129	
Dynamism × Rep 4	1.494	1.588	0.3469	
Dynamism × Rep 5	1.946	1.588	0.2204	
Dynamism × Rep 6	1.289	1.588	0.4170	
Dynamism × Gender [male]	9.445	3.524	0.0074	**
Rep 2 × Gender [male]	−7.855	29.909	0.7929	
Rep 3 × Gender [male]	−42.253	29.888	0.1576	
Rep 4 × Gender [male]	−0.687	29.888	0.9817	
Rep 5 × Gender [male]	41.087	29.888	0.1694	
Rep 6 × Gender [male]	52.896	29.916	0.0772	

Table A.17: LMER model of English shadowed /p/ tokens' VOT in Extr. Asp.
Part 1/2; Threshold for significance (adjusted with Bonferroni correction): p<0.0020

	Estimate	Std. Error	Pr(> |t|)
Dynamism × Rep 2 × Gender [male]	1.639	4.982	0.7422
Dynamism × Rep 3 × Gender [male]	8.145	4.979	0.1020
Dynamism × Rep 4 × Gender [male]	0.043	4.979	0.9931
Dynamism × Rep 5 × Gender [male]	−6.794	4.979	0.1725
Dynamism × Rep 6 × Gender [male]	−9.273	4.982	0.0629

Table A.18: LMER model of English shadowed /p/ tokens' VOT in Extr. Asp.
Part 2/2; Threshold for significance (adjusted with Bonferroni correction): p<0.0020

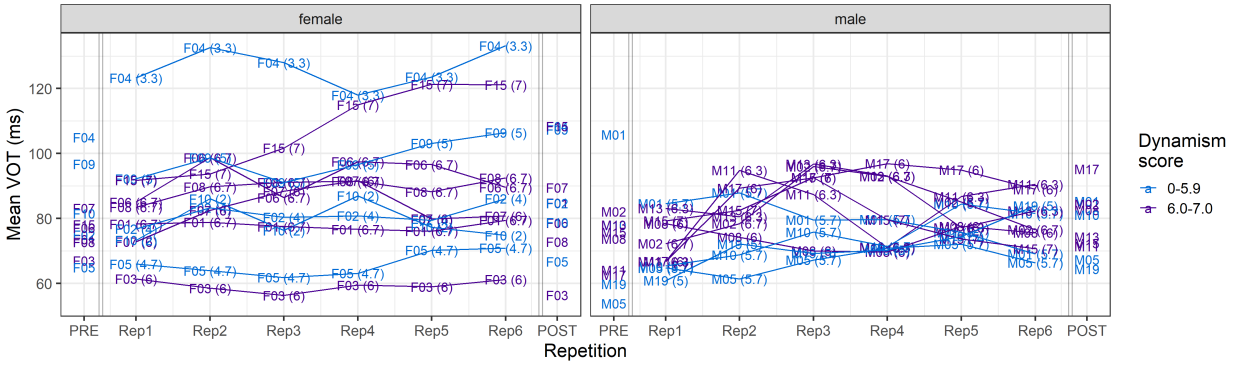

Figure A.18: Participants' shadowing trajectories by Dynamism rating in Extr. Asp.

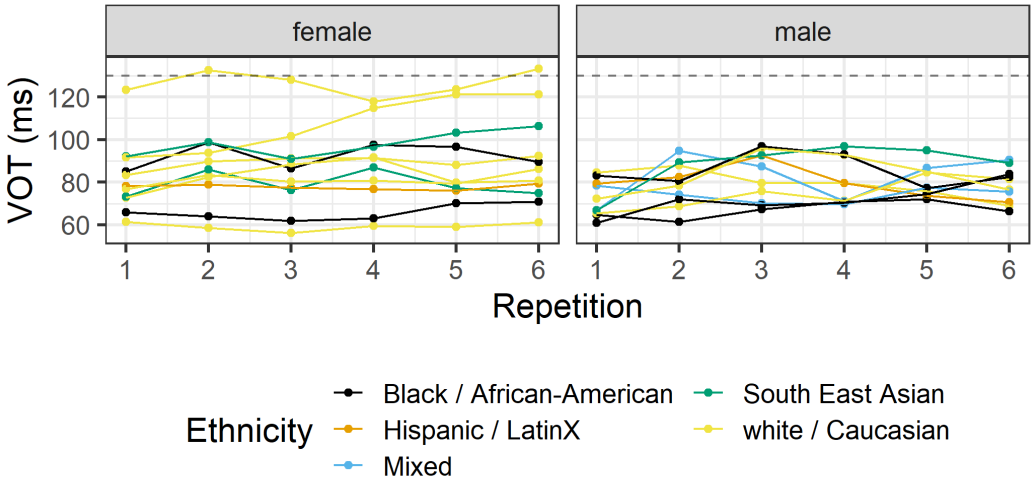

Figure A.19: Shadowing performance for /p/'s in Extr. Asp. by ethnicity
Gray dashed line indicates the model talker's VOT (130 ms)

	Estimate	Std. Error	Pr(> |t|)	
(Intercept)	2.947	8.523	0.7321	
Gender [male]	−18.636	12.108	0.1356	
Rep 2	−0.186	5.089	0.9708	
Rep 3	−2.468	5.089	0.6279	
Rep 4	−11.829	5.098	0.0204	.
Rep 5	−7.741	5.106	0.1297	
Rep 6	−9.710	5.114	0.0578	
Gender [male] × Rep 2	−7.654	7.383	0.3000	
Gender [male] × Rep 3	−5.644	7.383	0.4447	
Gender [male] × Rep 4	8.488	7.388	0.2507	
Gender [male] × Rep 5	4.648	7.407	0.5304	
Gender [male] × Rep 6	19.248	7.400	0.0094	.

Table A.19: LMER model of English shadowed /b/ productions in the Extreme Aspiration condition (target: 15 ms); Threshold for significance (adjusted with Bonferroni correction): p<0.0042

	Estimate	Std. Error	Pr(> |t|)	
(Intercept)	−4.913	23.086	0.8315	
Solidarity	0.092	3.478	0.9789	
Rep 2	−8.634	32.429	0.7901	
Rep 3	1.665	32.429	0.9591	
Rep 4	35.134	32.429	0.2789	
Rep 5	−28.711	32.429	0.3762	
Rep 6	−47.623	32.534	0.1435	
Gender [male]	7.054	35.642	0.8432	
Solidarity × Rep 2	1.200	4.911	0.8069	
Solidarity × Rep 3	−1.165	4.911	0.8126	
Solidarity × Rep 4	−8.180	4.911	0.0961	
Solidarity × Rep 5	2.836	4.911	0.5638	
Solidarity × Rep 6	5.383	4.922	0.2744	
Solidarity × Gender [male]	−5.916	5.286	0.2633	
Rep 2 × Gender [male]	−95.024	50.161	0.0584	
Rep 3 × Gender [male]	−44.862	50.161	0.3713	
Rep 4 × Gender [male]	−104.113	50.161	0.0382	.
Rep 5 × Gender [male]	−50.943	50.600	0.3143	
Rep 6 × Gender [male]	−21.386	50.229	0.6704	

Table A.20: LMER model of English shadowed /b/ tokens' VOT in Extr. Asp., those who "had room" to accommodate Part 1/2; Threshold for significance (adjusted with Bonferroni correction): p<0.0020

	Estimate	Std. Error	Pr(> |t|)	
Solidarity × Rep 2 × Gender [male]	12.363	7.431	0.0965	
Solidarity × Rep 3 × Gender [male]	7.679	7.431	0.3017	
Solidarity × Rep 4 × Gender [male]	20.278	7.431	0.0065	.
Solidarity × Rep 5 × Gender [male]	10.773	7.495	0.1509	
Solidarity × Rep 6 × Gender [male]	8.191	7.438	0.2711	

Table A.21: LMER model of English shadowed /b/ tokens' VOT in Extr. Asp., those who "had room" to accommodate Part 2/2; Threshold for significance (adjusted with Bonferroni correction): p<0.0020

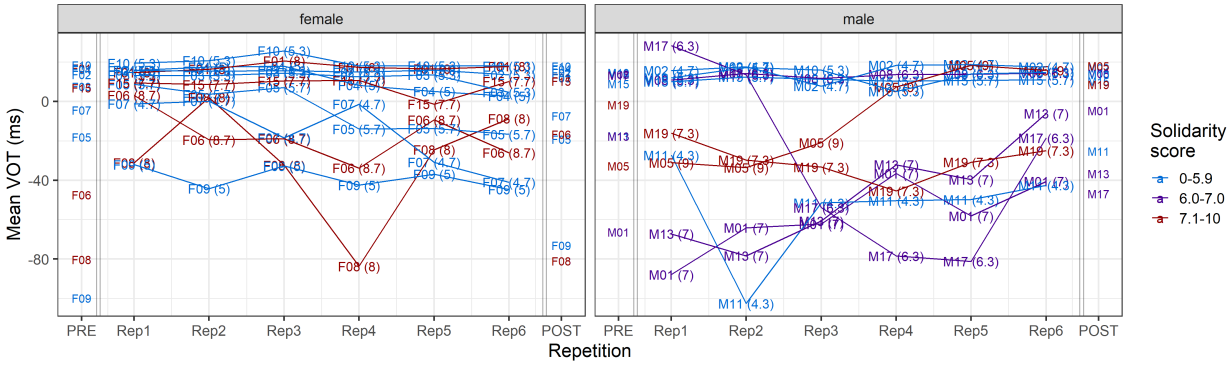

Figure A.20: Participants' /b/ shadowing trajectories by Solidarity rating in Extr. Asp.

	Estimate	Std. Error	Pr(> |t|)	
(Intercept)	−33.690	32.923	0.3064	
Superiority	4.537	5.005	0.3649	
Rep 2	−19.135	46.364	0.6799	
Rep 3	22.845	46.364	0.6223	
Rep 4	29.656	46.364	0.5226	
Rep 5	−21.791	46.364	0.6384	
Rep 6	−26.187	46.763	0.5756	
Gender [male]	−126.666	56.722	0.0257	.
Superiority × Rep 2	2.813	7.067	0.6907	
Superiority × Rep 3	−4.427	7.067	0.5312	
Superiority × Rep 4	−7.274	7.067	0.3035	
Superiority × Rep 5	1.746	7.067	0.8048	
Superiority × Rep 6	2.027	7.119	0.7759	
Superiority × Gender [male]	14.437	8.647	0.0953	
Rep 2 × Gender [male]	199.176	79.931	0.0129	.
Rep 3 × Gender [male]	78.686	79.931	0.3251	
Rep 4 × Gender [male]	143.407	79.931	0.0731	
Rep 5 × Gender [male]	169.751	80.391	0.0349	.
Rep 6 × Gender [male]	181.646	80.162	0.0236	.

Table A.22: LMER model of English shadowed /b/ tokens' VOT in Extr. Asp., those who "had room" to accommodate Part 1/2; Threshold for significance (adjusted with Bonferroni correction): p<0.0020

	Estimate	Std. Error	Pr(> |t|)	
Superiority × Rep 2 × Gender [male]	−32.430	12.190	0.0079	.
Superiority × Rep 3 × Gender [male]	−11.100	12.190	0.3627	
Superiority × Rep 4 × Gender [male]	−17.449	12.190	0.1526	
Superiority × Rep 5 × Gender [male]	−22.686	12.263	0.0646	
Superiority × Rep 6 × Gender [male]	−22.521	12.220	0.0656	

Table A.23: LMER model of English shadowed /b/ tokens' VOT in Extr. Asp., those who "had room" to accommodate Part 2/2; Threshold for significance (adjusted with Bonferroni correction): p<0.0020

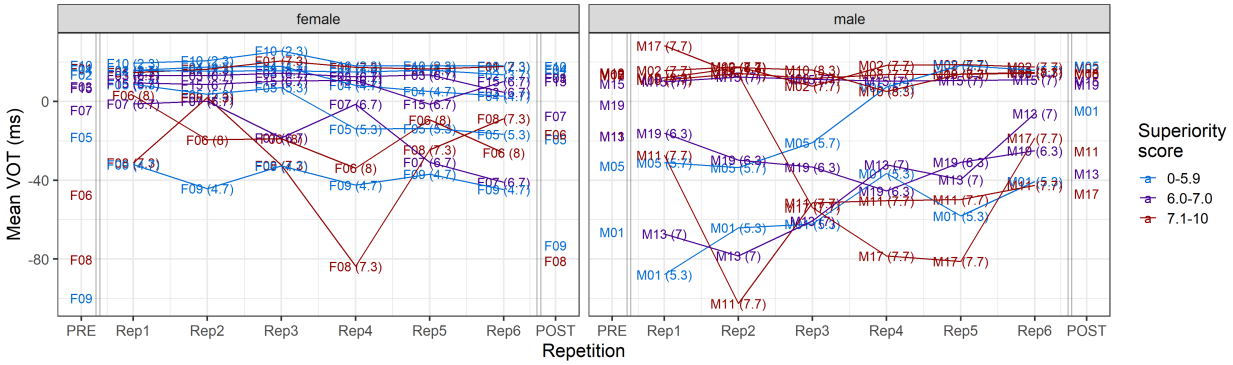

Figure A.21: Participants' /b/ shadowing trajectories by Superiority rating in Extr. Asp.

	Estimate	Std. Error	Pr(> |t|)	
(Intercept)	−33.690	32.923	0.3064	
Superiority	4.537	5.005	0.3649	
Rep 2	−19.135	46.364	0.6799	
Rep 3	22.845	46.364	0.6223	
Rep 4	29.656	46.364	0.5226	
Rep 5	−21.791	46.364	0.6384	
Rep 6	−26.187	46.763	0.5756	
Gender [male]	−126.666	56.722	0.0257	.
Superiority × Rep 2	2.813	7.067	0.6907	
Superiority × Rep 3	−4.427	7.067	0.5312	
Superiority × Rep 4	−7.274	7.067	0.3035	
Superiority × Rep 5	1.746	7.067	0.8048	
Superiority × Rep 6	2.027	7.119	0.7759	
Superiority × Gender [male]	14.437	8.647	0.0953	
Rep 2 × Gender [male]	199.176	79.931	0.0129	.
Rep 3 × Gender [male]	78.686	79.931	0.3251	
Rep 4 × Gender [male]	143.407	79.931	0.0731	
Rep 5 × Gender [male]	169.751	80.391	0.0349	.
Rep 6 × Gender [male]	181.646	80.162	0.0236	.

Table A.24: LMER model of English shadowed /b/ tokens' VOT in Extr. Asp., those who "had room" to accommodate Part 1/2; Threshold for significance (adjusted with Bonferroni correction): p<0.0020

	Estimate	Std. Error	Pr(> |t|)	
Superiority × Rep 2 × Gender [male]	−32.430	12.190	0.0079	.
Superiority × Rep 3 × Gender [male]	−11.100	12.190	0.3627	
Superiority × Rep 4 × Gender [male]	−17.449	12.190	0.1526	
Superiority × Rep 5 × Gender [male]	−22.686	12.263	0.0646	
Superiority × Rep 6 × Gender [male]	−22.521	12.220	0.0656	

Table A.25: LMER model of English shadowed /b/ tokens' VOT in Extr. Asp., those who "had room" to accommodate Part 2/2; Threshold for significance (adjusted with Bonferroni correction): p<0.0020

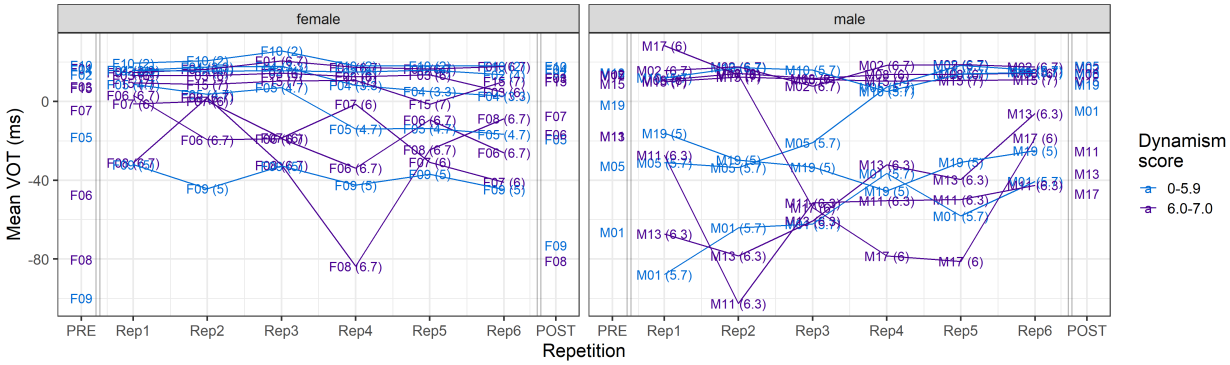

Figure A.22: Participants' /b/ shadowing trajectories by Dynamism rating in Extr. Asp.

English shadowing data: The Extreme Prevoicing condition

	Estimate	Std. Error	Pr(> |t|)	
(Intercept)	99.665	10.037	<0.0001	***
Solidarity	−3.043	1.591	0.0559	
Rep 2	8.174	13.985	0.559	
Rep 3	14.641	13.951	0.2941	
Rep 4	14.795	13.911	0.2877	
Rep 5	7.060	13.996	0.614	
Rep 6	25.243	14.120	0.074	
Gender [male]	13.116	15.778	0.4059	
Solidarity × Rep 2	−1.523	2.242	0.497	
Solidarity × Rep 3	−2.779	2.239	0.2148	
Solidarity × Rep 4	−2.419	2.228	0.2779	
Solidarity × Rep 5	−1.433	2.240	0.5225	
Solidarity × Rep 6	−4.306	2.260	0.0569	
Solidarity × Gender [male]	−3.919	2.394	0.1018	
Rep 2 × Gender [male]	−5.528	22.487	0.8059	
Rep 3 × Gender [male]	−10.524	22.592	0.6414	
Rep 4 × Gender [male]	2.990	22.216	0.893	
Rep 5 × Gender [male]	5.644	22.282	0.8001	
Rep 6 × Gender [male]	−31.342	22.365	0.1613	

Table A.26: LMER model of English shadowed /p/ tokens' VOT in Extr. Prev. Part 1/2; Threshold for significance (adjusted with Bonferroni correction): p<0.0020

	Estimate	Std. Error	Pr(> |t|)
Solidarity × Rep 2 × Gender [male]	0.761	3.409	0.8234
Solidarity × Rep 3 × Gender [male]	1.723	3.422	0.6146
Solidarity × Rep 4 × Gender [male]	−0.209	3.367	0.9504
Solidarity × Rep 5 × Gender [male]	−0.529	3.380	0.8756
Solidarity × Rep 6 × Gender [male]	5.202	3.393	0.1255

Table A.27: LMER model of English shadowed /p/ tokens' VOT in Extr. Prev.
Part 2/2: Threshold for significance (adjusted with Bonferroni correction): p<0.0020

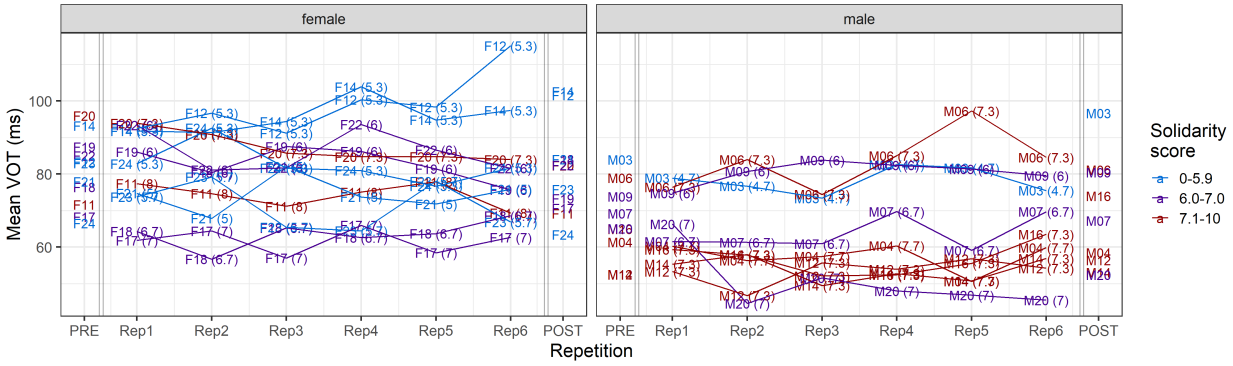

Figure A.23: Participants' /p/ shadowing trajectories by Solidarity rating in Extr. Prev.

	Estimate	Std. Error	Pr(> |t|)	
(Intercept)	64.797	8.129	<0.0001	***
Superiority	2.350	1.149	0.0409	.
Rep 2	3.314	11.210	0.7675	
Rep 3	−8.026	11.155	0.4720	
Rep 4	−11.402	11.203	0.3090	
Rep 5	2.574	11.217	0.8185	
Rep 6	−8.436	11.275	0.4544	
Gender [male]	25.183	11.276	0.0257	.
Superiority × Rep 2	−0.663	1.609	0.6803	
Superiority × Rep 3	0.827	1.602	0.6059	
Superiority × Rep 4	1.633	1.607	0.3096	
Superiority × Rep 5	−0.648	1.610	0.6874	
Superiority × Rep 6	1.014	1.616	0.5306	
Superiority × Gender [male]	−6.007	1.631	0.0002	***
Rep 2 × Gender [male]	10.533	15.928	0.5085	
Rep 3 × Gender [male]	7.086	15.932	0.6566	
Rep 4 × Gender [male]	32.641	15.835	0.0394	.
Rep 5 × Gender [male]	13.457	15.839	0.3957	
Rep 6 × Gender [male]	17.830	15.948	0.2637	

Table A.28: LMER model of English shadowed /p/ tokens' VOT in Extr. Prev. Part 1/2; Threshold for significance (adjusted with Bonferroni correction): p<0.0020

	Estimate	Std. Error	Pr(> |t|)	
Superiority × Rep 2 × Gender [male]	−1.731	2.298	0.4515	
Superiority × Rep 3 × Gender [male]	−1.122	2.298	0.6254	
Superiority × Rep 4 × Gender [male]	−4.768	2.284	0.0370	.
Superiority × Rep 5 × Gender [male]	−1.797	2.288	0.4325	
Superiority × Rep 6 × Gender [male]	−2.363	2.300	0.3045	

Table A.29: LMER model of English shadowed /p/ tokens' VOT in Extr. Prev. Part 2/2; Threshold for significance (adjusted with Bonferroni correction): p<0.0020

	Estimate	Std. Error	Pr(> |t|)	
(Intercept)	81.182	8.348	<0.0001	***
Dynamism	−0.048	1.383	0.9722	
Rep 2	15.198	11.616	0.1909	
Rep 3	5.370	11.513	0.6410	
Rep 4	−0.787	11.604	0.9459	
Rep 5	4.429	11.693	0.7049	
Rep 6	21.485	11.810	0.0691	
Gender [male]	8.929	10.709	0.4045	
Dynamism × Rep 2	−2.799	1.952	0.1518	
Dynamism × Rep 3	−1.321	1.933	0.4945	
Dynamism × Rep 4	0.106	1.948	0.9565	
Dynamism × Rep 5	−1.068	1.958	0.5855	
Dynamism × Rep 6	−3.888	1.979	0.0496	.
Dynamism × Gender [male]	−3.963	1.757	0.0242	.

Table A.30: LMER model of English shadowed /p/ tokens' VOT in Extr. Prev. Part 1/2; Threshold for significance (adjusted with Bonferroni correction): p<0.0020

	Estimate	Std. Error	Pr(> |t|)	
Rep 2 × Gender [male]	−28.565	15.256	0.0613	
Rep 3 × Gender [male]	−15.020	15.185	0.3227	
Rep 4 × Gender [male]	0.378	15.128	0.9801	
Rep 5 × Gender [male]	−0.267	15.225	0.9860	
Rep 6 × Gender [male]	−34.710	15.326	0.0237	.
Dynamism × Rep 2 × Gender [male]	4.549	2.503	0.0694	
Dynamism × Rep 3 × Gender [male]	2.385	2.483	0.3369	
Dynamism × Rep 4 × Gender [male]	−0.060	2.477	0.9806	
Dynamism × Rep 5 × Gender [male]	0.289	2.493	0.9077	
Dynamism × Rep 6 × Gender [male]	6.046	2.509	0.0161	.

Table A.31: LMER model of English shadowed /p/ tokens' VOT in Extr. Prev. Part 2/2; Threshold for significance (adjusted with Bonferroni correction): p<0.0020

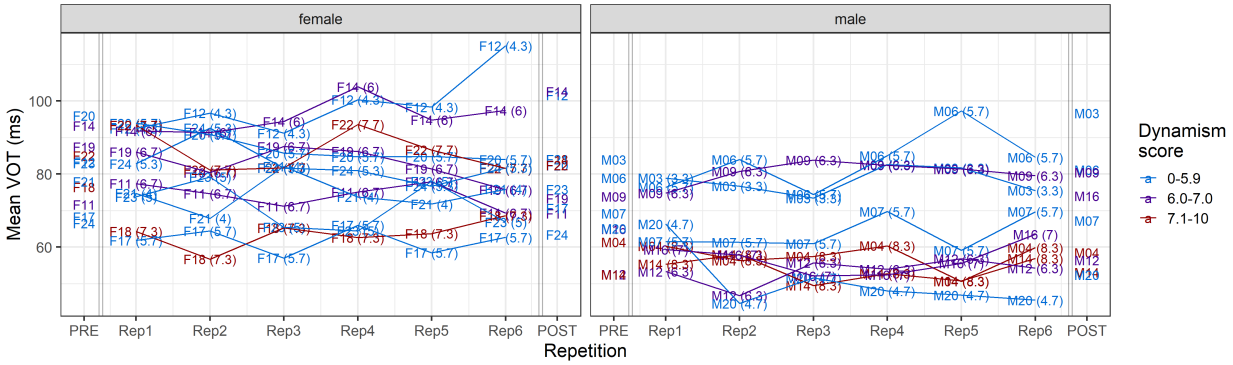

Figure A.24: Participants' /p/ shadowing trajectories by Dynamism rating in Extr. Prev.

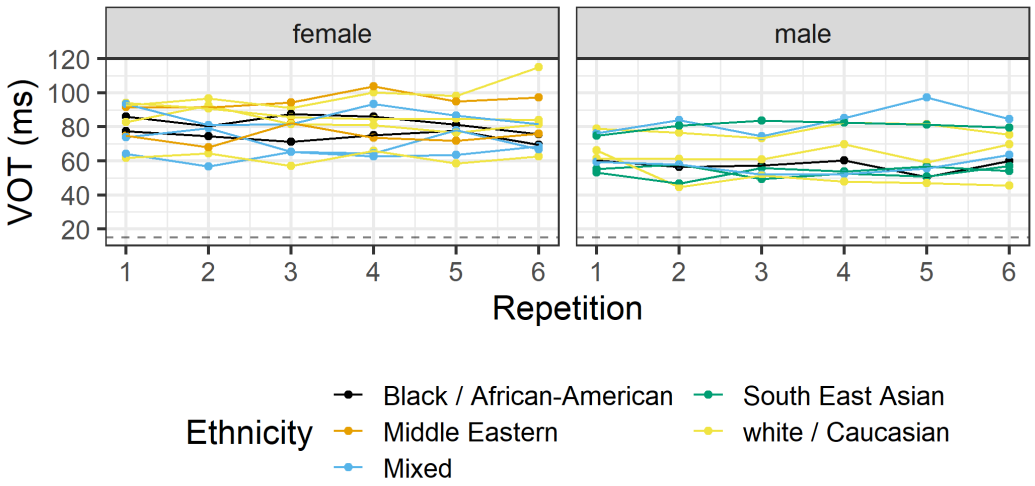

Figure A.25: Shadowing performance for /p/'s in Extr. Prev. by ethnicity
Gray dashed line indicates the model talker's VOT (15 ms)

	Estimate	Std. Error	Pr(> |t|)	
(Intercept)	−22.043	13.850	0.1271	
Gender [male]	14.826	20.556	0.4792	
Rep 2	−7.755	4.536	0.0875	
Rep 3	−12.900	4.522	0.0044	.
Rep 4	−3.292	4.529	0.4674	
Rep 5	−6.459	4.529	0.1540	
Rep 6	−8.306	4.529	0.0668	
Gender [male] × Rep 2	11.162	6.744	0.0981	
Gender [male] × Rep 3	9.530	6.763	0.1590	
Gender [male] × Rep 4	2.480	6.745	0.7132	
Gender [male] × Rep 5	4.694	6.745	0.4866	
Gender [male] × Rep 6	5.429	6.760	0.4220	

Table A.32: LMER model of English shadowed /b/ productions in the Extreme Prevoicing condition (target: −130 ms); Threshold for significance (adjusted with Bonferroni correction): p<0.0042

	Estimate	Std. Error	Pr(> |t|)
(Intercept)	−1.506	30.428	0.9605
Gender [male]	94.564	49.165	0.0546
Rep 2	−45.216	43.147	0.2948
Rep 3	−53.755	42.983	0.2112
Rep 4	−62.674	43.051	0.1456
Rep 5	−12.182	43.051	0.7772
Rep 6	−67.532	43.209	0.1183
Solidarity	−3.305	4.883	0.4987
Gender [male] × Rep 2	33.168	69.547	0.6335
Gender [male] × Rep 3	70.129	69.520	0.3132
Gender [male] × Rep 4	88.224	69.541	0.2047
Gender [male] × Rep 5	−3.792	69.541	0.9565
Gender [male] × Rep 6	69.568	69.661	0.3181
Gender [male] × Solidarity	−11.424	7.449	0.1253
Rep 2 × Solidarity	6.013	6.936	0.3861
Rep 3 × Solidarity	6.645	6.911	0.3364
Rep 4 × Solidarity	9.619	6.919	0.1646
Rep 5 × Solidarity	0.903	6.919	0.8961
Rep 6 × Solidarity	9.627	6.954	0.1664

Table A.33: LMER model of English shadowed /b/ tokens' VOT in Extr. Prev. Part 1/2; Threshold for significance (adjusted with Bonferroni correction): p<0.0020

	Estimate	Std. Error	Pr(> |t|)
Gender [male] × Rep 2 × Solidarity	−3.767	10.537	0.7208
Gender [male] × Rep 3 × Solidarity	−9.597	10.532	0.3623
Gender [male] × Rep 4 × Solidarity	−13.488	10.536	0.2007
Gender [male] × Rep 5 × Solidarity	1.182	10.536	0.9107
Gender [male] × Rep 6 × Solidarity	−10.275	10.563	0.3309

Table A.34: LMER model of English shadowed /b/ tokens' VOT in Extr. Prev. Part 2/2; Threshold for significance (adjusted with Bonferroni correction): p<0.0020

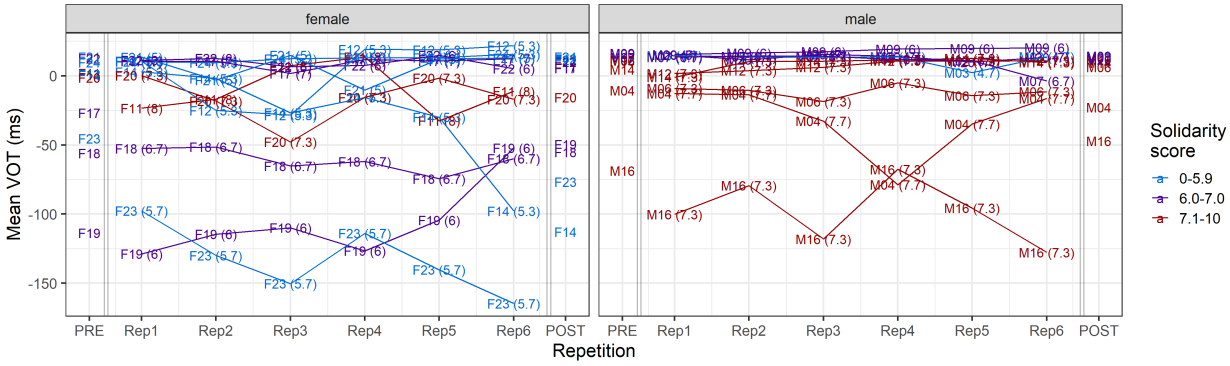

Figure A.26: Participants' /b/ shadowing trajectories by Solidarity rating in Extr. Prev.

	Estimate	Std. Error	Pr(> |t|)	
(Intercept)	−45.259	24.394	0.0637	
Gender [male]	115.585	34.404	0.0008	***
Rep 2	−33.003	34.464	0.3384	
Rep 3	−21.622	34.400	0.5297	
Rep 4	−22.534	34.402	0.5125	
Rep 5	−28.203	34.402	0.4124	
Rep 6	−35.312	34.411	0.3049	
Superiority	3.439	3.494	0.3252	
Gender [male] × Rep 2	27.041	48.677	0.5786	
Gender [male] × Rep 3	33.444	48.840	0.4936	
Gender [male] × Rep 4	22.772	48.643	0.6397	
Gender [male] × Rep 5	17.421	48.643	0.7203	
Gender [male] × Rep 6	35.461	48.751	0.4671	
Gender [male] × Superiority	−14.961	4.971	0.0027	.
Rep 2 × Superiority	3.604	4.947	0.4665	
Rep 3 × Superiority	1.271	4.941	0.7970	
Rep 4 × Superiority	2.780	4.942	0.5738	
Rep 5 × Superiority	3.146	4.942	0.5245	
Rep 6 × Superiority	3.931	4.944	0.4266	

Table A.35: LMER model of English shadowed /b/ tokens' VOT in Extr. Prev. Part 1/2; Threshold for significance (adjusted with Bonferroni correction): p<0.0020

	Estimate	Std. Error	Pr(> |t|)
Gender [male] × Rep 2 × Superiority	−2.236	7.027	0.7503
Gender [male] × Rep 3 × Superiority	−3.525	7.046	0.6169
Gender [male] × Rep 4 × Superiority	−2.936	7.026	0.6761
Gender [male] × Rep 5 × Superiority	−1.808	7.026	0.7969
Gender [male] × Rep 6 × Superiority	−4.302	7.043	0.5414

Table A.36: LMER model of English shadowed /b/ tokens' VOT in Extr. Prev. Part 2/2; Threshold for significance (adjusted with Bonferroni correction): p<0.0020

	Estimate	Std. Error	Pr(> |t|)
(Intercept)	−45.762	27.340	0.0945
Rep 2	−33.117	38.692	0.3923
Rep 3	−21.824	38.621	0.5721
Rep 4	−22.748	38.623	0.556
Rep 5	−28.418	38.623	0.462
Rep 6	−35.494	38.632	0.3584
Superiority	3.500	3.923	0.3725
Rep 2 × Superiority	3.623	5.554	0.5144
Rep 3 × Superiority	1.293	5.547	0.8157
Rep 4 × Superiority	2.806	5.548	0.6131
Rep 5 × Superiority	3.172	5.548	0.5677
Rep 6 × Superiority	3.949	5.550	0.4769

Table A.37: LMER model of English shadowed /b/ tokens' VOT from females in Extr. Prev. Threshold for significance (adjusted with Bonferroni correction): p<0.0042

	Estimate	Std. Error	Pr(> |t|)	
(Intercept)	77.964	29.968	0.0094	.
Rep 2	−1.070	42.612	0.9800	
Rep 3	47.794	42.446	0.2605	
Rep 4	−10.421	42.446	0.8061	
Rep 5	25.405	42.446	0.5497	
Rep 6	57.171	42.456	0.1785	
Superiority	−13.017	4.177	0.0019	**
Rep 2 × Superiority	−0.614	5.936	0.9177	
Rep 3 × Superiority	−7.979	5.917	0.1779	
Rep 4 × Superiority	1.173	5.918	0.8429	
Rep 5 × Superiority	−3.991	5.918	0.5002	
Rep 6 × Superiority	−8.404	5.920	0.1561	

Table A.38: LMER model of English shadowed /b/ tokens' VOT from females without F23 in Extr. Prev. Threshold for significance (adjusted with Bonferroni correction): p<0.0042

	Estimate	Std. Error	Pr(> |t|)	
(Intercept)	69.740	20.269	0.0006	***
Rep 2	−5.878	28.393	0.8360	
Rep 3	11.798	28.638	0.6805	
Rep 4	0.239	28.403	0.9933	
Rep 5	−10.783	28.403	0.7043	
Rep 6	0.339	28.524	0.9905	
Superiority	−11.426	2.921	<0.0001	***
Rep 2 × Superiority	1.349	4.122	0.7435	
Rep 3 × Superiority	−2.251	4.150	0.5876	
Rep 4 × Superiority	−0.156	4.125	0.9699	
Rep 5 × Superiority	1.338	4.125	0.7458	
Rep 6 × Superiority	−0.386	4.143	0.9258	

Table A.39: LMER model of English shadowed /b/ tokens' VOT from males in Extr. Prev. Threshold for significance (adjusted with Bonferroni correction): p<0.0042

	Estimate	Std. Error	Pr(> |t|)	
(Intercept)	41.832	24.197	0.0840	
Gender [male]	0.678	31.887	0.9830	
Rep 2	−68.173	34.273	0.0468	.
Rep 3	−63.634	34.093	0.0621	
Rep 4	−28.035	34.214	0.4127	
Rep 5	−10.629	34.214	0.7561	
Rep 6	−40.344	34.108	0.2370	
Dynamism	−10.882	4.051	0.0073	.
Gender [male] × Rep 2	57.412	45.174	0.2039	
Gender [male] × Rep 3	66.372	45.118	0.1415	
Gender [male] × Rep 4	51.957	45.174	0.2502	
Gender [male] × Rep 5	−1.376	45.174	0.9757	
Gender [male] × Rep 6	33.273	45.108	0.4608	
Gender [male] × Dynamism	2.824	5.210	0.5878	
Rep 2 × Dynamism	10.253	5.752	0.0748	
Rep 3 × Dynamism	8.681	5.727	0.1298	
Rep 4 × Dynamism	4.213	5.743	0.4633	
Rep 5 × Dynamism	0.701	5.743	0.9029	
Rep 6 × Dynamism	5.464	5.732	0.3406	

Table A.40: LMER model of English shadowed /b/ tokens' VOT in Extr. Prev. Part 1/2; Threshold for significance (adjusted with Bonferroni correction): p<0.0020

	Estimate	Std. Error	Pr(> |t|)
Gender [male] × Rep 2 × Dynamism	−7.984	7.374	0.2790
Gender [male] × Rep 3 × Dynamism	−9.721	7.367	0.1872
Gender [male] × Rep 4 × Dynamism	−8.212	7.378	0.2658
Gender [male] × Rep 5 × Dynamism	0.955	7.378	0.8971
Gender [male] × Rep 6 × Dynamism	−4.700	7.371	0.5238

Table A.41: LMER model of English shadowed /b/ tokens' VOT in Extr. Prev. Part 2/2; Threshold for significance (adjusted with Bonferroni correction): p<0.0020

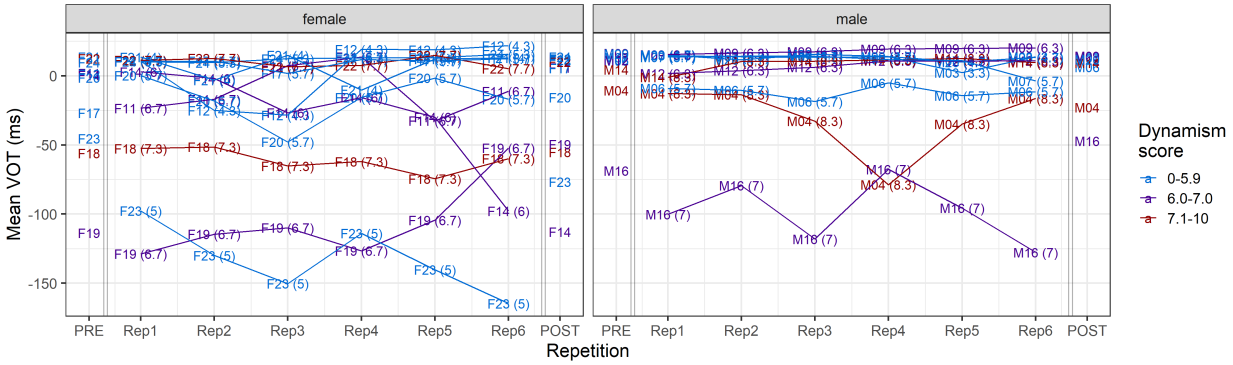

Figure A.27: Participants' /b/ shadowing trajectories by Dynamism rating in Extr. Prev.

Hungarian reading data: The Extreme Prevoicing condition

	Estimate	Std. Error	Pr(> |t|)	
(Intercept)	54.945	13.710	0.0008	***
Gender [male]	−29.896	20.740	0.1669	
Exposure [post]	−2.011	3.822	0.5988	
Superiority	−4.746	2.149	0.0407	.
Gender [male] × Exposure [post]	19.601	5.825	0.0008	***
Gender [male] × Sup.	5.625	3.091	0.0858	
Exposure [post] × Sup.	0.040	0.603	0.9469	
Gender [male] × Exposure [post] × Sup.	−3.075	0.868	0.0004	***

Table A.42: LMER model of Hungarian read /p/ tokens' VOT in Extr. Prev.; Threshold for significance (adjusted with Bonferroni correction): p<0.00625

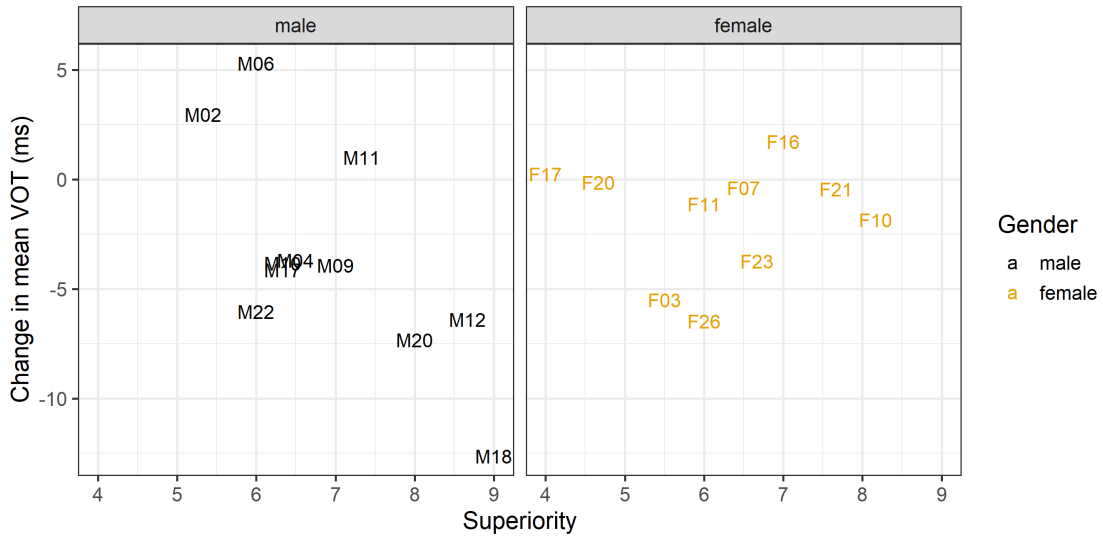

Figure A.28: Change in mean /p/ VOT in Extr. Prev. in the Hungarian reading task by gender and Superiority rating

	Estimate	Std. Error	Pr(> |t|)
(Intercept)	18.431	15.820	0.2590
Gender [male]	33.537	21.242	0.1320
Exposure [post]	−1.131	4.248	0.7900
Solidarity	1.034	2.285	0.6560
Gender [male] × Exposure [post]	3.219	5.737	0.5750
Gender [male] × Sol.	−4.045	3.057	0.2030
Exposure [post] × Sol.	−0.093	0.617	0.8800
Gender [male] × Exposure [post] × Sol.	−0.717	0.825	0.3850

Table A.43: LMER model of Hungarian read /p/ tokens' VOT in Extr. Prev.; Threshold for significance (adjusted with Bonferroni correction): p<0.00625

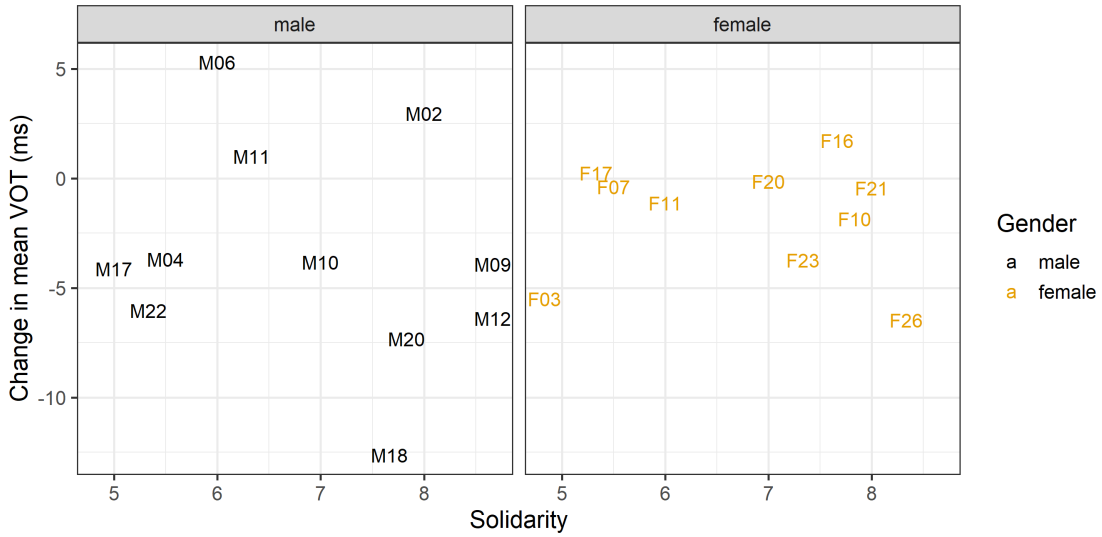

Figure A.29: Change in mean /p/ VOT in Extr. Prev. in the Hungarian reading task by gender and Solidarity rating

	Estimate	Std. Error	Pr(> |t|)	
(Intercept)	44.401	12.351	0.0020	**
Gender [male]	−17.235	15.488	0.2808	
Exposure [post]	−2.286	3.247	0.4815	
Dynamism	−3.017	1.898	0.1297	
Gender [male] × Exposure [post]	7.824	4.109	0.0571	
Gender [male] × Dyn.	3.659	2.397	0.1446	
Exposure [post] × Dyn.	0.083	0.503	0.8683	
Gender [male] × Exposure [post] × Dyn.	−1.537	0.636	0.0158	.

Table A.44: LMER model of Hungarian read /p/ tokens' VOT in Extr. Prev.; Threshold for significance (adjusted with Bonferroni correction): p<0.00625

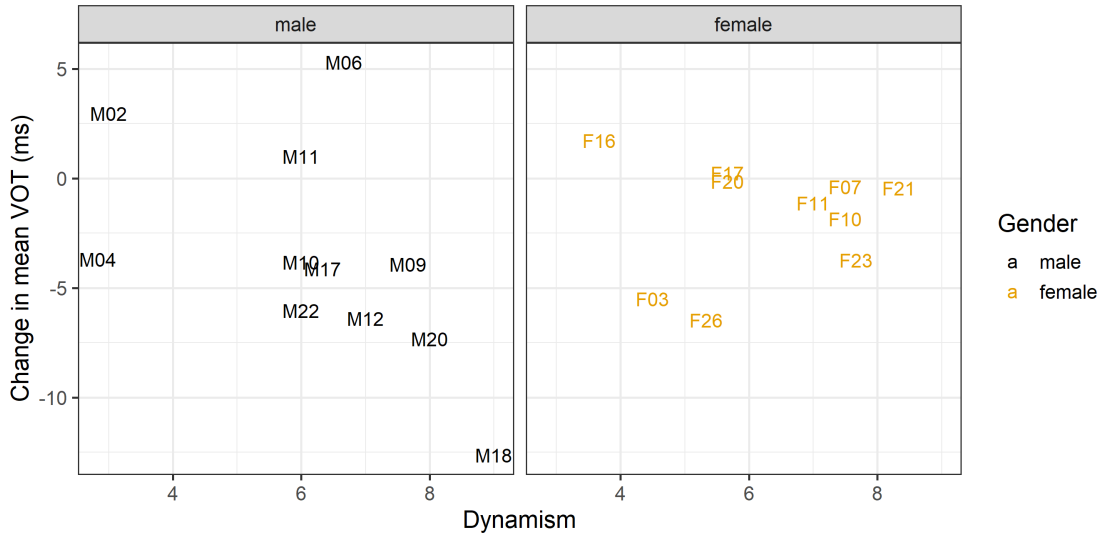

Figure A.30: Change in mean /p/ VOT in Extr. Prev. in the Hungarian reading task by gender and Dynamism rating

	Estimate	Std. Error	Pr(> |t|)	
(Intercept)	−44.156	37.727	0.2572	
Gender [male]	−15.208	57.441	0.7942	
Exposure [post]	−9.787	11.485	0.3942	
Superiority	−5.349	5.952	0.3808	
Gender [male] × Exposure [post]	58.602	17.514	0.0008	***
Gender [male] × Sup.	4.186	8.560	0.6308	
Exposure [post] × Sup.	0.860	1.813	0.6354	
Gender [male] × Exposure [post] × Sup.	−9.300	2.610	0.0004	***

Table A.45: LMER model of Hungarian read /b/ tokens' VOT in Extr. Prev.; Threshold for significance (adjusted with Bonferroni correction): p<0.00625

	Estimate	Std. Error	Pr(> |t|)	
(Intercept)	−98.395	39.932	0.0241	.
Gender [male]	65.881	53.872	0.2372	
Exposure [post]	−20.267	12.698	0.1107	
Solidarity	3.095	5.794	0.5998	
Gender [male] × Exposure [post]	53.722	17.154	0.0018	**
Gender [male] × Sol.	−8.151	7.752	0.3070	
Exposure [post] × Sol.	2.333	1.844	0.2059	
Gender [male] × Exposure [post] × Sol.	−8.605	2.468	0.0005	***

Table A.46: LMER model of Hungarian read /b/ tokens' VOT in Extr. Prev.; Threshold for significance (adjusted with Bonferroni correction): p<0.00625

Figure A.31: Change in mean /b/ VOT in Extr. Prev. in the Hungarian reading task by gender and Solidarity rating

	Estimate	Std. Error	Pr(> |t|)	
(Intercept)	−64.905	32.400	0.0606	
Gender [male]	0.077	40.949	0.9985	
Exposure [post]	19.321	9.738	0.0474	.
Dynamism	−1.987	5.018	0.6969	
Gender [male] × Exposure [post]	−4.930	12.323	0.6892	
Gender [male] × Dyn.	1.570	6.336	0.8071	
Exposure [post] × Dyn.	−3.782	1.510	0.0123	.
Gender [male] × Exposure [post] × Dyn.	−0.118	1.907	0.9509	

Table A.47: LMER model of Hungarian read /b/ tokens' VOT in Extr. Prev.;
Threshold for significance (adjusted with Bonferroni correction): p<0.00625

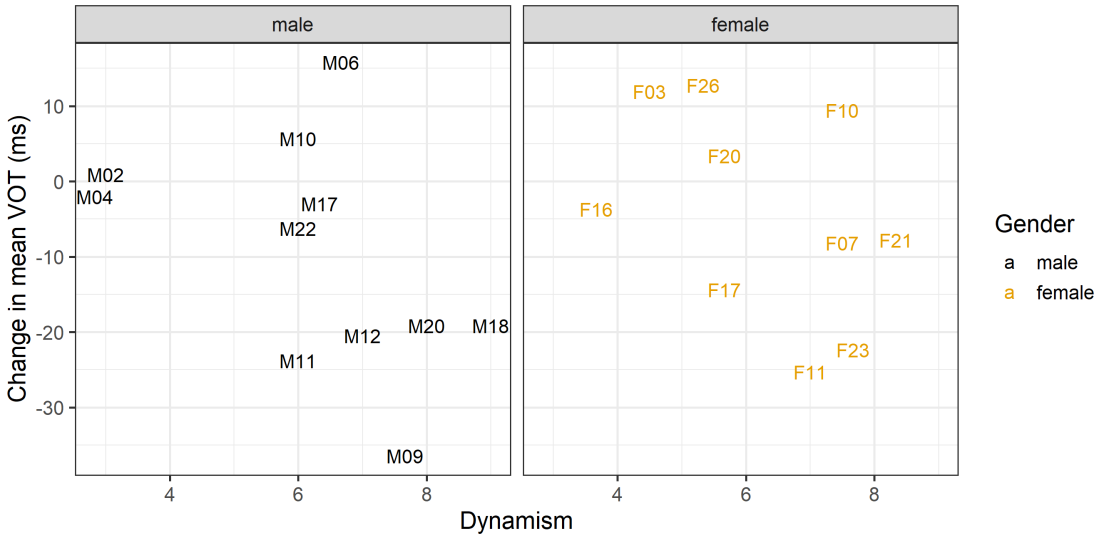

Figure A.32: Change in mean /b/ VOT in Extr. Prev. in the Hungarian reading task by gender and Dynamism rating

Hungarian reading data: The Extreme Aspirating condition

	Estimate	Std. Error	Pr(> |t|)	
(Intercept)	33.737	12.139	0.0129	.
Gender [male]	−8.892	19.219	0.6496	
Exposure [post]	−4.427	3.056	0.1477	
Superiority	−0.801	1.740	0.6512	
Gender [male] × Exposure [post]	7.747	4.867	0.1116	
Gender [male] × Sup.	0.890	2.863	0.7598	
Exposure [post] × Sup.	0.780	0.441	0.0773	
Gender [male] × Exposure [post] × Sup.	−1.464	0.725	0.0437	.

Table A.48: LMER model of Hungarian read /p/ tokens' VOT in Extr. Asp.;
Threshold for significance (adjusted with Bonferroni correction): p<0.00625

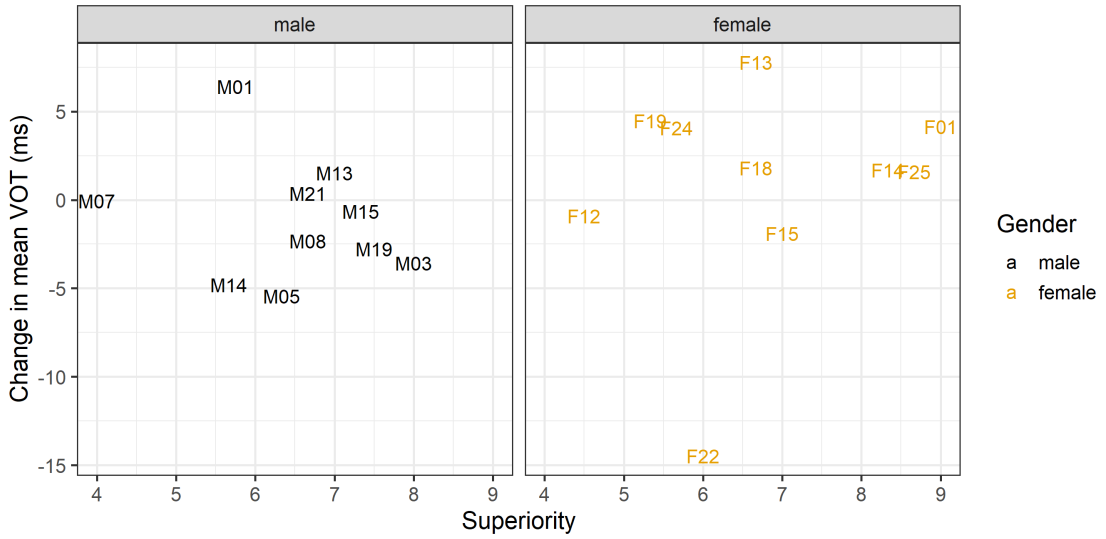

Figure A.33: Change in mean /p/ VOT in Extr. Asp. in the Hungarian reading task by gender and Superiority rating

	Estimate	Std. Error	Pr(> |t|)
(Intercept)	26.739	13.667	0.0671
Gender [male]	−12.189	16.564	0.4721
Exposure [post]	−3.396	3.667	0.3544
Solidarity	0.234	2.005	0.9086
Gender [male] × Exposure [post]	−2.899	4.466	0.5163
Gender [male] × Sol.	1.501	2.482	0.5536
Exposure [post] × Sol.	0.637	0.541	0.2388
Gender [male] × Exposure [post] × Sol.	0.189	0.670	0.7780

Table A.49: LMER model of Hungarian read /p/ tokens' VOT in Extr. Asp.;
Threshold for significance (adjusted with Bonferroni correction): p<0.00625

Figure A.34: Change in mean /p/ VOT in Extr. Asp. in the Hungarian reading task by gender and Solidarity rating

	Estimate	Std. Error	Pr(> |t|)	
(Intercept)	22.150	11.387	0.0684	
Gender [male]	−7.887	14.869	0.6029	
Exposure [post]	−7.545	3.020	0.0126	.
Dynamism	0.954	1.715	0.5855	
Gender [male] × Exposure [post]	9.285	3.972	0.0195	.
Gender [male] × Dyn.	0.916	2.326	0.6986	
Exposure [post] × Dyn.	1.304	0.458	0.0045	*
Gender [male] × Exposure [post] × Dyn.	−1.783	0.621	0.0042	*

Table A.50: LMER model of Hungarian read /p/ tokens' VOT in Extr. Asp. (did not converge); Threshold for significance (adjusted with Bonferroni correction): p<0.00625

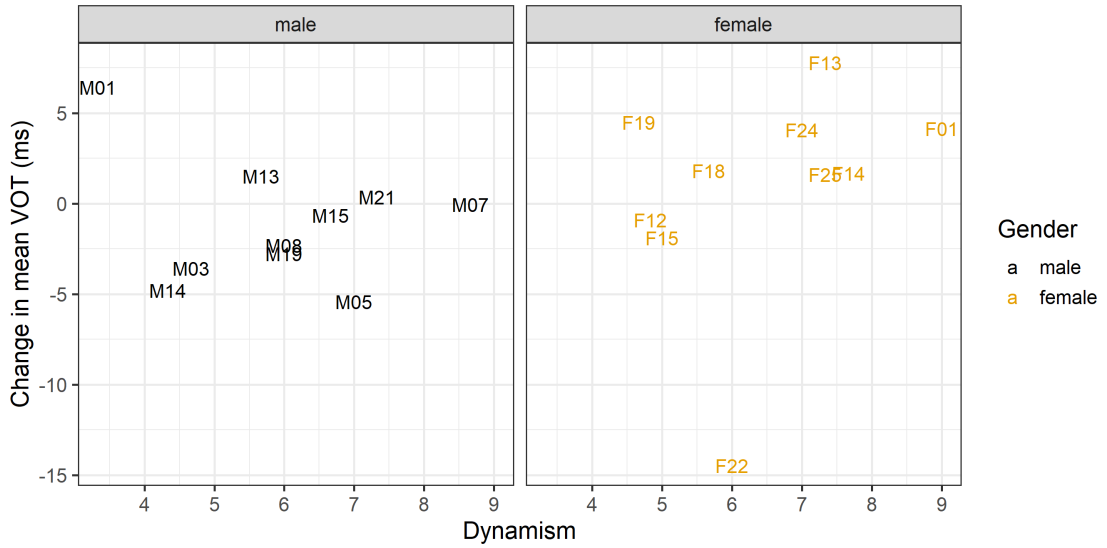

Figure A.35: Change in mean /p/ VOT in Extr. Asp. in the Hungarian reading task by gender and Dynamism rating

	Estimate	Std. Error	Pr(> |t|)	
(Intercept)	14.005	34.352	0.6886	
Gender [male]	−78.351	54.702	0.1704	
Exposure [post]	**−30.787**	**10.642**	**0.0039**	**
Superiority	−11.198	4.952	0.0373	.
Gender [male] × Exposure [post]	28.300	16.998	0.0961	
Gender [male] × Sup.	9.479	8.149	0.2610	
Exposure [post] × Sup.	**4.503**	**1.536**	**0.0034**	**
Gender [male] × Exposure [post] × Sup.	−5.527	2.532	0.0292	.

Table A.51: LMER model of Hungarian read /b/ tokens' VOT in Extr. Asp.;
Threshold for significance (adjusted with Bonferroni correction): p<0.00625

	Estimate	Std. Error	Pr(> |t|)	
(Intercept)	3.013	44.071	0.9463	
Gender [male]	−53.284	53.653	0.3348	
Exposure [post]	**−53.002**	12.715	<0.0001	***
Solidarity	−9.720	6.494	0.1531	
Gender [male] × Exposure [post]	18.849	15.488	0.2238	
Gender [male] × Sol.	5.694	8.040	0.4886	
Exposure [post] × Sol.	7.893	1.874	<0.0001	***
Gender [male] × Exposure [post] × Sol.	−3.899	2.321	0.0931	

Table A.52: LMER model of Hungarian read /b/ tokens' VOT in Extr. Asp.; Threshold for significance (adjusted with Bonferroni correction): p<0.00625

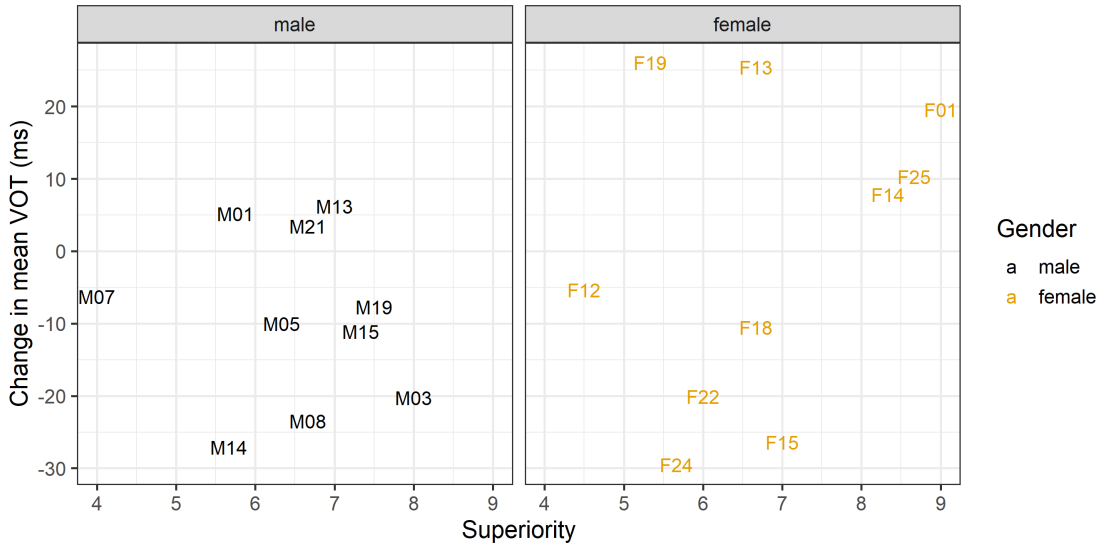

Figure A.36: Change in mean /b/ VOT in Extr. Asp. in the Hungarian reading task by gender and Superiority rating

	Estimate	Std. Error	Pr(> |t|)	
(Intercept)	−15.572	36.570	0.6757	
Gender [male]	−49.688	48.067	0.3160	
Exposure [post]	−32.308	10.545	0.0022	**
Dynamism	−7.189	5.542	0.2122	
Gender [male] × Exposure [post]	16.859	13.877	0.2246	
Gender [male] × Dyn.	5.473	7.518	0.4767	
Exposure [post] × Dyn.	4.970	1.599	0.0019	**
Gender [male] × Exposure [post] × Dyn.	−3.910	2.171	0.0719	

Table A.53: LMER model of Hungarian read /b/ tokens' VOT in Extr. Asp. (did not converge); Threshold for significance (adjusted with Bonferroni correction): p<0.00625

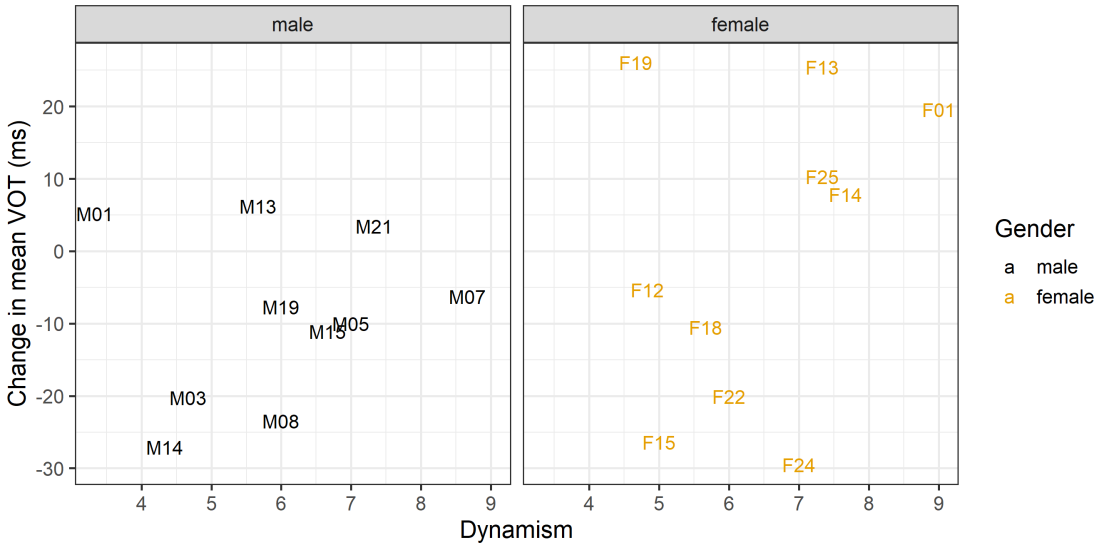

Figure A.37: Change in mean /b/ VOT in Extr. Asp. in the Hungarian reading task by gender and Dynamism rating

Hungarian shadowing data: The Extreme Prevoicing condition

	Estimate	Std. Error	Pr(> |t|)	
(Intercept)	27.522	3.688	<0.00001	***
Gender [male]	1.961	4.220	0.6490	
Rep 2	−1.099	1.630	0.5002	
Rep 3	1.107	1.630	0.4971	
Rep 4	0.892	1.630	0.5843	
Rep 5	−0.904	1.630	0.5794	
Rep 6	2.143	1.642	0.1921	
Gender [male] × Rep 2	0.101	2.033	0.9603	
Gender [male] × Rep 3	−2.564	2.034	0.2079	
Gender [male] × Rep 4	−0.926	2.033	0.6490	
Gender [male] × Rep 5	−0.792	2.033	0.6970	
Gender [male] × Rep 6	−4.908	2.042	0.0164	.

Table A.54: LMER model of shadowing /p/ productions in the Hungarian Extr. Prev. condition (target: 15 ms); Participants who "had room" to accommodate
Threshold for significance (adjusted with Bonferroni correction): p<0.0042

	Estimate	Std. Error	Pr(> |t|)	
(Intercept)	12.467	5.612	0.0265	.
Solidarity	1.418	0.790	0.0727	
Rep 2	−0.228	7.676	0.9763	
Rep 3	9.244	7.676	0.2287	
Rep 4	6.160	7.676	0.4224	
Rep 5	9.189	7.676	0.2314	
Rep 6	7.184	7.678	0.3496	
Gender [male]	35.021	7.346	<0.00001	***
Solidarity × Rep 2	0.049	1.114	0.9647	
Solidarity × Rep 3	−1.101	1.114	0.3231	
Solidarity × Rep 4	−0.741	1.114	0.5060	
Solidarity × Rep 5	−1.402	1.114	0.2086	
Solidarity × Rep 6	−0.846	1.115	0.4481	
Solidarity × Gender [male]	−4.370	1.058	<0.00001	***
Rep 2 × Gender [male]	−6.863	10.366	0.5080	
Rep 3 × Gender [male]	−13.769	10.376	0.1846	
Rep 4 × Gender [male]	−10.133	10.366	0.3284	
Rep 5 × Gender [male]	−15.526	10.366	0.1344	
Rep 6 × Gender [male]	−18.708	10.367	0.0713	

Table A.55: LMER model of English shadowed /p/ tokens' VOT in Hungarian Extr. Prev. Part 1/2; Threshold for significance (adjusted with Bonferroni correction): p<0.0020

	Estimate	Std. Error	Pr(> |t|)
Solidarity × Rep 2 × Gender [male]	0.861	1.492	0.5641
Solidarity × Rep 3 × Gender [male]	1.542	1.494	0.3020
Solidarity × Rep 4 × Gender [male]	1.324	1.492	0.3750
Solidarity × Rep 5 × Gender [male]	2.143	1.492	0.1509
Solidarity × Rep 6 × Gender [male]	2.181	1.492	0.1441

Table A.56: LMER model of English shadowed /p/ tokens' VOT in Hungarian Extr. Prev. Part 2/2; Threshold for significance (adjusted with Bonferroni correction): p<0.0020

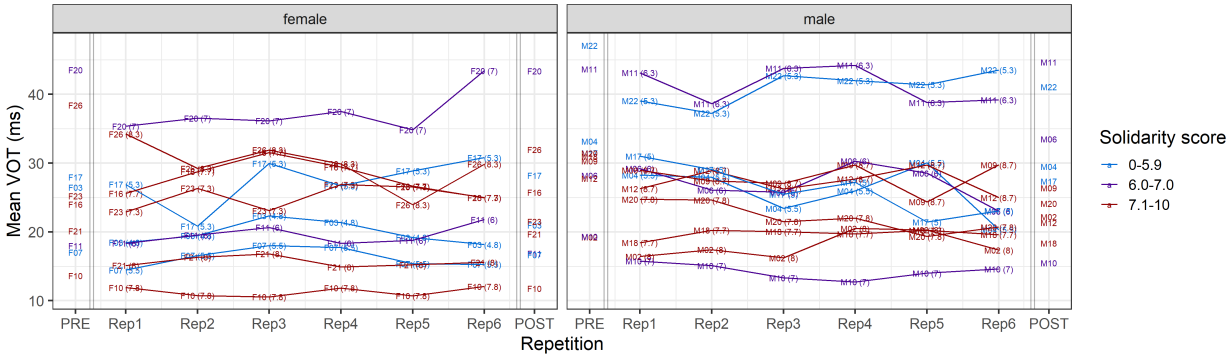

Figure A.38: Participants' /p/ shadowing trajectories by Solidarity rating in Hungarian Extr. Prev.

	Estimate	Std. Error	Pr(> |t|)	
(Intercept)	47.376	4.986	<0.00001	***
Superiority	−4.064	0.755	<0.00001	***
Rep 2	−6.631	6.756	0.3264	
Rep 3	3.922	6.756	0.5616	
Rep 4	0.568	6.756	0.9330	
Rep 5	0.259	6.756	0.9694	
Rep 6	9.347	6.756	0.1667	
Gender [male]	−17.031	7.320	0.0201	.
Superiority × Rep 2	1.084	1.067	0.3097	
Superiority × Rep 3	−0.346	1.067	0.7458	
Superiority × Rep 4	0.091	1.067	0.9324	
Superiority × Rep 5	−0.093	1.067	0.9303	
Superiority × Rep 6	−1.276	1.067	0.2318	
Superiority × Gender [male]	3.602	1.090	0.0010	***
Rep 2 × Gender [male]	−0.272	10.316	0.9789	
Rep 3 × Gender [male]	−9.462	10.332	0.3599	
Rep 4 × Gender [male]	0.344	10.316	0.9734	
Rep 5 × Gender [male]	−2.135	10.316	0.8361	
Rep 6 × Gender [male]	−14.044	10.316	0.1736	

Table A.57: LMER model of English shadowed /p/ tokens' VOT in Hungarian Extr. Prev. Part 1/2; Threshold for significance (adjusted with Bonferroni correction): p<0.0020

	Estimate	Std. Error	Pr(> |t|)
Superiority × Rep 2 × Gender [male]	−0.211	1.536	0.8909
Superiority × Rep 3 × Gender [male]	0.930	1.539	0.5458
Superiority × Rep 4 × Gender [male]	−0.218	1.536	0.8871
Superiority × Rep 5 × Gender [male]	0.185	1.536	0.9043
Superiority × Rep 6 × Gender [male]	1.617	1.537	0.2928

Table A.58: LMER model of English shadowed /p/ tokens' VOT in Hungarian Extr. Prev. Part 2/2; Threshold for significance (adjusted with Bonferroni correction): p<0.0020

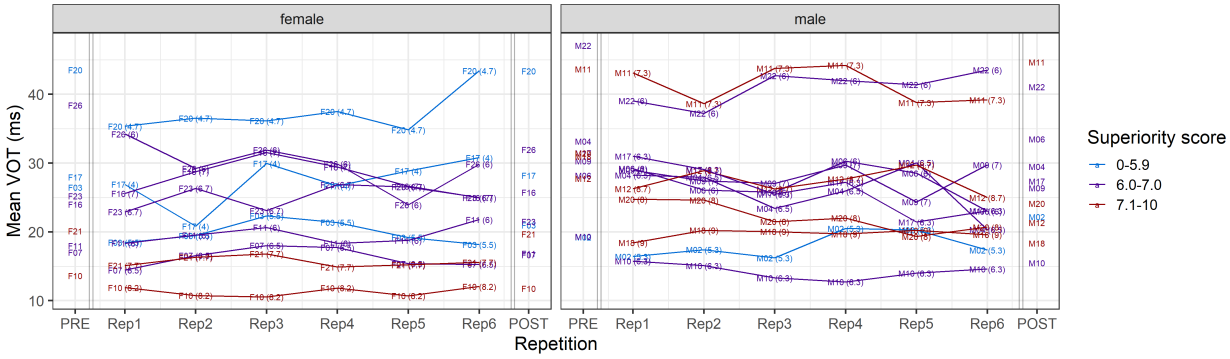

Figure A.39: Participants' /p/ shadowing trajectories by Superiority rating in Hungarian Extr. Prev.

	Estimate	Std. Error	Pr(> |t|)	
(Intercept)	39.332	4.355	<0.00001	***
Dynamism	−2.740	0.640	<0.00001	***
Rep 2	−1.983	5.833	0.7339	
Rep 3	6.058	5.833	0.2991	
Rep 4	2.097	5.833	0.7192	
Rep 5	−3.301	5.833	0.5715	
Rep 6	−0.683	5.842	0.9070	
Gender [male]	−12.128	5.247	0.0209	.
Dynamism × Rep 2	0.332	0.904	0.7131	
Dynamism × Rep 3	−0.682	0.904	0.4507	
Dynamism × Rep 4	−0.154	0.904	0.8650	
Dynamism × Rep 5	0.474	0.904	0.6001	
Dynamism × Rep 6	0.331	0.907	0.7152	
Dynamism × Gender [male]	2.726	0.811	0.0008	***
Rep 2 × Gender [male]	0.478	7.397	0.9485	
Rep 3 × Gender [male]	−9.602	7.397	0.1944	
Rep 4 × Gender [male]	−1.510	7.397	0.8383	
Rep 5 × Gender [male]	6.615	7.397	0.3713	
Rep 6 × Gender [male]	−4.798	7.404	0.5171	

Table A.59: LMER model of English shadowed /p/ tokens' VOT in Hungarian Extr. Prev. Part 1/2; Threshold for significance (adjusted with Bonferroni correction): p<0.0020

	Estimate	Std. Error	Pr(> |t|)
Dynamism × Rep 2 × Gender [male]	−0.224	1.144	0.8450
Dynamism × Rep 3 × Gender [male]	1.014	1.144	0.3753
Dynamism × Rep 4 × Gender [male]	0.064	1.144	0.9553
Dynamism × Rep 5 × Gender [male]	−1.205	1.144	0.2924
Dynamism × Rep 6 × Gender [male]	0.176	1.146	0.8780

Table A.60: LMER model of English shadowed /p/ tokens' VOT in Hungarian Extr. Prev. Part 2/2; Threshold for significance (adjusted with Bonferroni correction): p<0.0020

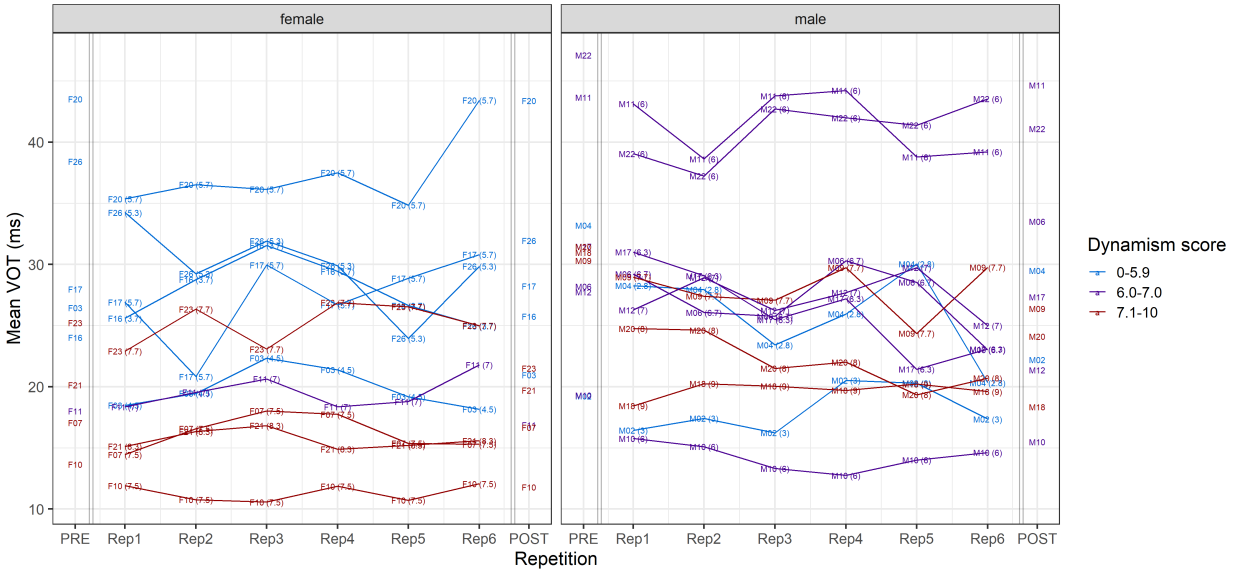

Figure A.40: Participants' /p/ shadowing trajectories by Dynamism rating in Hungarian Extr. Prev.

	Estimate	Std. Error	Pr(> |t|)	
(Intercept)	−128.196	18.730	<0.00001	***
Solidarity	5.960	2.711	0.028	.
Rep 2	−2.292	26.362	0.9307	
Rep 3	46.216	26.362	0.0797	
Rep 4	27.726	26.362	0.2931	
Rep 5	15.934	26.362	0.5456	
Rep 6	−5.701	26.362	0.8288	
Gender [male]	47.753	25.211	0.0584	
Solidarity × Rep 2	0.273	3.827	0.9431	
Solidarity × Rep 3	−7.074	3.827	0.0647	
Solidarity × Rep 4	−3.603	3.827	0.3467	
Solidarity × Rep 5	−2.282	3.827	0.5511	
Solidarity × Rep 6	0.472	3.827	0.9019	
Solidarity × Gender [male]	−7.492	3.628	0.039	.
Rep 2 × Gender [male]	19.263	35.588	0.5884	
Rep 3 × Gender [male]	−24.358	35.588	0.4938	
Rep 4 × Gender [male]	−23.118	35.588	0.516	
Rep 5 × Gender [male]	9.551	35.589	0.7885	
Rep 6 × Gender [male]	47.137	35.588	0.1855	

Table A.61: LMER model of English shadowed /b/ tokens' VOT in Hungarian Extr. Prev. Part 1/2; Threshold for significance (adjusted with Bonferroni correction): p<0.0020

	Estimate	Std. Error	Pr(> |t|)
Solidarity × Rep 2 × Gender [male]	−3.261	5.121	0.5244
Solidarity × Rep 3 × Gender [male]	3.253	5.121	0.5254
Solidarity × Rep 4 × Gender [male]	2.864	5.121	0.576
Solidarity × Rep 5 × Gender [male]	−1.385	5.121	0.7868
Solidarity × Rep 6 × Gender [male]	−6.829	5.121	0.1825

Table A.62: LMER model of English shadowed /b/ tokens' VOT in Hungarian Extr. Prev. Part 2/2; Threshold for significance (adjusted with Bonferroni correction): p<0.0020

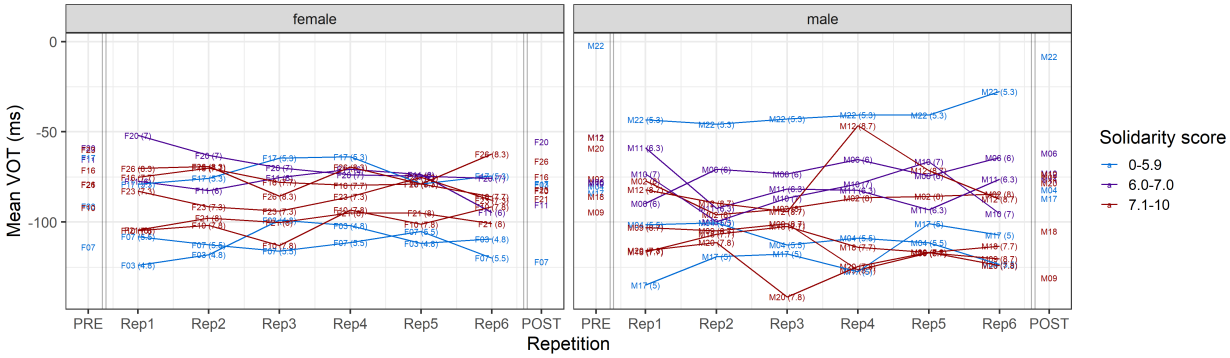

Figure A.41: Participants' /b/ shadowing trajectories by Solidarity rating in Hungarian Extr. Prev.

	Estimate	Std. Error	Pr(> |t|)	
(Intercept)	−42.288	16.791	0.0119	.
Superiority	−7.315	2.641	0.0057	.
Rep 2	−9.176	23.629	0.6978	
Rep 3	14.686	23.629	0.5343	
Rep 4	−2.196	23.629	0.9260	
Rep 5	−20.164	23.629	0.3936	
Rep 6	−17.216	23.629	0.4663	
Gender [male]	0.022	25.495	0.9993	
Superiority × Rep 2	1.406	3.730	0.7064	
Superiority × Rep 3	−2.647	3.730	0.4781	
Superiority × Rep 4	0.882	3.730	0.8131	
Superiority × Rep 5	3.317	3.730	0.3741	
Superiority × Rep 6	2.367	3.730	0.5258	
Superiority × Gender [male]	0.318	3.800	0.9333	
Rep 2 × Gender [male]	−4.390	36.001	0.9030	
Rep 3 × Gender [male]	−11.109	36.001	0.7577	
Rep 4 × Gender [male]	−17.629	36.001	0.6244	
Rep 5 × Gender [male]	31.363	36.001	0.3838	
Rep 6 × Gender [male]	34.722	36.001	0.3349	

Table A.63: LMER model of English shadowed /b/ tokens' VOT in Hungarian Extr. Prev. Part 1/2; Threshold for significance (adjusted with Bonferroni correction): p<0.0020

	Estimate	Std. Error	Pr(> |t|)
Superiority × Rep 2 × Gender [male]	0.017	5.365	0.9974
Superiority × Rep 3 × Gender [male]	1.479	5.365	0.7828
Superiority × Rep 4 × Gender [male]	1.898	5.365	0.7236
Superiority × Rep 5 × Gender [male]	−4.906	5.365	0.3606
Superiority × Rep 6 × Gender [male]	−5.242	5.365	0.3287

Table A.64: LMER model of English shadowed /b/ tokens' VOT in Hungarian Extr. Prev. Part 2/2; Threshold for significance (adjusted with Bonferroni correction): p<0.0020

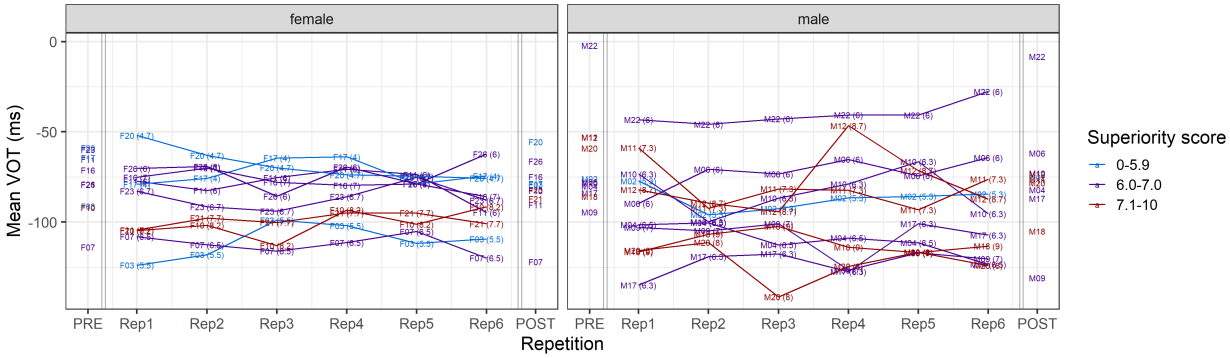

Figure A.42: Participants' /b/ shadowing trajectories by Superiority rating in Hungarian Extr. Prev.

	Estimate	Std. Error	Pr(> |t|)	
(Intercept)	−66.881	14.430	<0.00001	***
Dynamism	−3.325	2.227	0.1360	
Rep 2	6.726	20.297	0.7400	
Rep 3	11.814	20.297	0.5610	
Rep 4	1.524	20.297	0.9400	
Rep 5	−14.422	20.297	0.4770	
Rep 6	−1.202	20.297	0.9530	
Gender [male]	2.933	18.181	0.8720	
Dynamism × Rep 2	−1.140	3.147	0.7170	
Dynamism × Rep 3	−2.162	3.147	0.4920	
Dynamism × Rep 4	0.281	3.147	0.9290	
Dynamism × Rep 5	2.368	3.147	0.4520	
Dynamism × Rep 6	−0.207	3.147	0.9480	
Dynamism × Gender [male]	−1.014	2.813	0.7190	
Rep 2 × Gender [male]	−28.870	25.681	0.2610	
Rep 3 × Gender [male]	−32.138	25.681	0.2110	
Rep 4 × Gender [male]	−11.899	25.681	0.6430	
Rep 5 × Gender [male]	4.078	25.683	0.8740	
Rep 6 × Gender [male]	−15.486	25.681	0.5470	

Table A.65: LMER model of English shadowed /b/ tokens' VOT in Hungarian Extr. Prev. Part 1/2; Threshold for significance (adjusted with Bonferroni correction): p<0.0020

	Estimate	Std. Error	Pr(> |t|)
Dynamism × Rep 2 × Gender [male]	4.106	3.974	0.3020
Dynamism × Rep 3 × Gender [male]	4.695	3.974	0.2380
Dynamism × Rep 4 × Gender [male]	1.305	3.974	0.7430
Dynamism × Rep 5 × Gender [male]	−0.685	3.974	0.8630
Dynamism × Rep 6 × Gender [male]	2.486	3.974	0.5320

Table A.66: LMER model of English shadowed /b/ tokens' VOT in Hungarian Extr. Prev. Part 2/2; Threshold for significance (adjusted with Bonferroni correction): p<0.0020

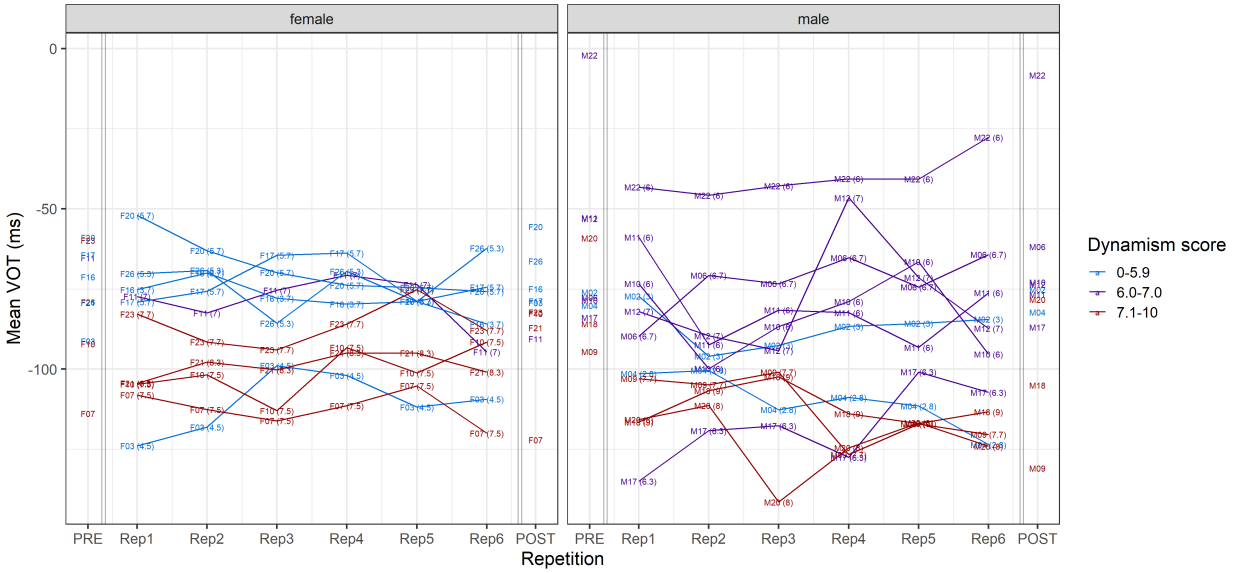

Figure A.43: Participants' /b/ shadowing trajectories by Dynamism rating in Hungarian Extr. Prev.

Hungarian shadowing data: The Extreme Aspiration condition

	Estimate	Std. Error	Pr(> |t|)	
(Intercept)	34.686	5.712	<0.00001	***
Solidarity	−1.236	0.816	0.1302	
Rep 2	−1.176	7.827	0.8806	
Rep 3	4.311	7.824	0.5817	
Rep 4	8.706	7.824	0.2659	
Rep 5	1.309	7.824	0.8672	
Rep 6	0.682	7.824	0.9306	
Gender [male]	−21.556	6.742	0.0014	**
Solidarity × Rep 2	0.110	1.154	0.9239	
Solidarity × Rep 3	−0.653	1.153	0.5715	
Solidarity × Rep 4	−1.014	1.153	0.3796	
Solidarity × Rep 5	−0.075	1.153	0.9484	
Solidarity × Rep 6	0.326	1.153	0.7774	
Solidarity × Gender [male]	2.919	1.010	0.0039	**
Rep 2 × Gender [male]	−5.189	9.533	0.5863	
Rep 3 × Gender [male]	−6.863	9.531	0.4715	
Rep 4 × Gender [male]	−9.579	9.531	0.3150	
Rep 5 × Gender [male]	−4.832	9.531	0.6122	
Rep 6 × Gender [male]	0.174	9.531	0.9855	

Table A.67: LMER model of English shadowed /p/ tokens' VOT in Hungarian Extr. Asp. Part 1/2; Threshold for significance (adjusted with Bonferroni correction): p<0.0020

	Estimate	Std. Error	Pr(> |t|)
Solidarity × Rep 2 × Gender [male]	0.981	1.428	0.4923
Solidarity × Rep 3 × Gender [male]	1.262	1.428	0.3771
Solidarity × Rep 4 × Gender [male]	1.418	1.428	0.3209
Solidarity × Rep 5 × Gender [male]	0.959	1.428	0.5018
Solidarity × Rep 6 × Gender [male]	−0.073	1.428	0.9593

Table A.68: LMER model of English shadowed /p/ tokens' VOT in Hungarian Extr. Asp. Part 2/2; Threshold for significance (adjusted with Bonferroni correction): p<0.0020

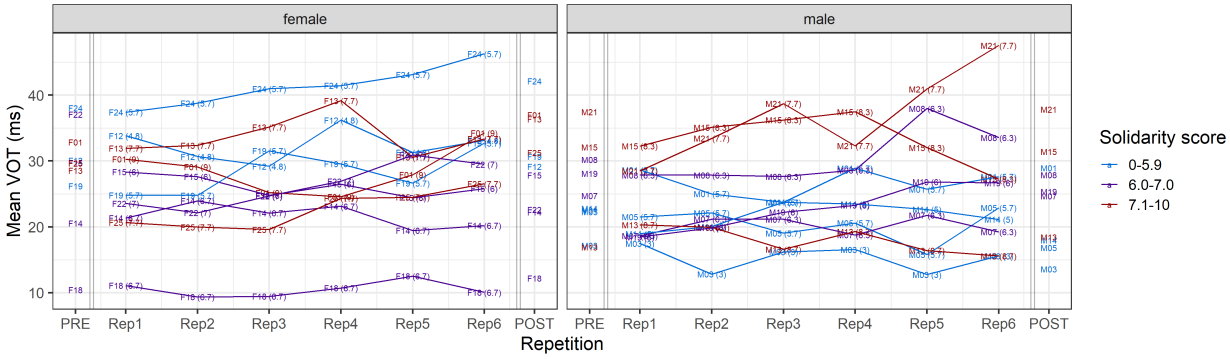

Figure A.44: Participants' /p/ shadowing trajectories by Solidarity rating in Hungarian Extr. Asp.

	Estimate	Std. Error	Pr(> |t|)	
(Intercept)	38.454	4.865	<0.00001	***
Superiority	−1.771	0.671	0.0084	.
Rep 2	−2.771	6.576	0.6735	
Rep 3	5.712	6.574	0.3850	
Rep 4	9.910	6.574	0.1319	
Rep 5	4.453	6.574	0.4983	
Rep 6	5.155	6.574	0.4330	
Gender [male]	−20.322	7.413	0.0062	.
Superiority × Rep 2	0.344	0.949	0.7170	
Superiority × Rep 3	−0.850	0.949	0.3705	
Superiority × Rep 4	−1.176	0.949	0.2151	
Superiority × Rep 5	−0.537	0.949	0.5714	
Superiority × Rep 6	−0.338	0.949	0.7215	
Superiority × Gender [male]	2.627	1.105	0.0175	.
Rep 2 × Gender [male]	7.186	10.478	0.4929	
Rep 3 × Gender [male]	−3.864	10.476	0.7123	
Rep 4 × Gender [male]	−10.515	10.476	0.3156	
Rep 5 × Gender [male]	0.144	10.476	0.9890	
Rep 6 × Gender [male]	−2.947	10.476	0.7785	

Table A.69: LMER model of English shadowed /p/ tokens' VOT in Hungarian Extr. Asp. Part 1/2; Threshold for significance (adjusted with Bonferroni correction): p<0.0020

	Estimate	Std. Error	Pr(> |t|)
Superiority × Rep 2 × Gender [male]	−0.951	1.561	0.5424
Superiority × Rep 3 × Gender [male]	0.760	1.561	0.6265
Superiority × Rep 4 × Gender [male]	1.525	1.561	0.3285
Superiority × Rep 5 × Gender [male]	0.140	1.561	0.9284
Superiority × Rep 6 × Gender [male]	0.374	1.561	0.8105

Table A.70: LMER model of English shadowed /p/ tokens' VOT in Hungarian Extr. Asp. Part 2/2; Threshold for significance (adjusted with Bonferroni correction): p<0.0020

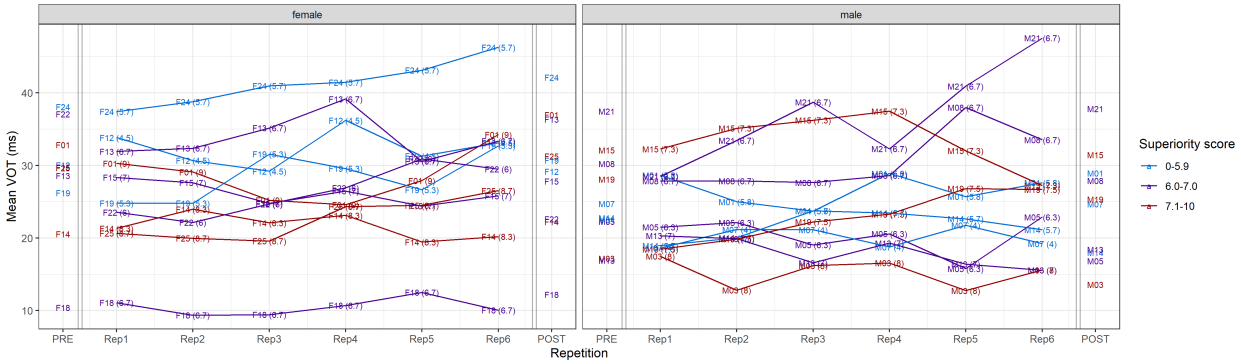

Figure A.45: Participants' /p/ shadowing trajectories by Superiority rating in Hungarian Extr. Asp.

	Estimate	Std. Error	Pr(> |t|)	
(Intercept)	23.194	4.878	<0.00001	***
Dynamism	0.503	0.709	0.4780	
Rep 2	−3.534	6.602	0.5930	
Rep 3	2.282	6.602	0.7300	
Rep 4	5.881	6.602	0.3730	
Rep 5	1.122	6.602	0.8650	
Rep 6	−0.273	6.602	0.9670	
Gender [male]	0.289	6.145	0.9620	
Dynamism × Rep 2	0.480	1.001	0.6320	
Dynamism × Rep 3	−0.362	1.001	0.7180	
Dynamism × Rep 4	−0.613	1.001	0.5410	
Dynamism × Rep 5	−0.048	1.001	0.9610	
Dynamism × Rep 6	0.486	1.001	0.6280	
Dynamism × Gender [male]	−0.470	0.962	0.6250	
Rep 2 × Gender [male]	−4.187	8.685	0.6300	
Rep 3 × Gender [male]	−8.868	8.685	0.3070	
Rep 4 × Gender [male]	−4.746	8.685	0.5850	
Rep 5 × Gender [male]	−6.945	8.685	0.4240	
Rep 6 × Gender [male]	−5.151	8.685	0.5530	

Table A.71: LMER model of English shadowed /p/ tokens' VOT in Hungarian Extr. Asp. Part 1/2; Threshold for significance (adjusted with Bonferroni correction): p<0.0020

	Estimate	Std. Error	Pr(> |t|)
Dynamism × Rep 2 × Gender [male]	0.894	1.358	0.5110
Dynamism × Rep 3 × Gender [male]	1.677	1.358	0.2170
Dynamism × Rep 4 × Gender [male]	0.701	1.358	0.6060
Dynamism × Rep 5 × Gender [male]	1.363	1.358	0.3160
Dynamism × Rep 6 × Gender [male]	0.832	1.358	0.5400

Table A.72: LMER model of English shadowed /p/ tokens' VOT in Hungarian Extr. Asp. Part 2/2; Threshold for significance (adjusted with Bonferroni correction): p<0.0020

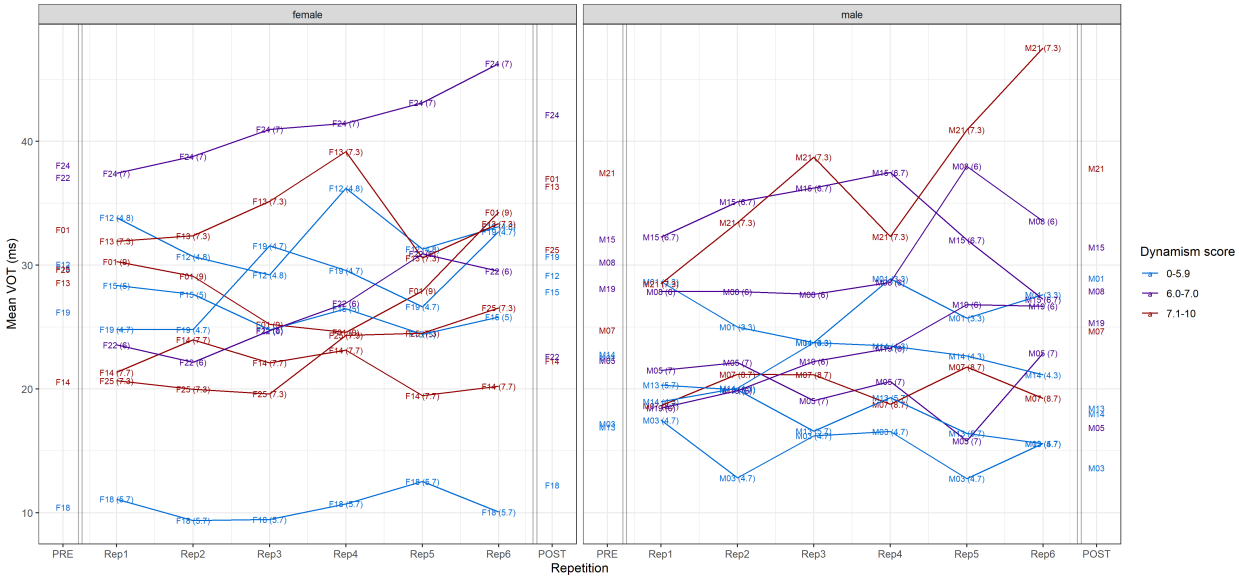

Figure A.46: Participants' /p/ shadowing trajectories by Dynamism rating in Hungarian Extr. Asp.

	Estimate	Std. Error	Pr(> |t|)	
(Intercept)	−69.702	10.001	<0.00001	***
Gender [male]	−18.512	13.945	0.1990	
Rep 2	2.271	3.998	0.5700	
Rep 3	−1.960	3.998	0.6240	
Rep 4	−5.155	4.019	0.2000	
Rep 5	−4.581	4.004	0.2530	
Rep 6	4.387	4.026	0.2760	
Gender [male] × Rep 2	−4.536	5.653	0.4220	
Gender [male] × Rep 3	2.593	5.653	0.6470	
Gender [male] × Rep 4	3.531	5.669	0.5330	
Gender [male] × Rep 5	−1.074	5.658	0.8490	
Gender [male] × Rep 6	−2.273	5.674	0.6890	

Table A.73: LMER model of shadowing /b/ productions in the Hungarian Extreme Aspirating condition (target: 130 ms);
Threshold for significance (adjusted with Bonferroni correction): p<0.0042

	Estimate	Std. Error	Pr(> |t|)	
(Intercept)	0.626	21.242	0.9765	
Solidarity	−10.507	3.121	0.0008	***
Rep 2	−13.441	29.903	0.6531	
Rep 3	−49.092	29.903	0.1008	
Rep 4	−35.149	30.421	0.2481	
Rep 5	−7.770	30.072	0.7961	
Rep 6	−46.032	30.648	0.1333	
Gender [male]	−70.933	25.792	0.0060	.
Solidarity × Rep 2	2.351	4.409	0.5939	
Solidarity × Rep 3	7.052	4.409	0.1099	
Solidarity × Rep 4	4.582	4.493	0.3080	
Solidarity × Rep 5	0.501	4.440	0.9102	
Solidarity × Rep 6	7.713	4.544	0.0898	
Solidarity × Gender [male]	7.645	3.865	0.0481	.
Rep 2 × Gender [male]	31.589	36.428	0.3860	
Rep 3 × Gender [male]	62.066	36.428	0.0886	
Rep 4 × Gender [male]	40.336	36.854	0.2739	
Rep 5 × Gender [male]	18.783	36.566	0.6075	
Rep 6 × Gender [male]	40.867	37.042	0.2701	

Table A.74: LMER model of English shadowed /b/ tokens' VOT in Hungarian Extr. Asp. Part 1/2; Threshold for significance (adjusted with Bonferroni correction): p<0.0020

	Estimate	Std. Error	Pr(> |t|)
Solidarity × Rep 2 × Gender [male]	−5.608	5.459	0.3044
Solidarity × Rep 3 × Gender [male]	−9.022	5.459	0.0986
Solidarity × Rep 4 × Gender [male]	−5.669	5.527	0.3052
Solidarity × Rep 5 × Gender [male]	−3.161	5.483	0.5644
Solidarity × Rep 6 × Gender [male]	−6.553	5.568	0.2395

Table A.75: LMER model of English shadowed /b/ tokens' VOT in Hungarian Extr. Asp. Part 2/2; Threshold for significance (adjusted with Bonferroni correction): p<0.0020

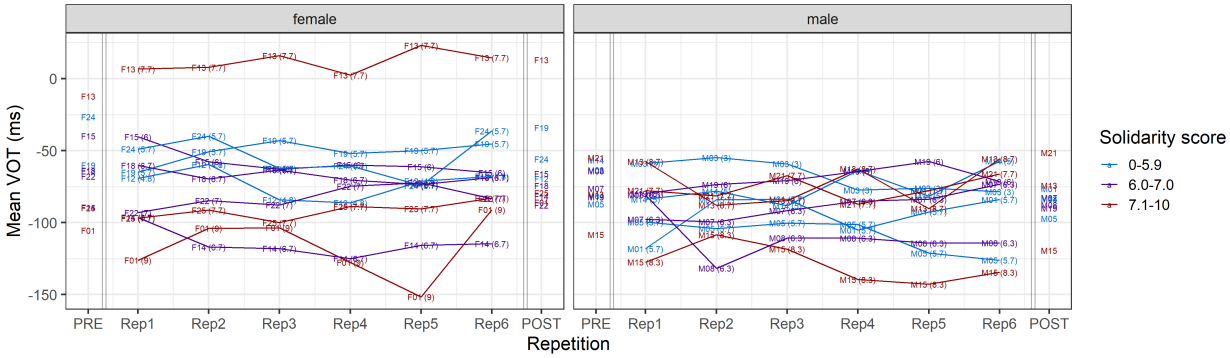

Figure A.47: Participants' /b/ shadowing trajectories by Solidarity rating in Hungarian Extr. Asp.

	Estimate	Std. Error	Pr(> |t|)	
(Intercept)	11.224	17.205	0.5143	
Superiority	−11.932	2.468	<0.00001	***
Rep 2	13.835	24.163	0.5670	
Rep 3	−7.925	24.163	0.7429	
Rep 4	0.033	24.451	0.9989	
Rep 5	22.058	24.236	0.3629	
Rep 6	−2.098	24.474	0.9317	
Gender [male]	−145.475	27.249	<0.00001	***
Superiority × Rep 2	−1.705	3.487	0.6250	
Superiority × Rep 3	0.879	3.487	0.8009	
Superiority × Rep 4	−0.686	3.533	0.8462	
Superiority × Rep 5	−3.916	3.501	0.2635	
Superiority × Rep 6	1.055	3.549	0.7663	
Superiority × Gender [male]	19.008	4.060	<0.00001	***
Rep 2 × Gender [male]	−4.448	38.506	0.9080	
Rep 3 × Gender [male]	29.321	38.506	0.4465	
Rep 4 × Gender [male]	47.344	38.688	0.2212	
Rep 5 × Gender [male]	22.282	38.552	0.5634	
Rep 6 × Gender [male]	78.023	38.704	0.0440	.

Table A.76: LMER model of English shadowed /b/ tokens' VOT in Hungarian Extr. Asp. Part 1/2; Threshold for significance (adjusted with Bonferroni correction): p<0.0020

	Estimate	Std. Error	Pr(> |t|)	
Superiority × Rep 2 × Gender [male]	−0.090	5.736	0.9875	
Superiority × Rep 3 × Gender [male]	−4.078	5.736	0.4773	
Superiority × Rep 4 × Gender [male]	−6.863	5.765	0.2340	
Superiority × Rep 5 × Gender [male]	−3.785	5.745	0.5101	
Superiority × Rep 6 × Gender [male]	−12.426	5.775	0.0316	.

Table A.77: LMER model of English shadowed /b/ tokens' VOT in Hungarian Extr. Asp. Part 2/2; Threshold for significance (adjusted with Bonferroni correction): p<0.0020

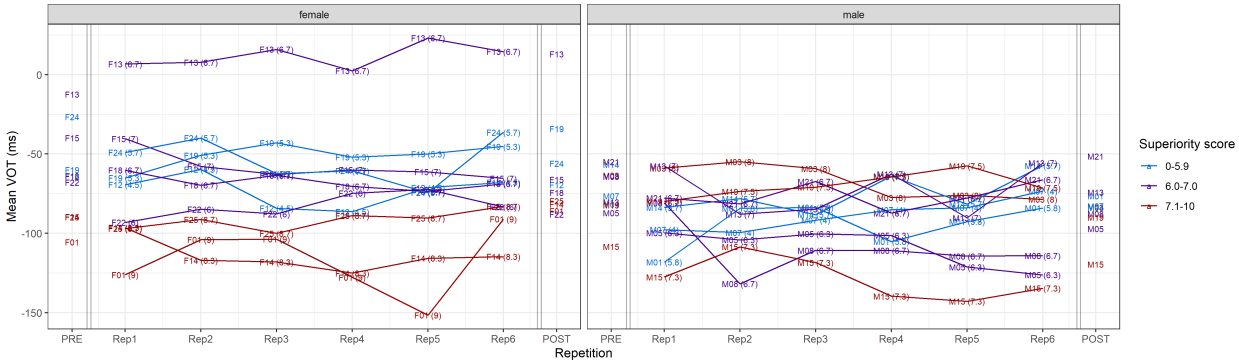

Figure A.48: Participants' /b/ shadowing trajectories by Superiority rating in Hungarian Extr. Asp.

	Estimate	Std. Error	Pr(> |t|)	
(Intercept)	−8.601	17.435	0.6219	
Dynamism	−9.466	2.631	0.0003	***
Rep 2	−8.258	24.499	0.7361	
Rep 3	−18.102	24.499	0.4601	
Rep 4	−4.452	24.812	0.8576	
Rep 5	19.077	24.611	0.4384	
Rep 6	−32.663	24.988	0.1913	
Gender [male]	−72.713	22.813	0.0015	**
Dynamism × Rep 2	1.632	3.716	0.6605	
Dynamism × Rep 3	2.503	3.716	0.5008	
Dynamism × Rep 4	−0.028	3.772	0.9941	
Dynamism × Rep 5	−3.655	3.739	0.3284	
Dynamism × Rep 6	5.909	3.814	0.1215	
Dynamism × Gender [male]	8.296	3.568	0.0202	.
Rep 2 × Gender [male]	32.228	32.229	0.3175	
Rep 3 × Gender [male]	29.365	32.229	0.3624	
Rep 4 × Gender [male]	11.109	32.468	0.7323	
Rep 5 × Gender [male]	−15.377	32.314	0.6342	
Rep 6 × Gender [male]	51.519	32.605	0.1143	

Table A.78: LMER model of English shadowed /b/ tokens' VOT in Hungarian Extr. Asp. Part 1/2; Threshold for significance (adjusted with Bonferroni correction): p<0.0020

	Estimate	Std. Error	Pr(> |t|)
Dynamism × Rep 2 × Gender [male]	−6.029	5.041	0.2318
Dynamism × Rep 3 × Gender [male]	−4.284	5.041	0.3955
Dynamism × Rep 4 × Gender [male]	−1.360	5.082	0.7890
Dynamism × Rep 5 × Gender [male]	2.087	5.057	0.6799
Dynamism × Rep 6 × Gender [male]	−8.715	5.114	0.0885

Table A.79: LMER model of English shadowed /b/ tokens' VOT in Hungarian Extr. Asp. Part 2/2; Threshold for significance (adjusted with Bonferroni correction): p<0.0020

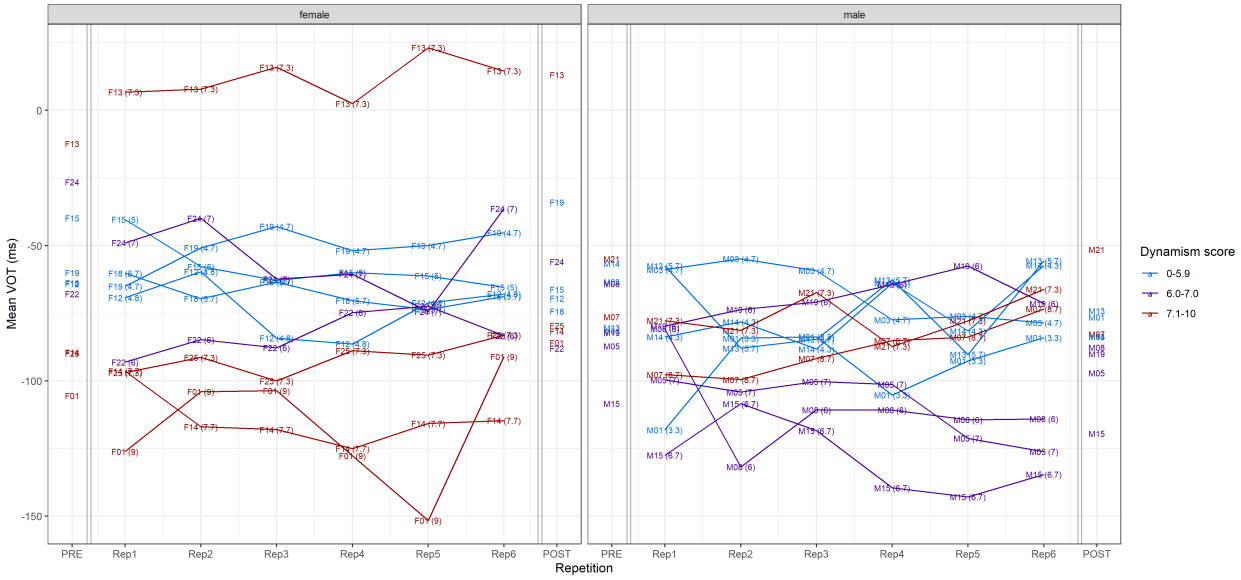

Figure A.49: Participants' /b/ shadowing trajectories by Dynamism rating in Hungarian Extr. Asp.

Bibliography

Abramson, Arthur S, and Douglas H Whalen. 2017. Voice Onset Time (VOT) at 50: Theoretical and practical issues in measuring voicing distinctions. *Journal of Phonetics* 63:75–86.

Allen, J Sean, Joanne L Miller, and David DeSteno. 2003. Individual talker differences in voice-onset-time. *The Journal of the Acoustical Society of America* 113:544–552.

Auer, Peter, and Frans Hinskens. 2005. The role of interpersonal accommodation in a theory of language change. In *Dialect change: Convergence and divergence in European languages*, ed. Peter Auer, Frans Hinskens, and Paul Kerswill, 335–357. Cambridge University Press.

Babel, Molly. 2009. Phonetic and social selectivity in speech accommodation. Doctoral Dissertation, University of California, Berkeley.

Babel, Molly. 2010. Dialect divergence and convergence in New Zealand English. *Language in Society* 39:437–456.

Babel, Molly. 2012. Evidence for phonetic and social selectivity in spontaneous phonetic imitation. *Journal of Phonetics* 40:177–189.

Babel, Molly, and Dasha Bulatov. 2012. The role of fundamental frequency in phonetic accommodation. *Language and Speech* 55:231–248.

Babel, Molly, Grant McGuire, Sophia Walters, and Alice Nicholls. 2014. Novelty and social preference in phonetic accommodation. *Laboratory Phonology* 5:123–150.

Babel, Molly, Brianne Senior, and Sophie Bishop. 2019. Do social preferences matter in lexical retuning? *Laboratory Phonology* 10:4.

Baese-Berk, Melissa M. 2010a. Learning novel phonetic categories in perception and production. *The Journal of the Acoustical Society of America* 128:2489–2489.

Baese-Berk, Melissa Michaud. 2010b. An examination of the relationship between speech perception and production. Doctoral Dissertation, Northwestern University.

Baese-Berk, Melissa Michaud. 2019. Interactions between speech perception and production during learning of novel phonemic categories. *Attention, Perception, & Psychophysics* 81:981–1005.

Ball, Martin J, and Joan Rahilly. 2014. *Phonetics: The science of speech*. Routledge.

Balog, Heather L, and Diane Brentari. 2008. The relationship between early gestures and intonation. *First Language* 28:141–163.

Bane, Max, Peter Graff, and Morgan Sonderegger. 2010. Longitudinal phonetic variation in a closed system. *Proceedings from the Annual Meeting of the Chicago Linguistic Society* 46:43–58.

Bauman, Carina. 2013. Social evaluation of Asian accented English. *University of Pennsylvania Working Papers in Linguistics* 19:3.

Beach, Elizabeth Francis, Denis Burnham, and Christine Kitamura. 2001. Bilingualism and the relationship between perception and production: Greek/English bilinguals and Thai bilabial stops. *International Journal of Bilingualism* 5:221–235.

Beckner, Clay, Péter Rácz, Jennifer Hay, Jürgen Brandstetter, and Christoph Bartneck. 2016. Participants Conform to Humans but Not to Humanoid Robots in an English Past Tense Formation Task. *Journal of Language and Social Psychology* 35:158–179.

Beebe, Leslie M. 1981. Social and situational factors affecting the communicative strategy of dialect code-switching. *International Journal of the Sociology of Language* 1981:139–149.

Bell, Allan. 1984. Language style as audience design. *Language in Society* 13:145–204.

Bell, Allan. 2002. Back in style: Reworking audience design. In *Style and Sociolinguistic Variation*, 139–169. Cambridge: Cambridge University Press.

Van den Berg, Marien E. 1986. Language planning and language use in Taiwan: Social identity, language accommodation, and language choice behavior. *International Journal of the Sociology of Language* 59:97–116.

Bilous, Frances R, and Robert M Krauss. 1988. Dominance and accommodation in the conversational behaviours of same- or mixed-gender dyads. *Language & Communication* 8:183–194.

Bloomfield, Leonard. 1933. *Language*. New York: Henry Holt.

Bock, J Kathryn. 1986. Syntactic Persistence in Language Production. *Cognitive Psychology* 18:355–387.

Brass, Marcel, Jan Derrfuss, and D Yves von Cramon. 2005. The inhibition of imitative and overlearned responses: a functional double dissociation. *Neuropsychologia* 43:89–98.

Brass, Marcel, Jan Derrfuss, Gabriele Matthes-von Cramon, and D Yves von Cramon. 2003. Imitative response tendencies in patients with frontal brain lesions. *Neuropsychology* 17:265.

Bybee, Joan. 1999. Usage-based phonology. *Functionalism and formalism in linguistics* 1:211–242.

Bybee, Joan. 2001. *Phonology and language use*, volume 94. Cambridge University Press.

Carford, J.C. 2001. *A practical introduction to phonetics*. Oxford: Clarendon Press.

Carranza, Michael A, and Ellen Bouchard Ryan. 1975. Evaluative reactions of bilingual Anglo and Mexican American adolescents toward speakers of English and Spanish. *International Journal of the Sociology of Language* 1975:83–104.

Chambers, Jack. 1992. Dialect acquisition. *Language* 68:673–705.

Chartrand, Tanya L, and John A Bargh. 1999. The Chameleon Effect: The Perception-Behavior Link and Social Interaction. *Journal of Personality and Social Psychology* 76:893–910.

Cho, Taehong, and Peter Ladefoged. 1999. Variation and universals in VOT: evidence from 18 languages. *Journal of Phonetics* 27:207–229.

Cho, Taehong, and James M McQueen. 2005. Prosodic influences on consonant production in Dutch: Effects of prosodic boundaries, phrasal accent and lexical stress. *Journal of Phonetics* 33:121–157.

Cho, Taehong, DH Whalen, and Gerard Docherty. 2019. Voice onset time and beyond: Exploring laryngeal contrast in 19 languages. *Journal of Phonetics* 72:52–65.

Chodroff, Eleanor, and Colin Wilson. 2017. Structure in talker-specific phonetic realization: Covariation of stop consonant VOT in American English. *Journal of Phonetics* 61:30–47.

Cohen Priva, Uriel, Lee Edelist, and Emily Gleason. 2017. Converging to the baseline: Corpus evidence for convergence in speech rate to interlocutor's baseline. *The Journal of the Acoustical Society of America* 141:2989–2996.

Cohn, Michelle, Bruno Ferenc Segedin, and Georgia Zellou. 2019. Imitating Siri: Socially-mediated vocal alignment to device and human voices. In *Proceedings of the 19th International Congress of Phonetic Sciences, Melbourne, Australia 2019*, ed. Sasha Calhoun, Paola Escudero, Marija Tabain, and Paul Warren, 1813–1817.

Cole, Jennifer, Hansook Choi, Heejin Kim, and Mark Hasegawa-Johnson. 2003. The effect of accent on the acoustic cues to stop voicing in Radio News speech. In *Proceedings of the 15th International Congress of Phonetic Sciences*, 15–18.

Coulston, Rachel, Sharon Oviatt, and Courtney Darves. 2002. Amplitude convergence in children's conversational speech with animated personas. In *7th International Conference on Spoken Language Processing*.

Cruttenden, Alan. 2013. *Gimson's pronunciation of English*. Routledge.

Davenport, Mike, and SJ Hannahs. 2010. *Introducing phonetics and phonology*. Oxford: Routledge.

Davidson, Lisa. 2016. Acoustic effects of phonetic conditions and laryngeal specification on phonation in English voiceless obstruents. *The Journal of the Acoustical Society of America* 140:3223–3224.

Davidson, Lisa. 2017. Phonation and laryngeal specification in American English voiceless obstruents. *Journal of the International Phonetic Association* 1–26.

Delvaux, Véronique, and Alain Soquet. 2007. The influence of ambient speech on adult speech productions through unintentional imitation. *Phonetica* 64:145–173.

Dijksterhuis, Ap, and John A Bargh. 2001. The perception-behavior expressway: Automatic effects of social perception on social behavior. In *Advances in Experimental Social Psychology*, volume 33, 1–40. Academic Press.

Docherty, Gerard, Dominic Watt, Carmen Llamas, Damien Hall, and Jennifer Nycz. 2011. Variation in Voice Onset Time along the Scottish-English Border. In *Proceedings of the 17th International Congress of Phonetic Sciences (ICPhS), Hong Kong*, 591–594. Online: https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2011/OnlineProceedings/RegularSession/Docherty/Docherty.pdf.

Docherty, Gerard J. 1992. *The timing of voicing in British English obstruents*. 9. Walter de Gruyter.

Estival, Dominique. 1985. Syntactic priming of the passive in English. *Text-Interdisciplinary Journal for the Study of Discourse* 5:7–22.

Fasold, Ralph W., William Labov, Fay Boyd Vaughn-Cooke, Guy Bailey, Walt Wolfram, Arthur K. Spears, and John Rickford. 1987. Are Black and White Vernaculars Diverging? Papers from the NWAVE XIV Panel Discussion. *American Speech* 62:3–80.

Feldstein, Stanley, and Cynthia L Crown. 1990. Oriental and Canadian conversational interactions: Chronographic structure and interpersonal perception. *Journal of Asian Pacific Communication* 1:247.

Flege, James Emil. 1982. Laryngeal timing and phonation onset in utterance-initial English stops. *Journal of Phonetics* 10:177–192.

Fletcher, Garth, Geoff Thomas, and Russil Durrant. 1999. Cognitive and Behavioral Accommodation in Close Relationships. *Journal of Social and Personal Relationships* 16:705–730.

Garner, Thurmon, and Donald L Rubin. 1986. Middle class Blacks' perceptions of dialect and style shifting: The case of southern attorneys. *Journal of Language and Social Psychology* 5:33–48.

Garrett, Andrew, and Keith Johnson. 2013. Phonetic bias in sound change. *Origins of Sound Change: Approaches to Phonologization* 51–97.

Garrod, Simon, and Anthony Anderson. 1987. Saying what you mean in dialogue: A study in conceptual and semantic co-ordination. *Cognition* 27:181–218.

Giles, Howard, Nikolas Coupland, and Justine Coupland. 1991. Accommodation theory: Communication, context, and consequence. *Contexts of accommodation: Developments in applied sociolinguistics* 1:1–68.

Giles, Howard, Anthony Mulac, James J Bradac, and Patricia Johnson. 1987. Speech accommodation theory: The first decade and beyond. *Annals of the International Communication Association* 10:13–48.

Gilliéron, Jules. 1918. *Généalogie des mots qui désignent l'abeille d'apres l'atlas linguistique de la france*. Paris: É. Champion.

Gimson, Alfred Charles. 1962. *An introduction to the pronunciation of English*. Edward Arnold, London.

Goldinger, Stephen D. 1997. Words and voices: Perception and production in an episodic lexicon. *Talker variability in speech processing* 33–66.

Goldinger, Stephen D. 1998. Echoes of echoes? An episodic theory of lexical access. *Psychological Review* 105:251.

Goldinger, Stephen D, and Tamiko Azuma. 2003. Puzzle-solving science: The quixotic quest for units in speech perception. *Journal of Phonetics* 31:305–320.

Goldinger, Stephen D, and Tamiko Azuma. 2004. Episodic memory reflected in printed word naming. *Psychonomic Bulletin & Review* 11:716–722.

Gordon, Matthew J. 2002. Investigating chain shifts and mergers. *The handbook of language variation and change* 244–266.

Gorter, Durk. 1987. Aspects of language choice in the Frisian-Dutch bilingual context: Neutrality and asymmetry. *Journal of Multilingual & Multicultural Development* 8:121–132.

Gósy, Mária. 2001. The VOT of the Hungarian voiceless plosives in words and in spontaneous Speech. *International Journal of Speech Technology* 4:75–85.

Gósy, Mária, and Catherine O Ringen. 2009. Everything you always wanted to know about VOT in Hungarian. In *Proceedings of the International Conference on the Structure of Hungarian*.

Gráczi, Etelka, Tekla. 2011. Explozívák a zöngésségi oppozíció függvényében. In *V. Alkalmazott Nyelvészeti Doktoranduszkonferencia*, ed. Váradi Tamás, 51–66. Budapest: MTA Nyelvtudományi Intézet.

Graff, Peter, Kie Zuraw, and Kuniko Nielsen. 2009. Investigating Preferential Imitation. In *Talk given at the LSA 2009 Annual Meeting*.

Gregory, Stanford, Stephen Webster, and Gang Huang. 1993. Voice pitch and amplitude convergence as a metric of quality in dyadic interviews. *Language & Communication* 13:195–217.

Gregory, Stanford W, Kelly Dagan, and Stephen Webster. 1997. Evaluating the relation of vocal accommodation in conversation partners' fundamental frequencies to perceptions of communication quality. *Journal of Nonverbal Behavior* 21:23–43.

Gregory, Stanford W, Brian E Green, Robert M Carrothers, Kelly A Dagan, and Stephen W Webster. 2001. Verifying the primacy of voice fundamental frequency in social status accommodation. *Language & Communication* 21:37–60.

Gregory, Stanford W, and Brian R Hoyt. 1982. Conversation partner mutual adaptation as demonstrated by Fourier series analysis. *Journal of Psycholinguistic Research* 11:35–46.

Gregory, Stanford W, and Stephen Webster. 1996. A nonverbal signal in voices of interview partners effectively predicts communication accommodation and social status perceptions. *Journal of Personality and Social Psychology* 70:1231–1240.

Grossberg, Stephen. 1980. How does a brain build a cognitive code? *Studies of Mind and Brain* 1–52.

Grossberg, Stephen. 1999. The link between brain learning, attention, and consciousness. *Consciousness and Cognition* 8:1–44.

Grossberg, Stephen. 2003. Resonant neural dynamics of speech perception. *Journal of Phonetics* 31:423–445.

Halácsy, Péter, András Kornai, László Németh, András Rung, István Szakadát, and Viktor Trón. 2004. Creating open language resources for Hungarian. In *Proceedings of Language Resources and Evaluation Conference*, 203–210.

Hanssen, Judith, Jörg Peters, and Carlos Gussenhoven. 2007. Phrase-final pitch accommodation effects in dutch. In *Proceedings of International Congress of Phonetic Sciences*, volume 2, 1077–1080.

Harrington, Jonathan. 2006. An acoustic analysis of 'happy-tensing' in the Queen's Christmas broadcasts. *Journal of Phonetics* 34:439–457.

Harrington, Jonathan. 2007. Evidence for a relationship between synchronic variability and diachronic change in the Queen's annual Christmas broadcasts. *Laboratory Phonology* 9:125–143.

Harrington, Jonathan, Sallyanne Palethorpe, and Catherine Watson. 2000. Monophthongal vowel changes in Received Pronunciation: an acoustic analysis of the Queen's Christmas broadcasts. *Journal of the International Phonetic Association* 30:63–78.

Harrington, Jonathan, Sallyanne Palethorpe, and Catherine I Watson. 2007. Age-related changes in fundamental frequency and formants: a longitudinal study of four speakers. In *Eighth Annual Conference of the International Speech Communication Association*.

Hosseini-Kivanani, Nina, Stephen J. Tobin, and Adamantios I. Gafos. 2019. Phonetic accommodation in the fundamental frequency of Korean-English bilinguals and English monolinguals.

In *Proceedings of the 19th International Congress of Phonetic Sciences, Melbourne, Australia 2019*, ed. Sasha Calhoun, Paola Escudero, Marija Tabain, and Paul Warren, 2610–2614.

Hutchinson, Amy, and Olga Dmitrieva. 2019. Stability of individual patterns in learning a second language voicing contrast. In *Proceedings of the 19th International Congress of Phonetic Sciences, Melbourne, Australia 2019*, ed. Sasha Calhoun, Paola Escudero, Marija Tabain, and Paul Warren, 959–963.

Jacewicz, Ewa, Robert Allen Fox, and Samantha Lyle. 2009. Variation in stop consonant voicing in two regional varieties of American English. *Journal of the International Phonetic Association* 39:313–334.

Johnson, Keith. 1997. Speech perception without speaker normalization: An exemplar model. In *Talker variability in speech processing*, ed. Keith Johnson and John W. Mullennix, 145–165.

Johnston, Paul. 1997. Regional variation. In *The edinburgh history of the scots language*, ed. Charles Jones, 433–513. Edinburgh: Edinburgh University Press.

Kane, John, Kinga Pápay, László Hunyadi, and Christer Gobl. 2011. On the use of creak in Hungarian spontaneous speech. In *Proceedings of International Congress of Phonetic Sciences*, 1014–1017.

Kaplan, Abby. 2011. How much homophony is normal? *Journal of linguistics* 47:631–671.

Kaschak, Michael, and Arthur Glenberg. 2004. Interactive alignment: Priming or memory retrieval? *Behavioral and Brain Sciences* 27:201–202.

Keating, Patricia A. 1984. Phonetic and phonological representation of stop consonant voicing. *Language* 286–319.

Keating, Patricia A, Michael J Mikoś, and William F III Ganong. 1981. A cross-language study of range of voice onset time in the perception of initial stop voicing. *The Journal of the Acoustical Society of America* 70:1261–1271.

Kharlamov, Viktor. 2018. Prevoicing and prenasalization in Russian initial plosives. *Journal of Phonetics* 71:215–228.

Kim, Midam. 2012. Phonetic accommodation after auditory exposure to native and nonnative speech. Doctoral Dissertation, Northwestern University.

Kim, Midam, William S Horton, and Ann R Bradlow. 2011. Phonetic convergence in spontaneous conversations as a function of interlocutor language distance. *Laboratory Phonology* 2:125–156.

Kim, Sahyang, Jiseung Kim, and Taehong Cho. 2018. Prosodic-structural modulation of stop voicing contrast along the VOT continuum in trochaic and iambic words in American English. *Journal of Phonetics* 71:65–80.

Koch, Lisa M, Alan M Gross, and Russell Kolts. 2001. Attitudes toward Black English and code switching. *Journal of Black psychology* 27:29–42.

Kontra, Miklós, and Mária Gósy. 1988. Approximation of the standard: A form of variability in bilingual speech. In *Methods in Dialectology*, 442–455. Multilingual Matters Clevedon.

Kraljic, Tanya, and Arthur G Samuel. 2006. Generalization in perceptual learning for speech. *Psychonomic Bulletin & Review* 13:262–268.

Kuhl, Patricia K. 1991. Human adults and human infants show a “perceptual magnet effect" for the prototypes of speech categories, monkeys do not. *Perception & Psychophysics* 50:93–107.

Kuhl, Patricia K, and Paul Iverson. 1995. Linguistic Experience and the “Perceptual Magnet Effect”. In *Speech Perception and Linguistic Experience: Issues in Cross-language Research*, ed. W. Strange, 121–154.

Kuhl, Patricia K, and James D Miller. 1978. Speech perception by the chinchilla: Identification functions for synthetic VOT stimuli. *The Journal of the Acoustical Society of America* 63:905–917.

Labov, William. 1972. *Sociolinguistic Patterns*. University of Pennsylvania Press.

Labov, William. 1990. The intersection of sex and social class in the course of linguistic change. *Language Variation and Change* 2:205–254.

Labov, William. 1994. *Principles of Linguistic Change: Internal Factors*, volume 1. Oxford: Blackwell.

Lacy, Karyn R. 2004. Black spaces, black places: Strategic assimilation and identity construction in middle-class suburbia. *Ethnic and Racial Studies* 27:908–930.

Lasky, Robert E, Ann Syrdal-Lasky, and Robert E Klein. 1975. VOT discrimination by four to six and a half month old infants from Spanish environments. *Journal of Experimental Child Psychology* 20:215–225.

Laturnus, Rebecca. 2017. The effect of lifetime exposure on perceptual adaptation to non-native speech. Poster presented at New Ways of Analyzing Variation, Madison, United States.

Le Page, Robert Brock, and Andrée Tabouret-Keller. 1985. *Acts of identity: Creole-based approaches to language and ethnicity*. Cambridge:Cambridge University Press.

Levi, Susannah V. 2015. Generalization of Phonetic Detail: Cross-Segmental, Within-Category Priming of VOT. *Language and Speech* 58:549–562.

Lewandowski, Natalie. 2012. Talent in nonnative phonetic convergence. Doctoral Dissertation, Universität Stuttgart.

Lin, Susan, Margaret Cychosz, Alice Shen, and Emily Cibelli. 2019. The effects of phonetic training and visual feedback on novel contrast production. In *Proceedings of the 19th International Congress of Phonetic Sciences, Melbourne, Australia 2019*, ed. Sasha Calhoun, Paola Escudero, Marija Tabain, and Paul Warren, 899–903.

Lisker, Leigh, and Arthur S Abramson. 1964. A Cross-Language Study of Voicing in Initial Stops: Acoustical Measurements. *Word* 20:384–422.

Lisker, Leigh, and Arthur S Abramson. 1967. Some effects of context on voice onset time in English stops. *Language and speech* 10:1–28.

Mack, Molly. 1989. Consonant and vowel perception and production: Early English-French bilinguals and English monolinguals. *Perception & Psychophysics* 46:187–200.

MacKenzie, Laurel. 2017. Frequency effects over the lifespan: a case study of Attenborough's r's. *Linguistics Vanguard* 3.

Martinet, André. 1952. Function, structure, and sound change. *Word* 8:1–32.

Masuya, Yoshiro. 1997. Voice onset time of the syllable-initial /p/, /t/and /k/ followed by an accented vowel in Lowland Scottish English. In *Phonetics and phonology: Selected papers*, 139–172.

Maye, Jessica, and LouAnn Gerken. 2000. Learning phonemes without minimal pairs. In *Proceedings of the 24th annual Boston University Conference on Language Development*, volume 2, 522–533.

Maye, Jessica, Janet F. Werker, and LouAnn Gerken. 2002. Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition* 82:B101–B111.

McFarland, David H. 2001. Respiratory markers of conversational interaction. *Journal of Speech, Language, and Hearing Research* 44:128–143.

McMurray, Bob, Richard N Aslin, and Joseph C Toscano. 2009. Statistical learning of phonetic categories: Insights from a computational approach. *Developmental Science* 12:369–378.

McQueen, James M, Anne Cutler, and Dennis Norris. 2006. Phonological Abstraction in the Mental Lexicon. *Cognitive Science* 30:1113–1126.

Mielke, Jeff. 2005. Modeling Distinctive Feature Emergence. In *Proceedings of the West Coast Conference on Formal Linguistics*, 281–289.

Mielke, Jeff, Kuniko Nielsen, and Lyra V Magloughlin. 2013. Phonetic imitation by individuals with Autism Spectrum Disorders: Investigating the role of procedural and declarative memory. *Proceedings of Meetings on Acoustics ICA2013* 19:e1–e8.

Millasseau, Julien, Laurence Bruggeman, Ivan Yuen, and Katherine Demuth. 2019. Durational cues to place and voicing contrasts in Australian English word-initial stops. In *Proceedings of the 19th International Congress of Phonetic Sciences, Melbourne, Australia 2019*, ed. Sasha Calhoun, Paola Escudero, Marija Tabain, and Paul Warren, 3759–3762.

Mitterer, Holger, and Mirjam Ernestus. 2008. The link between speech perception and production is phonological and abstract: Evidence from the shadowing task. *Cognition* 109:168–173.

Munro, Murray J, Tracey M Derwing, and James E Flege. 1999. Canadians in Alabama: A perceptual study of dialect acquisition in adults. *Journal of Phonetics* 27:385–403.

Nagle, Charles. 2019. Perception, imitation, production: Exploring a three-way perception-production link. In *Proceedings of the 19th International Congress of Phonetic Sciences, Melbourne, Australia 2019*, ed. Sasha Calhoun, Paola Escudero, Marija Tabain, and Paul Warren, 1248–1252.

Namy, Laura L, Lynne C Nygaard, and Denise Sauerteig. 2002. Gender differences in vocal accommodation: The role of perception. *Journal of Language and Social Psychology* 21:422–432.

Natale, Michael. 1975. Convergence of mean vocal intensity in dyadic communication as a function of social desirability. *Journal of Personality and Social Psychology* 32:790.

Nielsen, Kuniko. 2008. The specificity of allophonic variability and its implications for accounts of speech perception. Doctoral Dissertation, University of California, Los Angeles.

Nielsen, Kuniko. 2011. Specificity and abstractness of VOT imitation. *Journal of Phonetics* 39:132–142.

Nielsen, Kuniko, and Rebecca Scarborough. 2019. Perceptual Target of Phonetic Accommodation: A Pattern within a Speaker's Phonetic System or the Raw Acoustic Signal? In *Proceedings of the 19th International Congress of Phonetic Sciences, Melbourne, Australia 2019*, ed. Sasha Calhoun, Paola Escudero, Marija Tabain, and Paul Warren, 1635–1639.

Norscia, Ivan, and Elisabetta Palagi. 2011. Yawn contagion and empathy in homo sapiens. *PloS one* 6.

Nycz, Jennifer. 2013. Changing words or changing rules? Second dialect acquisition and phonological representation. *Journal of Pragmatics* 52:49–62.

Nygaard, Lynne C, and Jennifer S Queen. 2000. The role of sentential prosody in learning voices. *The Journal of the Acoustical Society of America* 107:2856–2856.

Ohala, John J. 1983. The origin of sound patterns in vocal tract constraints. In *The Production of Speech*, 189–216. Springer.

Ohala, John J. 1989. Sound change is drawn from a pool of synchronic variation. In *Language change: Contributions to the study of its causes*, 173–198.

Ohala, John J. 2011. Accommodation to the Aerodynamic Voicing Constraint and its Phonological Relevance. In *Proceedings of the 17th International Congress of Phonetic Sciences, Hong Kong*, 64–67.

Olmstead, Annie, Navin Viswanathan, M. Pilar Aivar, and Sarath Manuel. 2013. Comparison of native and non-native phone imitation by English and Spanish speakers. *Frontiers in Psychology* 4:475.

Osgood, Charles E. 1964. Semantic Differential Technique in the Comparative Study of Cultures. *American Anthropologist* 66:171–200.

Osgood, Charles Egerton, George J Suci, and Percy H Tannenbaum. 1957. *The Measurement of Meaning*. 47. University of Illinois Press.

Pardo, Jennifer S. 2006. On phonetic convergence during conversational interaction. *The Journal of the Acoustical Society of America* 119:2382–2393.

Pardo, Jennifer S. 2010. Expressing oneself in conversational interaction. In *Expressing oneself/expressing oneś self: Communication, cognition, language, and identity*, ed. Morsella, Ezequiel, 183–196. Psychology Press.

Pardo, Jennifer S, Isabel Cajori Jay, Risa Hoshino, Sara Maria Hasbun, Chantal Sowemimo-Coker, and Robert M Krauss. 2013a. Influence of role-switching on phonetic convergence in conversation. *Discourse Processes* 50:276–300.

Pardo, Jennifer S, Rachel Gibbons, Alexandra Suppes, and Robert M Krauss. 2012. Phonetic convergence in college roommates. *Journal of Phonetics* 40:190–197.

Pardo, Jennifer S, Kelly Jordan, Rolliene Mallari, Caitlin Scanlon, and Eva Lewandowski. 2013b. Phonetic convergence in shadowed speech: The relation between acoustic and perceptual measures. *Journal of Memory and Language* 69:183–195.

Pardo, Jennifer S, Adelya Urmanche, Sherilyn Wilman, and Jaclyn Wiener. 2017. Phonetic convergence across multiple measures and model talkers. *Attention, Perception, & Psychophysics* 79:637–659.

Payne, Arvilla. 1980. Factors controlling the acquisition of the Philadelphia dialect by out-of-state children. In *Locating language in time and space*, ed. William Labov, 143–78.

Peirce, Jonathan W. 2007. Psychopy—psychophysics software in python. *Journal of neuroscience methods* 162:8–13.

Pickering, Martin J, and Simon Garrod. 2004. Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences* 27:169–190.

Pierrehumbert, Janet. 2001. Exemplar dynamics: Word frequency, lenition, and contrast. In *Frequency effects and the emergence of lexical structure*, ed. Joan Bybee and Paul Hopper, 137–157. John Benjamins, Amsterdam.

Pierrehumbert, Janet B. 2006. The next toolkit. *Journal of phonetics* 34:516–530.

R Core Team. 2013. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria. Retrieved from this website.

Rahman, Jacquelyn. 2008. Middle-class African Americans: Reactions and attitudes toward African American English. *American Speech* 83:141–176.

Reichel, Uwe D, Štefan Beňuš, and Katalin Mády. 2018. Entrainment profiles: Comparison by gender, role, and feature set. *Speech Communication* 100:46–57.

Rojczyk, Arkadiusz. 2011. Perception of the English Voice Onset Time Continuum by Polish Learners. In *The Acquisition of L2 Phonology*, 37–58. Multilingual Matters.

Roon, Kevin D, and Adamantios I Gafos. 2013. A dynamical model of the speech perception-production link. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, 35.

Ros, Raquel, Alexandre Coninx, Yiannis Demiris, Georgios Patsis, Valentin Enescu, and Hichem Sahli. 2014. Behavioral Accommodation Towards a Dance Robot Tutor. In *Proceedings of the 2014 ACM/IEEE International Conference on Human-Robot Interaction*, 278–279.

Ryalls, Jack, Marni Simon, and Jerry Thomason. 2004. Voice Onset Time production in older Caucasian-and African-Americans. *Journal of Multilingual Communication Disorders* 2:61–67.

Ryalls, John, Allison Zipprer, and Penelope Baldauff. 1997. A Preliminary Investigation of the Effects of Gender and Race on Voice Onset Time. *Journal of Speech, Language, and Hearing Research* 40:642–645.

Ryan, Ellen B, and Miguel A Carranza. 1975. Evaluative reactions of adolescents toward speakers of standard English and Mexican American accented English. *Journal of Personality and Social Psychology* 31:855.

Sanker, Chelsea. 2020. Dimensions of Convergence. In *Talk given at the NYU Phonetics and Experimental Phonology Laboratory, February 14, 2020*.

Sankoff, Gillian, and Hélène Blondeau. 2007. Language change across the lifespan: /r/ in Montreal French. *Language* 83:560–588.

Scanlon, Michael, and Alicia Beckford Wassink. 2010. African American English in urban Seattle: Accommodation and intraspeaker variation in the Pacific Northwest. *American Speech* 85:205–224.

Schertz, Jessamyn, Melissa Paquette-Smith, and Elizabeth K. Johnson. 2019. The relationship between perceptual similarity judgments and VOT convergence in a shadowing task. In *Proceedings of the 19th International Congress of Phonetic Sciences, Melbourne, Australia 2019*,

ed. Sasha Calhoun, Paola Escudero, Marija Tabain, and Paul Warren, 3711–3715.

Schuhmann, Katharina S, and Marie K Huffman. 2019. Development of L2 Spanish VOT before and after a brief pronunciation training session. *Journal of Second Language Pronunciation* 5:402–434.

Schweitzer, Antje, Wolfgang Wokurek, and Peter Manfred Pützer. 2019. Convergence of Harmonic Voice Quality Parameters in Spontaneous Dialogues. In *Proceedings of the 19th International Congress of Phonetic Sciences, Melbourne, Australia 2019*, ed. Sasha Calhoun, Paola Escudero, Marija Tabain, and Paul Warren, 363–367.

Scobbie, James M. 2006. Flexibility in the face of incompatible English VOT systems. In *Laboratory Phonology 8: Varieties of Phonological Competence*, 367–392. Mouton de Gruyter.

Serniclaes, Willy. 2005. On the invariance of speech percepts. *ZAS Papers in Linguistics* 40:177–194.

Seyfarth, Scott, and Marc Garellek. 2018. Plosive voicing acoustics and voice quality in Yerevan Armenian. *Journal of Phonetics* 71:425–450.

Shockley, Kevin, Laura Sabadini, and Carol A Fowler. 2004. Imitation in shadowing words. *Perception & Psychophysics* 66:422–429.

Shultz, Amanda A, Alexander L Francis, and Fernando Llanos. 2012. Differential cue weighting in perception and production of consonant voicing. *The Journal of the Acoustical Society of America* 132:EL95–EL101.

Siegel, Jeff. 2010. *Second Dialect Acquisition*. Cambridge University Press.

Simonet, Miquel, Joseph V Casillas, and Yamile Díaz. 2014. The effects of stress/accent on VOT depend on language (English, Spanish), consonant (/d/,/t/) and linguistic experience (monolinguals, bilinguals). In *Speech Prosody 7: Proceedings of the 7th International Conference on Speech Prosody: Social and Linguistic Speech Prosody. Trinity College.*, 2333–2042.

Smith, Bruce L. 1978. Effects of place of articulation and vowel environment on voiced stop consonant production. *Glossa* 12:163–175.

Solanki, Vijay, Alessandro Vinciarelli, Jane Stuart-Smith, and Rachel Smith. 2015. Measuring mimicry in task-oriented conversations: degree of mimicry is related to task difficulty. In *Sixteenth Annual Conference of the International Speech Communication Association*.

Solé, Maria-Josep. 2018. Articulatory adjustments in initial voiced stops in Spanish, French and English. *Journal of Phonetics* 66:217–241.

Sonderegger, Morgan. 2012. Phonetic and phonological dynamics on reality television. Doctoral Dissertation, University of Chicago.

Sonderegger, Morgan. 2015. Trajectories of voice onset time in spontaneous speech on reality TV. *The Scottish Consortium for ICPhS* .

Sonderegger, Morgan, Jane Stuart-Smith, Thea Knowles, Rachel MacDonald, and Tamara Rathcke. 2020. Structured heterogeneity in Scottish stops over the twentieth century. *Language* 96:94–125.

Spengler, Stephanie, D Yves von Cramon, and Marcel Brass. 2010. Resisting motor mimicry: control of imitation involves processes central to social cognition in patients with frontal and temporo-parietal lesions. *Social Neuroscience* 5:401–416.

Stuart-Smith, Jane, Morgan Sonderegger, Tamara Rathcke, and Rachel Macdonald. 2015. The private life of stops: VOT in a real-time corpus of spontaneous Glaswegian. *Laboratory Phonology* 6:505–549.

Szabó, Ildikó Emese. 2019. Phonetic Selectivity in Accommodation: The Effect of Chronological Age. In *Proceedings of the 19th International Congress of Phonetic Sciences, Melbourne, Australia*, 3195–3199.

Szabó, Ildikó Emese. to appear. A computational model of phonotactic acquisition: Predictability of exceptional patterns in Hungarian. *Linguistic Variation* .

Takahashi, Chikako. 2020. Your perception changes how you say it?(!) Discrimination ability as a predicting factor in L1 phonetic drift. In *Poster given at the LSA 2020 Annual Meeting*.

Thorne, Barrie, and Nancy Henley. 1975. *Language and sex: Difference and Dominance*. Newbury House, Rowley, MA.

Tobin, Stephen. 2013. Phonetic accommodation in Spanish-English and Korean-English bilinguals. In *Proceedings of Meetings on Acoustics ICA2013, JASA*, volume 19.

Tobin, Stephen J. 2015. A dynamic approach to phonetic change. In *Proceedings of the 18th ICPhS, Glasgow, UK*, ed. The Scottish Consortium for ICPhS 2015.

Tobin, Stephen J, Hosung Nam, and Carol A Fowler. 2017. Phonetic drift in Spanish-English bilinguals: Experiment and a self-organizing model. *Journal of Phonetics* 65:45–59.

Trubetzkoy, Nikolai. 1939. *Grundzüge der Phonologie.*. Prague. [Bd 7, der Travaux du Cercle Linguistique de Prague.]. English transl. by C. Baltaxe (1969) Principles of phonology. Berkeley, CA: University of California Press.

Trudgill, Peter. 1981. Linguistic Accommodation: Sociolinguistic Observations on a Sociopsychological Theory. *Papers from the Parasession on Language and Behavior* 218—-237.

Trudgill, Peter. 1986. *Dialects in Contact*. Blackwell.

Trudgill, Peter. 2004. *New-dialect formation: The inevitability of colonial Englishes*. Oxford University Press, USA.

Trudgill, Peter. 2008. Colonial dialect contact in the history of European languages: On the irrelevance of identity to new-dialect formation. *Language in Society* 37:241–254.

Tyack, Peter L. 2008. Convergence of calls as animals form social bonds, active compensation for noisy communication channels, and the evolution of vocal learning in mammals. *Journal of Comparative Psychology* 122:319.

Vallabha, Gautam K, James L McClelland, Ferran Pons, Janet F. Werker, and Shigeaki Amano. 2007. Unsupervised learning of vowel categories from infant-directed speech. *Proceedings of the National Academy of Sciences* 104:13273–13278.

Wade, Lacey. 2020. The Linguistic and the Social Intertwined: Linguistic Convergence toward Southern Speech. Doctoral Dissertation, University of Pennsylvania.

Wagner, Anita, Mirjam Ernestus, and Anne Cutler. 2006. Formant transitions in fricative identification: The role of native fricative inventory. *The Journal of the Acoustical Society of America* 120:2267–2277.

Wang, Xuan. 2019. Phonetic convergence of Hong Kong English: sound salience and the exemplar-based account. In *Proceedings of the 19th International Congress of Phonetic Sciences, Melbourne, Australia 2019*, ed. Sasha Calhoun, Paola Escudero, Marija Tabain, and Paul Warren, 2334–2338.

Watson, Catherine I, Margaret Maclagan, and Jonathan Harrington. 2000. Acoustic evidence for vowel change in New Zealand English. *Language Variation and Change* 12:51–68.

Webb, James T. 1970. Interview synchrony: An investigation of two speech rate measures in an automated standardized interview. In *Studies in dyadic communication new york*, 55–70.

Pergamon, New York.

Wedel, Andrew, Scott Jackson, and Abby Kaplan. 2013a. Functional load and the lexicon: Evidence that syntactic category and frequency relationships in minimal lemma pairs predict the loss of phoneme contrasts in language change. *Language and speech* 56:395–417.

Wedel, Andrew, Abby Kaplan, and Scott Jackson. 2013b. High functional load inhibits phonological contrast loss: A corpus study. *Cognition* 128:179–186.

Welkowitz, Joan, Ronald N Bond, and Stanley Feldstein. 1984. Conversational time patterns of Japanese-American adults and children in same and mixed-gender dyads. *Journal of Language and Social Psychology* 3:127–138.

Westbury, John R, and Patricia A Keating. 1986. On the naturalness of stop consonant voicing. *Journal of Linguistics* 22:145–166.

Winn, Matthew B, Monita Chatterjee, and William J Idsardi. 2013. Roles of voice onset time and F0 in stop consonant voicing perception: Effects of masking noise and low-pass filtering. *Journal of Speech, Language, and Hearing Research* 56:1097–1107.

Wolfram, Walt. 2007. Sociolinguistic folklore in the study of African American English. *Language and Linguistics Compass* 1:292–313.

Xu, Yang, and David Reitter. 2016. Convergence of Syntactic Complexity in Conversation. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*, volume 2: Short Papers, 443–448.

Yaeger-Dror, Malcah. 1988. The influence of changing group vitality on convergence toward a dominant linguistic norm: An Israeli example. *Language & Communication* 8:285–305.

Yoon, Jennifer MD, and Claudio Tennie. 2010. Contagious yawning: a reflection of empathy, mimicry, or contagion? *Animal Behaviour* 79:e1–e3.

Ytsma, Johannes. 1988. Bilingual classroom interaction in friesland. In *Bilingualism and the Individual, Multilingual Matters*, 53–68. Philadelphia, PA, USA : Multilingual Matters.

Yu, Alan, Carissa Abrego-Collier, Rebekah Baglini, Tommy Grano, Martina Martinovic, Charles Otte III, Julia Thomas, and Jasmin Urban. 2011. Speaker attitude and sexual orientation affect phonetic imitation. *University of Pennsylvania Working Papers in Linguistics* 17:26.

Zahn, Christopher J, and Robert Hopper. 1985. Measuring language attitudes: The speech evaluation instrument. *Journal of Language and Social Psychology* 4:113–123.