SILENCE-CUED STOP PERCEPTION:
SPLIT DECISIONS

BRIDGET SAMUELS & BERT VAUX

University of Southern California
University of Cambridge

ABSTRACT

Bastian, Eimas, & Liberman (1961) found that listeners heard a [p] when a silence of more than 50ms was inserted between the [s] and the [l] in a recording of the word *slit.* It has long been known that silence is an important cue in stop consonant perception. Nevertheless, it is surprising that a short interval of silence can substitute for something as acoustically and articulatorily complex as a phoneme. In the present work, we replicate and expand upon this study to further examine the phenomenon of silence-cued stop perception. We demonstrate the 'Split Effect' in a previously unexplored set of environments, analyze factors that contribute to the identity of silence-cued stops, and lay the groundwork for further investigation of the acoustic and non-acoustic factors that contribute to this perceptual illusion. Our study demonstrates an experimental paradigm for studying the genesis of such effects synchronically and in a controlled setting.

[1] INTRODUCTION

Perceptual illusions that emerge during speech perception provide a window into the way humans process auditory input and the special manner in which sounds identified as speech are interpreted. For example, it has been shown that listeners can integrate monaurally administered stimuli (e.g. [s] presented in one ear and [pa] presented in the other, perceived as *spa*), but only if the subjects believe the stimuli to be linguistic (Liberman 1982). Conversely, Werker & Tees (1984) found that a non-native phonemic contrast could be reliably distinguished by English listeners in an acoustic discrimination task, but not when they understood they were hearing linguistic stimuli. These findings suggest that the perceptual system deals with speech in a highly specialized way, and that when this speech perception mechanism is at work, it triggers a number of speech-specific processes that cannot be consciously circumvented. In the present study, we investigate some of the factors that contribute to another perceptual

illusion in which listeners perceive a small duration of silence as a stop consonant, which we call the 'Split Effect'.

The Split Effect was first documented by Bastian et al. (1961). They reported that the syllable [slɪt] is heard as *split* when a short interval of silence (c. 50–80 ms) is introduced between the frication at the beginning of the syllable and the /l/. (This finding was subsequently confirmed by Fitch et al. 1980 using synthetic speech and by Best, Morrongielo, & Robson 1981 for *say/stay* using sine wave speech.) This silence-cued stop percept correlates with the fact that a stop consonant is made by completely obstructing the vocal tract, thus creating a brief silence in the acoustic signal. It has long been known that silence is an important cue in stop consonant perception (e.g. Bastian et al. 1961; Dorman, Raphael, & Liberman 1979; Bailey & Summerfield 1980; Fitch et al. 1980; Summerfield, Bailey, Seton, & Dorman 1981; Repp 1984a, 1984b, 1985). Nevertheless, it is surprising that inserting a short interval of silence into the middle of a speech string to which no other modifications have been made can prove to be a natural-sounding substitute for something as acoustically and articulatorily complex as a phoneme. It is noteworthy moreover that the illusion is automatic and involuntary; awareness of the effect does not eliminate it.

Perceptual phenomena like the Split Effect are complex, in part due to the multidimensional parametric space that defines phonemic categories (Holt & Lotto 2010). Additionally, one parameter may compensate for another in a system of "trading relations" (Oden & Massaro 1978; Repp 1982; Hawkins 2010), such that a token lacking in one particular acoustic property of a category may be considered less deviant if it strongly shows another relevant property (cf. Fitch et al. 1980 showing silence trading with formant transitions for the perception of stops; Ohde & Stevens 1983 on integration involving place of articulation; and Repp 1984a discussing integration over many different variables involved in perception). Some parameters are more perceptually salient than others, particularly in certain environments or listening conditions, leading to a hierarchy of importance for these parameters in the ultimate assignment of a sound to a particular phonemic category (Wright 2001). This complex system of categorization and abstraction enables the appropriate labeling of phonemes despite the wide fluctuations in the acoustic input received, but it also introduces the possibility of phenomena like the Split Effect.

In order to probe the phonetic and phonological factors that contribute to the Split Effect and the circumstances under which it occurs, we first sought to

replicate[1] the Split Effect with an expanded range of responses to include other places of articulation for the silence-cued stop, rather than a forced choice between *slit* and *split.* This modification allowed us to investigate the relative frequency of percepts other than *p.* Experiment 1 served to confirm the perceptual boundary reported in the literature and to establish that *t* and, to a lesser extent, *k* may also be perceived in this context. We also found that the length of the silent interval appeared to influence the quality of the percept. In Experiment 2, we used a series of /s_Vt/[2] frames to investigate the roles of the following segment quality, the length of the silent interval, and the lexical status/frequency of the relevant words in the silence-cued percept. We discovered that the dominant silence-cued stop percept varies depending on the quality of the following vowel and confirmed that the length of the silent interval and lexical frequency also contribute to the percepts.

Our studies improve our understanding of the factors that contribute to the Split Effect, extend the range of contexts in which it has been described, and provide an example of spontaneous 'normalization' of the speech signal (Ohala 1993) in an experimental setting. Moreover, studying perceptual illusions has the potential to improve our understanding of the complicated interplay among acoustic and non-acoustic factors that drives phonemic categorization, the process that makes speech perception unique and different from auditory perception in other contexts.

[2] METHODS

[2.1] *Experiment 1*

Participants

Thirty adults with no reported history of hearing or language abnormalities participated in Experiment 1. All were native speakers of American English. Participants varied in their experience with formal linguistics and phonetics. The responses of the subjects with linguistic training all fell within the range of those of the subjects without linguistic training.

---

[1] Bastian et al. (1961) published only a brief abstract describing their results. Their experimental procedures were not described sufficiently for replication. We therefore chose to follow the experimental paradigm used by Fitch et al. (1980) to investigate trading relations between the duration of the silent interval and spectral cues in creating a stop percept. Like Fitch et al., we prepared stimuli that varied in the amount of inserted silence in 8 ms increments. However, like Bastian et al., we used recordings of natural productions of slit rather than the synthetic speech employed by Fitch et al.

[2] We use an _underscore_ to represent an interval of silence. V stands for any vowel, and C stands for any consonant.

Stimuli

Tokens of /slɪt/ produced by two native speakers of American English, one male and one female, were recorded using Audacity. Utilizing both spectrogram and waveform data, we then located the end of the fricative noise associated with the /s/ and cut each digitized token at that location, making sure to make the cut at a zero point on the waveform in order to avoid unwanted perceptual artifacts. The experimental stimuli were created from these recordings by inserting silent intervals of 0-160 milliseconds, in increments of 8ms, at the cut point. Because the male recording showed anticipatory formant trails starting approximately halfway through the frication, which we hypothesized might affect the results of the experiment, we created a third series of stimuli using the same process, but first removing the portion of the /s/ containing the trails. Impressionistically these stimuli still sounded natural, and the data suggest that altering the frication in this way may have enhanced rather than reduced the perceptual illusion.

Procedure

Each set of stimuli (male, male with excised formant trails, and female) constituted a block in the experiment. The three blocks were presented in randomized order, and the presentation order of stimuli was also randomized within each block. The stimuli were presented via SuperLab software on a personal computer in a quiet room, with each stimulus being played three times in a row with approximately one second between repetitions. Subjects had a choice of pressing five keys, corresponding to hearing *sklit*, *split*, *stlit*, *slit*, or none of these. The five options were continuously presented on the computer monitor. Statistical analyzes were performed in Microsoft Excel and Matlab.

[2.2] *Experiment 2*

Participants

Thirty adults with no reported history of hearing or language abnormalities participated in Experiment 2. All were native speakers of American or British English. The participants in Experiment 2 did not previously participate in Experiment 1. Participants varied in their experience with formal linguistics and phonetics. The responses of the subjects with linguistic training all fell within the range of those of the subjects without linguistic training.

Stimuli

The stimuli for Experiment 2 (Table 1) were recorded by a male native speaker of American English using Audacity. The familiarization stimuli included one

word with a simplex [s] onset, one word beginning with [sk] onset, one beginning with [sp], and one beginning with [st]. Test stimuli included the vowels [æ, eɪ, ij, ɛ, ɪ, ʊ, oʊ, ɑ, uw], each presented in the frame /sVt/. Filler items included a total of nine English words with [sk], [sp], and [st] onsets that also conformed to the /sVt/ frame. The familiarization and filler items were presented as recorded. Silent intervals in test items were created as described for Experiment 1, but with silences of 0–120 milliseconds in increments of 30ms.

Familizarization	Test items	Filler items
song	sat (/k, p, t/)	scoot
skid	sate (/k, p, t/)	skate
spoon	seat (/k/)	skeet
scale	set (/t/)	spat
	sit (/k, p/)	spit
	soot	spot
	sote [soʊt] (/t/)	state
	sought (/k, p/)	stet
	suit (/k/)	stoat

TABLE 1: Familiarization, test, and filler items used in Experiment 2. Consonants in parentheses in the test item column indicate which insertions into each frame are attested words of English.

Procedure

Experiment 2 was presented online using the Qualtrics browser-based software package. Participants were able to use their own personal computers to complete the experiment. Instructions indicated to make sure to be in a quiet room, free of any distractions and with the volume set to a comfortable level. The experiment was divided into four blocks: an initial familiarization block followed by three test blocks. Prior to the familiarization block, subjects were presented with instructions to press the S key when an item started with a 'plain S sound' as in *sand*, the K key for items starting with an SK sound as in *scanned*, the P key for items starting with SP as in *spanned,* or the T key for an ST sound as in *stand.* In the familiarization phase, each participant was presented with each of the four familiarization items. Each item played automatically only once, though a playback button gave the option to repeat each stimulus if desired. Upon pressing any key, the software advanced to the next item with no possibility of backtracking. Participants performed with over 98% accuracy (118/120) in the familiarization phase.

The instructions were repeated but no accuracy feedback was given at the end of the familiarization phase. Each test block consisted of all 45 test items (9 stimuli with 0, 30, 60, 90, and 120 ms inserted silence) and 9 filler items presented in random order, in the same manner as in the familiarization block. Participants performed with at least 96% accuracy on all filler items. Statistical analyzes were performed in Microsoft Excel and R statistical software.

[3]   RESULTS AND DISCUSSION

[3.1] *Experiment 1*

Based on the prior literature on the Split Effect, we expected participants to show a categorical boundary, as evidenced by an abrupt shift from *slit* responses (for stimuli with shorter silent intervals) to responses indicating the perception of a silence-cued stop for stimuli with longer silent intervals. We analyzed each subject's responses in Experiment 1 individually to identify (a) the duration of the silent interval at which the participant *first* reported hearing a stop, which we call 'time to first C' in Table 2, and (b) the span throughout which the participant *only* reported hearing stops. We refer to the lower bound of (b) here as the 'time to only C responses'. For example, if a participant responded *slit* to the 0–32ms stimuli, *stlit* to the 40ms stimulus, *slit* to the 48 and 56ms stimuli, and some combination of *stlit* and *split* to the 64–160ms stimuli, we would say this participant had a time to only C responses of 64ms.

One participant was excluded in each group due to pressing invalid keys such that their responses could not be interpreted. All remaining participants showed the expected pattern, indicative of categorical perception, for the male series with excised formant trails. For the male and female series, 28/29 participants exhibited a threshold response, such that we could characterize a mean time to only C responses. Perhaps because of the lower strength of the formants in the female series, the variation was greater in that series. Paired two-tailed Student's T-tests were used for pairwise comparison of the mean time to first C response and categorical boundary across series. The only statistically significant difference was between the mean time to only C responses in the female series and the male series with excised formant trails ($p < 0.001$).

	Female	Male	Male cut /s/
Mean time to first C	49.7 ms	45.2 ms	45.0 ms
Median time to first C	56 ms	40 ms	48 ms
IQR of time to first C	32-64 ms	32-48 ms	40-56 ms
Mean time to only C	92.1 ms	72.0 ms	57.7 ms
Median time to only C	104 ms	56 ms	56 ms
IQR of time to only C	56-122 ms	40-96 ms	48-64 ms

TABLE 2: Mean, median, and interquartile range (IQR) of time to first reported silence-cued stop percept (time to first C); mean, median, and interquartile range (IQR) of time to time to only C responses for the female speaker, male speaker, and male speaker with excised formant trails (male cut /s/). N=29 per group.

Plotting the summed percentage of *p, t,* and *k* responses against the duration of the silent interval (Figure 1) yields an upward-sloping sigmoidal curve. This type of curve is considered to be indicative of the categorical nature of phoneme perception (McMurray, Spivey, & Aslin 2000).
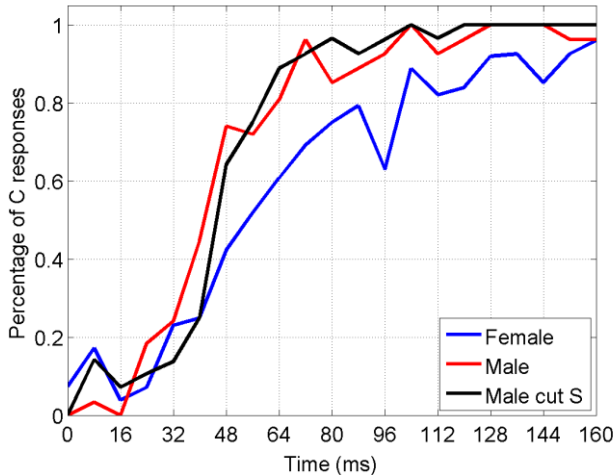

FIGURE 1: Summed percentages of consonant (C) responses plotted against the duration of the silent interval (time, ms) for the female speaker, male speaker, and male speaker with excised formant trails (male cut S).

The categorical boundary can be defined as the point at which the percentages of *slit* responses and C responses are each 50%. Judging by this criterion as well as the mean time to first C and mean time to only C, the results of our experiment fall within or near the range of threshold values reported by Fitch et al. (1981)

and Best et al. (1981) using the same criterion, namely from 50–80 ms. Some variation is expected due to differences in methodology. In particular, these earlier studies did not give subjects the option of reporting perceptions of three places of articulation. Repp (1984a) allowed subjects to report hearing *stlit* or *split*, but because *stlit* responses were reportedly infrequent among the small subject pool used in the experiment, they were combined with the *split* responses in the analyzes; additionally, the Repp study began with a token of *split* with a modified burst, not a token of *slit*, and thus may have biased the perceptual boundary towards a consonant percept since the burst contributes towards the perception of a consonant in this context (Repp 1984a).

Analyzing the trends in individual consonant responses over time (Figure 2) revealed some trends which we investigated further in Experiment 2. While *split* responses were in the majority, there were also non-trivial numbers of *stlit* and *sklit* responses, with the former outnumbering the latter. This indicates that factors other than lexical status (i.e., the fact that *split* is a word while *sklit* and *stlit* are not) may play a role in determining the identity of the silence-cued percept. It is tempting to suspect a connection to cluster frequency, since /spl/ is a common cluster in English whereas /skl/ only appears in a few infrequent words (e.g. *sclerosis*) and /stl/ is phonotactically illicit in all positions (Pierrehumbert 1994).
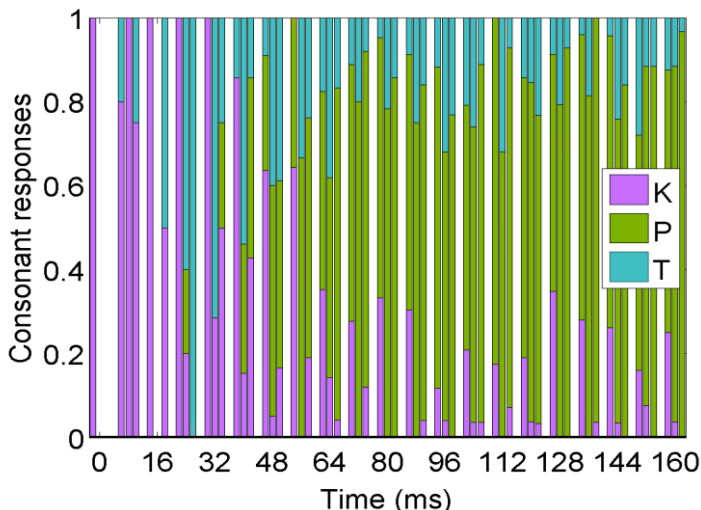

FIGURE 2: Normalized ratios of consonant responses (to all three series of stimuli) for silent intervals ranging in duration from 0-160 ms. At each time point the three series are plotted separately: female (left), male (middle), and male with excised formant trails (right). *K* responses appear in purple; *p* responses in green; *t* responses in blue.

[3.2] *Experiment 2*

We first analyzed how the pattern of responses varied as the duration of the silent interval increased (Figure 3). As expected based on the literature and on Experiment 1, the proportion of *s* (no epenthesis) responses decreased from nearly 100% for the unaltered stimuli to 16% for the stimuli with 120 ms of silence.

Generalized linear mixed models were used to investigate the effect of silent interval duration on consonant selection. This allowed subjects to be treated as a random effect, consistent with our repeated measures experimental design. The dependent variable was treated as a binomial distribution, with the response of interest coded as 1 and all other responses assigned a value of 0.

The two predictors included in these models were silent interval duration (0, 30, 60, 90, 120) and lexical frequency of the response, as reported by the spoken portion of the Corpus of Contemporary American English (http://corpus.byu.edu/coca/). Lexical frequency was operationalized as a proportion relative to the frequency of the other available options for that trial. For instance, in the /sæt/ frame, there were four possible responses: *sat*, *scat*, *spat*, and *stat*. *Sat* appears 4535 times in the corpus, while the other words appear 18, 103, and 57 times, respectively. Thus, the lexical frequency value assigned to a *sat* response was .96, .02 to a response of *spat*, and so on (non-words or words that did not appear in the corpus were assigned a value of 0). Word frequency was included in the models to account for the possible influence of lexical accessibility, independent of silent interval duration.
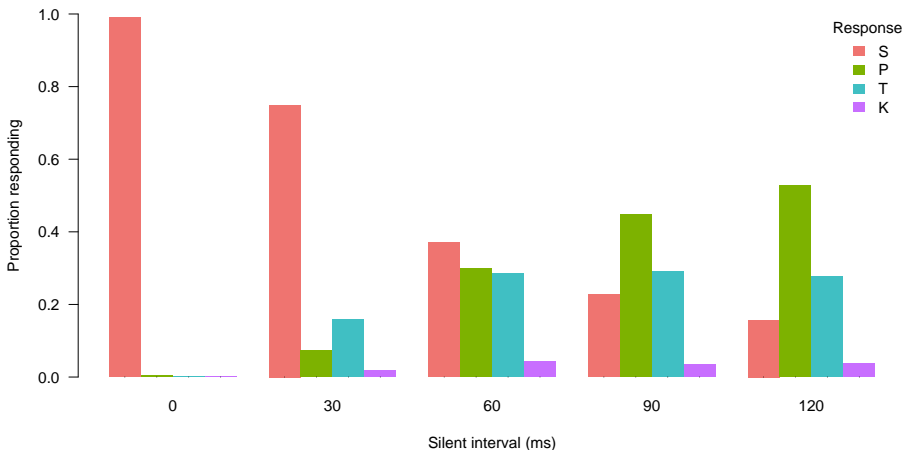

FIGURE 3: Proportion of responses plotted against the duration of the silent interval (ms). *S* responses appear in red; *k* responses in purple; *p* responses in green; *t* responses in blue.
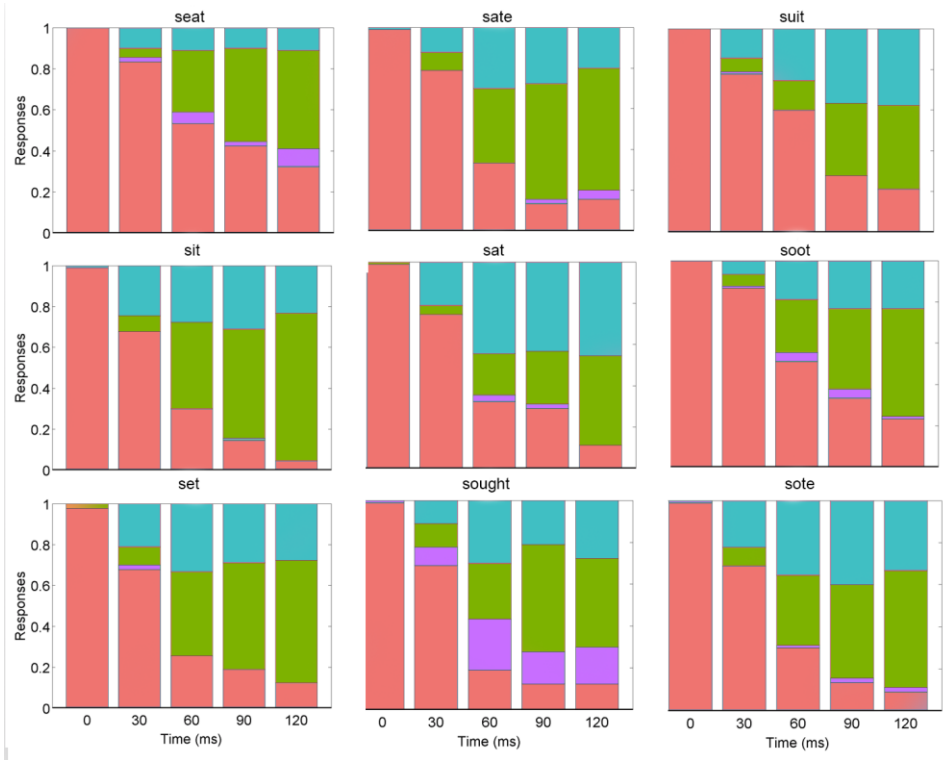
FIGURE 4: Proportion of responses plotted against the duration of the silent interval, separated by /sVt/ frame. Top row, left to right: seat, sate, suit. Middle row, left to right: sit, sat, soot. Bottom row, left to right: set, sought, sote. S responses appear in red; k responses in purple; p responses in green; t responses in blue.

We first set out to establish whether *s* responses decreased as the duration of the silent interval increased. This model, which treated *s* responses as the dependent measure, revealed that *s* responses indeed became significantly less frequent as silent interval duration increased (β = -.088, SE = .004, z = -21.74, p < .001). Again in agreement with prior studies and Experiment 1, the percentage of *s* responses was 75% for the stimuli with 30 ms of silence but dropped to 37% for the stimuli with 60 ms of silence. This is consistent with a threshold response or categorical boundary between 30–60ms. As expected, lexical frequency also contributed to the likelihood of *s* responses (β = 1.82, SE = .30, z = 6.11, p < .001) and the interaction between lexical frequency and silent interval duration was significant (β = .043, SE = .005, z = 8.88, p < .001). The split effect therefore cannot be accounted for by lexical frequency alone.

The lack of *k* responses noted in Experiment 1 became even more striking in

Experiment 2. The percentage of *k* responses ranged from 2–4% for the stimuli with 60–120 ms silent intervals, whereas the percentages of *p* and *t* responses ranged from 28–53% for those stimuli. Our models confirmed that there were more *p* (β = .044, SE = .002, z = 26.5, p < .001) and *t* (β = ., SE = ., z = , p < .001) responses as the duration of the silent interval increased. *K* responses were so infrequent that a model of their pattern could not be reliably estimated.

Both *p* and *t* responses increase as a function of silence interval. But are either of these responses especially popular for certain interval durations? Although our samples are not fully independent, Chi-square tests can give some indication. Chi-square tests lend support to the notion that *t* responses predominated over *p* responses for the stimuli with a 30 ms silent interval (X^2 = 25.19, df = 1, p < .001). Neither *t* nor *p* responses were predominant for the 60 ms silent interval (X^2 = .30). *P* responses were more common than *t* responses for the stimuli with 90 (X^2 = 26.93, df = 1, p < .001) and 120 (X^2 = 62.58, df = 1, p < .001) ms of silence (see also Figure 3). An explanation in terms of differing voice onset time (VOT) could be entertained here. However, this seems unlikely since a number of studies have shown that [p], [t], and [k] all have average VOT values in or near the 50–80ms range, though the average VOT for [p] tends to be on the shorter end of it (Lisker & Abramson 1965; Sweeting & Baken 1982; Hardcastle, Barry, & Clark 1985; Morris & Brown 1987; Brown, Morris, & Weiss 1993). The possibility remains that /t/ could be associated with a shorter VOT due to its allophone [ɾ] in English, which has an average duration of 10–40 ms. However, flapping does not occur in the interconsonantal environment in which the silent interval is presented here.

Analyzing the responses to stimuli created from each /sVt/ frame separately reveals that properties of the following vowel may have an effect on the silence-cued stop percept (Figure 4). While the effects of neighboring segments on silence-cued stop percepts remain to be investigated more fully, we can identify at least one point of interest here. Notably, 56% of the *k* responses (61/109) occurred with the *sought* frame (Figure 4, bottom middle panel).

One potential explanation for this unexpectedly high proportion of *k* responses in the *sought* frame could be lexical frequency. Table 4 presents the frequencies of English words that conform to each frame (e.g. *sat – scat – spat – stat*), as reported in the spoken portion of the Corpus of Contemporary American English (http://corpus.byu.edu/coca/). Within the *sought* frame, the variant containing [k] (e.g. *scot, Scott*) is the most frequent. Although the dominant percept was *p* in this frame, the frequency of *Scott/scot* may have boosted the number of *k* percepts, which were rare across all frames tested.

	% of total S	% of total K	% of total P	% of total T
sat	11%	5%	8%	17%
sate	11%	6%	13%	10%
seat	14%	16%	10%	5%
set	10%	2%	14%	12%
sit	10%	1%	14%	12%
soot	13%	9%	10%	8%
sote [sowt]	10%	6%	12%	14%
sought	9%	56%	11%	10%
suit	13%	1%	8%	13%
Total number of responses	2022	109	1096	823
Percentage of total responses	50%	3%	27%	20%

TABLE 3: Percentage of responses by vowel frame. The total number of responses to test items was 4050 (9 vowel frames x 5 silent intervals x 3 blocks x 30 participants).

	/sVt/	/skVt/	/spVt/	/stVt/
sat	4535	18	103	57
sate	17	356	76	53587
seat	4940	14	0	0
set	22950	0	0	1
sit	10727	114	407	0
soot	80	0	0	0
sote [sowt]	0	0	0	0
sought	1342	9903	3413	0
suit	3268	54	0	0

TABLE 4: Frequencies of /s(C)Vt/ words based on the spoken portion of the Corpus of Contemporary American English. The most frequent consonant response for each frame is shaded in gray.

[4]   GENERAL DISCUSSION AND CONCLUSIONS

Several general conclusions emerge from the experiments discussed in this paper. First, we established that even in the 'canonical' Split Effect environment (/s_lɪt/), silence-cued stops other than *p* are perceived. Interestingly, *k* is a very infrequent percept whereas *t* is relatively frequent, regardless of the fact that /stl/ is an illicit onset cluster in English. This dispreference for *k* was apparent in both Experiment 1 and 2. As Stevens & Blumstein (1978) report, the burst of

[k] (in the syllable [ki] 'has an "extra," non-robust feature: a compact mid-frequency spectral peak.' This additional feature likely serves to differentiate [k] from the other stop consonants. Moreover, the formant transitions from [k] to a vowel or lateral begin with a low F1 and high F2 and then converge, quite unlike the formant transitions seen with a alveolar or labial stop in this context, which are much more similar to each other than they are to those of a velar. Thus, the lack of uniquely salient cues that might indicate a velar stop likely reduces *k* percepts in experiments like the ones reported here (see also Plauche, Ohala, & Deloglu 1997 and Hawkins 2010).

Further investigation is needed to clarify why *p* is the dominant percept in some situations but *t* prevails in the remainder. One possible explanation has been discussed by Mann & Repp (1980) with regards to the perception of stops after the voiceless fricatives [s] and [ʃ]. Humans expect certain coarticulatory effects in speech and automatically 'undo' or 'normalize' them (Ohala 1993). Participants in tasks like ours may anticipate coarticulation between vowels and the stops preceding them, and in the process of attempting to reverse it, color their perceptions of the silence-cued stops. The reliance on cues from the following vocalic portion rather than from the preceding consonant is perhaps predicted from the preference listeners give to CV transitions over VC transitions when the two conflict, as reported by Fujimura, Macchi, & Streeter (1978) and Ohala (1990). Normalization processes like these contribute to the phonologization of phonetic patterns, and its fossilized effects are the subject of study in listener-based models of sound change (e.g. Blevins 2004). Our study demonstrates an experimental paradigm in which the genesis of these effects can be studied synchronically and in a controlled setting.

In Experiment 2, we tested the Split Effect in nine /s_Vt/ frames. We found that the identity of the vowel following the silent interval and lexical frequency play roles in shaping the stop percept. The results of Experiment 2 collectively call into question the hypothesis that *split* is the dominant percept in the canonical Split Effect context due to the fact that it is a word in English while [sklɪt] and [stlɪt] are not, or because /spl/ is a more frequent cluster than the rare /skl/ and the illicit /stl/. This conclusion is strengthened in light of our finding in Experiment 1 that there were a significant number of *stlit* responses. We found that lexical frequency plays a role in determining the silence-cued stop percept, but alone is insufficient to explain the results. More research is needed to tease apart the different types of frequency effects (e.g. cluster, lexical, or lexical neighborhood) that may be at play. Another factor that likely contributed to our results was the nature of the response task: since we asked about the identity of the onset in Experiment 2, the subjects were essentially performing

phonemic monitoring/detection. Cutler, Mehler, Norris & Segui (1987) report on several studies, some of them phonemic restoration tasks similar to ours, which suggest that phonemic monitoring direct subjects' attention to the pre-lexical level and therefore inhibits lexical effects. Furthermore, Vitevitch & Luce (1999) show that neighborhood density plays less of a role in processing of monosyllabic spoken stimuli than in longer stimuli.

The present study demonstrates the Split Effect in a previously unexplored set of environments, analyzes factors that contribute to the identity of silence-cued stops for the first time, and lays the groundwork for further investigation of the acoustic and non-acoustic factors that contribute to the Split Effect. In further studies, we plan to investigate the roles of phonotactics, prosodic structure, and other phonological factors in this phenomenon. For example, inserting a silent interval in a position where a stop is phonotactically illicit may not reliably generate a silence-cued stop percept. This may lead to language-specific differences in the circumstances under which the Split Effect can be observed. Investigations of perceptual illusions like these advance our understanding of the complex set of parameters that interact to determine phonemic categorization, the crucial process that makes speech perception different from non-speech audition.

REFERENCES

Bailey, Peter J., & Quentin Summerfield. 1980. Information in speech: observations on the perception of [s]-stop clusters. *Journal of Experimental Psychology: Human Perception and Performance* 6(3): 536–563.

Bastian, Jarvis, Peter D. Eimas , & Alvin M. Liberman. 1961. Identification and

discrimination of phonemic contrast induced by silent interval. *Journal of the Acoustical Society of America* 33: 842.

Best, Catherine T., Barbara Morrongiello, & Rick Robson. 1981. Perceptual equivalence of acoustic cues in speech and nonspeech perception. *Perception and Psychophysics* 29(3): 191–211.

Blevins, Juliette. 2004. *Evolutionary phonology.* Cambridge: Cambridge University Press.

Brown Jr., W.S., Richard. J. Morris, & Rudolf Weiss. 1993. Comparative methods for measurement of VOT. *Journal of Phonetics* 21(3): 329–336.

Cutler, Anne, Jacques Mehler, Dennis Norris, & Juan Segui. 1987. Phoneme identification and the lexicon. *Cognitive Psychology* 19: 141–177.

Dorman, Michael F., Lawrence J. Raphael, & Alvin M. Liberman. 1979. Some experiments on the sound of silence in phonetic perception. *Journal of the Acoustical Society of America* 65(6): 1518–1532.

Fitch, Hollis L., Terry Halwes, Donna M. Erickson, & Alvin M. Liberman. 1980. Perceptual equivalence of two acoustic cues for stop-consonant manner. *Perception and Psychophysics* 27(4): 343–350.

Fujimura, Osamu, Marian J. Macchi, & Lynn A. Streeter. 1978. Perception of stop consonants with conflicting transitional cues: a cross-linguistic study. *Language & Speech* 21(4): 337–346.

Hardcastle, William J., Rosemarie A. Morgan Barry, & C.J. Clark. 1985. Articulatory and voicing characteristics of adult dysarthric and verbal dyspraxic speakers: an instrumental study. *The British Journal of Disorders of Communication* 20(3): 249–270.

Hawkins, Sarah. 2010. Phonological features, auditory objects, and illusions. *Journal of Phonetics* 38(1): 60–89.

Holt, Lori L. & Andrew J. Lotto. 2010. Speech perception as categorization. *Attention, Perception, & Psychophysics* 72, 1218–1227.

Liberman, Alvin M. 1982. On finding that speech is special. *American Psychologist* 37(2): 148–167.

Lisker, Leigh & Arthur S. Abramson. 1964. A cross-language study of voicing in initial stops: acoustical measurements. *Word* 20: 384–422.

Mann, Virginia A., & Bruno H. Repp. 1980. Influence of vocalic context on perception of the [ʃ]–[s] continuum. *Perception & Psychophysics* 28(3): 213–228.

Oden, Gregg C., & Dominic W. Massaro. 1978. Integration of featural information in speech perception. *Psychological Review* 85(3): 172–191.

Ohala, John J. 1990. The phonetics and phonology of aspects of assimilation. In *Papers in Laboratory Phonology I: Between the Grammar and the Physics of Speech,* edited by John Kingston & Mary E. Beckman, 258–275. Cambridge: Cambridge University Press.

Ohala, John J. 1993. Coarticulation and phonology. *Language and Speech* 36(2,3): 155–170.

Ohde, Ralph N., & Kenneth N. Stevens. 1983. Effect of burst amplitude on the perception of stop consonant place of articulation. *Journal of the Acoustical Society of America* 74(3): 706–714.

Pierrehumbert, Janet. 1994. Syllable structure and word structure. In *Papers in Laboratory Phonology III*, edited by Patricia A. Keating, 168–188. Cambridge: Cambridge University Press.

Plauché, Madelaine, John J. Ohala, & Christina Delogu. 1997. Asymmetries in consonant confusion. *Proceedings of Eurospeech 1997: 5th European Conference on Speech Communication and Technology* 4: 2187–2190. Rhodes, Greece.

Repp, Bruno H. 1982. Phonetic trading relations and context effects: new experimental evidence for a speech mode of perception. *Psychological Bulletin* 92(1): 81–110.

Repp, Bruno H. 1984a. The role of release bursts in the perception of [s]-stop clusters. *Journal of the Acoustical Society of America* 75, 1219–1230.

Repp, Bruno H. 1984b. Effects of temporal stimulus properties on perception of the [sl]–[spl] distinction. *Phonetica* 41: 117–124.

Repp, Bruno H. 1985. Perceptual coherence of speech: stability of silence-cued stop consonants. *Journal of Experimental Psychology: Human Perception and Performance.* 11(6): 799–813.

Stevens, Kenneth N., & Sheila E. Blumstein. 1978. Invariant cues for place of articulation in stop consonants. *Journal of the Acoustical Society of America* 65: 1358–1368.

Summerfield, Quentin, Peter J. Bailey, John Seton, & Michael F. Dorman. 1981. Fricative envelope parameters and silent intervals in distinguishing 'slit' and 'split'. *Phonetica* 38: 181–192.

Sweeting, Patricia M., & Ronald J. Baken. 1982. Voice onset time in a normal-aged population. *Journal of Speech, Language, & Hearing Research* 25: 129–134.

Vitevich, Michael S. & Paul A. Luce. 1999. Probabilistic phonotactics and neighborhood activation in spoken word recognition. *Journal of Memory & Language* 40: 374–408.

Werker, Janet F. & Richard C. Tees. 1984. Phonemic and phonetic factors in adult cross-language speech perception. *Journal of the Acoustical Society of America, 75(6):* 1866–78.

Wright, Richard. 2001. Perceptual Cues in Contrast Maintenance. In *The Role of Speech Perception in Phonology*, edited by Elizabeth Hume & Keith Johnson, 251–277. San Diego: Academic Press.

CONTACTS

Bridget Samuels
University of Southern California
bdsamuel@usc.edu

Bert Vaux
University of Cambridge
bv230@cam.ac.uk