

1 Methodology

1.1 Participants

One hundred native speakers of English participated in the experiment and were compensated with either \$10 CAD or course credit. They were recruited from the UBC student population. Twenty additional native English speakers participated in a pretest to determine the most ambiguous sounds. Twenty five other native speakers of English participated for course credit in a control experiment.

1.2 Materials

One hundred and forty English words and 100 nonwords that were phonologically legal in English were used as exposure materials. The set of words consisted of 40 critical items, 20 control items and 60 filler words. Half of the critical items had an /s/ in the onset of the first syllable and half had an /s/ in the onset of the final syllable. All critical tokens formed nonwords if their /s/ was replaced with /ʃ/. Half the control items had an /ʃ/ in the onset of the first syllable and half had an /ʃ/ in the onset of the final syllable. Each critical item and control item contained just the one sibilant, with no other /s z ʃ ʒ tʃ dʒ/. Filler words and nonwords did not contain any sibilants. Frequencies and number of syllables across item types are in Table 1.

Table 1: Mean and standard deviations for frequencies and number of syllables of each item type

Item type	Frequency	Number of syllables
Filler words	1.81 (1.05)	2.4 (0.55)
/s/-initial	1.69 (0.85)	2.4 (0.59)
/s/-final	1.75 (1.11)	2.3 (0.47)
/ʃ/-initial	2.01 (1.17)	2.3 (0.48)
/ʃ/-final	1.60 (1.12)	2.4 (0.69)

Four monosyllabic minimal pairs of voiceless sibilants were selected as test items for categorization (*sack-shack*, *sigh-shy*, *sin-shin*, and *sock-shock*). Two of the pairs had a higher log frequency per million words (LFPM) from SUBTLEXus (Brysbaert and New, 2009) for the /s/ word, and two had higher LFPM for the /ʃ/ word, as shown in Table 2.

All words and nonwords were recorded by a male Vancouver English speaker in quiet room. Critical words for the exposure phase were recorded in pairs, once normally and once with the sibilant swapped forming a nonword. The speaker was instructed to produce both forms with comparable speech rate, speech style and prosody.

Table 2: Frequencies of words used in categorization continua

Continuum	/s/-word frequency	/ʃ/-word frequency
sack-shack	1.11	0.75
sigh-shy	0.53	1.26
sin-shin	1.20	0.48
sock-shock	0.95	1.46

For each critical item, the word and nonword versions were morphed together in an 11-step continuum (0%-100% of the nonword /ʃ/ recording, in steps of 10%) using STRAIGHT (Kawahara et al., 2008) in Matlab (The Mathworks, Inc.). Prior to morphing, the word and nonword versions were time aligned based on acoustic landmarks, like stop bursts, onset of F2, nasalization or frication, etc. All control items and filler words were processed and resynthesized by STRAIGHT to ensure a consistent quality across stimulus items.

1.3 Pretest

To determine which step of each continua would be used in exposure, a phonetic categorization experiment was conducted. Participants were presented with each step of each exposure word-nonword continuum and each categorization minimal pair continuum, resulting in 495 trials (40 exposure words plus five minimal pairs by 11 steps). The experiment was implemented in E-prime (cite). As half of the critical items had a sibilant in the middle of the word (onset of the final syllable), participants were asked to respond with word or non word rather than asking for the identity of the ambiguous sound, as in previous research (Reinisch et al., 2013).

The proportion of s-responses (or word responses for exposure items) at each step of each continuum was calculated and the most ambiguous step chosen. The threshold for the ambiguous step for this experiment was when the percentage of s-response dropped near 50%. A full list of steps chosen for each stimulus item is in the appendix. For the minimal pairs, six steps surrounding the 50% cross over point were selected for use in the phonetic categorization task. Due to experimenter error, the continuum for *seedling* was not included in the stimuli, so the chosen step was the average chosen step for the /s/-initial words. The average step chosen for /s/-initial words was 6.8 ($SD = 0.5$), and for /s/-final words the average step was 7.7 ($SD = 0.8$).

1.4 Procedure

Participants in the experimental conditions completed two tasks, an exposure task and a categorization task. The exposure task was a lexical decision task, where participants heard auditory stimuli and were instructed to respond with either "word" if they thought what they heard was a word or "nonword" if they didn't think it was a word. The buttons corresponding to "word" and "nonword" were counterbalanced across participants. Trial

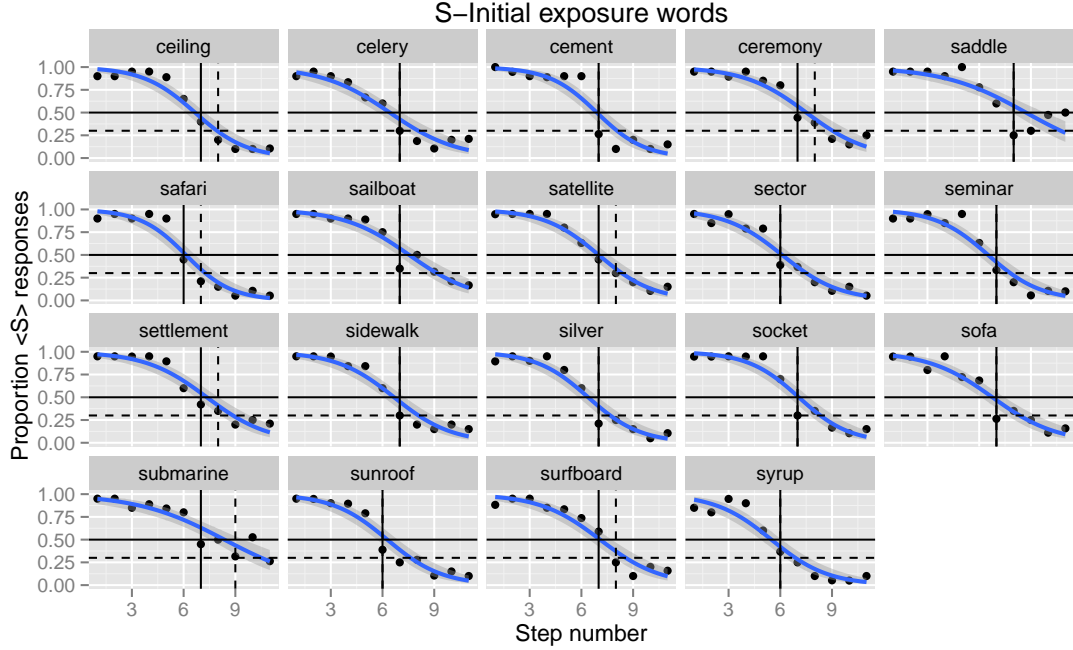


Figure 1: Proportion of word-responses for /s/-initial exposure words. Solid lines represent Experiment 1 selection criteria (50% word-response rate) and dashed lines represent Experiment 2 selection criteria (30% word-response rate). Dots are averaged word-response across subjects, and the blue line is a binomial model constructed from the responses.

order was pseudorandom, with no critical or control items appearing in the first six trials, and no critical or control trials in a row, but random otherwise, following Reinisch et al. (2013).

In the categorization task, participants heard an auditory stimulus and had to categorize it as one of two words, differing only in the onset sibilant (s vs sh). The buttons corresponding to the words were counterbalanced across participants. The six most ambiguous steps of the minimal pair continua were used with seven repetitions each, giving a total of 168 trials. Participants were instructed that there would be two tasks in the experiment, and both tasks were explained at the beginning to remove experimenter interaction between exposure and categorization.

Participants were assigned to one of four conditions. Two of the conditions exposed participants to only critical items that began with /s/, and the other two exposed them to only critical items that had an /s/ in the onset of the final syllable, giving a consistent 200 trials in all exposure phases with control and filler items shared across all participants. Additionally, participants in half the conditions received additional instructions that the speaker’s “s” sounds were sometimes ambiguous, and to listen carefully to ensure correct responses in the lexical decision.

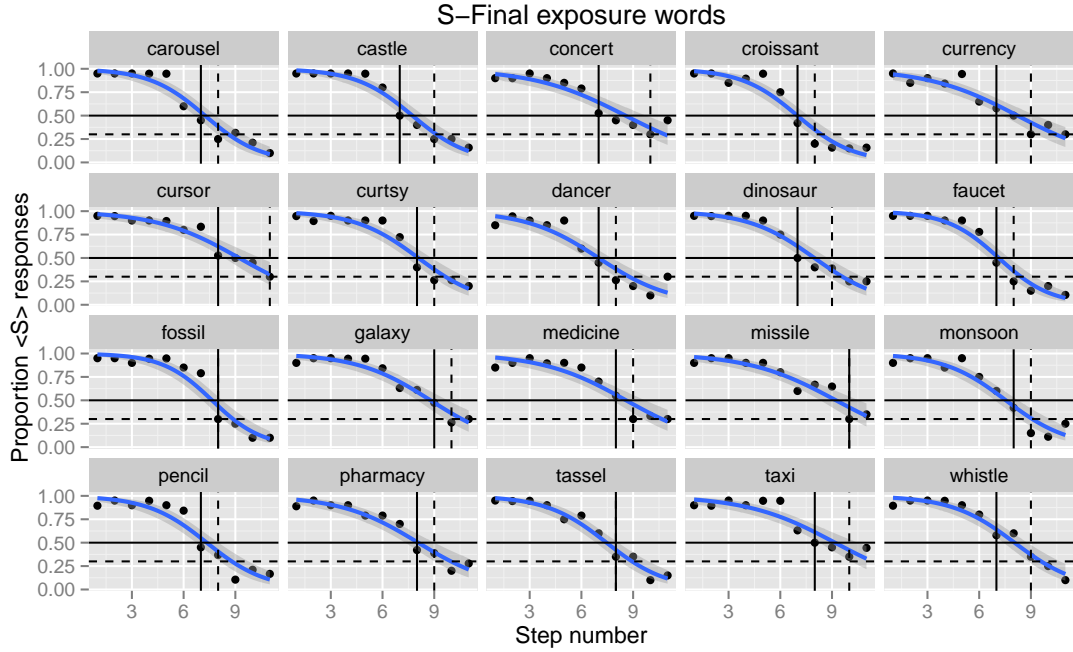


Figure 2: Proportion of word-responses for /s/-final exposure words. Solid lines represent Experiment 1 selection criteria (50% word-response rate) and dashed lines represent Experiment 2 selection criteria (30% word-response rate). Dots are averaged word-response across subjects, and the blue line is a binomial model constructed from the responses

References

- Brysbaert, M. and New, B. (2009). Moving beyond Kucera and Francis: a critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior research methods*, 41(4):977–990.
- Kawahara, H., Morise, M., Takahashi, T., Nisimura, R., Irino, T., and Banno, H. (2008). Tandem-straight: A temporally stable power spectral representation for periodic signals and applications to interference-free spectrum, F0, and aperiodicity estimation. In *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, pages 3933–3936.
- Reinisch, E., Weber, A., and Mitterer, H. (2013). Listeners retune phoneme categories across languages. *Journal of experimental psychology. Human perception and performance*, 39(1):75–86.