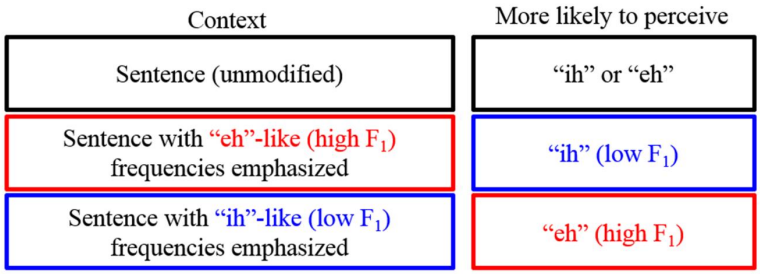
"Parameterizing effects of precursor spectral characteristics on categorization of subsequent vowel sounds" by Christian E. Stilp

INTRODUCTION

- When spectra differ between earlier (precursor) and later (target) sounds, listeners perceive larger spectral changes than are physically present. For example, when a precursor sentence possesses relatively higher frequencies, the target vowel is perceived as possessing relatively lower frequencies, and vice versa. These are known as **spectral contrast effects (SCEs)**.



- Precursor stimuli in these studies are often speech (e.g., a sentence). Typically, the researcher uses a filter to amplify frequencies in the precursor that are contrastive with frequencies in the target sound, producing the SCE.
- Spectral properties of speech change rapidly throughout its duration. While the researcher used a filter to add a peak to the precursor's long-term spectrum, s/he ceded control as to when that peak occurs and for how long. This obscures key characteristics of precisely how earlier sounds influence perception of later sounds.
- The present study systematically varied key spectral properties of precursor sounds to vary their influence (via SCEs) on vowel categorization.

METHODS

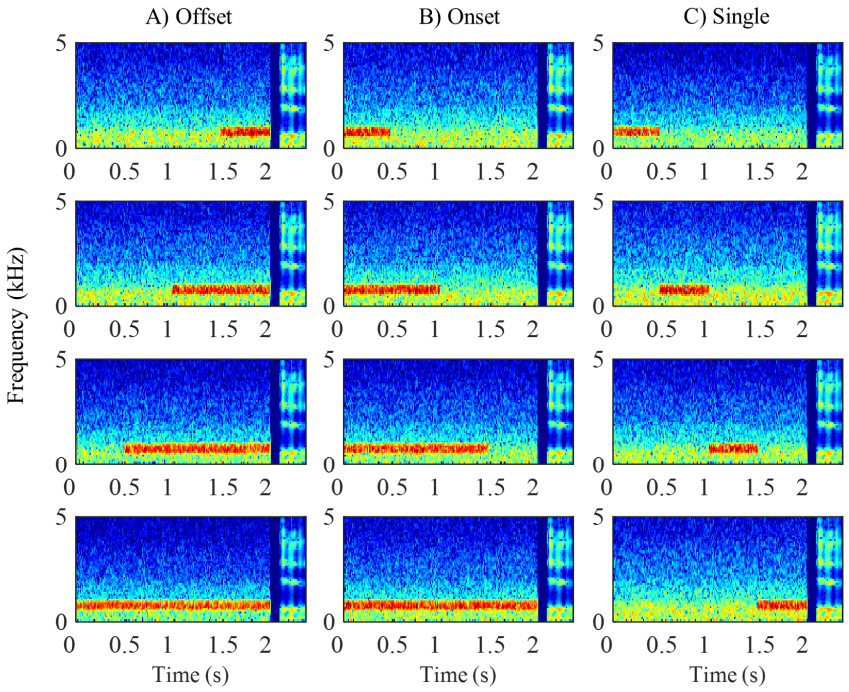
Participants

- 26 native English speakers were recruited; 17 completed the entire experiment.

Stimuli

- **Contexts** = 2 sec of speech shaped noise
- Part or all of context was processed by either low-F1 (100-400 Hz) or high-F1 (550-850 Hz) filters with +10 dB or +20 dB filter gain

- *Offset condition*: spectral peak occurred in last 500/1000/1500/entire 2000 ms of the context
- *Onset condition*: spectral peak occurred in first 500/1000/1500/entire 2000 ms of the context
- *Single condition*: spectral peak occurred in one 500-ms epoch only (0-500/500-1000/1000-1500/1500-2000 ms)



Sample trials showing the different experimental paradigms. In each, high-F1 frequencies (550-850 Hz) in the noise context are amplified by +20 dB.

- **Vowels** = same /ɑ̈-to-/ʊ̈/ continuum as previously tested by Stilp et al. (2015), separated from contexts by a 50-ms ISI.
- Target vowels and unfiltered noise were presented at 64 dB SPL. The level of filtered noise varied depending on how long the spectral peak was present in the context (66-81 dB SPL).

Procedure

Visit 1:

- Informed consent

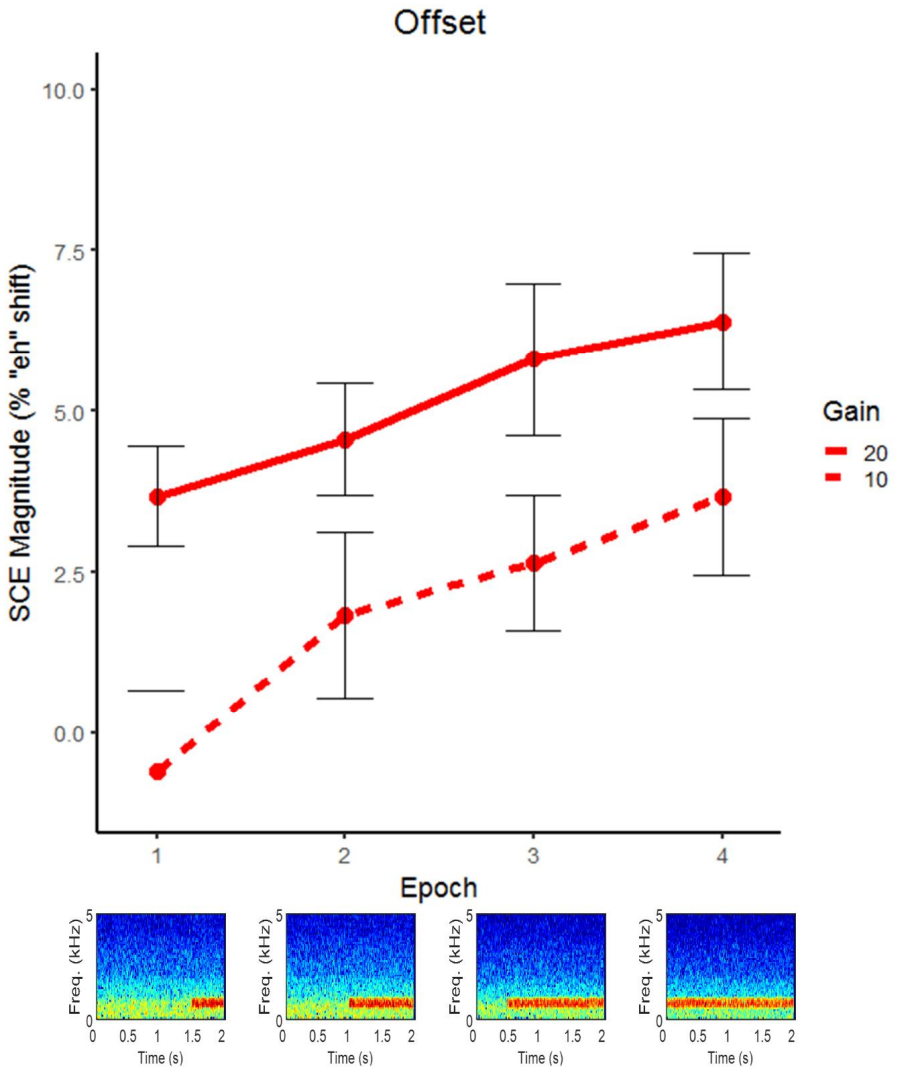
- Abridged LEAP-Q language questionnaire to confirm English was their native language (n=3 removed)
- Practice trials to ensure consistent identification of endpoints of the target vowel continuum (80% correct criterion; no participants removed)
- Hearing screen at octave frequencies from 125 to 8000 Hz to assure healthy hearing (≤ 20 dB HL; n=6 removed, leaving final sample of n=17)

Visits 2-7:

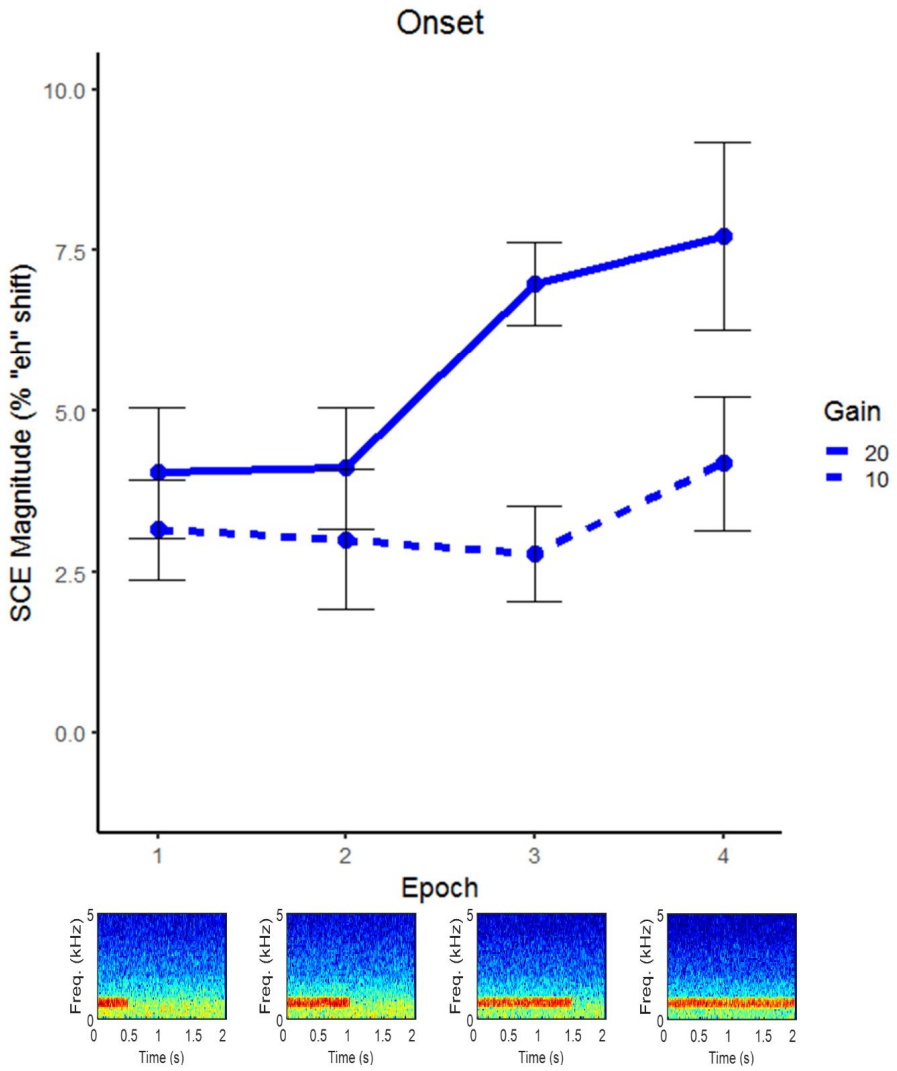
- One of six hourlong experimental blocks: three paradigms (Offset, Onset, Single) x two levels of filter gain (+10 dB, +20 dB) in counterbalanced orders.

RESULTS

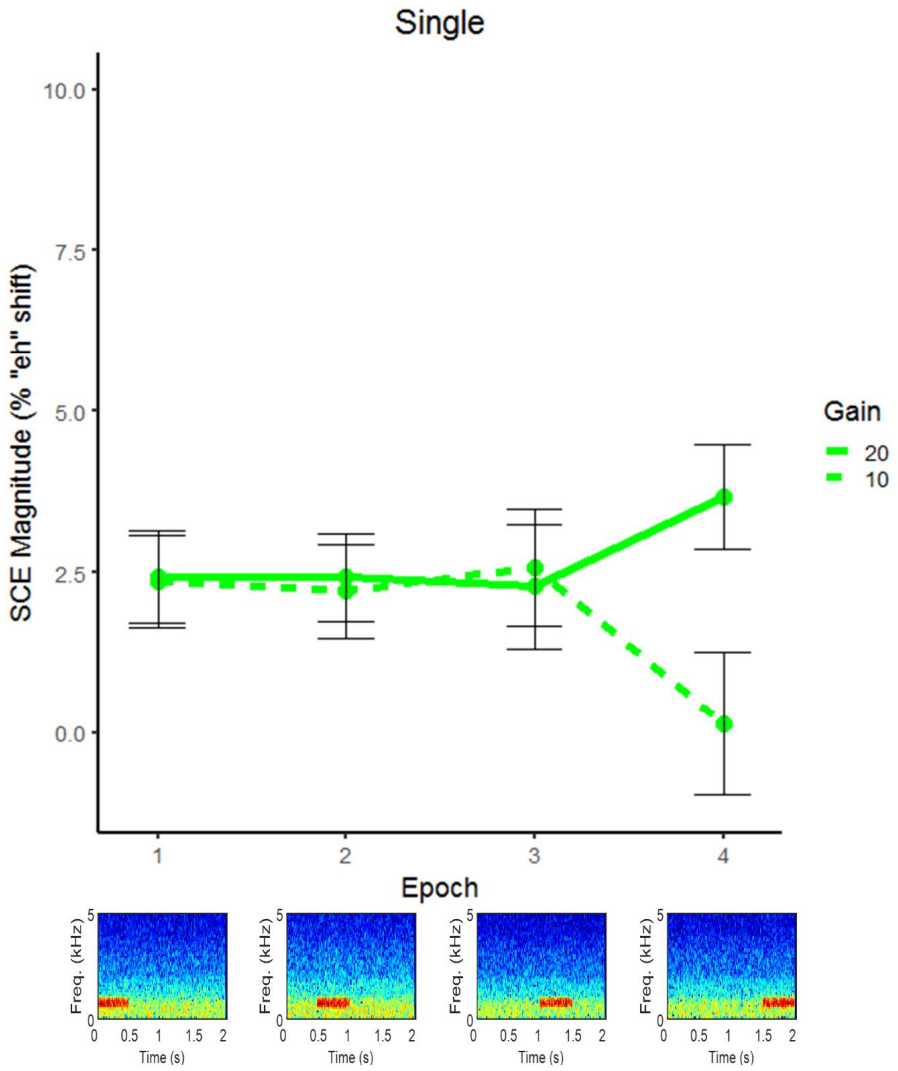
- SCEs were measured as the percent shift in ñehò (high-frequency) responses following lower-frequency-amplified (higher %) vs. higher-frequency-amplified (lower %) precursor sentences. SCEs were calculated for each listener in each block of each paradigm. Means and standard errors are plotted below.
- SCEs were calculated for each subject in each block (paradigm x epoch x filter gain), then were entered as the dependent measure in linear mixed-effects models for each paradigm.
    - Fixed effects: filter gain (mean-centered), epoch number (factor; epoch 1 used as reference level), and their interaction
    - Random effects: random slope for filter gain, random intercept for listeners



- SCEs were larger following +20 dB gain filtering than +10 dB gain filtering ($t = 2.55$, $p < .02$)
- SCEs were marginally larger in epoch 2 than epoch 1 ($t = 1.69$, $p = .10$), but were significantly larger in epoch 3 ($t = 2.73$, $p < .01$) and epoch 4 ($t = 3.56$, $p < .001$)
- No significant gain-by-epoch interactions



- SCEs occurred even when spectrally neutral (unfiltered) noise separated the spectral peak in the precursor from the target vowel (epochs 1-3)
- No significant influence of filter gain
- Relative to epoch 1, SCEs were only significantly larger in epoch 4 (t = 2.52, p < .02)
- The gain-by-epoch (epoch 1 vs. epoch 3) interaction approached significance (t = 1.78, p = .08); less so for epoch 1 vs. epoch 4 (t = 1.42, p = .16)



- SCEs occurred even when spectrally neutral (unfiltered) noise separated the spectral peak in the precursor from the target vowel (epochs 1-3)
- No significant influence of filter gain
- No significant influence of epoch
- The gain-by-epoch interaction was only significant when comparing epoch 1 to epoch 4 (t = 2.21, p = .03)

DISCUSSION

- Neural mechanisms related to adaptation are thought to underlie SCEs (Delgutte, 1996; Delgutte et al., 1996). Neurons adapt to frequency components that are present in earlier (precursor) sounds, particularly to large spectral peaks. These adapted neurons would

then be less responsive to these same frequency components in later (target) sounds. Conversely, neurons that encode other frequencies would be unadapted / less adapted, making them relatively more responsive to frequencies in later (target) sounds.

- F1 frequency regions were either amplified (addition of a spectral peak via filtering) or not (speech-shaped noise). In both cases, acoustic energy was sufficiently continuous throughout the precursor to unintentionally produce pedestals (Smith & Zwislocki, 1975; Smith, 1977). In those studies, anesthetized auditory nerve fibers in guinea pigs were responsive to increments in stimulus intensity even while adapted. Here, constant stimulation from the speech-shaped noise context might retain some sensitivity to an earlier spectral peak that is temporally nonadjacent to the target vowel, producing an SCE (see Onset and Single data for epochs 1-3).

- Filtered noise precursors produced smaller SCEs than other studies that used speech precursors (e.g., +20 dB peaks added to sentence contexts and these vowels produced SCE shifts of 14-17%; Assgari & Stilp, 2015). This may be due to upward spread of masking, as ever-present acoustic energy at frequencies just above the added spectral peaks could have masked them, reducing their contributions towards SCEs.

- Previously, larger filter gain (adding larger spectral peaks to the precursors) resulted in systematically larger SCEs (Stilp et al., 2015; Stilp & Assgari, 2017). Here, effects of filter gain differed by paradigm. Higher filter gain increased SCE magnitudes across the board in the Offset paradigm, only for epochs 3 and 4 in the Onset paradigm, and only for epoch 4 in the Single paradigm. Thus, temporal characteristics of spectral peaks in precursor sounds must also be considered when assessing their influence on subsequent speech categorization.

REFERENCES

Assgari, A. A., & Stilp, C. E. (2015). Talker information influences spectral contrast effects in speech categorization. The Journal of the Acoustical Society of America, 138(5), 3023-3032.

Delgutte, B. (1996). Auditory neural processing of speech. In W. J. Hardcastle & J. Laver (Eds.), The Handbook of Phonetic Sciences (pp. 507ï 538). Oxford: Blackwell Publishing Ltd.

Delgutte, B., Hammond, B.M.,Kalluri, S., Litvak, L.M.,&Cariani, P.A. (1996). Neural encoding of temporal envelope and temporal interactions in speech. In W. Ainsworth & S. Greenberg (Eds.), Proceedings of Auditory Basis of Speech Perception. European Speech Communication Association.

Marian, V., Blumenfeld, H. K., & Kaushanskaya, M. (2007). The Language Experience and Proficiency Questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals. Journal of Speech, Language, and Hearing Research., 50(4), 940-967.

Smith, R. L. (1977). Short-term adaptation in single auditory nerve fibers: some poststimulatory effects. Journal of Neurophysiology, 40(5), 1098-1111.

Smith, R. L., & Zwislocki, J. J. (1975). Short-term adaptation and incremental responses of single auditory-nerve fibers. Biological Cybernetics, 17(3), 169-182.

Stilp, C. E., Anderson, P. W., & Winn, M. B. (2015). Predicting contrast effects following reliable spectral properties in speech perception. The Journal of the Acoustical Society of America, 137(6), 3466-3476.

Stilp, C. E., & Assgari, A. A. (2017). Consonant categorization exhibits a graded influence of surrounding spectral context. The Journal of the Acoustical Society of America, 141(2), EL153-EL158.