INTRODUCTION

- Perception of sounds ï including phonemes ï is affected by the sounds that immediately precede them.

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

- SCEs occur on different timescales; the precursor can be temporally adjacent to the target (i.e., proximal) or nonadjacent (i.e., distal).

- Previous studies of SCEs for precursor and target sounds separated by a silent gap:
    - Holt and Lotto (2002): precursor syllable (/al/ or /ar/) preceded the target consonant (/ga/-/da/); SCEs were extinguished with 400-ms ISIs
    - Broadbent and Ladefoged (1960): precursor sentence preceded the target vowel (/ɑ̈-/ʊ̀/, as in ñbitò and ñbetò); some participants showed SCEs with 10-second intervals between precursor and target
    - These studies used very different stimuli, target frequency regions, and precursor-target silent intervals

- The present study systematically measured the effect of precursor-target interval on SCEs in vowel categorization and consonant categorization using a uniform range of ISI durations.

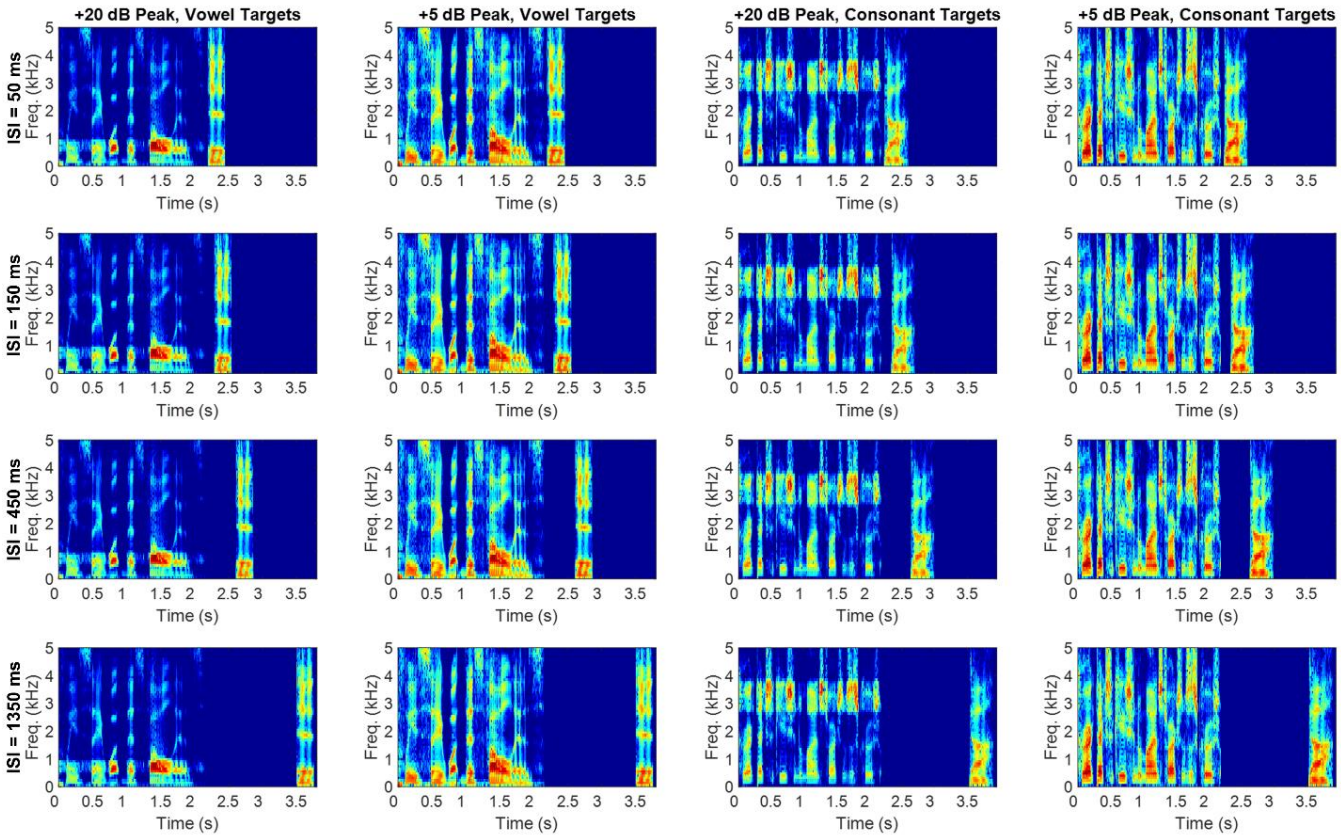
METHODS

- Four groups of 20 normal-hearing native English speaking listeners participated.

- Baseline screen: 20 trials to ensure consistent identification of the endpoints of 10-step vowel/consonant target series.

- Test: Identification of vowel/consonant continuum with four durations of silent separation between precursor and target stimuli (ISI = 50, 150, 450, 1350 ms) tested in counterbalanced orders.

- All stimuli were presented to one ear at 70 dB SPL



Sample trials. In all cases, the precursor sentence has been filtered to amplify the upper region of the frequency band needed to identify the target sound (left two columns: 550-850 Hz; right two columns: 2700-3700 Hz).

<u>Vowel categorization tasks</u>

- Precursor sentence and vowel targets (/ɑ̈-/ʊ̈) taken from Stilp, Anderson, & Winn (2015)
    - These vowels are contrasted by the frequency range 100 ï 850 Hz
- In each block, listeners heard precursors processed to amplify low-F1 (100-400 Hz) or high-F1 (550-850 Hz) regions.

- Each listener group heard precursors processed at one level of filter gain (+5 dB or +20 dB)

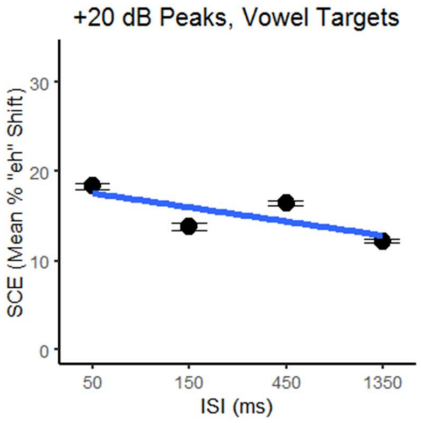
Consonant categorization tasks

- Precursor sentence and consonant targets (/gŏ/-/dŏ/) taken from Stilp (2020)
    - These consonants are contrasted by the frequency range between 1700 and 3700 Hz
- In each block, listeners heard precursors processed to amplify low-F3 (1700-2700 Hz) or high-F3 (2700-3700 Hz) regions
- Each listener group heard precursors processed at one level of filter gain (+5 dB or +20 dB)

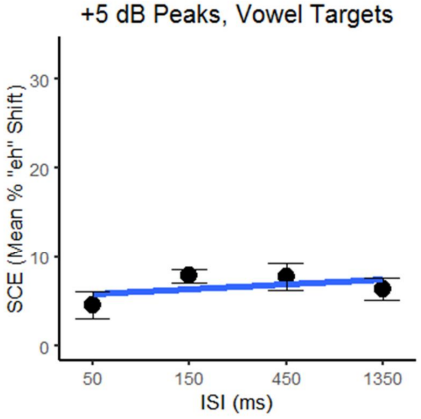
RESULTS

- Generalized linear mixed-effects models were used to identify the continuum step where perception crossed 50% between one phoneme and the other.
    - Fixed effects: Target continuum step (slope; mean-centered), Filter Frequency (categorical), ISI (mean-centered), and all interactions
    - Random effects: subject-nested random effects for Intercept, Target slope, Filter Frequency, and ISI.

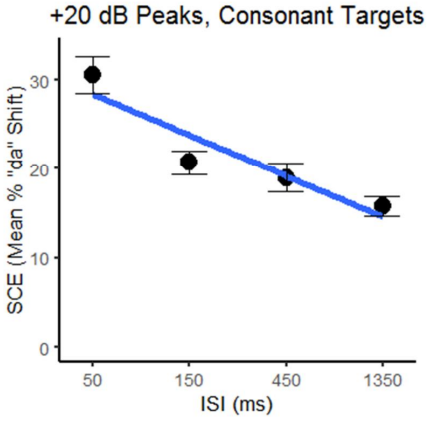
- All tasks produced SCEs (perceptual bias toward hearing phonemes with more high-frequency energy following precursor with lower frequencies amplified).
- SCEs are plotted below as the mean shift in listeners' higher-frequency responses ("eh", "da") due to precursor filtering. Error bars indicate standard error, and linear regressions are fit to each data set.



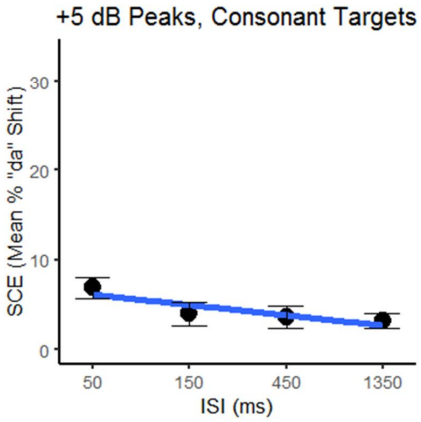
Mixed effects model: SCEs decrease with increasing ISI (Z = –2.106, p = .035)



Mixed effects model: SCEs do not seem to vary as a function of ISI (Z = 0.746, p = .456)



Mixed effects model: SCEs decrease with increasing ISI (Z = –5.091, p < .001)



Mixed effects model: SCEs do not seem to vary as a function of ISI (Z = 1.226, p = .220)

- With large spectral peaks (+20 dB filter gain): SCEs decreased as log ISI duration increased

- With small spectral peaks (+5 dB filter gain): SCEs did not systematically vary as a function of log ISI
    - In vowel categorization, an unexpected pattern of results was observed with unusually small SCEs at 50 ms ISIs. In consonant categorization, SCE magnitudes dropped initially (paired t-test on 50-ms ISIs and 150-ms ISIs; t_{19} = 2.11, p = .048), then plateaued.

DISCUSSION

- Significant relationships and large SCE magnitudes at 1350 ms ISI suggest that effects would continue to decrease at even longer ISIs, as reported by Broadbent and Ladefoged (1960). Spectral properties of precursor sentences used in their study were highly different from each other, closely resembling +20 dB filter gain conditions here.

- SCEs decayed more rapidly in results collected by Holt and Lotto (2002). This may be due to their use of shorter-duration precursors (a syllable rather than a sentence). Longer-duration contexts can produce larger SCEs (Holt, 2006), which the present data for +20 dB conditions show persist over longer timecourses.

- Results are consistent with the mechanisms thought to underlie SCEs: neural adaptation in both the peripheral and central auditory system (Stilp, 2020).
    - Larger spectral peaks in the precursor are expected to produce larger differences in neural activation profiles, and thus larger SCEs.
    - Sentence-length precursors are expected to produce sufficient neural adaptation to observe SCEs in both the periphery and the central auditory system.
    - Increasing ISI duration would progressively decrease the contributions from peripheral adaptation (where time constants of adaptation are generally shorter), thus decreasing SCE magnitudes.

- Existing models of neural responses to speech sounds largely consider these sounds in isolation, leaving realistic contextualized sound processing unexplained. Future research utilizing computational models of auditory nerve activity will articulate and clarify the

low-level processes related to neural adaptation proposed to underlie these context effects.

- Auditory enhancement effects (EEs) also decrease linearly with increasing log ISI (Wilson 1970; Viemeister 1980). The fact that SCEs (at +20 dB filter gain) pattern similarly is perhaps not surprising given that EEs and SCEs are closely related (Stilp, 2019; 2020).

REFERENCES

Broadbent, D. E., & Ladefoged, P. (1960). Vowel judgements and adaptation level. Proceedings of the Royal Society B: Biological Sciences, 151, 384ï 399.

Holt, L. L. (2006). The mean matters: Effects of statistically defined nonspeech spectral distributions on speech categorization. Journal of the Acoustical Society of America, 120(5), 2801ï 2817.

Holt, L. L., & Lotto, A. J. (2002). Behavioral examinations of the level of auditory processing of speech context effects. Hearing Research, 167(1ï 2), 156ï 169.

Stilp, C. E. (2019). Auditory enhancement and spectral contrast effects in speech perception. The Journal of the Acoustical Society of America, 146(2), 1503ï 1517.

Stilp, C. E. (2020). Evaluating peripheral versus central contributions to spectral context effects in speech perception. Hearing Research, 392, 1ï 12. https://doi.org/10.1016/j.heares.2020.107983

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), 3466ï 3476.

Viemeister, N. F. (1980). Adaptation of masking. In G. v. d. Brink & F. A. Bilsen (Eds.), Psychophysical, Physiological and Behavioural Studies in Hearing (pp. 190ï 198). Delft: University Press.

Wilson, J. P. (1970). An auditory afterimage. In R. Plomp & G. F. Smoorenberg (Eds.), Frequency Analysis and Psychophysics of Hearing (pp. 303-315). Sijthoff, Leiden, Netherlands.