Spectro-temporal weighting of interaural time differences in speech

Boston University College of Health
& Rehabilitation Sciences: Sargent College
Department of Speech, Language & Hearing Sciences

Lucas S. Baltzell, Jayaganesh Swaminathan, and Virginia Best
Department of Speech, Language & Hearing Sciences
Boston University and University of the Pacific

BACKGROUND

- Numerous studies have demonstrated that the perceptual weighting of interaural time differences is non-uniform in time and frequency.
  - ITDs around 600 Hz dominate the spatial percept of broadband sounds. The "spectral dominance" region appears to follow from ITD sensitivity.
  - ITDs at sound onset dominate the spatial percept of continuous sounds ("onset dominance").
  - Spectro-temporal interactions can occur for stimuli with competing ITD cues.
- To our knowledge, no previous attempts have been made to characterize spectro-temporal dominance regions for ITDs in speech, which was the goal of this study. **Specifically, we wanted to determine if the perceptual weighting of ITDs (1) followed straightforwardly from ITD sensitivity of individual spectro-temporal regions (Figure 1), and/or (2) favored ITDs at syllable onset.**

METHODS

- 2 speech tokens: "two" and "eight"
- Filtered into 8 frequency bins and 2 time bins
- ITD sensitivity task
  - ITD thresholds measured for each time-frequency bin presented in isolation
  - Adaptive 2AFC left-right discrimination procedure
- Lateralization weights task
  - ITDs for each time-frequency bin drawn independently from a normal distribution with a mean of 0 and a standard deviation of 200 µs.
  - Listeners indicated left or right (500 trials)
- 11 normal hearing listeners

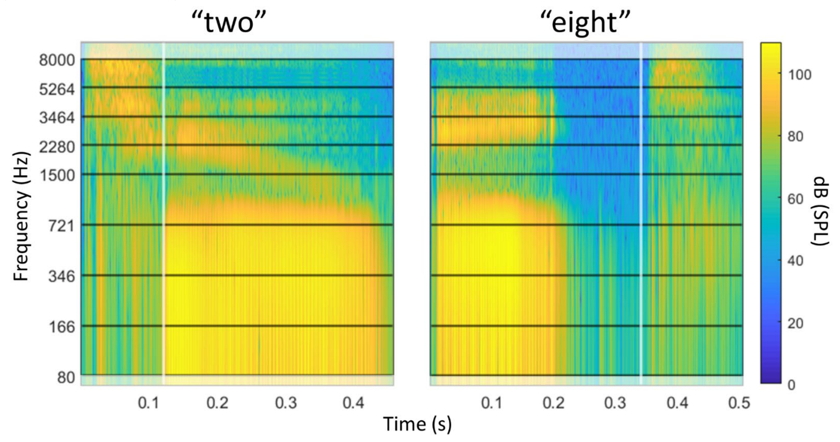

Figure 1: Tokens were filtered into 8 frequency bins and 2 time bins. Grids overlaid on the spectrogram represent the boundaries between these bins. The overall intensity for each token was 70 dB SPL.

LATERALIZATION WEIGHTS

- Lateralization weights were obtained for each listener by modelling their binary lateralization judgments (0 or 1) with a logistic regression using the ITDs in each time-frequency bin as predictors:

$$y_i = logit^{-1}(\beta_0 + \beta_j x_{i,j} + \varepsilon_i),$$

where y_i is lateralization judgement on the i^{th} trial (i = 1:500) and $x_{i,j}$ is the vector of $\boldsymbol{j}$ ITDs across time-frequency bins (j = 1:16) on the i^{th} trial.

- Once raw β-weights were obtained for the ITD predictors, individual β-weights were *normalized* by the sum over all β-weights:

$$\hat{\beta}_j = \frac{\beta_j}{K}, \quad \text{where } K = \sum_{j=1}^{16} \beta_j, \quad \text{such that } \sum_{j=1}^{16} \hat{\beta}_j = 1$$

- Intensity-scaled ITD sensitivity was computed for each listener:

$$\xi_j = norm(\text{ITD sensitivity in a.u.})_j * norm(\text{stimulus intensity in dB SL})_j$$

and was used to evaluate whether weight patterns could be accounted for by ITD sensitivity

CLASSIFICATION ACCURACY

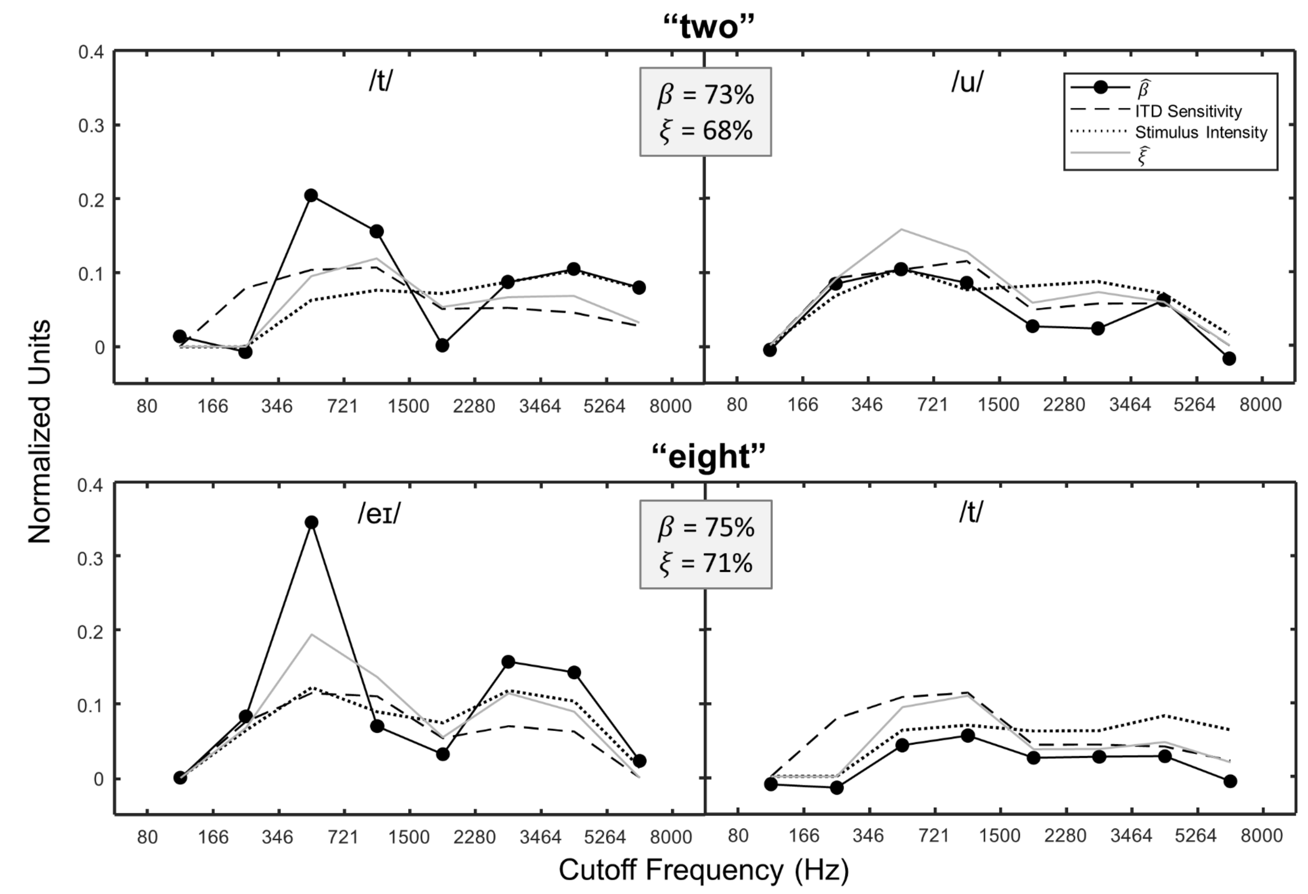

Figure 2: Classification accuracy for the regression model was assessed for each listener using a method that is immune to constant rescaling. To determine the classification accuracy of the normalized weights $\hat{\beta}$, we found $\hat{y} = logit^{-1}(\hat{\beta}_0 +$

HIERARCHICAL BAYESIAN ANALYSIS

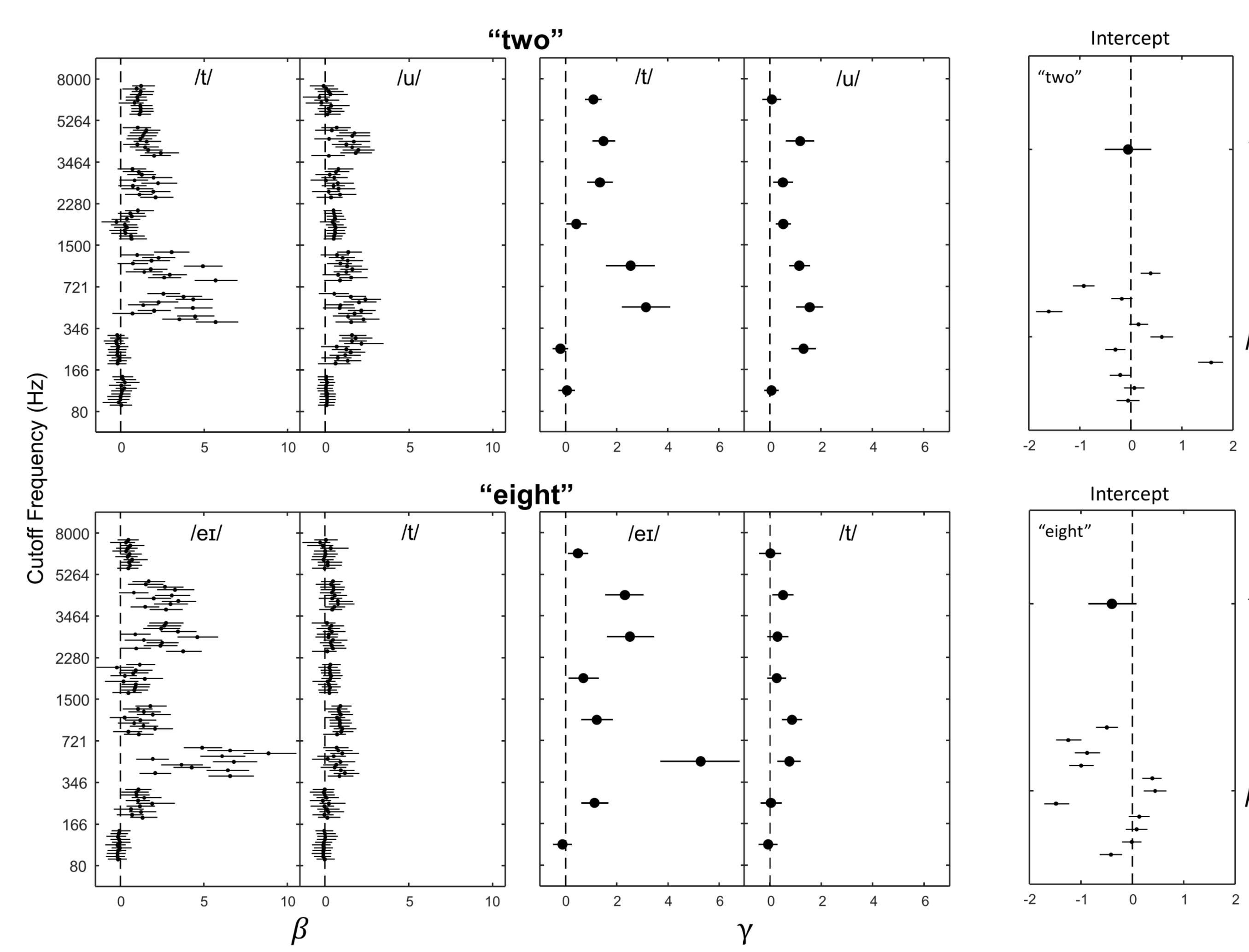

Figure 3: Statistical significance was assessed with a Bayesian hierarchical model, and due to the inherent differences across speech tokens, a separate model was fit for each speech token. For each time-frequency bin, a posterior distribution of β-weight values were obtained for each listener, as well as a posterior distribution for the group-level mean γ. The median of these distributions are shown here, along with 97.5% credible intervals. The intercept terms reflect response bias for lateralization judgments across all time-frequency bins. As in the regression model, a $logit^{-1}$ linking function was used to relate the weighted predictors $\beta_{j,l}X_{j,l}$ to the lateralization judgments Y_l (l = 1:11).

CONCLUSIONS

- **(1) ITDs carried by phonemes in the first position of the syllable contribute more strongly to lateralization judgments than ITDs carried by phonemes in the second position.**
  - The strong effect of syllable position in our data (Figure 3) is consistent with a central mechanism, preferentially weighting ITD cues at the global onset of the syllable rather than at local acoustic onsets within frequency channels (Figure 4).
  - It is unclear whether this effect represents a central decision process or a central adaptation process
- **(2) Lateralization judgments can be reasonably well accounted for by a combination of ITD sensitivity and stimulus intensity across time-frequency bins.**
  - While we found that β-weights were significantly better than ξ-weights at classifying lateralization judgements, the difference was not as dramatic as might be expected based on visual inspection of the weighting functions (Figure 2)
  - Since ITD sensitivity was measured in isolation, any effect of syllable position would not be accounted for by ξ-weights. We therefore suggest that the difference in classification accuracy is driven primarily by the effect of syllable position.
- **(3) ITDs at low frequencies (vowels) appear to contribute more to lateralization judgments than ITDs at high frequencies (unvoiced stop consonants)**
  - This result is subtle and could not be quantified statistically (different models).

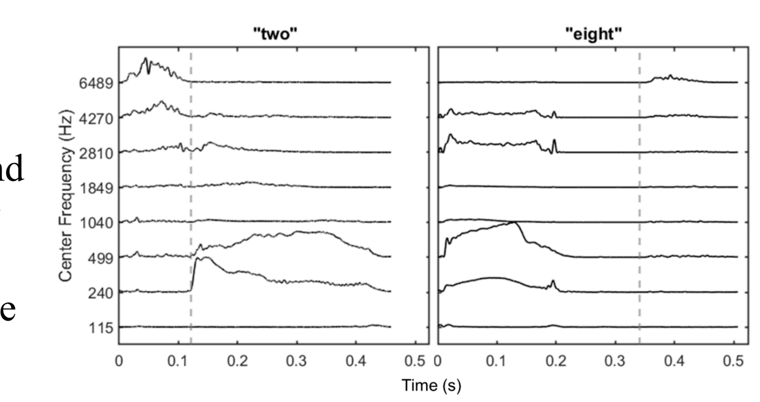

Figure 4: Stimulus envelopes for each time-frequency bin. Envelopes were computed by taking the magnitude of the analytic signal, and low pass filtering the resulting signal (150-Hz cutoff). Frequency bins are indicated by their center frequency, and boundaries between time bins are indicated with dashed gray lines.

CHALLENGES AND DIRECTIONS

- While we are interested in deriving general conclusions about the perceptual weighting of ITDs for classes of speech sounds, the significant natural variability both within and across these classes presents a challenge. More definitive conclusions will likely require a large sample of stimuli and a variety of complementary approaches.
- Different weight patterns may be observed under different listening conditions (background noise, reverberation, etc.) and for listeners with hearing impairment (or cochlear implant users).
- Linked bilateral hearing aids offer the possibility of preserving or enhancing binaural cues. Binaural signal processing strategies are often faced with a tradeoff between providing monaural and binaural benefit, and can theoretically be improved with a better understanding of which spectro-temporal regions of speech contain the most perceptually relevant binaural cues