



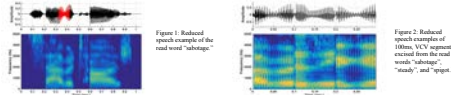
Acoustic Modeling of the Perception of Place Information in Incomplete Stops

Megan Willi & Brad Story
Speech, Language, and Hearing Sciences, The University of Arizona

Presented at the 169th Meeting of the
Acoustical Society of America
19 May 2015
Session 2pSC
Pittsburgh, PA

Introduction

- Previous research on stop consonant production found that *less than 60%* of the stops sampled from a connected speech corpus contained a clearly defined hold duration followed by a plosive release [Crystal & House, JASA, 1988]. How listeners perceive the remaining portion of incomplete stop consonants is not well understood.
- **Prior pilot research** demonstrated that participants could identify place information (i.e. /b-d-g/) in reduced, 100 ms vowel-consonant-vowel (VCV) segments excised from the Arizona English Recording Corpus.
- The **purpose of the current study** is to investigate whether relative formant deflection patterns, a potential model of acoustic invariance proposed by Story and Buntun (2010), are capable of predicting listeners' perceptions of place information in acoustically continuous, voiced stop consonants.
- Listeners identified speech stimuli simulated using a **computational model of speech production** and model parameters based on x-ray microbeam articulatory data from VCV utterances [Story, JASA, 2009].



Aims

- Evaluate participants' perceptions of place-of-articulation information in stimuli that simulate:
- 1) incomplete closure in reduced voiced stop consonants.
 - 2) proportionally reduced consonant and vowel timing functions in stimuli with incomplete stop consonant closure.

Method

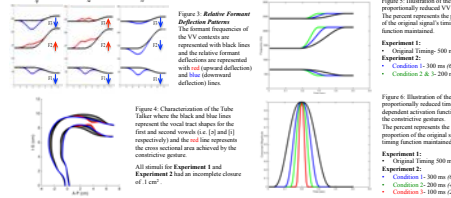
Participants: 10 native English speakers (Exp. 1) and 5 native English speakers (Exp. 2)

Task: Forced Choice Test (i.e. /b-d-g/)

Materials: All stimuli were 500 ms, vowel-consonant-vowel (VCV) utterances simulated using a voice source model based on the kinematic representation of the medial surfaces of the vocal folds and an airway modulation model of the vocal tract (aka "Tube Talker"). VCV continua were created for 3 underlying vowel-to-vowel transition (VV) contexts (i.e. [ai], [aɪ], and [au]) by incrementally moving the constriction location from the lips toward the velar part of the vocal tract in 20 (Exp. 1) and 15 (Exp. 2) discrete 0.4-cm steps. Experiment specific manipulations are described below.

Design: Stimuli were randomly presented 5 times (Exp. 1) and 3 times (Exp. 2) in a block design where only one vowel context was presented per block.

Analysis: Participants ID curve boundaries were compared to the perceptual boundaries predicted by the contour plots of the relative formant deflection patterns.



Example Stimuli

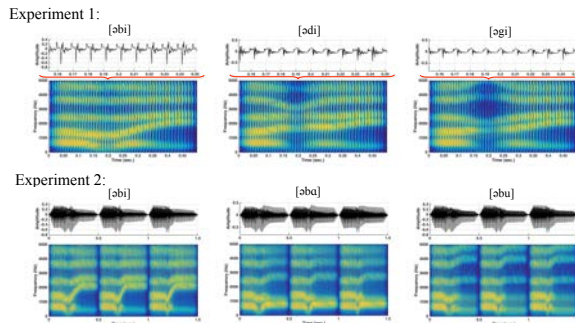


Figure 7: (Experiment 1) Example stimuli from one VV context (i.e. [ai]) at three different vocal tract locations (i.e. 17.5 cm, 13.9 cm, 11.9 cm respectively for /b-d-g/). (Experiment 2) Example stimuli at vocal tract location 17.5 cm (i.e. /b/) for conditions 1, 2, and 3 for each vowel context (i.e. [ai], [aɪ], [au] respectively).

Results: Experiment 1

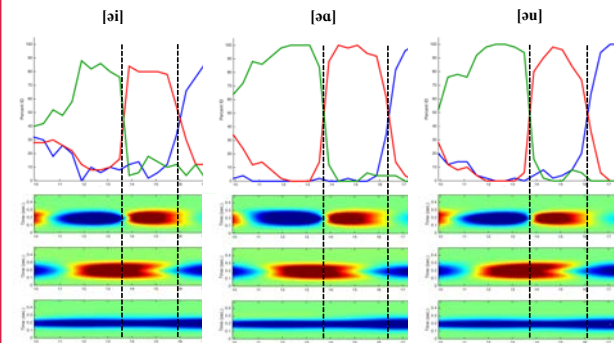


Figure 8: (Top) Identity curves averaged across all participants for the Forced Choice Test: /b/-(blue), /d/-(red), and /g/-(green). (Bottom) Contour plots depicting the relative formant deflection directions: upwards (red) and downwards (blue). The three panels correspond to F1 (lower panel), F2 (middle panel), and F3 (upper panel). The black lines indicate participants' phonetic boundaries defined as a 50% crossover point on the ID curve.

Results: Experiment 2

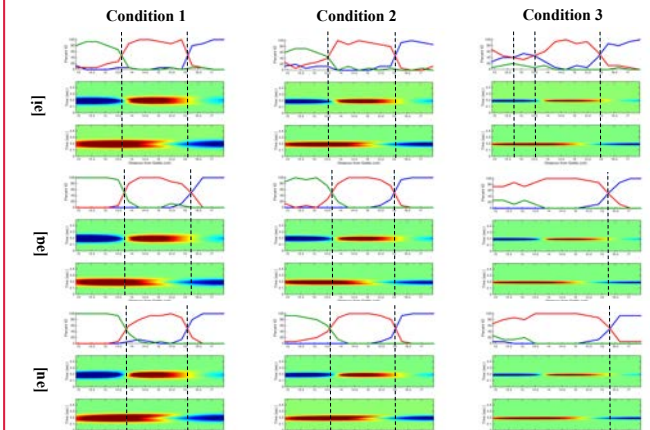


Figure 9: Average participant ID curves and contour plots (F2 lower panel and F3 upper panel) for three vowel contexts (i.e. [ai], [aɪ], [au]) from top to bottom respectively for Condition 1, 2, and 3 (left to right respectively). See Figure 8 for a detailed description.

Discussion

- Listeners' phonetic boundaries in Experiment 1 and Experiment 2 indicate that place-of-articulation information is present in incomplete, voiced stop consonant VCV stimuli lacking canonical hold duration and burst cues.
- Listeners' phonetic boundaries coincide with the proposed relative formant deflection patterns for all three places-of-articulation (i.e. /b-d-g/) across vowel contexts (i.e. [ai], [aɪ], [au]) and timing function manipulations (i.e. Conditions 1, 2, 3) except for the velar position in Condition 3.
- The results suggest that listeners may be sensitive to changes along this relative acoustic dimension and that relative formant deflection patterns could potentially explain the perception of place-of-articulation information in natural, reduced speech contexts. However, further investigation of the perceptual limits of this cue with respect to place is necessary.

References

Crystal, T. H., & House, A. S. (1988). Segmental durations in connected-speech signals: Current results. *The Journal of the Acoustical Society of America*, 83(4), 1553-1573.

Story, B. H. (2009). Vowel and consonant contributions to vocal tract shape. *The Journal of the Acoustical Society of America*, 126, 825-836.

Story, B. H., & Buntun, K. (2010). Relation of vocal tract shape, formant transitions, and stop consonant identification. *Journal of Speech, Language, and Hearing Research*, 53(6), 1514-1528.

Warner, N., & Tucker, B. V. (2011). Phonetic variability of stops and flaps in spontaneous and careful speech. *The Journal of the Acoustical Society of America*, 130(3), 1606-1617.

Acknowledgements

This research was supported by the Grunewald Foundation Fellowship and NIH R01-DC011275.