An analysis of the articulatory factors contributing to perceptual asymmetry

Ian Calloway (University of Michigan)

Background and Goals: In perceptual asymmetries involving consonants (as observed in the laboratory), two consonants are highly confusable with each other, but confusions exhibit a strong bias toward one member of the pair. Although members of each consonant pair differ in active articulator and place of articulation, they tend to show high confusability in a specific vocalic context. This study utilizes a corpus of real-time MRI data (Sorensen et al. 2017) to assess the extent to which confusability can be predicted from articulatory information. The targeted asymmetries, in Table 1, are three consonant pairs whose confusability depends on a specific context and one that appears to be context-independent. By relating the spatial dynamics of stop and fricative productions to their spectral acoustic properties (acknowledging that other acoustic parameters may also drive confusability) the study's goal is to offer insight into why some, but not other, confusion patterns show sensitivity to phonetic context.

Confusions between…	**favor…**	**in the context of…**
/k/ and /t/	/t/	High front vowels (e.g., Guion 1998)
/p/ and /t/	/t/	High front vowels (Winitz et al. 1972)
/k/ and /p/	/p/	High vowels (Winitz et al. 1972, Plauché 2001)
/θ/ and /f/	/f/	*No context attributed* (e.g., Miller & Nicely 1955)

Table 1 – Examples of perceptual asymmetry explored in this study

Method: The corpus consists of 17 American English speakers producing words, statements, and nonsense syllables. The productions used were V_1CV_1 disyllables – with C= /p/, /t/, /k/, /f/, or /θ/, and flanking V = /ɑ/, /i/, or /u/. For each production, the first frame after stop release (= up to 30ms after constriction release) or the last frame before fricative constriction release was manually extracted for analysis. Each speaker had two useable repetitions of each consonant in each vocalic context, for 510 tokens. From each video frame, a 30-point cross-sectional vocal tract (VT) area function was automatically generated by adapting a process described in Takemoto et al. 2006.

Consistent with the tube-model of the vocal tract, the spectral acoustics of frication and stop bursts show sensitivity to the shape of vocal tract anterior to the consonantal constriction. Accordingly, for each speaker, consonant pair, and vocalic context, the average difference in the anterior cavity lengths of the two consonants was computed. The spectral acoustics of aspiration, however, show sensitivity to the vocal tract cavities anterior and posterior to the constriction. For each area function, both lengths were computed, defining a 2D value. For each unique condition the average Euclidean distance between the consonant values was generated.

Results: Omnibus ANOVAs were computed for each consonant pair with difference and distance measures as DVs and vocalic context as the IV. For all pairs but /p/-/t/, the F-tests were significant for both DVs, so post-hoc pairwise comparisons by vocalic context were computed for each pair with Bonferroni correction. Table 2 gives the predictions for each pair (based on the patterns in Table 1) and the results of the comparisons.

Consonant Pair	Environment predicted to show smallest difference	Diff. in Anterior Cavity Length (for Burst and Frication)	Euclidean Distance Measure (for Aspiration)

/k/ & /t/	/i/	/i/ < /u/ < /ɑ/	/i/ < /u/ < /ɑ/
/p/ & /t/	/i/	No significant difference	No significant difference
/k/ & /p/	/i,u/	/i,u/ < /ɑ/	/i,u/ < /ɑ/
/θ/ & /f/	*No context attributed*	/u,ɑ/ < /i/	/u,ɑ/ < /i/

Table 2 – Predictions and results for each consonant pair

As predicted, for /k/-/p/ and /k/-/t/, the spatial measures corresponding to burst and aspiration acoustics were most similar in the vocalic contexts associated with higher confusability. These results offer the clearest support for the role of coarticulatory overlap in the spectral similarity of these consonant pairs. In /k/ and /t/, for example, the palatal constriction from /i/ reduces the size of the anterior cavity of /k/ closer to that of /t/, facilitating similarity in burst and aspiration spectral characteristics. The spatial measures of /p/ and /t/ did not differ significantly with respect to vocalic context. Though contrary to prediction, these results suggest that coarticulatory effects depend on relative positioning of the two interacting components. The anterior cavity lengths of /p/ and /t/, which are anterior to the vocalic constrictions explored in this study, do not differ with respect to vocalic context, and so the spatial measures addressed in this study would likely not differ either. These results also appear to suggest that the perceptual similarity of /p/ and /t/ may not be driven by their consonantal spectral acoustics.

Prior literature did not indicate a vocalic context favoring confusions between /θ/ and /f/, but /i/ shows the greatest spatial difference in these data. In ongoing research, random forest classifiers trained on the long-term average spectra of /f/ and /θ/ achieve a higher accuracy when trained on tokens of /f/ and /θ/ preceding high front vowels (/ɪ,i/) than when trained on tokens preceding low (/ɑ/) or high back vowels (/u,ʊ/). These articulatory and acoustic results offer a prediction to be tested perceptually.

Discussion: This project makes two contributions. It outlines the contribution of articulatory factors to the confusability of consonant pairs that show perceptual asymmetry. It also lays out a method to generate testable predictions of phonetic environments that facilitate confusability.

References

[1] Guion, S. G. (1998). The role of perception in the sound change of velar palatalization. *Phonetica, 55*, 18-52.

[2] Miller, G. A. & Nicely, P. E. (1955). An analysis of perceptual confusions among some English consonants. *J. Acoust. Soc. Am. 27*, 338-352.

[3] Plauche, Madelaine Claire. (2001). *Acoustic cues in the directionality of stop consonant confusions* (Dissertation/Thesis). ProQuest Dissertations Publishing.

[4] Sorensen, T., Skordilis, Z. I. et al. (2017). Database of Volumetric and Real-Time Vocal Tract MRI for Speech Science. *INTERSPEECH* pp. 645-649.

[5] Takemoto, H., Honda, K., et al. (2006). Measurement of temporal changes in vocal tract area function from 3D cine-MRI data. *J. Acoust. Soc. Am. 119*, 1037-1049.

[6] Winitz, H., Scheib, M. E., & Reeds, J. A. 1972. Identification of stops and vowels for the burst portion of/p, t, k/isolated from conversational speech. *J. Acoust. Soc. Am. 51*, 1309-1317.