# CONTEXTUALLY DEPENDENT CUE WEIGHTING FOR A LARYNGEAL CONTRAST IN SHANGHAI WU

Jie Zhang and Hanbo Yan

The University of Kansas
zhang@ku.edu, yanhanbo@ku.edu

## ABSTRACT

Phonological categories are often differentiated by multiple phonetic cues. This paper reports a production and perception study of a laryngeal contrast in Shanghai Wu that is not only cued in multiple dimensions, but also cued differently on different manners (stops, fricatives, sonorants) and in different positions (non-sandhi, sandhi). Acoustic results showed that although this contrast has been described as phonatory in earlier literature, it is primarily a tone contrast. Phonation correlates only appear in fricatives, and tone sandhi neutralizes the f0 difference. Our perception results were largely consistent with the aggregate acoustic results, indicating that speakers adjust the perceptual weights of individual cues for a contrast according to contexts. These findings support the position that phonological contrasts are formed by the integration of multiple cues in a language-specific, context-specific fashion and should be represented as such.

**Keywords**: Shanghai Wu, tone, tone sandhi, phonation, cue weighting.

## 1. INTRODUCTION

Phonological categories are often differentiated by multiple phonetic cues (e.g., see [16] and [20] for English voicing). A large body of work has been devoted to the understanding of how different acoustic cues are weighted in the perception of contrasts and how the weighting is affected by the acoustic dimensions along which the cues vary, the distributional characteristics of the acoustic cues, and the speakers' language background [11, 13, 14, 18, 22, 25]. In addition, the acoustic cues for the same contrast often depend on the phonological context in which the contrast appears. This paper presents a comprehensive case study on the contextual dependency of cue realization and cue weighting by examining the acoustics and perception of a laryngeal contrast in different contexts in Shanghai Wu.

Shanghai has a three-way laryngeal contrast among voiceless aspirated, voiceless unaspirated, and voiced stops. The voiced series, however, is known as "voiceless with voiced aspiration" [9, 30]. On fricatives, there is a two-way voicing contrast, whereby the voiced ones are truly voiced, and on sonorants, there is a modal-murmured distinction that corresponds to the voiceless-voiced distinction in obstruents [28, 30].

Shanghai Wu is also tonal. There are three phonetic tones on open or sonorant-closed syllables — 53, 34, and 13 (5 = high, 1 = low), and two phonetic tones on ʔ-closed syllables — 55 and 12. But there is a co-occurrence restriction between tones and onsets in that the higher tones 53, 34, and 55 only occur with voiceless/modal onsets and the lower tones 13 and 12 only occur with voiced/murmured onsets [28, 30].

Further complicating the issue is tone sandhi in Shanghai, which extends the tone on the first syllable of a polysyllabic compound over the entire compound domain [28, 30]. E.g., *tɔ*$^{34}$ and *dɔ*$^{13}$, when appearing in non-initial position of a compound, are reported to lose their tonal difference due to tone sandhi: /pɔ$^{34}$-tɔ$^{34}$/→[pɔ$^{33}$-tɔ$^{44}$] 'to check in'; /pɔ$^{34}$-dɔ$^{13}$/→[pɔ$^{33}$-dɔ$^{44}$] 'to report news'.

The data pattern in Shanghai, therefore, allows us to investigate both the weighting of multiple phonetic cues and how these cues change according to contexts. A production experiment and a perception experiment were designed to address what the acoustic cues are for the laryngeal contrast, how they vary according to the manner and position of the consonant, and how Shanghai speakers weight these cues in their perception.

## 2. PRODUCTION EXPERIMENT

### 2.1. Methods

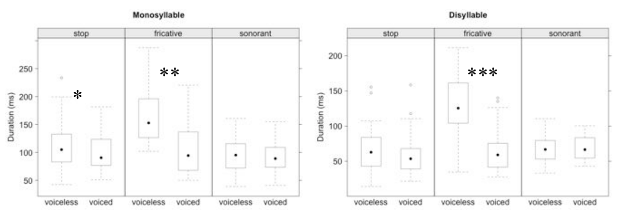
For our production study, 13 monosyllabic minimal pairs with rising tones (6 for stops, 4 for fricatives, 3 for sonorants) were used for the non-sandhi context (e.g., *pu*$^{34}$ and *bu*$^{13}$). The same pairs were then used as σ2 of disyllabic compounds with matched σ1 for the sandhi context (e.g., *fə̃*$^{51}$*-pu*$^{34}$ and *fə̃*$^{51}$*-bu*$^{13}$). All test words were embedded in a carrier sentence. Ten native speakers (5F) with a mean age of 25 were recorded in Shanghai, each reading the stimuli twice. Consonant durations were measured in Praat [6] and analyzed with Linear Mixed-Effects models [4, 19]. Spectral properties (H1-H2, H1-A1, H1-A2, H1-A3, CPP) of the following vowels were measured in *VoiceSauce* [24], and f0 on the vowels was measured using *ProsodyPro* in Praat [29]. Growth Curve Analyses [17] were conducted on the phonation and f0 curves using cubic orthogonal polynomials. The models were built up from the base model that only included subject, item, and subject-by-condition (laryngeal feature) random effects. Condition and its interaction with the time terms were then added step-wise. Different manners and positions were analyzed separately, and the voiceless/modal category was always the baseline.

### 2.2. Results

#### *2.2.1. Consonant duration*

The consonant duration results are given in Figure 1. The effect of voicing is significant for stops (*Estimate*=-10.539, *SE*=4.473, *t*=-2.356, *p*=.018) and fricatives (*Estimate*=-58.713, *SE*=21.523, *t*=-2.728, *p*=0.006) in monosyllables and for fricative in σ2 of disyllables (*Estimate*=-66.048, *SE*=15.987, *t*=-4.131, *p*<.001). No other laryngeal comparisons reached significance. For stops in σ2 of disyllables, however, the entire closure duration is voiced.

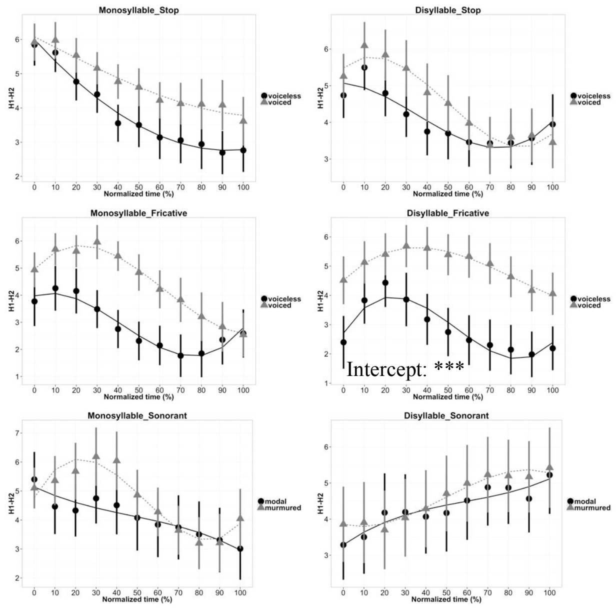
**Figure 1**: Duration of onset consonants in monosyllables (left) and σ2 of disyllables (right). *: p<.05; **: p<.01; ***: p<.001.




#### *2.2.2. Vowel phonation*

The H1-H2 results are given in Figure 2. For monosyllables, the model did not improve with the addition of the laryngeal feature or its interactions with the time terms for any manner (all *p*>.15). For σ2 of disyllables, stops and sonorants again did not exhibit any difference based on their laryngeal features (all *p*>.18). For fricatives, however, the effect of the laryngeal feature on the intercept was significant ($\chi^2(1)$=9.564, *p*=.002), and parameter estimates (*Estimate*=2.241, *SE*=0.568, *t*=3.942, *p*<.001) indicated that voiceless fricatives induced a lower H1-H2 than voiced fricatives.

**Figure 2**: H1-H2 results for monosyllables (left) and σ2 of disyllables (right). Symbols represent observed data with ±SE and lines represent growth curve models using cubic orthogonal polynomials.



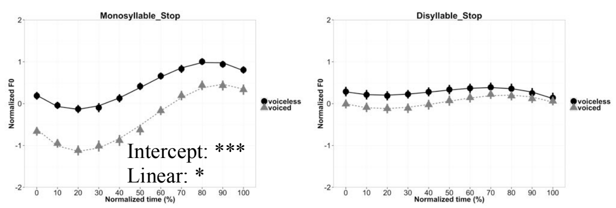

Results for other phonatory measures suggest similar generalizations and are not included here due to space limitation. But in the CPP results, voiceless fricatives in non-sandhi context exhibited a higher and sharper peak than voiced fricatives, indicating that there was a phonatory difference on the following vowel based on fricative voicing.

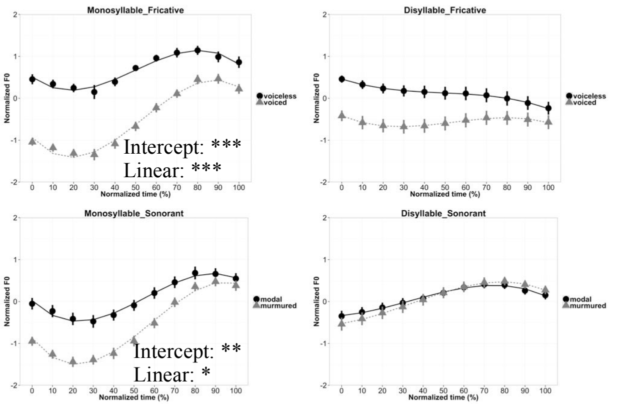
#### *2.2.3. F0*

The f0 results are given in Figure 3. For monosyllables, the addition of the laryngeal feature improved the model for the stops ($\chi^2(1)$=8.350, *p*=.004) and fricatives ($\chi^2(1)$=15.153, *p*<.001), and the addition of its interaction with the linear time term improved the model for the fricatives ($\chi^2(1)$=11.224, *p*<.001) and sonorants ($\chi^2(1)$=4.472, *p*=.034). Parameter estimates indicated that voiceless/modal consonants induced a higher f0 and a shallower f0 slope on the following vowel for all three manners.

For σ2 in disyllables, however, only for the fricatives did the laryngeal feature have a significant effect ($\chi^2(1)$=9.564, *p*=.050). No other model comparisons were significant (all *p*>.12). Parameter estimates indicated that the laryngeal effects on the intercept or higher time terms were not significant for any manner, including the fricatives.

**Figure 3**: F0 results for monosyllables (left) and σ2 of disyllables (right).






### 2.3. Discussion

Our production results indicate that the laryngeal contrast in Shanghai is primarily a tone contrast in the non-sandhi context, as although the H1-H2 comparison was generally in the expected direction [1, 5, 12, 15, 27], the difference did not reach significance under the Growth Curve Analysis; f0 curves, however, differed significantly on both the intercept and the slope for all three manners according to the consonant feature. In the sandhi context, the f0 difference was neutralized, but the stops and fricatives exhibited both duration and voicing differences. For the sonorants, however, no acoustic difference between the modal and murmured categories was detected. These results indeed show that the acoustic cues for the contrast vary by context: different manners and different positions cue the contrast differently.

Our results on consonant duration and voicing are consistent with those of earlier research [8, 10, 21, 23, 26]; but unlike [8] and [21], our phonation measures did not show a significant effect of the laryngeal feature. For f0, although we showed that it significantly covaried with the consonant feature in the non-sandhi context — a result shared by all previous research, we did not find incomplete neutralization in the sandhi context (cf. [10, 21, 26]). There are two potential reasons for the disparity. One is that, given our speakers were considerably younger than the speakers used in earlier studies, it is possible that Shanghai is gradually losing the phonation difference. Another possibility is that the different results are at least partially due to the different statistical methods used. For instance, when the H1-H2 result for stops in monosyllables was analysed in Repeated-Measures ANOVA, the effects of both voicing and its interaction with the time points were significant. But no such effects were found in the Growth Curve Analyses with a maximal random effects structure [2].

## 3. PERCEPTION EXPERIMENT

### 3.1. Methods

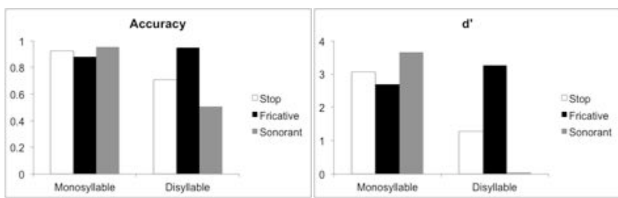
The stimuli for our perception study were monosyllabic and disyllabic words in which the target syllables were manipulated via cross-splicing and super-imposing three sets of cues for the contrast — consonant properties, vowel phonation, and f0. For instance, from two base tokens [pu$^{34}$] and [bu$^{13}$], six additional stimuli were constructed. One of them had the consonant and phonation properties of [pu$^{34}$] and the f0 properties of [bu$^{13}$] as the result of superimposing the f0 of [bu$^{13}$] onto [pu$^{34}$]; another had the consonant properties of [pu$^{34}$] and the phonation and f0 properties of [bu$^{13}$] by cross-splicing the consonant portion of [pu$^{34}$] to the vocalic portion of [bu$^{13}$], etc. The base tokens were selected from a female speaker's production data that were representative of the overall acoustic patterns, and all test stimuli were embedded in the same carrier sentence and presented to the subjects for a 2AFC task, for which they had to choose on a monitor the Chinese character(s) they heard. 41 native speakers (25F) with a mean age of 24.4 participated in the experiment in Shanghai.

For each manner and position, a logistic regression model with consonant, phonation, and f0 cues as predictors was fitted for each subject, and a Classification and Regression Tree (CART) analysis [3, 7] was also conducted to investigate how the listeners classified the stimuli based on these cues.

### 3.2. Results

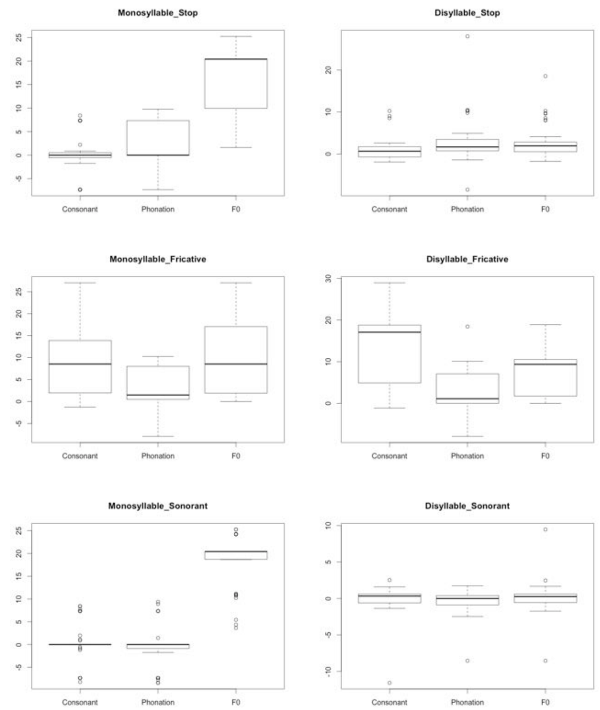
The accuracy and d' results for the listeners' classification of the natural tokens are given in Figure 4. These results indicate that the subjects had near perfect identification of the laryngeal class in the non-sandhi context regardless of manner and in the sandhi context for fricatives. For stops in the sandhi context, the identification was weaker, but well above chance; for sonorants, however, identification was at chance.

**Figure 4**: Accuracy and d' results for the natural tokens.




The coefficients for the consonant, phonation, and f0 cues in the logistic regressions for different manners and positions are given in Figure 5. The results indicate that for the monosyllables, f0 was the primary cue for the laryngeal classification of stops and sonorants, as indicated by its high positive coefficients, and the f0 and consonant cues were more important than the phonation cues for the classification of the fricatives. For σ2 of disyllables, the results for the fricatives were similar to monosyllables, but for stops and sonorants, there did not seem to be any cue that the listeners consistently relied on in classification, as the coefficients for all cues were near zero for most speakers.

**Figure 5**: Coefficients for the consonant, phonation, and f0 cues in logistic regressions for different manners and positions.




These generalizations are further supported by the CART analyses given in Figure 6. The available trees indicate that for stops and sonorants in monosyllables, f0 was the sole significant predictor for the subjects' classification; for fricatives in monosyllables, f0, consonant, and phonation cues all contributed, but their roles differed: f0 > consonant > phonation; for fricatives in disyllables, only the consonant and f0 cues were relevant, and the former was more important. The lack of classification for stops and sonorants in disyllables again shows that there was no reliable cue in this context.

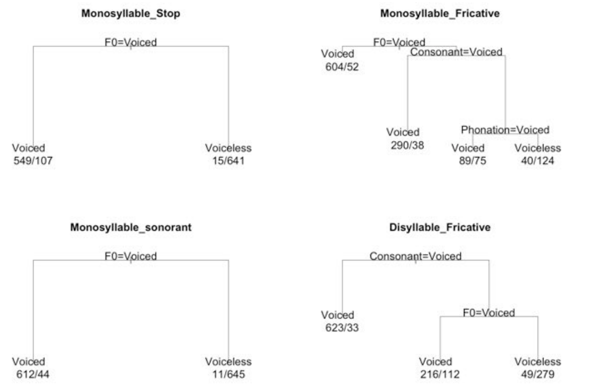



**Figure 6**: CART analyses for the three manners in monosyllables and the fricatives in σ2 of disyllables.

### 3.3. Discussion

The perception results were generally consistent with the aggregate production results: the laryngeal contrast in Shanghai was primarily cued by f0 in the non-sandhi context, and the f0 cue could override conflicting cues in the consonant or phonation; for the sandhi context, f0 became ineffective for stops and sonorants, but still had an effect on fricative classification. Different manners cued the contrast differently, and classification was the most robust for fricatives. For stops in the sandhi context, the listeners were able to classify the natural tokens at a relatively high rate, indicating the relevance of the consonant cue, but the effect of the cue was not strong enough to override conflicting cues from f0 and phonation, if any. For sonorants in this context, however, both the natural token identification and the classification of all stimuli demonstrate that there was simply no reliable cue for the laryngeal contrast.

## 4. CONCLUSION

We presented in this paper a case study for how a phonological contrast is cued in multiple phonetic dimensions. What is of particular interest is that the contrast in question — a laryngeal contrast in Shanghai Wu — is cued differently when realized on different manners (stops, fricatives, sonorants) and in different positions (non-sandhi, sandhi). Our acoustic results showed that although this contrast has been described as phonatory in earlier literature, it is primarily a tone contrast, at least in the younger speakers that we tested. Phonation correlates only appear in fricatives, and tone sandhi neutralizes the f0 difference. Our perception results were largely consistent with the aggregate acoustic results, indicating that speakers adjust the perceptual weights of individual cues for a contrast according to contexts. These findings support the position that phonological contrasts are formed by the integration of multiple cues in a language-specific, context-specific fashion and should be represented as such.

## 5. REFERENCES


[1] Andruski, J., Ratliff, M. 2000. Phonation types in production of phonological tone: The case of Green Mong. *J Int. Phon. Assoc.* 30, 37-61.

[2] Barr, D. J., Levy, R., Scheepers, C., Tily, H. J. 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *J. Mem. Lang.* 23: 255-278.

[3] Baayen, R. H. 2008. *Analyzing Linguistic Data: A Practical Introduction to Statistics Using R*. Cambridge: Cambridge University Press.

[4] Bates, D., Maechler, M., Bolker, B., Walker, S. 2014. *lme4: Linear mixed-effects models using Eigen and S4*. R package version 1.1-7. Available at http://CRAN.R-project.org/package=lme4.

[5] Blankenship, B. 1997. *The Time Course of Breathiness and Laryngealization in Vowels*. Ph.D. diss., UCLA.

[6] Boersma, P., Weenink, D. 2012. *Praat: doing phonetics by computer (Computer program)*. Available at http://www.praat.org/. Accessed 5 July 2012.

[7] Breiman, L., Friedman, J. H., Olshen, R. A., Stone, C. J. 1984. *Classification and Regression Trees*. Belmont, CA: Wadsworth.

[8] Cao, J.-F., Maddieson, I. 1992. An exploration of phonation types in Wu dialects of Chinese. *J. Phonetics* 20, 77-92.

[9] Chao, Y.-R. 1967. Contrastive aspects of the Wu dialects. *Language* 43, 92-101.

[10] Chen, Y.-Y. 2011. How does phonology guide phonetics in segment-f0 interaction? *J. Phonetics* 39, 612-625.

[11] DiCanio, C. 2014. Cue weight in the perception of Trique glottal consonants. *J. Acoust. Soc. Am.* 135, 884-895.

[12] Esposito, C. M. 2006. *The Effects of Linguistic Experience on the Perception of Phonation*. Ph.D. diss., UCLA.

[13] Flege, J. E., Wang, C. 1989. Native-language phonotactic constraints affect how well Chinese subjects perceive the word-final /t/-/d/ contrast. *J. Phonetics* 17, 299-315.

[14] Holt, L., Lotto, A. 2006. Cue weighting in auditory categorization: Implications for first and second language acquisition. *J. Acoust. Soc. Am.* 119, 3059-3071.

[15] Khan, S. D. 2012. The phonetics of contrastive phonation in Gujarati. *J. Phonetics* 40, 780-795.

[16] Lisker, L. 1986. “Voicing” in English: A catalogue of acoustic features signalling /b/ versus /p/ in trochees. *Language and Speech* 29, 3-11.

[17] Mirman, D. 2014. *Growth Curve Analysis and Visualization Using R*. Boca Raton: CRC Press.

[18] Oden, G. C., Massaro, D. W. 1978. Integration of featural information in speech perception. *Psych. Rev.* 85, 172-191.

[19] R Core Team. 2014. *R: A Language and Environment for Statistical Computing*, version 3.1.0. Vienna, Austria: R Foundation for Statistical Computing. http://www.R-project.org/.

[20] Raphael, L. J. 1972. Preceding vowel duration as a cue to the perception of the voicing characteristic of word-final consonants in American English. *J. Acoust. Soc. Am.* 51, 1296-1303.

[21] Ren, N.-Q. 1992. *Phonation Types and Stop Consonant Distinctions: Shanghai Chinese*. Ph.D. dissertation, University of Connecticut, Storrs.

[22] Repp, B. H. 1983. Trading relations among acoustic cues in speech perception are largely a result of phonetic categorization. *Speech Communication* 2, 341-361.

[23] Shen, Z.-W., Wang, W. S. 1995. Wuyu zhuoseyin de yanjiu — Tongji shang de fenxi he lilun shang de kaolü (A study of voiced stops in thje Wu dialects — Statistical analysis and theoretical considerations). In: Zee, E. (ed), *Wuyu Yanjiu (Studies of the Wu Dialects)*. Hong Kong: New Asia Books, 219-238.

[24] Shue, Y.-L., Keating, P., Vicenik, C., Yu, K. 2011. *VoiceSauce: A Program for Voice Analysis*. Available at http://www.ee.ucla.edu/~spapl/voicesauce/.

[25] Shultz, A. A., Francis, A. L., Llanos, F. 2012. Differential cue weighting in perception and production of consonant voicing. *J. Acoust. Soc. Am.* 132, EL95-EL101.

[26] Wang, Yizhi. 2011. *Acoustic Measurements and Perceptual Studies on Initial Stops in Wu Dialects — Take Shanghainese for Example*. Ph.D. diss., Zhejing University, Zhejiang, China.

[27] Wayland, R., Jongman, A. 2003. Acoustic correlates of breathy and clear vowels: The case of Khmer. *J. Phonetics* 31, 181-201.

[28] Xu, B.-H., Tang, Z.-Z. 1988. *Shanghai Shiqu Fangyan Zhi* (A Description of the Urban Shanghai Dialect). Shanghai Education Press.

[29] Xu, Y. 2005-2013. *ProsodyPro.praat.* Available at http://www.phon.ucl.ac.uk/home/yi/ProsodyPro/.

[30] Zhu, X.-N. 1999. *Shanghai Tonetics*. München: Lincom Europa.