THE PHONETIC REALIZATION OF INTERVOCALIC OBSTRUENTS IN SUZHOU WU CHINESE

Chunyu Ge, Peggy Mok

The Chinese University of Hong Kong
chunyuge@cuhk.edu.hk, peggymok@cuhk.edu.hk

ABSTRACT

Suzhou Wu Chinese has a three-way voicing contrast in obstruents and the realization of voiced obstruents is position-dependent. This paper investigated the phonetic realization of intervocalic voiceless unaspirated and voiced obstruents in Suzhou Wu. Acoustic data of disyllabic words from three age groups were collected. Closure duration, proportion of voicing, and partial voicing shape of intervocalic obstruents were analyzed. The results showed that although there were slight differences across age groups in the realization of intervocalic obstruents, all age groups produced only a small portion of voiced obstruents as fully voiced. Moreover, for both voiced and voiceless obstruents, most tokens were realized with the bleed pattern (voicing extending from the previous vowel), which was also consistent across age groups. The results contribute to our understanding of the voicing contrast in obstruents cross-linguistically and further study is needed to explore more factors influencing the realization of voicing.

Keywords: obstruents, phonetic realization, voicing, Wu Chinese

1. INTRODUCTION

The voicing contrast of obstruents can be effectively captured by the time between voice onset and the release burst of the obstruents (VOT), which can distinguish voiced (negative VOT), voiceless unaspirated (short positive VOT) and voiceless aspirated (long positive VOT) obstruents [1, 2]. However, some languages pose challenges to this unidimensional approach. For example, the voiced unaspirated stops and the voiced aspirated stops in Hindi both have negative VOT [3]. Another example is Korean, in which VOT is not sufficient to distinguish lax stops from aspirated stops, at least for younger speakers [4]. Therefore, more fine-grained description of voicing contrast is needed.

Furthermore, to describe the implementation of voicing in more detail is motivated by the fact that the realization of the voicing contrast in obstruents is influenced by various factors. [5] is an early exploration into this direction, which investigated the voicing profiles of stop closures in five languages using the probability of voicing during stop closure. They found that the voicing probability can be influenced by neighbouring sounds, as well as positional and prosodic factors. [6, 7] are more fine-grained in that the author not only analysed the voicing proportion during closure in American English, but also analysed the partial voicing shape of obstruents in intervocalic positions. She classified the partial voicing shape into bleed (voicing extending from previous sounds), prevoicing, hump (voicing strongest in the middle of the closure) and trough (voicing weaker in the middle of the closure than both ends) patterns and found that bleed is the most common implementation pattern of voiced obstruents in American English. Recently, [8] demonstrated that the position-dependent lenition of voiceless stops in Mixetc was modulated by prosodic conditions.

The language investigated in this study, Wu Chinese, has a three-way voicing contrast in obstruents (including stops, fricatives, and affricates) and the implementation of voiced obstruents is position-dependent. Wu Chinese distinguishes voiceless unaspirated, voiceless aspirated, and voiced obstruents [9]. Previous studies have shown that voiced obstruents in Wu Chinese are phonetically voiceless but with breathy voice in the following vowel in isolation and at word-initial positions, but are truly voiced (i.e., with voicing during consonant closure) at word-medial/intervocalic positions. [10] demonstrated with spectrogram that voiced obstruents in Suzhou Wu were not different in VOT from voiceless unaspirated obstruents at word-initial positions, but at intervocalic positions, voiced obstruents were phonetically voiced during closure, while voiceless unaspirated obstruents were voiceless. Consistent results were obtained in Shanghai Wu by [11] using fiberscope, showing that the vocal folds were vibrating during closure for voiced obstruents at intervocalic positions but not at word-initial positions. Therefore, the phonetic realization of the voicing of voiced obstruents in Wu Chinese is dependent on word position. A limitation of the previous studies is that they were usually based on visual inspections without measurements and they have not investigated the shape of the voicing, which may be more informative. Another background of this study is that our earlier studies [12, 13] showed that voiced obstruents in Suzhou Wu are changing in that younger speakers produce less breathy voice of voiced obstruents in isolation. As our earlier studies concentrated on voiced obstruents in isolation, this study aims to further explore the realization of obstruents at intervocalic positions across age groups and to investigate whether younger speakers also produce voiced obstruents differently from older and middle-aged speakers.

To sum up, this paper investigates fine-grained patterns of the realization of intervocalic obstruents in Suzhou Wu, and also to explore the difference of the realization across age groups. This paper hopes to contribute valuable empirical data to the cross-linguistic patterns of voicing contrast in obstruents.

2. METHOD

To compare the realization of intervocalic obstruents across age groups, native Suzhou Wu speakers of three age groups were recruited to participate in a word reading production experiment. Table 1 shows the number of speakers in each age group and their age range.

Table 1: Number of speakers in three age groups and their age range

Age group	Female	Male	Age range
older	5	5	65-84
middle-aged	10	6	43-54
younger	5	5	21-29

The test words were disyllabic words and the target syllable was the second one. The first syllables were all in CV structure, without any nasal or consonantal coda. The target syllables were in CV or CVN structure, where C included bilabial or alveolar stops /p b t d/ and alveo-palatal affricates /tɕ dʑ/. Due to phonotactic constraints, affricate onsets can only co-occur with rhymes starting with the high front vowel /i/, and the rhyme /in/ was selected for both the voiceless and voiced affricates. Likewise, the rhyme /in/ was also used for bilabial stops. But for alveolar stops, the rhyme /e/ was used due to the absence of suitable minimal pairs with the rhyme /in/. The participants were asked to read the words in isolation. Each word was repeated three times. The sounds were recorded using a Behringer ECM8000 microphone through the Scarlett Solo Studio 2nd Gen audio interface to a laptop, with a sampling frequency of 44,100 Hz.

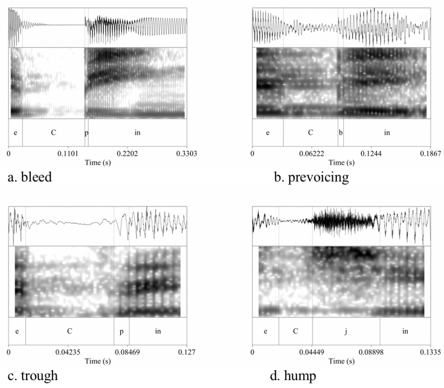

Figure 1: Waveforms and spectrograms of partial voicing shape patterns

The sounds were segmented manually in Praat [14]. The target element is the closure of the obstruent onset of the second syllable. The start of closure was determined as the end of the previous vowel's F2 and the end was the release burst of the stop or affricate. After segmentation, a Praat script was run to get the duration of the closure and the proportion of voicing during the closure using the *fraction of locally unvoiced frames* value available in *Voice Report*. The proportion of voicing was calculated by subtracting the fraction of locally unvoiced frames from one. The pitch ceiling was set at 500 Hz for female speakers and 300 for males, and the floor was at 100 Hz for females and 75 for males. Also, the closure was cut into three equal length intervals, and the proportion of voicing was measured during each interval. The partial voicing shape was determined based on the proportions of voicing of these three intervals (see [6] for more details of the definition). Before determining the partial voicing shape, all tokens which are fully voiceless were excluded. If the proportion of voicing decreased from the first interval to the second and the last, it is a bleed pattern. It is a prevoicing pattern, if the proportion increased from the first to the last interval. If the second interval had the greatest proportion of voicing, it is a hump pattern. Contrarily, it is a trough if the second interval had the smallest proportion of voicing. Figure 1 shows examples of each pattern.

3. RESULTS

3.1. Closure duration

Table 2 shows the mean closure duration of intervocalic obstruents across age groups. The closure duration was analyzed using linear mixed-effects models, with phonological Voicing, Manner of articulation and Age group as fixed effects and also as random slope by speaker. Stops generally had longer closure than affricates (Estimate = -36.01, $p < 0.01$). Voiceless obstruents had longer closure than voiced obstruents (Estimate = 16.39, $p < 0.01$). However, the inclusion of Age group into the model did not significantly improve the model fit ($\chi^2(2) = 0.49$, $p = 0.78$). Therefore, there is no statistically significant difference among older, middle-aged and younger speakers in closure duration of intervocalic obstruents.

Table 2: Mean closure duration (in ms) of intervocalic obstruents across age groups

laryngeal	manner	older	middle	younger
voiced	stop	111	123	118
	affricate	81	88	93
voiceless	stop	140	140	131
	affricate	97	95	104

3.2. Voicing proportion over the total closure duration

The tokens were divided into fully voiced, partially voiced and fully voiceless, according to the proportion of voicing during the entire closure duration. Those with more than 90% voicing were fully voiced, while those with less than 10% were fully voiceless, and partially voiced with voicing ratio in-between, following [6]. Figure 2 shows the percentage of the three realizations of stops and affricates across age groups. Although for both voiced and voiceless obstruents, most tokens were partially voiced, voiced obstruents had more fully voiced tokens than those of voiceless obstruents, and likewise, less fully voiceless tokens than those of voiceless obstruents. The difference between voiced and voiceless obstruents was also slightly smaller for younger speakers than older speakers.

To substantiate the observations, generalized linear mixed-effects models were fitted on the type of overall voicing, which was coded as a binary variable with fully voiced as 1 and partial voicing and fully voiceless as 0. Phonological Voicing and Manner were included as fixed effects and also as random slope by speaker. Closure was also included

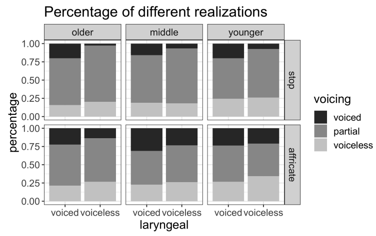

Figure 2: Percentage of fully voiced, partially voiced, and fully voiceless stops and affricates across age groups

as a fixed effect. Voiceless obstruents were less likely to be fully voiced than voiced obstruents (Estimate = -0.22, $p < 0.01$), and the longer the closure duration was, the less likely it was to be fully voiced (Estimate = -0.06, $p < 0.01$). But again, the inclusion of Age group did not significantly improve the model fit ($\chi^2(2) = 1.53$, $p = 0.46$). Also, no significant difference was found between stops and affricates ($\chi^2(1) = 1.45$, $p = 0.23$). Therefore, intervocalic voiced obstruents were more likely to be fully voiced than voiceless obstruents, and there was no significant difference among older, middle-aged and younger speakers.

3.3. Partial voicing shape

Apart from the overall proportion of voicing, the proportion of voicing was also measured during the three equal length intervals of the closure, in order to investigate the fine pattern of the implementation of voicing. Figure 3 shows the mean proportion of voicing during each interval for voiceless and voiced obstruents across age groups. The data were also analyzed using linear mixed-effects models with phonological Voicing, Manner, Interval, Age group, and their interactions as fixed effects, whereas Manner and Voicing were also included as random slopes by speaker. The effect of these factors was determined by model comparison. The most remarkable effect was interval. The first interval had higher proportion of voicing than both the second and third intervals. Once again, Age group did not improve the model fit significantly ($\chi^2(2) = 1.07$, $p = 0.59$). What is interesting here is that there was a significant interaction between Interval and Voicing ($\chi^2(2) = 10.41$, $p < 0.01$). For all three intervals, voiced obstruents had significantly higher voicing proportion than voiceless obstruents.

The partial voicing shape was also determined using the criteria mentioned in the Method and the proportion of each partial voicing shape for voiced and voiceless obstruents across age groups was shown in Figure 4. It is quite clear that the bleed pattern took up the largest proportion for both stops and affricates, regardless of their voicing. This observation was consistent among all three age groups. The bleed pattern indicates that the voicing in the closure is extending from the previous vowel. Among the other three patterns, the prevoicing took up the largest proportion, while the hump and trough patterns were extremely rare.

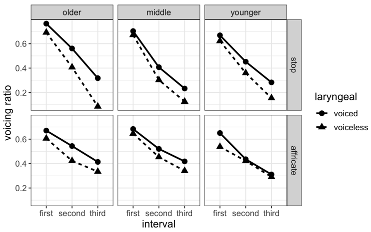

Figure 3: Proportion of voicing by interval of intervocalic stops and affricates across age groups

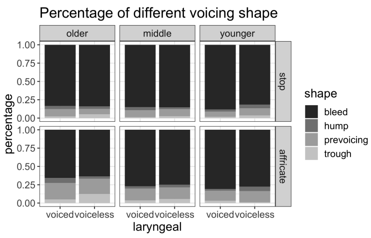

Figure 4: Percentage of different voice shape of intervocalic stops and affricates across age groups

4. DISCUSSION

This paper investigated the implementation of intervocalic voiceless unaspirated and voiced obstruents in Suzhou Wu across age groups. The most striking finding is that only a small portion of intervocalic voiced obstruents were implemented as fully voiced, which is different from the observation of previous studies that intervocalic obstruents in Wu Chinese were usually truly voiced. It is also found that the most frequent partial voicing pattern was the bleed pattern, in which the voicing was extending from the previous vowel. Moreover, younger speakers did not have different realizations from older and middle-aged speakers, neither in closure duration nor in voicing proportion. Furthermore, listeners may be sensitive to the greater proportion of voicing in voiced obstruents, which can be further investigated using a perception experiment.

Although only a small portion of voiced obstruents was realized as fully voiced, it is still premature to draw the conclusion that the voiced obstruents in Suzhou Wu has changed into voiceless even in the intervocalic position. First, the probability of being fully voiced was larger for voiced than for voiceless obstruents and the proportion of voicing was greater for voiced than for voiceless obstruents for all three intervals during closure. Therefore, speakers still make more efforts to maintain voicing in voiced obstruents than in voiceless obstruents. This can be seen in Figure 3, at the third interval, voiced obstruents had larger proportion of voicing than voiceless obstruents and the proportion of voicing of voiceless stops nearly dropped to zero for older and middle-aged speakers. Another finding is that the bleed pattern is the most common realization pattern of both voiceless and voiced obstruents in Suzhou Wu, which is consistent with the findings in American English [6, 7]. This could be evidence that the prevalence of bleed in obstruents is a cross-linguistic trend, which has phonetic groundings. For instance, [15] showed that voicing will continue into closure until the oral pressure builds up to inhibit voicing.

One of the conditions this study failed to control is the prosodic structure of the disyllabic word. As mentioned in the Introduction, prosodic conditions can influence the realization of stops [8]. In Suzhou Wu, as in other Wu dialects, prosodic structure can influence the pattern of tone sandhi [16]. Left-dominant tone sandhi applies in prosodic words and right-dominant tone sandhi applies in prosodic phrases. Left-dominant tone sandhi implies a closer relationship between two syllables than right-dominant tone sandhi. However, the speakers in this study read most of the words with right-dominant tone sandhi. It is possible that the realization of intervocalic obstruents can be influenced by tone sandhi, and they are more likely to be fully voiced only in those disyllabic words which undergo left-dominant tone sandhi. This would be an interesting topic for further study.

5. CONCLUSION

The phonetic realization of the voicing of intervocalic obstruents in Suzhou Wu Chinese was investigated in this paper. Although only a small portion of voiced obstruents was realized as fully voiced, their probability of being fully voiced and their proportion of voicing were greater than voiceless obstruents. Younger speakers had similar realizations of obstruents with older and middle-aged speakers, both in terms of closure duration and voicing. Prosodic factors should be taken into consideration for further investigation of the implementation of voicing in Suzhou Wu obstruents.

6. ACKNOWLEDGEMENTS

This study was supported by the RGC General Research Fund 2019/20 Project Reference 14607619 awarded to the second author and the CUHK Research Fellowship Scheme 2019-20, 2020-21 awarded to the first author.

7. REFERENCES

[1] L. Lisker and A. S. Abramson, "A Cross-Language Study of Voicing in Initial Stops: Acoustical Measurements," *Word*, vol. 20, no. 3, pp. 384–422, Jan. 1964.

[2] T. Cho, D. H. Whalen, and G. Docherty, "Voice onset time and beyond: Exploring laryngeal contrast in 19 languages," *Journal of Phonetics*, vol. 72, pp. 52–65, Jan. 2019.

[3] J. Schertz and S. Khan, "Acoustic cues in production and perception of the four-way stop laryngeal contrast in Hindi and Urdu," *Journal of Phonetics*, vol. 81, p. 100979, Jul. 2020.

[4] D. J. Silva, "Acoustic evidence for the emergence of tonal contrast in contemporary Korean," *Phonology*, vol. 23, no. 2, pp. 287–308, Aug. 2006.

[5] C. Shih, B. Möbius, and B. Narasimhan, "Contextual effects on consonant voicing profiles: A cross-linguistic study," in *Proceedings of the 14th International Congress of Phonetic Sciences*, vol. 2, San Francisco, 1999, pp. 989–992.

[6] L. Davidson, "Variability in the implementation of voicing in American English obstruents," *Journal of Phonetics*, vol. 54, pp. 35–50, Jan. 2016.

[7] ——, "Phonation and laryngeal specification in American English voiceless obstruents," *Journal of the International Phonetic Association*, vol. 48, no. 3, pp. 331–356, Dec. 2018.

[8] C. DiCanio, W.-R. Chen, J. Benn, J. D. Amith, and R. Castillo García, "Extreme stop allophony in Mixtec spontaneous speech: Data, word prosody, and modelling," *Journal of Phonetics*, vol. 92, p. 101147, May 2022.

[9] J. Cao and I. Maddieson, "An Exploration of Phonation Types in Wu Dialects of Chinese," *Journal of Phonetics*, vol. 20, pp. 77–92, Jan. 1992.

[10] F. Shi, "Acoustic characteristics of voiced obstruents in Suzhou Chinese [in Chinese]," *Studies in Language and Linguistics (Yuyan Yanjiu)*, no. 01, pp. 49–83, 1983.

[11] N. Ren, "Phonation types and stop consonant distinctions-Shanghai Chinese," Ph.D. dissertation, The University of Connecticut, 1992.

[12] C. Ge, W. Xu, W. Gu, and P. Mok, "An electroglottographic study of phonation types in tones of Suzhou Wu Chinese," in *Speech Prosody 2022*. ISCA, May 2022, pp. 570–574.

[13] C. Ge, W. Xu, W. Gu, and P. P. K. Mok, "The change in breathy voice after tone split: A production study of Suzhou Wu Chinese," *Journal of Phonetics*, vol. 98, p. 101239, May 2023.

[14] P. Boersma, "Praat: Doing phonetics with computers," 2020.

[15] J. J. Ohala, "The Origin of Sound Patterns in Vocal Tract Constraints," in *The Production of Speech*, P. F. MacNeilage, Ed. New York, NY: Springer, 1983, pp. 189–216.

[16] X. Li, "Tone sandhi in the Suzhou dialect: Synchronic and diachronic perspectives," *Journal of Chinese Linguistics Monograph Series*, no. 24, pp. 21–32, 2010.