Whispered Speech in Shanghainese

Xiaoya Wang [*]

August 23, 2023

Abstract

Shanghai Chinese is a Wu dialect, also known as Shanghainese and Shanghai Wu, mainly spoken in and around one of the largest cities in China - Shanghai, where 24 million people reside. Shanghai Wu is an underresearched language than other tonal languages in China (e.i. Mandarin, Cantonese). The present study aims to investigate the phonetic properties of Shanghai Wu with a concentration on breathy vowels, and how they are produced in the speech mode of whisper compared to normal speech, more importantly, the cue realization and cue weighting in whispered speech as cross-linguistic phenomena.

[*]Department of Linguistics, University of California, Berkeley. E-mail: xiaoyawang@berkeley.edu

Contents

1 Introduction

Shanghainese, also known as **Shanghai Wu** and **Shanghai Chinese**, is a Wu dialect mainly spoken in and around one of the largesst cities in China - Shanghai. (Shao 2013) Shanghainese is an underresearched tonal language compared to other Chinese languages (i.e. Mandarin & Cantonese). Also, linguists have a divergence of descriptions about phonetic and phonological phenomena in Shanghai Chinese as well.

According to Penn Language center [1], Shanghainese has 15 vowels, and 8 are phonemic as follows: [i][y][ɪ][e][ø][ɛ][ə][ɐ][a][ɑ][ɔ][ɤ][o][ʊ][u]. Sorted by surrounding environment, 9 vowels ([i][y][ø][ɛ][a][ɔ][ɤ][o][u]) occur in open syllables, while 6 vowels occur in closed syllables. (Chen & Gussenhoven 2015)

As a tone language, Shanghainese has a register tone system consisting of 5 tones, which is different from contour tone system like Mandarin and Cantonese. Defined by five-scale pitch system (Chao 1930), tones are Tone 1 (53), Tone 2 (34), Tone 3 (23), Tone 4 (55), and Tone 5 (12), where Tone 3, 4 & 5 are lower-register and Tone 1, 2 & 4 are high-register. (Jiayin & Hallé 2017)

Compared to other Wu dialects, which also has language contact with Mandarin, Shanghainese has relatively less tones. (Rose 2015) The assumption is that as a cosmopolitan immigrant city, Shanghainese has been in language contact with non-tonal languages (i.e. Japanese & English), which may result in the merging of tones. (Ballard 1984)

Chinese linguists have described the consonant system of Shanghainese by high-low registers and voicing feature with terminology "Yin" and "Yang" for centuries. In phonology, Yin consonants are [+register] [−voice], while Yang consonants are [−register] [+voice]. (Jiayin & Hallé 2017, Gao et al. 2020)

Previous studies have attested that the occurrence of breathy vowels increases after low-register consonants. (Gao et al. 2020)

[1] Controversy of vowel system remains

The phonological rule to summarize the phonemenon would be:

$$[+\text{syllabic}, +\text{spread glottis}] \rightarrow [+syllabic, -spreadglottis]/[\text{-}syllabic,\text{-}register, +voice]\_$$

The present study aims to investigate the temporal and spectral properties of whispered and normal speech in Shanghainese.

2 Methodology

2.1 Task Design

Two word-list elicitation tasks were performed. The elicitation task was a list of CV structure syllables. The feature bundles of consonants were [−register] [+voice]. Two task shared the same word list (Table 1), which was used in a previous study. (Gao et al. 2020) The tasks only differed in speech mode.

Table 1. T2-T3 minimal pair monosyllabic words; there is no [nɛ34] syllable in Shanghainese.

	ø	stop	fricative	nasal
T2(34, yin)	ɛ‘love’	pɛ‘board’	fɛ‘reverse’	mɛ‘beautiful’
		tɛ‘gallbladder’	sɛ‘umbrella’	
T3(23, yang)	ɛ‘salty’	pɛ‘handle’	fɛ‘annoying’	mɛ‘plum’
		tɛ‘desk’	sɛ‘greedy’	

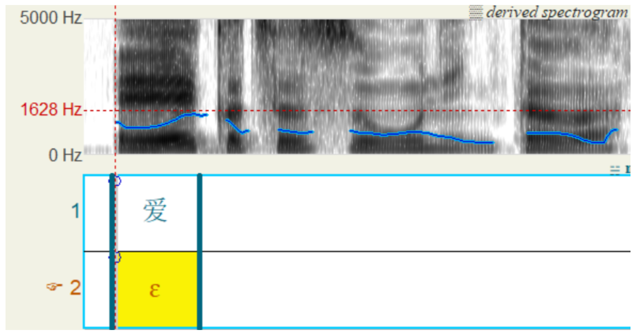

Figure 1: Spectrogram of Tone Sandhi.

2.2 Subjects

The participant was a senior male student studying at University of California, Berkeley. He was 22, born and raised in the downtown of Shanghai. Grandparents and parents were identified as native-Shanghainese speakers. The subject was native-speakers of Mandarin and Shanghainese with English as the second language.

The subject was required to speak the target word + "I know this word" as a sentence in Shanghai Chinese. The target words were located at the initial of sentence to avoid the effect of tone sandhi (Zee & Maddieson 1979) as shown in Figure 1. For the first time, the subject was asked to produce sentences in the speech mode of whisper. After the task and a ten-minute interval of break, he was told to speak the word list again in normal speech. Each sentence was repeated for six time, and three time respectively in each speech mode.

2.3 Measurements

According to related study, six parameters were selected for measurements (Heeren 2015):

1. f_0 in normal speech
2. Formant locations
3. Formant bandwidths
4. Center of Gravity (CoG)
5. Relative vowel duration
6. Mean intensity

3 Hypotheses

Here are some general hypotheses:

1. There is a cue trading relationship between fundamental frequency, intrinsic pitch and breathiness. (Heeren 2015)

2. Breathiness can be reflected as the increase of airflow and the decrease of pharyngeal cavity. (Maddieson & Ladefoged 1985)

3. Cues are universal. Shanghai Wu shares a similar pattern with other tonal languages in whisper speech. (Pfiffner 2023)

4 Prediction

In the whispered speech, the predictions for eight measurements are (Heeren 2015):

Center of Gravity (CoG) would go up. Mean vowel intensity would increase. Formant bandwidths would narrow down. Vowel duration would extend. Systematic change would be found.

5 Data Analysis

The segmentation and transcription was primarily processed in ELAN (*ELAN Version 6.4* 2022). The data analysis were mainly taken in Praat (Boersma, Paul and Weenink, David 2023).

5.1 F0 in normal speech

Before diving into the comparison between normal and whispered speech, the section shows how F0 varied between pitch targets for an individual (Figure 2, 3).

As we can tell from the two figures, T3 is more

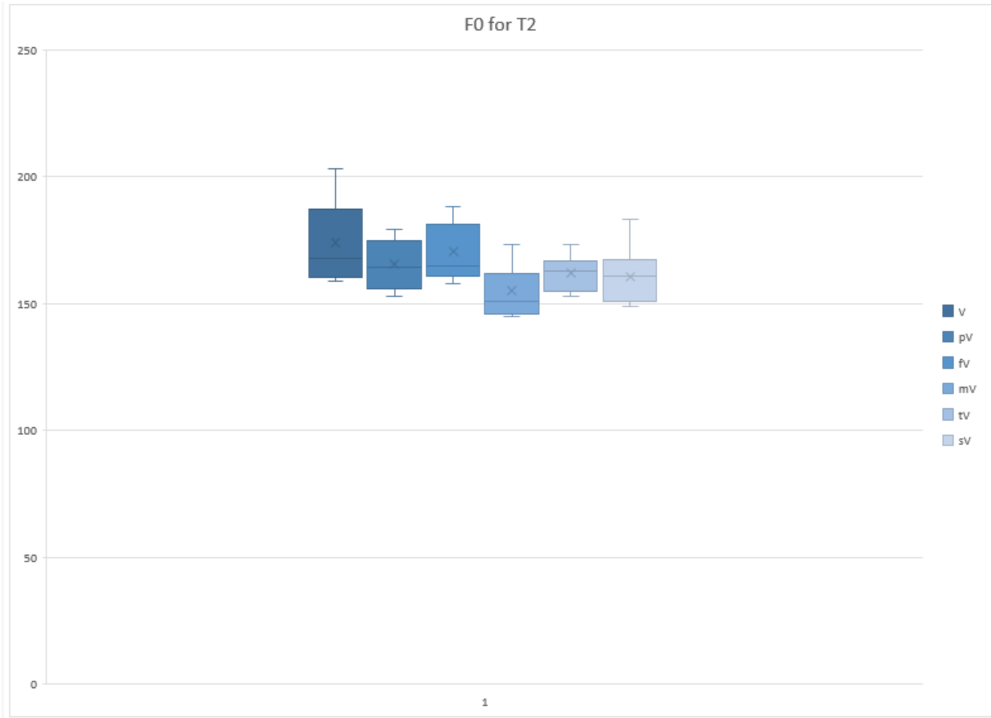

Figure 2: F0 for T2.

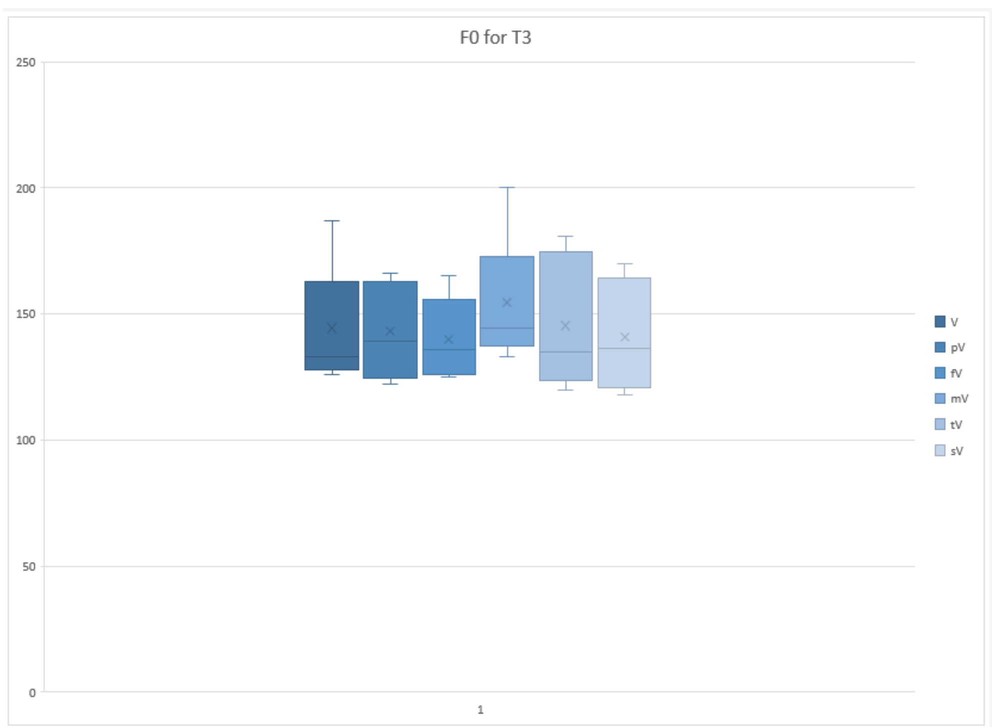

Figure 3: F0 for T3.

5.2 Spectural measures

5.2.1 Formant locations

Table 2. Mean and standard deviations (in the parentheses) across registers, for each of the register (Yin, Yang), for the first through third formants (F1-F3, in Hz), Center of Gravity (CoG, in Hz), for whispered speech and normal speech.

	Normal Speech		Whispered Speech	
	T2	T3	T2	T3
F1	455 (22)	482 (11.4)	708 (96)	861 (172.2)
F2	2085 (28)	2063 (55.2)	2225 (100.6)	2281 (85.9)
F3	2689 (78)	2597 (54.2)	2796 (209.2)	2802 (59.9)
CoG	547 (28)	592 (64.3)	581 (248.6)	428 (256)

In the Table 2, the formant locations were most significantly changed in F1 (increase 56% for the high-register, 79% for the low-register, in Hz). In whispered speech, the formant locations were more scattered.

5.2.2 Formant bandwidths

Table 3. Mean and standard deviations (in and out of the parentheses respectively) across registers, for each of the register (Yin, Yang), for the first through third bandwidths (B1-B3, in Hz), Intensity (in Hz), for whispered speech and normal speech.

	Normal Speech		Whispered Speech	
	T2	T3	T2	T3
B1	50 (9.8)	53 (15.1)	198 (116.3)	187 (74.3)
B2	156 (26.4)	191 (21.3)	451 (384)	1273 (996.2)
B3	263 (65.4)	252 (68)	342 (104)	329 (224)
Intensity	81 (1.6)	82 (0.5)	56 (6.3)	58 (3.6)

In Table 3, formant bandwidths increased greatly in whispered speech for Tone 3, and especially for B2. Contrary to the prediction, the formant bandwidths get wider.

5.2.3 Center of gravity

In the previous study, CoG was higher in whispered than in normal speech. (Heeren 2015) In this case, the CoG for lower-register was getting lower. The assumption is that in order to be differed from the high-register, the CoG may be a strong cue.

5.3 Temporal property

The vowel duration for whispered speech was longer than in normal speech. The five longest vowel duration was in whispered speech. The five shortest vowel duration was in normal speech. In whispered speech, lengthening the vowel duration may be a cue. My assumption is that when the intensity is lower, having more time to convey information is a good strategy.

Other study indicates that duration is the secondary cue for perception of voicing and tones in Shanghai Chinese. (Gao & Hallé 2013)

5.4 Mean intensity

In table 3, the experiment attested that intensity was lower in whispered speech than normal speech. Previous research shows that “intensity differed for different registers and fundamental frequencies". (Hirano et al. 1969) Present study indicates that intensity differed for different speech modes as well. ($p>0.05$).

6 Discussion

The measurements of airflow, glottal pressure and electroglottography can be taken in the future. More data from Shanghainese speakers, and speakers of other Wu dialects is required to have a speech pattern for individual variability and whispered speech in a language family. Also, the comparison of whispered speech with other tonal languages can be further explored.

7 Acknowledgment

I would like to thank my supervisor & instructor Professor Alexandra Pfiffner for her brilliant personality and introducing me to the wonderland of phonetics (physically openning the door to the PhonLab at the final week), Professor Keith Johnson for his expertise in the field of phonetics and sharing of literature review, Zijie Jiang for volunteering in the experiment and helping with task design, my peers in Ling 111 Phonology, Ling 113 Experimental Phonetics and Ling 210 Advanced Phonetics for the great fellowship this semester, the Society of Linguistics Undergraduate Students (SLUgS) for the sense of belonging, my friends in the church, and my parents for unconditional love.

References

Ballard, W. (1984), 'Wu, min and a little hakka lexical tone sandhi: right and left', *Cahiers de Linguistique Asie Orientale* **13**(1), 3–34.

Boersma, Paul and Weenink, David (2023), 'Praat: doing phonetics by computer', Computer program. Version 6.3.10, retrieved 3 May 2023 from http://www.praat.org/.

Chao, Y. R. (1930), ‘A system of ’tone-letters”, *Le Maître Phonétique* **45**, 24–27.

Chen, Y. & Gussenhoven, C. (2015), ‘Shanghai chinese’, *Journal of the International Phonetic Association* **45**(3), 321–337.

ELAN Version 6.4 (2022), Computer software. Nijmegen: Max Planck Institute for Psycholinguistics, The Language Archive.
URL: *https://archive.mpi.nl/tla/elan*

Gao, J. & Hallé, P. (2013), ‘Duration as a secondary cue for perception of voicing and tone in shanghai chinese’, pp. 3157–3161.

Gao, J., Hallé, P. & Draxler, C. (2020), ‘Breathy voice and low-register: A case of trading relation in shanghai chinese tone perception?’, *Language and Speech* **63**(3), 582–607.

Heeren, W. (2015), ‘Vocalic correlates of pitch in whispered versus normal speech’, *The Journal of the Acoustical Society of America* **138**(6), 3800–3810.

Hirano, M., Ohala, J. & Vennard, W. (1969), ‘The function of laryngeal muscles in regulating fundamental frequency and intensity of phonation’, *Journal of Speech and Hearing Research* **12**(3), 616–628.

Jiayin, G. & Hallé, P. (2017), ‘Phonetic and phonological properties of tones in shanghai chinese’, *Cahiers de Linguistique Asie Orientale* **46**(1), 1–31.

Maddieson, I. & Ladefoged, P. (1985), ‘“tense” and “lax” in four minority languages of china’, *Journal of Phonetics* **13**(4), 433–454.

Pfiffner, A. (2023), ‘Cue-based feature’.

Rose, P. (2015), ‘Tonation in three chinese wu dialects’.

Shao, Q. (2013), ‘Shanghai gone: Domicide and defiance in a chinese megacity’.

Zee, E. & Maddieson, I. (1979), 'Tones and tone sandhi in shanghai: Phonetic evidence and phonological analysis', *UCLA Working Papers in Phonetics* **45**, 93–129.