Proceedings of the International Conference on Global Politics and Socio-Humanities
DOI: 10.54254/2753-7064/17/20230774

Comparing the Discrimination of /b/-/p/-/pʰ/ among Shanghainese and Mandarin Native Speakers

Zixuan Kang[1,a,*]

[1]*Ningbo Xiaoshi High School, Ningbo, Zhejiang, China*
a. kangzixuan@nbxiaoshi.net
**corresponding author*

Abstract: Shanghainese speakers show a variety of phonological phenomena depending on the region and age of the speaker. Junior Shanghainese speakers exhibit a mixture of Mandarin and Shanghainese phonology and usually speak Shanghainese with many discernible differences compared to the old generation of Shanghainese speakers. Crucially, Shanghainese speakers often acquire Shanghainese as a second language and transfer Mandarin phonology to it. This paper aims to examine the ability to discriminate the three-way contrast /b/-/p/-/pʰ/ among senior Shanghainese speakers, junior Shanghainese speakers, and Mandarin speakers through an experimental study. The experiments show that: 1) Senior Shanghainese speakers perform less well when presented with isolated stop sounds, and they use other acoustic cues rather than plain VOT differences in everyday speech; 2) the acquisition of Shanghainese as a second language only has a limited effect on junior Shanghainese speakers' internal phonology. Therefore, the contribution of this study is that it has determined the importance of VOT in the discrimination of Shanghainese stops and has explored the degree of influence exerted by the acquisition of an L2.

Keywords: Shanghainese, Mandarin, VOT, stops, L2 acquisition

1. Introduction

Unlike many Romance languages that distinguish voiced stops and voiceless stops, Mandarin is a language that only contrasts aspiration in stops and does not have any voiced stops:

Table 1: Some Chinese contrasting pairs.

a.	拔 [pa35]	'to pull out'	爬 [pʰa35]	'to climb; to creep'
b.	大 [ta51]	'big'	踏 [tʰa51]	'to tread'
c.	该 [kai55]	'should; ought to'	开 [kʰai55]	'to open'

As is shown in Table 1, Mandarin implements aspiration contrast of voiceless stops but does not possess any voiced stops. The stops of Shanghainese are more complicated. Like other northern Wu dialects, the Shanghai dialect has voiced initials [b d g ɦ z v dʑ ʑ] [1]. Moreover, based on its development, Shanghainese can be divided into three major historical types, the so-called 'old', 'middle', and 'new' types. The old type is generally considered obsolete, and what we refer to as Shanghainese today only includes the middle and new types in the Urban Branch of Shanghainese

DOI: 10.54254/2753-7064/17/20230774

[2]. The middle type of Shanghainese is spoken by people born between 1940 to 1965, and the new type is spoken by people born after 1965. Different phonologies can be observed depending on the regional and type differences in Shanghainese. The phonetic quality of stops can vary depending on the environment. In real utterances, the preservation of vocal fold vibration in Shanghainese reduces the functional cost of neutralizing tone in second syllables [3]. In isolate monosyllables, it is generally considered that Shanghainese possesses the complete three-way contrast in stops (e.g., /b/-/p/-/pʰ/). Adopting a somewhat different schema, the three-way contrast can collapse into just two: aspirated vs. unaspirated and modal voice vs. breathy voice:

Table 2: Some Shanghainese words.

a. 边 [pi] 'side'	unaspirated voiceless stop + modal vowel
b. 偏 [pʰi] 'slanting; partial'	aspirated voiceless stop + modal vowel
c. 皮 [pi̤] 'skin'	unaspirated voiceless stop + breathy vowel

As is shown in Table 2, Shanghainese does possess voiced stops. When the stops are not isolated, they have more complicated behaviors. Under the first schema, it can be said that voiced stops are phonetically voiceless with slack voice phonation in stressed, word-initial position. These consonants are truly voiced in intervocalic positions. Under the second schema, it can be observed that when '皮' [pi̤] is at non-word-initial positions, it has two free variants: [bi̤] and [bi]. In the northern part of Shanghainese, the so-called "voiced stops" have a VOT greater than or approximately equal to zero, so people can only effectively distinguish '边' [pi] and '皮' [pi̤] by the phonation types of the vowels. When doing phonological analyses, if people want to simplify the treatment for vowels, they generally consider the underlying form of [pi̤] as /bi/, just as in the first schema; if people want to simplify the treatment for consonants, then they need to set two sets of vowels, modal and breathy (although some may analyze the non-modal phonation as "whispery voice" in place of "breathy voice" as is adopted here [4]) as in the second schema. Since only the speakers' discrimination ability is concerned about here and only isolated monosyllabic sounds were used in the experiments, the simplest, /b/-/p/-/pʰ/ three-way contrast is adopted henceforth.

It has been shown in many researches that the acquisition of L2 is influenced by L1 to some extent, especially phonologically [5,6]. To examine the categorical perception of /b/-/p/-/pʰ/ among native Mandarin and Shanghainese speakers and determine to what extent the acquisition of a lingua franca before the internalization of the dialect will affect one's phonological allocation of the phonemes /b/, /p/, and /pʰ/ among junior Shanghainese, experiments were designed to investigate these questions.

2. Problem Statement

There is a significant difference between the perception of the voicing of stops of Mandarin and Shanghainese Native Speakers. Senior Shanghainese speakers are raised in a full Shanghainese environment and preserve the voiced stops in their accents. However, younger generations living in Shanghai grew up in a bilingual environment and only speak Shanghainese at home. Fewer people can grasp Shanghainese, and this drop happens more quickly in younger generations [7]. Many often acquire Shanghainese in search of a cultural identity after the acquisition of Mandarin. Thus, their accents have already been influenced by Mandarin. Whether or not the junior Shanghainese speakers will be at a somewhat intermediate position between Mandarin speakers and senior Shanghainese in terms of the performance of discriminating [p] and [b] is to be examined in the experiments.

DOI: 10.54254/2753-7064/17/20230774

3. Method

The experiment is mainly a replication of Miyawaki et al., with slight adaptations [8].

3.1. Subject

Three groups of participants, consisting of native Mandarin speakers, native Shanghainese speakers who were over the age of 40, and native Shanghainese speakers who were less than 20, were assembled. 10 participants per group. The third group consisting of junior Shanghainese speakers that were less than 20 all acquired Shanghainese as a second language with average mastery (Please be aware that they were not those junior Shanghainese speakers who can speak the new type of Shanghainese as their first language. They only acquired the middle type of Shanghainese). None of the subjects had any prior knowledge or training in phonology or phonetics. None of the Mandarin speakers had any prior knowledge of Shanghainese. All subjects were reported to have normal hearing and were paid to take part in the experiment.

3.2. Material

The Abramson/Lisker VOT Stimuli at Haskins Laboratories were used throughout the experiment. The stimuli were created by a formant synthesizer, and VOT variants were synthesized ranging from voicing starting 150 ms before the release ("M" for minus) to 150 ms after the release ("P" for plus). Different languages have different settings of the VOT of their consonants. For instance, English, Spanish, and Mandarin are languages with two phonological stop categories per place of articulation but a different implementation of contrast in terms of VOT. Generally, it could be stated that in the word-initial position, Mandarin has the longest VOT values for fortis stops, followed by English, whereas Spanish presents the shortest values [9]. In the experiment, the average values in Mandarin and Shanghainese were considered, and only the stimuli from LABM040 to LABP040 were used.

3.3. Procedure

The experiment was divided into two major components.

The first part is aimed at determining the perceptual boundaries of the three groups of speakers. First, the subjects listened to the ordered series of stimuli from LABM040 to LABP040. Next, they will be told that the stimuli were several instances of some Shanghainese syllables [ba51], [pa51], and [p^ha51]. At this point, some rough explanations will be given to Mandarin speakers. Next, they were again presented with the ordered series, two stimuli at a time (e.g., LABM040 and LABM030, LABM030 and LABM020, etc.), with roughly one second of pause between each two stimuli, and they were instructed to tell whether the two sounds were different or not. The procedure ran three times. The data were collected.

The second part was designed to test the discrimination ability among the three groups of speakers. The series from LABM050 to LABP050 were divided into 6 pairs that were three steps apart (LABM040-LABM010, LABM030-LAB000, etc.). For each pair, 6 triads were constructed by duplicating one stimulus of the pair (ABB, BAB, BBA, BAA, ABA, AAB. See Liberman et al. for a detailed way of constructing the triads [10]). Each triad was then presented to the subjects, and the subjects were told that they could use any criterion to make judgments about which one in the triad was different, including guessing. Between every two stimuli were roughly ten seconds of pause, and the six triads were played with roughly 10 seconds between every two triads, during which the subject made decisions. There were six rounds like this since there were six groups of triads, each consisting of six triads. The process ran three times. The data were collected.

DOI: 10.54254/2753-7064/17/20230774

4. Hypothesis

The results from the first part of the experiment are expected to be threefold. First, the Mandarin speakers will successfully mark stimuli in the neighborhood of LABP020, which is supposed to be the boundary between /p/ and /pʰ/, but they will fail to mark stimuli in the neighborhood of LABM020, which is supposed to be the boundary between /b/ and /p/. In contrast, the senior speakers will successfully mark both boundaries mentioned above. In addition, the junior speakers will perform approximately as the Mandarin speakers, but slightly better at distinguishing the boundary between /b/ and /p/.

The results from the second part of the experiment are also expected to be threefold. The Mandarin speakers will perform well when discriminating triads containing LABP020 but will perform poorly when discriminating triads containing LABM020. However, the senior Shanghainese speakers will perform well in both tasks mentioned above. Also, the junior Shanghainese speakers will have about the same performance as the Mandarin speakers but will have a slightly higher percentage of correct answers when performing the second task.

5. Results

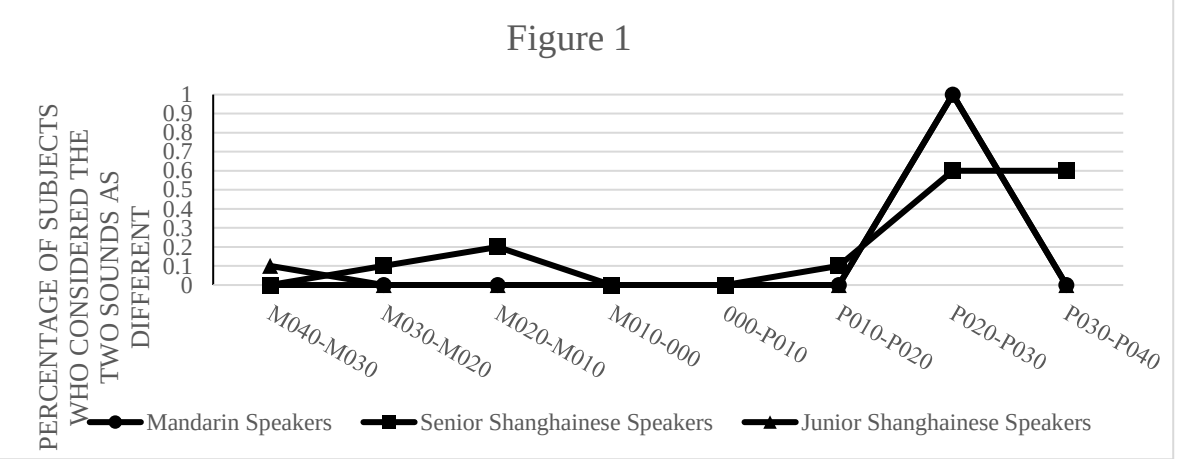

Figure 1: Results from the first part of the experiment.

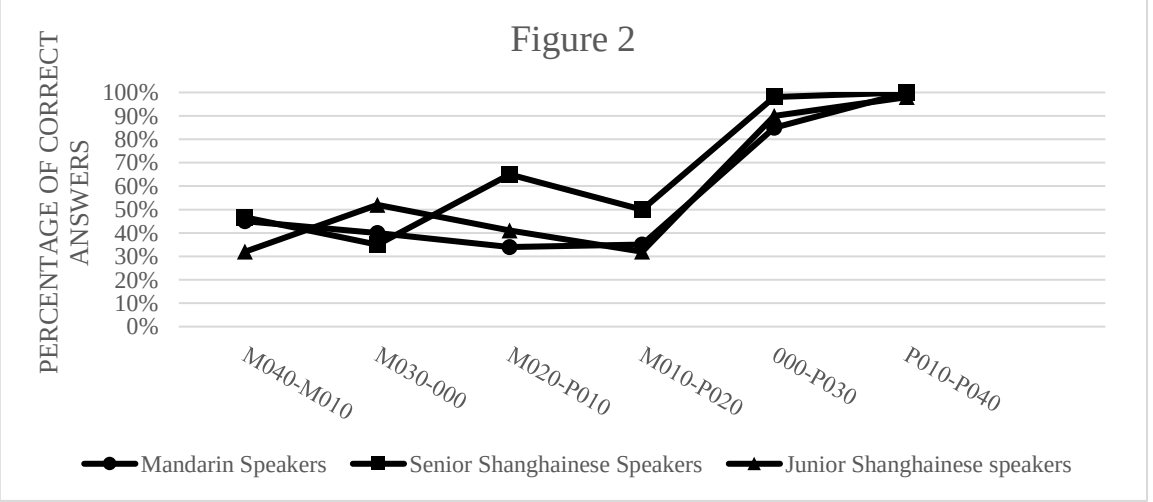

Figure 2: Results from the second part of the experiment.

DOI: 10.54254/2753-7064/17/20230774

The results of the first part of the experiment are shown in Figure 1. As can be seen, there are two peaks in the figure: one around M020-M010, the other around P020-P030 ("M" for the full name "LABM"; "P" for the full name "LABP", and henceforth). The peak around B020-B010 is formed only by Senior Shanghainese speakers, and neither Mandarin Speakers nor Junior Shanghainese Speakers demonstrated any discernible performance of identifying the stimuli in the neighborhood of M020 as "different from the previous one". It is also interesting to note that although there seem to be two peaks, they are not identical, the first one (up to 100% of participants) being significantly lower than the second one (only about 20% of participants).

The results of the second part of the experiment are shown in Figure 2. As shown in the figure, there is only one discernible peak, namely the one around LAB000-P030 and P010-P040 that is about 100% of correct answers. There is a less discernible peak around triads containing M020 formed by Senior Shanghainese Speakers, which is around 65% of correct answers. Mandarin Speakers and Junior Shanghainese Speakers, however, showed no clear peak of well-performance within this range.

6. Conclusion

There are three places in the results above that are contrary to the research hypothesis: the lower and less prominent peak in Figure 1, the absence of any clear peak formed by Mandarin Speakers and Junior Shanghainese Speakers around triads containing M020 in Figure 2, and the fact that the Junior Shanghainese Speakers seem to have performed exactly like Mandarin speakers.

The first two problematic phenomena can simply be explained by saying that the synthesized voice was less natural and more unfamiliar for the subjects, thus resulting in them not being able to show their normal level of performance. However, a more likely possibility could be that senior Shanghainese speakers rely more on other acoustic cues instead of plain VOT differences. These factors can be said to include tonal differences, phonation types, H1-H2, and so on [11]. In fact, this is exactly the case, as is stated in the introduction, and the deviance of the results from the hypothesis shows that these are unignorable factors. Therefore, senior Shanghainese speakers can only successfully discriminate /b/ and /p/ in normal speech and will perform poorly when presented with isolated speech sounds.

The third phenomenon can be used to forward the assertion that the acquisition of Shanghainese as a second language only has a limited effect on one's prior internal phonology, given that the results show no statistically significant difference between Mandarin Speakers and Junior Shanghainese Speakers.

The experiment also has some shortcomings. First and most importantly, it failed to include experiments involving the discrimination in real words, rather than mere isolated sounds. Second, although the experiment did not consider gender differences when selecting subjects, gender may in fact play an important role in terms of sound realization. We will leave these questions to further research.

References

[1] *Leyang, X. (2016). Phonological Comparison between Mandarin and Shanghainese.*
[2] *Chen, Y., & Gussenhoven, C. (2015). Shanghai Chinese. Journal of the International Phonetic Association, 45(3), 321-337.*
[3] *Wei, Z. (2022). Cue reweighting in Shanghainese sandhi patterns. ExLing 2022, 193.*
[4] *Tian, J., & Kuang, J. (2021). The phonetic properties of the non-modal phonation in Shanghainese. Journal of the International Phonetic Association, 51(2), 202-228.*
[5] *Ringbom, H. (1987). The role of the first language in foreign language learning. (No Title).*
[6] *Major, R. C. (2008). Transfer in second language phonology. Phonology and second language acquisition, 36, 63-94.*

DOI: 10.54254/2753-7064/17/20230774

[7] *Gui, T., & Zhou, Y. (2021). A survey of Shanghainese dialect: its current situation and future. Journal of Student Research, 10(2).*

[8] *Miyawaki, K., Jenkins, J. J., Strange, W., Liberman, A. M., Verbrugge, R., & Fujimura, O. (1975). An effect of linguistic experience: The discrimination of [r] and [l] by native speakers of Japanese and English. Perception & Psychophysics, 18(5), 331-340.*

[9] *Liu, Z. (2016). Exploring cross-linguistic influence: Perception and production of L1, L2 and L3 bilabial stops by Mandarin Chinese speakers.*

[10] *Liberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. Journal of experimental psychology, 54(5), 358.*

[11] *Chai, Y., & Garellek, M. (2022). On H1–H2 as an acoustic measure of linguistic phonation type. The Journal of the Acoustical Society of America, 152(3), 1856-1870.*