On the paths from voicing contrast to tonal contrast
Dissertation Proposal

Jia Tian

Submitted on May 10, 2019

Dissertation Supervisor: Jianjing Kuang
Proposal Committee: Eugene Buckley (chair), Mark Liberman, Don Ringe

Contents

1 Introduction

Most languages have obstruent voicing contrasts (Maddieson, 1984; Ladefoged & Maddieson, 1996), i.e., they have phonemic contrasts between classes of obstruents which differ in the laryngeal setting. That is, they differ in the mode of action of the larynx, or in the timing of laryngeal activity in relation to the oral articulation. Arguably, voicing is one of the most important linguistic features of human languages.

The voicing contrast shows considerable variation both within and across languages. Different sound changes related to the voicing contrast have been attested. In what follows I propose to examine the production and perception of the obstruent voicing contrast in contemporary Shanghainese.

Shanghainese is traditionally described as having a voiced vs. voiceless voicing contrast in obstruents. Word-initially, consonantal voicing cues have been lost to some extent. Phonation and f0 differences exist. Word-medially, however, consonantal voicing cues still play the most important role. Essentially, Shanghainese occupies the middle ground between, on the one hand, languages such as English in which there is only voicing contrast but no tonal contrast, and on the other hand, languages such as Mandarin in which there is no voicing contrast but only tonal contrast. I will show that the system in Shanghainese is evolving further towards a purely tonal system. The study of Shanghainese will provide important insights to questions related to tonogenesis, to the development path of the voicing contrast, and to speakers' phonological knowledge of the voicing contrast during sound change.

1.1 The voicing contrast

1.1.1 Cross-linguistic variation

Cross-linguistically, there is considerable variation in the phonological representation and the phonetic realization of the voicing contrast. First, languages can differ at the phonological level. Two-way voicing contrasts such as those used in English (Lisker & Abramson, 1964; Keating, 1984; Lisker, 1986), German (Jessen, 1998; Kleber, 2018), French (Caramazza & Yeni-Komshian, 1974; Kirby & Ladd, 2016), and Spanish (Williams, 1977; Schmidt & Flege, 1996) are most commonly found in the world's languages (Maddieson, 1984). Three-way voicing contrasts are also common, for example in Korean (Lisker & Abramson, 1964; Cho et al., 2002) and Thai (Lisker & Abramson, 1964; Abramson, 1989). Four-way voicing contrasts are rare cross-linguistically, but common in Indic languages such as Hindi (Lisker & Abramson, 1964; Benguerel & Bhatia, 1980; Dixit, 1989), Urdu (Hussain, 2018) and Marathi (Lisker & Abramson, 1964; Berkson, 2013, 2019). Five-way and six-way voicing contrasts are extremely rare, but also attested (e.g., Sindhi and Siraiki have five-way contrasts (Hussain, 2018); Owerri Igbo has a six-way contrast (Ladefoged et al., 1976)).

Second, sometimes even though two languages are thought to have the same phonological contrast, they may differ at the phonetic level, either because different phonetic categories (e.g., {voiced}, {voiceless unaspirated}, or {voiceless aspirated}) are chosen to realize the same phonological contrast or because different phonetic realizations are adopted to realize the same phonetic category. (Following Keating (1984), curly brackets "{}" are used to refer to phonetic categories as opposed to phonological features [+/-voice].) For example, although both English and Spanish have a two-way voicing contrast which may be phonologically represented as [+voice] vs. [-voice] (Keating, 1984; Kingston & Diehl, 1994; Wetzels & Mascaró, 2001), they differ significantly in the phonetic categories chosen to realize the two voicing categories. On the one hand, the phonologically voiced stops in Spanish are prevoiced (i.e., there is vocal fold vibration during stop closure; there is a substantial voicing lead), while the phonologically voiceless stops are produced with a short voicing lag. Therefore, Spanish is classified as a "true voicing" language which contrasts {voiced} stops with {voiceless unaspirated} ones. On the other hand, the phonologically voiceless stops in English are aspirated, while the phonologically voiced stops show short voicing lag in phrase-initial position. Therefore, English is said to be an "aspirating" language which contrasts {voiceless aspirated} stops with {voiceless unaspirated} ones. (Though see Iverson & Salmons (1995); Jessen & Roux (2002); J. Beckman et al. (2013) for the suggestion that

"true voicing" and "aspirating" languages differ at the phonological level: "true voicing" languages distinguish marked voiced stops ([voice]) from unmarked voiceless unaspirated stops ([ ], a blank space indicating the absence of a phonological specification), whereas an "aspirating" language distinguishes marked aspirated stops ([spread glottis]) from unmarked often passively voiced stops ([ ]).) It was originally thought that languages with two-way voicing contrasts are either "true voicing" or "aspirating" (Iverson & Salmons, 1995). In other words, no language with a two-way voicing contrast contrasts prevoiced stops with voiceless aspirated ones. However, it has been shown that some varieties of Scottish English (Catford, 2001; Watt & Yurkova, 2007) and Swedish (Karlsson et al., 2004; Helgason & Ringen, 2008) contrast prevoiced stops with aspirated ones in utterance-initial position.

Moreover, for any given phonetic category, a wide range of phonetic realizations exists. For example, a cross-linguistic survey of Voice Onset Time (VOT, defined as "the relative timing of events at the glottis and at the place of oral occlusion" (Lisker & Abramson, 1964)) in 18 languages showed that mean VOT of voiceless aspirated velar stops ranged from 73 ms to 154 ms in the languages being investigated (Cho & Ladefoged, 1999).

Cases above are languages whose voicing contrasts can be captured by the timing of laryngeal activity in relation to the oral articulation. Some languages, however, have voicing contrasts in which VOT does not play a role in making the distinction. No languages have more than three VOT distinctions (Cho & Ladefoged, 1999). Therefore, languages with more than three voicing categories, such as Hindi or Owerri Igbo, must use some other action of the larynx to make the additional contrasts. Hindi has a four-way voicing contrast. In addition to voiced, voiceless unaspirated, and voiceless aspirated categories which are well differentiated in VOT, a fourth category, namely the voiced aspirated category, is not distinguished from the voiced category by VOT (Lisker & Abramson, 1964; Benguerel & Bhatia, 1980; Dixit, 1989). Rather, voiced and voiced aspirated stops are distinguished by voice quality (phonation type). Here the phonemic contrast of the two consonant categories are phonetically manifested on the following vowels. Vowels following voiced aspirated stops show breathier phonation.

Even for languages that have a two-way or three-way voicing contrast, VOT may not play a role in making the distinction. For example, Swiss German has a native two-way voicing contrast in stops (there is a third category, namely aspirated, that occurs mainly in loanwords) which is not signaled by VOT – both types are unaspirated – but by closure duration (Ladd & Schmid, 2018). Another example is Korean. Korean has a three-way voicing contrast which cannot be fully captured by VOT alone, at least in Accentual Phrase

(AP)-initial position (Lisker & Abramson, 1964; Cho et al., 2002). The three-way contrast is reflected in other phonetic parameters, such as F0, H1-H2, acoustic burst energy, intraoral pressure and airflow (Cho et al., 2002).

To sum up, the voicing contrast, arguably one of the most important features of human languages, shows very complicated variations in both phonological representation and phonetic realization in the world's languages.

1.1.2 Contextual variation

In addition to cross-linguistic variation, within the same language, the phonetic realization of voicing categories is highly affected by phonetic contexts such as prosodic conditions and adjacent sounds. For example, speech sounds are generally articulated more strongly in stronger prosodic positions (lexically stressed or marked with phrasal accent, or at the beginning of a prosodic domain) (M. E. Beckman & Edwards, 1994; de Jong, 1995; Fougeron & Keating, 1997; Cho & Keating, 2001; Keating et al., 2003; Cho & McQueen, 2005; Keating, 2006; Cho & Keating, 2009). With regard to different voicing categories, this means that they are typically produced with stronger contact between the articulators, longer gestural durations, and less coarticulation, though not necessarily simultaneously. Acoustically, there is longer segment duration, longer closure duration, increased VOT, and increased burst amplitude (Fougeron & Keating, 1997; Turk & White, 1999; Cho & Keating, 2001; Cho & McQueen, 2005; Cole et al., 2007; Cho & Keating, 2009; Kuzla & Ernestus, 2011; Davidson, 2016, 2018). To give another example, obstruents after vowels, approximants, and nasals are more likely to be phonetically voiced compared to those after obstruents (Davidson, 2016, 2018).

Importantly, some of the acoustic characteristics affected by phonetic contexts are important cues to phonological contrasts. Sometimes different voicing categories are affected differently in different phonetic contexts. As a result, the same phonological contrast shows different phonetic realizations in different phonetic contexts. For instance, the voiced stops in American English are mostly voiceless unaspirated phrase-initially, while they are often voiced in non-initial positions. The voiceless stops are mostly aspirated in both positions. Therefore, the contrast between voiced and voiceless stops in American English is realized as a contrast between voiceless unaspirated and voiceless aspirated stops phrase-initially, while the same phonological contrast is realized as a contrast between voiceless aspirated and voiced stops in non-initial positions (Lisker & Abramson, 1964; Keating, 1984; Lisker, 1986; Davidson, 2016, 2018). Similarly, the three-way voicing contrast (fortis, lenis, aspirated)

in Seoul Korean is realized differently depending on the sound's position in an accentual phrase. In traditional descriptions, aspirated stops are always aspirated, and fortis stops are always voiceless unaspirated, but lenis stops are slightly aspirated AP-initially and fully voiced AP-medially (Cho & Keating, 2001; Jun, 2006). (Seoul Korean is currently undergoing a tonogenesis-like sound change. Younger speakers produce the contrast differently. See Kang (2014); Bang et al. (2018) among others for details of this change.)

To sum up, within the same language, the voicing contrast shows contextual variation in different phonetic contexts.

1.1.3 Related sound changes

The voicing contrast is complicated: it differs from language to language in phonological representation and phonetic realization, and shows contextual variation within the same language. As a result, the sound change of voicing contrast can develop different paths both in different languages, and within the same language, depending on the phonetic context.

Some changes happen at the phonological level, and the voicing contrast is lost. If the voicing contrast is replaced by other contrasts, the lexical contrast remains intact despite the loss of the voicing distinction. This process is called cue shifting or transphonologization. Tonal contrast originated in onset voicing contrast is the most well-known example of it. If the original voicing contrast is not replaced by other contrasts, the language loses its lexical contrast. This process is called merger. Sometimes there is no change at the phonological level – the voicing contrast is maintained – but its phonetic realization changes. This process can be called boundary shift or chain shift. All these changes have been attested. I will describe each of these three changes in more detail in the following sections.

1.1.3.1 Cue shifting

The best-known cue shifting is the development of contrastive tones (or "tonogenesis", Matisoff, 1973) from the loss of the voicing distinction between syllable-initial voiced and voiceless obstruents. Relatively higher tones develop after voiceless onsets, and relatively lower tones develop after voiced ones. Tonogenesis caused by onset voicing contrast is widely attested in East and Southeast Asian languages (Edkins, 1853; Maspero, 1912; Karlgren, 1926; Haudricourt, 1954, 1961, 1965; Matisoff, 1973; Maran, 1973; Mazaudon, 1977; Abramson & Luangthongkum, 2009; Hyslop, 2009; Kingston, 2011; Kang, 2014; Brunelle & Kirby, 2016; Bang et al., 2018, among many others). It is also observed in, for example, Afrikaans, a West Germanic language (Coetzee et al., 2018).

Although there is little disagreement about the claim that onset voicing contrast causes tonal split, there is considerable controversy about how it happens exactly. According to Haudricourt (Haudricourt, 1954, 1961, 1965) and Matisoff (1973), voiced and voiceless onsets merge and cause confusion first. It is the confusion that triggers the split in the tonal system to "protect its contrasts". Ohala (1981) criticized such claims by saying that it is difficult to defend such claims even on logical grounds. If speakers have such control over the way their pronunciation changes, then why did they let the loss of the original contrast happen in the first place? Later models suggest that the secondary cue becomes primary before the loss of the original cue. For example, according to Hyman (1976), tonogenesis involves three steps: (1) intrinsic-phonetic, (2) extrinsic-phonological, and (3) distinctive-phonemic. In stage one, voiced and voiceless consonants determine the f0 perturbations on following vowel as the result of coarticulation. In this stage the f0 perturbations are unplanned byproducts of voicing states; however, the f0 perturbations are indeed perceptible. In stage two, f0 perturbations are exaggerated to such an extent that it cannot be entirely predicted on the basis of the universal phonetic effect of a preceding consonant. In this stage, pitch replaces the voicing distinction as the primary cue. In stage three, the consonant voicing distinction is lost. Hyman (1976) called the process from stage 1 to 2 "phonologization", and the process from stage 2 to 3 "phonemicization". Similar to Hyman (1976), Maran (1973) proposed that f0 distinction starts to be exaggerated before consonant distinction starts to be lost. Maran (1973) differs from Hyman (1976) in that his model contains a period in which consonant voicing distinction is in trading relationship with f0 contrast. Hyman's model, however, does not have a such period. Previous studies on tonogenesis have found both patterns (Silva, 2006; Kang & Han, 2013; Kang, 2014; Bang et al., 2018; Coetzee et al., 2018). In particular, studies even found both patterns in one single language (Seoul Korean).

There is also controversy whether consonant voicing contrast directly gives rise to tonal contrast. According to Haudricourt (1965), the loss of the onset voicing distinction evolve in either of two ways. In languages without tones, the onset voicing contrast leads to contrastive voice registers, in which voice quality plays the most important role, but differences in f0, vowel quality, and perhaps vowel length also often occur. On the other hand, in tonal languages, the onset voicing contrast leads to the split of the tonal system. Thurgood (Thurgood, 2002, 2007), however, suggests that in most, if not all cases, the shift from consonant voicing contrast to tonal contrast is mediated through a voice quality (phonation) stage in which voice quality plays the most important role. However, to my knowledge, no experimental study has ever observed that pitch replaces voice quality as the primary cue.

The domain of tonogenesis has rarely received any attention. Most studies simply discuss tonogenesis caused by the voicing contrast of syllable-initial obstruents as if tonogenesis happens to all syllables. Matisoff (1973) explicitly asserts that tonogenesis only happens in languages that have a monosyllabic structure ("to become truly tonal a language must have a basically monosyllabic structure (i.e. the morphemes must be only one syllable long)"). However, ongoing tonogenesis in Seoul Korean (Silva, 2006; Kang & Han, 2013; Kang, 2014; Bang et al., 2018), Afrikaans (Coetzee et al., 2018) and Kurtöp (Hyslop, 2009) demonstrate that the domain of tonogenesis can be larger than a syllable, e.g., AP in Seoul Korean and word in Afrikaans and Kurtöp. The fact that words in languages such as Chinese were monosyllabic at the time of tonogenesis have prevented the researchers from seeing other possibilities.

It remains unclear how non-initial voicing contrast which does not give rise to tonal contrast will evolve. Korean and Afrikaans are at a too early stage to provide an answer to this question. As we will see, the study of Shanghainese will show that non-initial voicing contrast which does not give rise to tonal contrast will be lost through merger.

1.1.3.2 Merger

The merger of laryngeal contrasts is a phenomenon widely attested across the world's languages. An ongoing merger of word-initial voiced and voiceless fricatives can be found in Dutch (Pinget, 2015). From several decades ago, word-initial voiced fricatives were produced more and more often as voiceless. This change originates in central and Northern part of the Netherlands, and is now in progress over the entire Dutch speaking area. The merger is almost complete in the Northern part of the Netherlands now, but West-Flanders still have the contrast. Dutch is also undergoing merger of word-initial voiced and voiceless stops. The devoicing of stops in word-initial position was recently found in Flanders but not in the Netherlands. This is an incipient sound change.

The merger of word-initial and word-medial voiced and voiceless stops has happened historically in High German dialects ("interior High German consonant weakening") (Zhirmunskiĭ, 1962). In the northern part of the Upper Saxon and the neighboring northeastern Thuringian dialects, the Proto-Germanic voiceless and voiced stops have merged completely both word-initially and word-medially, whereas in a much larger area merger happened only word-medially.

The merger of voicing contrast in word-final position is a common diachronic change widely attested in the world's languages (e.g., German, Polish, Turkish, Russian, to name a

few) (Iverson & Salmons, 2011; Hualde, 2011; Roettger et al., 2014). It is often called "final devoicing", though Iverson & Salmons (2011) pointed out that this term conflates a number of different processes – deletion and addition of laryngeal features – and that it is better to call this phenomenon "final laryngeal neutralization".

It has long been debated whether the neutralization of the final voicing distinction is complete in final devoicing. In traditional formal analysis, the voiced and voiceless categories are thought to be indistinguishable both in production and perception, but numerous experimental studies have argued that there are small acoustic and articulatory differences. Further studies suggest that listeners can distinguish between the two voicing categories with above-chance accuracy. However, the effect sizes found in both production and perception are often extremely small, so there have been a lot of criticisms on methodological grounds. It has been pointed out that the debate has become increasingly about methodology rather than the phenomenon per se (see Roettger et al., 2014, for a detailed review).

1.1.3.3 Boundary shift/chain shift

Boundary shift is common in vowel (e.g., the Great Vowel Shift, Wolfe, 1972), but it seems relatively rare in consonant. It has been found that English dialects vary in the phonetic realization of voiced and voiceless stops. For example, Catford (2001) noted that in varieties of English spoken in Scotland and the North of England, phrase-initial /p/, /t/, and /k/ are unaspirated. Watt & Yurkova (2007) found in Aberdeen English, a Scottish English dialect, that phrase-initial voiced stops are prevoiced, while voiceless stops are voiceless aspirated. These patterns differ from those in British and American English, in which phrase-initial voiced stops are voiceless unaspirated and voiceless stops are voiceless aspirated. Since these English dialects share the same ancestor, boundary shift must have happened. Boundary shift of voicing categories has also been found in Phay, Samre, and Khasi, three Mon-Khmer languages, where voiceless unaspirated stops became aspirated, while voiced stops became voiceless unaspirated (Haudricourt, 1965).

1.2 Listeners and speakers' roles in sound change

It is generally accepted that listeners play important roles in sound change. Ohala's model of sound change (Ohala, 1981, 1993) and the extensions thereof (e.g., Solé (2014)) proposed that the driving force of sound change is listeners' unintentional error. According to these proposals, sound change starts when the listener fails to compensate for the effects of contextual coarticulation or when they over compensate. Lindblom's hyper- and

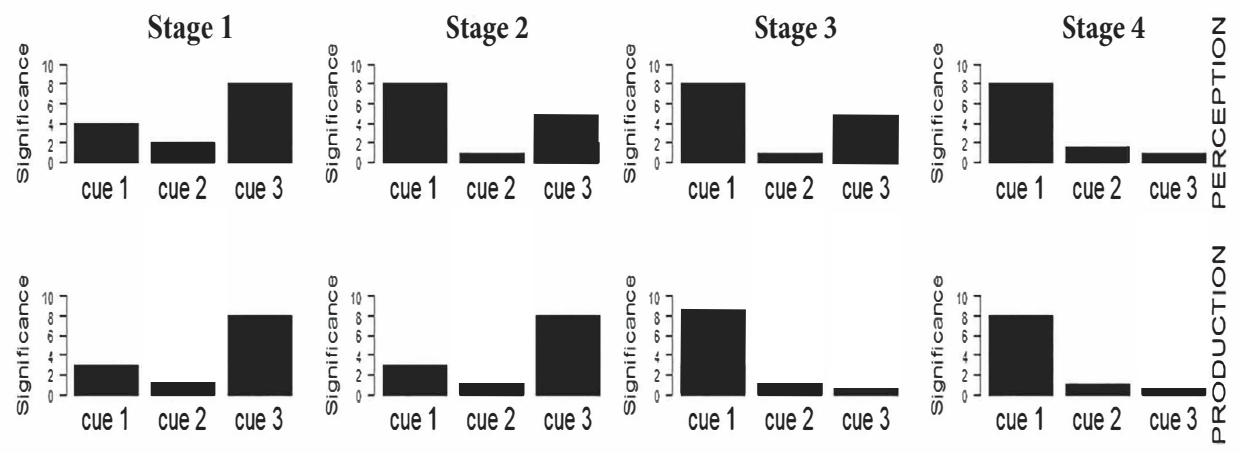
hypo-articulation (H&H) theory of speaker-listener interactions share with Ohala the view that unnormalized coarticulated variants can lead to sound change, but speakers' roles are emphasized, and listeners' misperceptions are de-emphasized (Lindblom, 1990; Lindblom et al., 1995). According to this theory, the ideal speaker makes a running estimate of the listener's informational needs and adjusts the production accordingly. When the listener's informational needs are estimated to be high, the speaker articulates hyper-forms, where the contrastive differences are more distinct. When the listener's informational needs are estimated to be low, the speaker articulates hypo-forms, where the contrastive differences are less distinct. Sound change happens when listeners correctly identify the lexical item when they hear the hypo-forms and then use the hypo-forms they hear in production. Beddor (2009) also attributes sound change to listeners, but she suggests that listeners do not make mistake. Instead, they simply attend to other cues because multiple interpretations are fully consistent with the coarticulated input. Some listeners simply place more weight on the effect than on the source of the coarticulation.

Recently, several studies have provided more empirical data on listener and speaker's role during sound change. Harrington (2012) found in /ʊ/- and /u/-fronting (vowel boundary shift) in Standard Southern British English that change occurs first in perception, i.e., speakers perceive new cue before they produce it. Pinget (2015) found in the merger of voiced and voiceless onsets in Dutch that merger occurs first in perception. Change in production starts later but reaches completion first. Coetzee et al. (2018) found in the final stage of Afrikaans tonogenesis that production and perception are generally aligned. When there is misalignment, production is in lead. Kuang & Cui (2018) found in a cue shifting in Southern Yi that new cues are used in perception first. Kuang & Cui (2018) hypothesize that production and perception have different mappings at different stages of sound change. At the very beginning of a sound change, perception initiates the change. When a sound change is close to completion, listeners are likely to retain sensitivity to the old cue for some time. They call for more empirical studies on different types of sound change to test their hypothesis. As one of the goals of the current study, I will examine the mapping between production and perception of Shanghainese speakers in different stages of the change to shed light on the relationship between production and perception during sound change.

The process of a cue shifting sound change according to the hypothesis in Kuang & Cui (2018) is illustrated in Figure 1. In Stage 1, before the change is initiated, cue 3 is the primary cue in both production and perception. In Stage 2, change in perception starts first, and the primary cue in perception has shifted to cue 1. Though cue 3 is no longer the

primary cue in perception, listeners are still sensitive to it. There is no change in production. In Stage 3, change in production happens, and the primary cue in production has shifted to cue 1. Cue 3, the original cue, is lost. No change happens in perception. In this stage, the change has completed in production, but not yet in perception. In Stage 4, cue 3 is lost in perception. The change has completed in both production and perception.

Figure 1: Hypothetical mapping between production and perception during different stages of a cue shifting sound change.



1.3 Shanghainese

This study focuses on two ongoing sound changes of the voicing contrast in Shanghainese. Shanghainese contrasts voiced and voiceless obstruents. Currently, word-initial onset voicing contrast is giving rise to contrastive tones. Word-medial voiced and voiceless stops are merging.

1.3.1 From voicing contrast to tonal contrast in Chinese dialects

It is hypothesized and widely accepted that between Middle Chinese (500 A. D.) and the present-day Chinese dialects, the four tones in Middle Chinese (Ping "level", Shang "rising", Qu "departing", Ru "entering") split into lower and higher reflexes depending on voicing of the initial consonants (Haudricourt, 1954; Tsu-lin, 1970; Pulleyblank, 1984; Sagart, 1999; Kingston, 2011). Relatively lower tones develop on vowels following voiced consonants, and relatively higher tones develop on vowels following voiceless ones. After the split, the voicing contrast is often completely lost, i.e., there is no distinction between voiced and voiceless onsets.

1.3.2 Two ongoing sound changes of the voicing contrast in Shanghainese

While the historical voiced vs. voiceless voicing contrast is lost after the tonal split in other Chinese dialects, it is still systematically retained in the Wu dialects of Chinese. Wu is the second to the largest group of Chinese dialects (Mandarin is the largest group). According to a 2017 census, the Wu dialects are spoken by a population of about 80 million (Eberhard et al., 2019). Wu dialects are still in the intermediate stage in the development of tone. The tonal contrast has appeared, but the voicing contrast on the onsets has not yet disappeared. The tonal inventory of Songjiang, a Wu dialect spoken in the suburb of Shanghai, is given in Table 1. The eight tones in Songjiang are organized by two voicing categories/tonal registers and four contours, corresponding to two voicing categories and four tones in Middle Chinese.

Table 1: Tone inventory of Songjiang, spoken in the suburb of Shanghai (Bao, 1999). The transcriptions adopt the five-scale pitch system developed by Chao (1930). This system divides speakers pitch range into five scales with 5 indicating the highest end and 1 the lowest.

		Contour			
Voicing	Register	Ping	Shang	Qu	Ru
Voiceless	Upper	T1 (53)	T3 (44)	T5 (35)	T7 (5)
Voiced	Lower	T2 (31)	T4 (22)	T6 (13)	T8 (3)

The focus of this study is Urban Shanghainese, the variety of Shanghainese spoken in the urban districts of Shanghai, the most thoroughly studied Wu dialect. Moreover, it is the most advanced in various sound changes (e.g., vowel merger, tone merger, tone sandhi development (Chao, 1928)) compared to other Wu dialects. In what follows, I refer to this language simply as "Shanghainese". The tonal system of Shanghainese is given in Table 2. The system of Shanghainese is very similar to that of Songjiang, the difference being some tones have merged in Shanghainese. Of the five tones of Shanghainese, T1, T2 and T4, the three upper register tones, have phonologically voiceless consonants as onsets, and start within the relatively higher f0 range. T3 and T5, the two lower register tones, have phonologically voiced consonants as onsets, and start within the relatively lower f0 range.

The voicing contrast in Shanghainese is currently undergoing two sound changes. Word-initially, onset voicing is giving rise to contrastive tones. This change started in the remote past, but has not completed yet. Word-medially, voiced and voiceless stops are merging. This

Table 2: Tone inventory of Shanghainese. The transcriptions in the parentheses are based on Xu & Tang (1988), a classic description of Shanghainese, which adopts the five-scale pitch system developed by Chao (1930). This system divides speakers pitch range into five scales with 5 indicating the highest end and 1 the lowest.

		Contour			
Voicing	Register	Ping	Shang	Qu	Ru
Voiceless	Upper	T1 (53)		T2 (34)	T4 (55)
Voiced	Lower			T3 (23)	T5 (12)

change started fairly recently and has never been reported in the literature. The processes involved are shown in Table 3.

Table 3: Stages of sound change in Shanghainese. Each cell shows whether there is difference between the two voicing categories for a given cue.

	Stage I		Stage II		Stage III	
	Initial	Medial	Initial	Medial	Initial	Medial
Consonant voicing difference	Yes		Depending on utterance position and manner	Yes	No	
Phonation difference	Minor		Yes	No	No	
Pitch difference	Minor		Yes	Yes	Yes	No

1.3.2.1 Stage I

In Stage I, consonant voicing difference is the most important cue to the contrast. Presumably phonation and pitch difference also exist, but they are due to unintentional onset perturbation, and the effect is minor. Although this stage existed so far in the past that it has never been directly observed, it is certain that a stage where there was only consonant voicing contrast once existed, according to historical documents (rhyme books and rhyme tables). However, it should be noted that the actual phonetic nature of the voiced category is obscure. The voiced stops might be prevoiced. It is also possible that they are slightly aspirated, just as the lenis stop in Korean. Other possibilities also exist. For fricatives, it is very likely that voiced fricatives show more voicing proportion (v%) and

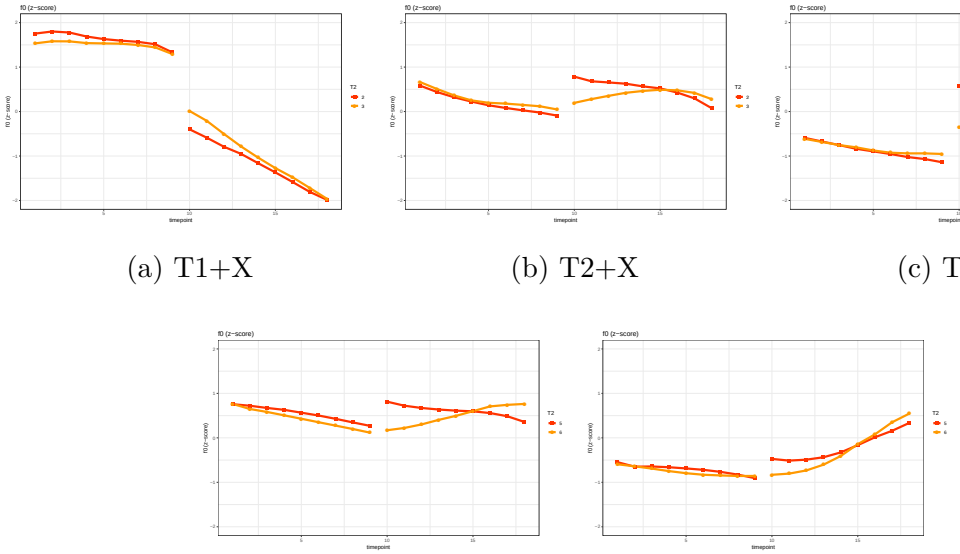
shorter duration than voiceless ones, a pattern we still see in current Shanghainese speakers. It should also be noted that it is unclear whether the language was monosyllabic at that time. In other words, it is unclear whether word-medial onsets existed in Stage I.

1.3.2.2 Stage II

The phonetic realization of the historical voicing contrast has received much attention since the 1850s. Studies reveal a different pattern from that in Stage I, which I call Stage II here. Many Shanghainese speakers, especially the relatively older ones, are still in this stage. Before I describe Stage II in detail, I need to give more information about a crucial concept in Shanghainese (and other Northern Wu dialects as well, Kuang et al., 2018): the phonological word. In Shanghainese, phonological words usually contain 1-3 syllables. 4-7 syllables are also possible, but the longer the domain, the rarer the phonological word. The formation of phonological words in Shanghainese is sensitive to both prosodic organization and morphosyntactic structures (Zee & Maddieson, 1979; Xu & Tang, 1988; Selkirk & Shen, 1990; Zee, 2004; Zhang & Meng, 2016). Monosyllabic modifier + monosyllabic noun structure obligatorily forms a phonological word, while monosyllabic verb + monosyllabic noun structure almost never forms a phonological word. When the components are disyllabic or trisyllabic, the pattern is more complicated, but almost always predictable (for detailed documentation, see Zee, 2004). In Shanghainese, when the phonological word is longer than one syllable, its tonal pattern is largely determined by the tone of its first syllable. This process is called tone sandhi. The tone sandhi pattern for disyllabic words in Shanghainese is given in (1). Note that in these rules, the tonal pattern of the disyllabic word is the same regardless of what the second syllable is. However, this is an oversimplification. As shown in Figure 2, f0 of the whole word is largely determined by the tone of the first syllable. However, voicing of the second syllable also plays important roles. When the first syllable is higher in pitch than the second syllable (figure (a)), the second syllable with voiced onsets has higher pitch than the second syllable with voiceless onsets (orange line is higher than the red line on the second syllable, at least in the first half of the syllable). However, when the first syllable is lower in pitch than the second syllable (figure (b) to (e)), the trend is reversed: the second syllable with voiceless onsets is higher in pitch. In what follows, I refer to the phonological word in Shanghainese simply as "word".

I should also describe the syllable structure in Shanghainese. Except for syllabic nasals, syllables in Shanghainese are (C)V(X). Each syllable contains a vowel. The syllable can start either with a consonant or directly with the vowel. Some syllables are open. Some

Figure 2: f0 (z-scored) of disyllabic words produced by four Shanghainese speakers born in the 1940s. In each subgraph, left lines are the first syllable, and right lines are the second syllable. Red lines when X = voiceless; Orange lines when X = voiced.



(a) T1+X (b) T2+X (c) T3+X

(d) T4+X (e) T5+X

syllables are closed by a nasal or glottal stop coda. Therefore, when I talk about word-medial obstruents in disyllabic words, the obstruents are the onsets of the second syllable.

(1) Shanghainese tone sandhi rules. X represents any tone.

1. T1+X: 53+X → 55-31
2. T2+X: 34+X → 33-44
3. T3+X: 23+X → 22-44
4. T4+X: 55+X → 33-44
5. T5+X: 12+X → 11-13

In stage II, the intermediate stage, word-medially, both stops and fricatives show difference in consonantal cues. The voiced category shows more proportion of voicing and shorter duration. There is no phonation difference word-medially. f0 of the following vowels differs, as we have seen in Figure 2.

Word-initially, vowels after voiced stops show breathier phonation and lower pitch (Chao, 1928; Cao & Maddieson, 1992; Ren, 1992; Wang, 2012; Tian & Kuang, 2016; Gao & Hallé, 2017). Consonant voicing cues show complicated patterns word-initially. Their realization depends on the position of the sound in an utterance, and the manner of the sound.

For stops, utterance-initially, there is no VOT distinction – both categories are phonetically voiceless unaspirated – and the historical voicing contrast is realized as phonation and pitch differences on the following vowels. Many studies attempted to analyze the breathier phonation after utterance-initial word-initial voiced stops as a consonantal property (Cao & Maddieson, 1992; Ren, 1992, among many others), because otherwise there is no consonantal cue in this position. However, this approach is criticized by a number of studies which considered the breathier phonation a property of the phonation/tonal register (Shen et al., 1987; Yip, 1993; Zhu, 1995).

Utterance-medially, some impressionistic observations suggest that both voiced and voiceless stops are phonetically voiceless, just as utterance-initial ones (Shen et al., 1987; Cao & Maddieson, 1992). However, these impressionistic observations are extremely doubtful because voicing is phonetically natural in intersonorant positions. A later study with much younger speakers indeed found that utterance-medial word-initial voiced stops (and also fricatives) show more voicing than voiceless ones (Zhang & Yan, 2018). In addition to proportion of voicing, Shen et al. (1987) also reported longer closure duration for utterance-medial word-initial voiced stops, although these results are not replicated in later studies with younger speakers (Wang, 2012; Zhang & Yan, 2018). There are two possibilities for the discrepancy. First, all of them are right; the language has changed. Second, Shen was wrong; there was never closure duration difference. This will be tested in the current study.

Fricatives have received very little attention in early studies. Two more recent studies (Gao & Hallé, 2017; Zhang & Yan, 2018) found that younger speakers produce shorter frication duration and higher proportion of voicing for word-initial voiced fricatives both utterance-initially and utterance-medially. Presumably older speakers also did no. In other words, word-initial fricatives show consonantal voicing differences both utterance-initially and utterance-medially in Stage II, while stops show consonantal voicing differences only utterance-medially. Therefore, fricatives are more conservative than stops in that they retain more voicing cues than stops utterance-initially.

1.3.2.3 Stage III

In stage III, the voicing contrast further shifts to tonal contrast. Word-initially, previous studies have shown that speakers born in the 1980s and later no longer produce phonation difference for both stops and fricatives (Gao, 2016; Tian & Kuang, 2016; Zhang & Yan, 2018). I also noticed that some younger speakers devoice word-initial voiced fricatives. Word-medially, I noticed that younger speakers are merging the voicing contrast in stops.

Fricatives do not seem to merge. These changes happen fairly recently. My impressionistic observations should be validated by experimental data.

1.3.3 The phonological representation of the historical voicing contrast in Shanghainese

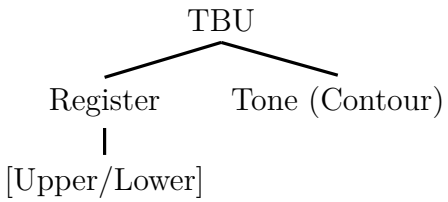
In Stage I where there is only consonant voicing distinction (there must be some phonation and f0 differences, but these differences are unintended perturbations from onset voicing), the contrast is undoubtedly a consonant voicing contrast. In Stage III where there is only tonal contrast, the contrast is undoubtedly not a consonant voicing contrast. In Stage II, however, consonant voicing, phonation, and f0 distinctions cooccur making it less obvious how the contrast should be analyzed. Many efforts have been devoted to the underlying phonological representation of this contrast in Stage II. In what follows I review important claims.

1.3.3.1 Yip (Underlyingly phonation contrast)

Yip proposed thatthe historical voicing contrast in Shanghainese is underlyingly a phonation contrast in Stage II. Onset voicing and pitch height are derived from the underlying phonation contrast.

Yip (1980) proposed a general model for tone representation. Her model is given in (2). This is the first time the feature Register is introduced in the representation of tone. According to this model, each tone bearing unit (TBU) is defined by two features: Register and Tone. Register determines the overall pitch height. Upper register tones are higher in pitch than the lower register tones. Note that the upper and lower registers have also been referred to as [+upper] and [-upper] in the literature. Throughout this proposal, I will use "upper" and "lower" for the sake of convenience. The feature tone determines the contour of the TBU.

(2) Yip (1980)



In the analysis of Shanghainese, she noticed that there is a complicated relation among voicing, phonation, and pitch, which is shown in Table 4. For obstruents, once one of the three is known, the other two can be predicted. However, this corresponding relation does

not hold for nasals and liquids. All nasals and liquids are voiced, so it is impossible to predict phonation or pitch from voicing. However, it is still possible to predict voicing from phonation or pitch: however the phonation or pitch is, they are always voiced. The one to one mapping between phonation and pitch holds for all syllables, including nor only nasals and liquids, but also zeros, the onsetless syllables. It should be noted that it is unclear why nasals and liquids show phonation and pitch differences in Shanghainese, despite the fact that nasals and liquids did not contrast in voicing in Middle Chinese. Most nasals and liquids are breathier in phonation and lower in pitch. Modal ones are rare but do exist. In fact, in many other Wu dialects, all nasals and liquids are breathier in phonation, and lower in pitch Chao (1928). Therefore, it is very likely that at some point, some nasals and liquids became higher register.

Table 4: The relation among voicing, phonation, and pitch for different onsets.

	Obstruent		Nasal and liquid		Zero (Onsetless)	
Voicing	Voiced	Voiceless	Voiced		-	
Phonation	Breathy	Modal	Breathy	Modal	Breathy	Modal
Pitch	Lower	Higher	Lower	Higher	Lower	Higher

Given the one to one mapping between phonation and pitch in all Shanghainese syllables, and the fact that pitch is not always predictable from voicing, Yip (1980) took the position that "murmur (i.e., breathier phonation) and [-upper] (i.e., lower register) are one and the same" (see also Yip, 1993). Yip (1980) did not explicitly say whether obstructs in Shanghainese contrast underlying in voicing; she probably wouldn't.

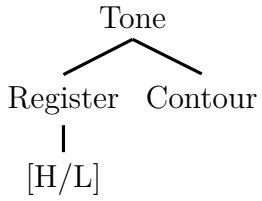
Yip (1993) provided a modified analysis of Shanghainese. It is suggested that Shanghainese has the feature [murmur]. This feature causes obstruent voicing ([murmur] $\rightarrow$ [voice]), so there is no need for a distinctive feature [voice] underlyingly in Shanghainese. This feature also conditions pitch height. To my knowledge, Yip (1993) is the only study which suggests that there is no need for a distinctive feature [voice] underlyingly in Shanghainese.

1.3.3.2 Bao and Duanmu (Underlyingly onset voicing contrast)

Bao (1999) proposed that the historical voicing contrast in Shanghainese is underlyingly a voicing contrast in Stage II, and that pitch height is derived from the underlying voicing contrast. He did not mention phonation, but presumably he would think that phonation is also derived from voicing.

The general model for tone representation proposed by Bao (1999) is given in (3). This model is the same as that proposed in Yip (1980). But when applying this model to Songjiang, a suburban variety of Shanghainese that has basically the same tonal system as Urban Shanghainese (cf. Table 1 and Table 2), Bao suggested that for this language, the register node needs not be specified, because they are predictable from the syllable-initial obstruents. Unlike Yip, Bao did not mention nasals or liquids.

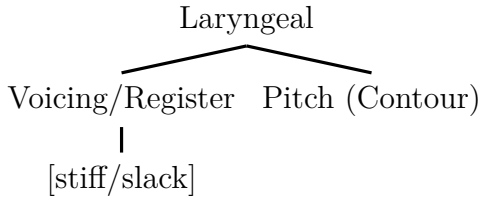
(3) Bao (1999)



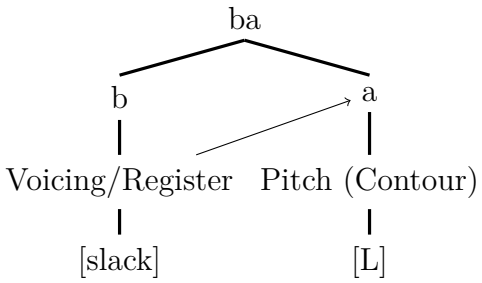
Similar to Bao, Duanmu (1990) also proposed that the historical voicing contrast in Shanghainese is underlyingly an onset voicing contrast in Stage II. However, details slightly differ.

The tonal model proposed by Duanmu (1990) is given in (4). In this model, the V/R (Voicing/Register) node represents both consonant voicing and tonal register. The Pitch node represents contour. Underlyingly, onsets are specified for V/R, but unspecified for Pitch. Rime segments (vowel+coda) are unspecified for V/R, but specified for Pitch. When onset and rime combine to form a syllable, the V/R node of the onset is spread to the nuclear vowel, and probably further to the coda. An example of /ba/ in Shanghainese is given in (5) (Duanmu used /zoŋ/ (p. 129), but I show /ba/ here for the sake of convenience).

(4) Duanmu (1990)



(5) The derivation of /ba/



In the analysis of Shanghainese, Duanmu suggests that tone split happens in initial and isolated syllables. Voiceless onsets give rise to the upper register, and voiced onset give rise to the lower register. Register affects both phonation and f0. Then the onset voicing contrast is neutralized. In non-initial syllables, the onset voicing contrast is maintained, and there is no tone split.

1.3.3.3 A summary

Different phonological representations have been proposed to analyze the historical voicing contrast in Shanghainese in Stage II. Although opinions differ, all analyses share the view that the system in Shanghainese is definitely not purely tonal. Instead, it occupies the middle ground between, on the one hand, languages such as English in which there is only voicing contrast but no tonal contrast, and on the other hand, languages such as Mandarin in which there is no voicing contrast but only tonal contrast.

In this dissertation, I will show that the system in Shanghainese is moving further towards a purely tonal one.

1.4 Research questions

This study aims at investigating the phonological representation of the historical voicing contrast in contemporary Shanghainese. Note that throughout this proposal, I refer to the historical (Middle Chinese) voicing contrast in Shanghainese as "voicing contrast", "phonological voicing contrast", or "historical voicing contrast", and the historically voiced/voiceless obstruents in Shanghainese as "voiced/voiceless", "phonologically voiced/voiceless", or "historically voiced/voiceless". These designations are used for the sake of convenience and are not intended to indicate that the contrast should necessarily be analyzed as a voicing contrast in contemporary Shanghainese. Nor do these designations indicate that the historical contrast has the same underlying representation for all Shanghainese speakers at this point.

Specifically, I ask: how is the historical voicing contrast represented in contemporary Shanghainese speakers' mind? Does the phonological representation change over time? In the following sections, I break these question down, and propose different experiments and analyses (production, perception, relation between production and perception) to answer them.

Shanghainese provides us with an important case study to better understand tonogenesis, the development path of the voicing contrast, and speakers' phonological knowledge of the voicing contrast during sound change.

2 The production of the historical voicing contrast in Shanghainese

2.1 Questions and hypotheses

In this section, I propose to test the production of the historical voicing contrast in different phonetic environments of different generations of Shanghainese speakers to investigate the phonological representation of the historical voicing contrast in contemporary Shanghainese. Specifically, I ask:

1. In production, is there any evidence that the historical voicing contrast is still represented as a consonant voicing contrast in Shanghainese?

    (a) Is there any difference in consonant voicing cues in any of the following phonetic environments?

        i. word position, i.e., the position of the target obstruent in the word (word-initial vs. word-medial)

        ii. utterance position, i.e., the position of the target word in the utterance (utterance-initial, utterance-medial, utterance-final)

        iii. manner of articulation of the target obstruent (stop, fricative)

    (b) Does the phonetic realization of the historical voicing contrast differ among the phonetic environments mentioned above?

    (c) Do speakers of different age and gender differ?

    For older speakers, I expect to replicate previous findings that word-initially, whether there is difference in consonant voicing cues depends on manner and utterance position.

I have summarized the pattern in Table 3 Stage II and described the pattern in detail in Section 1.3.2.2.

I expect that in all phonetic environments, younger speakers show less difference in consonantal voicing cues. Female speakers might be more advanced in this change, i.e., they produce less difference in consonant voicing cues than male speakers.

It is generally accepted that phonological contrasts are enhanced in prosodically stronger positions (Hsu & Jun, 1998; Cho & Jun, 2000; Cho, 2005; Cho & McQueen, 2005; Cole et al., 2007). I speculate that enhanced contrasts are more likely to be preserved. Following this idea, I should expect more differences in consonantal voicing cues in the utterance-initial position than non-initial positions. This might be true for word-medial obstruents and word-initial fricatives. However, existing results for word-initial stops have already found that difference in consonant voicing cues is better preserved in non-utterance-initial positions.

It has been found that fricatives retain more difference in consonant voicing cues than stops in the utterance-initial word-initial position, I expect that fricatives might also retain more consonant voicing cues in other positions.

2. Given that consonantal voicing cues are less and less important in the historical voicing contrast, are difference in other cues exaggerated?

   (a) Is there any difference in suprasegmental cues (phonation, f0) in any of the following phonetic environments (the same phonetic environments as I test for consonantal voicing cues)?

      i. word position, i.e., the position of the target obstruent in the word (word-initial vs. word-medial)

      ii. utterance position, i.e., the position of the target word in the utterance (utterance-initial, utterance-medial, utterance-final)

      iii. manner of articulation of the target obstruent (stop, fricative)

   (b) Does the phonetic realization of the historical voicing contrast differ among the phonetic environments mentioned above?

   (c) Do speakers of different age and gender differ?

   For older speakers, I expect the pattern summarized in Table 3 Stage II. Word-initially, both f0 and phonation differences exist. Word-medially, there is only f0 difference but no phonation difference.

I expect that younger speakers show less phonation difference than older speakers word-initially, but their f0 difference remains robust. Younger speakers show less f0 difference word-medially. Female speakers might be more advanced in the change.

Utterance-initial position may show larger f0 and phonation differences.

Manner do not seem to affect f0 or phonation.

3. How do cues interplay during tonogenesis in production?

   (a) What is the relative importance of different cues (consonantal, phonation, f0) in production in different stages of tonogenesis? Do stops and fricatives differ?

   I expect that older speakers are in Stage II. Both phonation and f0 contribute to the contrast. Consonantal voicing cues also contribute to the contrast, except for utterance-initial stops. I have no basis to predict the relative importance of difference cues. The relative importance of consonantal voicing cues and phonation cues gradually declines, and younger speakers rely on f0 only (Stage III).

   I really hope that the oldest speakers I recruit use phonation or consonantal cues as the primary cue. If so, I will be able to understand how f0 becomes primary during tonogenesis. However, it is very likely that the shift has long happened, and we are not able to observe it.

   (b) Is there trading relation between cues in production during tonogenesis?

   A number of sound change models have given different prediction. Some suggest that tonogenesis includes a stage of cue trading (Maran, 1973; Beddor, 2009), while other models suggest that it is only when the original secondary cue is fully developed that the original primary cue starts to be lost (Hyman, 1976). Previous studies on tonogenesis and other cue shifting changes have found both patterns (Silva, 2006; Kang & Han, 2013; Kang, 2014; Bang et al., 2018; Coetzee et al., 2018; Kuang & Cui, 2018). In particular, studies even found both patterns in one single language (Seoul Korean). Therefore, I leave the prediction open.

4. What is the relative importance of different cues (consonantal, phonation, f0) in production in different stages of the merger?

   Phonation should play no role. I have no basis to predict whether consonantal cues or f0 play the most important role before the merger. During the merger, the importance of consonantal cues and f0 gradually declines.

2.2 Participants

All experiments proposed in this study will be conducted in Shanghai, China. The tentative plan is to recruit 70 native Shanghainese speakers in total, with 5 male and 5 female speakers in each of the 10-year bins starting from 1940 to 2000. The tentative plan is illustrated in Table 5. If it turns out that it is too difficult to find enough speakers that are evenly distributed in age and gender, I will recruit 60 speakers (15 older females, 15 older males, 15 younger females, 15 younger males) instead. Speakers born before 1990 are classified as older speakers. The alternative plan is given in Table 6. All experiments (both production and perception) will be conducted with the same speakers.

Table 5: Planned participants.

	1940s	1950s	1960s	1970s	1980s	1990s	2000s	Total
Female	5	5	5	5	5	5	5	35
Male	5	5	5	5	5	5	5	35

Table 6: The alternative plan of participant recruitment.

	Older	Younger	Total
Female	15	15	30
Male	15	15	30

2.3 Stimuli

All participants will be asked to participate in two production experiments, a sentence reading experiment and a map task. Most of the previous studies on Shanghainese are done using isolated monosyllables. This study differs from them in two important ways: 1) connected speech and 2) disyllabic target words are used.

First, this is because connected speech is more natural and better reflects how speakers use the language in their daily life. Second, while researchers generally use monosyllabic words to study word-initial obstruents, monosyllabic words are word-final, utterance-initial, and utterance-final at the same time. Therefore, it is impossible to guarantee that the effect is indeed due to the sound's being word-initial. By using disyllabic words in different utterance positions, I will be able to tease these effects apart. By designing sentences in

which the factors that I am interested in are systematically manipulated, I will be able to quickly collect data that contain full combination of all factors that I am interested in. However, speakers' performance is generally less natural in reading task. Therefore, I will also collect data from a map task to elicit more natural speech.

2.3.1 Sentence reading

In the sentence reading experiment, speakers will read 150 sentences, 100 of which for word-initial obstruents, and 100 of which for word-medial obstruents. 50 sentences are used for both word-initial and word-medial obstruents. The design of the target words is given in Table 7. All syllables are open. All following and preceding vowels are low. All words are high frequency.

In the 100 sentences that contain word-initial obstruents, the first syllable of the word bears T2 or T3, and cover stop (/p/, /b/, /t/, /d/, /k/, /g/) and fricative (/f/, /v/, /s/, /z/). Each sound has 10 different syllables (10 /p/, 10 /b/, 10 /t/ etc). The second syllable of the word is always T2, and starts with /p/, /t/, /k/, /f/, /s/. There does not have to be mapping relation between the manner of articulation of the two syllables.

In the 100 sentences that contain word-medial obstruents, the second syllable of the word bears T2 or T3, and cover the sounds mentioned above. The first syllable of the word is always T3, and starts with /b/, /d/, /g/, /v/, /z/. There does not have to be mapping relation between the manner of articulation of the two syllables.

Of the 150 sentences, 50 that contain T3 as the first syllable and T2 as the second syllable are used both for word-initial and word-medial obstruents.

In each sentence, the target word occurs three times, once sentence-initially, once sentence-medially, and once sentence-finally. Sentence corresponds to utterance.

Table 7: Stimuli in the sentence reading task.

	Word-initial		Word-medial	
	1st syllable	2nd syllable	1st syllable	2nd syllable
Tone (#)	T2/T3 (50/50)	T2 (100)	T3 (100)	T2/T3 (50/50)

2.3.2 Map task

Recall that in the sentence reading experiment, I use 10 syllables for each sound, which gives rise to 150 words in 150 sentences. In the map task, I choose one of the 10 syllables,

and ask participants to use the resulting 15 words in the map task.

During the map task, speakers will be given a map which contains the target word. They will be asked to describe what they see on the map to me. I will interact with them and try to elicit more tokens.

2.4 Recording procedures

The sentences will be printed on a piece of paper and be given to the participants. When the participant is ready, recordings will be made in a quiet room (or a sound booth, if I have access to one). The recordings will be made using OpenSesame. Each sentence will be displayed on the monitor, and speakers will be asked to read the sentences aloud. When the speakers make mistakes (judged by either the speaker or me), they are allowed/asked to read the sentence again until they read it correctly. Each sentence will be read once.

2.5 Measures

Praat textgrids will be created using a forced aligner. Then I will adjust boundaries by hand. Consonantal cues, phonation, and f0 of the target syllables will be measured. The measures are described in the following sections. Measurements will be within-speaker z-scored.

2.5.1 Consonantal measures

Different consonantal measures will be taken depending on the manner of the target sound, the word and the sentence position. Consonantal measures are summarized in Table 8. For sentence-initial word-initial stops, closure duration (cd) cannot be measured, because it is impossible to determine the beginning of the closure when there is no preceding sound. VOT will be measured in this position. For stops in other positions, VOT cannot be measured, and cd will be measured. Following Davidson (2016, 2018), the offset of the second formant of the preceding vowel will be used to determine the left edge of the closure boundaries. The onset of the burst will be used as the right edge of the closure, except where there is no visible burst, in which case the onset of F2 of the following vowel will be taken as the edge of the sound. For fricatives, fricative duration (fd) will be measured. In the sentence-initial word-initial position, the onset of frication will be used to determine the left edge of the fricative boundaries, while in other positions, the offset of the second formant of the preceding vowel will be used to determine the left edge of the boundaries, the same as what I do for stops.

The right edge of the fricatives is taken to be the onset of F2 of the following vowel. While this segmenting method could potentially increase the proportion of tokens that are coded as having some voicing during the closure or frication, it was adopted because I do not want to erroneously exclude tokens that do contain voicing.

Closure and fricative duration will be extracted using a Praat script (Boersma & Weenink, 2018). In addition to the duration of the obstruent interval, the script also measures the proportion of voicing (v%) in the interval. The v% measure will be obtained by using the "fraction of locally unvoiced frames" in Praat's Voice Report. Praat's defaults will be used, and the pitch settings will be optimized for voice analysis as described in Eager (2015).

Table 8: Consonantal Measures.

	Stop		Fricative	
	Sentence-initial	Sentence-medial	Sentence-initial	Sentence-medial
Word-initial	VOT	cd, v%	fd, v%	fd, v%
Word-medial	cd, v%	cd, v%	fd, v%	fd, v%

2.5.2 Suprasegmental measures

2.5.2.1 Phonation

Phonation will be measured in VoiceSauce (Shue et al., 2011). Phonation measures include three spectral measures (the relative amplitude difference between the fundamental and the most prominent harmonics around the first three formants, i.e., H1*-A1*, H1*-A2*, and H1*-A3*) and one noise measure (Cepstral Peak Prominence, CPP, Hillenbrand et al. (1994)). The spectral measures are thought to correlate with different aspects of glottal constriction. It is proposed that H1*-A1* is related to posterior glottal opening at the arytenoids (Hanson et al., 2001), and H1*-A2* and H1*-A3* are correlated with the abruptness of vocal fold closure (Stevens, 1977; Holmberg et al., 1995; Hanson et al., 2001; Cho et al., 2002). The relatively less constricted glottis during breathier phonation results in a glottal waveform with greater low-frequency and weaker high-frequency components, which can be quantified by higher values of these three measures. These measures have be found to be reliable cues for the phonation contrast in many of the world's languages (Garellek & Keating, 2011; Esposito, 2012; Khan, 2012; Kuang & Keating, 2014; Kelterer, 2017), and importantly, in Shanghainese (Cao & Maddieson, 1992; Ren, 1992; Tian & Kuang, 2016;

Gao, 2016; Gao & Hallé, 2017). H1*-H2*, the amplitude difference between the first and the second harmonic, will not be included because it has been shown to be not useful in Shanghainese and other related Wu dialects such as Jiashan (Tian & Kuang, 2016; Jiang & Kuang, 2016). All spectral measures are corrected for formant frequencies, using the correction algorithm in Iseli et al. (2007). Here and elsewhere in the literature, corrected measures are written with asterisks.

Cepstral Peak Prominence (CPP) is a measure of vocal aperiodicity and spectral noise levels. The less constricted glottis in breathier voice increases the chance for glottal air turbulence, resulting in lower CPP values. CPP is useful for distinguishing phonation contrasts in many languages, including Jalapa Mazatec (Garellek & Keating, 2011), Southern Yi (Kuang, 2011), While Hmong (Esposito, 2012), Gujarati (Khan, 2012), and importantly, Shanghainese (Tian & Kuang, 2016; Gao, 2016; Gao & Hallé, 2017).

Following Tian & Kuang (2016), all phonation measurements will be taken in the first third interval of the following vowel, because this is where the phonation distinction is the strongest.

2.5.2.2 f0

f0 will be measured in VoiceSauce using the Straight algorithm, using VoiceSauce's default settings (Kawahara et al., 2009). Like phonation measures, f0 will be taken in the first third interval of the following vowel.

2.6 Statistical modeling

In cases where more than one measures are taken (consonantal cues except when sentence-initial word-initial stops are measured, phonation cues), after the within-speaker z-score normalization, a principal component analysis (PCA) will be run to combine multiple cues. The first principal component (PC1) will be used to represent the consonantal/phonation cues. For example, two consonantal measures (fd and v%) were taken on sentence-initial word-initial fricatives. Therefore, I will run PCA on these two measures to obtain the PC1, and this PC1 is used to represent the consonantal cues for fricatives in this position. Likewise, because there are four phonation measures, a PCA will be run to extract the PC1 for these phonation cues, and the extracted PC1 will be used to represent the phonation cues. As a result, I start with multiple consonantal and phonation cues, but end up having only three cues: a PC1 for consonantal cues (PC1_C), a PC1 for phonation cues (PC1_P), and f0. These three measures are subject to statistical modeling.

Linear mixed-effects models will be used to evaluate whether the two voicing categories differ, and whether there is age and gender effect. Mixed-effects logistic regression models will be used to evaluate the relative importance of different cues.

2.7 Results

In this section I show preliminary findings on stops.

2.7.1 Word-initial stops

2.7.1.1 Materials

The data presented here are from recordings I collected in Shanghai in the summer of 2016. A total of 107 speakers were recorded. Their information is given in Table 9. During the recording, speakers were asked to read many words, sentences, and passages. The analyses in this section are based on read isolated monosyllabic words. Speakers were asked to read three pairs of words: /pa 34/ vs. /ba 23/, /ta 34/ vs. /da 23/, and /ka 34/ vs /ga 23/. All words bear a rising tone. Each word was read once.

Phonation and pitch were measured using the method described in section 2.5. Phonation measures include three spectral measures (H1*-A1*, H1*-A2*, H1*-A3*) and a noise measure (CPP). Since phonation is correlated with multiple measures, a principal component analysis was run on the phonation measures. The first principal component (PC1) accounted for 65% of the variance, and was used to represent the phonation cues. I did not measure VOT because it is well established that VOT does not play any role in isolated monosyllables. I will nevertheless measure VOT in the dissertation. Voicing proportion (v%) and closure duration (CD) cannot be measured because there is no preceding syllable.

Table 9: Information of the 107 speakers.

	1930s	1940s	1950s	1960s	1970s	1980s	1990s	2000s	Total
Female	2	5	8	10	7	8	8	5	53
Male	0	5	5	6	4	6	19	9	54

2.7.1.2 Results

As shown in Fig 3, younger speakers show smaller phonation difference between voiced and voiceless stops. Results of linear mixed-effects model show that for the youngest female

Figure 3: Phonation cues of word-initial stops.

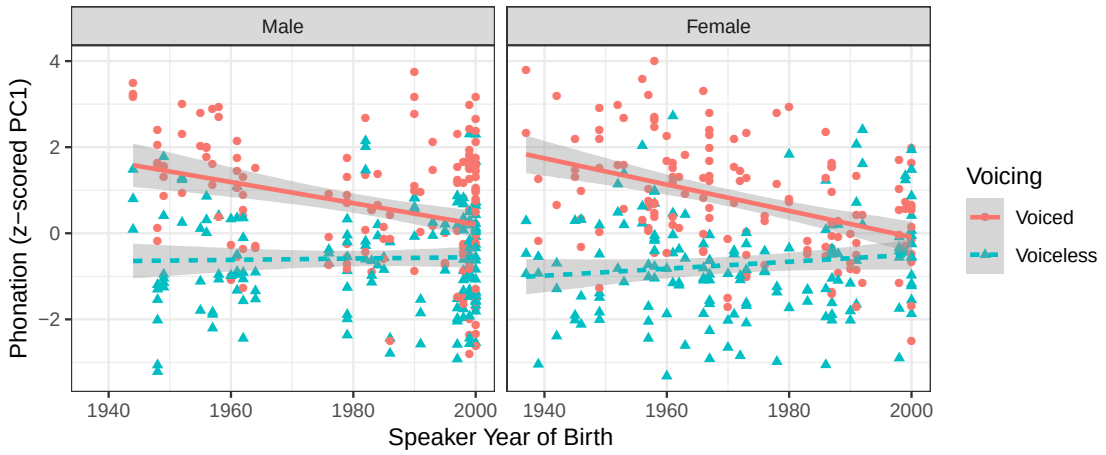
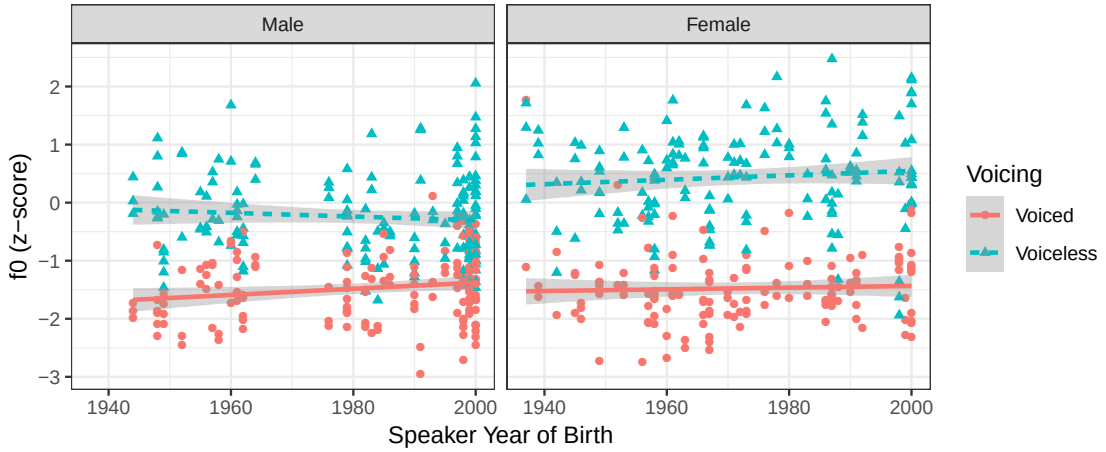


Figure 4: f0 of word-initial stops.



speakers (those born in the year 2000), there is no phonation difference between the two voicing categories (t = -1.631, p > .05). There is a significant year of birth by voicing interaction, indicating that older female speakers show larger phonation distinctions between the two voicing categories (t = -4.299, p < .05). Moreover, there is no gender interaction, indicating that younger male speakers have also lost the phonation contrast completely.

Fig 3 shows that for all speakers, there is robust f0 distinction between the two voicing categories. Younger speakers do not further expand f0 contrast, nor do they show smaller f0 distinction. These observations are confirmed by the results of the statistical model. For the youngest female speakers, voiceless stops show higher f0 (t = 15.901, p < .05). There is no year of birth by voicing interaction, indicating that older and younger female speakers do not differ in the f0 difference between the two voicing categories. There is a voicing by gender

interaction (t = -5.891, p < .05), indicating that male speakers show smaller f0 differences. However, this is probably due to the fact that female speakers show larger f0 range.

Logistic regression models were run on older (born before 1990) and younger (born in and after 1990) speakers to examine whether phonation or f0 is used as the primary cue to the contrast. Results show that for all four groups (older female, older male, younger female, younger male), f0 is the most important cue. Phonation plays secondary roles for older speakers. Younger speakers do not use phonation.

2.7.2 Word-medial stops

2.7.2.1 Materials

The analyses in this section are also based on the recordings I made in Shanghai in the summer of 2016. However, only 40 speakers are examined in this section due to limited time. The information of these 40 speakers are given in Table 10. The analyses in this section are based on read passages. Speakers typically spent 15 minutes reading these passages. The recordings are forced aligned by the Montreal Forced Aligner[1]. The target sounds are word-medial stops with a preceding T3 open syllable. All words are utterance-initial disyllabic words. Voicing proportion (v%) and closure duration (cd) of the stops are measured in Praat (Boersma & Weenink, 2018). Phonation and f0 of the following vowels are measured in VoiceSauce. All measures are within-speaker z-scored. Phonation and f0 are measured on the first third interval of the following vowel. Since there are two consonantal cues, a principal component analysis was run to extract PC1. PC1 accounted for 76% of the variance in consonantal cues. Principal component analysis was also run on phonation measures to extract PC1. PC1 accounted for 43% of the variance in phonation.

Table 10: Information of the 40 speakers examined in this section.

	1930s	1940s	1950s	1960s	1970s	1980s	1990s	2000s	Total
Female	1	2	2	2	3	0	6	4	20
Male	0	2	4	3	1	0	8	2	20

2.7.2.2 Results

Unfortunately, all mixed effects models in the word-medial position show singular fit, so I can only eyeball the figures. This problem is probably because the data are too few. It

[1]https://montreal-forced-aligner.readthedocs.io/en/latest/

Figure 5: Consonantal cues of word-medial stops.

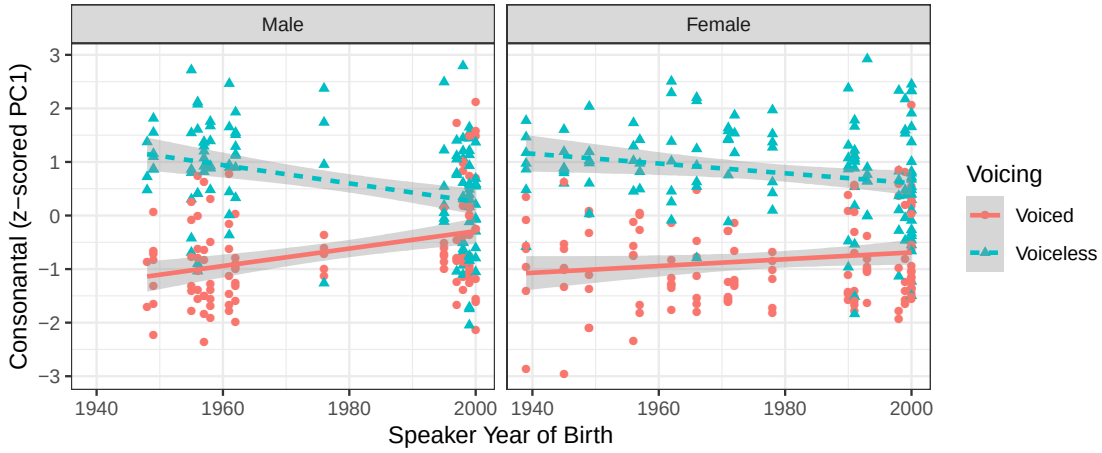
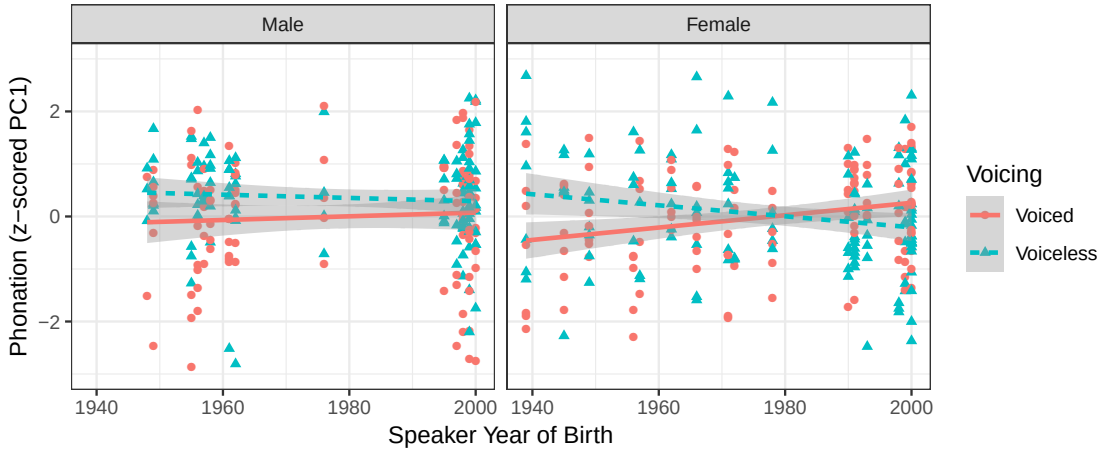


Figure 6: Phonation cues of word-medial stops.



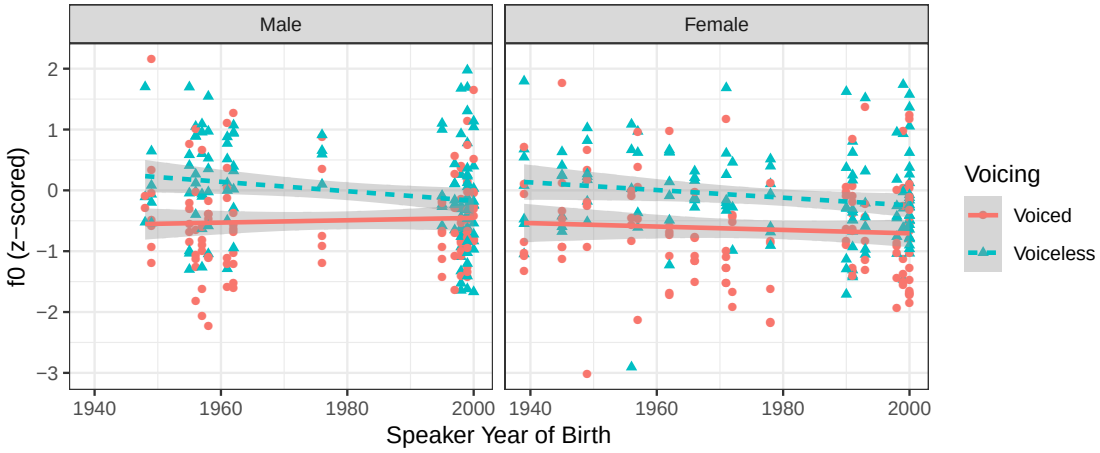
should be solved by collecting more data. Fig 5 shows that younger speakers show smaller consonantal differences between voiced and voiceless stops. Younger male speakers seem to be more advanced in this change.

Fig 6 shows that there is no phonation contrast for any speaker. This is compatible with previous findings that phonation contrast only exists in the word-initial position.

Fig 7 shows that there are some f0 differences in older speakers, but younger speakers do not have f0 difference.

Results of logistic regression models show that for older speakers (born before 1990), the most important cue is the consonantal cue. Phonation is useless. f0 plays secondary roles. Younger speakers show smaller consonantal contrast, and no cue is exaggerated, indicating that this is a merger. For younger male speakers, there is hardly any contrast. Younger

Figure 7: f0 of word-medial stops.



female speakers still show some differences in consonantal cues.

2.8 Discussion

Results from the production experiments show that consonantal voicing cues are less and less important in the historical voicing contrast in Shanghainese. For utterance-initial word-initial stops, consonantal voicing cues have long been replaced by f0 and completely lost. Older speakers use phonation contrast, but the phonation contrast is gradually lost by the younger generation. There is no f0 exaggeration accompanying the loss of phonation distinction, i.e., there is no trading relation between phonation and f0.

Consonantal voicing cues are also lost to a great extent word-medially, and no other cues emerge to maintain the original contrast. These results suggest that for languages that undergo tonogenesis, the non-initial voicing contrast that does not give rise to contrastive tones will be lost through merger.

I would like to emphasize that the merger in the word-medial position is particularly important. Previously, though the primary cue in the word-initial position has long shifted to f0, because there is still contrast word-medially, and consonant voicing plays the primary role, we could still say that the language still maintains its voicing contrast. But from this point on, no evidence in production could support the claim that Shanghainese still has consonant voicing contrast.

3 The perception of the historical voicing contrast in Shanghainese

3.1 Questions and hypotheses

The production results suggest that consonantal voicing cues are gradually lost in younger speakers' production. This raises the question whether the historical voicing contrast is still represented as a voicing contrast in Shanghainese speakers' mind. To answer this question, I need to test younger speakers' perception. As long as consonant voicing cues are the primary cue in perception, I cannot claim that the contrast has become purely tonal. This section aims at providing answers to the following questions:

1. In perception, is there any evidence that the historical voicing contrast is still represented as a consonant voicing contrast in Shanghainese?

   (a) Do speakers use difference in consonant voicing cues to identify the voicing categories in the following phonetic environments?

      i. word position, i.e., the position of the target obstruent in the word (word-initial vs. word-medial)

      ii. manner of articulation of the target obstruent (stop, fricative)

   (b) Do speakers of different age and gender differ?

   I expect that younger speakers will rely less and less on consonant voicing cues in perception. Female speakers may be more advanced in this change. Fricatives may be less advanced in this change than stops.

2. Is it possible that the historical voicing contrast is represented as a suprasegmental (phonation or f0) contrast in Shanghainese?

   (a) Do speakers use difference in suprasegmental cues (phonation, f0) to identify the voicing categories in the following phonetic environments?

      i. word position, i.e., the position of the target obstruent in the word (word-initial vs. word-medial)

      ii. manner of articulation of the target obstruent (stop, fricative)

   (b) Do speakers of different age and gender differ?

Word-initially, I expect that all speakers rely on f0 to perceive the historical voicing contrast. Phonation cues are expected to play some roles for older speakers, but they gradually lose importance in younger speakers' perception. The change is more advanced in female. The change is more advanced in stops than fricatives.

Word-medially, I expect that younger speakers are more and more likely to identify the voicing categories by chance. Phonation does not play any role for any speaker. f0 plays some roles for older speakers, but f0 gradually lose importance in the younger generation. The change is more advanced in female. The change is more advanced in stops than fricatives.

3. How do cues interplay during tonogenesis in perception?

   (a) What is the relative importance of different cues (consonantal, phonation, f0) in perception in different stages of tonogenesis?

   I expect that for older speakers who are in Stage II, both phonation and f0 contribute to the contrast in perception. Consonantal voicing cues also contribute to the contrast, and this might also be true for utterance-initial stops. For most speakers, if not all, f0 is the most important cue. The relative importance of consonantal voicing cues and phonation cues gradually declines, and younger speakers rely on f0 only (Stage III). I hope that I can find older speakers who use phonation or consonantal cues as the primary cue in perception. If so, I will be able to understand how secondary cue becomes primary during tonogenesis in perception. However, it is very likely that the shift has long occurred, and we are not able to observe it.

   (b) Is there trading relation between cues in perception in tonogenesis?

   Whether or not there is trading relation between cues in perception has received much less attention compared to that in production. I leave the prediction open.

4. What is the relative importance of different cues (consonantal, phonation, f0) in perception in different stages of the merger?

   I expect that before the merger, consonantal cues play the most important role. f0 plays a secondary role. Phonation does not play any role. During the merger, the importance of consonantal cues and f0 gradually declines, and are finally lost.

3.2 Participants

Participants in the perception experiment will be the same as the participants in the production experiment.

3.3 Stimuli

All participants will be asked to participate in two identification experiments. The first experiment uses natural stimuli, and tests whether the participants maintain the lexical contrast. The second experiment uses synthesized stimuli, and tests how participants make the distinction. All stimuli are isolated disyllabic words.

3.3.1 Natural stimuli

This experiment will use the naturally produced minimal pairs from a previous production experiment. All words are high frequency. Half of the minimal pairs differ only in the word-initial obstruent, e.g., /ba sa/ vs. /pa sa/. Another half of the minimal pairs differ in the word-medial obstruent, e.g., /sa ba/ vs. /sa pa/. Five pairs will be tested (/ba/~/pa/, /da/~/ta/, /ga/~/ka/, /ze/~/se/, /ve/~/fe/). Stimuli from 3 older female speakers and 3 older male speakers that maintain the lexical contrast both word-initially and word-medially will be used. There will be 120 tokens in total: 2 categories * 5 pairs * 2 position * 6 speakers.

3.3.2 Synthesized stimuli

3.3.2.1 Word-initial

The synthesized stimuli for word-initial will be based on minimal pairs of monosyllables and disyllabic words produced by an older male speaker. Five minimal pairs will be selected (that containing /ba/~/pa/, /da/~/ta/, /ga/~/ka/, /ze/~/se/, /ve/~/fe/). All words are high frequency. For each of pair, the duration of the first syllable will be normalized to the same value. The duration of the second syllable will also be normalized to the same value (a value different from the word-initial one).

3.3.2.1.1 Stop

For each pair of /ba/~/pa/, /da/~/ta/, and /ga/~/ka/, there will be 18 synthesized stimuli in total (3 VOT * 2 phonation * 3 f0).

For each pair (e.g., /ba/~/pa/), a minimal pair of words will be selected (e.g., /ba sa/ vs. /pa sa/). The tone of the second syllable is T2. Phonation will be measured to ensure that there is indeed phonation distinction. This gives two steps in phonation: modal vs. breathier. VOT will be manipulated to be varying in three steps: 10 ms, -70 ms, and -120 ms. -70 ms is a value common for slightly prevoiced stops, and -120 ms is a value common for heavily prevoiced stops (Kessinger & Blumstein, 1997). f0 will be manipulated to be varying in 3 steps. The construction of the f0 continua will be based on a previous study from which this male speaker's voice is chosen. There are 60 speakers in total (15 older female, 15 older male, 15 younger female, 15 younger male). I will calculate the natural stimuli's average pitch in semitone, expand the ends a little bit, and construct three steps.

3.3.2.1.2 Fricative

For each pair of /ze/~/se/ and /ve/~/fe/), there will be 54 resynthesized stimuli (3 fd * 3 v% * 2 phonation * 3 f0).

Word-initial fricative stimuli are created in similar manner as the stop stimuli. The original stimuli come from the same older male speaker. Two phonation types come from the original token. f0 steps and values are the same as stops. The manipulated consonantal cues are fricative duration and proportion of voicing. Fricative duration has 3 steps: short (100 ms), medium (130 ms), and long (160 ms). v% has 3 steps, 0%, 30%, and 60%. These values are based on naturally produced stimuli.

3.3.2.2 Word-medial

For both stop and fricative, there will be 54 resynthesized stimuli (3 fd/cd, 3 v% * 2 phonation * 3 f0).

Based on the natural stimuli, duration of the obstruent will have 3 steps: 60 ms, 90 ms, 120 ms. There will be 3 steps in v%: 0% (fully devoiced), 50% (partially voiced), and 100% (fully voiced). Again there will be 2 phonation types from naturally produced stimuli. Because speakers do not produce phonation contrast word-medially, word-initially syllables that have been used for the word-initial stimuli will be used again in the word-medial position. Preceding syllables will be taken from other words and be added. f0 of the initial syllable, and f0 at the end of the second syllable will be the same. There will be three f0 steps at the beginning of the second vowel.

3.4 Test procedures

Participants first take part in the experiment that uses natural stimuli. They then take part in the experiment that uses synthesized stimuli. All participants listen to all stimuli. The stimuli will be presented in random order in Praat. In the identification task, each stimulus will be presented in isolation. Participants can listen to the stimuli as many times as they want. The choices will be a minimal pair that differs only in the voicing of the target sound. The choices will be presented in Chinese character.

3.5 Statistical modeling

For each phonetic environment, a mixed-effect logistic regression model will be used to examine whether each cue is used, which cue is more important, and whether different speakers differ.

4 The mapping between production and perception

4.1 Questions and hypotheses

In this section, I will compare the relative cue weighting of different cues in production and perception. The goal is to better understand whether it is the listener or the speaker that leads sound change. Specifically, I ask:

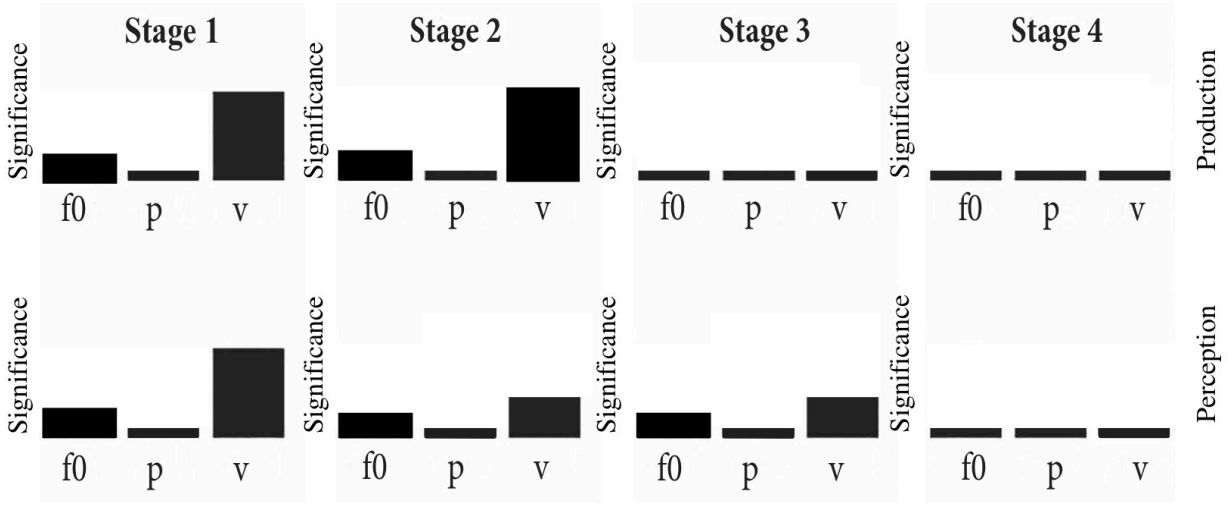
1. What is the mapping between the relative cue weighting in production and perception?
    (a) in different word positions
    (b) for different manner of articulation
    (c) for speakers of different age and gender

My prediction is that results will follow the hypothesis in Kuang & Cui (2018) that at the beginning of the sound change, perception changes first. Change in production begins later but completes first. Finally, the change completes in perception.

The specific prediction for word-medial obstruents is given in Figure 8. In Stage 1 (note that these four stages are not the same as Stage I, II, and III in previous sections), consonant voicing plays the most important role. f0 also plays some roles, but its contribution is secondary. Cue weighting in production and perception is aligned. Older speakers show

the pattern in Stage 1. In Stage 2, voicing becomes less important in perception, but there is no change in production. This is the pattern for slightly younger speakers. In Stage 3, both voicing and f0 become completely useless in production, but speakers' perception is still the same as Stage 2. Some even younger speakers should be in this Stage. Finally, in Stage 4, voicing and f0 are lost in perception. The youngest speakers might show the pattern in Stage 4.

Figure 8: Predicted mapping between production and perception during different stages of merger of word-medial obstruents. p = phonation. v = consonant voicing cues.



It is extremely hard to draw a figure like Figure 8 for word-initial obstruents, not because I do not have a prediction about the mapping between production and perception — I predict that perception lags behind in such a late stage sound change – but because it is unclear what the relative importance of different cues is in production.

5 Discussion

I have not answered my very first question: what is the phonological representation of the historical voicing contrast in Shanghainese? Does the phonological representation change over time? I cannot provide a definite answer to these questions at this point because I have not carried out all the proposed experiments yet. In what follows I discuss some possibilities.

At one end of the spectrum, if both production and perception data show that consonantal voicing cues do not play any role in any phonetic environment for a given speaker, I am safe to draw the conclusion that the historical voicing contrast has completely shifted to

tonal contrast for this speaker.

At the other end of the spectrum, for a given speaker, if in either production or perception, consonant voicing cues are the primary cue in any phonetic environment, then the shift from the historical voicing contrast to tonal contrast has not completed yet. The speaker must have some knowledge about the voicing contrast.

If consonantal voicing cues only play some very minor roles in some phonetic environments in either production or perception, and it is f0 that plays the most important role in these cases, I could still conclude that the historical voicing contrast has completely shifted to tonal contrast. The produced and/or perceived consonantal cues are due to unintentional perturbation from f0, which is universal and not language-specific, just as consonantal cues can perturb f0.

However, if in any phonetic environment either in production or perception, consonantal voicing cues are found to be important enough so that it cannot be explained by universal perturbation from f0, then the shift from the historical voicing contrast to tonal contrast has not completed yet. The speaker must still retain some knowledge about the voicing contrast.

My preliminary production results show that older speakers use consonant voicing as the primary cue in the word-medial position, so they must retain some knowledge about consonant voicing, and the shift from consonant voicing contrast to tonal contrast has not completed. The production results show that younger speakers do not produce difference in consonant voicing cues in this position, so they have moved further towards a purely tonal system. I am eager to examine other environments in production, and to investigate if younger speakers stop using consonant voicing cues in perception either to provide a more complete picture of the speakers' phonological knowledge of the historical voicing contrast.

6 Future work

While the preliminary production results in this proposal look promising, many questions are unanswered. Future work should carry out all proposed experiments to verify the preliminary findings, and to answer the unanswered questions.

I believe that this work will provide a better understanding of tonogenesis, and how speakers lose their phonological knowledge of the voicing contrast during sound change.

7 Planned structure of the dissertation

- Chapter 1 Introduction
- Chapter 2 The production of the historical voicing contrast in Shanghainese
- Chapter 3 The perception of the historical voicing contrast in Shanghainese
- Chapter 4 The mapping between production and perception
- Chapter 5 Discussion and conclusion

8 Time line

I plan to finalize the experiment design in June 2019. I will collect data in Shanghai from July to August in 2019. I will analyze the data in September to December in 2019. I will write one chapter each month from January to May in 2020. I plan to defend the dissertation in August 2020.

References

Abramson, A. S. (1989). Laryngeal control in the plosives of Standard Thai. Pasaa: Notes and news about language teaching and linguistics in Thailand, 19, 85–93.

Abramson, A. S., & Luangthongkum, T. (2009). A fuzzy boundary between tone languages and voice-register languages. In G. Fant, H. Fujisaki, & J. Shen (Eds.), Frontiers in phonetics and speech science (pp. 149–155). Beijing: The Commercial Press.

Bang, H.-Y., Sonderegger, M., Kang, Y., Clayards, M., & Yoon, T.-J. (2018). The emergence, progress, and impact of sound change in progress in Seoul Korean: Implications for mechanisms of tonogenesis. Journal of Phonetics, 66, 120–144.

Bao, Z. (1999). The Structure of Tone. New York: Oxford University Press.

Beckman, J., Jessen, M., & Ringen, C. (2013). Empirical evidence for laryngeal features: Aspirating vs. true voice languages. Journal of Linguistics, 49, 259–284.

Beckman, M. E., & Edwards, J. (1994). Articulatory evidence for differentiating stress categories. In Papers in laboratory phonology iii: Phonological structure and phonetic form (pp. 7–33). Cambridge: Cambridge University Press.

Beddor, P. S. (2009). A Coarticulatory Path to Sound Change. Language, 85(4), 785–821.

Benguerel, A.-P., & Bhatia, T. K. (1980). Hindi Stop Consonants: an Acoustic and Fiberscopic Study. Phonetica, 37, 134–148.

Berkson, K. H. (2013). Phonation Types in Marathi: An Acoustic Investigation (Unpublished doctoral dissertation).

Berkson, K. H. (2019). Acoustic correlates of breathy sonorants in Marathi. Journal of Phonetics, 73, 70–90.

Boersma, P., & Weenink, D. (2018). Praat: doing phonetics by computer.

Brunelle, M., & Kirby, J. (2016). Tone and Phonation in Southeast Asian Languages. Linguistics and Language Compass, 10(4), 191–207. doi: 10.1111/lnc3.12182

Cao, J., & Maddieson, I. (1992). An exploration of phonation types in Wu dialects of Chinese. Journal of Phonetics, 20, 77–92.

Caramazza, A., & Yeni-Komshian, G. H. (1974). Voice onset time in two French dialects. Journal of phonetics, 2(3), 239–245.

Catford, J. C. (2001). A Practical Introduction to Phonetics (Second ed.). Oxford University Press.

Chao, Y. R. (1928). Studies in the Modern Wu-Dialects [现代吴语的研究]. Peking: Tsing Hua College Research Institute.

Chao, Y. R. (1930). A system of tone-letters. Le Maître Phonétique, 30, 24–27.

Cho, T. (2005). Prosodic strengthening and featural enhancement: Evidence from acoustic and articulatory realizations of / ,i/ in English. The Journal of the Acoustical Society of America, 117(6), 3867–3878.

Cho, T., & Jun, S.-A. (2000). Domain-initial strengthening as enhancement of laryngeal features: Aerodynamic evidence from Korean. Chicago Linguistics Society, 36, 31–44.

Cho, T., Jun, S.-A., & Ladefoged, P. (2002). Acoustic and aerodynamic correlates of Korean stops and fricatives. Journal of Phonetics, 30, 193–228.

Cho, T., & Keating, P. (2001). Articulatory and acoustic studies on domain-initial strengthening in Korean. Journal of Phonetics, 29, 155–190.

Cho, T., & Keating, P. (2009). Effects of initial position versus prominence in English. Journal of Phonetics, 37, 466–485.

Cho, T., & Ladefoged, P. (1999). Variation and universals in VOT: Evidence from 18 languages. Journal of Phonetics, 27, 207–229.

Cho, T., & McQueen, J. M. (2005). Prosodic influences on consonant production in Dutch: Effects of prosodic boundaries, phrasal accent and lexical stress. Journal of Phonetics, 33, 121–157.

Coetzee, A. W., Beddor, P. S., Shedden, K., Styler, W., & Wissing, D. (2018). Plosive voicing in Afrikaans: Differential cue weighting and tonogenesis. Journal of Phonetics, 66, 185–216.

Cole, J., Kim, H., Choi, H., & Hasegawa-Johnson, M. (2007). Prosodic effects on acoustic cues to stop voicing and place of articulation: Evidence from Radio News speech. Journal of Phonetics, 35(2), 180–209.

Davidson, L. (2016). Variability in the implementation of voicing in American English obstruents. Journal of Phonetics, 54, 35–50.

Davidson, L. (2018). Phonation and laryngeal specification in American English voiceless obstruents. Journal of the International Phonetic Association, 48(3), 331–356.

de Jong, K. J. (1995). The supraglottal articulation of prominence in English: Linguistic stress as localized hyperarticulation. The Journal of the Acoustical Society of America, 97(1), 491–504.

Dixit, R. P. (1989). Glottal gestures in Hindi plosives. Journal of Phonetics, 17, 213–237.

Duanmu, S. (1990). A formal study of syllable, tone, stress and domain in Chinese languages (Unpublished doctoral dissertation). Massachusetts Institute of Technology.

Eager, C. D. (2015). Automated voicing analysis in Praat : statistically equivalent to manual segmentation. Proceedings of the 18th International Congress of Phonetic Sciences. doi: 10.1007/BF02382000

Eberhard, D. M., Simons, G. F., & Fennig, C. D. (Eds.). (2019). Ethnologue: Languages of the World (Twenty-sec ed.). Dallas, Texas: SIL International. Retrieved from http:// www.ethnologue.com

Edkins, J. (1853). A grammar of colloquial Chinese as exhibited in the Shanghai dialect. Shanghai: Presbyterian Mission Press.

Esposito, C. M. (2012). An acoustic and electroglottographic study of White Hmong tone and phonation. Journal of Phonetics, 40(3), 466–476.

Fougeron, C., & Keating, P. (1997). Articulatory strengthening at edges of prosodic domains. The Journal of the Acoustical Society of America, 101(6), 3728–3740.

Gao, J. (2016). Sociolinguistic motivations in sound change : on-going loss of low tone breathy voice in Shanghai Chinese. Papers in Historical Phonology, 1, 166–186.

Gao, J., & Hallé, P. (2017). Phonetic and phonological properties of tones in Shanghai Chinese. Cahiers de Linguistique Asie Orientale, 46, 1–31.

Garellek, M., & Keating, P. (2011). The acoustic consequences of phonation and tone interactions in Jalapa Mazatec. Journal of the International Phonetic Association, 41(2), 185–205.

Hanson, H. M., Stevens, K. N., Kuo, H.-K. J., Chen, M. Y., & Slifka, J. (2001). Towards models of phonation. Journal of Phonetics, 29, 451–480.

Harrington, J. (2012). The coarticulatory basis of diachronic high back vowel fronting. In M.-J. Solé & D. Recasens (Eds.), The initiation of sound change: Perception, production and social factors (pp. 103–122). John Benjamins Publishing Company.

Haudricourt, A.-G. (1954). De l'origine des tons en vietnamien (The origin of tones in Vietnamese). Journal Asiatique, 242, 69–82.

Haudricourt, A.-G. (1961). Bipartition et Tripartition des Systèmes de Tons dans quelques Langues d' Extrême-Orient (Two-way and Three-way Splitting of Tonal Systems in Some Far-Eastern Languages). Bulletin de la Société de Linguistique de Paris, 56(163-180).

Haudricourt, A.-G. (1965). Les mutations consonantiques des occlusives initiales en môn-khmer (Consonant shifts in Mon-Khmer initial stops). Bulletin de la Société de Linguistique de Paris, 60(1), 160–172.

Helgason, P., & Ringen, C. (2008). Voicing and aspiration in Swedish stops. Journal of Phonetics, 36, 607–628.

Hillenbrand, J., Cleveland, R. A., & Erickson, R. L. (1994). Acoustic Correlates of Breathy Vocal Quality. Journal of Speech and Hearing Research, 37, 769–778.

Holmberg, E. B., Hillman, R. E., Perkell, J. S., Guiod, P. C., & Goldman, S. L. (1995). Comparisons among aerodynamic, electroglottographic, and acoustic spectral measures of female voice. Journal of Speech and Hearing Research, 38, 1212–1223.

Hsu, C.-S. K., & Jun, S.-A. (1998). Prosodic Strengthening in Taiwanese: Syntagmatic or Paradigmatic? UCLA Working Papers in Phonetics, 96, 69–89.

Hualde, J. I. (2011). Sound Change. In The blackwell companion to phonology (pp. 1–22). Wiley-Blackwell.

Hussain, Q. (2018). A typological study of Voice Onset Time (VOT) in Indo-Iranian languages. Journal of Phonetics, 71, 284–305.

Hyman, L. M. (1976). Phonologization. In A. Juilland (Ed.), Linguistic studies offered to joseph greenberg on the occasion of his sixtieth birthday (pp. 407–418). Saratoga: Anma Libri.

Hyslop, G. (2009). Kurtöp Tone: A tonogenetic case study. Lingua, 119(6), 827–845.

Iseli, M., Shue, Y.-L., & Alwan, A. (2007). Age, sex, and vowel dependencies of acoustic measures related to the voice source. Journal of the Acoustical Society of America, 121(4), 2283–2295.

Iverson, G. K., & Salmons, J. C. (1995). Aspiration and laryngeal representation in Germanic. Phonology, 12, 369–396.

Iverson, G. K., & Salmons, J. C. (2011). Final Devoicing and Final Laryngeal Neutralization. In The blackwell companion to phonology (pp. 1–22).

Jessen, M. (1998). Phonetics and Phonology of Tense and Lax Obstruents in German. Amsterdam, Philadelphia: John Benjamins Publishing Company.

Jessen, M., & Roux, J. C. (2002). Voice quality differences associated with stops and clicks in Xhosa. Journal of Phonetics, 30, 1–52.

Jiang, B., & Kuang, J. (2016). Consonant effects on tonal registers in Jiashan Wu. Proceedings of the Linguistic Society of America, 1, 1–13.

Jun, S.-A. (2006). Prosodic Typology: The Phonology of Intonation and Phrasing (S.-A. Jun, Ed.). Oxford: Oxford University Press.

Kang, Y. (2014). Voice Onset Time merger and development of tonal contrast in Seoul Korean stops: A corpus study. Journal of Phonetics, 45, 76–90.

Kang, Y., & Han, S. (2013). Tonogenesis in early Contemporary Seoul Korean: A longitudinal case study. Lingua, 134, 62–74.

Karlgren, B. (1926). Études sur la phonologie Chinoise. Leiden: E.J. Brill: Archives d' études Orientales 15.

Karlsson, F., Zetterholm, E., & Sullivan, K. P. H. (2004). Development of a Gender Difference in Voice Onset Time. In Proceedings of the 10th australian international conference on speech science & technology (pp. 316–321). Sydney, Australia.

Kawahara, H., Takahashi, T., Morise, M., & Hideki, B. (2009). Development of exploratory research tools based on TANDEM-STRAIGHT. In Proceedings of 2009 apsipa annual summit and conference (pp. 111–120). Sapporo, Japan. Retrieved from http://eprints.lib.hokudai.ac.jp/dspace/handle/2115/39651

Keating, P. (1984). Phonetic and Phonological Representation of Stop Consonant. Language, 60(2), 286–319.

Keating, P. (2006). Phonetic Encoding of Prosodic Structure. In J. Harrington & M. Tabain (Eds.), Speech production: Models, phonetic processes, and techniques (pp. 167–186). New York and Hove: Psychology Press.

Keating, P., Cho, T., Fougeron, C., & Hsu, C.-S. (2003). Domain-initial articulatory strengthening in four languages. In J. Local, R. Ogden, & R. Temple (Eds.), Phonetic interpretation: Papers in laboratory phonology vi (pp. 145–163). Cambridge University Press.

Kelterer, A. (2017). Non-modal voice quality in Chichimeco (Unpublished doctoral dissertation). Lund University.

Kessinger, R. H., & Blumstein, S. E. (1997). Effects of speaking rate on voice-onset time in Thai, French, and English. Journal of Phonetics, 25, 143–168.

Khan, S. u. D. (2012). The phonetics of contrastive phonation in Gujarati. Journal of Phonetics, 40(6), 780–795.

Kingston, J. (2011). Tonogenesis. In The blackwell companion to phonology (Vol. 97, pp. 1–30).

Kingston, J., & Diehl, R. L. (1994). Phonetic Knowledge. Language, 70(3), 419–454.

Kirby, J. P., & Ladd, D. R. (2016). Effects of obstruent voicing on vowel F0: Evidence from "true voicing" languages. The Journal of the Acoustical Society of America, 140(4), 2400–2411.

Kleber, F. (2018). VOT or quantity: What matters more for the voicing contrast in German regional varieties? Results from apparent-time analyses. Journal of Phonetics, 71, 468–486.

Kuang, J. (2011). Production and Perception of the Phonation Contrast in Yi (Unpublished doctoral dissertation). UCLA.

Kuang, J., & Cui, A. (2018). Relative cue weighting in perception and production of a sound change in progress. Journal of Phonetics, 71, 194–214.

Kuang, J., & Keating, P. (2014). Vocal fold vibratory patterns in tense versus lax phonation contrasts. Journal of the Acoustical Society of America, 136(5), 2784–2797.

Kuang, J., Tian, J., & Zhou, Y. (2018). The common word prosody in Northern Wu. In Proc. tal2018, sixth international symposium on tonal aspects of languages (pp. 7–11). Berlin.

Kuzla, C., & Ernestus, M. (2011). Prosodic conditioning of phonetic detail in German plosives. Journal of Phonetics, 39(2), 143–155.

Ladd, D. R., & Schmid, S. (2018). Obstruent voicing effects on F0, but without voicing: Phonetic correlates of Swiss German lenis, fortis, and aspirated stops. Journal of Phonetics, 71, 229–248.

Ladefoged, P., & Maddieson, I. (1996). The Sounds of the World's Languages. Oxford: Blackwell Publishers.

Ladefoged, P., Williamson, K., Benjamin, O. E., & Uwulaka, A. (1976). The stops of Owerri Igbo. Studies in African Linguistics Supplement, 6, 147–163.

Lindblom, B. (1990). Explaining Phonetic Variation: A Sketch of the H&H Theory. In W. J. Hardcastle & A. Marchal (Eds.), Speech production and speech modeling (pp. 403–439). Dordrecht: Kluwer: Kluwer Academic Publishers. doi: 10.1016/0165-1684(91)90075-t

Lindblom, B., Guion, S., Hura, S., Moon, S.-J., & Willerman, R. (1995). Is sound change adaptive? Revista di linguistica, 7(1), 5–37.

Lisker, L. (1986). "Voicing" in English: A Catalogue of Acoustic Features Signaling /b/ Versus /p/ in Trochees. Language and Speech, 29(1), 3–11.

Lisker, L., & Abramson, A. S. (1964). A Cross-Language Study of Voicing in Initial Stops: Acoustical Measurements. Word, 20(3), 384–422.

Maddieson, I. (1984). Patterns of sounds. Cambridge University Press.

Maran, L. R. (1973). On becoming a tone language: A Tibeto-Burman model of tonogenesis. In Consonant types and tone: Southern california occasional papers in linguistics, no. 1 (Vol. 1, pp. 98–114).

Maspero, H. (1912). Etudes sur la phonétique historique de la langue annamite. Les initiales. Bulletin de l'École française d'Extrême-Orient, 12(1), 1–124.

Matisoff, J. A. (1973). Tonogenesis in Southeast Asia. In L. Hyman (Ed.), Consonant types and tone: Southern california occasional papers in linguistics, no. 1 (pp. 71–95). Los Angeles: University of Southern California.

Mazaudon, M. (1977). Tibeto-Burman tonogenetics. Linguistics of the Tibeto-Burman Area, 3(2), 1–123.

Ohala, J. J. (1981). The listener as a source of sound change (C. S. Masek, R. A. Hendrick, & M. F. Miller, Eds.). Chicago: Chicago Linguistic Society.

Ohala, J. J. (1993). The phonetics of sound change. In C. Jones (Ed.), Historical linguistics: Problems and perspectives (pp. 237–278). London: Longman.

Pinget, A.-F. C. H. (2015). The actuation of sound change (Unpublished doctoral dissertation). Utrecht University.

Pulleyblank, E. G. (1984). Middle Chinese: A Study in Historical Phonology. Vancouver: University of British Columbia Press.

Ren, N. (1992). Phonation types and stop consonant distinctions: Shanghai Chinese (Unpublished doctoral dissertation). The University of Connecticut.

Roettger, T. B., Winter, B., Grawunder, S., Kirby, J. P., & Grice, M. (2014). Assessing incomplete neutralization of final devoicing in German. Journal of Phonetics, 43(1), 11–25.

Sagart, L. (1999). The origin of Chinese tones. In Proceedings of the symposium/cross-linguistic studies of tonal phenomena/tonogenesis, typology and related topics (pp. 91–104). Tokyo, Japan.

Schmidt, A. M., & Flege, J. E. (1996). Speaking rate effects on stops produced by Spanish and English monolinguals and Spanish/English bilinguals. Phonetica, 53(3), 162–179.

Selkirk, E., & Shen, T. (1990). Prosodie Domains in Shanghai Chinese. In Phonology-syntax connection (pp. 313–337). Chicago: University of Chicago Press.

Shen, Z., Wooters, C., & Wang, W. S.-Y. (1987). Closure Duration in the Classification of Stops: A statistical analysis. OSU Working Papers in Linguistics, 35, 197–209.

Shue, Y.-L., Keating, P., Vicenik, C., & Yu, K. (2011). VoiceSauce: A program for voice analysis. In Proceedings of the 17th international congress of phonetic sciences (pp. 1846–1849). Hong Kong.

Silva, D. J. (2006). Acoustic emergence in evidence of tonal for the contrast contemporary Korean. Phonology, 23(2), 287–308.

Solé, M.-J. (2014). The perception of voice-initiating gestures. Laboratory Phonology, 5(1), 37–68.

Stevens, K. N. (1977). Physics of Laryngeal Behavior and Larynx Modes. Phonetica, 34, 264–279.

Thurgood, G. (2002). Vietnamese and tonogenesis: Revising the model and the analysis. Diachronica, 19(2), 333–363.

Thurgood, G. (2007). Tonogenesis Revisited: Revising the Model and the Analysis. In J. G. Harris, S. Burusphat, & J. E. Harris (Eds.), Studies in tai and southeast asian linguistics (pp. 263–291). Bangkok: Ek Phim Thai Ltd.

Tian, J., & Kuang, J. (2016). Revisiting the Register Contrast in Shanghai Chinese. In Tonal aspects of languages 2016 (pp. 147–151).

Tsu-lin, M. (1970). Tones and Prosody in Middle Chinese and The Origin of The Rising Tone Author. Harvard Journal of Asiatic Studies, 30, 86–110.

Turk, A. E., & White, L. (1999). Structural influences on accentual lengthening in English. Journal of Phonetics, 27, 171–206.

Wang, Y. (2012). 吴语塞音声母的声学和感知研究——以上海话为例 [Acoustic Measurements and Perceptual Studies on Initial Stops in Wu-Dialects ——Take Shanghainese for example] (Unpublished doctoral dissertation). Zhejiang University.

Watt, D., & Yurkova, J. (2007). Voice Onset Time and the Scottish Vowel Length Rule in Aberdeen English. In Proceedings of the 16th international congress of phonetic sciences (pp. 1521–1524). Saarbrücken.

Wetzels, W. L., & Mascaró, J. (2001). The Typology of Voicing and Devoicing. Language, 77(2), 207–244.

Williams, L. (1977). The voicing contrast in Spanish. Journal of phonetics, 5(2), 169–184.

Wolfe, P. M. (1972). Linguistic Change and the Great Vowel Shift in English. University of California Press.

Xu, B., & Tang, Z. (1988). 上海市区方言志 [A Description of the Urban Shanghai Dialect]. Shanghai: Shanghai Educational Publishing House.

Yip, M. (1980). The Tonal Phonology of Chinese (Unpublished doctoral dissertation). Massachusetts Institute of Technology.

Yip, M. (1993). Tonal Register in East Asian Languages. In K. Snider & H. van der Hulst (Eds.), The phonology of tone: The representation of tonal register (pp. 245–268). DE GRUYTER MOUTON.

Zee, E. (2004). Tone and Syntax in Shanghai Dialect..

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

Zhang, J., & Meng, Y. (2016). Structure-dependent tone sandhi in real and nonce disyllables in Shanghai Wu. Journal of Phonetics, 54, 169–201.

Zhang, J., & Yan, H. (2018). Contextually dependent cue realization and cue weighting for a laryngeal contrast in Shanghai Wu. Journal of the Acoustical Society of America, 144(3), 1293–1308.

Zhirmunskiĭ, V. M. (1962). Notes on the "binnenhochdeutsche Konsonantenschwächung" (interior High German consonant weakening) and the mergers it caused. Berlin: Akademie-Verlag.

Zhu, X. (1995). Shanghai tonetics (Unpublished doctoral dissertation). The Australian National University.