BRILL

CAHIERS DE LINGUISTIQUE ASIE ORIENTALE
EAST ASIAN LANGUAGES AND LINGUISTICS
46 (2017) 1–31

CAHIERS DE LINGUISTIQUE ASIE ORIENTALE
brill.com/clao

Phonetic and phonological properties of tones in Shanghai Chinese

Jiayin GAO
Laboratoire de Langue et civilisation à tradition orale, CNRS–Université Paris 3
Laboratoire de Phonétique et Phonologie, CNRS–Université Paris 3
jiayin.gao@univ-paris3.fr

Pierre HALLÉ
Laboratoire de Phonétique et Phonologie, CNRS–Université Paris 3
Laboratoire Mémoire et cognition, INSERM–Université Paris 5
pierre.halle@univ-paris3.fr

Abstract

This study investigates the relations between tone, voicing, and voice quality in modern Shanghai Chinese. In low tone syllables, word-initial obstruent onsets are traditionally described as voiceless and breathy, and sonorant onsets as voiced and breathy.

Our study is based on acoustic and electroglottographic (EGG) data from speakers of two age groups (20–30 vs. 60–80 years). Our results are globally in line with previous studies, but with notable differences. In low tone syllables, while word-initial stops are phonetically voiceless most of the time, fricatives are quite often phonetically voiced. While low tone obstruent onsets are followed by breathier vowels than high tone onsets, this pattern is not clear-cut for nasal onsets. Furthermore, our transversal data show that low tone breathiness is more systematically produced by elderly – especially male – speakers, rather than young speakers, suggesting an on-going change towards the loss of breathiness.

 | DOI: 10.1163/19606028-04601001

Keywords

phonetics – phonology – Shanghai Chinese – tone – voicing – breathy voice – electroglottography

Résumé

Cette étude traite des relations entre ton, voisement et qualité de voix en shanghaïen. Dans les syllabes au ton bas, les initiales obstruantes sont décrites traditionnellement comme non-voisées avec voix soufflée et les sonantes comme voisées avec voix soufflée.

Notre étude analyse les données acoustiques et électroglottographiques de locuteurs de deux groupes d'âge (20–30 vs. 60–80 ans). Nos résultats concordent avec les études passées, avec toutefois quelques différences notables. Aux tons bas, les initiales occlusives sont le plus souvent non-voisées tandis que les fricatives sont souvent voisées. Les initiales obstruantes sont suivies par des voyelles plus soufflées aux tons bas que hauts. Cette tendance n'est pas nette pour les nasales. D'autre part, nos données transversales montrent que la qualité de voix soufflée est plus systématique chez les locuteurs âgés – surtout les hommes – que chez les jeunes, suggérant un changement en cours vers la perte de voix soufflée.

Mots-clés

phonétique – phonologie – Shanghaïen – ton – voisement – voix soufflée – electroglottographie

1 Introduction

1.1 *Background*

Shanghai Chinese, in the broad sense, comprises different varieties spoken in the urban and suburban areas of Shanghai. In this study, we use Shanghai Chinese in the narrow sense, that is, the variety spoken in the urban area. Cross-age variations are quite important in Shanghai Chinese, to the extent that three generational varieties are classically defined: the Old variety (*lǎopài* 老派, generation born before the 1930s), the Middle variety (*zhōngpài* 中派, born between the 1940s and the 1960s), and the New variety (*xīnpài* 新派, born between the 1970s and 1990s) (Xu & Tang 1988; Qian 2003). Chen and Gussenhoven (2015) mentioned large variations in the New variety, suggesting con-

TABLE 1 *Phonological co-occurrence of tone register and obstruent onset in Shanghai Chinese*

Tone		Obstruent onset		Label
High	T1, T2, T4	Voiceless	p pʰ t tʰ k kʰ ts tsʰ tɕ tɕʰ f s ɕ	yin
Low	*T3, T5*	Voiced	b d g dʑ v z ʑ	yang

tinuous evolution from the older speakers of the New variety to the youngest speakers, the latter being probably more strongly influenced by Standard Chinese.

Shanghai Chinese belongs to the Wu dialect group, the second largest Chinese dialect group after the Mandarin group. Wu dialects are spoken in the southern part of the Jiangsu province, in Shanghai, in most of the Zhejiang province, as well as in a few counties in other provinces. One of the common characteristics to all Wu dialects is their uniform retention of the three-way laryngeal contrast of stop consonants of Middle Chinese (Li 1937). Today's Shanghai Chinese has a three-way laryngeal contrast for stop consonants (voiceless aspirated, voiceless unaspirated, and voiced) and a two-way laryngeal contrast for fricative consonants (voiceless vs. voiced) (see Table 1).

The five syllable-citation tones in Shanghai Chinese are numbered tones 1–5, and referred to as T1–5. Their citation forms (i.e., as produced on a single syllable) are 52 (˥˨), 34 (˧˦), 23 (˨˧), 55 (˥), and 12 (˩˨), respectively, according to Xu & Tang (1988), using Chao's tone-letter 1–5 scale (Chao 1930). Slightly different values have been proposed, such as 51 instead of 52 (T1), or 13 instead of 23 (T3). T4 and T5 are "checked tones," that is, they are carried by a short syllable ending with a glottal stop in pre-pausal position. Although every syllable except some grammatical particles (which cannot occur word-initially) has a tone identity, that is, carries a base syllable-tone, Shanghai Chinese has a word-tone system: In phonological words, the tone contour of the whole word is determined by the tone identity of the initial syllable. In the case of disyllabic words, for instance, the tone contour of the first syllable spreads to the two syllables of the whole word. For example, if the first syllable's base tone is T1 (52), the two-syllable contour becomes 55+22, regardless of the second syllable's base tone.

Voicing and tone. Phonological voicing of the Shanghai syllables' onset consonants is related to tone register, in that phonologically voiceless obstruent onsets co-occur with a high tone register comprising T1, T2, and T4, whereas phonologically voiced obstruent onsets co-occur with a low tone register comprising *T3*, and *T5*, as shown in Table 1. For sake of clarity, we use in the following *T3* and *T5* (italics) for the low register tones to distinguish them from high register tones. Sonorant onsets, which are unspecified for voicing and are always phonetically voiced, may co-occur with either high or low tones, but much more frequently with low than high tones. High- and low-register tones are traditionally called "yin" (阴) and "yang" (阳) tones, respectively. In this paper, we use "yin" and "yang" to label high and low-register tones, respectively. In addition, as far as obstruent onsets are concerned, we consider that high-register tone and [-voice] form one phonological category, which we call "yin," and that low-register tone and [+voice] form another phonological category, which we call "yang" (Table 1). (For the "yin" category, this paper only focuses on the voiceless *unaspirated* stop series; Shanghai Chinese also has a voiceless aspirated series of stops, which belongs to the high-register tone or "yin" category.) Realization of the tone/voicing contrast is complementary according to the syllable position in a phonological word. As explained above, word-initial syllables maintain their tonal identity, although tone shape is modified by the spreading process. In this context, the "yin/yang" contrast is realized as high vs. low register, and the obstruent consonant onset of the word is always phonetically voiceless according to both impressionistic descriptions and acoustic analyses (Liu 1925; Chao 1928; Cao & Maddieson 1992). Non-initial syllables lose their original tone identity. In this context, the "yin" vs. "yang" contrast is realized as a *phonetic* voicing contrast, that is, with vs. without glottal pulsing.

Voicing and voice quality. Shanghai phonologically voiced stops ("yang" stops) have long been described as "clear sound with muddy aspiration" (*qīngyīn zhuóliú* 清音浊流) in word-initial position (Liu 1925; Chao 1928). While in Chinese terminology, "clear" and "muddy" are used to describe voiceless and voiced sounds, respectively, muddy aspiration is usually interpreted as "breathy voice." Although the voiced aspiration is not as strong as in Hindi, or even "trop faible pour mériter d'être désignée" ['too weak to deserve the [breathy] qualification'] (Karlgren 1926: 260), recent experimental studies mostly confirmed that "yang" stops are breathier than "yin" stops. Ren Nianqi's (1988; 1992) research supported this observation with fiberoptic transillumination data from two speakers in their late 20s-mid 30s at the time of the recording: his data showed that the peak of glottal opening occurred well before stop release for a "yin" voiceless non-aspirated stop onset but at vowel onset for a "yang" stop as well as for a voiceless aspirated stop. This suggested greater airflow at stop release for "yang" stops, presumably corresponding to breathiness. Gao et al. (2011), however, did not observe the same pattern with a young speaker aged 22, in that the peak of the glottal opening for "yang" stops occurred before vowel onset, suggesting little or no breathiness. Cao

& Maddieson (1992) and Chen Yiya (2011) used H1–H2 (difference in amplitude between the first and second harmonics in the spectrum) as one indicator of voice quality – with higher values indicating a steeper spectral slope, that is, breathier voice (Fischer-Jørgensen 1967; Klatt & Klatt 1990; Gordon & Ladefoged 2001). Both studies found higher H1–H2 for "yang" than "yin" stops at stop release. H1–A1 (difference in amplitude between the first harmonic and the first formant) as well as aerodynamic measures showed similar but less robust tendencies (Cao & Maddieson 1992). The age of the speakers was not reported in the study of Cao and Maddieson (1992), and the speakers examined by Chen Yiya (2011) were relatively old (born between 1935 and 50).

Other articulatory and acoustic correlates of the "yin/yang" contrast have been reported, such as intensity and duration (Rose 1982a; Shen et al. 1987; Chen 2010; Gao & Hallé 2013). This study will only focus on the laryngeal properties mentioned above, that is, F0, phonetic voicing, and voice quality.

1.2 *Goal of the study*

The main goal of this study is to give a detailed description of the phonetic correlates of Shanghai tones – F0 contours, phonetic voicing of the onset, and voice quality – (1) for monosyllabic and disyllabic words, (2) for zero, stop, fricative and nasal onsets (previous studies mostly focused on stop onsets), (3) using acoustic and electroglottographic (EGG) data, and (4) using speakers of two age groups. Note again that the voiceless aspirated stop series is out of the scope of this study.

The central questions addressed in this study are:

> *How can we define the relationship between (phonologically) voiced onsets and low tones on one hand, and between voiced onsets and breathy voice on the other hand? Are these relationships phonetic or phonological?*

Besides, we also aim at answering the following specific questions:

1. *How might phonetic voicing and voice quality interact with onset-type?* In particular, how do zero, nasal, and fricative onsets compare to stop onsets? Previous studies mostly focused on stop onsets and described the other onset types only impressionistically.
2. *How does voice quality vary according to within-word syllable position?*
3. *What is the phonetic domain of low tone breathiness?* Two main proposals have been put forward: Breathy voice has been claimed to be a property of the syllable onset (Cao & Maddieson 1992; Ren 1992), or a property of

> the entire syllable (Sherard 1972: 87; Ramsey 1987: 91; Rose 1989; 2002). While some authors attribute breathiness to the entire syllable based on auditory impressions and phonological arguments, Cao & Maddieson (1992) examined the phonetic domain of breathiness based on acoustic evidence, using three time points within a syllable, comparing "yin" and "yang" tone syllables. They showed that the voice quality difference was only apparent in the first of the three time points. In our study, we aim to examine the phonetic domain of breathiness, using five time points instead of three in order to refine the measurement.

Finally, we used transversal data, including the productions of two age groups, for the purpose of understanding how Shanghai tone production evolves over time. Under the tremendous influence of Standard Chinese, as well as migrant dialects, urban Shanghai Chinese has undergone rapid and great changes since more than a century ago. At the phonological level, the consonantal, vocalic, and tonal systems underwent two main types of change: (1) simplification, accomplished by the mergers of certain phonological categories, e.g., the merger of "yin shang 阴上" and "yin qu 阴去" tonal categories into the latter, today's T2 category (Qian 2003); (2) especially among the youngest generation, neutralization, or, on the contrary, creation of a new contrast in compliance with the Standard Chinese system (e.g., loss of /ŋ/ in syllable onset position, as in /ŋa/ *T3* > /a/ *T3* 外 'outside,' syllable onset /ŋ/ not being permissible in Standard Chinese). In this study, we investigate whether and how sound changes occur at the phonetic level in tone production.

2 Experiment

2.1 *Methods*

2.1.1 Participants

Twenty-two native speakers of Shanghai Chinese participated in the recordings. Two age groups were tested. The young group included 12 speakers (6 male and 6 female) from 21 to 29 years of age (mean 24.9); the elderly group included 10 speakers (4 male and 6 female) from 61 to 79 years of age (mean 68.7) at the time of the recordings.

The participants were given a detailed questionnaire on their linguistic background. All the speakers were born and raised in Shanghai urban or suburban areas, except one elderly male speaker who was born in the Jiangsu province but moved to Shanghai before the age of one. All of them had spent most of their lifetime in the Shanghai urban area. Two of the elderly speakers, a couple, had lived in the Chongming county for nearly 30 years, where the Chongming variety of Shanghai Chinese is spoken.

All the young speakers had learned Standard Chinese before the age of eight; all learned English at school; one spoke some German and three spoke some French; two of them spoke one or two other Wu dialects, and the others did not speak any other dialect than Shanghai Chinese and Standard Chinese. As for the elderly speakers, all had learned Standard Chinese at primary school or at adult age. None of them spoke any foreign language, but six of them spoke another Wu dialect than Shanghai Chinese.

We will report the acoustic data of all the 22 participants, and the EGG data of 10 of the 22 participants, including 6 young speakers (3 males and 3 females, mean age 24.3, range 24–25), and 4 elderly speakers (3 males and 1 female, mean age 67.3, range 64–72), of whom the EGG signals were the least noisy.

2.1.2 Speech materials and design

We used syllables of Shanghai Chinese in all the five lexical tones T1 to *T5*, with the following onset types: zero onset, stops, fricatives, and nasals, that is, all manners of articulation except glides and affricates. We used a monosyllabic context in order to examine the tones in their citation forms, and two disyllabic contexts for the purpose of examining sandhi-modified realizations. In the disyllabic contexts, the target syllable was either the first syllable, which should partly maintain its tonal identity, or the second syllable, which should lose its tonal identity by virtue of a tone sandhi rule, whereby the tone contour of the non-initial syllables of a polysyllabic word is determined by the sole first syllable (Yip 1980; Zee & Maddieson 1979).

Thirty-two monosyllabic and sixty dissyllabic words were produced in the carrier sentence [__ gə ə zz̩ ŋo nin tə ə] (__这个字/词我认得的。‘__ this character/word, I know it’). The target word was elicited in sentence-initial position in order to avoid potential intervocalic voicing of “yang” obstruents in word-initial syllable. The target syllable appeared in three contexts: (1) monosyllabic target word, (2) first syllable of a dissyllabic target word (S1), and (3) second syllable of a disyllabic target word (S2). Each target syllable carried one of the five citation tones. The unchecked syllables (T1, T2, *T3*) shared the /ɛ/ rhyme and the checked syllables (T4, *T5*) shared the /aʔ/ rhyme.

(1) Monosyllabic words.

T1:	ɛ 哀	‘grief’	pɛ 杯	‘cup’	tɛ 堆	‘stack’
T2:	ɛ 爱	‘love’	pɛ 板	‘board’	tɛ 胆	‘gallbladder’
T3:	*ɛ* 咸	*‘salty’*	*bɛ* 办	*‘handle’*	*dɛ* 谈	*‘talk’*

T4:	aʔ 鸭	'duck'	paʔ 八	'eight'	taʔ 搭	'build'
T5:	*aʔ 盒*	*'box'*	*baʔ 白*	*'white'*	*daʔ 踏*	*'tread'*

T1:	fɛ 翻	'turn over'	sɛ 三	'three'	mɛ 蛮[1]	'quite'	nɛ 拿[2]	'take'
T2:	fɛ 反	'reverse'	sɛ 伞	'umbrella'	mɛ 美	'beautiful'		
T3:	*vɛ 饭*	*'rice'*	*zɛ 才*	*'talent'*	*mɛ 梅*	*'plum'*	*nɛ 难*	*'difficult'*
T4:	faʔ 发	'deliver'	saʔ 杀	'kill'				
T5:	*vaʔ 罚*	*'punish'*	*zaʔ 石*	*'stone'*	*maʔ 麦*	*'wheat'*	*naʔ 纳*	*'accept'*

(2) Disyllabic words with target syllable in syllable 1 (S1 context).

T1:	**pɛ**.tsz̩ 杯子	'cup'	**tɛ**.ɕiŋ 担心	'worry'
T2:	**pɛ**.tɕɪʔ 背脊	'back (of the body)'	**tɛ**.tsz̩ 胆子	'courage'
T3:	***bɛ**.koŋ 办公*	*'work'*	***dɛ**.tsz̩ 台子*	*'table'*
T4:	**paʔ**.paʔ 八百	'eight hundreds'	**taʔ**.sɛ 搭讪	'hit on (someone)'
T5:	***baʔ**.pɛ 白板*	*'white board'*	***daʔ**.zaʔ 踏实*	*'steady and sure'*

T1:	**fɛ**.ɪʔ 翻译	'translate'	**sɛ**.ɕi 三鲜	'shredded sea foods'
T2:	**fɛ**.wɛ 返回	'return'	**sɛ**.piŋ 伞柄	'handle of the umbrella'
T3:	***vɛ**.wø 饭碗*	*'bowl'*	***zɛ**.nəŋ 才能*	*'talent'*
T4:	**faʔ**.lɪʔ 法律[3]	'law'	**saʔ**.tsʰu 塞车	'traffic jam'
T5:	***vaʔ**.kø 罚款*	*'monetary penalty'*	***zaʔ**.dɤ 石头*	*'stone'*

(3) Disyllabic words with target syllable in syllable 2 (S2 context).

T2+**T1**→33+44:	tsɔ.**pɛ** 早班	'morning shift'	tɕi.**tɛ** 简单	'easy'
T1+**T1**→55+21:	ku.**pɛ** 科班	'professional training'	sz̩.**tɛ** 私单	'private deal'
T2+**T2**→33+44:	ɕi.**pɛ** 死板	'stubborn'	tɕiɔ.**tɛ** 校对	'proofread'
T1+**T2**→55+21:	kø.**pɛ** 干贝	'dried scallop'	tsz̩.**tɛ** 猪胆	'pig's gallbladder'
*T2+**T3**→33+44:*	*pʰɛ.**bɛ** 配备*	*'equip'*	*tsz̩.**dɛ** 子弹*	*'bullet'*
*T1+**T3**→55+21:*	*tsʰɔ.**bɛ** 操办*	*'manage'*	*tɕi.**dɛ** 鸡蛋*	*'chicken egg'*
T2+**T4**→33+44:	sz̩.**paʔ** 四百	'four hundreds'	pɔ.**taʔ** 报答	'return back'
T1+**T4**→55+21:	sɛ.**paʔ** 三百	'three hundreds'	i.**taʔ** 医德	'medical ethics'

1 This character has another reading with *T3*, so the lexical context was given to elicit the T1 reading.

2 There are at least two phonetic variants of this character, [nɛ] and [no] (both with T1). The speaker was instructed to produce the desired reading [nɛ] (T1).

3 This word is produced [faʔ.lɪʔ] by most elderly speakers, but [faʔ.lʏʔ] by all young speakers.

T2+T5→33+44:	*ɕiɔ.**baʔ*** 小白	(*a frequent nickname*)	*tsø.**daʔ*** 转达	*'transmit'*
T1+T5→55+21:	*kɔ.**baʔ*** 茭白	(*a Chinese food plant*)	*ɔ.**daʔ*** 凹凸	*'concavity'*

T2+**T1**→33+44:	tsʰɔ.**fɛ** 吵翻	'quarrel'	tʰɔ.**sɛ** 套衫	'sweaters'
T1+**T1**→55+21:	sɛ.**fɛ** 三番	'time and again'	i.**sɛ** 衣衫	'clothes'
T2+**T2**→33+44:	tɕʰi.**fɛ** 遣返	'repatriate'	tsɤ.**sɛ** 走散	'lost'
T1+**T2**→55+21:	sɛ.**fɛ** 三反	'three-anti campaigns'	ɕiɔ.**sɛ** 消散	'disappear'
*T2+**T3**→33+44:*	*tsɔ.**vɛ*** 早饭	*'breakfast'*	*tɕiɔ.**zɛ*** 教材	*'textbook'*
*T1+**T3**→55+21:*	*tsʰz̩.**vɛ*** 糍饭	*'stuffed rice ball'*	*tʰi.**zɛ*** 天才	*'genius'*
T2+**T4**→33+44:	zu.**faʔ** 做法	'method'	pɛ.**saʔ** 板刷	'scrubbing brush'
T1+**T4**→55+21:	kɛ.**faʔ** 开发	'develop'	tsz̩.**saʔ** 知识	'knowledge'
*T2+**T5**→33+44:*	*tʰi.**vaʔ*** 体罚	*'corporal punishment'*	*ɕy.**zaʔ*** 选择	*'choice'*
*T1+**T5**→55+21:*	*i.**vaʔ*** 衣物	*'clothing'*	*ɕy.**zaʔ*** 虚实	*'actual status'*

We used only two rhymes, /ɛ/ in unchecked syllables and /aʔ/ in checked syllables, in order to avoid the influence of heterogeneous vocalic context on our different phonetic measures. Note that non-high vowels should be privileged for measures of phonation type to avoid the influence of low-frequency formants on the first and second harmonics (Hanson & Chuang 1999; Hanson 1997). The /ɛ/ and /aʔ/ rhymes were also chosen because they both occur after almost all the onset consonants and for almost all the tones we used in this study. We could not use the same vowel /a/ or /ɛ/ for both checked and unchecked syllables, because /a/ does not occur after /f, v/ in unchecked syllables, and /ɛ/ does not occur in checked syllables.

In the monosyllabic context, the onset could be a zero onset or a stop, fricative, or nasal onset with either labial or dental place of articulation, that is, belonged to the /∅ (zero), p, (b), t, (d), f, (v), s, (z), m, n/ set. Symbols within parentheses indicate phonologically voiced obstruents, which only co-occur with *T3* and *T5*, as mentioned in § 1.1. The velar place of articulation was not included because there are very few velar stop/nasal onset syllables and no velar fricatives. There is no T2 /nɛ/ syllable, nor T4 /maʔ/ or /naʔ/ syllable. This made a total of 32 (= 7 onsets × 5 tones – 3) monosyllables. The rhyme was /ɛ/ or /aʔ/ for unchecked or checked syllables, respectively. Each of the 22 speakers repeated the word list twice, except one young female speaker and one elderly male speaker who read the word list only once due to technical problems. In the S1 context, the possible onsets were restricted to the /p, (b), t, (d), f, (v), s, (z)/ set. The rhyme was, again, /ɛ/ or /aʔ/ for unchecked or checked syllables, respectively. This made a total of 20 (= 4 onsets × 5 tones) S1 syllables. In the S2 context, the tone value realized on the target syllable depends on the base tone of the first syllable. It is high level (44) when the preceding syllable's tone is

T2, *T3*, or *T5*; it is low (21 or 23) when the preceding syllable's tone is T1 or T4. We therefore studied two sub-contexts in the S2 context: first syllable in tone T2 (second syllable should be high-level pitch) or in tone T1 (second syllable should be low pitch). As in the S1 context, the set of possible onsets was /p, (b), t, (d), f, (v), s, (z)/ the rime was /ɛ/ or /aʔ/ for unchecked or checked syllables. This made a total of 40 (= 4 onsets × 5 tones × 2 preceding tones) S2 syllables.

2.1.3 Apparatus

Simultaneous audio and electroglottographic (EGG) data were collected. Speakers were recorded individually in a quiet room. The audio recordings were made with a high quality headband microphone through an external soundboard connected to a laptop in stereo mode: one channel for the audio signal, and the other for the EGG signal. The EGG signals were recorded with a VoceVista EGG system. Both signals were sampled at 44.1 kHz, with 16-bit resolution.

2.2 *Analyses and results*

2.2.1 Acoustic data

2.2.1.1 *Analyses*

Fundamental frequency (F0). For the estimation of F0, we used the cross-correlation method as implemented in Praat (Boersma 2001), setting the default F0 range to [60, 400 Hz] (which covered the F0 range of both male and female voices) and the analysis time step at 5 ms. We used default settings for all other parameters. (For a few male speakers, we had to modify the settings F0 minimum and 'octave cost' in order to avoid errors resulting in half the correct F0.) For each target syllable's rime (/ɛ/ or /aʔ/), we computed mean F0 values over five consecutive equal time intervals covering the entire vowel. That is, we time-normalized the F0 contour data. Because the durations of the rhymes ranged from ~70 to ~370 ms, the mean time interval in these F0 contours ranged from about ~14 to ~74 ms.

Voicing. Two measures related to phonetic voicing were used: voice onset time (VOT) for word-initial stops, and voicing-ratio (or v-ratio) for fricatives and word-medial stops. VOT intervals corresponded to the duration between onset of the release burst visible on the spectrogram and the zero-crossing point preceding the periodic waveform of the following vowel. V-ratio is the proportion of the voiced part duration out of the total duration of the consonant: the voiced part was determined by the detection of F0, as calculated in Praat (Boersma 2001) using the cross-correlation method. For the v-ratio measurement, the F0 range was set from 60 to 400 Hz, and the time step at 2 ms.

Voice quality. For consonant onsets, we measured HNR (harmonics-to-noise ratio) during stop release, or entire consonant duration for fricatives and nasals, in order to estimate the amount of aperiodic noise possibly correlated with breathy phonation. We used the "Harmonicity (cc)" function in Praat (Boersma 2001) with the following parameters: 2 ms time step, 60 Hz pitch floor, 0.03 silence threshold, and 4.5 periods per window. For the vowel portion of target syllables, we used several measures of spectral tilt, including the difference in amplitude between the first and second harmonics (H1–H2), the first harmonic and the first formant (H1–A1), or the first harmonic and the second formant (H1–A2). We also computed HNR, as for consonant onsets, expecting that breathier vowel portions would be noisier, hence exhibit lower HNR values. We only report here the H1–H2 data, since we found it to be the most sensitive measure distinguishing breathy from modal voice vowels in Shanghai Chinese.

2.2.1.2 *Results*

F0 contours of citation tones. We begin with summarizing the five-point F0 contours for monosyllables shown in Figure 1,[4] according to tone and speaker group, for the unchecked (upper panel) and checked (lower panel) tones. For the unchecked tones, T1 starts high then falls down steeply; T2 starts lower than T1 but higher than *T3*, then rises very gently until a mid-high endpoint (very often, the contour can be rather flat, see Figure 2); *T3* starts very low and then rises until a mid-high endpoint similar in height to that of T2. The main difference between T1 and the two other unchecked tones is its contour shape, whereas the main difference between T2 and *T3* is their F0 onset. Concerning the checked tones T4 and *T5*, the main difference between them also lies in their F0 onset: T4 starts high and *T5* starts low. T4 is slightly falling and *T5* slightly rising; their final F0 contour is somewhat chaotic due to final glottalization, hence the larger variability on the final time points.

The F0 range, that is, the range of F0 variation between maximum and minimum in F0 contours, according to speaker group is worth noting. It is larger for female than male, especially young speakers, when measured in Hertz. If converted to semitones, however, elderly male speakers exhibit an F0 range which is very close to that of female speakers, whereas young male speakers have a much smaller range than the other groups (see Table 2). Such a narrow tonal space might have consequences on the distinction between tones, especially those with similar F0 contours, such as T2 and *T3*: additional cues may be used to maintain perceptibility, as we will discuss in the following. Young male speakers exhibit a difference of only 22 Hz (or ~3 semi tones) between T2 and *T3* onset, which is just sufficient for perceptual contrast in languages ('t Hart 1981).

4 All error bars in all figures throughout the paper represent standard errors.

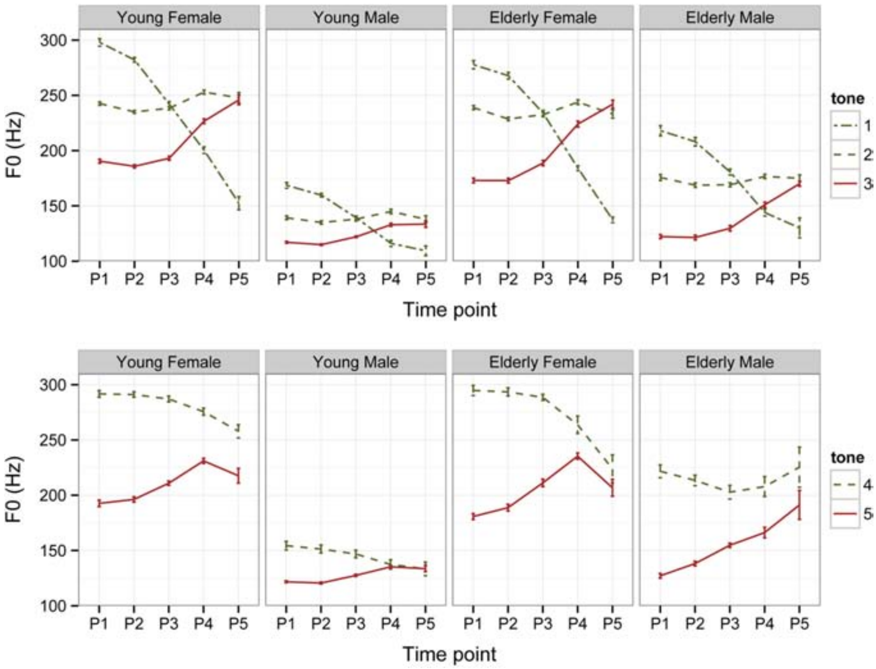

FIGURE 1 *Average FO contours of the five monosyllabic tones according to speaker group, syllable or tone type (upper panel: unchecked tones; lower panel: checked tones). "Muddy" (brown) colors/solid lines stand for "yang" tones.*

For young speakers, the FO ranges of male vs. female speakers are clearly distinct, with no overlap, whereas for elderly speakers, there is an overlap between male and female speakers' FO ranges. FO range is larger for elderly male speakers compared to young male speakers, but is similar between elderly female and young female speakers.

Voicing: VOT and v-ratio. For all the statistical analyses in this section, we ran two repeated-measure ANOVAs separately for the unchecked and checked syllable data, with *Manner* of articulation (stop vs. fricative), *Place* of articulation (labial vs. dental), *Tone* (unchecked: T1, T2, *T3*; checked: T4 vs. *T5*) as within-subject factors, *Gender* and *Age* as between-subject factors. The dependent variable was *VOT* or *v-ratio*. We will only report relevant results, that is, the effect of *Tone* on *VOT* or *v-ratio* and the interaction of *Tone* with other factors.[5]

As described in the literature, "yang" stops in word-initial position had positive VOTs, that is, were pronounced without pre-voicing, with rare exceptions

5 We report *p*-values with the classic though arbitrary assumption that $p < .05$ values point to significant effects (see Colquhoun 2014, for a criticism of such practice).

TABLE 2 *F0 range for elderly vs. young and male vs. female speakers*

	F0 range (Hz and semitones)	
	Young	**Elderly**
Male	60 Hz (7.8 st)	114 Hz (12.2 st)
Female	150 Hz (11.8 st)	155 Hz (13.2 st)

(~6‰). Among all the 22 speakers, only one young female speaker aged 24 produced "yang" stops with pre-voicing very occasionally (twice in monosyllables, and twice in the S1 context), and exclusively in checked syllables (selected recording: Sound 1).[6] The negative VOTs of this speaker were excluded from the VOT averaging. We did find a VOT difference between "yang" and "yin" unaspirated stops:[7] the former had slightly but overall significantly longer VOTs than the latter, in both monosyllabic and S1 contexts. No statistical interaction was found between *Tone* and *Gender* or *Tone* and *Age*. Therefore, we report averaged VOTs across all speakers in Table 3.

In the monosyllabic context, both ANOVAs (unchecked and checked syllables) showed significantly longer VOTs for "yang" than "yin" stops, with a mean difference of 3.6 ms between *T3* and T1–2 for unchecked syllables, $F(2,36) = 9.8$, $p = .0004$, and of 3.9 ms between *T5* and T4 for checked syllables, $F(1,18) = 19.3$, $p = .0004$. In the S1 context, *Tone* was significant only for unchecked syllables, showing longer VOTs for "yang" than "yin" stops, although only by 1.8 ms, $F(2,36) = 6.4$, $p = .0042$.

Table 4 shows the average voicing-ratio (v-ratio) of fricative onsets in word-initial position, according to tone, place of articulation, and context. (The v-ratio measure varies from 0 to 1, with higher v-ratios indicating more phonetic voicing.) The average v-ratio was close to zero for the phonologically voiceless (i.e., "yin") fricatives /f, s/ but considerably higher and more variable for the "yang" fricatives /v, z/, suggesting that "yang" fricatives in word-initial position were not systematically voiceless phonetically, as stop onsets were.

In both monosyllabic and S1 context, ANOVAs (for unchecked and checked syllables) showed a significant effect of *Tone*: *v-ratio* was significantly higher

6 All selected recordings can be found at http://booksandjournals.brillonline.com/content/journals/19606028/46/1.

7 Voiceless aspirated stops, which are not reported in the present study, clearly have longer VOTs than unaspirated stops (Gao et al. 2011).

TABLE 3 *Average VOTs in milliseconds (SD in parentheses) of stop onsets in first syllables (monosyllables or S1), according to place of articulation and syllable tone. Significance level (means): ** for* p < *.01. Grey shading indicates the "yang" stops: phonologically /b, d/; phonetically [p, t] in tones* T3 *or* T5.

		T1	T2	*T3*	T4	*T5*
Mono	[p]	14.4 (4.0)	13.9 (4.0)	16.9 (5.2)	11.6 (3.9)	15.7 (7.4)
	[t]	14.3 (3.9)	14.5 (4.1)	18.8 (7.6)	10.8 (3.7)	14.5 (4.7)
	Mean	**14.3**	**14.2**	**17.8 ****	**11.2**	**15.1 ****
S1	[p]	13.5 (4.7)	14.7 (5.3)	16.4 (5.9)	12.3 (4.6)	14.2 (6.9)
	[t]	14.5 (5.3)	15.5 (5.9)	16.5 (5.8)	14.8 (6.9)	15.6 (4.8)
	Mean	**14.0**	**15.1**	**16.4 ****	**13.6**	**15.0**

TABLE 4 *Average v-ratios (SD in parentheses) for fricative onsets of word-initial syllables (monosyllables or S1), according to context, place of articulation and tone. Significance level (means): ** for* p < *.01.*

		T1	T2	*T3*	T4	*T5*
Mono	[f]	0.11 (0.07)	0.09 (0.04)	0.61 (0.35)	0.07 (0.03)	0.61 (0.35)
	[s]	0.07 (0.02)	0.07 (0.03)	0.23 (0.31)	0.05 (0.02)	0.18 (0.26)
	Mean	**0.09**	**0.08**	**0.42 ***	**0.06**	**0.40 ****
S1	[f]	0.12 (0.07)	0.14 (0.09)	0.62 (0.37)	0.10 (0.05)	0.42 (0.32)
	[s]	0.08 (0.03)	0.08 (0.02)	0.14 (0.14)	0.06 (0.02)	0.07 (0.03)
	Mean	**0.10**	**0.11**	**0.38 ****	**0.08**	**0.24 ****

for "yang" than "yin" fricative onsets. For monosyllables, $\Delta = 0.34$ (unchecked), $F(2,36) = 64.0$, $p < .0001$, $\Delta = 0.34$ (checked), $F(1,18) = 52.1$, $p < .0001$; for S1 syllables, $\Delta = 0.27$ (unchecked), $F(2,36) = 43.4$, $p < .0001$, $\Delta = 0.16$ (checked), $F(1,18) = 21.8$, $p < .0001$. The *Tone* × *Place* interaction was also significant (mono unchecked syllables: $F(2,36) = 21.9$, $p < .0001$; mono checked syllables: $F(1,18) = 32.5$, $p < .0001$; S1 unchecked syllables, $F(2,36) = 31.8$, $p < .0001$; S1 checked syllables, $F(1,18) = 20.0$, $p = .0003$), with a larger "yang–yin" differential for labial than for dental fricatives in all cases.

The inter-speaker group difference was also confirmed by the ANOVAS. In monosyllables, for unchecked syllables, the *Tone* × *Age* × *Gender* interaction was significant, $F(2,36) = 3.7$, $p = .035$: the *Tone* × *Gender* interaction was signifi-

TABLE 5 *Average v-ratios (SD in parentheses) of obstruent onsets in non-initial (second) syllable, according to manner of articulation and tone. Significance level (means): ** for* p < *.01.*

	T1	T2	*T3*	T4	*T5*
stop	0.30 (0.14)	0.30 (0.13)	0.90 (0.21)	0.26 (0.20)	0.88 (0.18)
fricative	0.28 (0.11)	0.28 (0.12)	0.98 (0.09)	0.20 (0.10)	0.96 (0.15)
Mean	**0.29**	**0.29**	**0.94 ****	**0.23**	**0.92 ****

cant only for young speakers, $F(2,20) = 6.1$, $p = .0085$, reflecting a larger *T3* vs. T1–2 v-ratio differential for young female speakers ($\Delta = 0.56$) than for young male speakers ($\Delta = 0.27$). The same pattern obtained for the S1 context. The following only concerns the monosyllabic context. The *Tone* × *Age* interaction was significant only for female speakers, reflecting a larger *T3* vs. T1–2 differential for young female speakers ($\Delta = 0.56$) than for elderly female speakers ($\Delta = 0.24$). For checked syllables, *Tone* did not interact with *Gender*, $F(1,18) = 1.5$, $p = .24$., but the *Tone* × *Age* interaction was significant, $F(1,18) = 9.0$, $p = .0077$, reflecting a larger *T5* vs. T4 differential for young speakers ($\Delta = 0.49$) than for elderly speakers ($\Delta = 0.19$).

In summary, young speakers, especially young female speakers were more prone than other speakers to produce phonetically voiced word-initial fricatives, especially labial ones, for "yang" tones.

Table 5 shows the average v-ratio of obstruent onsets in word-medial position, according to tone, pooled across speaker groups and places of articulation (The same pattern of results was found for all speaker groups and both places of articulation). In average, v-ratio was close to 1 for "yang" onsets, much higher than for "yin" onsets, whether stops or fricatives, therefore suggesting that "yang" obstruents are phonetically voiced in word-medial position, as described in the literature. Occasionally (9% of our data), voiced stops were realized as spirantized in this position, that is, /b/ realized as a [ɸ]- or [β]-like fricative, and /d/ realized as a [z]- or [ð]-like fricative. These spirantized realizations were excluded from v-ratio analyses.

In both ANOVA analyses, *Tone* was significant: for unchecked syllables, *v-ratio* was significantly higher for *T3* than T1 and T2 syllables ($0.94 > 0.29$), $F(2,36) = 611.7$, $p < .0001$; for checked syllables, *v-ratio* was significantly higher for *T5* than T4 syllables ($0.92 > 0.23$), $F(1,18) = 1347.4$, $p < .0001$. The *Tone* × *Manner* interaction was significant, with a larger "yang–yin" differential for fricatives than for stops: for unchecked syllables, $\Delta = 0.70$ vs. 0.59 for fricatives vs.

stops, $F(2,36) = 24.0$, $p < .0001$; for checked syllables, $\Delta = 0.76$ vs. 0.62 for fricatives vs. stops, $F(1,18) = 22.4$, $p \leq .00025$. This was perhaps due to the fact that the burst-and-release portion of the stops was generally judged as voiceless by the F0 detection algorithm, thus reducing the v-ratio estimation. Contrary to the word-initial position, *Tone* did not interact with either *Gender* or *Age* in word-medial position: for unchecked syllables, $F\text{s} < 1$, n.s.; for checked syllables, *Tone* × *Gender*, $F < 1$, *Tone* × *Age*, $F(1,18) = 3.4$, $p = .082$.

Voice quality. In this section, we first provide illustrations of the "yin" vs. "yang" tone productions in the four speaker groups. Figure 2 shows waveforms and spectrograms of /pɛ/ T2 (left panels) and /bɛ/ *T3* (right panels) produced by a young female, a young male, an elderly female, and an elderly male speaker, from left top to right bottom (selected recordings: Sounds 2).

What catches our eye is that, while a noisy component, characteristic of a breathy quality, can be observed on the first half of the vowel in /bɛ/ *T3* for the two elderly speakers, it is virtually absent for the two young speakers. Moreover, variations are found especially in the young male and the elderly female groups (see § 2.2.2.2 for additional illustrations).

It seems that low tone breathiness is produced by elderly speakers, but not so much by young speakers. In order to quantify this observation, we report acoustic measurements of voice quality: a robust estimate of spectral tilt, H1–H2 (higher values indicate breathier phonation), and a measure of harmonics-to-noise ratio (HNR: lower HNR values suggest noisier/breathier phonation); we report H1–H2 computed on the vocalic portions and HNR on the onset consonant portions of the stimuli.

Figure 3 shows mean HNR values for the onset consonants of monosyllables. (A very similar pattern obtains for the S1 context.) For stop onsets, in both the monosyllabic and S1 contexts, mean HNR was lower for *T3* than T1 or T2 syllables. No difference was found between T1 and T2. Mean HNR was likewise lower for *T5* than T4. This indicated that stop release was noisier, thus breathier, for "yang" than "yin" stop onsets. For fricative onsets, the pattern was numerically reversed. However, this was not likely to reflect breathier phonation for "yin" than "yang" fricative onsets. Word-initial "yang" fricatives were often realized as phonetically voiced, as shown above, thus exhibiting harmonic structure, which yielded higher HNR values than is observed with "yin" fricatives, even though they might be produced with breathy phonation. For nasal onsets, little difference was observed between "yin" and "yang" tones, for all speaker groups, suggesting that in today's Shanghai Chinese, "yang" nasal onsets are not breathier than "yin" nasal onsets, contrary to descriptions in Chao (1928) or Rose (1989; 2002). We did not analyze the S2 context, since intervocalic "yang"

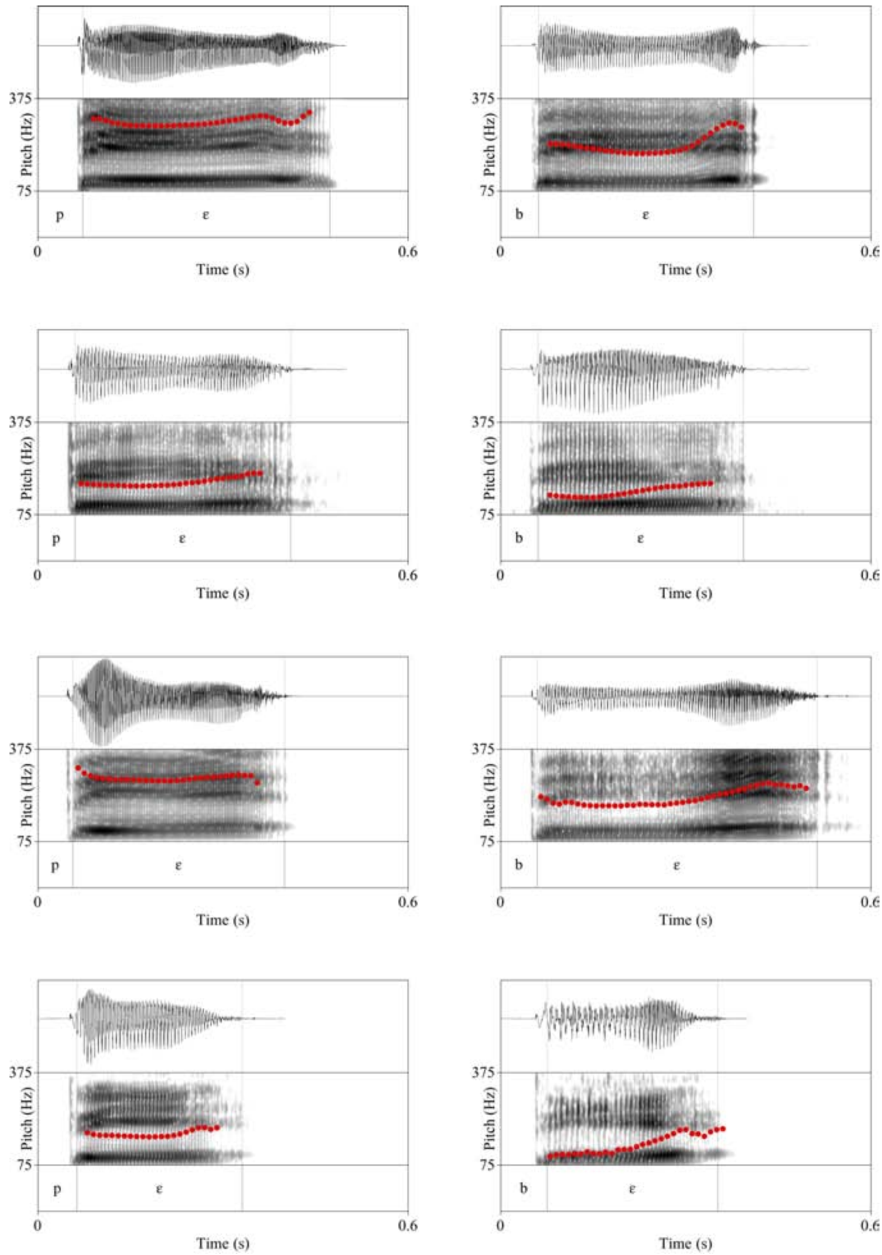

FIGURE 2 *Illustrations of waveforms and spectrograms (with F0 contours) for /pɛ/ T2 (left panels) and /bɛ/ T3 (right panels) syllables; from top to bottom: young female, young male, elderly female, and elderly male*

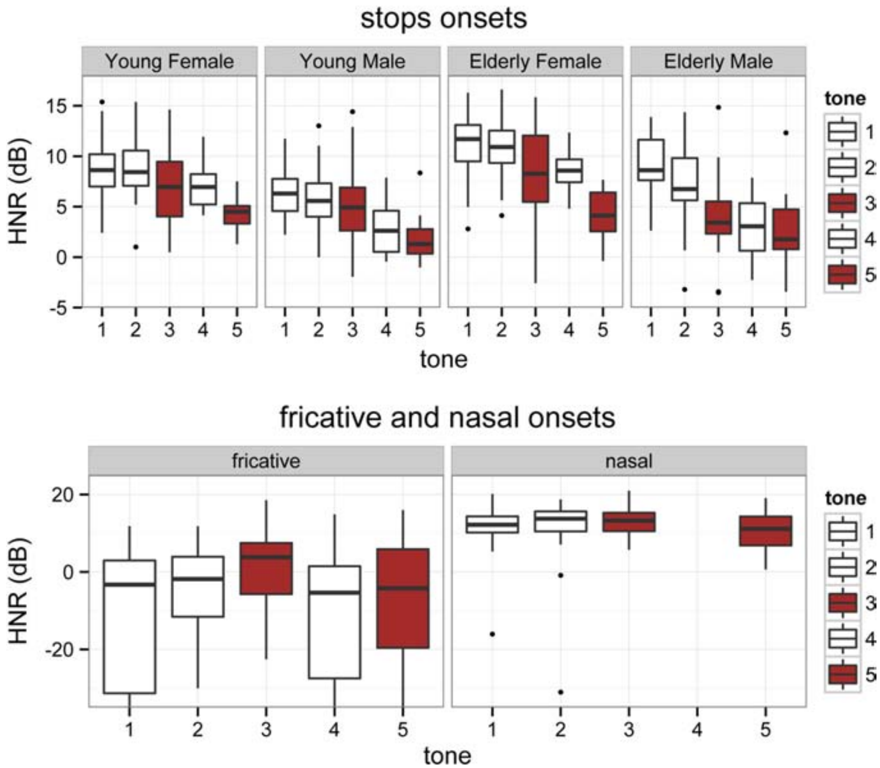

FIGURE 3 *Mean HNR (dB) in the monosyllabic context for stop onsets (top panels) according to tone and speaker group; for fricative onsets (bottom left panel) and for nasal onsets (bottom right panel) according to tone*

obstruents were phonetically voiced in this context and would necessarily yield higher HNR values than "yin" obstruents.

We report statistical analyses for stop onsets. In both ANOVA analyses (unchecked and checked syllables), *Tone* was significant overall, with lower HNR for "yang" than "yin" tones. For unchecked syllables, HNR was significantly lower for *T3* than T1 or T2 tones (4.3 < 7.6 dB), $F(2,36) = 28.7$, $p < .0001$; for checked syllables, HNR was significantly lower for *T5* than T4 (2.9 < 5.2 dB), $F(1,18) = 47.1$, $p < .0001$. For unchecked syllables, the *Tone* × *Age* interaction was significant, $F(2,36) = 6.1$, $p = 0.005$, reflecting a higher T1–2 vs. *T3* differential for elderly speakers ($\Delta = 4.9$) than for young speakers ($\Delta = 1.7$). For checked syllables, this interaction was not significant, $F(1,18) = 4.2$, $p = .055$.

Figure 4 shows H1–H2 in the vowel part of monosyllables, separately for unchecked and checked syllables, according to tone and time point, averaged across speakers and onsets. H1–H2 was higher for "yang" than "yin" tones from the beginning until around 80% into the vowel.

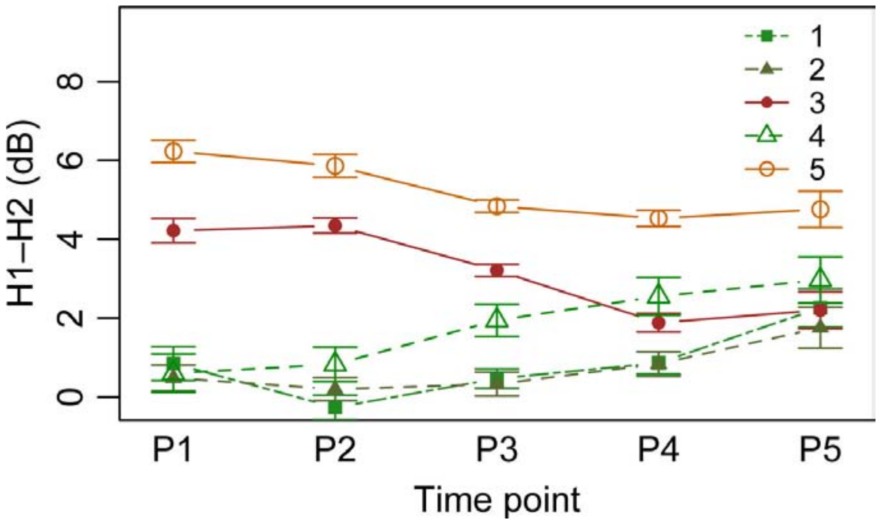

FIGURE 4 *H1–H2 in monosyllables according to tone and time point, averaged across speakers and onsets. "Muddy" (brown) colors/solid lines stand for "yang" tones.*

We applied two LME base models to fit to our H1–H2 data, separately for unchecked and checked syllables, using the *nlme* package (Pinheiro et al. 2015). The base models only included speaker random intercepts. Adding speaker-by-condition random slopes (for *Onset, Time point, Repetition*) improved the fit. *Tone*, *Onset*, *Time point*, *Repetition, Age, Gender*, and the interaction between *Tone* and all the other predictors were then step-wise added to the model. Comparison between these models showed a significant effect of *Tone* on H1–H2, for unchecked syllables, $\chi^2(8)=188.2$, $p<.0001$, and for checked syllables, $\chi^2(7)=200.5$, $p<.0001$. Contrasts further revealed that H1–H2 was significantly higher for *T3* than combined T1 and T2, 3.17 (*T3*) > 0.78 dB (T1–2), $t(2276)=13.9$, $p<.0001$, but was comparable for T1 and T2, 0.80 (T1) ≈ 0.76 dB (T2), $t(2276)=-0.4$, $p=.69$.

For both syllable types, the *Tone* × *Onset* interaction was significant (unchecked: $\chi^2(30)=38.2$, $p<.0001$; checked: $\chi^2(22)=27.2$, $p<.0001$). The effect of *Tone* was significant for all onsets, *p*s < .0001, but weaker for /m/, $\chi^2(8)=6.1$, $p=.048$, with a "yang–yin" differential of only 0.94 dB. Among the other onsets, for unchecked syllables, the "yang–yin" differential was largest for stop onsets (/p/: Δ = 3.85, /t/: Δ = 3.37), followed by zero onset (Δ = 2.36), and smallest for fricative onsets (/f/: Δ = 1.85; /s/: Δ = 1.75); for checked syllables, the "yang–yin" differential was larger for obstruent onsets (/p/: Δ = 3.74, /t/: 3.94; /f/: Δ = 4.93; /s/: Δ = 4.24) than for zero onsets (Δ = 1.39).

For both syllable types, the *Tone* × *Time point* interaction was significant as well (unchecked: $\chi^2(38)=106.6$, $p<.0001$; checked: $\chi^2(26)=72.5$, $p<.0001$), indi-

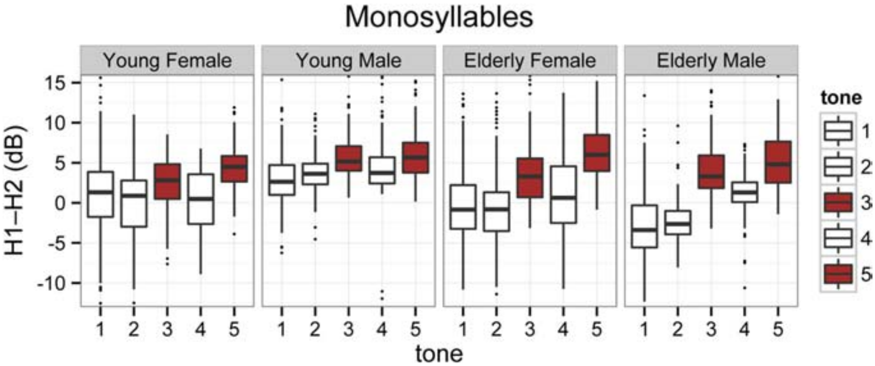

FIGURE 5 *H1–H2 in monosyllables according to tone and speaker group, averaged across speakers, onsets and time points P1–P3*

cating a change of H1–H2 difference over time, as can be observed in Figure 4. The "yang–yin" differential was largest on the first two time points and then decreased. *Tone* was significant on the first four time points of the vowel and not significant at the last, fifth time point (P5). For unchecked syllables, P1: $\Delta = 3.61$, P2: $\Delta = 4.41$, P3: $\Delta = 2.86$, P4: $\Delta = 0.83$, P5: $\Delta = -0.05$ (P1–3, *p*s < .0001; P4, $\chi^2(8) = 14.7$, $p = .0006$; P5, $\chi^2(8) = 0.94$, $p = .62$). For checked syllables, P1: $\Delta = 6.43$, P2: $\Delta = 5.46$, P3: $\Delta = 2.94$, P4: $\Delta = 1.64$, P5: $\Delta = 1.26$ (P1–3, *p*s < .0001; P4, $\chi^2(7) = 11.4$, $p = .0008$; P5, $\chi^2(7) = 2.7$, $p = .10$). Because the H1–H2 differential between "yang" and "yin" tones becomes weaker at the end of the rime, we chose to only include the first three time points in the analyses presented in the following.

The *Tone* × *Age* interaction and the *Tone* × *Gender* interaction were significant only for unchecked syllables (*Tone* × *Age*: $\chi^2(40) = 27.1$, $p < .0001$; *Tone* × *Gender*: $\chi^2(42) = 28.2$, $p < .0001$). The "yang–yin" differential was larger for elderly speakers ($\Delta = 3.32$) than for young speakers ($\Delta = 1.67$), and for male speakers ($\Delta = 3.02$) than for female speakers ($\Delta = 1.76$). Figure 5 shows H1–H2 according to tone and speaker group, averaged across speakers, onsets, and the first three time points, illustrating this between-speaker variation.

For the S1 context, H1–H2 was higher for "yang" than "yin" tones, but the difference was less clear-cut than for the monosyllable context. For the S2 context, H1–H2 appeared independent of the base tone of the syllable.

The same LME models as for the monosyllabic data were applied to the disyllabic data (S1 and S2 contexts).

Figure 6 shows H1–H2 in the vowel part of disyllabic words, according to tone, averaged across speakers, onsets, and the first three time points.

For the S1 context, the results essentially mimicked those for the monosyllable context. They showed a significant effect of *Tone* for both unchecked

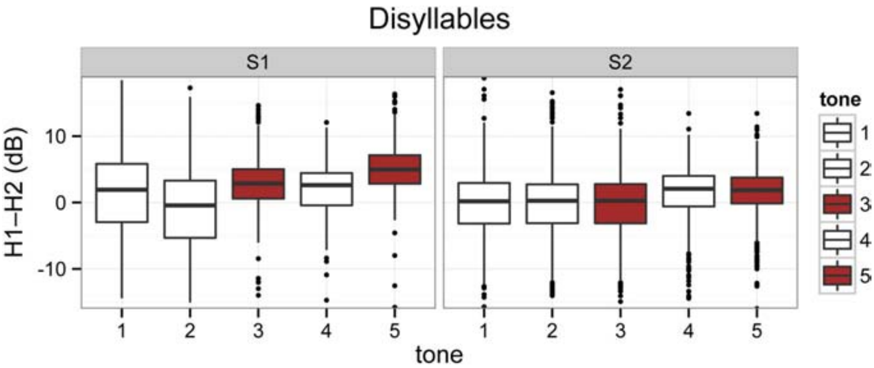

FIGURE 6 *H1–H2 in disyllables (left panel for S1 and right panel for S2 contexts) according to tone and target syllable, averaged across speakers, onsets and time points P1–P3*

syllables, $\chi^2(8)=152.6$, $p<.0001$, and checked syllables, $\chi^2(7)=103.7$, $p<.0001$. H1–H2 was higher for *T3* than combined T1 and T2, and for *T5* than T4 (unchecked tones: 2.49 (*T3*) > −0.35 dB (T1–2); checked tones: 4.61 (*T5*) > 1.79 dB (T4)). For unchecked syllables, the *Tone* × *Gender* interaction were significant, $\chi^2(30)=31.3$, $p<.0001$. However, the "yang–yin" differentials (*T3* vs. T1–2) were comparable for male ($\Delta=2.99$) and female ($\Delta=2.72$) speakers. This interaction must therefore be restricted to T1 and T2.

For the S2 context, the effect of *Tone* was not significant for unchecked syllables, $\chi^2(8)=1.90$, $p=.39$. Although *Tone* was significant for checked syllables, $\chi^2(7)=3.9$, $p=.048$, the "yang–yin" differential ($\Delta=0.27$) was quite modest.

2.2.2 EGG data

2.2.2.1 *Analyses*

We restricted the analyses of the EGG data to the monosyllabic data because they yielded the most clear-cut differences between breathy-and modal-voice productions in the acoustic data.

The thirty-two monosyllabic words described in § 2.1.2 were analyzed.

We adopted the dEGG (derivative of the EGG signal) method to detect the onset of glottal closing and opening (Henrich et al. 2004; Michaud 2005), using a semi-automatic Matlab program ("peakdet.m") written by Alexis Michaud.[8] This program first plots the EGG and (smoothed) dEGG signals, and then computes, for each glottal cycle, an F0 value based on the duration

8 Downloadable from COVAREP (A COoperative Voice Analysis REPository for speech technologies, Degottex et al. 2014): https://github.com/covarep/covarep/blob/master/glottalsource/egg/peakdet/peakdet.m.

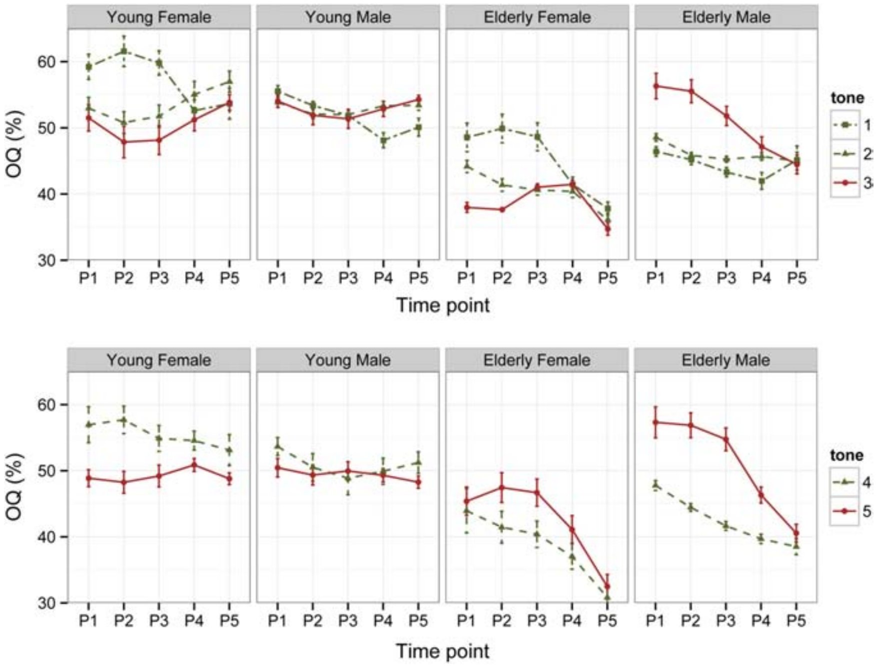

FIGURE 7 *OQ contours according to tone, time point and speaker group for unchecked (upper panel) and checked (lower panel) tones*

between two consecutive positive peaks on the dEGG signal, and finally the Open Quotient (OQ) value based on the duration between the negative and the positive peak divided by the duration of the entire glottal cycle.

2.2.2.2 *Results*

Open quotient. Figure 7 shows the OQ for the monosyllabic context according to tone, time point and speaker group. It shows higher OQ values in "yang" than "yin" syllables only for the elderly male speakers, indicating breathier phonation during the vowel after "yang" than "yin" onsets for this speaker group.

We applied the same LME models as in the previous section to evaluate the effect of each predictor, especially *Tone* and its interaction with other predictors. While the effect of *Tone* was not significant for unchecked syllables, $\chi^2(7) = 4.4$, $p = .11$, it was significant for checked syllables, $\chi^2(6) = 4.4$, $p = .037$. Both *Tone* × *Gender* interaction and *Tone* × *Age* interaction were significant (unchecked: *Tone* × *Gender*: $\chi^2(40) = 89.9$, $p < .0001$; *Tone* × *Age*: $\chi^2(38) = 35.9$, $p < .0001$; checked: *Tone* × *Gender*: $\chi^2(26) = 18.0$, $p < .0001$; *Tone* × *Age*: $\chi^2(25) = 74.6$, $p < .0001$).

For elderly male speakers, OQ was significantly higher for "yang" than "yin" tones (unchecked tones: 0.51 (*T3*) > 0.45 (T1–2), $\chi^2(7) = 45.0$, $p < .0001$; checked tones: 0.51 (*T5*) > 0.42 (T4), $\chi^2(6) = 55.8$, $p < .0001$). For both young and the elderly female speaker, OQ was, contrary to predictions, higher for "yin" than "yang" tones (young: 0.56 (T1–2) > 0.51 (*T3*), 0.55 (T4) > 0.49 (*T5*); elderly: 0.43 (T1–2) > 0.39 (*T3*)), except for checked tones for the elderly female speaker where OQ was significantly higher for *T5* (0.43) than T4 (0.39), $\chi^2(6) = 4.0$, $p = .044$. For young male speakers, OQ was similar for "yin" and "yang" tones (unchecked: 0.52 (T1–2) ≈ 0.53 (*T3*), $\chi^2(7) = 2.2$, $p = .33$; checked: 0.51 (T4) ≈ 0.49 (*T5*), $\chi^2(6) = 2.9$, $p = .09$).

We thus explored the interaction between *Tone* and the other factors only for elderly male speakers. *Tone* only interacted with *Time point*, (unchecked: $\chi^2(34) = 33.4$, checked: $\chi^2(22) = 40.5$, $ps < .0001$). Similarly to H1–H2 results, *Tone* was significant for time points P1–P4, but not P5, and the "yang–yin" differential was largest in the first half of the vowel. For unchecked syllables, P1: $\Delta = 0.08$, P2: $\Delta = 0.1$, P3: $\Delta = 0.07$, P4: $\Delta = 0.03$, P5: $\Delta = -0.01$ ($ps < .0001$ for P1–3; $p = .0045$ for P4; $p = .81$ for P5). For checked syllables, P1: $\Delta = 0.13$, P2: $\Delta = 0.16$, P3: $\Delta = 0.16$, P4: $\Delta = 0.07$, P5: $\Delta = 0.01$ ($ps < .0001$ for P1–3; $p = .0001$ for P4; $p = .51$ for P5).

The tradeoff between OQ and F0 range. Tone had no overall effect on OQ for the three young male speakers (aged 24–25), but if we look more carefully at their data, substantial inter-speaker variation can be observed in this speaker group. F0 and OQ contours of these speakers for unchecked syllables are shown in Figure 8.

A repeated-measure ANOVA was run for each male speaker, with OQ as the dependent variable, *Tone*, *Onset*, and *Time point* as within-subject factors. In particular, one of the three speakers (hereafter Speaker 1, selected recordings: Sounds 3) showed the same pattern as the elderly male speakers, that is, higher OQ for "yang" than "yin" vowels (0.56 > 0.50); another one (hereafter Speaker 2: same recordings as Sounds 2B) showed a similar pattern to that of the young female speakers, that is, significantly lower OQ for "yang" than "yin" vowels (0.48 < 0.54), $F(1,5) = 15.6$, $p = .011$; the third one (hereafter Speaker 3, selected recordings: Sounds 3) showed no significant OQ difference between "yin" and "yang" vowels (0.53 ≈ 0.54), $F(1,5) < 1$. If we look at both OQ patterns and F0 ranges, a trend for a tradeoff between F0 range and OQ seems to emerge in the young male speakers' data (but not in the other speakers' data). Speaker 1 had the smallest F0 range, and very close T2 and *T3* F0 onsets, but the highest positive "yang–yin" OQ differential: We may speculate that he produced rather weakly contrastive F0 contour information, which was compensated by clear phonation type differences between "yin" and "yang" tones. Speaker 2 had the

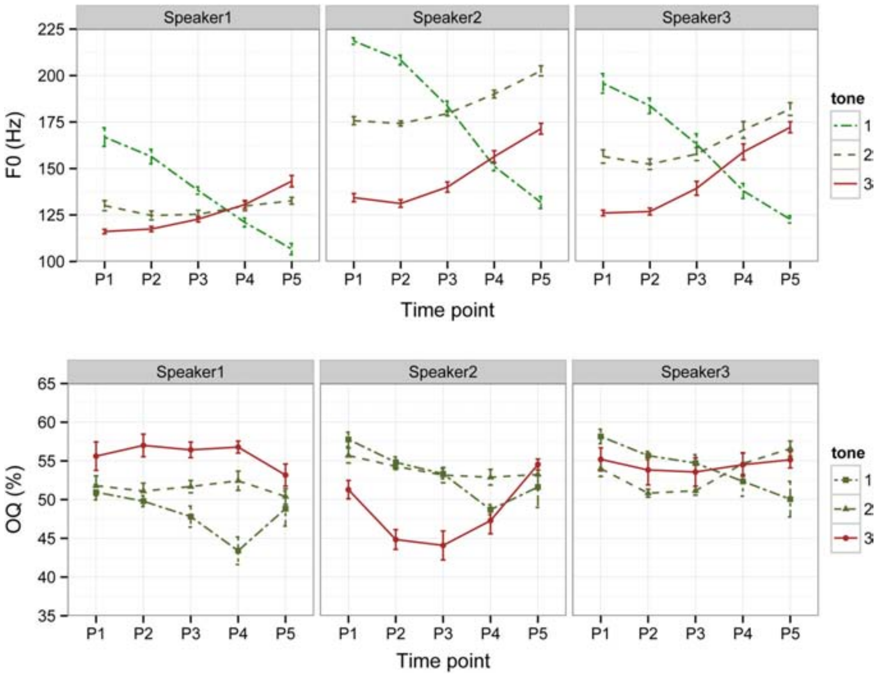

FIGURE 8 *FO (upper panel) and OQ (lower panel) contours averaged across onsets for unchecked syllables, for the three young male speakers*

largest FO range and a reversed OQ pattern compared to elderly male speakers: We may speculate that he did not have to contrast "yang" from "yin" rhymes by producing breathiness for the "yang" type. Between these two extremes, Speaker 3 had an intermediate FO range and did not exhibit OQ differences between "yin" and "yang" tones. A similar observation on FO and OQ range tradeoff has been made in Risiangku Tamang (Tibeto-Burman) spoken in Nepal (Mazaudon & Michaud 2008).

3 Discussion

3.1 *Voicing*

According to what is described in the literature, the "yang" obstruents are phonetically voiceless word-initially. Our results are in partial conformity with this description: this holds true for stops, but not always for fricatives.

Word-initial "yang" stops are realized predominantly without pre-voicing, with extremely rare exceptions, but "yin" and "yang" stops nevertheless differ in terms of VOT, most clearly in the monosyllabic context. "Yang" stops have slightly longer VOTs than "yin" stops. Many studies on Shanghai stops insisted on the positive VOT values of phonologically voiced stops but attach less importance to their VOT values relative to those of unaspirated voiceless stops. However, longer VOT values for "yang" than "yin" stops are also found in Suzhou, with a difference of 2.3–8.4 ms (Iwata et al. 1991) (but see Shi 1983, for contradictory results) and of 12 ms in Zhenhai (Rose 1982b: 268). In these Wu dialects, "yang" stops are all reported to be "breathy" or "whispery." It would thus seem that breathy phonation is correlated with somewhat longer VOTs, although to a much lesser degree in Shanghai Chinese (with a difference of only 3–4 ms), possibly due to its weaker breathiness. Similarly, in some languages with distinctive phonation differences, breathy/slack/lax stops have longer VOTs than their modal phonation counterparts, by around 20 ms for some speakers/languages and less consistently for others (Maddieson & Ladefoged 1985, for Kawa and Jingpho; Andruski & Ratliff 2000, for Green Mong). Our VOT measure for phonetically voiceless stops is equivalent to the ACT (after closure time) measure proposed by Mikukeit & Reetz (2006) to distinguish between the breathy and plain voiced stops of Indic languages, which cannot be distinguished by their negative and similar VOTs. ACT is found longer for breathy than plain voiced stops by ~25 ms in East Bengali (Mikukeit & Reetz 2006) and by ~10 ms in Marathi (Berkson 2012).[9]

Word-initial "yang" fricatives are quite often, although not systematically, phonetically realized as voiced or partially voiced. Labial fricatives are more prone to phonetic voicing than dental ones. Our data also show large inter- and intra-speaker variability in the realization of fricatives' voicing. In particular, young speakers tend to realize "yang" fricatives as phonetically voiced much more often than elderly speakers, and female speakers more often than male speakers.

Word-medial "yang" obstruent onsets are phonetically voiced as opposed to "yin" onsets, which are phonetically voiceless, as described in the literature. A new finding in our study is the occasional spirantization of word-medial voiced stops, in free variation with the full stop realization in this position.

9 Other durational measures such as NOT (noise offset time: Davis 1994) and PVI (pre-vocalic interval: Berkson 2012) have been proposed to capture the superimposed aspiration (SA, or part of it in the case of NOT) possibly occurring at the beginning of the vowel, that is, breathiness. In Shanghai Chinese, aspiration (or breathiness) is much weaker than in Indic languages, making the SA portion difficult to locate (hardly noticeable noise excitation or jump in amplitude), as shown in Figure 2D.

3.2 *Voice quality*

How might voice quality interact with onset-type? Breathy voice quality in "yang" tone syllables is most prominent after stop onsets, but is also observed after fricatives, as well as zero onsets. In fact, "yang" zero onset syllables are transcribed by some linguists with an initial [ɦ]. Moreover, a short breathy friction part distinct from the following vowel can be observed for some speakers, although this pattern is far from being systematic. Thus, breathy voice is not restricted to obstruent onset syllables. However, nasal onsets stand out as an exception. Both HNR and H1–H2 results show much smaller difference between "yin" and "yang" tones for nasal onsets. This is in line with the findings of Zhang & Yan (2015) based on acoustic analyses on the vowel part following sonorant onsets.

How does voice quality vary according to within-word syllable position? The breathy voice phonation difference between "yin" and "yang" tones is greatest in monosyllabic words. In disyllabic words, the difference is smaller when the target syllable is the first syllable; it is not conditioned at all by the underlying "yin" vs. "yang" tone register when the target syllable is the second syllable.

What is the phonetic domain of low tone breathiness? The HNR data show breathier/noisier phonation in the release part of word-initial "yang" than "yin" oral stops for both monosyllabic and disyllabic words. The H1–H2 data indicate that breathy phonation is most prominent at vowel onset and is maintained until at least the vowel's midpoint, before it decreases in the second half of the vowel. Hence, phonetically, breathy phonation is not limited to the consonant release and vowel onset parts, as proposed by Cao & Maddieson (1992) and Ren (1992).

It should be noted that, for our female speakers, OQ and H1–H2 data were not positively correlated, as has often been assumed it should (Klatt & Klatt 1990). H1–H2 is slightly higher for "yang" than "yin" syllables for these speakers, whereas OQ exhibits the reverse pattern. The EGG signal, however, could be analyzed for only four out of the 12 female speakers, among which a single elderly female speaker. (The EGG signal was too noisy for the remaining speakers; see Holmberg et al. 1995 for difficulties in recording EGG with female speakers.) The discrepancy between EGG and H1–H2 data for the female speaker groups might thus be at least partly due to the different female speaker samples. Another possible explanation for this discrepancy is the positive correlation between F0 and OQ that has been occasionally observed, for example in Dutch, a language in which voice quality is not contrastive (Koreman 1996). Our data are not sufficient to clarify this issue. We may conclude, however, that whereas H1–H2 indicated breathier "yang" than "yin" syllables for female speakers – this difference being weak for young female speakers – the OQ data of part of these female speakers were not in line with their H1–H2 data.

3.3 *Multiple cues, cue switching, and evolution*

Finally, we come to the central question: *How can we define the relationship between (phonologically) voiced onsets and low tones on one hand, and between voiced onsets and breathy voice on the other hand? Are these relationships phonetic or phonological?*

In Shanghai Chinese, pitch is undoubtedly the primary cue to tone in a prosodic word, as described in impressionistic and experimental reports (Chao 1928; Xu & Tang 1988; Chen & Gussenhoven 2015), but also as shown by the distinctive role of pitch in tone perception (Gao & Hallé 2013). However, the "yin" or "yang" identity is also maintained at the syllable level. When pitch is neutralized, that is, in word-medial position, this identity is expressed by onset phonetic voicing, except in zero and sonorant onsets. Thus, the primary cue to the "yin" vs. "yang" identity switches between pitch and phonetic voicing according to position within a prosodic word. Hence, the relationship between low pitch and onset voicing is phonological, both being the primary cue with a contextual cue switching.

Breathiness and onset phonetic voicing function as secondary cues to low tone in the initial syllable of a prosodic word, where pitch is primary. In this context, the phonetic voicing of fricative onsets occurs in free variation in low tone syllables, more frequently for labial than dental fricatives, and with young than elderly speakers.

The phonetic voicing of fricative onsets in low tone syllables is occasional, and we do not have a good explanation for it. Breathiness, however, is a redundant cue that has existed in Shanghai Chinese probably for about one thousand years. Breathiness contributes to the perception of low tone (Gao & Hallé 2015), and thus functions as an enhancing feature (Stevens & Keyser 2010). This is probably the reason why it has survived for such a long time in Shanghai Chinese. Note, however, that many Chinese dialects, including Cantonese, in which original low tones are still phonetically low, have lost breathiness.

Our study, comparing the acoustic and physiological data of today's elderly and young speakers, showed an ongoing disappearance of low tone breathiness. If low tone breathiness, as a redundant cue, is doomed to disappear, as it already did in most Chinese dialects, we speculate that the disappearance is accelerated by the impact of the promotion of Standard Chinese. Standard Chinese can be considered a "pure" tone language, where pitch is by far the most important cue and breathiness has no linguistic role. Our Shanghai young speakers, who all received school education in Standard Chinese, frequently use this prestigious variety at work and even with their friends and family. They generally feel the impact of Standard Chinese on Shanghai Chinese and judge themselves as speaking a "degraded" dialect. Some linguists make this same judgment on young speakers' performance. Zhu Xiaonong (2006: 2) once commented, "Quite a lot of young Shanghai people even cannot speak their native language properly." According to a survey conducted by Gao (2016), young speakers' self-evaluation of linguistic competence and frequency of usage is overall lower for Shanghai Chinese than Standard Chinese, while elderly speakers have generally a high evaluation of linguistic competence and frequency of usage in Shanghai Chinese.[10] This same study also shows a lower evaluation in Shanghai Chinese for female than male speakers, which is in line with previous sociolinguistic findings reporting women conform to prestigious linguistic forms more than men (Labov 1966: 313; Shuy et al. 1967). The results of the present study, showing a more advanced loss of low tone breathiness in female than male speakers, might thus be related to female speakers' stronger conformity to prestigious Standard Chinese.

Acknowledgements

We express our gratitude to Martine Mazaudon and Alexis Michaud for fruitful discussions, as well as the two anonymous reviewers for their constructive suggestions on an earlier draft of this paper. This study was funded by a subvention from LabEx EFL (Strand 1), a PhD student fellowship to the first author, and a JSPS Bridge fellowship to the second author.

References

Andruski, Jean E. & Ratliff, Martha. 2000. Phonation types in production of phonological tone: The case of Green Mong. *Journal of the International Phonetic Association* 30(1–2). 37–61.

10 Shanghai Chinese speakers rated their linguistic competence in, and frequency of usage of Shanghai vs. Standard Chinese on a 1–5 scale. Seven out of 12 young speakers rated their linguistic competence higher for Standard than Shanghai Chinese, and 6 out of 12 young speakers rated their frequency of usage higher for Standard than Shanghai Chinese. All of the 10 elderly speakers rated their linguistic competence and frequency of usage higher for Shanghai than Standard Chinese.

Berkson, Kelly Harper. 2012. Capturing breathy voice: Durational measures of oral stops in Marathi. *Kansas Working Papers in Linguistics* 33. 27–46.

Boersma, Paul. 2001. Praat, a system for doing phonetics by computer. *Glot International* 5(9/10). 341–345.

Cao, Jianfen & Maddieson, Ian. 1992. An exploration of phonation types in Wu dialects of Chinese. *Journal of Phonetics* 20. 77–92.

Chao, Yuen Ren (赵元任). 1928. *Xiàndài wúyǔ de yánjiū* (现代吴语的研究) [Studies in the modern Wu dialects]. Peking: Tsinghua College Research Institute.

Chao, Yuen Ren. 1930. A system of "tone-letters." *Le Maître Phonétique* 45. 24–27.

Chen, Yiya. 2011. How does phonology guide phonetics in segment-f_0 interaction? *Journal of Phonetics* 39(4). 612–625.

Chen, Yiya & Gussenhoven, Carlos. 2015. Shanghai Chinese. *Journal of the International Phonetic Association* 45(3). 321–337.

Chen, Zhongmin (陈忠敏). 2010. *Wūyǔ qīngyīn zhuóliú de shēngxué tèzhēng jí jiàndìng biāozhì: Yǐ shànghǎihuà wéilì* (吴语清音浊流的声学特征及鉴定标志: 以上海话为例) [An acoustic study of voiceless onset followed by breathiness of Wu dialects: based on the Shanghai dialect]. *Yǔyán Yánjiū* (语言研究) [Linguistic studies] 30(3). 20–34.

Colquhoun, David. 2014. An investigation of the false discovery rate and the misinterpretation of *p*-values. *Royal Society Open Science* 1. 1–16.

Davis, Katharine. (1994). Stop voicing in Hindi. *Journal of Phonetics* 22. 177–193.

Degottex, Gille & Kane, John & Drugman, Thomas & Raitio, Tuomo & Scherer, Stefan. 2014. COVAREP: A collaborative voice analysis repository for speech technologies. *2014 IEEE International Conference on Acoustics, Speech and Signal Processing* (*ICASSP*), 960–964.

Fischer-Jørgensen, Eli. 1967. Phonetic analysis of breathy (murmured) vowels in Gujarati. *Indian Linguistics* 28. 71–139.

Gao, Jiayin. 2016. Sociolinguistics motivations in sound change: On-going loss of low tone breathy voice in Shanghai Chinese. *Papers in Historical Phonology* 1. 166–186.

Gao, Jiayin & Hallé, Pierre. 2013. Duration as a secondary cue for perception of voicing and tone in Shanghai Chinese. *Proceedings of the Annual Conference of the International Speech Communication Association*, 3157–3161.

Gao, Jiayin & Hallé, Pierre. 2015. The role of voice quality in Shanghai tone perception. *Proceedings of the 18th International Congress of Phonetic Sciences*, Paper number 448. 1–5.

Gao, Jiayin & Hallé, Pierre & Honda, Kiyoshi & Maeda, Shinji & Toda, Martine. 2011. Shanghai slack voice: Acoustic and EPGG data. *Proceedings of the 17th International Congress of Phonetic Sciences*, 719–722.

Gordon, Matthew & Ladefoged, Peter. 2001. Phonation types: A cross-linguistic overview. *Journal of Phonetics* 29(4). 383–406.

Hanson, Helen M. 1997. Glottal characteristics of female speakers: Acoustic correlates. *The Journal of the Acoustical Society of America* 101. 466–481.

Hanson, Helen M. & Chuang, Erika S. 1999. Glottal characteristics of male speakers: Acoustic correlates and comparison with female data. *The Journal of the Acoustical Society of America* 106(2). 1064–1077.

't Hart, Johan. 1981. Differential sensitivity to pitch distance, particularly in speech. *The Journal of the Acoustical Society of America* 69(3). 811–821.

Henrich, Nathalie & d'Alessandro, Christophe & Doval, Boris & Castellengo, Michèle. 2004. On the use of the derivative of electroglottographic signals for characterization of nonpathological phonation. *The Journal of the Acoustical Society of America* 115(3). 1321–1332.

Holmberg, Eva B. & Hillman, Robert E. & Perkell, Joseph S. & Guiod, Peter C. & Goldman, Susan L. 1995. Comparisons among aerodynamic, electroglottographic, and acoustic spectral measures of female voice. *Journal of Speech, Language, and Hearing Research* 38(6). 1212–1223.

Iwata, Ray & Hirose, Hajime & Niimi, Seiji & Horiguchi, Satoshi. 1991. Physiological properties of "breathy" phonation in a Chinese dialect: A fiberoptic and electromyographic study on Suzhou dialect. *Proceedings of the 12th International Conference on Phonetic Sciences*. 162–165.

Karlgren, Berhard. 1926. *Études sur la phonologie chinoise*. Leiden: Brill.

Klatt, Dennis H. & Klatt, Laura C. 1990. Analysis, synthesis, and perception of voice quality variations among female and male talkers. *The Journal of the Acoustical Society of America* 87(2). 820–857.

Koreman, Jacobus Johannes Maria. 1996. *Decoding linguistic information in the glottal airflow*. Nijmegen: University of Nijmegen. (Doctoral dissertation.)

Labov, William. 1966. *The social stratification of English in New York City*. Washington, D.C.: Center for Applied Linguistics.

Li, Fang-Kuei. 1937. Languages and dialects of China. *The Chinese yearbook* 1. 1–13.

Liu, Fu. 1925. *Étude expérimentale sur les tons du chinois*. Paris: Les Belles Lettres.

Maddieson, Ian & Ladefoged, Peter. 1985. "Tense" and "lax" in four minority languages of China. *UCLA Working papers in Phonetics* 60. 59–83.

Mazaudon, Martine & Michaud, Alexis. 2008. Tonal contrasts and initial consonants: a case study of Tamang, a 'missing link' in tonogenesis. *Phonetica* 65(4). 231–256.

Michaud, Alexis. 2005. *Prosodie de langues à tons (naxi et vietnamien), prosodie de l'anglais: éclairages croisés*. Paris: Université Sorbonne Nouvelle – Paris 3. (Doctoral dissertation.)

Mikuteit, Simone & Reetz, Henning. 2007. Caught in the ACT: The timing of aspiration and voicing in East Bengali. *Language and Speech* 50, 247–277.

Pinheiro, Jose & Bates, Douglas & DebRoy, Saikat & Sarkar, Deepayan & R Development Core Team. 2015. nlme: Linear and nonlinear mixed effects models. *R package version 3.1–120*.

Qian, Nairong (钱乃荣). 2003. *Shànghǎi yǔyán fāzhǎnshǐ* (上海语言发展史) [A history of the language of Shanghai]. Shanghai: Shanghai Renmin Publishing House.

Ramsey, S. Robert. 1987. *The languages of China*. Princeton University Press.

Ren, Nianqi. 1988. A fiberoptic and transillumination study of Shanghai stops. Paper presented at International Conference on Wu Dialects. Hong Kong, 11–14 December 1988.

Ren, Nianqi. 1992. An acoustic study of Shanghai stops. Storrs: University of Connecticut. (Doctoral dissertation.)

Rose, Philip. 1982a. Acoustic characteristics of the Shanghai-Zhenhai syllable-types. In Bradley, David (ed.). *Papers in Southeast Asian Linguistic, No. 8: Tonation* (*Pacific Linguistics, Series A, N° 62*), 1–53. Canberra: Australian National University.

Rose, Philip. 1982b. *An acoustically based phonetic description of the syllable in the Zhenhai dialect*. Cambridge: University of Cambridge. (Doctoral dissertation.)

Rose, Philip. 1989. Phonetics and phonology of Yang tone: Phonation types in Zhenhai. *Cahiers de Linguistique Asie Orientale* 18(2). 229–245.

Rose, Philip. 2002. Independent depressor and register effects in Wu dialect tonology. *Journal of Chinese Linguistics* 30(1). 39–81.

Shen, Zhongwei & Wooters, Charles & Wang, William S.Y. 1987. Closure duration in the classification of stops: A statistical analysis. *Ohio State University Working Papers in Linguistics* 35. 197–209

Sherard, Michael. 1972. Shanghai phonology. Ithaca: Cornell University. (Doctoral dissertation.)

Shi, Feng (石锋). 1983. *Sūzhōuhuà zhuó sèyīn de shēngxué tèzhēng* (苏州话浊塞音的声学特征) [Acoustic features of muddy initials in Suzhou dialect]. *Yǔyán Yánjiū* (语言研究) 1. 49–83.

Shuy, Roger W. & Wolfram, Walt & Riley, William K. 1967. *A study of social dialects in Detroit*. Washington, D.C.

Stevens, Kenneth Noble & Keyser, Samuel Jay. 2010. Quantal theory, enhancement and overlap. *Journal of Phonetics* 38(1). 10–19.

Xu, Baohua (许宝华) & Tang, Zhenzhou (汤珍珠). 1988. *Shànghǎi shìqū fāngyánzhì* (上海市区方言志) [Urban Shanghai dialect records]. Shanghai: Shanghai Education Publication House.

Yip, Moira. 1980. The tonal phonology of Chinese. Cambridge: MIT. (Doctoral dissertation.)

Zee, Eric & Maddieson, Ian. 1979. *Tones and tone sandhi in Shanghai: Phonetic evidence and phonological analysis*. UCLA *Working papers in Phonetics* 45. 93–129.

Zhang, Jie & Yan, Hanbo. 2015. Contextually dependent cue weighting for a laryngeal contrast in Shanghai Wu. *Proceedings of the 18th International Congress of Phonetic Sciences*, Paper number 362. 1–5.

Zhu, Xiaonong. 2006. *A grammar of Shanghai Wu*. Munich: Lincom Europa.