SHANGHAI SLACK VOICE: ACOUSTIC AND EPGG DATA

J-Y. Gao[a], P. Hallé[a,b], K. Honda[a], S. Maeda[c] & M. Toda[a]

[a]LPP-Paris 3, France; [b]LPNCog-Paris 5, France; [c]ENST Paris, France

jiayin.gao@gmail.com; pierre.halle@parisdescartes.fr; khonda@sannet.ne.jp;
maeda@tsi.enst.fr; martinetoda@yahoo.co.jp

ABSTRACT

From a representational viewpoint, the "voiced" series of obstruents in Shanghai dialect can be specified in terms of complementary, context-conditioned tonal and segmental features: either low tone or glottal pulsing. Yet, some studies have proposed that, when the "voiced" obstruents can only be signaled by low tone (stress-initially), they retain something of segmental voicedness. This somewhat mysterious "something" has often been identified to a moderate degree of breathiness after stop release, or "slack voice." In this study we revisit this issue and find that Shanghai obstruents, as produced today by young Shanghai people, indeed retain some characteristics of plain voiced obstruents but breathiness does not appear as the sole one. We propose that articulatory timing relationships are the main determinant to the mysterious voiced quality of Shanghai obstruents.

Keywords: Shanghai slack voice, low vs. high tone, acoustic/physiological breathiness, EPGG

1. INTRODUCTION

Four tones are assumed in early Middle Chinese, traditionally labeled 平, 上, 去, 入 (*ping, shang, qu, ru*). Middle Chinese distinguished two series of obstruents, described as 浊音 (*zhuoyin* 'muddy voice') and 清音 (*qingyin* 'clear voice') in the Chinese linguistic tradition, probably voiceless and voiced, respectively. Segmental tonogenesis led to the general Middle Chinese "tonal split" into 阴 vs. 阳 tones (yin vs. yang: 'low' vs. 'high' register tones), from the clear vs. muddy series. The clear-muddy distinction thus became redundant with pitch register, motivating the disappearance in most late Middle Chinese dialects of the 'muddy' obstruents, replaced with voiceless aspirated or non-aspirated ones of the remaining 'clear' series. The retention of the 'muddy' obstruents –as well as their associated low, yang tones– is traditionally taken as a defining feature of the Wu dialects. Indeed, all the dialects believed to belong to the Wu family have a phonologically voiced series in addition to two voiceless series. Yet, some southern dialects classified as Wu on that criterion clearly possess features typical of the Min dialects.

The phonetic characterization of 'muddy' obstruents in Wu dialects, as well as cross-dialectal differences in this respect, are controversial. Early impressionistic descriptions simply assumed plain voicing, that is glottal pulsing throughout the production of these obstruents. In the late twenties, a more detailed account was proposed by Liu Fu and Chao Yuanren [4, 8] for Shanghai dialect among others: the closure of the 'muddy' stops is voiceless but their release is voiced and breathy. This proposal, known by Chinese linguists as 清音浊流 (*qingyin zhuoliu*: 'clear sound then muddy aspiration'), motivates transcriptions such as [t$^\text{ɦ}$]. However, this description may hold only for stress-initial stops in the dialects examined so far, less clearly perhaps for southern than northern dialects (e.g., Whenzhou vs. Shanghai dialects [3]). In non-initial, unstressed syllables, muddy stops tend to be fully voiced in Shanghai dialect among others ([2, 3]; also see [10]). More precisely, muddy stops in this context either are plain voiced stops but their syllable loses its yang tone (due to pervasive tone sandhi in Wu dialects) or are voiceless but yang tone is retained. A possible phonological account of "muddiness" is thus that muddy stops contrast with the others by either a segmental [+voice] or a suprasegmental [-high tone] feature, depending on stress context. By this account, muddy stops need not differ segmentally from voiceless unaspirated stops in stress-initial position. Yet, even then, muddy stops are felt to retain aspects of phonetic voicedness. Breathy phonation has been proposed as one such aspect in northern Wu dialects such as Shanghai (cf. [3] for a review); whispery phonation has been proposed for Zhenhai dialect [11]. The specific phonation of these stops is also called "slack voice," suggesting a loose tension/adduction of the vocal folds. Studies conducted in the late eighties found both acoustic and physiological cues to breathiness in the release portion of muddy stops [3, 9, 10]: H1 relative salience [1, 6], oral airflow, and glottal opening as measured by fiberoptic transillumination [11]. The

degree of breathiness revealed by these studies is however far from that found in, for example, Hindi. Recently, Chinese linguists proposed that moderate breathiness was a feature attached to the entire syllable or its rime rather than to the onset consonant, but this view is still debated [3, 5, 13]. Our study addresses that issue in using zero onset and nasal onsets.

In this paper, we focus on the Shanghai dialect, as spoken by young, educated native speakers. The literature on Wu "slack voice" has exclusively focused, as far as we know, on oral stop syllable onsets. In this paper, we also examine fricatives (/f, v, s, z/). Because nasal and zero onsets may bear either yang or yin tone, we also examine yin-yang pairs with nasal and zero onsets: were slack voice characterizing the entire syllable, it should be found in nasal or zero onset yang but not yin tone syllables. (Zero onset yang tone syllables may be transcribed with an /ɦ/ whose motivation is morphophonemic rather than phonetic as in 雨 /ɦy/ [y] 'rain'.) Our report on the production of slack voice syllables covers acoustic measurements such as H1–H2, and glottal opening estimations obtained with EPGG, a novel technique similar to photoglottography but with an external lightening source [7].

2. EXPERIMENT

The main goal of the experiment was to determine the time course of glottal opening before and after the onset consonants under scrutiny, along with possibly related acoustic measurements such as H1–H2. Several syllable pairs differing in tone (yin vs. yang) and onset (nasal onsets excepted) were compared. We added voiceless aspirated stops for sake of comparison with previous studies.

2.1. Method

2.1.1. Participant

The first author, a young woman aged 23 years, native speaker of Shanghai dialect, raised in Shanghainese-speaking family environment, was recorded on the speech materials.

2.1.2. Speech materials

Sixteen syllables, sharing the rime /ɛ/, in sentence-initial position within the frame sentence *X gə ə zi / ŋo nintə ə* ('X', this character, I know) were produced five times in succession to ensure ease of production. The onsets were /p, t, f, s, m, n, Ø, pʰ, tʰ/ (yin tone) and /b, d, v, z, m, n, ɦ/ (yang tone).

2.1.3. Procedure and apparatus

The session was conducted in a soundproof booth, using a Dash 8 multi-channel data acquisition device recording three channels: audio, EPGG, and oral airflow, all sampled at 20 kHz with 16 bit resolution. The EPGG and airflow channels were low-pass filtered at 500 and 80 Hz, respectively. The signals were transferred to computer as wave files, segmented into utterances and processed.

2.1.4. Physiological analyses

For any given syllable, five EPGG signals (only three for /pɛ, tɛ, pʰɛ, tʰɛ/) were lined up on the onset of /ɛ/, then averaged together; 600 ms before line-up point and 200 ms after sentence offset were included in those averages so that the physiological activity before and after the target syllable could be tracked. For a large number of syllables, the first repetition exhibited extra-wide glottal opening, presumably due to the speaker taking her breath to produce the 5 repetitions, and thus was not retained in the averaging process.

Table 1: The 16 syllables investigated.

pɛ	tɛ	fɛ	sɛ	mɛ	nɛ	ɛ	pʰɛ	tʰɛ
背	呆	翻	三	蛮	拿	爱	配	摊
bɛ	dɛ	vɛ	zɛ	mɛ	nɛ	ɦɛ		
办	蛋	烦	才	慢	难	害		

We estimated the location (relative to vowel onset) and amplitude (arbitrary unit) of the glottal opening maximum ahead of the target syllable for each utterance individually. Averages of these measurements are presented in the following.

2.1.5. Acoustic analyses

For each target syllable, the following acoustic parameters were measured: VOT (stop onsets), onset duration (onsets other than stops and zero), vowel duration, H1 and H2 amplitudes, hence the H1–H2 difference, and harmonic to noise ratio (HNR). H1–H2 and HNR were computed on a 30 ms window at vowel /ɛ/'s onset, middle, and offset. For /pʰɛ/ and /tʰɛ/, only VOT was measured.

H1-H2 values are usually taken as a cue to loose setting of vocal folds and/or breathiness; lower HNR values indicate noisier speech.

2.2. Results

2.2.1. H1–H2 and HNR

Table 2 shows the differences in H1–H2 between yang and yin tone paired syllables (e.g., /bɛ/ and

720

/pɛ/). They are all significantly positive at vowel onset, except for /zɛ/-/sɛ/, indicating that yang tone is breathier. The yang tone advantage reduces slightly but is still significant at vowel middle; it is not observed at vowel offset.

Table 2: Differences in H1–H2 between yang and yin tone paired syllables at three locations; significance: * for p<.05, ** for p<.01.

Pair	V onset	V middle	V offset
bɛ-pɛ	7.45**	5.85**	0.30
dɛ-tɛ	5.01**	3.96**	0.11
mɛ$_L$- mɛ$_H$	3.87**	1.90**	0.34
nɛ$_L$- nɛ$_H$	2.43**	2.82**	0.04
ɦɛ-ɛ	8.22**	1.04**	2.21
vɛ-fɛ	3.85**	4.86**	-2.01
zɛ-sɛ	-0.86**	1.76**	-5.83
means	4.28**	3.17**	-0.69

HNR is significantly higher (p<.01) for yin than yang tone syllables at vowel onset (14.3>10.3 dB), except for the "zero" onset pair (suggesting /ɦ/ in /ɦɛ/ tends to be realized [ɦ]). Yin tone advantage decays at vowel middle (11.0≈8.7 dB) and tends to reverse at vowel offset (11.1≈14.8 dB).

2.2.2. Voice onset times

There was no prevoicing for "voiced" stops. Their VOT was similar to that of voiceless unaspirated stops: ~10 and ~20 ms for labial and dental stops, respectively. Voiceless aspirated stops had much longer VOTs (/pʰ/: 117 ms, / tʰ/: 101 ms).

2.2.3. Durations

Table 3 shows the onset consonant and total duration of target syllables with fricative or nasal onset. Total duration is similar for yin and yang tone syllables but onset duration is dramatically shorter for yang tone syllables (*p*<.0001)

Table 3: Yin vs. yang tone syllables' onset consonant duration and whole syllable durations (ms).

Syllable	yin		yang	
	onset	syllable	onset	syllable
mɛ$_L$- mɛ$_H$	78	300	55	296
nɛ$_L$- nɛ$_H$	86	308	65	335
fɛ-vɛ	148	367	74	360
sɛ-zɛ	193	400	113	401
means	126	344	77	348

2.2.4. Glottal opening

Table 4 shows the estimated location relative to /ɛ/ onset and the amplitude of the peak of glottal opening preceding target syllable. This peak occurs earlier (148 vs. 93 ms before /ɛ/ onset, *p*<.001), and its amplitude is lower for yang than yin tone

syllables. For /pʰɛ, tʰɛ/, a larger peak (0.49) occurs about 50-100 ms before /ɛ/ onset.

Table 4: Peak of glottal opening before target Cɛ: location relative to /ɛ/ onset (ms) and amplitude (arbitrary unit), for non-aspirated C.

Syllable	Peak Location		Peak Amplitude	
Pair	yang	yin	yang	yin
bɛ-pɛ	-169	-116	0.32	0.35
dɛ-tɛ	-164	-87	0.35	0.52
mɛ$_L$- mɛ$_H$	-180	-82	0.33	0.42
nɛ$_L$- nɛ$_H$	-152	-100	0.33	0.41
vɛ-fɛ	-111	+12	0.32	0.34
zɛ-sɛ	-165	-71	0.34	0.33
ɦɛ-ɛ	-92	-204	0.39	0.45
means	-148	-93	0.34	0.40

Figure 1 shows EPGG data for /fɛ/ and /vɛ/.

Figure 1A-D: EPGG data for /fɛ/ and /vɛ/ sentences; the thick lines are averages for individual repetitions lined-up on /ɛ/ onset (dotted vertical lines); A-B: examples of speech waveform for (A) /fɛ/ and (B) /vɛ/; C-D: EPGG data for (A) /fɛ/ and (B) /vɛ/.
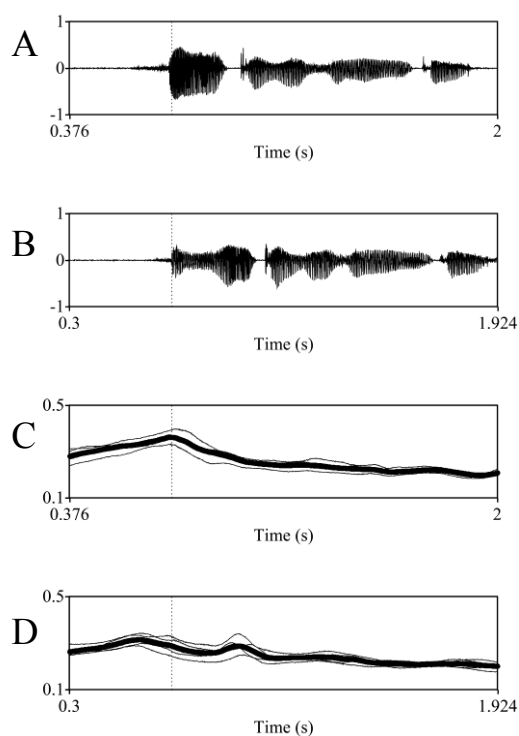
3. DISCUSSION

In this study, we examined both physiological and acoustic characteristics that possibly accompany the so-called "slack voice" that characterize the historically voiced series of yang tone obstruents in Shanghai dialect, and perhaps also generally apply to *all* yang tone syllable onsets: we thus extended the investigation to nasal and zero onset syllables.

On the acoustic side, our results, although preliminary inasmuch they are limited to a single speaker, largely confirm the previous findings of breathiness, not only in the "voiced" series of obstruents but also in yang tone syllables with a nasal or zero onset. In our data, the most salient index of breathiness was the H1–H2 amplitude difference: it was larger across the board for yang tone syllables at vowel onset, signaling the smoother glottal pulsing typical of breathiness. This index of breathiness, however was weaker inside the syllable, and had disappeared at syllable offset. Our H1-H2 data are therefore intermediate between those of Cao and Maddieson [3] and Chen [5], who found breathiness extremely transient versus pervading the entire syllable, respectively. A stronger support to the latter opinion from our data is that breathiness is found even in syllables with a sonorant onset. Thus, breathiness might be attached to the entire syllable rather than just its segmental onset. Our HNR data are less consistent but show some trend toward noisier vowel onsets in yang than yin tone. Yet, moderate breathiness does not necessarily produce much turbulent noise. Apart from breathiness, a robust acoustic difference between yin and yang tone syllables lies in their consonant onset durations. They are consistently shorter for yang than yin tone, even for /m/ and /n/. The difference is spectacular for the fricative onsets, with /s, f/ about twice as long as /z, v/. For stops, we could not of course measure closure durations. But we may suspect they are shorter for yang tone syllables, as has been reported in a previous study for both word-medial and word-initial Shanghai stops [12]. On the basis of our data and [12], we therefore propose that the perceived voiced quality of Shanghai "slack voice" obstruents might be due to within-syllable timing relationships between onset and rime favoring the rime, that would universally tend to convey a voiced percept.

How do these results fare with the physiological data? The glottal opening timing patterns are rather consistent across all onset types. As a rule, the glottal opening-closing gesture is phased earlier relative to vowel onset for yang than yin tone syllables (except /ε/</ɦɛ/). Its peak amplitude is slightly but consistently lower for yang than yin tone syllables. These findings run contrary to those of Ren [2] who found larger and earlier peaks of glottal opening for voiceless unaspirated than "voiced" stops. However, glottal opening is phased with vowel very similarly in Ren's data and ours for aspirated vs. unaspirated voiceless stops. The two sets of data differ with respect to slack voice onsets only: in Ren's data, glottal opening peaked around release for voiced stops. Ren wondered "How can voicing start so early in initial position" [with such a late-occurring glottal opening] [2]? In contrast, our data are not puzzling. The relatively early offset of the glottal opening-closing gesture in yang tone allows for a more complex laryngeal setting for yang than yin tone syllables. In sum, this "yang" setting may be common to all types of onset and be laryngeal in nature, rather than segmental and restricted to one type of segments: obstruents. Early occurring glottal opening-closing gesture, together with the short onset syllable timing, also suggest this gesture offset tends to be phased with vowel onset for yin syllables and with consonant onset for yang syllables. Thus, slack voice in yang tone is characterized, we believe, not only by a specific kind of laryngeal maneuver but also by specific segmental phasing relationships.

4. REFERENCES

[1] Bickley, C. 1982. Acoustic analysis and perception of breathy vowels. *MIT Speech Communication Working Papers* 1, 71-81.

[2] Cao, J.F. 1987. The ancient initial "voiced" consonants in modern Wu dialects. *Proc. 11th ICPhS* Tallinn, 169-172.

[3] Cao, J.F., Maddieson, I. 1992. An exploration of phonation types in Wu dialects of Chinese. *J. Phon.* 20, 77-92.

[4] Chao, Y.R. 1928. *Studies in the Wu Dialects*. Beijing: Qinghua University Press.

[5] Chen, Z.M. 2010. An acoustic study of voiceless onset followed by breathiness of Wu dialects: Based on the Shanghai dialect. *Yuyan Yanjiu* 30(3), 20-34.

[6] Fant, G. 1983. Preliminaries to analysis of the human voice source. *Quat. Progress Report* Speech Trans. Lab., Royal Institute of Technology, 4, 1-27.

[7] Honda, K., Maeda, S. 2008. Glottal-opening and airflow pattern during production of voiceless fricatives: A new non-invasive instrumentation. *J. Acoust. Soc. America* 123, 3738.

[8] Liu, F. 1925. Étude Expérimentale sur le tons du Chinois. *Collection de L'Institut de Phonétique et des Archives de la Parole, Fascicule 1.* Paris: Société d'Édition.

[9] Ren, N.Q. 1987. *An Acoustic Study of Shanghai Stops.* Unpublished Ph.D. Dissertation, Univ. of Connecticut.

[10] Ren, N.Q. 1988. A fiberoptic and transillumination study of Shanghai stops. *Int. Conf. on Wu Dialects* Hong Kong.

[11] Rose, P. 1989. Phonetics and phonology of yang tone phonation types in Zhenhai. *CLAO* 58, 229-245.

[12] Shen, Z.W., Wooters, C., Wang, W.S.Y. 1987. Closure duration in the classification of stops: A statistical analysis. *Ohio State Univ. Working Papers in Linguistics* 35, 197-209.

[13] Zhu, X.N. 2010. Slack voice: the phonetic nature of the voiced consonants in ancient Chinese. *Yuyan Yanjiu* 30(3), 1-19.