Articulation of the nasal vowels in Shanghai Chinese

Changhe Chen, Jonathan Havenhill

Department of Linguistics, The University of Hong Kong
riverch8@connect.hku.hk, jhavenhill@hku.hk

ABSTRACT

This paper presents an articulatory and acoustic study of the nasal vowels [ãᵑ ɒ̃ᵑ] in Shanghainese. Synchronous ultrasound, audio, nasalance, and EGG recordings from 4 young native speakers were analyzed. Data show that the two nasal vowels in Shanghainese are merged and generally produced with a retracted and lower tongue position compared with the oral vowel [a]. The nasal vowels are significantly nasalized from about half of the duration. Unlike previous studies of Chaoshan Chinese, French, and Yi, a decrease in contact quotient during nasal vowels was not observed. Rather, one speaker is observed to increase the contact quotient (less breathy) for nasal vowels carrying modal rising tone [˩˧]. Given the acoustic similarities of nasality and breathiness, this adjustment may serve to avoid misperception with the breathy rising tone [˨˧]. This study betters our understanding of the nasal vowel articulation in Shanghainese and provides additional data required for cross-linguistic comparisons.

Keywords: nasal vowel, Shanghai Chinese, ultrasound, EGG, nasalance.

1. INTRODUCTION

Existing studies have shown that nasal vowels are both acoustically complex and multidimensional in articulation [1]. Compared to their oral counterparts, nasal vowels show additional nasal formants and anti-formants, broader F1 bandwidth, lower F1 amplitude, and higher spectral tilt (e.g., [2, 3]). These acoustic features can be attributed to velopharyngeal coupling which occurs as a result of lowering the velum. In addition to a lowered velum, nasal vowel production may also involve distinct labial and/or lingual gestures [1, 4-7], breathier voice [1, 8], and different pharyngeal configurations [5, 7]. Such complexity means that formant frequency analysis alone is less reliable for determining the underlying articulatory gestures, necessitating the use of tools like ultrasound tongue imaging to observe the gestures directly.

This study investigates the articulation of nasal vowels in Shanghai Chinese (Shanghainese) from the perspectives of lingual articulation, nasalance, and voice quality. Formant frequency data will also be reported. Shanghainese differs from Shanghai Mandarin, which is a Mandarin dialect spoken in Shanghai. Shanghainese and Shanghai Mandarin are not mutually intelligible. Studies on the sound system of Shanghainese (e.g., [9-15]) are abundant, but comprehensive articulatory and acoustic study of its nasal vowels remains lacking.

The Shanghainese low vowels, both oral and nasal, have been given a range of transcriptions, as shown in Table 1. In Chinese literature, the low central vowel [ä] is usually transcribed with the non-IPA symbol [A], reproduced here. Shanghainese has one low oral vowel, typically transcribed in open syllables as [a], [ɑ], or [A]. Some (but not all) studies report raising of this vowel to [ɐ] in checked syllables. While Shanghainese historically had two low nasal vowel phonemes (/ã/ and /ɒ̃/), recent work reports a merger of the two categories among young speakers.

	In CV	In CVʔ	In CṼ
Chao [9] (ca. 1910s)	[ɑ]	[A]	[ã], [ɒ̃]
Xu & Tang [10] (ca. 1930s-1950s)	[A]	[A]	[ã], [ɒ̃]
Xu & Tang [10] (ca. 1960s-later)	[A]	[ɐ]	[Ã]
Qian [11] (ca. 1980s)	[A]	[ɐ]	[Ãᶮ] or [Ã] or [ã ɒ̃]
Qian [11] (ca. 2000s)	[A]	[ɐ]	[Ã]
Zee [12] (ca. 1950s)	[a]	[a]	[aᵑ], [ɔᵑ] ([ãᵑ] young speakers)
Chen & Gussenhoven [13] (ca. 1950s-1960s)	[a]	[ɐ]	[ɐ], [ɑ] (nasalized before [ŋ])
Chen [14] (ca. 1970s-1980s)	[a]	[a]	[ã ɑ̃]
Chen [14] (ca. 1990s)	[a]	[ɐ]	[Ã]
Shen [15] (young speakers in 1981)	[a]	[a]	[ã], [ɔ̃]

Table 1: Low vowel transcriptions in previous studies, with approximate speaker birth years.

Previous work regarding the degree to which the nasal vowels are nasalized is insufficient. Only Fung & Lau [16] report nasal airflow data from one Shanghainese speaker and show that the two nasal vowels are fully nasalized throughout their duration.

Given the limited data available, this study will examine the time course of vowel nasalization through nasalance.

Some acoustic features of nasal vowels, such as broader F1 bandwidth and higher spectral tilt, resemble those of breathy voice [17]. Previous studies have found that nasal/nasalized vowels in Chaoshan Chinese [8], French [1], and Yi [18] can be produced with breathier voice, due either to misperception or acoustic enhancement [1, 18]. Interaction of voice quality and nasality is of particular interest for Shanghainese, given that tone [˩˧] is produced with breathy (whispery) voice [19]. Although some studies report the loss of breathy voice on [˩˧] among young speakers [20], this study will also investigate the potential interaction between vowel nasality and breathy voice in syllables bearing the tone [˩˧]. Acoustic measures typically used to detect breathier voicing, e.g., H1-H2 and H1-A3, are unreliable with the presence of nasality, and thus voice quality will be analyzed through electroglottography.

Through a combination of synchronous acoustic, ultrasound, electroglottographic, and nasalance data, this study will investigate:

1. The tongue positions and acoustic features of the Shanghainese vowels [ɐ, ãᵑ, ɒ̃ᵑ] in comparison with [a].
2. The time course of nasalization for the nasal vowels, to determine whether the vowels are fully or partially nasalized.
3. Whether speakers of Shanghainese also use lower contact quotient (breathier voice) in producing the nasal vowels, and how the breathy voice of tone [˩˧] interacts with nasality.

2. METHODS

2.1. Speakers

Data from 4 native speakers of Shanghainese (2 men, 2 women) were analyzed. They are university students in their early twenties and lived in Shanghai prior to age 18. Urban Shanghainese is the primary language spoken in their families, although they are also proficient in Mandarin and English. None of the speakers reported speech or hearing disorders.

2.2. Materials

The word list contained 17 (near) minimal sets (68 words plus 5 additional (C)Vʔ syllables) of the low vowels in (C)(w)V, (C)Vʔ and (C)(w)Ṽ syllables carrying all the Shanghainese tones /˥˨ ˧˦ ˩˧ ˩˨ ˥/. Tones /˥˨ ˧˦ ˩˧/ are carried by (C)V, (C)Ṽ and (C)Vŋ syllables, while tones /˩˨ ˥/ are only carried by (C)Vʔ syllables. The word list also contained 57 fillers with rimes [i u o ɪŋ ʊŋ əŋ]. Sample sets are provided in Table 2.

CV	CVʔ	CṼ
[ka˥˨] 街 "road"	[kɐʔ˥] 夾 "clip"	[kãᵑ˥˨] 羹 "thick soup" [kɒ̃ᵑ˥˨] 江 "large river"
[tsa˧˦] 債 "debt"	[tsɐʔ˥] 著 "to wear"	[tsãᵑ˧˦] 帳 "accounts" [tsɒ̃ᵑ˧˦] 壯 "chubby"
[a˩˧] 鞋 "shoes"	[ɐʔ˩˨] 匣 "box"	[ãᵑ˩˧] 杏 "apricot" [ɒ̃ᵑ˩˧] 項 "item"

Table 2: Sample (near) minimal sets in the word list.

2.3. Data collection

We use a similar equipment setup to that of Carignan [1]. Synchronous ultrasound, audio, lip video, electroglottograph (EGG) and nasalance recordings were made in a sound-attenuated booth at the University of Hong Kong.

Ultrasound images of the tongue were recorded at a typical rate of 84 frames per second using a SonoSpeech Micro ultrasound system. Participants wore a headset [21] to prevent movement of the cameras and ultrasound transducer. Audio recordings were captured using an Earthworks Ethos supercardioid condenser microphone and a Sound Devices USBPre2 preamplifer. Audio was recorded to a Denon F650R solid state recorder at a 44.1 kHz sampling rate and 16 bit sampling depth. Ultrasound, audio, and lip video recordings were synchronized in Articulate Assistant Advanced (AAA) [22]. EGG recordings were made using a Voce Vista electroglottograph (model: 7050A). Nasalance was recorded using the Glottal Enterprises Nasalance Separator Handle. This system comprises a plate held against the participant's upper lip to separate the oral and nasal audio signals, rather than a Rothenberg mask which seals the nose and mouth, so high fidelity audio recordings can be made at the same time. Output from the EGG and Nasalance meter, as well as the speech and synchronization audio signals from AAA, were recorded simultaneously on an iMac using Audio Hijack.

Syllables written in simplified Chinese characters were presented in AAA in a unique pseudorandom order for each participant. Participants were instructed to read four successive repetitions of each syllable in isolation with a short pause between repetitions.

2.3. Data analysis

Tongue contours and formant frequencies were extracted at a single point during the steady state portion of the vowel. Tongue contours were tracked in AAA using a pre-trained MobileNet1.0-based neural network in DeepLabCut [23]. Data were analyzed with polar SSANOVA [24].

The nasalance recordings provide two audio channels, including both a nasal channel (A_N) and an oral channel (A_O). To see the time course of vowel nasalization, nasal consonants following the nasal vowels were excluded when present. RMS amplitude values were measured at 20% intervals during the vowel for both channels (A_O and A_N). Nasalance was calculated using the formula $A_N / (A_O + A_N)$, where a higher value indicates a higher degree of nasalization.

For the EGG data, since lower contact quotient (CQ) is related to higher spectral tilt [25], a shared acoustic feature of nasality and breathy voice, CQ was measured in nasal vowels; a lower CQ value indicates breathier voice. The data were obtained using EggWorks [26] (hybrid method [27]). Data within each fifth were averaged in VoiceSauce [28]. Each tone was analyzed separately, given that tones may be produced with different voice qualities. Tone [˥˩] is produced with glottalization at the end, making it difficult to obtain reliable CQ values, so this tone was excluded. Both nasalance and CQ values were analyzed with SSANOVA. Shading around the splines indicates 95% confidence interval and overlap of the shading suggests no significant difference.

Some data were excluded due to mispronunciation or clipping of the EGG and/or nasalance recordings. The numbers of tokens included are shown in Table 3 (following references).

3. RESULTS

This section presents tongue contours and formant frequencies of the low vowels, as well as nasalance and contact quotient data for the nasal vowels. Although it has been reported that the two nasal vowels [ãᵑ õᵑ] have merged for young speakers [10, 11, 12, 14], we report results for each vowel separately, given that the merger is relatively recent and previous studies are mainly impressionistic.

SSANOVA analysis of tongue position (Figure 1) reveals that, for SH01F, SH01M and SH02F, the oral vowel [a] is higher than [ɐ ãᵑ õᵑ]. For all speakers, the nasal vowels are more retracted than the oral vowels.

Acoustic data reveal that the F1 of [ɐ] tends to be lower than that of [a], although the ellipses for the two vowels still overlap considerably, particularly for SH02F. Consistent with the tongue contours, [ãᵑ õᵑ] tend to have a lower F2 than the oral vowels, with the exception of SH02F. Although the nasal vowels show a lower tongue position than the oral vowels, the F1 of the two nasal vowels usually does not exceed that of [a], as would be predicted by the difference in tongue position. This effect can be attributed to the lowering effect that nasalization has on F1 [29]. In accordance with previous work, this study confirms that the two nasal vowels have merged in production.

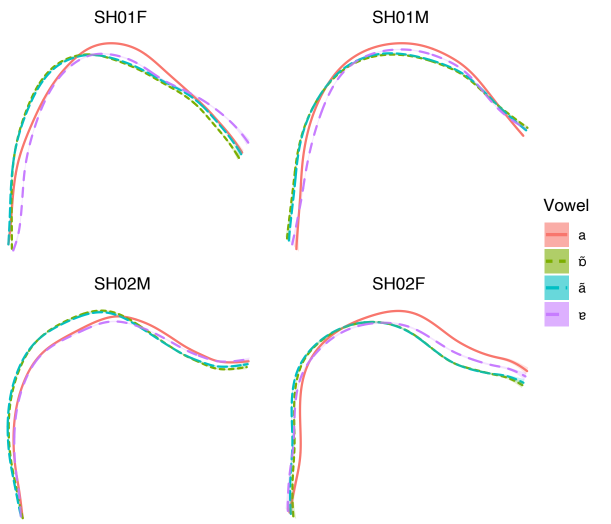

Figure 1: Polar SSANOVA tongue contours. The tongue root is to the left, shading represents 95% CI.

Nasalance data (Figure 3) reveal an increasing degree of nasality throughout the vowel duration. The nasal vowels become significantly nasalized after one fifth (SH02F), or two fifths (SH01M and SH02M) or three fifths (SH01F) of the vowels. This result differs from previous work on French, which shows a high degree of nasalization from the vowel onset [1].

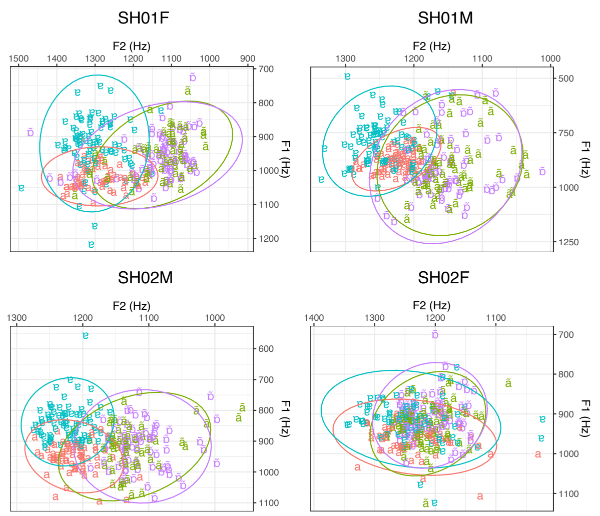

Figure 2: F1 and F2 values of the low vowels.

CQ measurements (Figure 4 and 5) indicate that none of the speakers used a lower CQ (breathier voice) to produce the nasal vowels. On the contrary, SH01M shows higher CQ values (less breathy) for the two nasal vowels carrying tone [˦˥] (Figure 4).

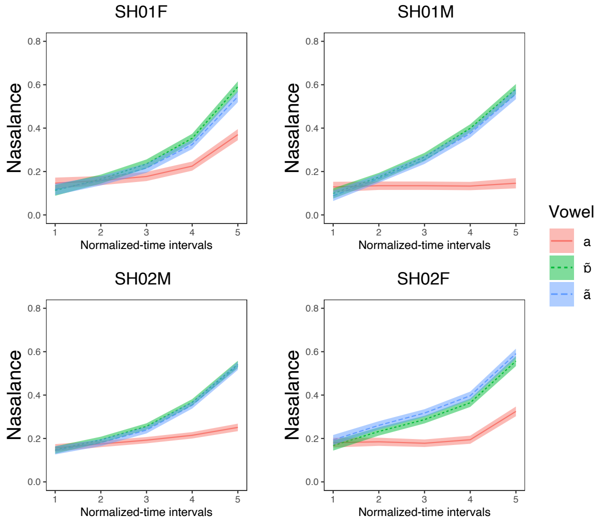

Figure 3: SSANOVAs for nasalance measurements.

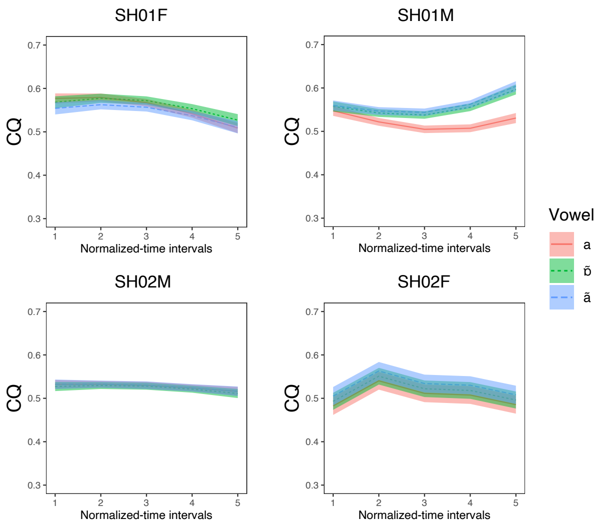

Figure 4: SSANOVAs for CQ values of tone [˦˥].

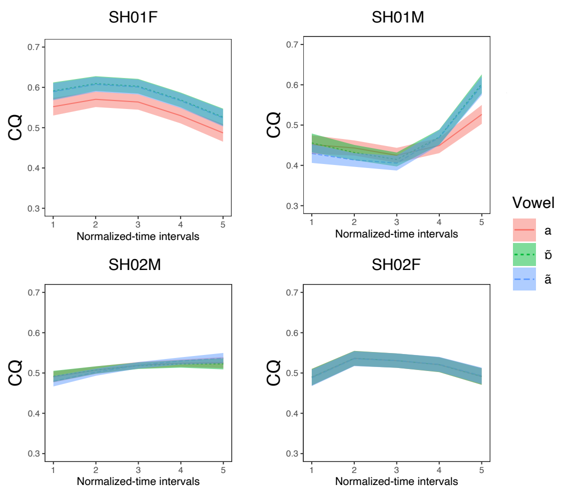

Figure 5: SSANOVAs for CQ values of tone [˩˧].

4. DISCUSSION AND CONCLUSIONS

Ultrasound data show that the nasal vowel is produced with a lower and more retracted tongue position than its oral counterpart [a], and can be transcribed as [ã̠ᵑ]. Retracted tongue position is believed to enhance the acoustic effects of nasalization in French [4]. However, it is not yet clear whether a similar interpretation can also be applied to Shanghainese given that the remaining nasal vowel is the result of merger between [ãᵑ] and [ɒ̃ᵑ] and may partially retain the tongue position used for [ɒ̃ᵑ]. Moreover, the weak coda nasal [ᵑ] may exert coarticulatory influence on the preceding vowel. These possibilities may be confirmed through future studies which more closely examine interspeaker and/or apparent-time variability in terms of the presence/absence of the nasal coda as well as merger/non-merger of the two nasal vowels.

With respect to the vowel height, the lower tongue position used by three speakers counteracts the F1 lowering effect of nasalization on low vowels [29], such that the oral and nasal vowels show similar F1 values. The tongue position used to produce the nasal vowels may therefore also reflect compensation for the F1 lowering effect of nasalization, thereby maintaining an auditorily low quality. While the vowel [ɐ] is produced with a lower tongue height when compared to [a], it tends to have lower F1. This observation suggests that transcription of this vowel as [ɐ] may better reflect the auditory quality of the vowel height rather than the actual tongue height.

Nasalance data show that the two nasal vowels are produced with similar degrees of nasalization, further supporting the conclusion that the two nasal vowels have merged. Unlike what has been observed for nasal vowels in French [1], the nasal vowel of Shanghainese shows only partial nasalization.

One speaker used higher CQ to produce the nasal vowels with tone [˦˥]. Gao et al. [30] showed that breathy voice can bias the tone perception towards tone [˩˧] in Shanghainese. The use of higher CQ may serve to prevent confusion of the modal rising tone [˦˥] with the breathy rising tone [˩˧], given the acoustic similarities between the nasality and breathy voice. Perceptual data are needed to verify this hypothesis.

Through the use of multimodal articulatory data, this study shows the multidimensionality of oral-nasal vowel contrast as well as the language- and speaker-specific features of nasal vowel production.

5. REFERENCES

[1] C. Carignan, "Covariation of nasalization, tongue height, and breathiness in the realization of F1 of Southern French nasal vowels," *Journal of Phonetics*, vol. 63, pp. 87–105, Jul. 2017.

[2] S. Hawkins and K. N. Stevens, "Acoustic and perceptual correlates of the non-nasal–nasal distinction for vowels," The Journal of the Acoustical Society of America, vol. 77, no. 4, pp. 1560–1575, Apr. 1985.
[3] W. Styler, "On the acoustical features of vowel nasality in English and French," The Journal of the Acoustical Society of America, vol. 142, no. 4, pp. 2469–2482, Oct. 2017.
[4] C. Carignan, "An acoustic and articulatory examination of the 'oral' in 'nasal': The oral articulations of French nasal vowels are not arbitrary," *Journal of Phonetics*, vol. 46, pp. 23–33, Sep. 2014.
[5] C. Carignan, R. K. Shosted, M. Fu, Z.-P. Liang, and B. P. Sutton, "A real-time MRI investigation of the role of lingual and pharyngeal articulation in the production of the nasal vowel system of French," *Journal of Phonetics*, vol. 50, pp. 34–51, May 2015.
[6] R. Shosted, C. Carignan, and P. Rong, "Managing the distinctiveness of phonemic nasal vowels: Articulatory evidence from Hindi," *The Journal of the Acoustical Society of America*, vol. 131, no. 1, pp. 455–465, Jan. 2012.
[7] M. Barlaz, R. Shosted, M. Fu, and B. Sutton, "Oropharygneal articulation of phonemic and phonetic nasalization in Brazilian Portuguese," *Journal of Phonetics*, vol. 71, pp. 81–97, Nov. 2018.
[8] C. Chen, "Voice Quality of the Nasal Vowels in Chaoshan Chinese," in *SST 2022*, Canberra, Dec. 2022, pp. 76-80.
[9] Y. R. Chao, 現代吳語的研究 [*Studies of the modern Wu dialects*]. Beijing: Tsing Hua College Research Institute, 1928.
[10] B. Xu and Z. Tang, 上海市區方言志 [*A description of urban Shanghainese*]. Shanghai: Shanghai Education Press, 1988.
[11] N. Qian, 上海語言發展史 [*The history and development of Shanghainese*]. Shanghai: Shanghai People's Publishing House, 2003.
[12] E. Zee, "Shanghai Phonology", in *The Sino-Tibetan Languages*, Routledge, pp. 185-192, 2017.
[13] Y. Chen and C. Gussenhoven, "Shanghai Chinese," *Journal of the International Phonetic Association*, vol. 45, no. 3, pp. 321–337, Dec. 2015.
[14] Z. Chen, "開埠以來上海城市方言語音演變 [*The historical developments of sound system since the beginning of Concession in the Shanghai urban dialect*]", Bulletin of Linguistic Studies, vol. 24, pp. 280-313.
[15] T. Shen, "上海話老派新派的差别 [The linguistic differences between the old and young speakers of Shanghainese]", Dialect, vol. 4, pp. 275-283, 1981.
[16] W. N. Fung and S. L. Lau, "A Physiological Analysis of Nasals and Nasalization in Chinese," in *Eurospeech 2001*, Aalborg, Sep. 2001, pp. 325–328.
[17] T. Arai, "Cue parsing between nasality and breathiness in speech perception," Acoust. Sci. & Tech., vol. 27, no. 5, pp. 298–301, 2006.
[18] M. Garellek, A. Ritchart, and J. Kuang, "Breathy voice during nasality: A cross-linguistic study," *Journal of Phonetics*, vol. 59, pp. 110–121, Nov. 2016.
[19] J. Tian and J. Kuang, "The phonetic properties of the non-modal phonation in Shanghainese," *Journal of the International Phonetic Association*, pp. 1–27, Jul. 2019.
[20] J. Tian and J. Kuang, "Revisiting the register contrast in Shanghai Chinese," in *Proceedings of Tonal Aspects of Languages 2016*, Buffalo, NY, 2016, pp. 147–151.
[21] Articulate Instruments Ltd, *Ultrasound Stabilisation Headset User's Manual: Revision 1.4*. Edinburgh, UK: Articulate Instruments Ltd, 2008.
[22] Articulate Instruments Ltd, *Articulate Assistant Advanced User's Guide: Version 2.14*. Edinburgh, UK: Articulate Instruments Ltd, 2012.
[23] A. Wrench and J. Balch-Tomes, "Beyond the edge: Markerless pose estimation of speech articulators from ultrasound and camera images using DeepLabCut," *Sensors*, vol. 22, 2022.
[24] J. Mielke, "An ultrasound study of Canadian French rhotic vowels with polar smoothing spline comparisons," *The Journal of the Acoustical Society of America*, vol. 137, no. 5, pp. 2858–2869, May 2015.
[25] E. Holmberg, R. Hillman, J. Perkell, P. Guiod, & S. Goldman, "Comparisons among aerodynamic, electroglottographic, and acoustic spectral measures of female voice", *Journal of Speech and Hearing Research*, vol. 38, pp. 1212-1223, 1995.
[26] H. Tehrani, "EGGWorks: A Program for Automated Analysis of EGG Signals", available online: https://appsobabble.com/functions/EGGWorks.aspx.
[27] D. M. Howard, "Variation of electrolaryngographically derived closed quotient for trained and untrained adult female singers", *Journal of Voice*, vol. 9, pp. 163–72, 1995.
[28] Y. Shue, P. Keating, C. Vicenik and K. Yu, "VoiceSauce: a program for voice analysis," in *Proceedings of the 17th International Congress of Phonetic Sciences*, Hong Kong, 2011, pp. 1846–1849.
[29] C. Carignan, "Using ultrasound and nasalance to separate oral and nasal contributions to formant frequencies of nasalized vowels," *The Journal of the Acoustical Society of America*, vol. 143, no. 5, pp. 2588–2601, May 2018.
[30] J. Gao, P. Hallé, and C. Draxler, "Breathy voice and low-register: A case of trading relation in Shanghai Chinese tone perception?" Language and Speech, vol. 63, no. 3, pp. 582–607, Sep. 2020.

		Ultrasound	Nasalance	EGG
01F	[a]	64	56	44
	[ɐ]	80	-	-
	[ã^{ŋ}]	56	48	35
	[ɒ̃^{ŋ}]	72	60	37
01M	[a]	67	44	44
	[ɐ]	87	-	-
	[ã^{ŋ}]	68	43	44
	[ɒ̃^{ŋ}]	68	45	44
02M	[a]	68	60	36
	[ɐ]	84	-	-
	[ã^{ŋ}]	56	60	32
	[ɒ̃^{ŋ}]	68	44	32
02F	[a]	68	60	26
	[ɐ]	80	-	-
	[ã^{ŋ}]	68	60	30
	[ɒ̃^{ŋ}]	68	48	33

Table 3: The numbers of tokens included.