INTERSPEECH 2016
September 8–12, 2016, San Francisco, USA

Diphthongization of Nuclear Vowels and the Emergence of a Tetraphthong in Hetang Cantonese

Wenqi Hu [1], *Fang Hu* [2], *Jian Jin* [3]

[1, 3] Sun Yat-Sen University, Canton, China
[2] Institute of Linguistics, Chinese Academy of Social Sciences, Beijing, China
371480218@qq.com, hufang@hotmail.com, imjinjian@163.com

Abstract

This paper is an acoustic phonetic description of vowels in Hetang Cantonese, and focuses on the diphthongization of nuclear vowels. Different to the representative dialect such as Guangzhou or Hong Kong Cantonese, the Hetang dialect exhibits its unique characteristics regarding the phonetics and phonology of vowels. A noticeable phenomenon is the diphthongization of nuclear vowesl. And, a tetraphthong [uɔᵄi] emerges when the nuclear vowel is diphthongized in a triphthong.

Index Terms: vowels, diphthongization, tetraphthong, syllable structure, Cantonese, the Hetang dialect

1. Introduction

Syllables in Chinese and its dialects are straightforward, as each syllable is separately a written unit. In the classification of languages into stress-timing versus syllable-timing ([1]), Chinese and its dialects belong to syllable-timing languages in general. And Cantonese is often identified as a typical one. First, Cantonese dialects usually have no neutral tones and less speech reduction. For instance, speech rhythmic measures indicate that Hong Kong Cantonese is more syllable-timed than Beijing Mandarin Chinese ([2]). Second, the lack of tone sandhi makes Cantonese more syllable-based than dialects such as Wu Chinese. Shanghai Wu, for instance, is more tonal domain-based than syllable-based, and its speech rhythm becomes closer to that in Japanese, a mora-timing language (see e.g., [3]).

Chinese syllables have a simple structure, and could be represented sequentially as CGVC, namely initial consonant, on-glide, vowel nucleus, and coda ([4]). The coda is limited to a nasal N [-m -n -ŋ], an unreleased stop S [-p -t -k], or a vocalic off-glide G [-i -u] in Cantonese. The representative dialect of Cantonese, Guangzhou or Hong Kong, is well known for the lack of on-glide G ([5], [6], [7], [8], [9]). However, the Hetang dialect, which belongs to the Siyi (the four counties) sub-family of Cantonese dialects, does have an on-glide G [i u y] ([10]). Whereas the initial consonant, on-glide and coda are all optional, the nuclear vocalic element V is obligatory for a legitimate syllable. Syllables that are composed of a syllabic nasal are exceptional cases and are not relevant to the discussion here. The on-glide and nuclear vowel form a rising diphthong in CGV syllables; the nuclear vowel and off-glide form a falling diphthong in CVG syllables; the on-glide, nuclear vowel and off-glide form a triphthong in CGVG syllables; the monophthong and rising diphthong could have a consonantal coda and thus form CVN, CVS, CGVN, and CGVS syllables. And the vowel nuclei could contrast in vowel length and differ in quality at the same time.

The Hetang dialect has 7 monophthongs [i u y a ɛ œ ɔ] in CV syllables, 9 falling diphthongs [iu ui ɔi ou ai au ɐi ɐu ɵy] in CVG syllables, 3 rising diphthongs [ui iɛ ua] in CGV syllables, 4 triphthongs [yui iɐu iai uai] in CGVG syllables, 25 possible rimes of CVS/N syllables [ɐp ɐm ɐŋ ɐn ɐt ɐk ap an at ak aŋ am œŋ œk ɛk ɵn ɵt ɔŋ ɔk oŋ ok un ut it in], and 25 possible rimes of CGVS/N syllables [uan iaŋ uaŋ iap iak uak uɐn iɐŋ uɐŋ uɐt iɐk iɛm iɛn iɛŋ iɛp iɛt iɛk yœk yœŋ yɔŋ yɔk uɔŋ uɔk yok yoŋ]. For the sake of convenience, the long versus short vowel contrast was transcribed as differences in quality: [a]-[ɐ], [ɔ]-[o], [ɛ]-[e], and [œ]-[ɵ].

A less documented phenomenon for Cantonese phonology in particular and vowel phonology in general is that the 2 nuclear vowels [i ɔ] become diphthongized in 12 rimes in Hetang Cantonese: [ɔᵄi] in the CVG syllable, [iᵉp ɔᵄt] in CVS syllables, [uɔᵄt yɔᵄt] in CGVS syllables, [iᵉm ɔᵄn ɔᵄŋ] in CVN syllables, [uɔᵄŋ uɔᵄn yɔᵄn] in CGVN syllables, and [uɔᵄi] in the CGVG syllable, which results in a tetraphthong.

This paper gives an acoustic phonetic analysis of the nuclear vowels in different syllable types in the Hetang dialect, with an emphasis on the diphthongization of nuclear vowels.

2. METHODOLOGY

Meaningful monosyllabic words containing all the target nuclear vowels in different types of syllables were used as test words. Each target vowel/rime has two test words; ideally one has a zero initial and the other has a bilabial stop initial. And when possible, each test word is associated with a high level tone.

4 Hetang male native speakers provided the speech data. All speakers were born and raised in Hetang and did not have a history of speech or hearing problems. At the time of recording, the ages of the speakers ranged from 56 to 70.

The speech data were recorded during the first author's fieldwork in Hetang in the summer of 2015. The 16-bit audio was recorded directly into a laptop PC by using a Terratec 6 Fire USB digital audio sound card and a Shure SM-86 microphone. The sampling rate is 11,025 Hz. Each test word was placed in a carrier sentence [ŋɔ^{35} iɛu^{33} tok^{2}_ pi^{55} ni^{33} tiɛŋ^{24}] 'I am reading _ for you'. The test words in the word list were randomized, and the speakers were asked to read both the citation form and the word in the carrier sentence in a natural way and at a normal speech rate. Four repetitions were recorded.

Audio data were analyzed in Praat ([11]). Each element in a target rime was annotated, and then the lowest three

http://dx.doi.org/10.21437/Interspeech.2016-61

formants (F1, F2, F3) were extracted from the midpoint of the stead states of target elements. And the duration for each component of the rime was also measured. The acquired formant and duration data were used for further analysis. And this study focuses on nuclear vowel and its following diphthongized element, if any.

3. Results

3.1. Nuclear vowels

Figures 1-7 show the distributions of the Hetang nuclear vowels, which occur in different syllable structures respectively, in acoustic vowel spaces. Each 2-sigma ellipse was based on 16 data points (4 repetitions × 4 speakers), and plotted in an acoustic plane (F1 against F2), with the origin of the axes to the top right of the plot. The axis scale has been converted to Bark, and the scale on the ordinate is double that on the abscissa. Therefore, the acoustic vowel plane could better approximate the perceptual distances between vowels in such a way that the plots are more in accord with the auditory judgments of vowels. However, the values along the axes still correspond to the original values in Hertz.

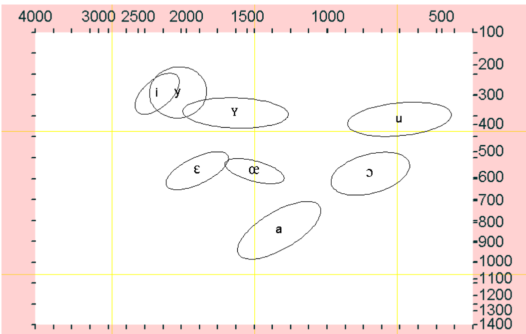

Figure 1: *2-sigma ellipses of Hetang monophthongs.*

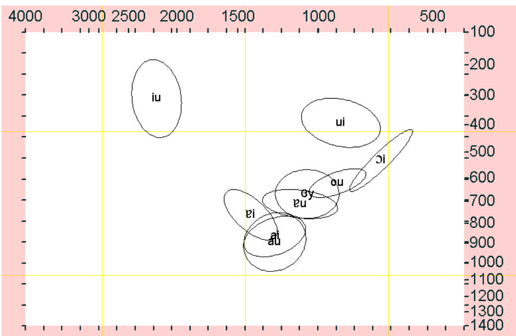

Figure 2: *2-sigma ellipses of nuclear vowels in falling diphthongs in Hetang Cantonese.*

As shown in Figure 1, the 7 monophthongs in Hetang display a triangular distribution with three levels of height and two levels of backness, namely the high vowels [i y u], the mid vowels [ɛ œ ɔ], the front vowels [i y ɛ œ], the back vowels [u ɔ], and the low vowel [a] that does not contrast in backness. While back vowels are predictably rounded, front vowels in Hetang contrast in lip rounding for high and mid vowels: [y] is a rounded high vowel, and [œ] is a rounded mid vowel. Please be noted that [ʏ] is a variant of [y] when it is preceded by a velar or glottal initial [k kʰ h].

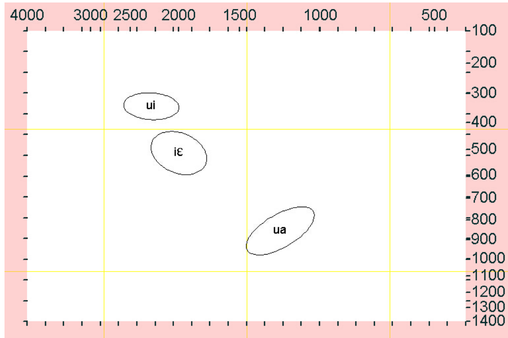

Figure 3: *2-sigma ellipses of nuclear vowels in rising diphthongs in Hetang Cantonese.*

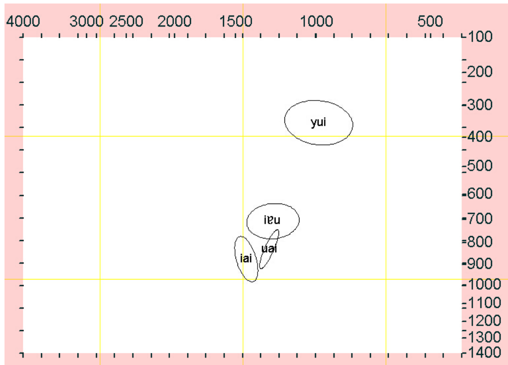

Figure 4: *2-sigma ellipses of nuclear vowels in triphthongs in Hetang Cantonese.*

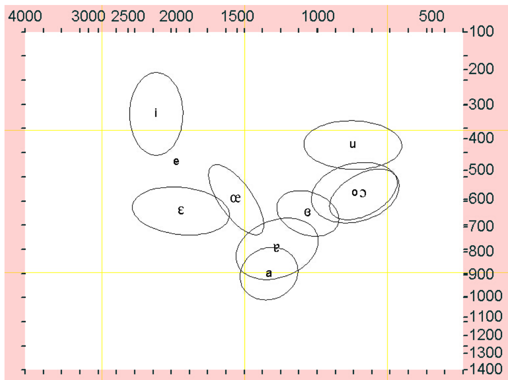

Figure 5: *2-sigma ellipses of nuclear vowels in CVS/N syllables in Hetang Cantonese.*

The nuclear vowels in 9 falling diphthongs do not show a balanced distribution in the acoustic vowel space in Figure 2. Rather, they are crowded in the region of back and low vowels. This is because the nuclear vowels in falling diphthongs contrast in quantity as well as quality. In addition to three unpaired falling diphthongs [iu ui ɵy], there are three pairs of the long versus short vowels in Hetang: [ai]-[ɐi], [au]-[ɐu], and [ɔi]-[ou]. It should be noted that there are only 8 data points for [ɔi], since the rest is diphthongized to [ɔᵊi].

Hetang only has three rising diphthongs [ui iɛ ua]. It is interesting to observe from Figure 3 that the nuclear vowels in rising diphthongs are all distributed in the region of front vowels, which is complementary to the distribution of the nuclear vowels in falling diphthongs in general.

Figure 4 shows the nuclear vowels of the 4 triphthongs [iai uai iɐu yui] in Hetang. The nuclear vowels in triphthongs also contrast in duration, namely the long [a] in [iai uai] versus short [ɐ] in [iɐu]. It should be noted that another triphthong [uɔi], which is reported in the literature ([9]), has changed into a tetraphthong [uɔ^{ɐ}i], as its nuclear vowel [ɔ] is diphthongized [ɔ^{ɐ}] (see 3.2 for details).

As shown in Figure 5, the nuclear vowels in CVS and CVN syllables have a balanced distribution in the acoustic vowel space in general. The 2-sigma ellipses were based on all data points from relevant rimes, since statistics did not yield significant difference in quality or quantity for the nuclear vowels between CVS and CVN syllables. Please note that [em] has only one data point, and the rest is diphthongized into [i^{e}m]. What is comparable to diphthongal environments is that there are three pairs of the long versus short vowels: [a]-[ɐ], [ɔ]-[o], and [œ]-[ɵ]. And it is interesting to observe that the ellipses for [ɔ] and [o] extensively overlapped with each other, suggesting no difference in quality. Table 1 summarized means and standard deviations in parentheses of F1, F2 and duration for the nuclear vowels [ɔ]-[o] in CVS and CVN syllables, respectively. Independent samples t-test yielded no significant difference for F1 and F2 in CVS syllables and no significant difference for F1 in CVN syllables, but significant difference for duration in both CVS and CVN syllables. In summary, [ɔ o] contrast in quantity but not in quality in general.

Table 1. *Means and standard deviations (in parentheses) of F1 and duration for the nuclear vowels [ɔ o] in CVS and CVN syllables in Hetang.*

	[ɔ]	n.	[o]	n.	t
F1 (CVS)	582 (49)	103	576 (58)	30	0.586
F2 (CVS)	770 (75)	103	781 (47)	30	0.330
Dur. (CVS)	227 (38)	103	128 (41)	30	0.00*
F1 (CVN)	576 (43)	54	574 (52)	32	0.868
F2 (CVN)	759 (65)	54	849 (144)	32	0.002*
Dur. (CVN)	264 (57)	54	112 (32)	32	0.00*

Table 2. *Means and standard deviations (in parentheses) of F1, F2 and duration for the nuclear vowels [ɔ o] in CGVS and CGVN syllables in Hetang.*

	[ɔ]	n.	[o]	n.	t
F1 (CGVS)	594 (29)	19	588 (55)	12	0.716
F2 (CGVS)	780 (47)	19	846 (107)	12	0.061
Dur. (CGVS)	256 (82)	19	106 (33)	12	0.00*
F1 (CGVN)	564 (40)	35	519 (33)	12	0.001*
F2 (CGVN)	821 (85)	35	877 (140)	12	0.212
Dur. (CGVN)	226 (78)	35	75 (20)	12	0.00*

Figures 6 and 7 show ellipses for nuclear vowels in CGVS and CGVN syllables, respectively. The data were pooled across all nasal codas [-m -n –ŋ], represented by –N in the figure, and across all stop codas [-p -t –k], represented by –R in the figure, respectively. Different to those in CVS and CVN syllables, nuclear vowels in CGVS and CGVN are distributed in mid and low vowels' region in the acoustic vowel space. And there are two pairs of the long and short vowels: [a]-[ɐ] [ɔ]-[o]. As shown in the acoustic vowel spaces, [a ɐ] exhibit a difference in quality between [iaS uaS iaN uaN] and [iɐS uɐS iɐN uɐN] respectively. But the ellipses for [ɔ] and [o], again, heavily overlapped with each other. As summarized in Table 2, independent samples t-test yielded significant difference for duration in CGVS syllables and for F1 and duration in CGVN syllables, but no significant difference for F1 and F2 in CGVS syllables and for F2 in CGVN syllables. In summary, [ɔ o] contrast in quantity in CGVS syllables, and in CGVN syllables, they still differ in quantity more than in quality.

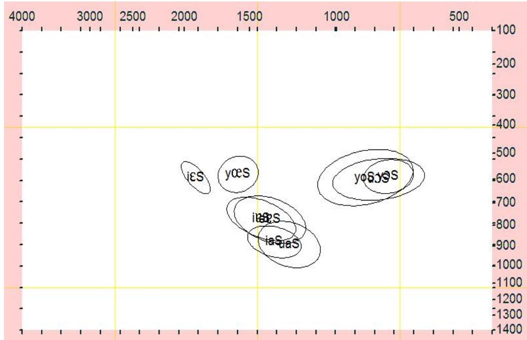

Figure 6: *2-sigma ellipses of nuclear vowels in CGVS syllables in Hetang Cantonese.*

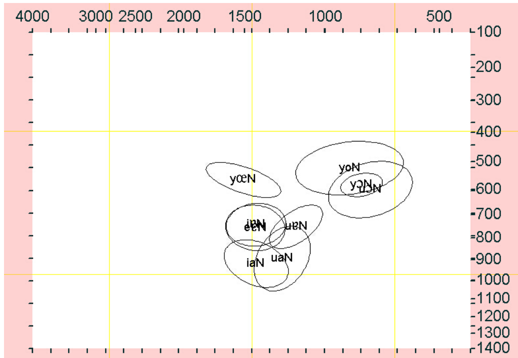

Figure 7: *2-sigma ellipses of nuclear vowels in CGVN syllables in Hetang Cantonese.*

It should be noted that the nuclear [ɔ] is diphthongized in a number of stop or nasal coda rimes [uɔ^{ɐ}t uɔ^{ɐ}n uɔ^{ɐ}ŋ yɔ^{ɐ}t yɔ^{ɐ}n]. [uɔŋ] is retained in a few samples, and [yɔR yɔN] in Figure 6 and 7 are variants of [uɔk uɔn] that are preceded by laminal initial consonants.

3.2. Diphthongization of nuclear vowels

Diphthongization is an ongoing process in Hetang Cantonese. The 2 nuclear vowels [i ɔ] tend to become diphthongized into [i^{e} ɔ^{ɐ}] respectively in 12 rimes [i^{e}m i^{e}p yɔ^{ɐ}t yɔ^{ɐ}n ɔ^{ɐ}n ɔ^{ɐ}ŋ ɔ^{ɐ}t uɔ^{ɐ}ŋ uɔ^{ɐ}n uɔ^{ɐ}t ɔ^{ɐ}i uɔ^{ɐ}i]. There are, however, variations regarding the process of diphthongization. Table 3 summarized frequencies and percentages of diphthongized realizations for each rime. Other realizations of the target rime were listed in the column of "notes". Please be noted that the diphthongized forms in the column of "notes" were included in the calculation of percentage of the realization of diphthongization, since it concerns with the same nuclear vowel.

As can be seen from the table, the process is overwhelming in the environment of labial, alveolar, or palatal codas [-p -m -t -n -i], namely 100% in 7 rimes and 84% to 95% in 3 rimes. By contrast, less diphthongization is observed in the environment of the velar coda [-ŋ], namely 25% in [ɔ^{ɐ}ŋ] and 27% in [uɔᵄŋ]. Moreover, the nuclear vowel [ɔ] in the rimes [ɔk uɔk yɔk yɔŋ] has not been observed for diphthongization. The other nuclear vowel [i] demonstrates no diphthongization in the environment of alveolar and velar codas [it in iu]. It seems that nuclear diphthongization is a gradient process in Hetang. The process could be vowel-dependent, and sensitive to the environment of following codas. Especially, the velar codas [-u -k -ŋ] have a blocking effect on the diphthongization of nuclear vowel.

Table 3. *Variations of diphthongized rimes in Hetang.*

rime	diph.	%	n.	notes
[iᵉp]	16	100%	16	
[iᵉm]	27	84%	32	[em]: 1
[ɔᵄi]	141	95%	155	[uɔᵄi]: 6
[uɔᵄi]	34	100%	80	[ɔᵄi]: 46
[ɔᵄt]	24	85%	65	[uɔᵄt]: 31
[uɔᵄt]	38	100%	147	[ɔᵄt]: 91; [yɔᵄt]: 18
[ɔᵄn]	44	100%	80	[uɔᵄn]: 35; [uᵄn]: 1
[uɔᵄn]	39	100%	168	[ɔᵄn]: 93; [uᵄn]: 11; [yɔᵄn]: 25
[yɔᵄt]	32	100%	32	
[yɔᵄn]	32	100%	32	
[ɔᵄŋ]	10	25%	160	[uɔᵄŋ]: 30; [uɔŋ]: 33; [yɔŋ]: 18
[uɔᵄŋ]	4	27%	52	[ɔᵄŋ]: 10; [ɔŋ]: 27

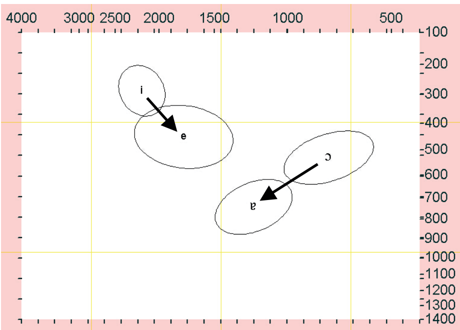

Figure 8: *Diphthongization of nuclear vowels [i] > [iᵉ] and [ɔ] > [ɔᵄ] in Hetang Cantonese.*

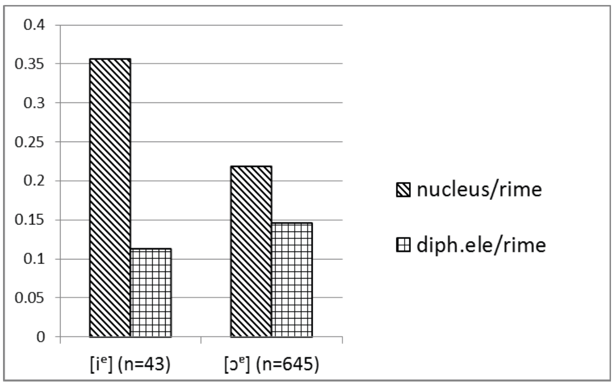

Figure 9: *Mean duration in percentage for the nucleus and diphthongized element in [iᵉ] and [ɔᵄ].*

Figure 8 shows the two processes of diphthongization and the 2-sigma ellipses in the figure are based on the pooled data from all repetitions of all target rimes in all speakers. And Figure 9 summarized the mean durations in percentage for nuclear vowels and the diphthongized elements, namely the percentage of the target element in the entire rime.

[iᵉ ɔᵄ] share a commonality in that they both have a small course of diphthongization in the acoustic vowel space, namely from [i] to [e] for [iᵉ] and from [ɔ] to [ɐ] for [ɔᵄ]. But [iᵉ ɔᵄ] differ in both temporal structure and spectral property. The duration of diphthongized element [e] is about one third of the nucleus [i]; and [e] has a more variable distribution than [i] does in the acoustic vowel space. [ɐ] is about 7% shorter than the nuclear [ɔ], and their ellipses are comparable. These differences suggest that [iᵉ ɔᵄ] could probably be in different stages of diphthongization. [ɔᵄ] is more stable than [iᵉ], as the former has a longer duration and less variable spectral target for the diphthongized element than the latter.

4. DISCUSSION AND CONCLUSIONS

Representative Cantonese dialects such as Guangzhou and Hong Kong dialects are well documented. But regional variations of Cantonese dialect family are less investigated, especially in the framework of experimental approaches to phonetics and phonology. This paper reports acoustic phonetic aspects of vowels with an emphasis on the diphthongization of nuclear vowels in Hetang Cantonese.

The 7 monophthongs [i u y a ɛ œ ɔ] contrast in vowel height, backness and lip rounding in CV syllables in Hetang. The nuclear vowels additionally contrast in quantity, which is generally accompanied by a difference in quality and is transcribed as [a ɔ ɛ œ] versus [ɐ o e ɵ] respectively, in CVG, CGVG, CVS and CVN syllables.

The two nuclear vowels [i ɔ] in Hetang Cantonese tend to become diphthongized in 12 rimes. The observed variations, however, suggest that diphthongization is a gradient process in Hetang. First, the process of diphthongization is vowel-dependent. The other nuclear vowels are not diphthongized. Even [i ɔ] are different. [i] is diphthongized in two rimes [iᵉp iᵉm], but not in the other rimes [it in iu]. [ɔ] is diphthongized overwhelmingly in all the ten rimes. As compared to [iᵉ], [ɔᵄ] has a more stable temporal structure and a better controlled spectral target for the diphthongized element. Second, the process of diphthongization shows a coda-dependent tendency, namely the velar codas [-u -k -ŋ] have a blocking effect on the process.

Syllables in Chinese dialects usually have a simple structure, for there is no consonant cluster in general. Meanwhile there are complex vowels and rimes as well as tones in syllables in Chinese dialects. Hetang Cantonese is among the most complex cases. A number of issues remain unclear and require further inquiries. First is the motivation of diphthongization. What is the initiative actuation? Second is the cost of phonology. [iᵉ] could be reanalyzed as [ie], i.e. a new rising diphthong, but [ɔɐ] is not a legitimate diphthong in Cantonese phonology. And moreover, a tetraphthong [uɔᵄi] emerges when [ɔ] is diphthongized in the triphthong [uɔi].

5. Acknowledgements

This work is supported by Chinese National Social Science Foundation (Project number: 15BYY073; P.I.: Dr. HU Fang).

6. References

[1] Abercrombie, D. *Elements of General Phonetics*, Edinburgh: Edinburgh University Press, 1967.

[2] Mok, P. "On the syllable-timing of Cantonese and Beijing Mandarin", *Journal of Chinese Phonetics*, 2: 148-154, 2009.

[3] Warner, N. and Arai, T. "Japanese mora-timing: a review", *Phonetica*, 58, 1-25, 2001.

[4] Duanmu, S. "Syllable and syllable structure in Chinese", to appear in *Encyclopedia of Chinese Language and Linguistics*, ed. by R. Sybesma (editor-in-chief), W. Behr, Y. Gu, Z. Handel, C.-T. J. Huang, and J. Myers, Leiden: Brill, expected 2016.

[5] Yue-Hashimoto, A. O.-K. *Studies in Yue Dialects 1: Phonology of Cantonese*, Cambridge University Press, 1972.

[6] Bauer, Robert S. & Benedict, Paul K. *Modern Cantonese phonology*, Berlin: Mouton de Gruyter, 1997.

[7] Zee, E. "Chinese (Hong Kong Cantonese)", Handbook of the International Phonetic Association, pp. 58-60, Cambridge, U.K.: Cambridge University Press, 1999.

[8] Li, X., Huang, J., Shi, Q., Mai, Y., & Chen, D. *A study on Cantonese* [in Chinese], Canton: Guangdong People's Publishing House, 1995.

[9] Gui, M. C. *The phonology of Guangzhou Cantonese*, Munich: Lincom Europa, 2005.

[10] Rong, H. *A Study on the Phonology of the Jiangmen Hetang Dialect* [in Chinese]. M.A. thesis, Sun Yat-sen University, 2006.

[11] Boersma, Paul & Weenink, David "Praat: doing phonetics by computer [Computer program]", Version 4.6, 2014.