



# Vowels and Diphthongs in Cangnan Southern Min Chinese Dialect

Fang Hu<sup>1</sup>, Chunyu Ge<sup>2</sup>

<sup>1</sup> Institute of Linguistics, Chinese Academy of Social Sciences

<sup>2</sup> Department of Linguistics, Graduate School, Chinese Academy of Social Sciences

hufang@cass.org.cn, gechunyu92@hotmail.com

## Abstract

This paper gives an acoustic phonetic description of the vowels and diphthongs in Cangnan Southern Min Chinese dialect. Vowel formant data from 10 speakers (5 male and 5 female) show that the distribution of Cangnan monophthongs in the acoustic vowel space is of particular typological interest. Diphthong production is examined in terms of temporal organization, spectral property, and dynamic aspects. Results suggest that the production of falling diphthongs tends to be a single articulatory event with a dynamic spectral target, while the production of rising diphthongs and level diphthongs is a sequence of two spectral targets.

**Index Terms:** vowel production, monophthongs, falling diphthongs, rising diphthongs, Cangnan Southern Min Chinese dialect

## 1. Introduction

Syllables in Chinese dialects are straightforward, as each syllable is a separate written unit. Chinese syllables have a simple structure, which could basically be represented as CGVC (initial consonant, on-glide, vowel nucleus, and coda), but there is controversy regarding the hierarchy of structure (see [1], [2] for reviews). The key issue concerns the phonetics and phonology of vowels and diphthongs. Generally there are three sequential possibilities for diphthong or triphthong formation in Chinese dialects. First, vowel nucleus and vocalic coda could form a falling diphthong, for instance [a + i > ai]; second, the glide and vowel nucleus could form a rising diphthong, for instance [i + a > ia]; third, the glide, vowel nucleus and vocalic coda could form a triphthong, for instance [u + a + i > uai]. But there is no consensus regarding the definition of diphthong. Some phoneticians view a diphthong as a single vowel with a phonetically complex nucleus ([3], [4], [5]), while others treat a diphthong as a sequence of two vowels or a combination of one vowel and one semivowel ([6], [7]). Even for English, for instance, there is controversy regarding what a diphthong is ([8], [9], [10]).

There are various proposals for the phonology of Chinese vowels and diphthongs in the literature ([11], [12], [13], [14], [15], [16], [17], [1]). But previous studies are based on phonemic analyses in general, and thus have non-unique solutions ([18]). An insightful observation is that falling and rising diphthongs could be different. For instance, Chao pointed out that falling diphthongs are true diphthongs in Wu dialects, but rising diphthongs are not ([19]).

Recent researches renewed the phonetics and phonology of vowels and diphthongs in Chinese dialects. On the basis of acoustic and lingual kinematic data from the Ningbo Wu dialect, [20] argued that falling diphthongs have one dynamic

target, while rising diphthongs are composed of two targets. That is, the diphthong [ai], for instance, is not a sequence of [a] and [i], but a single dynamic articulatory event, and should be treated as being distinctive to the monophthongal vowel [a]; in contrast, the diphthong [ia] is a sequence of [i] and [a]. And acoustic data from southwestern Mandarin ([21]), Jin ([22]), and Hui dialects ([23], [24]) demonstrate similar results. In summary, falling diphthongs and monophthongs are single-event articulations, while rising diphthongs are sequences of two articulation events.

This paper describes the phonetics and phonology of the vowels and diphthongs in Cangnan Southern Min Chinese dialect. The Cangnan County is located in the coastal south of Zhejiang province to the border of Fujian (Hokkien) province. Cangnan has more than four mutually unintelligible local dialects, of which the Cangnan southern Min is the dominant and prestigious one ([25]). Cangnan southern Min is adjacent to the northern Min dialects across the Zhejiang-Fujian border, but not to the motherland southern Min dialects in south Fujian. In other words, Cangnan southern Min is a linguistic exclave.

Table 1. *The vowel inventory of Cangnan Min dialect.*

Vowel		CV	CVN	C $\bar{V}$
Monophthongs		i u e a ɔ ə ɐ o u	a(n/ŋ) i(n/ŋ) u(n) ɔ(ŋ) ə(n)	ĩ ẽ ǣ
Diphthongs	Rising	ie ia iɔ ue ua ue	ia(n) ua(n) iɔ(ŋ)	ĩã uã
	Falling	əu ai au		ǣĩ
	Level	ui iu		ũĩ iũ
Triphthongs		iəu iau uai		

As listed in Table 1, the Cangnan southern Min has 9 monophthongs, 11 diphthongs, and 3 triphthongs in open syllables. However, only a limited number of vowels and rising diphthongs occur in syllables with a nasal coda; and a limited number of vowels and diphthongs occur in nasalized syllables. Due to the space limit, this paper focuses on monophthongs and diphthongs in open syllables.

## 2. Methodology

10 native speakers, 5 male and 5 female, provided speech data. All of them are born and raised up in Cangnan and have no reported speech or language disorders.

Two meaningful monosyllabic words, preferably one with a labial consonant and the other without a consonant, were

chosen as test words for each target vowel. The word list was randomized and the speakers were instructed to read the word list naturally in a normal speech rate. The 16-bit audio sounds were recorded directly into a laptop PC with a Lexicon I-O 22 sound card and through a SHURE SM86 microphone. The sampling rate is 11,025 Hz. Five repetitions were recorded.

The speech data were analyzed in praat 5.3.61 ([26]). If there are steady states, a diphthong is annotated as being composed of an onset element, an offset element, and the transition. The temporal structures were then measured for diphthongs, and the lowest four formants were extracted from the spectrogram in the mid-point of the onset and offset segments. And the range and rate of spectral change (the second formant, F2) was calculated as a measure of spectral dynamics for each diphthong.

### 3. Results

#### 3.1. Monophthongs

Figure 1 shows the distribution of the nine monophthongs [i u e a ɔ ə ɐ o u] of Cangnan in the acoustic vowel plane by using the first formant (F1) as ordinate and second formant (F2) as abscissa with the origin of the axes to the top right. The axes are Bark-scaled ([27]), while the values along the coordinates are still labeled in Hertz. And each 2-sigma ellipse in the figures is based on 50 data points (2 words  $\times$  5 repetitions  $\times$  5 speakers).

The Cangnan monophthongs exhibit a triangular distribution in the acoustic vowel space. There are four levels of vowel height, namely high vowels [i u], mid-high [e ə ɔ], mid-low [ɐ ɐ], and the low vowel [a]. There are three levels of vowel backness, namely front vowels [i e], central vowels [ə ɐ a], and back vowels [u o ɔ]. The low vowel [a] does not contrast with any vowel in backness. As a southern Min dialect, Cangnan vowels are of typological interest. First, there is no contrast in lip rounding: the front vowels are unrounded; the central vowels are unrounded; and the back vowels are rounded. The lack of the rounded high front [y] is a well-known accent of southern Min. And note that [u] is grouped with central vowels in the acoustic vowel plane, although it is transcribed with an unrounded back high vowel symbol. Second, there are less front vowels than back vowels and central vowels, i.e. less distinction of vowel height in front vowels than in back vowels and central vowels (c.f.: [28]). Third, there is no apical vowel, which is common in Chinese dialects. But this is also a characteristic of southern Min.

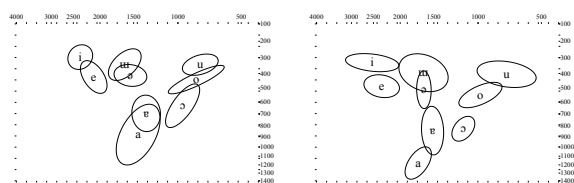


Figure 1: 2-sigma ellipses for the Cangnan monophthongs in acoustic F1/F2 vowel spaces in male (left) and female (right) speakers.

#### 3.2. Diphthongs

The 11 Cangnan diphthongs, 6 rising diphthongs, 3 falling diphthongs and 2 level diphthongs were examined in terms of temporal organization, spectral property, and dynamic aspects.

##### 3.2.1. Temporal organizations

Figure 2 summarized the mean durations in millisecond for components of diphthongs and Figure 3 summarized the duration data in percentage.

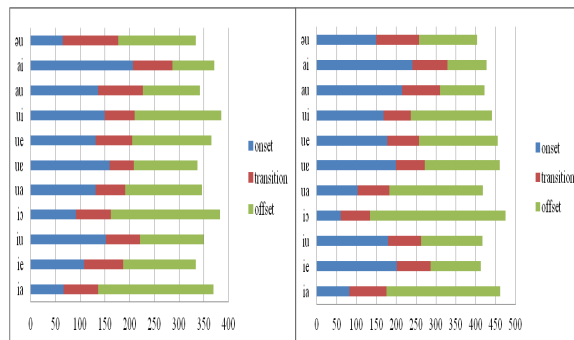


Figure 2: Mean durations in millisecond for the diphthongs in male (left) and female (right) speakers.

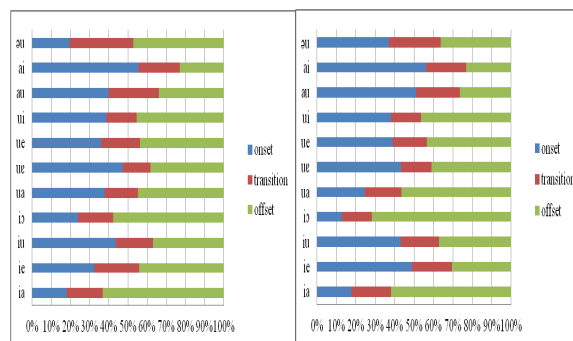


Figure 3: Mean durations in percentage for the diphthongs in male (left) and female (right) speakers.

Although no clear pattern, temporal organizations of Cangnan diphthongs show certain general tendencies. First, both onset and offset elements in a diphthong have steady states, and the transitions are generally short in all diphthongs (c.f.: [20], [21]). Second, the offset element in a rising diphthong is usually longer than its onset counterpart. The offset element takes more than 40% of duration while the onset element usually takes 20-40% in a rising diphthong. And in [iə], the onset element takes only 12% of duration for male and 18% for female speakers. Third, the onset tends to be the longest element, which takes 40% and 56% of duration in falling diphthongs [ai au] respectively. However, the falling diphthong [əu] has a temporal organization similar to rising diphthongs. Fourth, the onset and offset elements have a balanced duration in level diphthongs: both take about 40% of duration. Last, diphthongs are about 100 ms longer in female speakers than in male speakers: the durations of diphthongs are approximately 330-390 ms in male speakers and 400-480 ms in female speakers.

##### 3.2.2. Spectral properties

The formant data of the onset and offset elements in a diphthong were compared with their monophthongal counterparts. It is assumed that a diphthong element tends to have a spectral target, if it has a comparable distribution to its corresponding monophthong in an acoustic vowel plane, whereas the diphthong element tends not to have a target, if it

has a more variable distribution than its corresponding monophthong ([20]).

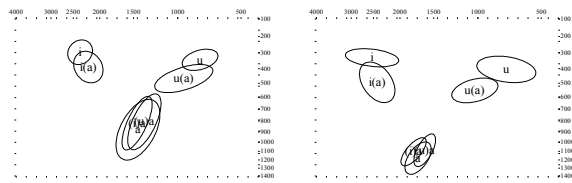


Figure 4: Vowel ellipses for the onset and offset elements of the rising diphthongs [ia ua] and their corresponding monophthongs in male (left) and female (right) speakers.

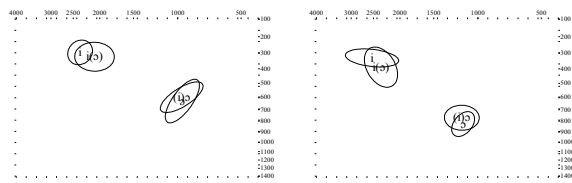


Figure 5: Vowel ellipses for the onset and offset elements of the rising diphthong [io] and their corresponding monophthongs in male (left) and female (right) speakers.

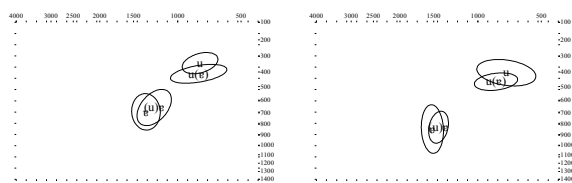


Figure 6: Vowel ellipses for the onset and offset elements of the rising diphthong [ue] and their corresponding monophthongs in male (left) and female (right) speakers.

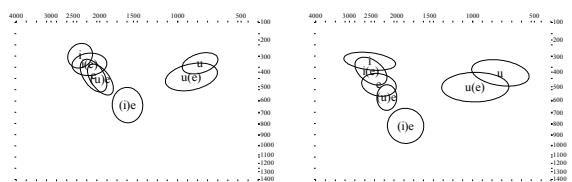


Figure 4: Vowel ellipses for the onset and offset elements of the rising diphthongs [ie ue] and their corresponding monophthongs in male (left) and female (right) speakers.

Figures 4-7 show the ellipses for diphthong elements of rising diphthongs [ie ia io ue ua ue] and their corresponding monophthongs in acoustic F1/F2 vowel planes. The monophthongs were labeled by plain IPAs, and the diphthong elements were indicated by the other diphthong elements in parentheses.

All rising diphthongs exhibit similar patterns except for [ie ue]. First, the offset elements of rising diphthongs and their monophthongal counterparts extensively overlap with each other. Second, the onset elements of rising diphthongs exhibit certain degrees of displacement to the corresponding

monophthongs, but still have comparable ellipses. It seems that the onset element is more likely to make spectral adjustment than is the offset element in a rising diphthong. In other words, anticipatory coarticulation is more salient than progressive coarticulation in the production of rising diphthongs. Nevertheless, both onset and offset elements in a rising diphthong tend to have spectral targets, since they are as well controlled as their monophthongal counterparts are.

[ie ue] are exceptional, but are not counterexamples. First, similar to those in other rising diphthongs, the onset elements in [ie ue] are subject to anticipatory coarticulation. Second, the fact that the offset element in [ie ue] displays a clear shift from its corresponding monophthong [e] in acoustic vowel planes could be attributable to coarticulation effect, too. However, the detected progressive coarticulation could probably be diphthong specific. Third, both onset and offset elements have comparable ellipses to the corresponding monophthongs, which suggests that the production of both onset and offset elements in [ie ue] is as well controlled as in the production of their corresponding monophthongs.

In summary, the production of rising diphthongs is a sequence of two spectral targets that are, however, subject to certain coarticulatory effects.

Figures 8-10 show the ellipses for diphthong elements of the three falling diphthongs and two level diphthongs and their corresponding monophthongs.

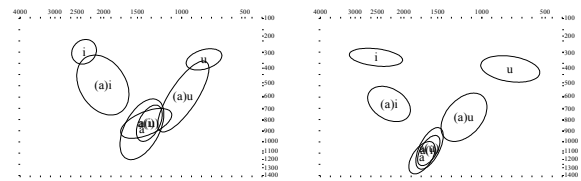


Figure 5: Vowel ellipses for the onset and offset elements of the falling diphthongs [ai au] and their corresponding monophthongs in male (left) and female (right) speakers.

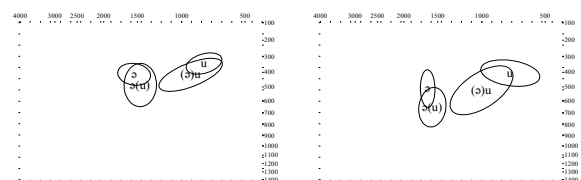


Figure 6: Vowel ellipses for the onset and offset elements of the falling diphthong [əu] and their corresponding monophthongs in male (left) and female (right) speakers.

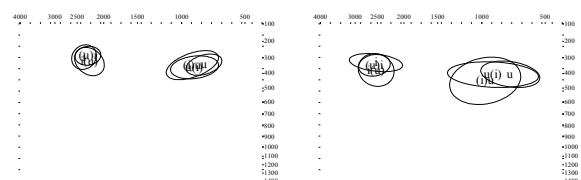


Figure 7: Vowel ellipses for the onset and offset elements of the level diphthongs [iu ui] and their corresponding monophthongs in male (left) and female (right) speakers.

As shown in Figure 8, the ellipses for the onset element of falling diphthongs [ai au] and the corresponding monophthong [a] heavily overlap. But the ellipses for the offset elements have very different and dispersed distributions to their monophthongal counterparts. And the enlarged ellipses suggest higher variability in the production of diphthong offsets than in that of monophthongs. It seems that in the production of falling diphthongs [ai au], only the onset element tends to have a spectral target and the offset element do not have one. Rather, the production of the offset element is constrained by the dynamics of diphthong production. In other words, the production of falling diphthongs is a single articulatory event with a dynamic spectral target, rather than a sequence of two spectral targets as in rising diphthongs. As shown in Figure 9, another falling diphthong [əu] is different from [ai au] in that the ellipse of the onset element has a dispersed distribution in the acoustic vowel plane, too. This is probably due to the innate property of the schwa-like target [ə] ([29]). Whether the onset elements have clear targets or not, falling diphthongs tend to have a dynamic spectral target.

As can be seen from Figure 10, the ellipses for both onset and offset elements of level diphthongs [iu ui] and their monophthongal counterparts heavily overlap with each other. That is, level diphthongs tend to have two spectral targets.

### 3.2.3. Dynamic aspects

Whether a dynamic spectral target or a sequence of two spectral targets, the production of diphthongs concerns a transition from an onset to offset component. It is thus of interest to check if dynamic aspects could characterize diphthong production. Following [30], [31], dynamic aspects were expressed as range and rate of spectral change for the second formant (F2). Table 2 summarized mean range and rate of F2 change for the diphthongs in Cangnan.

Table 2. Mean range in Hz and rate in Hz/ms of the F2 change for the diphthongs in Cangnan.

Diphthongs		male		female	
		ΔF2	rate	ΔF2	rate
Rising	ia	754	10.8	642	6.87
	ie	606	7.74	647	7.61
	iɔ	1132	16.17	1185	16.12
	ua	413	6.86	586	7.35
	ue	430	8.83	596	8.16
	ue	1040	14.08	1200	14.93
Falling	au	360	3.99	422	4.43
	ai	631	7.95	656	7.42
	əu	539	4.83	559	5.19
Level	iu	1353	19.79	1551	18.77
	ui	1402	23.37	1654	24.12

Previous studies show that F2 range and rate of change could characterize diphthongs when a language has a simple inventory of diphthongs ([30], [31]), but would have problems when a language has a complex inventory of diphthongs ([32], [33], [34]). [35] further points out that F2 range and rate of change could characterize diphthongs when falling diphthongs and rising diphthongs are separated; but it would have problems when falling and rising diphthongs are taken into account together.

The data of falling and level diphthongs in Cangnan are consistent with previous studies in general. The three falling diphthongs have different F2 range and rate of change: [ai] >

[əu] > [au] in both male and female speakers. And the two level diphthongs [iu ui] also have different F2 range and rate of change: [ui] > [iu]. But there is controversy regarding rising diphthongs. First, the data of F2 range is not consistent with that of F2 rate. Second, the data in male speakers is not consistent with that in female speakers. Third, a number of diphthong pairs such as [ua ue] and [ue iɔ] in both male and female speakers and [ia ie] in female speakers have similar F2 range of change. It seems that F2 rate of change is a better measure of spectral dynamics in diphthong production. Except for [ia] in female speakers, rising diphthongs exhibit a general ordering of F2 rate of change: [iɔ] > [ue] > [ia] > [ue] > [ie] > [ua].

In summary, different falling, level, and rising diphthongs are generally characterized by F2 rate of change within their own diphthong category in Cangnan.

## 4. Conclusions

This paper presents an acoustic phonetic description of the vowels and diphthongs in open syllables in Cangnan southern Min Chinese dialect. The results suggest a phonetically based vowel phonology of the Cangnan dialect.

First, it is of typological interest to observe a balanced but highly marked distribution of Cangnan monophthongs in the acoustic vowel space. It is balanced in a sense that the monophthongal vowels are quasi-equally distributed in backness dimension, and front, central and back vowels have a predictable lip rounding (c.f.: [28]). It is, however, highly marked regarding vowel height, namely there is more distinction of vowel height in central vowels than in back vowels, and further than in front vowels.

Second, the Cangnan diphthong data corroborates the proposal in [20] that diphthongs should be differentiated in terms of spectral dynamics. Both onset and offset elements have spectral steady states in all Cangnan diphthongs. But the results suggest that rising and level diphthongs tend to have spectral targets for both onset and offset components, whereas falling diphthongs only have one dynamic spectral target. In other words, the production of falling diphthongs is a single articulatory event as that of monophthongs, while the production of rising and level diphthongs is a sequence of two spectral targets. Phonologically, a falling diphthong should be treated as a single vowel and being contrastive to its corresponding monophthong in spectral dynamics, while a rising or level diphthong is simply a sequence of monophthongal components. That is, there are three distinctive vowels [a] in Cangnan, one static [a], i.e. the monophthong, and two dynamic [a]s, i.e. the falling diphthongs [ai au]; and there are two distinctive [ə]s, namely the static monophthongal [ə] and the dynamic diphthongal [əu].

Last, but not least, it should be emphasized that vowel phonology could be language-specific across languages on the one hand, and it is, therefore, of theoretical interest to differentiate what is universal by adopting an evidence-based approach on the other hand (c.f.: e.g. [36], [37]).

## 5. Acknowledgements

This work is supported by Chinese National Social Science Foundation (Project number: 15BYY073) and CASS Innovation Project “Language universals and diversities: an experimental approach to the phonetics and grammar in Chinese dialects” (P.I.: Dr. HU Fang).

## 6. References

- [1] Duanmu, S. *Syllable Structure: The Limits of Variation*. Oxford University Press, 2008.
- [2] 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.
- [3] Malmberg, B. *Structural linguistics and human communication*, Berlin: Springer-Verlag, 1963.
- [4] Abercrombie, D. *Elements of general phonetics*, Edinburgh: Edinburgh University Press, 1967.
- [5] Catford, I. *Fundamental problems in phonetics*, Edinburgh: Edinburgh University Press, 1977.
- [6] Sweet, H. A *handbook of phonetics including a popular exposition of the principles of spelling reform*, Oxford: Clarendon Press, 1877.
- [7] Jones, D. *Outline of English phonetics* (2nd Ed.), New York: E. P. Dutton, 1922.
- [8] Pike, K. L. "On the phonemic status of English diphthongs", *Language*, 23: 151-159, 1947.
- [9] Lehiste, I. & Peterson, G. E. "Transitions, glides, and diphthongs", *Journal of the Acoustical Society of America*, 33: 268-277, 1961.
- [10] Holbrook, A. and Fairbanks, G. "Diphthong formants and their movements", *Journal of Speech and Hearing Research*, 5: 38-58, 1962.
- [11] Hartman, L. M. "The segmental phonemes of the Peiping dialect", *Language*, 20: 28-42, 1944.
- [12] Luo, C. and Wang, J. *An Outline of General Phonetics* [in Chinese], Beijing: China Science Press, 1957.
- [13] Cheng, R. L. "Mandarin phonological structure", *Journal of Linguistics*, 2, 135-158, 1966.
- [14] Cheng, C. C. "A synchronic phonology of Mandarin Chinese", *Monograph on Linguistic Analysis*, No. 4, The Hague: Mouton, 1973.
- [15] Chao, Y.-R. *A grammar of spoken Chinese*, Berkeley: University of California Press, 1968.
- [16] You, R., Qian, N. and Gao, Z. "On the phonemic system of Putonghua" [in Chinese], *Chinese Language*, 5, 328-334, 1980.
- [17] Duanmu, S. *The Phonology of Standard Chinese* (2<sup>nd</sup> ed.), Oxford: Oxford University Press, 2007.
- [18] Chao, Y.-R. "The non-uniqueness of phonemic solutions of phonetic systems", *Bulletin of the Institute of History and Philology, Academia Sinica*, 4, 363-397, 1934.
- [19] Chao, Y.-R. *Studies in the Modern Wu Dialects*, Peking: Tsinghua University Research Institute Monograph, 4, 1928.
- [20] Hu, F. "Falling diphthongs have a dynamic target while rising diphthongs have two targets: Acoustics and articulation of the diphthong production in Ningbo Chinese", *Journal of the Acoustical Society of America*, Vol. 134, No. 5, Pt. 2, 4199, Nov 2013; full paper in Chinese in *Bulletin of Linguistic Studies*, 10: 12-37, Shanghai: Shanghai Lexicography Press, 2013.
- [21] Qiu, Y. and Hu, F. "Longchang vowels" [in Chinese], *Bulletin of Linguistic Studies*, 10: 38-51, Shanghai: Shanghai Lexicography Press, 2013.
- [22] Xia, L. and Hu, F. "Vowels and diphthongs in the Taiyuan Jin Chinese dialect", submitted.
- [23] Hu, F. and Zhang, M. "Diphthongized vowels in the Yi county Hui Chinese dialect", *Proceedings of Interspeech 2014*, 1703-1707, Singapore, 2014.
- [24] Hu, F. and Zhang, M. "On the diphthongized vowels in Qimen Hui Chinese dialect", *Proceedings of the 18<sup>th</sup> International Congress of Phonetic Sciences*, paper number 0100, Glasgow, U.K.: the University of Glasgow, 2015.
- [25] Wen, D. *The Cangnan Dialect* [in Chinese], Beijing: Language and Culture Press, 1991.
- [26] Boersma, P. & Weenink, D. "Praat: doing phonetics by computer" [Computer program]. Version 5.3.61, 2015.
- [27] Smith, J. O. and Abel, J. S. "Bark and ERB Bilinear Transforms", *IEEE Trans. Speech and Audio Proc.*, 7(6): 697-708, 1999.
- [28] Zhang, Q. and Hu, F. "The Vowel Inventory in the Xinfeng (Tieshikou) Hakka Dialect", *Proceedings of the 18th International Congress of Phonetic Sciences*, paper number 0228, Glasgow, U.K.: the University of Glasgow, 2015.
- [29] Barry, W. J. "Schwa vs. schwa +(r) in German", *Phonetica*, 52: 228-235, 1995.
- [30] Gay, T. "Effects of speaking rate on diphthong formant movements", *Journal of the Acoustical Society of America*, 44: 1570-3, 1968.
- [31] Gay, T. "A perceptual study of American English diphthongs", *Language and Speech*, 13: 65-88, 1970.
- [32] Manrique, A. "Acoustic analysis of the Spanish diphthongs", *Phonetica*, 36: 194-206, 1979.
- [33] Jha, S. K. "Acoustic analysis of the Maithili diphthongs", *Journal of phonetics*, 13: 223-230, 1986.
- [34] Bladon, A. "Diphthongs: A case study of dynamic auditory processing", *Speech Communication*, 4: 145-54, 1985.
- [35] Hu, F. "An acoustic analysis of diphthongs in Ningbo Chinese", *Proceedings of Eurospeech 2003*, pp. 801-804, Geneva, Switzerland, 2003.
- [36] Ren, H. On the Acoustic Structure of Diphthongal Syllables, Ph. D. dissertation, UCLA. Published in *UCLA Working Papers in Phonetics* 65, 1986.
- [37] Hu, F. *A Phonetic Study of the Vowels in Ningbo Chinese*, Ph. D. dissertation, City University of Hong Kong, 2005, published by China Social Sciences Press, Beijing, 2014.