INTERSPEECH 2019
September 15–19, 2019, Graz, Austria

Vowels and Diphthongs in the Xupu Xiang Chinese Dialect

Zhenrui Zhang[1], Fang Hu[2,1]

[1]Department of Linguistics, Graduate School, University of Chinese Academy of Social Sciences
[2]Institute of Linguistics, Chinese Academy of Social Sciences
1661829094@qq.com, hufang@cass.org.cn

Abstract

Based on an acoustic analysis of speech data from 10 speakers, 5 male and 5 female, this paper describes the phonetics and phonology of the vowels and diphthongs in the Xupu Xiang Chinese dialect. Results suggest that monophthongs and falling diphthongs should be grouped together, since the production of them is a single articulatory event. Falling diphthongs are composed of a dynamic spectral target, while monophthongs are composed of a static spectral target. But rising diphthongs are sequences of two spectral targets.

Index Terms: vowel (monophthong), falling diphthong, rising diphthong, the Xupu Xiang Chinese dialect.

1. Introduction

Chinese dialects generally have a simple syllable structure, which could basically be represented as CGVC (initial consonant, on-glide, vowel nucleus, and coda). However, there is debate about the structure of rimes (see [1], [2]), and a key issue concerns the phonetics and phonology of vowels and diphthongs. Some linguists 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 phonological analyses of Chinese vowels and diphthongs in the literature ([1], [11], [12], [13], [14], [15], [16], [17]). 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 are different. Chao pointed out that falling diphthongs are true diphthongs in Wu dialects, but rising diphthongs are not ([19]).

Recent acoustic and articulatory researches from several major Chinese dialects renewed the issue of vowels and diphthongs. Accumulative evidence from Wu dialects ([20], [21]), southwestern Mandarin ([22]), Jin ([23]), Hui dialects ([24], [25]), and Cangnan southern Min ([26]) suggested a dynamic theory of vowel production: monophhtongs have a static spectral target, diphthongized vowels and falling diphthongs have a dynamic target, and rising diphthongs are composed of two spectral targets ([27]). Thus, the diphthong [ai], is not a sequence of [a] and [i], but a single dynamic articulatory event, and should be treated as being phonologically contrastive to the monophthong [a]; by contrast, [ia] is a sequence of [i] and [a].

The Xupu (Lufeng) dialect belongs to the old varieties of the Xiang dialect family ([28]). The Xupu county (Lufeng town) is located in the west of the Hunan province. It is well known that the old varieties of Xiang and the Wu dialect family share some phonological commonalities. For instance, they both have voiced obstruents. However, the vowel phonology of the old varieties of Xiang is quite different. As shown in Table 1, Xupu has 8 monophthongs [ɿ ʅ i y a ɒ ɯ u], 4 falling diphthongs [ae ao ei əɯ], 6 rising diphthongs [ia iu iɛ ua uɛ yɛ], and 3 triphthongs [uei iəɯ iao] in open syllables. Note that the falling diphthong [ae] alternates with [ɛ]. That is, they are free variants. Less vowel distinctions are found in syllables with a nasal coda. And falling diphthongs only occur in open syllables. Due to the space limit, this paper focuses on monophthongs and diphthongs in open syllables.

Table 1: *Lufeng vowel inventory.*

vowels	C(G)V	C(G)VN
Monophthongs	i u ɯ y a ɒ ɿ ʅ	ẽ aŋ eŋ
Rising Diphthongs	iɛ iɒ iu uɛ ua yɛ	iẽ iaŋ uẽ uaŋ yẽ
Falling Diphthongs	ei əɯ ae ao	
Triphthongs	iəɯ iao uei	

2. Methodology

10 native speakers, 5 males and 5 females, provided speech data during the first author's fieldwork trip in January 2019. All of them were born and raised up in Xupu, and had no reported history of speech or hearing disorders. 3 Meaningful monosyllabic words were used as test words for each target vowel. Target words are onset-less syllables or syllables with a bilabial initial stop [p]. Each test word was placed in a carrier sentence [X, tʰəɯ^{34} X kɯ^{13} ni^{13} tʰẽ^{44}] "X, read X for you to listen". The speakers were instructed to read the randomized word list naturally in a normal speech rate. The 16-bit audio sounds were recorded through an AKG microphone directly into a laptop PC through a Lexicon I-O 22 USB sound card. The sampling rate is 11,025 Hz. 3 repetitions were recorded.

The target vowels were labelled in praat 6.0.29 ([29]). The diphthongs were annotated as being composed of three segments, i.e. an onset, an offset, and the transition connecting them. The lowest four formant frequencies were extracted from the midpoint of the steady state for the onset or offset segment. If there is no steady state, first formant (F1) minimum was taken as the target point for high vowels and F1 maximum was taken as the target point for low vowels. The duration of each segment was measured. And F2 range and rate of change were calculated for each diphthong.

http://dx.doi.org/10.21437/Interspeech.2019-1174

3. Results

3.1. Monophthongs

Figure 1 shows the distribution of 8 Xupu monophthongs [ɿ ʅ i u ɯ y a ɒ] in the acoustic vowel plane in male (upper) and female (lower) speakers. The acoustic vowel space is determined by the first formant (F1) as ordinate and second formant (F2) as abscissa with the origin of the coordinate to the top right. Coordinates are bark-scaled ([30]), but still labelled in Hertz. Each two-sigma ellipse is based on 45 sampled data points (3 test words × 3 repetitions × 5 speakers).

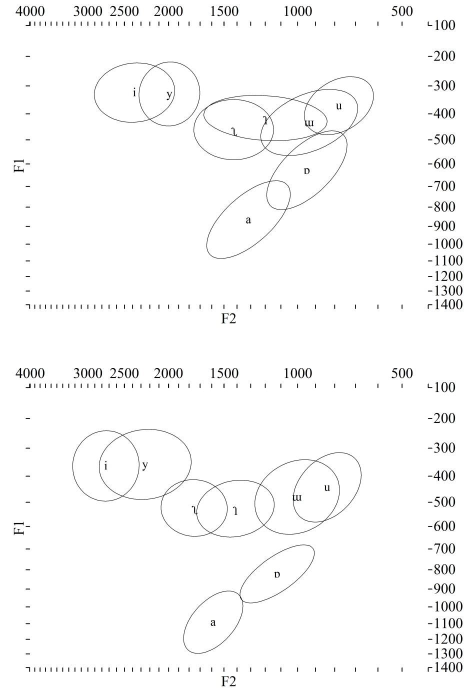

Figure 1: *2-sigma ellipses for the Xupu monophthongs in male (upper) and female speakers (lower).*

Table 2: *Means (in Hz) and one-way ANOVAs for F1, F2, and F3 of [ɯ u].*

	MALE		FEMALE	
	ɯ	u	ɯ	u
F1	408	354	454	420
	F=106.09, P<0.0001		F=23.99, P<0.0001	
F2	903	768	958	826
	F=162.92, P<0.0001		F=143.34, P<0.0001	
F3	2689	2491	3142	2958
	F=133.16, P<0.0001		F=75.59, P<0.0001	

Xupu distinguishes 2 levels of vowel height. There are 6 high vowels [ɿ ʅ i u ɯ y], and 2 low vowels [a ɒ]. The low vowels distinguish in backness. However, the rounded low back vowel [ɒ] is a bit higher than the unrounded low central vowel [a]. The 6 high vowels distinguish 2 levels of vowel backness. [i y] are front vowels, and [u ɯ] are back vowels. Both high front and back vowels distinguish in lip rounding.

The unrounded high back vowel [ɯ] was described as a rounded high back vowel [u] with certain degree of frication, and the rounded [u] was transcribed as a rounded semi-high back vowel [ʊ] in previous dialectological works ([28]). In other words, they ([u ʊ]) have same degree of backness and lip rounding, but distinguish in height. As shown in Figure 1, [u] ([ʊ]) is actually a bit higher than [ɯ] ([u]). That is probably due to the effect of lip rounding, since lip rounding lowers all formants. As shown in Table 2, all the first three formants, F1, F2, and F3, are significantly smaller in [u] than in [ɯ]. That is, [u ɯ] distinguish in lip rounding, rather than in vowel height.

Finally, the two apical vowels [ɿ ʅ] occupies a central high position in the acoustic F1/F2 vowel plane. This is a typical characteristic of apical vowels commonly observed in Chinese dialects, as they are historically originated from high front vowels in general ([31], [32]).

3.2. Diphthongs

The 4 falling diphthongs [ae ei əɯ ao] and the 6 rising diphthongs [iɛ iɒ iu uɛ ua yɛ] in Xupu were examined in terms of temporal organization, spectral property, and dynamic aspects.

3.2.1. *Temporal organization*

Figures 2 and 3 show mean durations in millisecond and in percentage respectively for diphthong components in falling and rising diphthongs in Xupu.

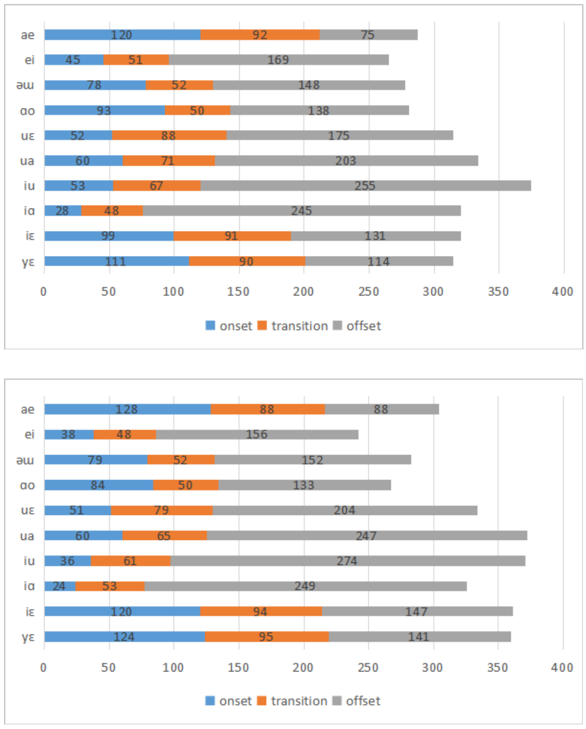

Figure 2: *Temporal organization in millisecond for diphthong components in male (upper) and female speakers (lower).*

It is apparent from the figures that temporal organization is quite consistent between male and female data. First, there is no clear pattern of temporal organization for falling diphthongs. [ae] has a quite balanced structure of temporal organization. The onset element in [ao əɯ] become shorter, and consequently the offset element become longer. The onset element becomes even shorter in [ei], so that the temporal organization for [ei] resembles those for rising diphthongs [iɒ iu uɛ ua], rather than for the other falling diphthongs. And [ao əɯ ei] also share a commonality in that they all have a short duration of transition.

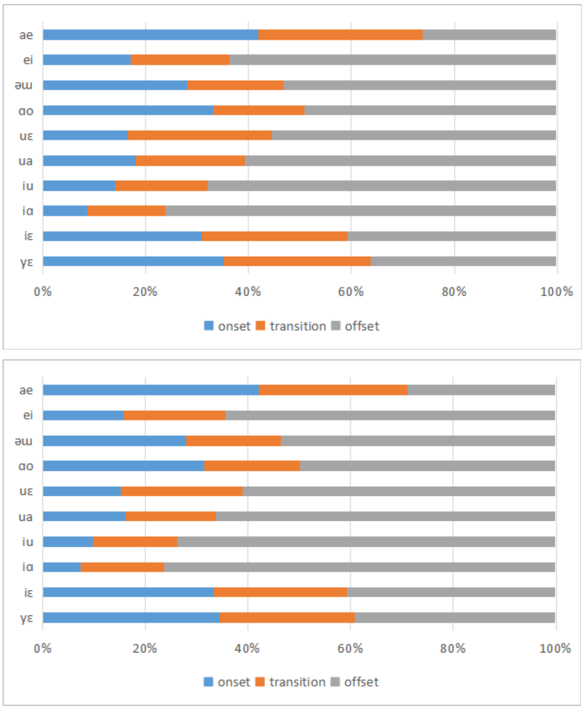

Figure 3: *Temporal organization in percentage for diphthong components in male (upper) and female (lower) speakers.*

The four rising diphthongs [iɒ iu uɛ ua] in Xupu exhibit a common pattern of temporal organization. That is, the offset element is the dominating component, which takes more than 60% of entire diphthong duration, and the onset element has a short duration, about 24-60 milliseconds, which takes about 10%-20% of entire diphthong duration. [iɛ yɛ] are exceptions. They have a relatively balanced structure of temporal organization, resembling the falling diphthong [ae].

3.2.2. Spectral properties

Figures 4-5 compare 2-sigma ellipses of the onset and offset elements in rising diphthongs and their monophthongal counterparts in the acoustic F1/F2 vowel plane. The IPAs with parentheses denote diphthong components.

As can be observed from the figures, ellipses for diphthong onsets [i u y] and offsets [a ɒ ɛ ɯ] extensively overlap with their corresponding monophthongs respectively, although there could be certain degree of coarticulatory effects. Exceptional cases are from [i] in [iɑ], and [ɛ] in [iɛ uɛ yɛ] (especially in male speakrs). Suffice it to conclude that rising diphthongs in Xupu tend to have two spectral targets, as both onset and offset elements have comparable spectral distributions to their monophthongal counterparts in general.

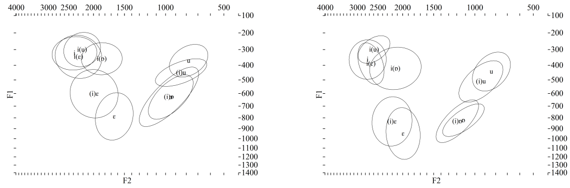

Figure 4: *2-sigma ellipses for the onset and offset of [iɒ iɛ iu] and corresponding monophthongs in male(left) and female (right) speakers.*

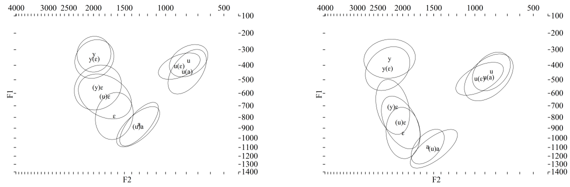

Figure 5: *2-sigma ellipses for the onset and offset of [ua uɛ yɛ] and corresponding monophthongs in male(left) and female (right) speakers.*

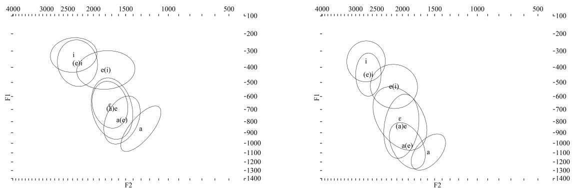

Figure 6: *2-sigma ellipses for the onset and offset of [ae ei] and corresponding monophthongs in male(left) and female (right) speakers.*

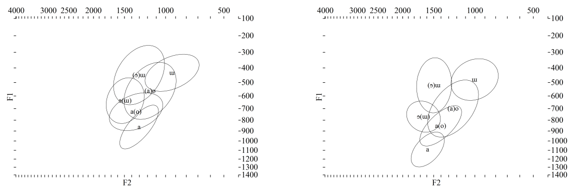

Figure 7: *2-sigma ellipses for the onset and offset of [əɯ ao] and corresponding monophthongs in male(left) and female (right) speakers.*

Figures 6-7 compare 2-sigma ellipses for the components in falling diphthongs and their monophthongal counterparts in the acoustic F1/F2 vowel plane. Diphthong components were denoted by the IPAs with parentheses.

First, falling diphthongs [ae ei ao əɯ] in Xupu all have short spectral movements, so that the ellipses for the onset and offset in a same falling diphthong extensively overlap with each other. As mentioned earlier, falling diphthong [ae] and the monophthong [ɛ] are free variants. 22 out of 45 samples in male speakers and 25 out of 45 samples in female speakers are pronounced as [ɛ]; the rest of them are produced as [ae]. It can be seen from the figures that [ae] and [ɛ] occupies a similar position in the acoustic vowel plane. In other words, they are both mid-low vowels, although [ae] are phonetically diphthongal and [ɛ] are monophthongal. It is apparent from

Figure 6 that the other front diphthong [ei] occupies a mid-high position in the acoustic vowel plane. Recall that Xupu monophthongs only have a high versus low distinction of vowel height, and there are no mid monophthongs. The situation is that the spectral regions for the mid-high and mid-low distinctions are occupied by dynamic vowels, i.e. the falling diphthongs. The phonetic evidence for this kind of phonological treatment is that diphthong elements [e] in [ae] and [ei], [o] in [ao], and [ə] in [əɯ] do not have comparable monophthongal counterparts. Moreover, the other diphthong element that has comparable monophthongal counterpart in falling diphthongs is quite variable. To conclude, it seems that the production of falling diphthongs in Xupu concerns with a dynamic spectral target, rather than a sequence of two spectral targets.

Second, the distinction between [ao əɯ] in the back vowel series is not in vowel height, but in peripherality. As can be seen from Figure 7, [ao] is peripheral, and [əɯ] is non-peripheral.

In summary, the production of falling diphthongs is a single articulatory event with a dynamic spectral target, while the production of rising diphthongs is a sequence of two articulatory events with their own spectral targets.

3.2.3. Spectral dynamics

Whether a diphthong is composed of a dynamic spectral target or a sequence of two spectral targets, the production of diphthongs includes a transition connecting the onset and offset components. It is proposed in the literature that spectral dynamics characterizes diphthong production in terms of the range and rate of second formant (F2) change ([33], [34]). Table 3 summarized mean range and rate of F2 change for the diphthongs in Xupu.

Table 3: *Mean ranges in HZ and rates in HZ/ms of the F2 change for the diphthongs in Xupu.*

Diphthongs		Male		Female	
		ΔF2	rate	ΔF2	rate
Falling	ei	487	9.55	494	10.29
	əɯ	262	5.04	245	4.71
	ao	221	4.42	209	4.18
	ae	209	2.27	233	2.65
Rising	iu	1461	21.81	1748	28.66
	iɑ	911	18.98	944	17.81
	uɛ	751	8.53	1050	13.29
	ua	486	6.85	523	8.05
	iɛ	431	4.74	466	4.96
	yɛ	274	3.04	276	2.91

First, it is generally observed from the table that a greater F2 range of change is correlated with a greater F2 rate of change. And the data is consistent across male and female speakers. The two exceptions are from the falling diphthong [ao] and the rising diphthong [iɑ] in female speakers: [ao] has a smaller F2 range of change but a greater F2 rate of change than [ae]; [iɑ] has a smaller F2 range of change but a greater F2 rate of change than [uɛ]. Second, the F2 rate of change serves as a better measure than the F2 range of change does. Falling and rising diphthongs have the same ordering in both male and female speakers in terms of F2 rate of change respectively: [ei] > [əɯ] > [ao] > [ae] and [iu] > [iɑ] > [uɛ] > [ua] > [iɛ] > [yɛ]. But the data is not consistent across male and female speakers in terms of F2 range of change. Third, falling and rising diphthongs should be taken into account separately, rather than together, as they concern spectral dynamics with opposite directions of movement in general. It has been reported in the literature that spectral dynamics would have problems in diphthong characterization in languages with a complex inventory of diphthongs, especially in those languages with both falling and rising diphthongs ([20], [22], [35], [36], [37], [38]).

4. Conclusion

This paper gives an acoustic phonetic description of the vowels and diphthongs in the Xupu Xiang Chinese dialect. Based on the acoustic phonetic data, an outline of the vowel phonology is given.

Monophthongs have a high versus low distinction, and there is no mid monophthongal vowels in Xupu. Interestingly, the acoustic phonetic data support a phonological analysis of the four falling diphthongs [ae ei ao əɯ] as dynamic mid vowels, as the production of falling diphthongs in Xupu is a single articulatory event with a dynamic spectral target. [ae] is a dynamic mid-low front vowel, and it alternates with the monophthong [ɛ]; [ei] is a dynamic mid-high front vowel; [ao] is a dynamic peripheral back vowel; [əɯ] is a dynamic non-peripheral back vowel. Therefore, Xupu vowels actually have a four-level distinction of vowel height. In contrast, rising diphthongs in Xupu are composed of two spectral targets, since both onset and offset elements in rising diphthongs have comparable spectral distributions to their monophthongal counterparts. In other words, in the Xupu dialect, a rising diphthong is a sequence of monophthongal components, while a falling diphthong should phonologically be treated as a single vowel and being contrastive to monophthongs. Finally, spectral dynamics help characterize diphthong production especially in terms of F2 rate of change, given that falling and rising diphthongs are accounted for separately.

5. Acknowledgements

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

6. References

[1] S. Duanmu, *Syllable Structure: The Limits of Variation.* Oxford University Press, 2008.
[2] S. Duanmu, *Syllable and syllable structure in Chinese.* In R. Sybesma (editor-in-chief), W. Behr, Y. Gu, Z. Handel, C.-T. J. Huang, and J. Myers, (eds.) *Encyclopedia of Chinese Language and Linguistics.* Leiden: Brill, 2016.
[3] B. Malmberg, *Structural linguistics and human communication*, Berlin: Springer-Verlag, 1963.
[4] D. Abercrombie, *Elements of general phonetics*, Edinburgh: Edinburgh University Press, 1967.
[5] I. Catford, *Fundamental problems in phonetics*, Edinburgh: Edinburgh University Press, 1977.
[6] H. Sweet, *A handbook of phonetics including a popular exposition of the principles of spelling reform.* Oxford: Clarendon Press, 1877.
[7] D. Jones, *Outline of English phonetics (2nd Ed.).* New York: E. P. Dutton, 1922.

[8] K. L. Pike, "On the phonemic status of English diphthongs," *Language,* 23, pp. 151-159, 1947.
[9] I. Lehiste, and G. E. Peterson, "Transitions, glides, and diphthongs," *Journal of the Acoustical Society of America*, 33, pp. 268-277, 1961.
[10] A. Holbrook, and G. Fairbanks, "Diphthong formants and their movements," *Journal of Speech and Hearing Research* 5, pp. 38-58, 1962.
[11] L. M. Hartman, "The segmental phonemes of the Peiping dialect," *Language*, 20, pp. 28-42, 1944.
[12] C. Luo, and J. Wang, *An Outline of General Phonetics* [in Chinese]. Beijing: China Science Press, 1957.
[13] R. L. Cheng, "Mandarin phonological structure," *Journal of Linguistics*, 2, pp. 135-158, 1966.
[14] C. C. Cheng, *A synchronic phonology of Mandarin Chinese. Monograph on Linguistic Analysis*, No. 4, The Hague: Mouton.
[15] Y.-R. Chao, 1968. "A grammar of spoken Chinese," Berkeley: University of California Press, 1973.
[16] R. You, N. Qian, and Z. Gao, "On the phonemic system of Putonghua" [in Chinese], *Chinese Language*, no. 5, pp. 328-334, 1980.
[17] C. Luo, and J. Wang, "An Outline of General Phonetics" [in Chinese], Beijing: China Science Press, 1957.
[18] Y.-R. Chao, "The non-uniqueness of phonemic solutions of phonetic systems," *Bulletin of the Institute of History and Philology, Academia Sinica,* 4, pp. 363-397, 1934
[19] Y.-R. Chao, *Studies in the Modern Wu Dialects,* Peking: Tsinghua University Research Institute Monograph, 4, 1928.
[20] F. Hu, "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, pp. 4199, Nov 2013; full paper in Chinese in Bulletin of Linguistic Studies, 10, pp. 12-37, Shanghai: Shanghai Lexicography Press, 2013
[21] Y. Yue and F. Hu, "Vowels and diphthongs in Hangzhou Wu Chinese dialect," *Proceedings of Interspeech 2018,* Hyderabad, pp. 207-211, 2018.
[22] Y. Qiu, and F. Hu, "Longchang vowels" [in Chinese], *Bulletin of Linguistic Studies,* 10, pp. 38-51, Shanghai: Shanghai Lexicography Press, 2013.
[23] L. Xia, and F. Hu, "Vowels and diphthongs in the Taiyuan Jin Chinese dialect," *Proceedings of Interspeech 2016*, pp. 993-997, San Francisco, USA, 2016.
[24] F. Hu, and M. Zhang, "Diphthongized vowels in the Yi county Hui Chinese dialect," *Proceedings of Interspeech 2014*, pp. 1703-1707, Singapore, 2014.
[25] F. Hu, and M. Zhang, "On the diphthongized vowels in Qimen Hui Chinese dialect," *Proceedings of the 18th International Congress of Phonetic Sciences*, paper number 0100, Glasgow, U.K.: the University of Glasgow, 2015
[26] F. Hu & C. Ge, "Vowels and Diphthongs in Cangnan Southern Min Chinese Dialect," *Proceedings of Interspeech 2016*, pp. 978-982, San Francisco, USA, 2016.
[27] F. Hu, "Toward a dynamic theory of vowel production," *The Journal of the Acoustical Society of America*, 142 (4), Pt.2, pp. 2579, 2017.
[28] K. He, *Studies on the Xupu Dialect* [in Chinese], Changsha: Hunan Educational Press, 1999.
[29] P. Boersma, and D. Weenink, Praat: doing phonetics by computer [Computer program]. Version 6.0.29, 2018.
[30] J. O. Smith and J. S. Abel, "Bark and ERB Bilinear Transforms." *IEEE Trans. Speech and Audio Proc.*, 7(6), pp. 697-708, 1999.
[31] F. Hu, "On the distinctive features for the high front vowels in Ningbo and Suzhou Wu Chinese – with reference to sound changes of high vowels" [in Chinese], *Zhongguo Yuwen (Chinese Language)*, no. 5, pp. 455-465, 2007.
[32] F. Hu and F. Ling, "Fricative vowels as an intermediate stage of vowel apicalization." *Language and Linguistics*, vol. 20, no. 1, pp.1-14, 2019.
[33] T. Gay, "Effects of speaking rate on diphthong formant movements," *Journal of the Acoustical Society of America,* 44, 1570-1573, 1968.
[34] T. Gay, "A perceptual study of American English diphthongs," *Language and Speech* 13, 65-88. 1970.
[35] A. Manrique, "Acoustic analysis of American English diphthongs," *Language and Speech,* 13, 223-230, 1986.
[36] S. K. Jha, "Acoustic analysis of diphthongs," *Journal of phonetics*, 13, 223-230, 1986.
[37] A. Bladon, "Diphthongs: A case study of dynamic aoditoryproceeding," *Speech Communication*, 4, 145-154, 1985.
[38] F. Hu, "An acoustic analysis of diphthongs in Ningbo Chinese," *Proceedings of Eurospeech 2003*, 801-804, Geneva, Switzerland, 2003.