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

# Vowels and Diphthongs in the Taiyuan Jin Chinese Dialect

*Liping Xia, Fang Hu*
Institute of Linguistics, Chinese Academy of Social Sciences
xlphg@163.com, hufang@cass.org.cn

## Abstract

On the basis of an acoustic phonetic analysis of monophthongs and diphthongs, this paper describes vowel phonology in the Taiyuan Jin dialect. The results show that Taiyuan has a comparable but different vowel inventory for C(G)V versus C(G)VN syllables. And the vowel contrast is dramatically reduced in checked syllables. The asymmetry between falling and rising diphthongs suggests a dynamic account of vowels, rather than a sequential taxonomy of vowels into monophthongs and diphthongs. Phonetically, monophthongs are composed of a static spectral target, falling diphthongs are composed of a dynamic spectral target, and rising diphthongs are sequences of two spectral targets. Phonologically, falling diphthongs are grouped with monophthongs, rather than rising diphthongs.

**Index Terms**: vowels, monophthongs, rising diphthongs, falling diphthongs, the Taiyuan Jin dialect

## 1. Introduction

Taiyuan is the capital city of Shanxi province in northern China. The Taiyuan dialect belongs to the Jin family of Chinese dialects. The Jin dialect family are known for its checked syllables, which distinguishes them from the surrounding Mandarin dialects ([1]). Taiyuan phonology was described in dialectological works ([2], [3], [4]). But little phonetic work has been done for Taiyuan vowels.

Similar to other Chinese dialects, Taiyuan Jin has a simple syllable structure that can be described sequentially as CGVC ([5], [6]). First, syllable-initial consonant is optional, and there is no consonant cluster. Second, the coda is optional, too. It could be a vocalic off-glide G, a nasal N [ŋ] (or nasalized vowel), or a glottal stop S [ʔ]. Third, the nuclear vowel V and the preceding on-glide G form a rising diphthong in CGV, CGVN, and CGVS syllables; the nuclear vowel V and the following off-glide form a falling diphthong in CVG syllables. Fourth, Taiyuan Jin does not contrast in vowel length, but vowels and rising diphthongs are phonetically short in CVS, CVN, CGVS, and CGVN syllables. Last, CVGN/S is not legitimate. That is, Taiyuan does not have falling diphthongs in close syllables.

Table 1. *Vowel inventory of Taiyuan dialect.*

| | -vocalic | -ŋ | -ʔ |
|---|---|---|---|
| monoph. | i y ɿ u<br>ɤ<br>(ɐ) a | i y o<br>ə ɒ | ə<br>ɐ |
| rising diph. | uɤ (uo)<br>ia ua | iɛ̃ yɛ̃ iɒ̃ uɒ̃ | iə yə uə<br>ia ua |
| falling diph. | ei əu<br>ai au | | |
| triphthongs | uei iəu<br>uai iau | | |

Table 1 summarizes the vowel inventory in the Taiyuan Jin. There are 6 monophthongal vowels [ɿ i u y a ɤ] in CV syllables. [ɿ] is an apical vowel that co-occurs with an coronal sibilant [ts ts$^h$ s]. And there is an additional [ɐ] for rhoticized syllables, which is a result of a morphological process. Only two vowels [a ɤ] occur as nuclear elements in CGV syllables.

There are usually less vowel distinctions in restricted environments. There is no apical vowel in CVN syllables, but a new vowel [ɛ] in CGVN syllables. Therefore Taiyuan maintains a 6-vowel system in nasal-coda syllables. In contrast, there is a 2-vowel contrast [ə ɐ] in CVS syllables.

Taiyuan has four falling diphthongs and four paired triphthongs. Due to the space limit, this paper focuses on vowels and diphthongs in Taiyuan.

## 2. Methodology

10 speakers, 5 males and 5 females, were recorded during the fieldwork trip in the spring of 2014. They were all native adult speakers without any reported history of speech and hearing disorders.

Meaningful monosyllabic words were used as test words. Each target vowel has 2 test words, preferably one with a zero-initial and the other with an initial of labial stop. Each test word was placed in a carrier sentence [X, uo ŋiɛ̃ lə kə X ʈʂɿ] "X, I have read an X". 5 repetitions were recorded.

The 16-bit audio sound was recorded directly into a laptop PC through a TerraTec DMX 6Fire USB sound card with a SHURE SM86 microphone. The sample rate is 11,025 Hz.

The target vowels were labelled in Praat 5.3.48. Each diphthong was annotated as being composed of a steady state of onset and offset elements, if any, and the connecting transition. The lowest four formants were extracted in the midpoint of each target element. The duration of each segment was also measured. And ranges and rates of spectral changes, namely ranges and rates of the second formant (F2) changes, were further calculated for each diphthong.

## 3. Results

### 3.1. Monophthongs

Figure 1 shows the distribution of the 6 Taiyuan monophthongs [ɿ i u y a ɤ] for male (left) and female (right) speakers respectively in a two-dimensional acoustic plane ( $F_1$ against $F_2$) with the origin of the axes to the top right of the plot. The ordinates are bark scaled, but the values along the axes are still labelled in Hz. Each 2-sigma ellipse was based on 50 data points (5 repetitions × 5 speakers × 2 test syllables).

The six monophthongs have a triangular distribution in the acoustic vowel space. There is a 2-way vowel height, a 2-way backness, and a 2-way lip rounding. [ɿ i y u] are high vowels, and [a] is low; [i y] are front, and [ɤ u] are back; [ɿ i a ɤ] are unrounded, and [y u] are rounded.

Figure 2 plots rhoticized vowel [ɐ], and two short vowels [ɐʔ əʔ] with the monophthongs. It's clear from the acoustic F1/F2 vowel space that both rhoticized [ɐ] and short [ɐʔ] occupy a low vowel position, as their ellipses heavily overlap

Copyright © 2016 ISCA

http://dx.doi.org/10.21437/Interspeech.2016-249

with that for the corresponding monophthong [a]. The other short vowel [əʔ] is distributed in a mid central position in the acoustic vowel space and is higher than the low vowel. It should be pointed out that the rhoticized [ɐ] is usually realized as a plain low vowel in Taiyuan.

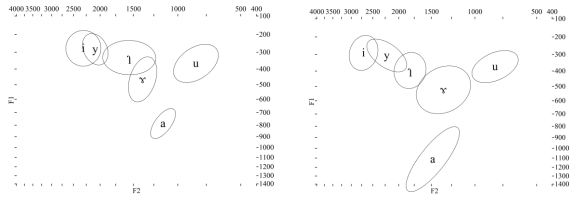

Figure 1: *2-sigma ellipses for the Taiyuan monophthongs in male speakers (left) and female speakers (right).*

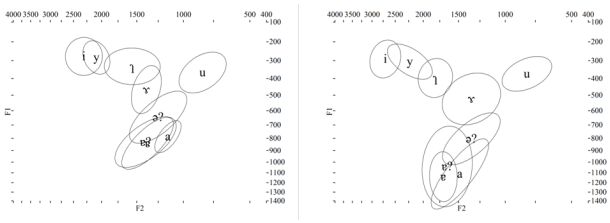

Figure 2: *2-sigma ellipses for the Taiyuan monophthongs, rhoticized vowel, and short vowels in male speakers (left) and female speakers (right).*

The short vowels are distributed in checked syllables, which is complementary to open syllables. The six-vowel distinction in height, backness and lip rounding in CV syllables is reduced to a 2-way distinction in height in checked syllables. The low vowel [ɐʔ] is retained, and other distinctions are neutralized into a short schwa [əʔ]. Moreover, the syllable-final glottal stop is often deleted in connected speech. It seems that duration would be the only difference between the long [a] and short [ɐʔ], as they share a similar low vowel quality.

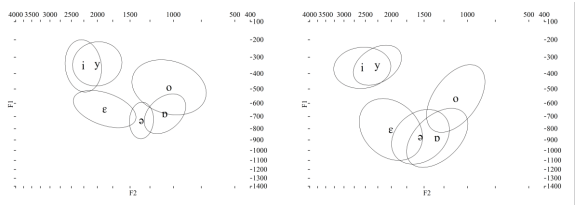

Figure 3: *2-sigma ellipses for the Taiyuan vowels in CVN syllables in male speakers (left) and female speakers (right).*

Figure 3 shows the distribution of vowels in CVN syllables in acoustic vowel spaces. Taiyuan maintains a six-vowel distinction in CVN syllables. There is a 2-way distinction in height [i y o] versus [ɛ ə ɒ], a 2-way distinction in backness [i y ɛ] versus [o ə ɒ], and a 2-way distinction of lip rounding [i ɛ ə] versus [y o ɒ].

It's interesting to note that [ə] occurs only in close syllables, [ɒ] occurs only in CVN and nasalized syllables, and [ɛ] occurs only in nasalized syllables. That is, they are not found in open CV syllables. Synchronically, differences in vowel distinction in different phonetic environments add complexity to vowel phonology of the Taiyuan Jin. Diachronically, losses of nasal ending and/or syllable-final glottal stop are common phonetic and/or phonological processes in Taiyuan Jin as well as in other Chinese dialects. And thus the emergent [ɛ ə ɒ] could probably add one more level, i.e. the level of mid vowels, to the current high-low distinction of vowels in the future.

### 3.2. Diphthongs

There are usually both falling and rising diphthongs in Chinese dialects. However, there is controversy regarding the nature of diphthongs. The key issue is to differentiate if a diphthong is a single phonemic unit or a sequence of vowels. For instance, [7] proposed that genuine diphthongs are single phonemes whereas the pseudo ones consist of two phonemes each. [8] indicated that only falling diphthongs are genuine diphthongs while rising diphthongs are pseudo in Wu Chinese dialects. Phonologically, falling diphthongs alternate with monophthongs, for instance [ai]/[ɛ] and [au]/[ɔ] or [ɒ] in a number of Chinese dialects ([9]), but rising diphthongs seldom alternate with monophthongs. [10] renewed the dichotomy between falling and rising diphthongs on the basis of acoustic and lingual kinematic data, and 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 phonologically distinctive to the monophthongal vowel [a]; in contrast, the diphthong [ia] is a sequence of [i] and [a]. And acoustic studies from Southwestern Mandarin ([11]), Southern Min ([12]), and Hui dialects ([13], [14]) all demonstrate similar results.

There is a 2-way distinction of height between the nuclear vowels [a] and [e]/[ə] in falling diphthongs [ai au ei əu] in Taiyuan. [ia ua] are typical rising diphthongs, but [uo/uɤ] is phonologically different. [ɤ uɤ uo] are sociolinguistic variants of one phoneme /ɤ/: [ɤ] from the old variety, [uo] from the new variety, and [uɤ] from an intermediate one. The discussion of [ɤ] is included in the section of monophthongs. There are few samples of [uɤ], and therefore the following discussion is based on the samples of [uo].

#### *3.2.1. Temporal structure*

The mean durations in millisecond and in percentage of the two elements and transition of diphthongs were summarized in the bar charts in Figure 4 and 5 respectively. As mentioned earlier, each target diphthong has two test syllables, with and without an initial consonant. The duration data is based on syllables with an initial consonant.

It's apparent from the figure that falling and rising diphthongs differ in temporal organization. In rising diphthongs [ia ua], the nuclear vowel [a] is the dominating element and has the longest duration, about 70% of the total duration, and the on-glide [i] or [u] has a very short duration, about 22-37 ms. In contrast, falling diphthongs have a relatively balanced temporal organization. Both nuclear vowel and the off-glide generally occupy about 30-40% of total duration in a falling diphthong. The transitions occupy about 20-30% of the duration in most cases of both falling and rising diphthongs in Taiyuan. This is quite short, as compared to the cases in other Chinese dialects such as Mandarin and Wu ([15], [16]). [uo] occurs as an exceptional case, as it has a temporal

structure similar to falling diphthongs, rather than rising diphthongs.

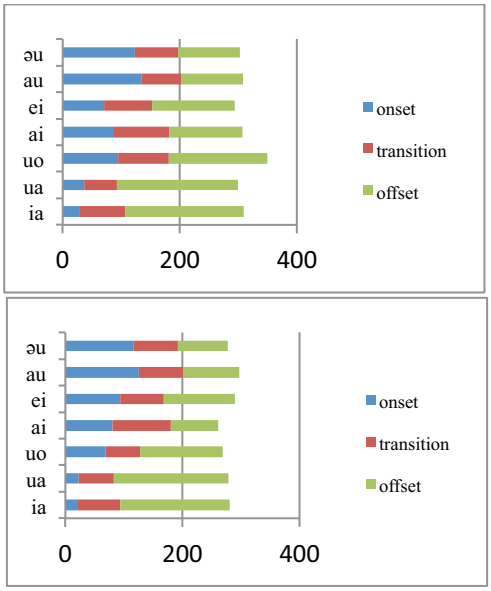


Figure 4: *Temporal structures (in millisecond) of the Taiyuan diphthongs in male (upper) and female (lower) speakers.*

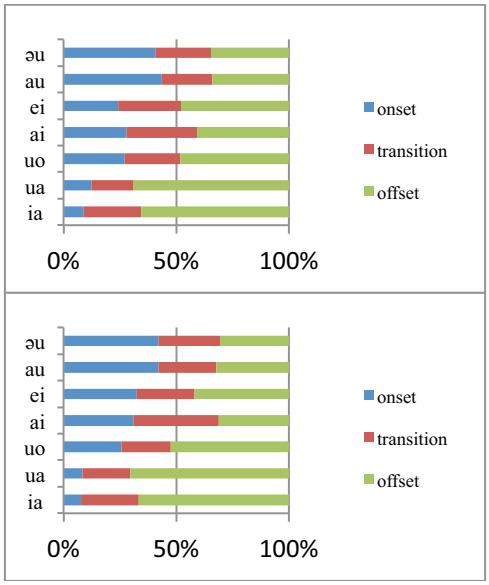


Figure 5: *Temporal structures (in percentage) of the Taiyuan diphthongs in male (upper) and female (lower) speakers.*

### 3.2.2. *Spectral properties*

Figure 6 shows the distribution of 2-sigma ellipses for the vocalic elements in the Taiyuan falling diphthongs in acoustic vowel spaces. The ellipses for corresponding monophthongs [i u a] were superimposed for comparison.

It can be seen from the figures that both nuclear vowel and off-glide of the falling diphthongs [ai ei] and [au əu] occupy a somewhat neutralized position along their onset and offset target monophthongs [a i] and [a u] respectively. Generally, the production of [ai ei] concerns a spectral change from [a] to [i], and the production of [au əu] from [a] to [u]. In summary, comparisons between vocalic elements in falling diphthongs and the corresponding monophthongs do not suggest that falling diphthongs are composed of two spectral targets. Rather, falling diphthongs are spectrally dynamic vowels.

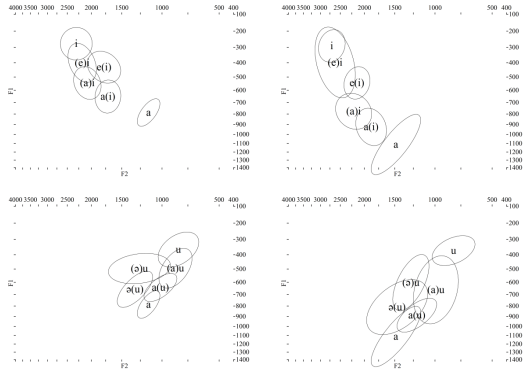

Figure 6: *2-sigma ellipses for the vocalic elements in Taiyuan falling diphthongs [*ai ei*] (upper) and [*au əu*] in male (left) and female (right) speakers.*

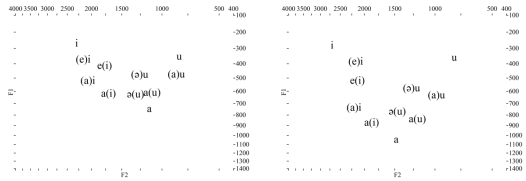

Figure 7: *Falling diphthongs [*ai ei au əu*] and the monophthongs [*i a u*] in male (left) and female (right) speakers in Taiyuan.*

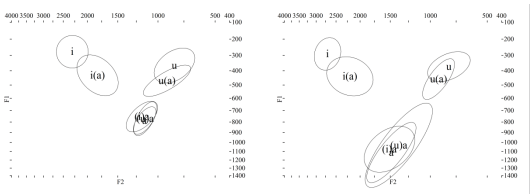

Figure 8: *2-sigma Ellipses for the vocalic elements in Taiyuan rising diphthongs in male (left) and female (right) speakers.*

From the production point of view, falling diphthongs and monophthongs share a commonality in that both involve a single articulatory event, and they differ in that the latter is a static event while the former is a dynamic one. Phonologically falling diphthongs are contrastive to monophthongs. There are two dynamic front vowels and two dynamic back vowels in Taiyuan: [ai ei] are dynamic front vowels, and [au əu] are dynamic back vowels. As mentioned earlier, monophthongs in Taiyuan have a 2-way distinction in height. Figure 7 plots the falling diphthongs onto the acoustic vowel triangle [i a u] based on the mean formant data. It's interesting to observe from the figure that the two dynamic front vowels [ai ei] differ in vowel height, namely the mid-high [ei] versus mid-low [ai],

whereas the two dynamic back vowels [au əu] differ in peripherality, namely the peripheral [au] versus a centralized [əu].

Figure 8 compares the distribution of the vocalic elements in rising diphthongs [ai au] with their corresponding monophthongs [i a u] in the acoustic $F_1/F_2$ plane. The 2-sigma ellipses for diphthong elements were based on samples in CV syllables. That is, each has 25 data points. As mentioned earlier, [uo] is a socio-phonetic variant of /ʁ/, and is thus excluded from discussion.

The production of rising diphthongs in Taiyuan demonstrates a different pattern to that of falling diphthongs. First of all, the ellipses for the nuclear vowel [a] in both [ia] and [ua] heavily overlap with the monophthong [a]. This suggests that, as comparable to the production of monophthong [a], there is a spectral target for the nuclear vowel [a] in the production of rising diphthongs in Taiyuan. Second, there is a certain degree of overlap between the ellipses for the on-glides in [ia ua] and their corresponding monophthongs respectively, but the ellipses for the onglides are meanwhile shifted towards the nuclear [a] in the acoustic vowel spaces. It seems that the production of the on-glides in [ia ua] exhibits an effect of anticipatory coarticulation.

From the above discussion, we can see that the rising diphthongs in Taiyuan dialect have a clear distinction between onsets and offsets. The offsets get each target very well, while the onsets have a position more or less under their targets. Furthermore, the onsets in V syllables are more frontless than in CV syllables, it can reflect that in V syllables, the [i] [u] are more like semi-vowels than monophthongs.

### *3.2.3. Spectral dynamics*

Spectral dynamics was expressed as range and rate of spectral change for the second formant (F2) ([17], [18]). Table 2 summarized mean range and rate of F2 change for the diphthongs in Taiyuan.

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

| male | ia | ai | ei | ua | au | əu |
|---|---|---|---|---|---|---|
| ΔF2 | 641 | 304 | 351 | 231 | 120 | 58 |
| rate | 8.1 | 3.2 | 4.3 | 3.4 | 1.7 | 0.8 |
| female | ia | ai | ei | ua | au | əu |
| ΔF2 | 690 | 553 | 443 | 458 | 221 | 206 |
| rate | 9.6 | 5.6 | 6.0 | 8.2 | 2.9 | 2.8 |

Previous studies suggest that F2 range and rate of change characterize diphthongs in a language with a simple inventory of diphthongs ([17], [18]), but would have problems in a language with a complex inventory of diphthongs ([19], [20], [21]). [22] further points out that F2 range and rate of change face with problems when falling and rising diphthongs are taken into account together.

The Taiyuan case has its own characteristics. First, falling diphthongs could hardly be characterized by F2 range or rate of change. [au əu] have similar F2 range and rate of change in female speakers. Second, F2 range of change is not consistent across speakers. The F2 range of change is smaller than for [ei] in male speakers, but is bigger for [ai] than for [ei] in female speakers. Third, falling diphthongs [ai au] have a substantially smaller F2 range and rate of change than their corresponding rising diphthongs [ia ua] respectively in both male and female speakers. That is, F2 rate of change serves as a parameter to distinguish falling diphthongs from their corresponding rising diphthongs. This could be attributed to the fact that the falling diphthongs [ai au] and their corresponding rising diphthongs [ia ua] are very different in terms of temporal organization as well as spectral targets in Taiyuan. It seems that spectral dynamics is not universal for the characterization of diphthongs, but could be diphthong-dependent and language-dependent.

## 4. Conclusion

This paper gives an acoustic phonetic analysis of the monophthongs and diphthongs in the Taiyuan Jin dialect. Results show that Taiyuan monophthongs demonstrate a 2-way distinction in height, backness and lip rounding. Vowel inventories are comparable in C(G)V and C(G)VN syllables, whereas vowel distinctions are dramatically reduced in checked syllables.

Fine-gained phonetic details reveal that falling diphthongs and rising diphthongs differ in temporal organization, spectral property, and spectral dynamics. The asymmetry between falling and rising diphthongs suggests a dynamic account of vowels, rather than a sequential taxonomy of vowels into monophthongs and diphthongs. Phonetically, monophthongs are composed of a static spectral target, falling diphthongs are composed of a dynamic spectral target, and rising diphthongs are sequences of two spectral targets. Phonologically, falling diphthongs are grouped with monophthongs, rather than rising diphthongs. That is, falling diphthongs and monophthongs are contrastive vowel phonemes in Taiyuan; rising diphthongs are sequences of vowels. And it is intriguing to observe that the 2 dynamic vowels, i.e. falling diphthongs [ai ei] contrast in vowel height, and the other 2 dynamic vowels [au əu] contrast in peripherality.

## 5. Acknowledgement

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 dialect" (P.I.: Dr. HU Fang).

## 6. Reference

[1] Institute of Linguistics & Institute of Ethnology and Anthropology, Chinese Academy of Social Sciences; Language Information Sciences Research Centre, City University of Hongkong. *Language Atlas of China* (2nd ed.) [In Chinese], Beijing: The Commercial Press, 2012.

[2] Wen, D. The homophony syllabary of Taiyuan dialect [In Chinese], *Linguistic Researches*, 3: 46-51, 1992.

[3] Shen, M. Re-classification of Jin dialect family [In Chinese], *Dialect*, 4: 343-356, 2006.

[4] Shen, M. *Dictionary of Taiyuan Dialect* [In Chinese], Nanjing: Jiangsu Education Publishing, Ltd, 1998.

[5] Duanmu, S. *Syllable Structure: The Limits of Variation*. Oxford University Press, 2008.

[6] 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.

[7] Collier, R. "Some acoustic and physiological observations on diphthongs", *Language and speech*, 25, 305-323, 1982.

[8] Chao, Y.-R. *Studies in the Modern Wu Dialects*, Peking: Tsinghua University Research Institute Monograph, 4, 1928.
[9] Yuan, J. *An outline of Chinese Dialects (2nd ed.)* [In Chinese], Beijing: Language & Culture Press, 2001.
[10] 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.
[11] Qiu, Y. and Hu, F. "Longchang vowels" [in Chinese], *Bulletin of Linguistic Studies*, 10: 38-51, Shanghai: Shanghai Lexicography Press, 2013.
[12] Hu, F. and Ge, C. "Vowels and diphthongs in the Cangnan Southern Min Chinese dialect", submitted.
[13] Hu, F. and Zhang, M. "Diphthongized vowels in the Yi county Hui Chinese dialect", *Proceedings of Interspeech 2014*, 1703-1707, Singapore, 2014.
[14] Hu, F. and Zhang, M. "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.
[15] Ren, H. On the Acoustic *Structure of Diphthongal Syllables*, Ph. D. dissertation, UCLA. Published in *UCLA Working Papers in Phonetics* 65, 1986.
[16] 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.
[17] Gay, T. "Effects of speaking rate on diphthong formant movements", *Journal of the Acoustical Society of America*, 44: 1570-3, 1968.
[18] Gay, T. "A perceptual study of American English diphthongs", *Language and Speech*, 13: 65-88, 1970.
[19] Manrique, A. "Acoustic analysis of the Spanish diphthongs", *Phonetica*, 36: 194-206, 1979.
[20] Jha, S. K. "Acoustic analysis of the Maithili diphthongs", *Journal of phonetics*, 13: 223-230, 1986.
[21] Bladon, A. "Diphthongs: A case study of dynamic auditory processing", *Speech Communication*, 4: 145-54, 1985.
[22] Hu, F. "An acoustic analysis of diphthongs in Ningbo Chinese". *Proceedings of Eurospeech 2003*, pp. 801-804, Geneva, Switzerland, 2003.