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

Effects of Subglottal-Coupling and Interdental-Space on Formant Trajectories during Front-to-Back Vowel Transitions in Chinese

Shuanglin Fan [1], Kiyoshi Honda [1], Jianwu Dang [1, 2], Hui Feng [3]

[1] Tianjin Key Laboratory of Cognitive Computation & its Applications, Tianjin University, Tianjin, China
[2] School of Information Science, Japan Advanced Institute of Science and Technology, Ishikawa, Japan
[3] School of Foreign Languages and Literature, Tianjin University, Tianjin, China
fshuanglin@126.com, khonda@sannet.ne.jp, jdang@jaist.ac.jp, fenghui@tju.edu.cn

Abstract

Discontinuity of the second formant (F2 discontinuity) is often found during back-to-front vowel transitions, and it has been thought due to two possible effects: acoustic coupling with the subglottal tract (subglottal-coupling effect, SCE) and traveling anti-resonance of the interdental space (interdental-space effect, ISE). Although both are possible to appear together, either of the two is common to find in many spectrographic observations, and how to distinguish from one another is often puzzling. This study aims at exploring manifestations of the two effects in Chinese triphthongs through acoustic analysis on front-to-back vowel sequences. Test utterances were recorded from five Chinese speakers with simultaneous measurement of subglottal resonance via a vibration sensor adhered to their necks. Results revealed that F2 discontinuity occurs near the second subglottal formant (SgF2) but not always, and discontinuity of both F2 and F3 is more common that occurs with a short time lag, suggesting predominance of ISE rather than SCE in the data.

Index Terms: F2 discontinuity, subglottal coupling, interdental space, anti-resonance

1. Introduction

The quantal theory proposed by Stevens noted nonlinear relations between acoustics and articulation [1]. Recent related discussions highlight the evidence of distinctive feature enhancement, which suggests articulatory-to-acoustic relations that supplement distinctive contrasts of sounds leading to perceptual saliency [2]. In the exploration of distinctive feature enhancement to contrast [$\pm$ back] features, brief discontinuity of the second formant frequency (F2 discontinuity) has been pointed out as evidence that contrasts the features. F2 discontinuity occurs near the frequency of the second subglottal formant (SgF2) during back-to-front vowel transitions [3], and it is called the subglottal-coupling effect (SCE). While SCE has been examined in previous studies [4] [5], another causal mechanism of F2 discontinuity is also proposed, which is called the interdental-space effect (ISE) [6]. ISE argues that it is the moving anti-resonance introduced into the vocal tract transfer functions that gives rise to formant discontinuity, but its role for feature enhancement has been underexplored.

The subglottal-coupling effect (SCE) refers to acoustic manifestation caused by interaction between the vocal and subglottal tracts, which is observed as F2 discontinuity on spectrographs. This discontinuity is due to resonance instability caused by the second subglottal formant (SgF2) and observed during transitions between front and back vowels [7] [8]. When F2 reaches the stationary frequency region of SgF2 in such a transition, F2 is suppressed by SgF2, resulting in brief discontinuity of the formant trajectory [9]. In contrast, the interdental-space effect (ISE) is the effect of side-branches of the vocal tract. The interdental space (between the upper and lower dental arches) forms bilateral side branches in the oral cavity when the tongue is in front, while it is included into the oral cavity when the tongue is in back. Thus, anti-resonance of the interdental space changes in frequency between front and back vowels, resulting in a traveling anti-resonance over a wide frequency range that crosses both F2 and F3 [6]. Both effects are thought to contribute a boundary to vowel-to-vowel transitions in agreement to the feature enhancement theory.

As both effects claim to cause F2 discontinuity in different ways during front-to-back vowel transitions, the primary question is how to distinguish the two in acoustic analysis. Whether the feature enhancement by F2 discontinuity can apply to Chinese utterances is another concern in this study. To answer the questions, this study aims at examining manifestations of the two effects on formant trajectories during front-to-back vowel transitions by analyzing speech and subglottal signals with examination of occurrence of formant discontinuity by the two effects.

2. Methods

Recordings of speech and subglottal resonance signals were conducted using a microphone (AT810E, AudioTechnica) and a vibration sensor (Hot Spot, K&K Sound). Our procedure for the recordings is similar to a previous study [10], as shown in Figure 1. The vibration sensor was adhered to a speaker's neck to measure resonances of the subglottal tract. Both microphone and vibration sensor signals were recorded through a USB audio interface (Scarlett 2i2, Focusrite) connected to a PC with 16-kHz sampling and 16-bit resolution.

2.1. Participants

Five native Chinese speakers of 23-24 years (3 male and 2 female subjects, labeled M1-M3 and F1-F2) participated in this experiment. No subjects reported any history of speech disorders.

http://dx.doi.org/10.21437/Interspeech.2016-1054

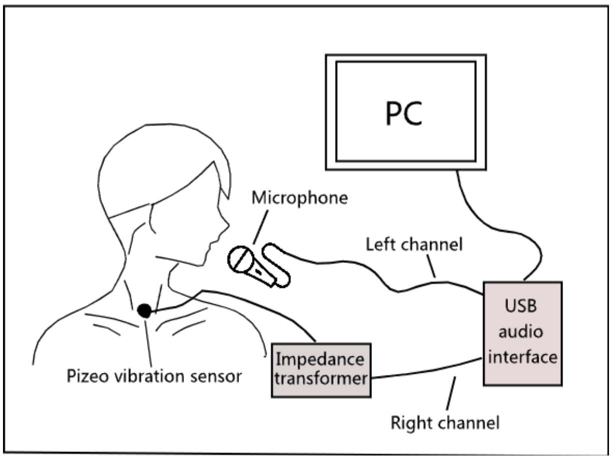

Figure 1: *Experimental setup for recording speakers' speech and subglottal resonance. The signal from the piezoelectric vibration sensor was conditioned by a transformer for impedance matching.*

2.2. Speech materials and data collection

Two triphthongs containing front-to-back vowel sequences were selected to examine the two effects on formants in transitions during articulation to contrast [± back] vowels. In Chinese, a vowel has many allophones in different phonetic environment, thus having varying [± back] feature values. In our study, the feature values of [± back] in Chinese triphthongs are derived from the vowel charts shown in previous studies [11] [12]. Chinese diphthongs were not used because they lack [± back] contrasts.

Subjects were asked to read sentences displayed on a PC monitor. Chinese two-character consonant-triphthong (CT) words were embedded in a carrier sentence "我读了___这个词" (*"I read the word ___" in English*). Each character in the words contains the same triphthong. That is, two triphthongs in each sentence are the same. There are two sentences used for speaker to read: one contains triphthong /iao/, and the other contains /iou/. For instance, the sentence containing the triphthong /iao/ is "我读了 '逍遥' 这个词", where /iao/ appears twice in the utterance "uo tu lɤ 'ɕiɑʊ iɑʊ' tʂɤ kɤ tsʰɿ" in IPA. Table 1 shows the information of our selected CT words used in the experiment. In our experiment, ten repetitions of each sentence were recorded from each speaker, and a total of 40 recordings were obtained for each of them. An informal listening test was performed to remove error samples from the total recordings. Table 2 shows the number of valid samples.

2.3. Acoustic analysis

Triphthong segments were collected from the valid samples. The second and third formants (F2, F3) with their corresponding amplitudes (A2, A3) were extracted from each segment semi-automatically by using the Imai's improved cepstral method [13]. The cepstral analysis were conducted by applying a time window of 32 ms long with a 16-ms shift length. Subglottal formants (SgF1, SgF2 and SgF3) were extracted from the subglottal resonance signals. The three subglottal formants for females are on average 599 Hz, 1581 Hz and 2438 Hz. For males, those averaged subglottal resonances are 505 Hz, 1359 Hz and 2211 Hz.

Table 1. *Chinese triphthongs, two-character CT words in pinyin and Chinese characters with English meaning. IPA symbols and features are also shown.*

Triphthong	iao	iou
CT words	xiāo yáo *(逍遥, " leisurely")*	yōu xiù *(优秀, "excellent")*
IPA [front/back]	/iɑʊ/ front/back/back	/iou/ front/back/back

Table 2. *Number of valid triphthongs from five subjects.*

Triphthong	M1	M2	M3	F1	F2
iao	15	15	10	10	10
iou	15	10	10	12	9

3. Results

Acoustic properties extracted from speech and subglottal signals were analyzed to examine manifestations of the two effects on formant trajectories during vowel transitions.

3.1. Occurrence of formant discontinuity

A brief attenuation of A2 or A3 exhibits an observable discontinuity of F2 or F3 on spectrograms, which suggests that damping of A2 and A3 is a measure for the occurrence of instability or a gap in the corresponding formant trajectories. Frequency of occurrence of F2 and F3 discontinuities was counted for each speaker with spectrographic observations and analyses of A2 and A3 damping. Table 3 shows the frequency of occurrence among the valid samples from all the subjects. It is seen that F2 and F3 discontinuities were not always discerned among the samples from the five speakers. Also, co-occurrence of F2 and F3 discontinuities was predominant in triphthong /iao/. In /iou/, F2 discontinuity including "F2-only" and "F2&F3" cases occurs more frequently than F3 discontinuity including "F3-only" and "F2&F3" cases. Thus, both individual variation and token-to-token variation were observed in the data.

3.2. Relationship between F2 gap frequency and SgF2

The relationship between F2 gap frequency and SgF2 was examined to explore which effect is causal to F2 discontinuity among all the samples exhibiting F2 discontinuities. F2 gap frequency was defined as the frequency where F2 discontinuity occurs with obvious amplitude (A2) attenuation. The frequencies of F2 gap and SgF2 were plotted in Figure 2 for two triphthongs separately. As seen in Figure 2(a), in triphthong /iao/, F2 gap frequencies are consistently higher than their corresponding SgF2s. In /iou/, in Figure 2(b), F2 gap frequencies are not always distant from SgF2s, often showing that both are close in frequency. Hence, not all the F2 discontinuities occur near the frequency of SgF2, though the SgF2 lies between the F2s of front and back vowels for both triphthongs.

3.3. Variation of F2 and A2 near SgF2

The results shown in section 3.2 reveal that when F2 discontinuity occurs, A2 diminishes but not always near SgF2. When F2 discontinuity is seen in the vicinity of SgF2, they exhibit attenuation in A2 near SgF2. Figure 3(a) and 3(b) show such a case exhibiting a rapid fall of F2 with A2 attenuation.

Table 3. *Occurrences of formant discontinuity for five speakers. "F2-only" and "F3-only" represent discontinuity occurring only in F2 or F3, respectively. "F2&F3" indicates co-occurrence of F2 and F3 discontinuity. "None" are the cases when F2 and F3 change continuously, and "Others" for ambiguous cases. The occurrence number is shown with the slash dividing two triphthongs.*

Speaker	Triphthong	Total	F2-only	F3-only	F2&F3	None	Others
M1	iao/iou	15/15	3/10	2/0	8/1	1/0	1/4
M2	iao/iou	15/10	4/2	3/0	5/6	0/1	3/1
M3	iao/iou	10/10	0/2	3/0	0/2	5/6	2/0
F1	iao/iou	10/12	2/5	1/0	4/5	0/1	3/1
F2	iao/iou	10/9	0/4	2/0	5/2	0/0	3/3
Total	iao/iou	60/56	9/23	11/0	22/16	6/8	12/9

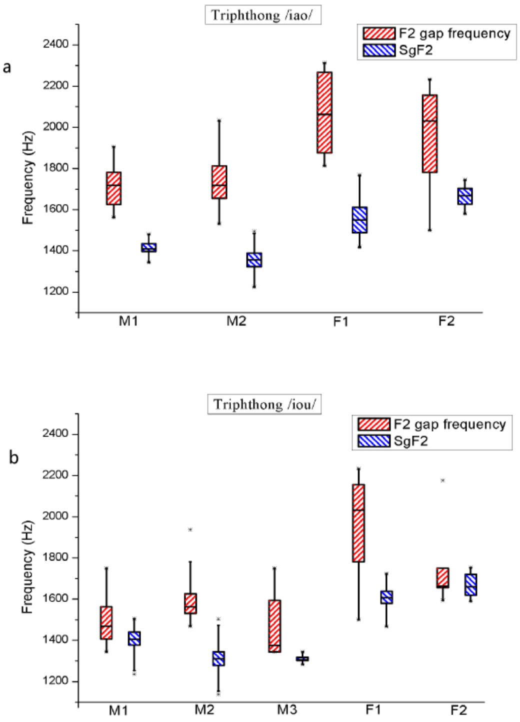

Figure 2: *Statistical summary of F2 gap frequency and SgF2. (a) Results of triphthong /iao/ for 4 speakers, excluding M2 who had no samples of F2, and (b) Results of triphthong /iou/ for 5 speakers.*

3.4. F2 and F3 discontinuities caused by ISE

The anti-resonance of ISE moves crossing F2 and F3 obliquely. In the case of back-to-front vowel transitions, F3 discontinuity is seen before F2 discontinuity. In contrast, during front-to-back vowel transitions, F2 discontinuity occurs first, as shown in Figure 4(a). Therefore, the time lag between F2 and F3 discontinuity is one of the indices of ISE in addition to co-occurrence of F2 and F3 discontinuities. Accordingly, F2 and F3 frequency gaps were examined for the temporal order of their appearance by measuring t2 and t3, which indicate when the corresponding discontinuities occur. It was found that F2 discontinuity occurs first with a time lag to F3 discontinuity among 26 samples. Figure 4(a) and 4(b) show an example of co-occurrence of F2 and F3 discontinuities. Spectrogram in Figure 4(a) and amplitude tracks in Figure 4(b) both show a time lag between occurrences of F2 and F3 discontinuity.

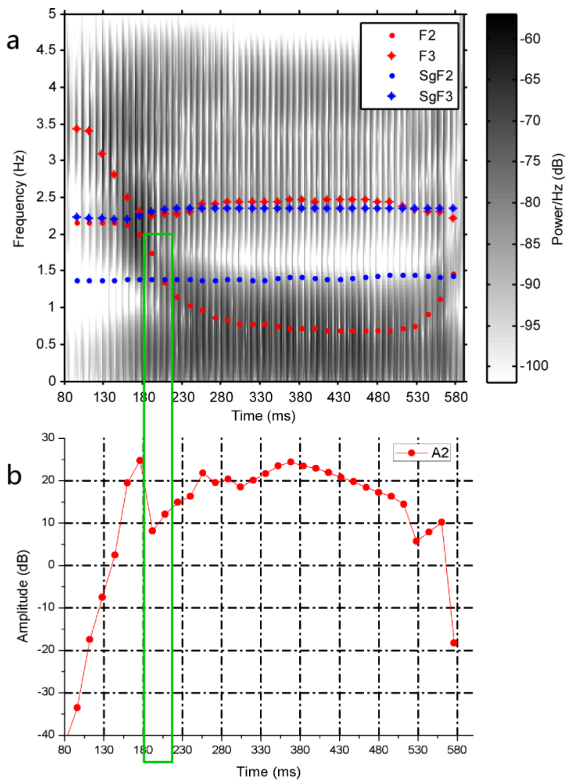

Figure 3: *Spectrogram and formant amplitude curve in /iou/ from subject M1 with marking on F2 discontinuity (green rectangle). (a) Trajectories of formants (F2 and F3) and subglottal resonances (SgF2 and SgF3) on a spectrogram, and (b) A2 amplitude curve showing a rapid fall of -16.6 dB at the discontinuity.*

4. Discussion

The effects of the subglottal tract and interdental space on formants during front-to-back vowel transitions were explored for Chinese triphthongs recorded from five speakers. In this study, it was observed from spectrograms that F2 discontinuity due to SCE occurs near SgF2 but not all the time. Discontinuities of F2 and F3 caused by ISE often occur with a time lag. The discontinuity of formants is not constantly observed among the subjects and tokens. When they are seen,

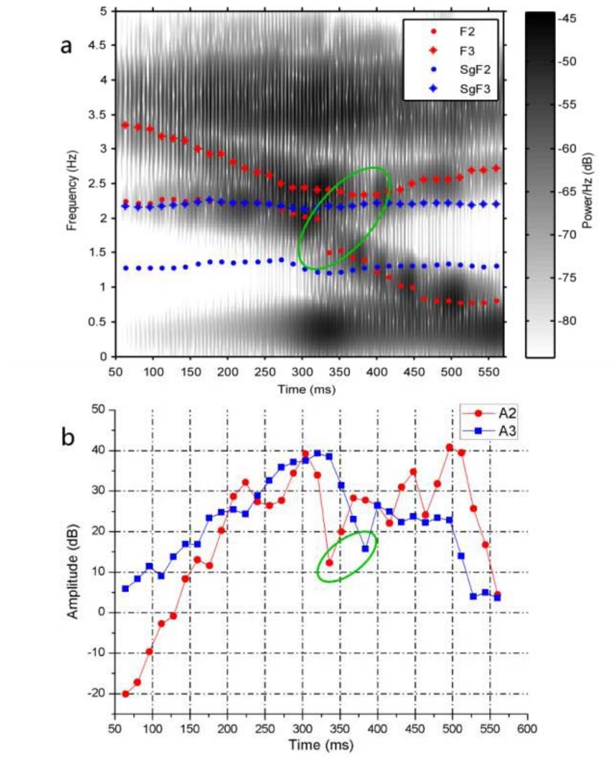

Figure 4: *Spectrogram and formant amplitude curves in /iou/ from subject F2. (a) Trajectories of F2, F3, SgF2, SgF3, A2 and A3 on a spectrogram, and (b) A2 and A3 amplitude tracks showing the time lag between F2 and F3 gaps.*

the effect on formant discontinuities due to ISE is more prominent than that of SCE.

The relationship between F2 gap frequency and SgF2 suggests that, when obvious F2 discontinuity occurs in the vicinity of SgF2 (Figure 2(b)), events of A2 attenuation and F2 frequency fall take place together. This observation supports SCE because SgF2 modulates vocal tract transfer functions due to the coupling between the vocal tract and subglottal airway. In this case, the constant acoustic attribute SgF2 tends to help enhance distinctive contrast between the defining acoustic attributes F2s of front and back vowels. However, discontinuity of F2 is also found to occur in the vicinity distant from SgF2 (Figure 2(a)). As the stationary anti-resonance is thought to influence a narrow frequency band in vowel spectra, formants would not be affected by the zero when they are away from the narrow band. This suggests that other factors may also cause F2 discontinuity.

Many spectrographic samples in this study show a bright frequency band that crosses F2 and F3 obliquely. This is a traveling anti-resonance caused by one or two side-branches in the vocal tract that changes in length during vowel articulation. The most likely structure of the causal branches in the vocal tract is the interdental space between the upper and lower dental arches [6]. The interdental space takes a form of bilateral side-branches in /i/, while it is included into the oral cavity in /a/. Between the two vowels, the branch length varies, thus producing a traveling anti-resonance. Figure 5 shows a typical example of the moving anti-resonance that forms obliquely narrow brighter frequency bands crossing formants at the edges

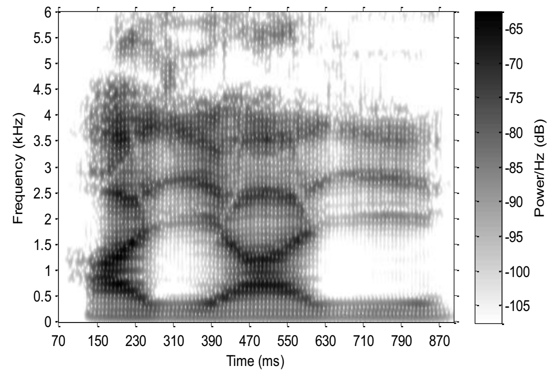

Figure 5: *Spectrogram showing the traveling anti-resonance during /aiai/ in a male speaker. Anti-resonance (narrow brighter band) falls from /a/ to /i/ crossing F3 and F2 and rises again in /a/ crossing F2 first.*

of adjacent contrast vowels with time lags. As a consequence, F2 and F3 discontinuities are observed in succession with a short time lag (Figure 4(a)). In addition, the moving anti-resonance would have an influence on vowel spectra within a wide frequency range, which may account for the cases where that F2 discontinuity occurs far away from the frequency of SgF2. Those observations support formant discontinuity due to ISE that exhibits visible boundaries on F2 and F3 between front and back vowels, thus providing another evidence to support feature enhancement in our data.

In our data, observations supporting ISE show differences across tokens and speakers. One possible explanation for the variation may be given to the side-branches of a complex shape formed by speakers' dental arches. Whether F3 discontinuity is a part of ISE or the SCE is question for another study.

In summary, formant trajectories are not only influenced by SCE, but also disturbed by ISE during the Chinese triphthongs with front-to-back vowel sequences. F2-only and F2&F3 discontinuity were typical acoustic manifestations of SCE and ISE, respectively. Comparing the two acoustic perturbation effects supporting the feature enhancement theory, formant instability caused by ISE was the major finding in this study. Causal mechanisms for the both remain as issues for further studies.

5. Acknowledgements

This work was supported by National NSF Program of China (No. 61573254) and National Basic Research Program of China (No. 2013CB329301). We would like to thank the subjects who participated in this study for help with the data collection.

6. References

[1] K. N. Stevens, "On the quantal nature of speech," *Journal of Phonetics*, pp. 3-45, 1989.

[2] S. J. Keyser, and K. N. Stevens, "Enhancement and Overlap in the Speech Chain," *Language*, vol.82, no.1, pp. 33-63, 2006.

[3] K. N. Stevens, "Toward a model for lexical access based on acoustic landmarks and distinctive features," *Journal of the Acoustical Society of America*, vol.111, no.4, pp. 1872-1891, 2002.

[4] X. Chi, and M. A. Sonderegger, "Subglottal coupling and its influence on vowel formants," *Journal of the Acoustical Society of America*, vol.122, no.3, pp. 1735-1745, 2007.

[5] S. M. Lulich, “Subglottal resonances and distinctive features,” *Journal of Phonetics*, vol.38, no.1, pp20-32, 2010.
[6] K. Honda, S. Takano, and H. Takemoto, “Effects of side cavities and tongue stabilization: Possible extensions of the quantal theory,” *Journal of Phonetics*, vol.38, no.1, pp. 33-43, 2010.
[7] M. A. Sonderegger, “Subglottal coupling and vowel space: an investigation in quantal theory,” B.S. dissertation, Massachusetts Institute of Technology, 2004.
[8] Y. Jung, “Acoustic articulatory evidence for quantal vowel categories: the features [low] and [back],” Ph.D, dissertation, Massachusetts Institute of Technology, 2010.
[9] K. N. Stevens, and S. J. Keyser, “Quantal theory, enhancement and overlap,” *Journal of Phonetics*, vol.38, no.1, pp. 10-19, 2010.
[10] S. M. Lulich, and J. R. Morton, “Subglottal resonances of adult male and female native speakers of American English,” *Journal of the Acoustical Society of America*, vol.132, no.4, pp. 2592-2602, 2012.
[11] L. Cai, and J. Kong, *Modern Chinese phonetic dictionary*, Beijing: Tsinghua University Press, 2014, pp. 686-687
[12] L. Lin, and L. Wang, *A Course in Phonetics*. Beiijng: Peking University Press, 2013, pp. 233-239
[13] S. Imai, and Y. Abe, “Spectral envelope extraction by improved cepstral method,” *Electro. And Commun. in Japan*, vol.62-A, no.4, pp. 441-444, 1978.