VOWEL LENGTH PERCEPTION IN CANTONESE

Ling Zhang

Department of Chinese and Bilingual Studies, The Hong Kong Polytechnic University, Hong Kong
zhanglingcbs@gmail.com

ABSTRACT

This paper reports the acoustic experiment and perception test on a pair of Cantonese long and short vowels /aː/ and /ɐ/ in the diphthong context.

The acoustic data show that [aːi] and [ɐi] have similar targets of vowel nucleus and coda, parallel transition glide, and comparable total duration. They are different in their durational structure: [aːi] has a long steady phrase of vowel nucleus while its coda is short. In contrast, the vowel nucleus of [ɐi] is short but its coda lasts long. The situation of [aːu] and [ɐu] is also similar.

In the perception test, the unidirectional gradual cutting method is adopted to change the duration ratio of vowel nucleus and coda. Progressive cutting [aːi]/[aːu] can make them sound like [ɐi]/[ɐu], while regressive cutting [ɐi]/[ɐu] can turn them into [aːi]/[aːu], which prove that these pairs of diphthongs are different in durational structure rather than vowel quality. Furthermore, borrowing the idea of the least square method, our re-analysis of the stimuli reveals that the length of vowel nucleus plays a more important role than the quality of transition glide and coda target in perception.

Keywords: Cantonese, diphthong, vowel length, perception

1. INTRODUCTION

It has been controversial whether there are real long and short vowels in Cantonese. The most debatable problem is whether vowel length (quantity) or vowel height (quality) is the key factor to differentiate these pairs of vowels. Actually most pairs of so-called long and short vowels in Cantonese are in complementary distribution and do not happen in the same context, except the pair of /aː/ and /ɐ/. As vowel nuclei, /aː/ and /ɐ/ are distinct in diphthongs, nasal rimes and occluded rimes. Since diphthongs have clear and continuous formant trajectories in spectrograms, our present study will carry out acoustic experiment and perception test that focus on diphthongs with /aː/ and /ɐ/. Through observing the trajectories of the first two formants – F_1 and F_2 across the time axis, we can infer the articulatory movement of the diphthongs. If the F_1 and F_2 values have great overlap for the targets of /aː/ and /ɐ/ as well as their coda targets in the diphthongs, we should consider whether they are differentiated by durational structure.

According to [3], two diphthongs can have similar targets and comparable total duration but vary in their auditory quality if they have unequal durations of the two targets and the glide relative to the total length of the diphthong. They also pointed out that if vowel length functions distinctively, it is the relative duration rather than the absolute duration that matters.

As far as the durational structure of Cantonese diphthongs are concerned, previous studies indicated that there is a kind of complementary inter-play between the vowel nucleus and coda. Chao [1] put forth that a coda is strongly or weakly articulated according as the vowel nucleus is short or long. Hashimoto [4] made it clearer that when the vowel nucleus is long, the coda is comparatively short, and when the vowel nucleus is short, the coda is comparatively long. Cheung [2] further elucidated this length complementarity, revealed the relationship between this mechanism and the prosodic feature of syllable isochrony, and provided a moraic analysis of the durational structure of the rimes in Cantonese. This feature of complementary lengthening is useful when we designed experiment on vowel length perception.

For the convenience of transcribing sounds in the present study, here we adopt the LSHK Cantonese Romanization [5]. The correspondence between IPA and this romanization system is listed below: [aː] = ‘aa’; [aːi] = ‘aai’; [ɐi] = ‘ai’; [ei] = ‘ei’; [i] = ‘i’; [aːu] = ‘aau’; [ɐu] = ‘au’; [ou] = ‘ou’; [u] = ‘u’, [k] = ‘g’.

2. EXPERIMENTAL METHOD

In the acoustic experiment, four male and four female native speakers of Cantonese aged 20 to 35 were asked to pronounce two groups of minimal pairs which contain the targeted diphthongs: one group with the null onset while the other group with the ‘g’ onset.

The parameters of duration, F_1 and F_2 were measured by the Praat software [6]. As introduced in Sec. 1, for the durational structure of diphthongs, the relative duration is more important than the absolute duration. Therefore, the time-normalized method was adopted in the present study: the total duration of the diphthong was measured and divided into 20 equal parts. Then the F_1 and F_2 values at every 5% duration point were obtained. Finally these 21 points were connected to get the F_1 and F_2 trajectories of the diphthong.

In the perception test, the null-onset syllables pronounced by one of the female informants were used as the original materials for synthesizing stimuli. These original syllables are about 600 ms long. According to the durational features of different types of diphthongs, different directions of gradual cutting method were applied to synthesizing stimuli.

For the diphthongs with long vowel nucleus (‘aai’ and ‘aau’), we used the progressive cutting method to shorten the vowel nucleus gradually. The cutting direction is from front to back, and every time we cut 20 ms off. The remaining part of the segment is a stimulus as well as the basis for the next stimulus. As the vowel nucleus is shortened and the ratio of the coda is increased gradually, it would be interesting to see whether the stimuli are heard as a diphthong with short vowel nucleus at certain cutting phase.

For the diphthongs with short vowel nucleus (‘ai’ and ‘au’), the regressive cutting method was adopted. The cutting procedure starts from the end of the coda. We cut off 20 ms backwards every time. This method can shorten the duration of coda thus raise the proportion of vowel nucleus over the whole diphthong. We can also observe whether the stimuli will sound like a diphthong with long vowel nucleus as the cutting procedure goes on.

The original order of the stimuli was shuffled. Between the stimuli, there was a break of 5 seconds for listeners to judge what they have heard and choose the corresponding choice on the questionnaire. There were 10 native speakers of Cantonese participating in this perception test.

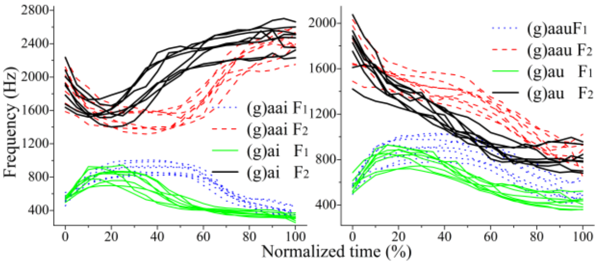
3. RESULTS AND DISCUSSION

3.1. Acoustic data of the diphthongs

The duration data of both ‘g’ and null onset groups show that the long or short vowel nucleus do not influence the overall duration of a diphthong. The difference within a minimal pair is less than 100 ms for a same speaker.

For the null-onset group, five speakers could control their syllables to be null onset while the other three pronounced the onset as another free variant [ŋ]. The comparison between the patterns of ‘g’ group and null onset group reveals that they are very similar except the onset-diphthong transition. Here we only display the contours of the ‘g’ group by all the speakers in Figure 1.

Figure 1: Time-normalized F_1 and F_2 trajectories of ‘(g)aai’ / ‘(g)ai’ and ‘(g)aau’ / ‘(g)au’.



In both graphs of Figure 1, about 10% of the beginning is the onset-diphthong transition. After this transition, the vowel nuclei start at about 10%~20%. It can be observed that the targets of vowel nuclei have great overlap between ‘aai’ and ‘ai’, ‘aau’ and ‘au’. The curves of ‘aai’ and ‘ai’ start to be apart from each other at about 20% of the time axis and converge again at about 80%. The coda targets of them are also quite similar. The factor that causes the separation of ‘aai’ and ‘ai’ is that the length of their nuclei is different. The vowel nucleus of ‘aai’ has a long steady phase while the vowel nucleus of ‘ai’ only lasts for a short period. Consequently, the transition glide from the vowel nucleus to the coda (VTC, henceforth) appears quite late for ‘aai’ while very early for ‘ai’. After the VTC ends, the coda only continues a very short time for ‘aai’ but maintains a long phase for ‘ai’.

As the targets of vowel nucleus and coda are the same, and the VTC curves are nearly parallel for ‘aai’ and ‘ai’, it can be concluded that their distinction is mainly caused by different proportions of their vowel nucleus and coda. Similar conclusions can be drawn from the acoustic data of ‘aau’ and ‘au’.

3.2. Perception results

Figure 2 and Figure 3 display the results of the perception test. The dashed lines denote that a certain sound (marked above the F_2 trajectory) reaches its perception peak at that time percentage. In other words, when cutting at that percentage of the original diphthong, most listeners (the percentage of them is shown in the parentheses) judge the remaining segment as a certain sound. It should be noticed that the remaining segment refers to the part on the right of the dashed line for progressive cutting, while on the left for regressive cutting.

Figure 2: The perception results of progressive cutting of ‘aai’ and ‘aau’.

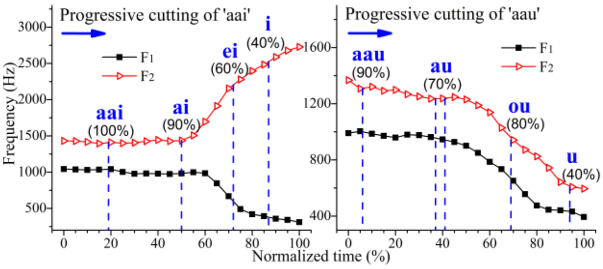
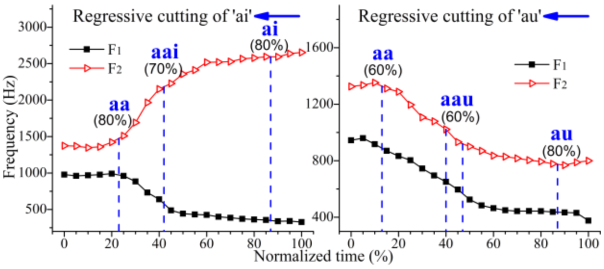


Figure 3: The perception results of regressive cutting of ‘ai’ and ‘au’.



From Figure 2 and Figure 3, we can observe that for ‘aai’ and ‘aau’, as the progressive cutting procedure goes on, the stimuli are heard as ‘aai’/ ‘aau’→ ‘ai’/ ‘au’ → ‘ei’/ ‘ou’ → ‘i’/ ‘u’. For ‘ai’ and ‘au’, as we proceed the regressive cutting, the stimuli are heard as ‘ai’/ ‘au’ → ‘aai’/ ‘aau’ → ‘aa’. As the unidirectional cutting method can swap ‘aai’ and ‘ai’, ‘aau’ and ‘au’, while this cutting method only changes the duration ratio of vowel nucleus and coda, it can be proved that ‘aai’ and ‘ai’, ‘aau’ and ‘au’ are different because of their durational structures rather than their vowel nuclear target quality. Moreover, as there is not a monophthong [ɐ] in Cantonese, the beginning part of [ɐi] and [ɐu] is judged as [aː] rather than other vowels, which also indicates that the quality of [ɐ] and [aː] is very similar.

3.3. Re-analysis of the stimuli

It can be hypothesized that the perception peaks are achieved when the stimulus and the targeted sound are most similar in Figure 2 and Figure 3. Does the similarity indicate that the overall F_1 and F_2 patterns are approximate, or that some phases are more important than other parts? To answer this question, we firstly re-normalize the time of the stimuli; then carry out theoretical calculations by borrowing the idea of the least square method from mathematics to find the theoretically optimal stimuli; and finally compare the theoretical calculation results with the perception results.

The least square method is conducted as below. Let the data of the targeted sound be represented by $\{t_i,y_i\}$, where t_i is the normalized time and y_i denotes for the value of F_1 or F_2. For the stimuli, the interpolated data after time renormalization is $\{t_i, f_i\}$, where f_i is the value of F_1 or F_2 at time t_i. To find the best fitted stimulus to the curve of the targeted sound, we define a parameter S as (1). Best fitting of one curve (either F_1 or F_2) is achieved when S is minimized.

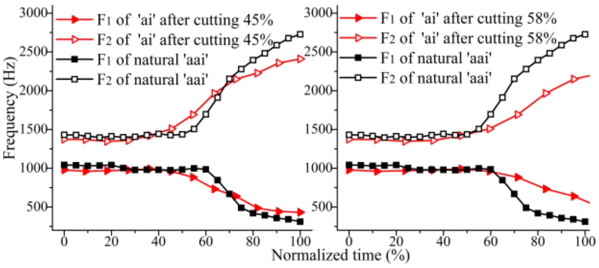
$$S = \sum_i \left(f_i - y_i\right)^2 \qquad (1)$$

For progressive cutting, the theoretical calculation agrees well with the perception results. After cutting 50% off the original ‘aai’, the remaining segment has the smallest S value and therefore is the most approximate to the natural ‘ai’. Meanwhile, it can be observed from Figure 2 that the perception peak of ‘ai’ also happens at 50% in the perception test. For progressive cutting of ‘aau’, the best fitted stimulus to ‘au’ and the actual perception peak of ‘au’ both occur at 40%.

For regressive cutting, the situation is more complicated and interesting. There is notable gap between our theoretical calculation and the perception test results. According to our theoretical calculation, if 45% of ‘ai’ is cut off backwards, the remaining segment mostly resembles the natural ‘aai’. However, as shown in Figure 3, the perception peak of ‘aai’ turns up when cutting 58% backwards off ‘ai’. The regressive cutting of ‘au’ also has similar phenomenon: the theoretical optimal stimulus is at 40%, while the actual perception peak of ‘aau’ is at 53%.

To more directly reveal the inconsistency of the theoretical calculation and the perception test of regressive cutting, we compare the natural ‘aai’ with ‘ai’ after regressive cutting 45% (theoretically optimal) and 58% (perceptually optimal) respectively in Figure 4. It is obvious that the overall curves of the 45% stimulus looks much more similar to the natural ‘aai’, while there is great gap between the 58% stimulus and the natural ‘aai’ on the right half, including the VTC and the coda target. However, in terms of the length of the vowel nucleus and the transition point between the vowel nucleus and the VTC, the 58% stimulus is more approximate to the natural ‘aai’ than the 45% stimulus. It can be inferred that different parts of a diphthong contribute unequally in perception. Vowel nucleus plays a more important role than VTC and coda target. From another point of view, the quantity feature – the duration ratio between vowel nucleus and the residual part is more important than the quality feature of VTC and coda target – they only need to be roughly realized.

Figure 4: The comparison between the targeted sound ‘aai ’ and the stimuli (‘ai’ after cutting 45% and 58%).



The mismatch between the theoretical and the actual point of perception is related to our cutting method. The coda phase in ‘ai’ actually has slight gradient from the VTC to the coda target. Cutting backwards changes the target of the coda. The more is cut off, the greater is deviated from the original coda target. It is generally observed that the /i/ in a diphthong is much lower than the prime target for a monophthong /i/. Although the coda targets of the 45% and the 58% stimuli are even lower, they are still judged as an /i/ coda in a diphthong. Thus, the vowel nucleus is more salient to distinct ‘aai’ and ‘ai’. It can be regarded that the 45% stimulus approximates the overall quality of the diphthong while the 58% stimulus resembles the durational structure. The inconsistency of the theoretical calculation and the perception test again proves that the durational structure is more important to distinguish ‘aai’ and ‘ai’.

Stevens [7] and [8] put forward the quantal theory, which can help us to understand the relationship of the acoustic parameters, the articulatory positions, and the auditory responses. He pointed out that the acoustic-articulatory and acoustic-auditory relations are not linear. Rather, there are plateau regions and abrupt change regions. The quality difference of VTC and coda target has little influence on the perception here because /i/ as a diphthong coda occupies the plateau of relative stability. In contrast, the durational ratio between the vowel nucleus and the residual part is in the abrupt change region of the auditory space thus listeners are more sensitive to the change of it.

4. CONCLUSIONS

The results of the acoustic experiment and perception test in our present study are consistent, which prove that /aː/ and /ɐ/ are distinct in vowel length rather than in vowel quality. Durational structure difference is the key factor to differentiate ‘aai’ and ‘ai’ as well as ‘aau’ and ‘au’.

Borrowing the idea of the least square method, we re-analyze the stimuli in the perception test and find the theoretically optimal stimuli. The theoretical calculation agrees well with the perception results for progressive cutting while has evident gap for regressive cutting. It can be inferred from the re-analysis that the length of vowel nucleus is more salient than the quality of VTC and coda target in perception. Further investigation should be conducted on this phenomenon.

5. ACKNOWLEDGEMENTS

I would like to thank Prof. Kwan-hin Cheung, Prof. John J. Ohala and Prof. Chusheng Zhuang for their valuable suggestions.

6. REFERENCES

[1] Chao, Y.R. 1947. *Cantonese Primer*. New York: Greenwood Press.

[2] Cheung, K.H. 1986. *The Phonology of Present-day Cantonese*. Doctoral dissertation. London: University College London.

[3] Clark, J., Yallop, C. 1990. *An Introduction to Phonetics and Phonology*. Oxford: Basil Blackwell.

[4] Hashimoto, A.O.Y. 1972. *Phonology of Cantonese*. Cambridge: Cambridge University Press.

[5] LSHK Cantonese Romanization. *http://www.lshk.org/cantonese.php*

[6] Praat (Version 5.1.05). Retrieved May 1, 2010, from *http://www.praat.org/*

[7] Stevens, K.N. 1972. The quantal nature of speech: Evidence from articulatory-acoustic data. In Denes, P.B., David, E.E. (eds.), *Human Communication: A Unified View*. NewYork: Mc Graw Hill, 51-66

[8] Stevens, K.N. 1989. On the quantal nature of speech. *Journal of Phonetics* 17, 3-46.