



Effect of language experience on the categorical perception of Cantonese vowel duration

Caicai Zhang, Gang Peng, William S-Y. Wang

Language Engineering Laboratory, The Chinese University of Hong Kong, Hong Kong

yzcelia@gmail.com, gpeng@ee.cuhk.edu.hk, wsywang@ee.cuhk.edu.hk

Abstract

This study investigated the effect of language experience on the categorical perception of Cantonese vowel duration distinction. By comparing Cantonese and Mandarin listeners' performances, we found that: (1) duration change elicited categorical perception in the performance of Cantonese listeners, but not in Mandarin listeners; (2) Cantonese listeners were affected by the vowel quality differences, whereas Mandarin subjects were generally unbiased towards the quality differences; (3) effect of duration was overridden by the vowel quality [a] condition in the performance of Cantonese listeners. Our findings suggested that vowel quality is incorporated as a phonological cue in Cantonese.

Index Terms: duration, vowel quality, categorical perception, Cantonese

1. Introduction

Duration is used to contrast lexical meanings in many languages in the world (e.g., Finnish, [1]). In Cantonese, duration was reported to differentiate two phonemes [a] and [ɐ] in closed syllables (such as [ai]-[ɛi], [aŋ]-[ɛŋ]) (cf. [2]). However, it has long been debated whether the [a]-[ɐ] distinction in Cantonese is encoded merely by duration difference, or by both duration and vowel quality differences.

This study aims to investigate this question from the perceptual perspective. Specifically, we aim to test whether vowel quality difference is an intrinsic difference associated with duration, or whether it is incorporated as a linguistic cue (together with duration) in the phonetic representation of [a] and [ɐ]. In other words, we wish to examine whether a change in vowel quality has a perceptual consequence. If vowel quality difference is intrinsic, it presumably has no effect on the perception (the listeners may not even discern the change in quality). But if quality is a linguistic cue, it is likely that a change in quality changes the perception.

Mixed results about the role of duration and quality differences were obtained in the literature. In terms of production, So and Wang [3] showed that Cantonese long-short vowels were significantly different in both duration and vowel quality, with short vowel [ɐ] being significantly shorter than [a] and also significantly lower in the acoustic F1-F2 plane. Zee's acoustic study [2] showed that the temporal organization of nucleus and glide distinguishes the long-short diphthongs, with the nucleus being longer in [ai] and [au], but the glide being longer in [ɛi] and [ɛu]. Moreover, Zee [4] found that [ɐ] and [a] also differed in vowel quality.

In terms of the perceptual studies, Lee [5] found that when the duration of nucleus and glide was manipulated to vary simultaneously in a continuum from 20 ms to 200 ms (20 ms per step), it is only the duration of nucleus that changed the perception in a categorical manner.

Lee [6] reported that both duration and vowel quality are effective cues for identifying Cantonese vowels, but this study targeted the perception of a non-contrastive long-short vowel

pair [ɛ:]-[ɛ] ([ɛ:] and [ɛ] are in complementary distribution, e.g., [sɛ:]-[sɛk]).

Shi and Liu [7] found that vowel quality indeed affected the identification of [a] and [ɐ] in some conditions. When the duration of [ɐ] was lengthened to that of [a] via duplication of the steady state portion of [ɐ], the quality of [ɐ] suppressed the effect of duration so that the lengthened [ɐ] failed to be perceived as [a] (identification rate of [a] < 50%). But in the shortened [a] condition, more than 80% of the stimuli were identified as [ɐ], consistent with the prediction from the duration perspective. In addition, this study reported a trading relation between duration and vowel quality in production, i.e. the smaller the duration difference between [a] and [ɐ], the greater the quality difference. Accordingly, greater quality difference tended to limit the effect of duration in perception.

According to Lehiste [8], there is an intrinsic difference in vowel duration associated with the degree of opening: other factors being equal, a high vowel is shorter than a low vowel. Such duration difference seems to be physiologically conditioned and thus constitutes a phonetic universal. Given the intrinsic connection between vowel duration and openness, it seems reasonable to expect the short vowel [ɐ] to be less open than [a] in Cantonese (which is exactly the case according to [3]). Therefore, it is likely that the vowel quality change in [a]-[ɐ] vowel pair originated from an intrinsic difference associated with vowel length.

Although Shi and Liu [7] provided some evidence that Cantonese listeners noticed the vowel quality difference in their perception, it is unclear whether this quality difference remains intrinsic or it has been used as a linguistic cue. To explore this question, we recruited Mandarin speakers as controls in this study. If a similar pattern is elicited in the performances of both Cantonese and Mandarin listeners, it lends some support to the hypothesis that this vowel quality difference has a universal psychophysical basis. If different patterns are found, it supports the hypothesis that this vowel quality difference is language-specific.

2. Methodology

2.1. Subjects

Nine Cantonese speakers (6 M, 3 F; mean age=25.9 yr, s.d.=1.8) and five Mandarin speakers (2 M, 3 F; mean age=28.3 yr, s.d.=5.6) participated in this experiment. On average Mandarin subjects have stayed in Hong Kong for 2.1 years (s.d.=1.5), meaning that they have had some exposure to Cantonese. No subject reported hearing difficulties or speech disorders.

2.2. Stimuli

Production of four pairs of meaningful words were elicited from a male Cantonese speaker, i.e. 嗌 /ai3/ 'to call' - 縊 /ɛi3/ 'to strangle', 歎 /ai2/ 'sighs' - 矮 /ɛi2/ 'short', 坳 /au3/ 'cavity' - 滷 /ɛu3/ 'to soak', 拗 /au2/ 'to bend' - 嘔 /ɛu2/ 'to vomit', each word being repeated nine times. Words bearing Tone 2

(T2) and T3 were included in the word list because of the available lexical contrasts of [a] and [ɛ] in these two tones.

Duration, F1 and F2 were measured from the nuclei ([a] and [ɛ]) and following glides (/i/ and /u/) of each diphthong. F1 and F2 were measured from 21 sampling points, 0%, 5%, 10% ... 95%, 100% of the nucleus and the glide respectively. Table 1 depicts the mean duration of nuclei and glides, and average F1 and F2 values of nuclei. Paired samples t-tests comparing [a] and [ɛ] found significant differences between these two vowels in both duration and vowel quality in all conditions, irrespective of the glide type (/i/ or /u/) and tone category (T2 or T3) ($p < 0.001$ in all cases).

Table 1. Duration, F1 and F2 of diphthongs [ai]-[ɛi] and [au]-[ɛu]. Numbers in brackets refer to s.d..

Diphthong	Duration (ms)		F1 (Hz)	F2 (Hz)
	Nucleus	Glide		
[ɛi]	125.760 (22.109)	337.576 (38.937)	846.166 (21.048)	1287.911 (58.203)
[ɛu]	173.302 (30.805)	424.466 (149.382)	791.378 (107.011)	1109.446 (145.791)
<i>Mean</i>	149.531	381.021	818.772	1198.679
[ai]	321.296 (51.936)	206.038 (26.310)	875.385 (24.153)	1262.286 (38.156)
[au]	260.405 (29.903)	235.977 (32.512)	934.794 (34.912)	1332.512 (50.913)
<i>Mean</i>	290.851	221.008	905.090	1297.399

F1 and F2 values measured at the midpoint of the nucleus (Figure 1) show that [ɛ] tends to be more centralized than [a] on the F1-F2 plane. However, the degree of centralization seems to differ slightly depending on the following glide. For example, [ai] and [ɛi] are largely overlapping, whereas [ɛu] obviously has a more central distribution than [au]. Results of acoustic analysis in this study were largely consistent with [3].

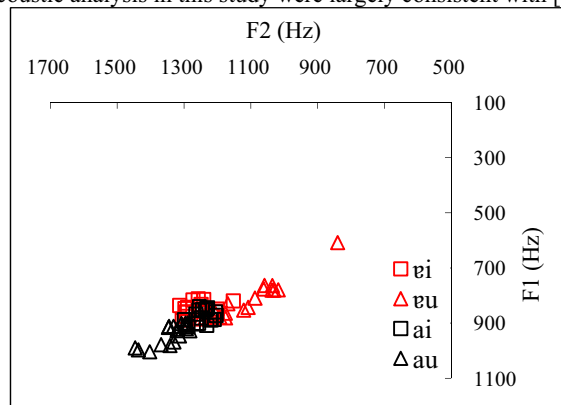


Figure 1: F1 and F2 values of /ai/, /ɛi/, /au/ and /ɛu/ at midpoint of the nucleus.

One clear token of each word was selected from the nine repetitions for manipulation. Duration of the nuclei was manipulated in two directions, lengthening and shortening. In the former condition, the duration of [ɛ] was lengthened to be as long as that of [a] by copying and pasting the steady state portion of [ɛ] period by period. In the latter condition, the duration of [a] was shortened to be similar to that of [ɛ] by removing one period from every two consecutive periods. Care was taken in the manipulation to avoid creating discontinuities. In this way, vowel quality of [a] is preserved in the shortening condition and quality of [ɛ] was preserved in the lengthening condition. No F0 smoothing was applied after the manipulation, as the stimuli were judged to preserve its

naturalness and original tone identity by both the experimenter and the male speaker who produced these tokens.

For both directions, the duration of nucleus was manipulated to vary in a continuum from 140 ms to 290 ms at 11 steps (around 15 ms per step; the step size differs slightly within the range of 13 ms and 17 ms because the manipulation was based on periods). Duration of the following glide was kept constant at 220 ms. A continuum of 11 stimulus was manipulated for each syllable (/ai3/, /ɛi3/, /ai2/, /ɛi2/, /au3/, /ɛu3/, /au2/, /au2/), giving rise to 88 stimuli in total.

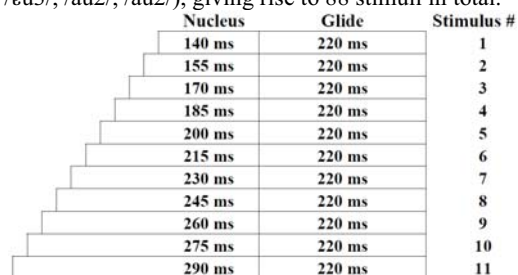


Figure 2. Schematic representation of stimuli duration.

2.3. Task

In the identification test, eight continuums of stimuli were presented in separated sessions. Because it is impossible to ask Mandarin subjects to identify Cantonese words, to ensure the same instruction for both groups of subjects, we drew two stimuli from the ends of each continuum - stimulus #1 and #11, and set them as the reference sounds. Subjects were required to memorize these two references before each session. Within a session, 11 stimuli were presented in random order and repeated two times. Subjects had to judge whether the heard stimulus is similar to reference 1 or 2 by pressing buttons on a keyboard. The identification test was self-paced: the next sound was played only after the subject had made a choice.

In the following AX discrimination test, two stimuli were presented in pairs with an inter-stimulus interval of 500 ms and the subjects were asked to judge whether these two sounds were identical or different. Only different sound pairs, 1-3, 2-4, 3-5, 4-6, 5-7, 6-8, 7-9, 8-10, 9-11, were included to control the experiment length. Each pair was repeated twice. The discrimination test was also self-paced by the subjects.

3. Results

For quality control purpose, only those subjects who correctly identified two references - stimulus #1 and #11 for more than 50% of the cases were included in the analysis. Two (1 M, 1 F) Cantonese subjects, and one female Mandarin subject were excluded given this criteria, leaving seven Cantonese subjects (5 M, 2 F) and four Mandarin subjects (2M, 2F) in the analysis.

3.1. Identification rates

A four-way repeated measures ANOVA was conducted on the identification rates of [ɛ] (the percentage that a stimulus was identified as [ɛ]), by indicating *language group* as a between-groups factor (Cantonese and Mandarin), and *vowel quality* ([a] and [ɛ]), *glide type* ([i] and [u]) and *tone category* (T2 and T3) as three within-subjects factors. As no significant main effects of *glide* and *tone* were found ($p > 0.1$), these two factors were disregarded in the following analyses. A three-way repeated measures ANOVA (*language group* × *vowel quality* × *stimulus no.*) was conducted on the same set of results.

The three-way ANOVA found a significant main effect of *language group* ($F(1, 42) = 5.361, p < 0.05$), suggesting an effect

of language experience on modulating listeners' perception. Moreover, there was a significant interaction of *vowel quality* by *language group* ($F(1, 42)=6.481, p<0.05$). It means that the vowel quality differences affected the performance of two language groups differently.

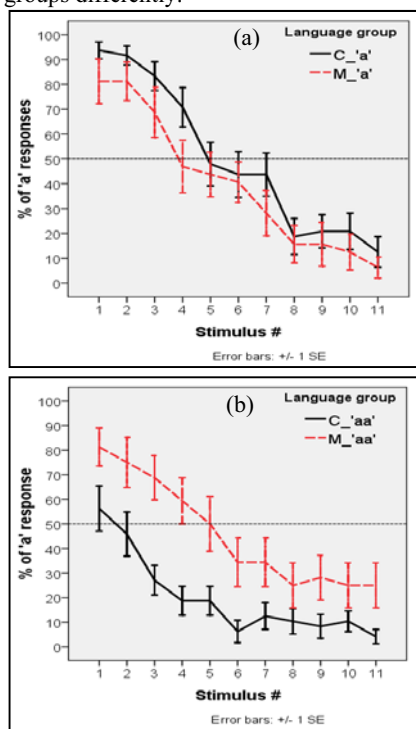


Figure 3. Mean identification rates of [v] plotted as a function of stimulus no. 'aa' and 'a' correspond to [a] and [v] respectively in the text. C=Cantonese; M=Mandarin. (a): vowel quality [v] condition (i.e. lengthening [v] to [a]); (b): the vowel quality [a] condition (i.e. shortening [a] to [v]).

The interaction of *quality* by *language* can be further explained by the different response patterns in the [v] and [a] quality conditions (Figure 3(a) and (b) respectively). As shown in Figure 3(a), there were overall more [v] responses in the performance of Cantonese listeners than that of Mandarin listeners. Moreover, the categorical boundary (50%) differed between these two groups of subjects. For Cantonese listeners, the categorical boundary was located closer to the longer end (stimulus #11) compared to Mandarin listeners. However, in the [a] quality condition (Figure 3(b)), it was Mandarin listeners, instead of Cantonese listeners who made more [v] responses. A closer examination of Cantonese listeners' responses suggested that the effect of duration was suppressed by vowel quality in this condition, because even the shortest stimulus (#1) failed to elicit a strong [v] response (< 60%). Starting from stimulus #2, the majority of identification shifted to [a], indicating that vowel quality, not duration, determined Cantonese listeners' perception in this condition.

In other words, Cantonese and Mandarin listeners responded significantly differently towards the vowel quality differences. Whereas the Cantonese subjects were sensitive to and affected by the vowel quality differences, Mandarin subjects were generally unbiased towards these two conditions.

Other factors that reached significance were *stimulus no.* ($F(4.493, 188.696)=48.242, p<0.001$), and *quality* by *stimulus no.* ($F(6.342, 266.382)=4.15, p<0.001$), meaning that the identification rates changed as a function of stimulus no. and vowel quality.

3.2. Discrimination accuracy

A *language group* \times *vowel quality* \times *stimulus pair* repeated measures ANOVA was conducted on the discrimination accuracy scores (the percentage of stimulus pairs identified as 'different'). The statistical results revealed significant main effects of *language group* ($F(1, 42)=23.177, p<0.001$), *quality* ($F(1,42)=9.595, p<0.01$), and a significant interaction of *quality* by *stimulus pair* by *language* ($F(8, 336)=2.142, p<0.05$). No other effects were significant.

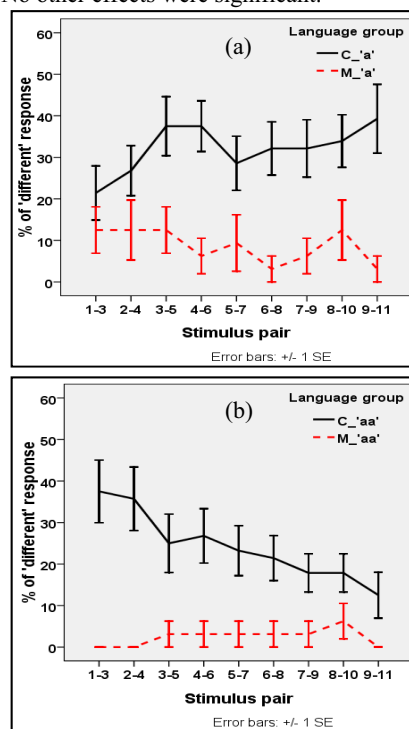


Figure 4. Mean percentage of 'different' responses. 'aa' and 'a' correspond to [a] and [v] in the text. C=Cantonese; M=Mandarin. (a): vowel quality [v] condition (i.e. lengthening [v] to [a]); (b): the vowel quality [a] condition (i.e. shortening [a] to [v]).

The main effect of *language group* can be explained by the overall higher accuracy in the discrimination of Cantonese listeners (see Figure 4). Vowel quality conditions also affected the subjects' perception, but this effect was more obvious in the responses of Cantonese listeners, as their discrimination tended to be more accurate at the longer end (pair 9-11) in the [v] condition, but more accurate at the shorter end (pair 1-3) in the [a] condition. On the other hand, the discrimination accuracy of Mandarin listeners remained low (<20%) regardless of vowel quality conditions and stimulus pairs. Indeed, some Mandarin subjects reported having difficulty in detecting a difference in most of the stimulus pairs.

3.3. Categorical perception

Categorical perception is defined as steep slopes in identification curves and a match between identification crossover section (50%) and discrimination accuracy peak location (cf. [9]). Based on such criteria, it seems that Cantonese listeners perceived duration change in a categorical manner in the [v] condition, but not in the [a] condition, while Mandarin listeners failed to show categorical perception in both conditions. The lack of categorical perception in the performance of Mandarin listeners is predictable given that

Mandarin has no length contrast [10]. We will focus on the performance of Cantonese listeners in the following discussion.

In the [ɐ] condition, Cantonese listeners' identification preference changed from [ɐ] to [a] at stimulus #5. Accordingly, their discrimination accuracy peaked at about stimulus pairs 4-6 and 3-5. Therefore the discrimination peaks were largely aligned with the identification crossover section. However, there was another peak at the end of the discrimination curve (pair 9-11), which had no counterpart in identification. At the current stage of study, it is not clear what factors caused this peak. It is possible that for example, two stimuli drawn from the longer end of the continuum (#9 and #11) may sound more prominent and therefore easier for the listeners to discriminate. However, this account cannot explain the lack of a similar peak in the response of Mandarin listeners. In the current analysis, only seven Cantonese and four Mandarin subjects were included. More subjects are needed to verify whether this peak is a language-specific phenomenon or not.

In the [a] condition, the discrimination peak was located at pair 1-3. In the identification curve, Cantonese listeners' perception changed from [ɐ] to [a] between stimulus #1 and #2. So there seems to be a rough match between identification and discrimination as well. Even though the overall identification rates failed to show an abrupt categorical change, the match of identification and discrimination might suggest a weak trend of categorical perception in this condition.

4. Discussion and conclusion

This study investigated the effect of language experience on the categorical perception of Cantonese vowel duration distinction. We found that: (1) duration change elicited categorical perception in the performance of Cantonese listeners, but not in that of Mandarin listeners; (2) Cantonese listeners were affected by the vowel quality differences, whereas Mandarin subjects were generally unbiased towards the quality differences; (3) vowel quality suppressed the effect of duration in the [a] condition in the performance of Cantonese listeners. Our findings confirmed that attention to the vowel quality difference is language-specific, meaning that quality is incorporated as a linguistic cue in Cantonese.

The finding that vowel quality overrode the effect of duration in some conditions can be explained by the trading relation of duration and quality. According to the temporal organization account, [ɐ] and [a] were different in terms of the relative ratio of nucleus duration versus glide duration [2]. Although [5] reported that glide duration did not change the perception categorically, if the temporal organization account is correct, it indicates that the relative ratio rather than the absolute duration of nucleus determines the perception.

In this study, duration of the following glide was kept constant at 220 ms, which affected the temporal organization of a stimulus. In the case of the shortest stimulus (#1), though the absolute duration of nucleus was close to the average [ɐ] duration (based on the data of the male speaker, see Table 1), the relative ratio of the manipulated stimuli (140 ms/220 ms=0.64) is not like that of the naturally produced [ɐ] diphthongs (150 ms/380 ms=0.39). The not so prototypical temporal organization could have been compensated by vowel quality, meaning that more weight was attached to vowel quality in this condition. This explanation predicts that [ɐ] quality yields more [ɐ] percepts than the [a] quality in this condition, which matched the results (see Figure 3(a) and (b)). More importantly, the trading relation of temporal cue and vowel quality was only found in the performance of Cantonese listeners, as there was no obvious compensation effect of quality in the performance of Mandarin listeners. This finding

again, confirmed the role of language experience in modulating listeners' perception.

Vowel quality differences, which presumably originated from intrinsic differences of duration, are captured by the ears of Cantonese listeners and granted importance in perception. Given the differences in speaking rates both within- and between-speakers, duration alone may not be an efficient cue in categorization. Therefore external reference, such as duration of the glide is used for distinguishing [a] and [ɐ], and quality differences are incorporated to enhance the distinction.

Given the low functional load of duration in Cantonese ([a]-[ɐ] being the only contrastive pair) and less efficient categorization of duration differences, it seems reasonable to speculate that the duration cue may give way to vowel quality differences in Cantonese in the long run.

The specific case of Cantonese reported here may also have implications for understanding the broader picture of duration distinction in the world's languages. If we list all the languages that contrast duration, these languages seem to fall in a spectrum with Finnish-like languages [1] on the one end and Cantonese-like languages on the other. In Finnish, duration is the primary cue for the distinction, and it systematically contrasts pairs of vowels and consonants [1]. On the other hand, Cantonese lacks systematic duration contrast and the distinction is co-encoded by both duration and quality differences. Somewhere in the middle of the spectrum lies English, in which both duration and quality underlie the phonological distinction, but still there are systematic contrasts in pairs of vowels like [i:]-[ɪ] and [u:]-[ʊ]. Despite the tremendous differences between Cantonese and other languages, our findings lead us to ask whether the path of Cantonese would be the future track of Finnish and English in the long-term as a result of reorganization of the internal structure of a language. General questions of this kind await more future research.

5. Acknowledgements

This study owes thanks to the nine Cantonese subjects and five Mandarin subjects who participated in this experiment.

6. References

- [1] Ylinen, S. Shestakova, A. Alku, P. and Huutilainen, M. "The perception of phonological quantity based on durational cues by native speakers, second-language users and non-speakers of Finnish", *Language and Speech*, 48:313-338, 2005.
- [2] Zee, E. "An acoustical analysis of the diphthongs in Cantonese", *Proceedings of the 14th International Congress of Phonetic Sciences*, 2:1101-1104, 1999.
- [3] So, L.K.H. and Wang, J. "Acoustic distinction between Cantonese long and short vowels", *SST Proc.*, 379-384, 1996.
- [4] Zee, E. "Frequency analysis of the vowels in Cantonese from 50 male and 50 female speakers", *Proceedings of the 15th International Congress of Phonetic Sciences*, 1117-1120, 2003.
- [5] Lee, P. Y. "The effect of duration on the perception of the Cantonese diphthongs", unpublished MPhil Thesis, the City University of Hong Kong, 2006.
- [6] Lee, W.-S. "Perceptual cues for identifying the vowels in Cantonese", *Zhongguo yuyan xuebao*, 1:212-220, 2008.
- [7] Shi, F., and Liu, Y. "Xianggang yueyu changduan yuanyin de tingbian shiyan ", in Pan, W.-Y. [Ed]. *Dongfang yuyan yu wenhua*, 98-109, 2002.
- [8] Lehiste, I. *Suprasegmentals*, M.I.T. Press, 1970.
- [9] Liberman, A.M., Harris, K.S. Hoffman, H.S. and Griffith, B.C. "The discrimination of speech sounds within and across phoneme boundaries", *Journal of Experimental Psychology*, 54:358-368, 1957.
- [10] Diehl, R. L., Lotto, A. J., and Holt, L. L. "Speech Perception," *Annual Review of Psychology*, 55:149-179, 2004.