DOES VOWEL INTRINSIC F0 AFFECT LEXICAL TONE?

Jessica Siddins, Jonathan Harrington

Institute of Phonetics and Speech Processing, Ludwig Maximilian University of Munich
jessica | jmh@phonetik.uni-muenchen.de

ABSTRACT

Diachronic change has often been linked with synchronic instability and the realignment of phonological categories according to fine phonetic detail. For example, the development of tones in some languages is widely believed to have been the result of coarticulatory influences of onset obstruent voicing contrasts on the fundamental frequency (f0) of the following vowel. This is just one example of a phenomenon known as intrinsic f0.

Our goal was to investigate whether lexical tone can be influenced by vowel intrinsic f0 effects. In a study of Hong Kong Cantonese, we found that f0 is significantly affected by both lexical tone and vowel openness. Vowel intrinsic f0 effects caused substantial overlap between the tones. We thus find evidence of synchronic instability leading to overlap between distinct phonological categories. These findings may contribute greatly to our understanding of the phonetic precursors to sound change.

Keywords: intrinsic f0, vowel height, lexical tone, sound change

1. INTRODUCTION

Ample evidence shows that coarticulatory effects cause small but perceivable f0 perturbations. This phenomenon is known as intrinsic f0 and has been found, for example, to raise the f0 onset following voiceless stops but lower it following voiced stops [11, 13]. In fact, such effects have been linked to a type of phonetically-driven sound change known as tonogenesis, in which an intrinsic f0 effect becomes phonemic and the underlying segmental contrast which caused it is neutralised [18].

Vowel intrinsic f0 refers to the lowering of f0 on open vowels and raising of f0 on close vowels [16, 22]. Such intrinsic f0 effects are supposed by many [23, 24] to be universal and thus indicative of a physiological or aerodynamic process common to all languages. Some authors, notably [12], have noted that there is very little evidence of diachronic changes based on vowel intrinsic f0 patterns. They argue that the relatively static effects of vowel openness on f0 are much less salient than those of consonant-induced f0 effects, which are believed to be more dynamic, and that vowel intrinsic f0 is therefore unlikely to be involved in sound change. However, to our knowledge, this has never been tested.

In addition, it remains largely unknown how certain types of intrinsic f0 effects operate in languages that use f0 linguistically as the primary cue to lexical tone. Following [20, 21], incomplete perceptual compensation for coarticulation may be the very first step of a sound change within a speech community. In a classic study of sung vowels in English, a non-tone language, [6] found evidence for only partial compensation for vowel intrinsic f0 by listeners. In tone languages, in which presumably both lexical tone and vowel openness influence f0, it may be more difficult to attribute f0 to its source.

We aim to establish whether vowel intrinsic f0 indeed affects the production of lexical tone. We will use Hong Kong Cantonese, which is well known for its crowded tone space, including three level tones, two rising tones and one falling tone. Of these six tones, four share the same f0 onset and several share the same f0 offsets. The Cantonese tones are presumed to be distinguished exclusively by f0 height and slope [7], except for the falling tone, which has also been linked with creaky voice [25, 26]. Thus, in Cantonese, a slight change in f0 height or slope can quickly encroach on the space of another tone [14, 3], independent of coarticulatory effects such as intrinsic f0. We predict that an interaction between lexical tone and vowel intrinsic f0 effects would make tone contrasts in Cantonese even less salient and thus susceptible to misperception, a possible precursor to sound change [20].

According to descriptions of the Hong Kong Cantonese tones [1, 4], there are five equally-spaced canonical tone heights at which a tone may begin and/or end, ranging from level 1 for very low tones to 5 for very high tones. The three level tones are situated at tone levels 22 (for the low tone), 33 (for the mid tone) and 55 (for the high tone). Accordingly, the low tone and the mid tone are closer together in the tone space than the mid tone and the high tone, which are separated by an extra level.

We thus predict the following patterns. On close vowels, the low tone and the mid tone should be raised towards the mid tone and the high tone, respectively. On open vowels, the high tone and the mid tone should fall towards the mid tone and the low tone, respectively. In addition, we predict more overlap between the low tone and the mid tone based on vowel intrinsic f0 effects than between the better separated high tone and mid tone.

2. METHOD

2.1. Stimuli

To investigate the effect of vowel openness, we chose two typical close vowels, /i/ and /y/, and two typical open vowels, /a/ and /aː/. The vowels were spoken on the low level tone, mid level tone and high level tone to test the effect of vowel openness on different ranges of each speaker's f0, as it has been reported that vowel intrinsic f0 effects disappear in the lower ranges of a speaker's f0 [23]. In order to minimise coarticulatory effects from the neighbouring consonants, all tokens began with a voiceless unaspirated alveolar stop and ended with a nasal coda (with alveolar place of articulation where possible).

Table 1: Stimuli

	open vowels		close vowels	
	/a/	/aː/	/i/	/y/
Tone 55	taŋ	taːn	tin	tyn
Tone 33	taŋ	taːn	tim	tyn
Tone 22	taŋ	taːn	tin	tyn

All targets are depicted in Table 1. Each token was produced ten times in isolation and in randomised order, resulting in 120 tokens per speaker.

2.2. Participants

Ten native Cantonese speakers (7 female) aged between 26 and 30 (mean 28 years) participated in this experiment. All participants were born and raised in Hong Kong but had lived in Munich for up to four months at the time of the experiment. The subjects were all active members of the local Cantonese-speaking community and spoke Cantonese on a daily basis. No speakers reported any history of speech, language or hearing disorders.

2.3. Procedure

The stimuli were presented in isolation using unambiguous Traditional Chinese characters for each token. Participants read the stimuli from a wall-mounted computer monitor and were recorded in SpeechRecorder [5] at a sampling rate of 44 100Hz with one channel audio and one channel EGG. Audio was recorded via a head-mounted microphone and EGG via a laryngograph.

2.4. Post-processing and analysis

The audio signal was automatically segmented and labelled using the WebMAUS forced alignment system [15].

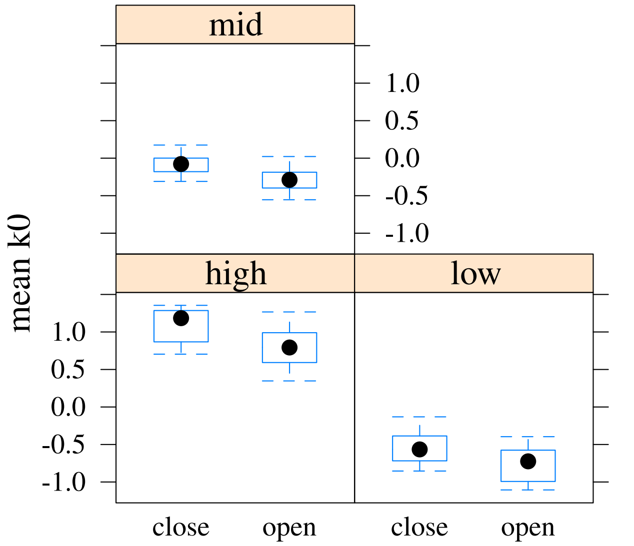
The EGG signal was first bandpass-filtered between 30 and 400Hz in Praat [2] and the f0 determined by calculating the short-term zero-crossing rate (25ms window and 10ms frame shift) of the filtered EGG signal in Emu [8]. In order to remove any effects of voicelessness or creak in the signal at vowel boundaries, the analysis was based on the central 80% of the voiced portion of the signal.

The f0 data were z-score normalised [17] to factor out as far as possible effects of gender and speaker. Given that Cantonese tones differ in height (high/mid/low) and slope (level/rising/falling) [7], we represented them in a two-dimensional space of f0 height and f0 slope. In order to do so, we applied the discrete cosine transformation separately to each vowel token's normalised f0 contour from which we obtained the two lowest-ordered DCT coefficients, k0 and k1, that can be shown to be proportional to the signal's mean and linear slope respectively [8].

To quantify the degree to which the tones overlapped we calculated the Mahalanobis distances between mid and high tones and between mid and low tones in the two-dimensional k0*k1 space. More specifically, in order to compute the distance between mid and high tones, we calculated the centroid (mean position) for the high tones and then the distances of all mid-toned vowels to this centroid. We then reversed the order by computing the distances from each high-toned vowel to the mid tone centroid. The same was done in calculating the distances between low and mid tones. The Mahalanobis distance is positive valued and a probability measure that is the square of the number ellipse standard-deviations in a two-dimensional space [9]. The closer a given tone is to the other tone's centroid, the lower the Mahalanobis distance: at a Mahalanobis distance of zero, the tone is at the other tone's centroid (i.e. at values of zero, the two tones overlap completely).

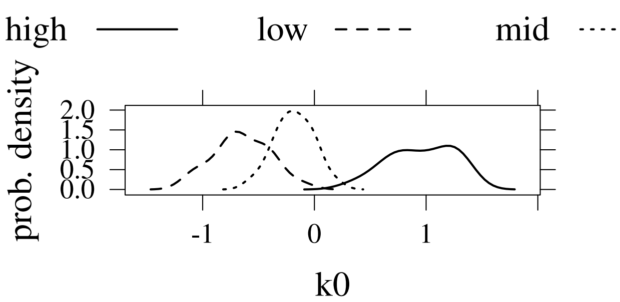
The Mahalanobis distances were categorised according to the following three factors. The factor TONE includes two levels: LM for the distances between low and mid tones, and MH for the distances between mid and high tones. The factor DIRECTION also includes two levels: UP to denote

Figure 1: Mean k0 (f0 height) as a function of tone (low, mid or high) and vowel openness (close, open) and averaged per speaker.



whether the distances were calculated from a lower to a higher tone centroid (from low to mid or from mid to high); DOWN for the other direction (from mid to low or from high to mid). Finally, the factor VOWEL is a recoding of vowel openness also with two levels (OVERLAP and SEPARATION) depending on whether or not vowel openness was expected to push two tonal categories together. For example, in calculating the distances of low tones to the mid-tone centroid, /i/ was coded as OVERLAP because the predicted higher intrinsic pitch due to the close vowel /i/ is expected to push the low tones towards the mid tone centroid. For the calculation of high tones to the mid-tone centroid, /i/ was on the other hand coded as SEPARATION because the increase in f0 on the high tones due to /i/ is expected to shift the high tones further away from the mid

Figure 2: Probability density of k0 (f0 height) as a function of tone height with vowel openness collapsed.



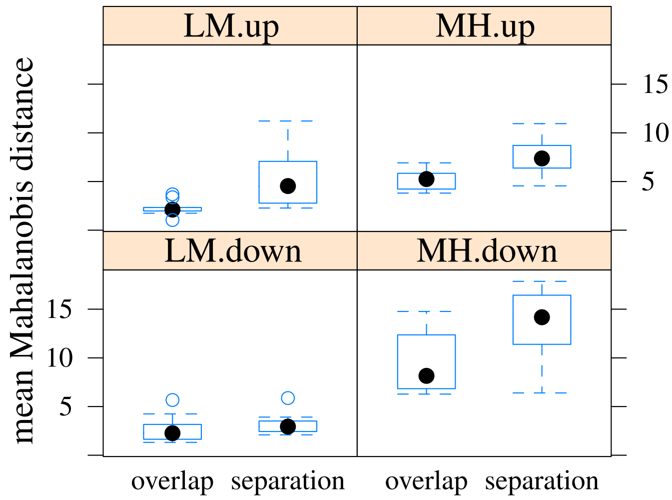
tone centroid. The main hypothesis to be tested in this paper is that there is an effect of VOWEL which would mean that tonal categorisation (the confusion between low and mid or between mid and high) is influenced by the intrinsic f0 consequences of vowel openness.

We ran a mixed model with Mahalanobis distance as the dependent variable, DIRECTION (2 levels: UP, DOWN), TONE (2 levels: LM, MH) and VOWEL (2 levels: SEPARATION, OVERLAP) as fixed factors and speaker as random factor.

3. RESULTS

Fig. 1 shows the separation of the tones by vowel openness based on k0 (f0 height). For each tone, the mean k0 is higher on close vowels than open vowels. In addition, when we consider the variation in Fig. 2, we see substantial overlap between neighbouring tones. The overlap appears to be strongest between "nearest" neighbours (i.e. the low and mid tones) than between tones separated by more than one tone level (i.e. the mid and high tones).

Figure 3: Mean Mahalanobis distances (one value per speaker) for each combination of fixed factors DIRECTION, TONE and VOWEL (see text). Vowel separation vs. overlap indicates the effect of vowel openness.



The Mahalanobis distances for the four combinations of factors TONE and DIRECTION are plotted in Fig. 3. A difference in Mahalanobis distance between the levels OVERLAP and SEPARATION indicates an effect of vowel openness. Thus, we see effects of vowel openness on the movement of the low tone relative to the mid tone (LM.UP), the mid tone relative to the high tone (MH.UP) and the high tone relative to the mid tone (MH.DOWN). There

Table 2: Summary statistics of a mixed model with the Mahalanobis distances as the dependent variable showing the significance of the fixed factors DIRECTION (UP, DOWN), TONE (LM, MH), VOWEL (OVERLAP, SEPARATION) and their interactions.

	Estimate	Std. Error	t value
(Intercept)	2.94	0.48	6.00
directionup	-0.67	0.71	-0.94
toneMH	6.52	0.89	7.33
vowelseparation	-0.06	0.74	-0.08
directionup:toneMH	-3.59	0.91	-3.95
directionup:vowelseparation	2.85	0.91	3.14
toneMH:vowelseparation	3.96	0.91	4.36
directionup:toneMH:vowelseparation	-4.42	1.29	-3.44

appears to be little effect of vowel height on the Mahalanobis distance of the mid tone relative to the low tone (LM.DOWN).

Assuming significance for t-statistics greater than 2, the results of the mixed model in Table 2 show a significant effect of factor TONE as well as significant effects for all interactions. In order to analyse the three-way interaction, we carried out post-hoc Tukey tests between all pairwise combinations of the levels of the fixed factors. Compatibly with Figs. 1 and 3, the results showed a significant effect for VOWEL in all contexts (MH.UP: $z = 3.15, p < 0.05$; MH.DOWN: $z = 5.26, p < 0.001$; LM.UP: $z = 3.76, p < 0.01$) except for LM.DOWN (cf. Fig. 3, in which there is little difference between OVERLAP and SEPARATION for this condition). These results indicate that vowel openness significantly affected the distance from the high tone to the mid tone (MH.DOWN), the mid tone to the high tone (MH.UP), and the low tone to the mid tone (LM.UP). However, vowel height did not influence the distance from the mid tone to the low tone (LM.DOWN).

4. DISCUSSION

Previously, vowel intrinsic f0 was not believed to be salient enough to be a phonetic trigger of sound change. Interestingly, however, our data show an effect of intrinsic f0 on lexical tone in Hong Kong Cantonese speech production.

Open vowel lowering appeared to have the strongest effect, but only on the high tone - it had no effect on the distance of the mid tone to the low tone. Closed vowel raising, however, affected the low and mid level tones equally. We thus find a stronger influence of vowel intrinsic f0 on the relationship between the mid and high level tones than on the low and mid level tones (cf. MH vs. LM panels of Fig. 3), in accordance with previous evidence that intrinsic f0 effects decrease in the lower f0 regions [23]. However, according to the canonical descriptions of the Cantonese level tones cited in Section 1 and the confirmation of this description in Fig. 2, the mid and high level tones are more clearly separated than the low and mid level tones and thus less susceptible to overlap and misclassification based on vowel intrinsic f0 effects. In fact, the low and mid tones are said to be undergoing a tone merger in Hong Kong Cantonese [19]. Thus, based on these data we might predict that it is the low tone that is likely to drift upwards toward the mid tone (rather than the mid tone sinking toward the low tone). This might have the benefit of freeing up space in the lower end of the Cantonese tone system, which has four tones (including the low level tone) starting at tone level 2. Thus, it may be that a combination of the close proximity between the low and mid tones combined with coarticulatory factors such as vowel intrinsic f0 are the driving factors behind this sound change.

In order for a sound change to occur, we first need to test listeners' ability to separate the effects of vowel openness from those of lexical tone on f0. If listeners are able to correctly attribute vowel intrinsic f0 effects to their source (i.e. differing vowel openness) rather than to f0 caused by lexical tone alone, we would not have reason to believe that coarticulatory effects of vowel openness on f0 would be involved in a sound change. However, if listeners' compensation is incomplete, we might expect slight misparsing of the phonetic effects of vowel openness as phonological effects of lexical tone and the beginning of a sound change [10, 20, 21].

5. ACKNOWLEDGEMENTS

This research was funded by European Research Council grant 295573 to Jonathan Harrington. Many thanks to our Cantonese informant, Siyi Li, for help choosing stimuli and carrying out the experiment.

6. REFERENCES

[1] Bauer, R. S., Benedict, P. K. 1997. *Modern Cantonese Phonology*. Number 102 in Trends in linguistics: Studies and monographs. Berlin: Mouton de Gruyter.

[2] Boersma, P., Weenink, D. 2012. *Praat: doing phonetics by computer* version 5.3.04 edition.

[3] Brunelle, M., Lim, J., Chow, D. September 2010. Tone identification and confusion in Cantonese. *Conference on the Typology of Tone and Intonation (TIE4)* Stockholm.

[4] Chao, Y. R. 1947. *Cantonese Primer*. Cambridge: Harvard University Press.

[5] Draxler, C., Jänsch, K. 2004. SpeechRecorder - a Universal Platform Independent Multi-Channel Audio Recording Software. *Proc. of the IV. International Conference on Language Resources and Evaluation* Lisbon, Portugal. 559–562.

[6] Fowler, C. A., Brown, J. M. 1997. Intrinsic f0 differences in spoken and sung vowels and their perception by listeners. *Perception & Psychophysics* 59(5), 729–738.

[7] Gandour, J. 1979. Perceptual dimensions of Cantonese tones: a multidimensional scaling reanalysis of Fok's tone confusion data. *South-east Asian Linguistic Studies* 4, 415–429.

[8] Harrington, J. 2010. *Phonetic Analysis of Speech Corpora*. Wiley Publishing.

[9] Harrington, J., Cassidy, S. 1999. *Techniques in Speech Acoustics*. Foris, Dordrecht: Kluwer Academic Publishers.

[10] Harrington, J., Kleber, F., Reubold, U. 2013. The effect of prosodic weakening on the production and perception of trans-consonantal vowel coarticulation in German. *Journal of the Acoustical Society of America* 134(1), 551–561.

[11] Haudricourt, A. G. 1972. Two-way and three-way splitting of tonal systems in some Far-Eastern languages. *Tai Phonetics and Phonology* (1), 58–86.

[12] Hombert, J.-M. 1977. Development of tones from vowel height? *Journal of Phonetics* 5, 9–16.

[13] Hombert, J.-M., Ohala, J. J., Ewan, W. G. 1979. Phonetical explanations for the develoment of tones. *Language* 55(1).

[14] Khouw, E., Ciocca, V. 2007. Perceptual correlates of Cantonese tones. *Journal of Phonetics* 35(1), 104–117.

[15] Kisler, T., Schiel, F., Sloetjes, H. 2012. Signal processing via web services: the use case WebMAUS. *Digital Humanities* Hamburg, Germany. 30–34.

[16] Lehiste, I., Peterson, G. E. 1961. Some basic considerations in the analysis of intonation. *The Journal of the Acoustical Society of America* 33(4), 419–425.

[17] Lobanov, B. 1971. Classification of Russian vowels spoken by different speakers. *Journal of the Acoustical Society of America* 49, 606–608.

[18] Matisoff, J. A. 1973. Tonogenesis in Southeast Asia. In: Hyman, L. M., (ed), *Consonant types and tones*. Southern California Occasional Papers in Linguistics.

[19] Mok, P., Zuo, D., Wong, P. W. Y. 10 2013. Production and perception of a sound change in progress: Tone merging in Hong Kong Cantonese. *Language Variation and Change* 25, 341–370.

[20] Ohala, J. J. 1981. The listener as a source of sound change. In: Masek, C. S., Hendrick, R. A., Miller, M. F., (eds), *Papers from the Parasession on Language and Behavior*. Chicago: Chicago Linguistic Society 178–203.

[21] Ohala, J. J. 1993. The phonetics of sound change. In: Jones, C., (ed), *Historical Linguistics: Problems and Perspectives*. London: Longman 237–278.

[22] Shadle, C. H., Pierrehumbert, J. B., Liberman, M. Y. 1979. The intrinsic pitch of vowels in sentence context. *The Journal of the Acoustical Society of America* 66(S1), S64–S64.

[23] Whalen, D., Levitt, A. G. 1995. The universality of intrinsic f0 of vowels. *Journal of Phonetics* 23, 349–366.

[24] Whalen, D., Levitt, A. G., Hsiao, P.-L., Smorodinsky, I. 1994. Intrinsic f0 of vowels in the babbling of 6-, 9- and 12-month-old French- and English-learning infants. *Haskins Laboratories Status Report on Speech Research* SR-117/118, 15–24.

[25] Yu, K. M., Lam, H. W. August 17-21 2011. The role of creaky voice in Cantonese tonal perception. *Proceedings of the 17th International Congress of Phonetic Sciences* Hong Kong. 2240–2243.

[26] Yu, K. M., Lam, H. W. 2014. The role of creaky voice in Cantonese tonal perception. *The Journal of the Acoustical Society of America* 136(3), 1320–1333.