# HIGH LEVEL TONE IS SPECIAL: VOCAL RESPONSES TO PITCH PERTURBATION IN CANTONESE

Li-Hsin Ning

Department of English (Linguistics Track), National Taiwan Normal University, Taiwan
lihsin@ntnu.edu.tw

## ABSTRACT

This paper investigates the vocal responses to pitch perturbation for the three level tones in Cantonese. Eight native Cantonese speakers were recruited and asked to produce three Cantonese words: *si55-ji33* 'be poetic,' *tsi33-ji33* 'to greet,' and *si11-ji33* 'to signal.' The three words had the same mid-level tone on the second syllable but differed in tone on the first syllable (thus abbreviated as HM, MM, and LM). The pitch of the first syllable was shifted for 200 ms, with ±350 cents for HM, +350/-50 cents for MM, and ±50 cents for LM. The manipulation of the pitch-shift magnitude was to overlap the shifted pitch with another lexical tone. The results show that the largest pitch-shift peak amplitudes were elicited when the H level tone of the HM word was downshifted 350 cents to the M level. This suggests that maintaining high level tone requires more effort and thus a small deviation from the target high pitch leads to larger compensation.

**Keywords**: pitch-shift paradigm, compensatory responses, high level tone, Cantonese.

## 1. INTRODUCTION

The pitch-shift paradigm where an artificial change in pitch (either upward or downward) was fed back to speakers during vocalizations has been used to investigate the online monitoring system in voice production. Previous research has shown that speakers typically compensate for the perturbation [2, 3, 7]. In other words, when the perceived pitch goes down, speakers tend to raise their pitch; when the perceived pitch goes up, speakers tend to lower their pitch. This compensation (or pitch-shift response) shows that an audio-vocal corrective mechanism is employed in the internal model in order to minimize the mismatch between the desired pitch and the actual pitch [5, 6].

Recent research interest has focused on the pitch-shift responses in different F0 heights [8, 9, 12]. All these studies found that larger and faster pitch-shift responses appeared at a high F0 voice compared to a low F0 voice (using the sustained vowel /a/). This suggests that high F0s which involve greater laryngeal effort are more susceptible to pitch perturbation. What remains interesting is whether this F0 height effect in pitch-shift responses would also appear when it applies to tonal categories. Ning's research [11] on the level tone production of Taiwanese Southern Min has shown that the magnitude of the pitch-shift responses was related to the degree of categorical perception. If the speakers cannot categorically perceive the pitch differences, they would not produce a corrective / compensatory response under the auditory perturbation. In this study, we followed the experimental protocol in Ning [11] and examined the pitch-shift responses for the three level tones in Cantonese (high, mid, and low) by overlapping one lexical tone with another. Since tonal changes contribute to meaning contrast in Cantonese, we would expect larger corrective pitch-shift responses in tonal word production when the desired tone is shifted.

## 2. METHODS

### 2.1. Participants

Eight native Cantonese speakers (4 females and 4 males, aged 20 to 24) from Macaw were recruited. A hearing test using an MAICO pure tone audiometer (model MA 25) was administered to all the participants. None of the participants reported a history of voice disorders and all passed a hearing test at 20 dB hearing level bilaterally at 250, 500, 750, 1000, 2000, 3000, and 4000 Hz. All the participants gave informed consent and received financial compensation for their time.

### 2.2. Materials and procedures

The experimental protocol follows Ning's study on Taiwanese Southern Min [11]. There were three sessions of recordings. In each session, the participants had to produce one of the three Cantonese words: *si*$^{55}$-*ji*$^{33}$ 'be poetic,' *tsi*$^{33}$-*ji*$^{33}$ 'to greet,' and *si*$^{11}$-*ji*$^{33}$ 'to signal.' The three words are minimal-like in a way that the second syllable has a mid-level tone, but the first syllable carries a high, mid, and low level tone, respectively. The three stimuli were symbolized as HM (for *si*$^{55}$-*ji*$^{33}$), MM (for *tsi*$^{33}$-*ji*$^{33}$), and LM (for *si*$^{11}$-*ji*$^{33}$).

An additional 10 native Cantonese speakers (5 females and 5 males) were recruited to estimate the pitch differences between the level tones. They were asked to produce each word ten times. The F0 value of the first syllable was then converted into cents using the formula in (1):

(1) cents = 1200*log2(f0/baseline)

where the baseline was the F0 value of the corresponding second syllable.

The results of the 10 additional recordings show that on average, the H level tone was 350 cents higher than the M level tone, whereas the L level tone was 50 cents lower than the M level tone.

The pitch perturbation paradigm is illustrated in Figure 1. The waveform shows an example of the disyllabic Cantonese word production. At 100 ms after production onset, the pitch signal in the auditory feedback was shifted for 200 ms, without smooth transition. In other words, only the first syllable was affected by the pitch-shifts.

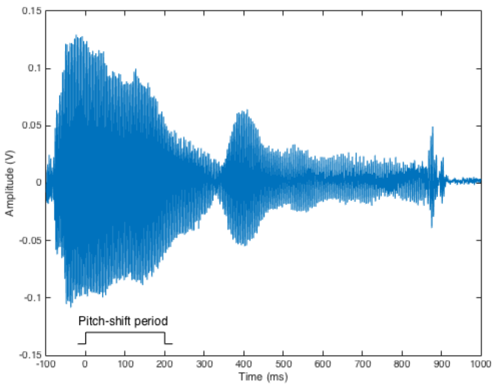
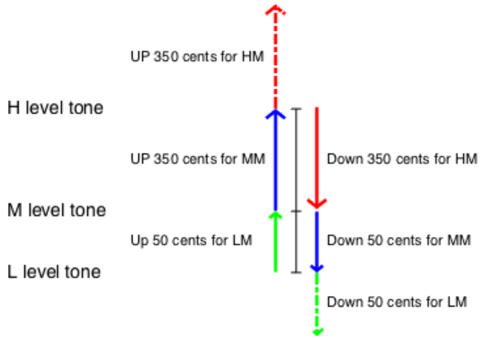


**Figure 1**: **The pitch-shit stimulus design.** The pitch-shift signal in the auditory feedback appeared at 100 ms after vocalization onset and remained for 200 ms. In other words, only the first syllable was affected by pitch perturbation.

The magnitude of pitch perturbation was based on the pitch differences obtained from the 10 additional recordings. The H level tone in HM was shifted either 350 cents upward or 350 cents downward (see the red arrows in Figure 2). The first M level tone in MM was shifted either 350 cents upward or 50 cents downward (see the blue arrows in Figure 2). The L level tone in LM was shifted either 50 cents upward or 50 cents downward (see the green arrows in Figure 2). These pitch-shifts could trick the speakers into thinking that they had made an error in tone production. We hypothesize that pitch-shift responses would be larger and the onset of the peak responses should be faster when the shifted pitch was overlapped with another lexical tone (indicated as the solid arrows in Figure 2). No significant compensation was expected in cases where the pitch-shifts did not change the tonal categories (indicated as the dashed arrows in Figure 2).

In each recording session, participants were asked to produce one of the three Cantonese words 60 times (for example, uttering the HM word 60 times in session 1, the MM word 60 times in session 2, and the LM word 60 times in session 3). This leads to 180 productions in total per person. The order of the three Cantonese words was counterbalanced. The shift direction could be upward, downward, or no shift (as a control). The order of three shift directions was randomized so that the participants would not be able to predict it.

**Figure 2**: **The shift magnitude and direction for HM, MM, and LM.** The solid arrows indicate the shifted pitch overlaps with another lexical tone in Cantonese. The dotted arrows indicate the shifted pitch goes beyond the tonal category in Cantonese. The red arrows stand for the perturbation in HM, the blue arrows stand for the perturbation in MM, and the green arrows stand for the perturbation in LM.



### 2.3. Data analysis

The pitch values were converted into cents using the formula in (1), where the baseline is the mean F0 of the 100 ms pre-shift production. Only compensatory responses were taken into account. Following responses which had the same direction as the pitch-shifts and non-responses which failed to show the obvious compensatory or following pattern were excluded prior to data averaging. The exclusion rate was 22%. Difference waves were generated by taking the difference between the upward/downward shift and the control. The peak amplitudes and peak latencies were then measured from the difference waves. Repeated measures ANOVAs fitting the linear model TONE (3 levels: HM, MM, and LM) x DIRECTION (2 levels: upward and downward) were performed on the difference waves. In order to examine the DIRECTION effect, the absolute peak amplitudes were used.

## 3. RESULTS

### 3.1. Peak amplitudes

The ANOVA results are summarized in Table 1. There was a significant main effect of TONE ($F(2,42)=20.711$, $p<.001$) and a significant main effect of DIRECTION ($F(1,42)=11.032$, $p<.01$). Post-hoc comparisons with the Tukey procedure for controlling the 95% family-wise confidence level show that larger peak amplitudes were elicited in HM than in MM and LM. Word productions with downward shifts also demonstrated larger peak amplitudes than those with upward shifts.

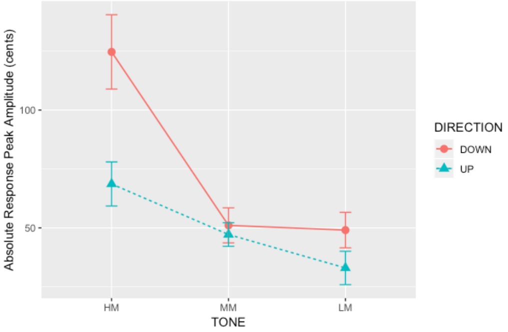
A significant interaction between TONE and DIRECTION was observed ($F(2,42)=4.267$, $p<.05$; see Figure 3). Simple main effect analyses were conducted at each level of TONE and each level of DIRECTION. There was a significant DIRECTION effect for HM ($F(1,14)=9.342$, $p<.01$), but not for MM ($F(1,14)=.188$, $p=.671$) and LM ($F(1,14)=2.414$, $p=.143$). In HM, downward shifts experienced larger peak amplitudes than upward shifts, suggesting that speakers were susceptible to pitch perturbation particularly when the H level tone was overlapped by the M level tone. On the other hand, simple TONE effects were observed in both downward shifts and upward shifts, where HM showed the largest peak amplitudes.

**Table 1**: The ANOVA results for the peak amplitudes

| **Factor** | **Statistics** | **Pairwise comparison** |
|---|---|---|
| TONE | $F(2,42)=20.711$, $p<.001$ | HM > MM = LM |
| DIRECTION | $F(1,42)=11.032$, $p<.01$ | DOWN > UP |
| TONE x DIRECTION | $F(2,42)=4.267$, $p<.05$ | |
| | DIRECTION effect for HM: $F(1,14)=9.342$, $p<.01$ | DOWN > UP for HM |
| | DIRECTION effect for MM: $F(1,14)=.188$, $p=.671$ | DOWN = UP for MM |
| | DIRECTION effect for LM: $F(1,14)=2.414$, $p=.143$ | DOWN = UP for LM |
| | TONE effect for downward: $F(2,21)=15.44$, $p<.001$ | HM > MM = LM for DOWN |
| | TONE effect for upward: $F(2,21)=5.94$, $p<.01$ | HM > LM<br>HM = MM<br>LM = MM<br>for UP |

*Note*. The '>' signs in the pairwise comparisons indicate significance, while the '=' signs indicate insignificance.

**Figure 3**: **The absolute pitch-shift response peak amplitudes (cents).**



### 3.2. Peak latencies

No main effects of TONE ($F(2,42)=.978$, $p=.385$) and DIRECTION ($F(1,42)=.077$, $p=.783$) were found for peak latencies. There was no significant interaction between TONE and DIRECTION ($F(2,42)=.059$, $p=.942$), either.

### 3.3. Planned comparisons

We conducted the planned comparisons on two things: the perturbations between H and M level tones (the blue solid up arrow vs. the red solid down arrow in Figure 2), and the perturbations between M and L level tones (the green solid up arrow vs. the blue solid down arrow in Figure 2). The results show that shifting H down to the M level tone (the red solid down arrow) had larger peak amplitudes than shifting M up to the H level tone (the blue solid up arrow) ($F(1,42)=21.93$, $p<.001$). However, no significant differences were found in the comparison of shifting M down to the L level tone and shifting L up to the M level tone. The findings suggest that the H level tone is special in speech motor control and overlapping H with the M level tone gives rise to a larger compensatory / corrective response.

## 4. DISCUSSION

This study investigates the compensatory pitch-shift responses for the three level tones in Cantonese. Compensation which appears when there is a mismatch between the expected pitch and perceived pitch has been claimed as a corrective mechanism in the internal model. The purpose of the compensation is to reduce the perceived error. Our results show that shifting H to the M level tone had the largest compensatory responses.

Our initial expectation was that larger pitch-shift responses would appear when the perceived pitch had been overlapped with another lexical tone in

Cantonese (i.e., the solid arrows in Figure 2). However, in our data, larger compensation was only seen in the downward shifts of H in HM. This corresponds to the previous studies which argue for greater pitch-shift response amplitudes at higher F0 levels (using a sustained vowel /a/) [8, 9, 11]. It seems that the H level tone involving higher laryngeal muscle effort is special. The audio-vocal control mechanism is sensitive to the unexpected changes that are linguistically meaningful. Therefore, shifting the H tone 350 cents down triggers large pitch-shift responses to compensate the perceived errors during word production. The lack of significance in the upward shift of H in HM can be attributed to the fact that it lies outside of the tonal category range in Cantonese. Thus, the audio-vocal mechanism ignores the linguistically non-meaningful pitch change.

Consider the M level tone in MM and the L level tone in LM. No significant DIRECTION effect was observed. It appears that the M level tone and L level tone are close in pitch (with only 50 cents difference) Acoustically, it may not be easy for listeners or speakers to differentiate the M and the L. Therefore, a small pitch shift (50 cents) on the M level tone or the L level tone has no effect on the audio-vocal corrective system. What remains unknown is why shifting M to the H level tone (350 cents) did not generate a larger compensatory response. We argue that like in Taiwanese Southern Min [11], the M level tone in Cantonese could be less categorically perceptible. The less degree of categorical perception could make it less susceptible to pitch perturbation.

In general, downward shifts lead to larger pitch-shift responses than upward shifts. This has been evidenced in many previous research [1, 4, 10]. The cause is still unknown. We speculate that the audio-vocal mechanism is designed for stabilizing high F0 and maintaining the high target. Therefore, downward shifts deviating from high target frequencies give rise to larger responses, while upward shifts enhancing the frequencies are favoured and are not regarded as errors.

Overall, compensatory responses have been claimed to be a corrective mechanism that is used to modify the mismatch between the expected motor output and the actual perceived or sensed output [5, 6]. The compensatory pitch-shift responses typically are no larger than the magnitudes of pitch-shifts and are tuned for small perturbations. The results in the present study adds to our knowledge that the degree of compensation does not only depend on the physical properties of the pitch-shifts (such as the shifted magnitude) but also on the linguistic meaning.

## 5. CONCLUSION

High level tone is special. Larger compensatory responses were observed when the H level tone was masked by the M level tone. This suggests that our audio-vocal mechanism aims at maintaining and stabilizing high frequencies.

## 6. ACKNOWLEDGMENTS

The author gratefully acknowledges the assistance of Jian Leat Siah in preparing the test stimuli and Jia-Chen Lin in the data collection.

## 7. REFERENCES

[1] Behroozmand, R., Karvelis, L., Liu, H., & Larson, C. 2009. Vocalization-induced enhancement of the auditory cortex responsiveness during voice F0 feedback perturbation. *Clinical Neurophysiology* 120, 1303-1312.

[2] Burnett, T., Freeland, M., Larson, C. 1998. Voice F0 responses to manipulations in pitch feedback. *Journal of the Acoustical Society of America* 103(6), 3153-3161.

[3] Burnett, T., & Larson, C. 2002. Early pitch-shift response is active in both steady and dynamic voice pitch control. *Journal of the Acoustical Society of America* 112(3), 1058-1063.

[4] Chen, S., Liu, H., Xu, Y., Larson, C. 2007. Voice F0 responses to pitch-shifted voice feedback during English speech. *Journal of the Acoustical Society of America* 121(2), 1157-1163.

[5] Guenther, F., Hampson, M., Johnson, D. 1998. A theoretical investigation of reference frames for the planning of speech movements. *Psychological Review* 105, 611-633.

[6] Hickok, G., Houde, J., Rong, F. 2011. Sensorimotor integration in speech processing: Computational basis and neural organization. *Neuron* 69(3), 407-422.

[7] Larson, C. (1998). Cross-modality influences in speech motor control The use of pitch shifting for the study of F0 control. *Journal of Communication Disorders* 31, 489-503.

[8] Liu, H., Auger, J., Larson, C. 2010. Voice fundamental frequency modulates vocal response to pitch perturbations during English speech. *Journal of the Acoustical Society of America* 127(1), EL1-EL5.

[9] Liu, H., Larson, C. 2007. Effects of perturbation magnitude and voice F0 level on the pitch-shift reflex. *Journal of the Acoustical Society of America* 122(6), 3671-3677.

[10] Liu, H., Meshman, M., Behroozmand, R., Larson, C. 2011. Differential effects of perturbation direction and magnitude on the neural processing of voice pitch feedback. *Clinical Neurophysiology* 122, 951-957.

[11] Ning. L.-H. 2019. Pitch-shift responses as an online monitoring mechanism during level tone production. *Journal of the Acoustical Society of America, 145(4),* 2192-2197.

[12] Sturgeon, B., Hubbard, R., Schmidt, S., Loucks, T. 2015. High F0 and musicianship make a difference: Pitch-shift responses across the vocal range. *Journal of Phonetics* 51, 70-81.