Sensorimotor adaptation and aftereffect to frequency-altered feedback in Mandarin-speaking vocalists and non-vocalists

[中文母語聲樂家與非聲樂家對於音頻擾動反饋所產生的感覺動作適性化與後效]

Li-Hsin Ning [甯俐馨]

National Taiwan Normal University [國立臺灣師範大學]

This research examined sensorimotor adaptation and aftereffect in trained vocalists and non-vocalists whose native language is Mandarin. The adaptive frequency-altered feedback paradigm involving a baseline of normal auditory feedback, a training phase of incrementally or decrementally changed feedback, and a test phase of normal auditory feedback was administered. The participants were asked to produce the sustained vowel /a/, Mandarin /ma1/ (“mother”), and Mandarin /ma2/ (“hemp”). The results show that the vocalists compensated less than the non-vocalists, suggesting that the vocalists’ audio-motor representations for pitch could be more entrenched than the non-vocalists’. All the participants displayed sensorimotor adaptation, indicating that online recalibration is an innate and automatic process. The presence of the aftereffect, however, depended on the stimulus type and vocal training experience. It appeared in all speakers’ responses to downward shift of /ma1/ and /ma2/, but only in the non-vocalists’ responses to /a/.

Keywords: sensorimotor adaptation, frequency-altered feedback paradigm, vocal training, tone production

關鍵詞：感覺動作適性化、音頻擾動反饋典範、聲樂訓練、聲調發音

1. Introduction

Motor control and learning is an essential aspect of life. From birth, we learn to walk, manipulate objects, and to communicate. Movement-making capacity is critical for survival. To regulate our movements, we heavily rely on sensory information (or movement-produced sensory feedback). This tells us about the state of the environment and the state of our body. For example, visual feedback obtained from the angle of elevation between a fielder and the ball provides him with information about where to go to intercept the ball. You can imagine how difficult it is to maintain posture and balance during movement with the eyes closed, and visual feedback is also essential for controlling stability. Audition is to some extent similar to vision. For instance, bats flying at night use acoustic information (including sounds from objects and from their own movements) to orient themselves. Speakers tend to increase their volume in a noisy environment (called the Lombard effect (Lane & Tranel 1971)) in order to enhance audibility, which would not occur if auditory feedback was not actively functioning. All these examples reveal the fact that sensorimotor (i.e., sensory-to-motor) relations significantly influence the regulation of skilled and functional behavior (including speech).

Auditory feedback is particularly important for learning to sing or speak in that it is associated with checking vocal output against a copy of the expected articulatory movement. For pre-linguistically deaf children, absence of auditory feedback poses a challenge in learning an oral language, since they lose the ability to check their articulation against the intended speech output through auditory feedback (Oller & Eilers 1988). For post-lingually deaf people, absence of auditory feedback leaves it difficult to regulate suprasegmental features such as loudness and intonation (Waldstein 1990, Lane & Webster 1991, Cowie & Douglas-Cowie 1992).

Laboratory studies have empirically examined this audio-motor relationship by perturbing auditory feedback during vocalization. Auditory feedback perturbation has been used to enhance speech fluency in adults who stutter (Sparks et al. 2002, Van Borsel, Reunes & Van den Bergh 2003, Chon et al. 2013) and to explore the impairment of sensorimotor integration in children with speech sound disorders and the elderly with Parkinson's disease (Terband, van Brenk & van Doornik-van der Zee 2014, Mollaei et al. 2016). In an auditory feedback perturbation paradigm, an artificial change in segments (such as vowel formants or spectral noise in fricative) or suprasegmentals (such as pitch, loudness or duration) was fed back to speakers while they were vocalizing. Speakers typically produce an automatic and reflex-like response that is in the opposite direction of the perturbation stimulus (i.e., a compensatory response). The **compensation** is to minimize the perceived error that happens in the auditory feedback. With contin-

uous and incrementally changed perturbation in auditory feedback, speakers can demonstrate the **sensorimotor adaptation**, in which they keep updating or adapting the feedforward commands of the motoric plans (Houde & Jordan 1998, 2002, Max, Wallace & Vincent 2003). Compensation and adaptation can persist for a period of time even when the perturbation is removed, which is called the **aftereffect**. The aftereffect suggests that plasticity is involved in speech motor planning (see Parrella & Houde 2019 for a complete review).

In what follows, I will review the potential factors that influence the patterns of compensation, sensorimotor adaptation, and aftereffect.

1.1 Compensation: The effects of vocal experience and language experience

Evidence from auditory feedback perturbation studies has consistently indicated that musical experience and (tone) language experience make a difference to speakers' **compensatory** responses to pitch perturbation. For instance, Zarate & Zatorre (2005, 2008) showed that experienced singers were able to suppress responses to altered feedback when asked to ignore 200-cent shifts. However, such suppression was not observed in non-singers. A similar suppression effect was discovered in Ning, Loucks & Shih (2015) where trained English-speaking vocalists, compared to English-speaking non-vocalists, had attenuated pitch-shifted responses to 200-cent shifts in the /a/ prolongation. Behroozmand et al. (2013) further examined the response patterns of musicians and non-musicians, and found that non-musicians failed to return to the baseline F0 after the offset of brief pitch-shifts (200 ms) while musicians did.

As for the language experience effect, smaller pitch-shift responses were observed in Cantonese speakers than in Mandarin speakers when they heard pitch-shift stimuli of ± 200 and ± 500 cents (Liu et al. 2010), suggesting that language-dependent modulation is involved in voice F0 control. As shown in Ning, Shih & Loucks (2014), second language learners of Mandarin were more susceptible to pitch perturbation than native Mandarin speakers. While native Mandarin speakers can suppress pitch perturbation, second language learners of Mandarin showed larger compensation to perturbed auditory feedback.

1.2 Sensorimotor adaptation and aftereffect

The preceding paragraphs reveal the importance of vocal training experience and tonal language experience in **compensation** to pitch perturbation. These two factors can also be influential in **sensorimotor adaptation** and **aftereffect**. To examine the **sensorimotor adaptation** and **aftereffect** in voice F0 control, the adaptive frequency-altered feedback (FAF for short) paradigm is used. The adaptive FAF

paradigm involves a **baseline** phase (where normal auditory feedback is provided), a **training** phase (where pitch values are shifted either upward or downward successively by a small amount, such as 2 cents, until a certain value, such as 100 cents, is reached), and a **test** phase (where auditory feedback returns to normal). A variant of the adaptive FAF paradigm contains a constant pitch shift in the training phase. Unlike the typical (or non-adaptive) FAF paradigm where pitch-shifts appear briefly (for example, 200 ms) in the middle of vocalizations (for example, 1000–1500 ms after vocal onset), auditory feedback in the adaptive FAF paradigm is shifted from vocal onset to vocal offset. Therefore, in the adaptive FAF paradigm, speakers may not be aware of any change in auditory feedback throughout the experiment. As noticing the presence of auditory perturbation has nothing to do with the compensatory response pattern (Purcell & Munhall 2006), we are able to investigate possible and unconscious changes in speech motor commands within a short period of experimental session.

1.2.1 The effect of vocal training experience

Previous adaptation studies have examined singers' and non-singers' voices singing /ta/ or /a/ at a given pitch height (for example, 349 Hz (F₄), 392 Hz (G₄), or 440 Hz (A₄) in Jones & Keough 2008 and Keough & Jones 2009, 220 Hz (A₃), 246.94 Hz (B₃), 293.66 Hz (D₄), 329.63 Hz (E₄) in Keough, Hawco & Jones 2013). They found that both singers and non-singers compensated for the FAF trials in the training phase. However, non-singers could compensate more than singers (Jones & Keough 2008). Sensorimotor adaptation was observed for both singers and non-singers, suggesting that both singers and non-singers updated their internal models and adjusted their F₀ values in order to get closer to the intended pitch (Keough, Hawco & Jones 2013, Keough & Jones 2009). The level of sensorimotor adaptation was unaffected by the type of instruction (either to ignore or to compensate) (Keough, Hawco & Jones 2013). Nevertheless, adaptive control is still qualitatively different between singers and non-singers. In the FAF trials, singers were more accurate than non-singers in the pitch vocalizations, as the singers' productions were closer to the desired pitch target. Non-singers, due to a lack of vocal training, had difficulty uttering a correct pitch target and thus treated the continuous changes in pitch as random errors. This suggests that singers' audio-motor representations for pitch, obtained through years of practice, are more entrenched than non-singers'.

Research has shown inconsistent results on the aftereffect of singers and non-singers, an effect that appears in the test phase where speakers responded (compensated) as if they still heard the pitch perturbation. Jones & Keough (2008) argue that singers but not non-singers showed the aftereffect; however, Keough & Jones (2009) claim that the aftereffect appeared in both singers and non-singers.

One potential reason is the difference in the experimental designs between the two studies. The perturbation magnitude in the FAF trials was constant in Jones & Keough (2008) but gradually changed in Keough & Jones (2009). It is likely the aftereffect co-occurs with adaptive control to gradual changes in the FAF trials.

1.2.2 Tone production

Sensorimotor adaptation and aftereffect was also found in tone production. Jones & Munhall (2002, 2005) tested native Mandarin speakers' /ma1/ (with a high level tone, 'mother') and /ma2/ (with a rising tone, 'hemp') and the participants showed evidence of sensorimotor adaptation after being exposed to the FAF. A negative aftereffect was observed for both /ma1/ and /ma2/ in the test phase. Instead of producing compensatory responses as if they were still hearing the pitch shifts (i.e., the (positive) aftereffect), the speakers overcorrected by changing pitch in an opposite direction to the compensatory responses (i.e., the negative aftereffect). This overcorrection made the F0 values of /ma1/ and /ma2/ even higher than the baseline after receiving upward shifts but even lower than the baseline after receiving downward shifts.

However, Feng et al. (2017) showed a different pattern of aftereffect in their native Mandarin speakers' tone production. In downshifting the high level tone (/ma1/), a positive aftereffect was observed as a persistent compensation occurred after the auditory perturbation was removed. In addition to single word production, they tested simultaneous perturbations in an epoch of two or four lexical tones. A significant adaptation to upshifting /ma3/ (with a falling-rising/dipping tone, 'horse') occurred only when /ma2/ was included in the epoch. They thus suggest that adaptation serves to maintain tone contrast in a block of words.

1.2.3 The effect of tonal language experience

In Ning (2018), L2 learners of Mandarin (beginners and advanced learners) and native Mandarin speakers were asked to produce the sustained vowel /a/ and the Mandarin /ma1/. Sensorimotor adaptation was found in all three groups, though the L2 learners compensated more than the native Mandarin speakers. This suggests that the online recalibration is innate and automatic, regardless of Mandarin proficiency. Unlike Jones & Munhall (2002, 2005) who found a **negative** aftereffect in tone production, Ning (2018) discovered a (positive) task-dependent aftereffect across speakers. The aftereffect was present in the downward shift /ma1/, but absent in the sustained /a/ and the upward shift /ma1/. This indicates that the compensation persists for some period of time only when the perturbation generates an obvious deviation from the linguistically meaningful pitch target (i.e., downshifting the high level tone in /ma1/). Overall, a significant group difference was only observed in the degree of compensation, with more compensation in the

L2 learners than in the native Mandarin speakers, which echoes the findings in Ning, Shih & Loucks (2014). The absence of group differences (L2 learners and native Mandarin speakers) in the sensorimotor adaptation and aftereffect suggests that plasticity is involved in everyone's speech motor acts.

1.3 Motivation and hypotheses

The abovementioned review tells us three things: (1) The sensorimotor adaptation is innate and automatic, which is present in all speakers' vocalizations; (2) the degree of compensation depends on vocal training experience and tonal language experience, with more experienced speakers associated with more suppression; (3) the aftereffect is affected by vocal training experience and linguistic meaningfulness of the task. For the last point (i.e., the aftereffect), previous research of vocal training experience has been limited to investigating English-speaking singers and non-singers; previous research of tone production task has been limited to investigating Mandarin-speaking non-singers. This makes us wonder if the combination of the two, vocal and tone, would give us a different pattern in the persistence of compensation. Therefore, the present study recruited Mandarin-speaking vocalists and non-vocalists.

As for the task, both vowel /a/ prolongation and lexical tone production were used. To avoid vocal fatigue, we chose two Mandarin lexical tones, /ma1/ and /ma2/.¹ The reason for choosing /ma1/ and /ma2/ was two-fold: (1) It followed the design in Jones & Munhall (2002, 2005) so that we could do cross-study comparisons; (2) as indicated in Ning, Loucks & Shih (2015), under single perturbation, only /ma1/ and /ma2/ had apparent compensation across groups.

We hypothesized that the sensorimotor adaptation would appear in all speakers and all tasks. Secondly, Mandarin-speaking vocalists having better control over their voice F0 should demonstrate smaller compensation than Mandarin-speaking non-vocalists, particularly in the singing-like /a/ prolongation. Finally, the aftereffect should be task-dependent. Following the results in Ning (2018), we hypothesized that the (positive) aftereffect would be absent in the /a/ prolongation but present in downward shifts of /ma1/ and /ma2/ (i.e., when the downward shifts were in an opposite direction to the intended high-going pitch).

1. It took approximately 8 minutes to complete one session of recording. When we piloted the study with four lexical tones and the vowel /a/, we had 10 sessions in total (5 stimuli × 2 directions). The majority of the participants in the pilot study reported that they had vocal fatigue. Therefore, we had to cut down the length of the experiment.

2. Methods

2.1 Participants

Fifteen vocalists (age: 20–24; 7 males) and fifteen non-vocalists (age: 20–24; 5 males) were recruited. The vocalists were all music majors in universities with a specialization in vocal music. The non-vocalists were not music majors and all indicated in the background questionnaire that they had never received vocal training. The 30 participants are all native speakers of Mandarin and none reported a history of voice disorders. A hearing test using an MAICO pure tone audiometer (model MA 25) was given to all the participants. All passed the hearing test at 20 dB hearing level bilaterally at 250, 500, 750, 1000, 2000, 3000, and 4000 Hz. The participants gave informed consent and received financial compensation for their time.

2.2 Materials and procedures

The experimental apparatus and protocol followed Ning (2018). A written instruction sheet was given to the participants and they were asked to produce the following stimuli: the vowel /a/, the Chinese word /ma1/ ('mother', with a high level tone), and the Chinese word /ma2/ ('hemp', with a rising tone). The vowel /a/ was chosen because it is an unmarked vowel and the articulatory gesture is simple (i.e., jaw opening) with no tongue constriction. The Mandarin syllable /ma/ was selected due to its high frequency and the consistency of unmarked vowel /a/. For all the three tested stimuli, the direction of pitch-shifts can be either upward or downward. This gave rise to 6 (3 stimuli $\times$ 2 directions) recording sessions in total.

Each recording session consisted of three phases: **baseline**, **ramp**, and **test** (see Figure 1). In the **baseline** phase (Trial 1 to Trial 10), participants produced the test stimuli (/a/, /ma1/ or /ma2/; hereafter AH, MA1, and MA2 for short) 10 times while receiving normal auditory feedback. In the **ramp** phase (Trial 11 to Trial 61), auditory feedback was incrementally or decrementally changed at a step of 2 cents. In other words, the pitch perturbation started at 0 cents (Trial 11) and terminated when the shifted pitch was 100 cents above or below the participant's F0 (Trial 61). In each FAF trial (Trial 11 to Trial 61), the auditory feedback was shifted from utterance onset to utterance offset. During the **test** phase (Trial 62 to Trial 81), participants produced 20 more utterances while receiving normal auditory feedback. The intertrial interval was set to be 1 second.

In the production of AH, participants were asked to hold a steady and comfortable pitch for 2 seconds. In the productions of MA1 and MA2, participants

Figure 1. Illustration of the experimental phases

were told to utter the words at a slower pace as if they were teaching students how to pronounce the words. This was to ensure that we could get a clear steady pitch (MA1) and a clear rising pitch (MA2) in their vocalizations. They were told to ignore any changes in the auditory feedback if heard. To reduce the potential influence of bone conducted signal, each participant’s voice was amplified with a McLELLAND MAR-16P Headphone Amplifier and played back with a gain of 10 dB SPL louder.

2.3 Data analysis

The raw signals were imported to MATLAB. Pitch estimation was conducted with the SWIPE algorithm (10 ms interval) (Camacho & Harris 2008, Scheerer & Jones 2012). The F0 values were then converted into cents with the formula in (1):

(1) $\text{cents} = 1200 * \log_2(\text{F0} / \text{baseline F0})$, where the baseline F0 is the mean F0 in the baseline phase.

Following Ning’s (2018) analysis, the trials were grouped into 8 blocks: 1 block of baseline trials (Trial 1–Trial 10), 5 blocks of FAF trials (Trial 11–Trial 21, Trial 22–Trial 31, Trial 32–Trial 41, Trial 42–Trial 51, Trial 52–Trial 61), and 2 blocks

of test trials (Trial 62– Trial 71, Trial 72–Trial 82). The first block of FAF trials had 11 trials instead of 10 because the pitch perturbation started at 0 cents. Two effects were explored: **sensorimotor adaptation** and **aftereffect**. First, **sensorimotor adaptation** denotes adaptive control in motor learning and measures whether speakers update their sensorimotor representations. To investigate the **sensorimotor adaptation**, the five blocks of FAF trials were considered and analyzed with a repeated measures analysis of variance (ANOVA). For AH and MA1, the mean cents of the vocalizations were measured and the ANOVA fitted the linear model of GROUP (vocalists and non-vocalists) $\times$ BLOCK (5 blocks of FAF trials) $\times$ TASK (AH, MA1) $\times$ DIRECTION (upward and downward shifts). For MA2, the min/max cents of the vocalizations were obtained and the ANOVA fitted the linear model of GROUP (vocalists and non-vocalists) $\times$ BLOCK (5 blocks of FAF trials) $\times$ DIRECTION (upward and downward shifts).

Second, the **aftereffect** indicates the persistence of compensation and shows that speakers could respond as if they were hearing pitch-shifts when auditory feedback returned to normal. To examine the **aftereffect**, a repeated measures ANOVA with the same independent variables (GROUP, BLOCK, TASK, and DIRECTION) and dependent variables (mean cents for AH and MA1, but min/max cents for MA2) was carried out for a comparison of the baseline trials (1–10), the first block of test trials (62–71) and the second block of test trials (72–18).

In order to evaluate the effect of DIRECTION in **sensorimotor adaptation** and **aftereffect**, absolute pitch values were used in the above-mentioned ANOVA analyses. For post-hoc analyses, Tukey's honestly significant difference (HSD) test was used with a family-wise alpha level of .05.

3. Results

3.1 Compensation and sensorimotor adaptation

3.1.1 The compensation and adaptation in the productions of AH and MA1

The mean pitch values (in cents) of the FAF trials are depicted in Figure 2. The error bars represent the standard error of the mean. As the pitch values were normalized to the mean of the baseline, 0 cents indicates that no difference from the baseline was observed. In general, both vocalists and non-vocalists showed increasing compensation (i.e., adaptive control) across the blocks. Visual inspection of the graph shows that non-vocalists produced larger compensation than vocalists, irrespective of the TASK.

The ANOVA results show that there were significant main effects of GROUP ($F(1,560)=9.484$, $p<.01$), BLOCK ($F(4,560)=47.835$, $p<.001$), TASK ($F(1,560)=$

Figure 2. The mean F0 values (in cents) in the 5 blocks of FAF trials for the productions of AH and MA1

20.727, $p < .001$), and DIRECTION ($F(1, 560) = 7.904$, $p < .01$). Post-hoc comparisons indicate that non-vocalists had significantly larger compensation (52 cents) than vocalists (42 cents) ($p < .01$). The compensation gradually and significantly rose from the first block to the fifth block of FAF trials (Trials 11–21 < Trials 22–31 < Trials 32–41 < Trials 42–51 < Trials 52–61, $p < .05$). This suggests that speakers kept updating their sensorimotor representations. Post-hoc analyses on the main effect of TASK indicate that larger compensation was elicited by MA1 (54 cents) than by AH (40 cents) ($p < .001$). As for the DIRECTION effect, larger compensation was found in response to downward shifts (51 cents) than to upward shifts (43 cents) ($p < .01$).

There was a significant interaction between GROUP and TASK ($F(1, 560) = 10.102$, $p < .01$). The simple main effect analyses indicate that in general both vocalists and non-vocalists had greater compensation for MA1 than for AH (MA1 > AH; vocalists: $F(1, 298) = 8.498$, $p < .01$; non-vocalists: $F(1, 298) = 7.384$, $p < .01$). The GROUP difference only occurred in the production of AH (non-vocalist > vocalist; $F(1, 298) = 6.362$, $p < .05$), not in the production of MA1 (non-vocalist = vocalist; $F(1, 298) = 2.787$, $p = .096$). This finding supports the second hypothesis addressed in Section 1.3. Vocalists with years of vocal training have better control over their voice F0 and thus are able to suppress compensation

in the steady-pitched /a/ prolongation. No other significant interactions were observed.

3.1.2 The compensation and adaptation in the production of MA2

The min and max pitch values (in cents) in the FAF trials are depicted in Figure 3. The error bars represent the standard error of the mean. Generally speaking, as under the AH and MA1 conditions, both vocalists and non-vocalists showed increasing compensation (adaptive) across the blocks for the production of MA2.

Figure 3. The min and max F0 values (in cents) in the 5 blocks of FAF trials for the production of MA2

For the min cents, which represent the onset of pitch rising in MA2, the ANOVA results show that there were significant main effects of GROUP ($F(1, 280) = 5.313$, $p < .05$), BLOCK ($F(4, 280) = 4.978$, $p < .001$), and DIRECTION ($F(1, 280) = 10.301$, $p < .01$). Post-hoc comparisons indicate that in the rising pitch of MA2, vocalists (63 cents) compensated more at the onset than non-vocalists (50 cents) ($p < .05$). The significant difference between the blocks started from the Trials 42–51, making the last two blocks of FAF trials significantly different from the first three blocks ($p < .01$). As for the DIRECTION effect, the responses to downward shifts (66 cents) were greater than those to upward shifts (47 cents) ($p < .01$). No significant interactions were found for the min cents in the production of MA2.

For the max cents, which indicate the offset of pitch rising in MA2, the ANOVA results showed a significant main effect of BLOCK ($F(4, 280) = 2.678$, $p < .05$). The BLOCK difference was in the comparison of the first block and the last block of the FAF trials ($p < .05$). No other main effects and interactions were observed for the max cents in the production of MA2. Together with the results in the min cents, these findings suggest vocal training and shift direction make a difference at the onset of vocalization rather than at the offset.

3.1.3 Comparison across the three tasks (AH, MA1, and MA2) regarding compensation and adaptation

The max cents of MA2 and the mean cents of AH and MA1 were selected for cross-task comparisons. We did not use the mean cents for MA2 because two rising pitch contours can have the same mean value even though their slopes are entirely divergent. We, instead, followed Jones & Munhall (2005), using the max cents for MA2 to compare the mean cents of AH and MA1. As the max cents represent the ultimate pitch height in the rising tone, it is comparable to the target pitch height in the level tone and steady vowel.

The same set of variables (GROUP, BLOCK, TASK, and DIRECTION) was fitted as the independent variables in the repeated measures ANOVA. The ANOVA results show that there were significant main effects of GROUP ($F(1, 840) = 3.856$, $p < .05$), BLOCK ($F(4, 840) = 29.473$, $p < .001$), and TASK ($F(2, 840) = 88.290$, $p < .001$). Post-hoc comparisons indicate that the non-vocalists in general had greater compensation (65 cents) than vocalists (58 cents) across the tasks ($p < .05$). Similar to as found in Sections 3.1.1 and 3.1.2, the compensation gradually and significantly rose from the first block to the fifth block of FAF trials (Trials 11–21 < Trials 22–31 < Trials 32–41 < Trials 42–51 < Trials 52–61, $p < .05$). As for the TASK effect, the greatest compensation was elicited by MA2 (91 cents), followed by MA1 (54 cents) and then AH (40 cents) ($p < .01$). No other main effect or interactions were observed.

3.2 Aftereffect

3.2.1 The aftereffect in the productions of AH and MA1

The mean pitch values (in cents) in the baseline trials and test trials are depicted in Figure 4. The error bars represent the standard error of the mean. The aftereffect is manifested in the significant difference between the test trials and the baseline trials. Figure 4 shows that the aftereffect seems to appear in the downward shift MA1 condition for all participants, and in the non-vocalists' AH production.

The ANOVA results show that there were significant main effects of GROUP ($F(1, 336) = 7.276$, $p < .01$), BLOCK ($F(2, 336) = 55.434$, $p < .001$), TASK ($F(1, 336) =$

Figure 4. The mean F0 values (in cents) in the baseline trials, the first block of test trials, and the second block of test trials for the production of AH and MA1

8.727, $p < .01$), and DIRECTION ($F(1, 336) = 5.976$, $p < .05$). Post-hoc comparisons for the factor GROUP show that the aftereffect was more apparent in non-vocalists (36 cents) than in vocalists (25 cents) ($p < .01$). In general, there was no significant difference between the first block of test trials (43 cents) and the second block of test trials (48 cents) ($p = .511$). The only BLOCK difference was in the comparison of test trials and baseline trials ($p < .001$). As for the TASK, the aftereffect was more apparent for MA1 (36 cents) than for AH (24 cents) ($p < .01$). As for the DIRECTION, a more apparent aftereffect was found in downward shifts (35 cents) than in upward shifts (25 cents) ($p < .05$).

There was a significant interaction between GROUP and TASK ($F(1, 336) = 25.278$, $p < .001$). The simple main effect analyses indicate that the TASK effect appeared in vocalists (more apparent aftereffect in MA1 than in AH; $F(1, 178) = 29.57$, $p < .001$) but not in non-vocalists (MA1=AH; $F(1, 178) = 1.322$, $p = .252$). The GROUP effect occurred in the production of AH (a more apparent aftereffect in non-vocalists than in vocalists; $F(1, 178) = 32$, $p < .001$), but not in the production of MA1 (vocalists=non-vocalists; $F(1, 178) = 1.533$, $p = .217$). No other significant interactions were observed.

3.2.2 The aftereffect in the production of MA2

The min and max pitch values (in cents) in the baseline trials and test trials are depicted in Figure 5. The error bars represent the standard error of the mean. Figure 5 shows that the aftereffect in the production of MA2 seems to appear in the vocalists’ production and in the non-vocalists’ responses to downward shifts.

Figure 5. The min and max F0 values (in cents) in the baseline trials, the first block of test trials, and the second block of test trials for the production of MA2

For the min cents, the ANOVA results show that there were significant main effects of BLOCK ($F(2,168)=51.005, p<.001$) and DIRECTION ($F(1,168)=23.848, p<.001$). The interaction between BLOCK and DIRECTION was also significant ($F(2,168)=7.270, p<.001$). Post-hoc analyses indicate that the differences between the test trials and the baseline trials were in response to both downward ($F(2,87)=35.1, p<.001$) and upward shifts ($F(2,87)=17.9, p<.001$). However, the DIRECTION effect, i.e., a more apparent aftereffect in downward shifts than in upward shifts, only occurred in the two blocks of test trials (Test Phase 1: $F(1,58)=21.8, p<.001$; Test Phase 2: $F(1,58)=7.374, p<.01$). No other significant main effect and interactions were found.

For the max cents, the ANOVA results show that there was a significant main effect of BLOCK ($F(2,168)=41.699, p<.001$), with a considerable difference

between the test trials and the baseline trials ($p < .001$). No other significant main effects and interactions were found.

3.2.3 Comparison across the three tasks (AH, MA₁, and MA₂) in the aftereffect

Following the same logic stated in Section 3.1.3, we compared the max cents of MA₂ with the mean cents of AH and MA₁ in the aftereffect. The ANOVA results show that there were significant main effects of GROUP ($F(1, 504) = 4.832$, $p < .05$), BLOCK ($F(2, 504) = 94.554$, $p < .001$), and TASK ($F(2, 504) = 31.807$, $p < .001$). The interaction between GROUP and TASK was also significant ($F(2, 504) = 8.037$, $p < .001$). The simple main effects analysis show that for non-vocalists, the aftereffect was most apparent in MA₂, then in AH and least in MA₁ (MA₂ > AH > MA₁ in non-vocalists, $F(2, 267) = 8.444$, $p < .001$). For the vocalists, the aftereffect was most apparent in MA₂, then in MA₁ and least in AH (MA₂ > MA₁ > AH in vocalists, $F(2, 267) = 21.690$, $p < .001$). Post-hoc analyses of the BLOCK effect indicate that the pitch values were significantly larger in the test phase than in the baseline phase while no distinction between the two test phases was observed (Test Phase 1 = Test Phase 2 > Baseline, $p < .01$).

4. Discussion

4.1 Sensorimotor adaptation

This study investigated whether vocalists and non-vocalists adapt to a continuous change in perceived pitch and whether a perturbation effect persists after auditory feedback returns to normal. Our results of the significant block effect in the three stimuli show that all the participants were able to adapt to the incremental or decremental change in perceived pitch by updating their audio-motor representations. The first hypothesis, in which the **sensorimotor adaptation** is an innate and automatic process, has thus been supported. The trial-by-trial adaptation data (see Figures 6–7) suggest that the produced pitch was initiated at a similar F₀ to the one in the previous trial.

Figure 6. The mean F0 values (in cents) in each trial for the production of AH and MA1

Figure 7. The min and max F0 values (in cents) in each trial for the production of MA2

4.2 Compensation

Echoing the second hypothesis, the degree of compensation differed across groups and tasks. Significantly smaller compensation was observed in the vocalists' AH, suggesting that vocal training may have an effect on the singing-like production (sustaining /a/ for 2 seconds). The results in the current study and Ning (2018) suggest that experience matters in the compensation of AH. Speakers with vocal training or tone learning experience had a smaller (or can suppress) compensation than their counterparts (vocalists (42 cents) < non-vocalists (52 cents) in the current study; native Mandarin speakers (56 cents) < L2 advanced learners (61 cents) < L2 beginners (70 cents) in Ning 2018).

The vocal training effect observed in the AH production during the present study also corresponds to the results in Jones & Keough (2008) and Keough & Jones (2009), who found that non-singers had difficulty in finding the correct pitch target and thus were more susceptible to pitch perturbation. The difficulty of finding the correct pitch was also discovered in our non-vocalists' AH. As can be seen in the trial-by-trial adaptation process in Figure 6 (the upper panel), while the vocalists had a clear self-defined pitch target for AH and updated their audio-motor representations in a linear fashion, the non-vocalists seemed to have unfixed pitch targets for AH and thus had a non-linear adaptation. Even though the participants were not told to vocalize AH at an exact pitch value, the vocalists were more capable of finding the intended and consistent pitch.

The questions then remain why there was no significant group difference in the compensation of MA1 (vocalists = non-vocalists) and why there was an opposite group pattern in the compensation of MA2 (vocalists > non-vocalists at the onset of the vocalization). It is apparent that the three stimuli involve different audio-motor representations. Even though MA1 and AH share the characteristic of a high steady pitch, the participants were instructed to produce MA1 as a Chinese word but AH as a simple vowel. The speech motor commands for the high level tone (MA1) are entrenched in native Mandarin speakers regardless of vocal training background. Therefore, both non-vocalists and vocalists demonstrated a similar degree of compensation and updated the audio-motor representations in a linear fashion (see the lower panel of Figure 6). In addition to the lexical difference between MA1 and AH, the two stimuli were produced over different durations (2 seconds for AH but unspecified for MA1). It is likely that duration had an effect on the degree of compensation. To account for or rule out this possibility, future research may have to examine AH and MA1 produced over the same duration.

As for the rising tone in MA2, vocalists compensated more than non-vocalists did at the onset of the pitch rising (i.e., the min cents). However, the trial-by-trial

adaptation (see the upper panel of Figure 7) shows that the vocalists started the compensation in the later blocks of FAF trials whereas the non-vocalists' compensation started earlier (from the second block of FAF trials). It is unclear why the vocalists had greater compensation in the later blocks of FAF trials. Nonetheless, the trial-by-trial process still suggests that in response to smaller shifts (in the first two blocks of FAF trials), the vocalists were still less affected than the non-vocalists.

Two more interesting patterns in the degree of compensation are worthy of discussion. First, in general, greater compensation was elicited by MA2, followed by MA1 and then AH ($MA2 > MA1 > AH$). As mentioned above, the participants produced MA2 and MA1 as Chinese words but AH as a simple vowel. This task difference is consistent with previous studies in which larger compensation was observed in speech than in sustained vowels (Burnett et al. 1998, Hain et al. 2000, Larson et al. 2001, Natke, Donath & Kalveram 2003). Second, in general, greater compensation was seen in downward shifts than in upward shifts. This direction effect (down > up) in the degree of compensation also appeared in previous studies in which sustained vowels were used (Chen et al. 2007, Liu et al. 2011, Behroozmand et al. 2013). The true cause for the direction effect is still unclear. However, we speculate that the large compensation in the downward shifts comes from the fact that downward shifts generate an obvious deviation from the intended high and steady pitch. The huge mismatch between the heard pitch and the intended high pitch seems to give rise to the large corrective response.

4.3 Aftereffect

As for the aftereffect (i.e., how long compensation persists), the patterns were more complicated but in general were consistent with our third hypothesis (i.e., the aftereffect is task-dependent). More precisely, the results show that the aftereffect was dependent on both tasks and participants ($MA2 > AH > MA1$ in non-vocalists, but $MA2 > MA1 > AH$ in vocalists). In the production of AH, non-vocalists, but not the vocalists, had an apparent aftereffect for both downward shifts and upward shifts (see Figure 4). This suggests that the non-vocalists continued the compensation after the pitch perturbation was removed, whereas the vocalists returned to normal without any delay. The absence of the aftereffect in the vocalists may suggest that they have more precise motor commands and predictions of self-produced pitch.

In the production of MA1, both the vocalists and non-vocalists showed an aftereffect, but only to the downward shift MA1 (see Figure 4). In the production of MA2, the non-vocalists showed an aftereffect in the downward shifts as they

did in MA1, whereas the vocalists had the aftereffect in both upward and downward shifts. This suggests that the rising pitch was manipulated differently by the vocalists and non-vocalists.

Take a further look at the aftereffect in MA2. The vocalists produced different slopes of MA2 for upward shifts and downward shifts in the test phase. As schematized in Figure 8, when MA2 was shifted downwards, the vocalists, like the non-vocalists, uplifted the entire trajectory in the test phase, which is a typical compensatory response. They all responded as if they were still hearing pitch perturbation, an evidence of aftereffect. However, when MA2 was shifted upwards, vocalists made the slope level out instead of pulling down the entire trajectory. There was a negative aftereffect at the onset of the upward shift MA2. It is likely that the vocalists may be aware that they, for some reason, had made more compensation at the onset of the vocalization in the FAF trials, and therefore they overshot their production when perturbation was removed. This overshoot is very like the overcorrection observed in Jones & Munhall's (2002, 2005) /ma1/ and /ma2/; however, it only appears in the vocalists' onset of MA2 in the current study. We speculate that the reason for the negative aftereffect in the upward shift MA2 is to maintain tone contrast, echoing the claim in Feng et al. (2017). If the vocalists corrected for the 'more' compensation by pulling down the entire pitch trajectory in the test phase of upward shift MA2, they would have lost the contrast between Tone 2 and Tone 3 (the rising part). To avoid tone confusion, the vocalists therefore overshot the production at the onset of MA2.

Figure 8. The schematized aftereffect in MA2 for the non-vocalists and the vocalists

Finally, there was a clear direction effect (down > up) in the aftereffect for all MA1's and the non-vocalists' MA2. However, the direction effect did not appear in the production of AH and in the vocalists' MA2. Ning (2018) also obtained the same finding, where all participants showed a significant aftereffect in the downward shift MA1, but not in the upward shift MA1. The downward shifts made the perceived pitch deviate considerably from the intended high or rising pitch. Thus, a greater effort was made to compensate for the error and the effect continued even when perturbation was no longer present. As for the absence of aftereffect in the upward shifts for all MA1's and the non-vocalists' MA2, one possible explanation is that the speakers started at a fairly low fundamental frequency, and it therefore became challenging for them to lower F0 any further when facing upward shifts. However, our data show that the participants initiated their F0 within roughly the same range for downward shifts (standard deviation: ± 3.55 Hz) and upward shifts (standard deviation: ± 3.88 Hz). That means they did not change their initial pitch level over the duration of the experiment. Then, it is likely that the absence of aftereffect in the upward shifts is because upshifting the pitch corresponds to the goal of maintaining or reaching a high pitch in tone production. This goal-matching characteristic could make the speakers unaware of the perturbation and thus no further correction was made when it was removed.

5. Conclusion

In general, it seems that both vocalists and non-vocalists have the ability to remap audio-motor representations within a short period of perturbation. However, the degree of compensation and sensorimotor adaptation depends on vocal training, with vocalists compensating less than non-vocalists. The aftereffect is stimulus-specific and dependent on vocal training, present in the lexical tone productions when perturbed pitch goes in an opposite direction to the intended pitch trajectory or in sustained vowels when speakers (i.e., non-vocalists) have no entrenched feedforward commands for pitch regulation.

Acknowledgements

The author gratefully acknowledges the assistance of Chien-Min Kuo and Jian Leat Siah in the data collection. This research was supported by the Ministry of Science and Technology, Taiwan (Grant No. MOST 105-2410-H-003-001).

References

- Behroozmand, Roozbeh, Nadine Ibrahim, Oleg Korzyukov, Donald A. Robin, and Charles R. Larson. 2013. Left-hemisphere activation is associated with enhanced vocal pitch error detection in musicians with absolute pitch. *Brain and Cognition* 84.1:97–108. <https://doi.org/10.1016/j.bandc.2013.11.007>
- Burnett, Theresa A., Marcia B. Freeland, Charles R. Larson, and Timothy C. Hain. 1998. Voice F0 responses to manipulations in pitch feedback. *The Journal of the Acoustical Society of America* 103.6:3153–3161. <https://doi.org/10.1121/1.423073>
- Camacho, Arturo, and John G. Harris. 2008. A sawtooth waveform inspired pitch estimator for speech and music. *The Journal of the Acoustical Society of America* 124.3:1638–1652. <https://doi.org/10.1121/1.2951592>
- Chen, Stephanie H., Hanjun Liu, Yi Xu, and Charles R. Larson. 2007. Voice F0 responses to pitch-shifted voice feedback during English speech. *The Journal of the Acoustical Society of America* 121.2:1157–1163. <https://doi.org/10.1121/1.2404624>
- Chon, Hee Cheong, Shelly Jo Kraft, Jingfei Zhang, Torrey Loucks, and Nicoline G. Ambrose. 2013. Individual variability in delayed auditory feedback effects on speech fluency and rate in normally fluent adults. *Journal of Speech, Language, and Hearing Research* 56.2:489–504. [https://doi.org/10.1044/1092-4388\(2012/11-0303\)](https://doi.org/10.1044/1092-4388(2012/11-0303))
- Cowie, Roddy, and Ellen Douglas-Cowie. 1992. *Postlingually Acquired Deafness*. New York & Berlin: Mouton de Gruyter. <https://doi.org/10.1515/9783110869125>
- Feng, Yongqiang, Yan Xiao, Yonghong Yan, and Ludo Max. 2017. Adaptation in Mandarin tone production with pitch-shifted auditory feedback: Influence of tonal contrast requirements. *Language, Cognition and Neuroscience* 33.6:734–749. <https://doi.org/10.1080/23273798.2017.1421317>
- Hain, Timothy C., Theresa A. Burnett, Swathi Kiran, Charles R. Larson, Shajila Singh, and Mary K. Kenney. 2000. Instructing subjects to make a voluntary response reveals the presence of two components to the audio-vocal reflex. *Experimental Brain Research* 130.2:133–141. <https://doi.org/10.1007/s002219900237>
- Houde, John F., and Michael I. Jordan. 1998. Sensorimotor adaptation in speech production. *Science* 279.5354:1213–1216. <https://doi.org/10.1126/science.279.5354.1213>
- Houde, John F., and Michael I. Jordan. 2002. Sensorimotor adaptation of speech I: Compensation and adaptation. *Journal of Speech, Language, and Hearing Research* 45.2:295–310. [https://doi.org/10.1044/1092-4388\(2002/023\)](https://doi.org/10.1044/1092-4388(2002/023))
- Jones, Jeffery A., and Dwayne Keough. 2008. Auditory-motor mapping for pitch control in singers and nonsingers. *Experimental Brain Research* 190.3:279–287. <https://doi.org/10.1007/s00221-008-1473-y>
- Jones, Jeffery A., and Kevin G. Munhall. 2002. The role of auditory feedback during phonation: Studies of Mandarin tone production. *Journal of Phonetics* 30.3:303–320. <https://doi.org/10.1006/jpho.2001.0160>
- Jones, Jeffery A., and Kevin G. Munhall. 2005. Remapping auditory-motor representations in voice production. *Current Biology* 15.19:1768–1772. <https://doi.org/10.1016/j.cub.2005.08.063>
- Keough, Dwayne, Colin Hawco, and Jeffery A. Jones. 2013. Auditory-motor adaptation to frequency-altered auditory feedback occurs when participants ignore feedback. *BMC Neuroscience* 14, article no. 25. <https://doi.org/10.1186/1471-2202-14-25>

- Keough, Dwayne, and Jeffery A. Jones. 2009. The sensitivity of auditory-motor representations to subtle changes in auditory feedback while singing. *The Journal of the Acoustical Society of America* 126.2:837–846. <https://doi.org/10.1121/1.3158600>
- Lane, Harlan, and Bernard Tranel. 1971. The Lombard sign and the role of hearing in speech. *Journal of Speech and Hearing Research* 14.4:677–709. <https://doi.org/10.1044/jshr.1404.677>
- Lane, Harlan, and Jane Wozniak Webster. 1991. Speech deterioration in postlingually deafened adults. *The Journal of the Acoustical Society of America* 89.2:859–866. <https://doi.org/10.1121/1.1894647>
- Larson, Charles R., Theresa A. Burnett, Jay J. Bauer, Swathi Kiran, and Timothy C. Hain. 2001. Comparison of voice F0 responses to pitch-shift onset and offset conditions. *The Journal of the Acoustical Society of America* 110.6:2845–2848. <https://doi.org/10.1121/1.1417527>
- Liu, Hanjun, Emily Q. Wang, Zhaocong Chen, Peng Liu, Charles R. Larson, and Dongfeng Huang. 2010. Effects of tonal native language on voice fundamental frequency responses to pitch feedback perturbations during sustained vocalizations. *The Journal of the Acoustical Society of America* 128.6:3739–3746. <https://doi.org/10.1121/1.3500675>
- Liu, Hanjun, Michelle Meshman, Roozbeh Behroozmand, and Charles. R. Larson. 2011. Differential effects of perturbation direction and magnitude on the neural processing of voice pitch feedback. *Clinical Neurophysiology* 122.5:951–957. <https://doi.org/10.1016/j.clinph.2010.08.010>
- Max, Ludo, Marie E. Wallace, and Irena Vincent. 2003. Sensorimotor adaptation to auditory perturbations during speech: Acoustic and kinematic experiments. *Proceedings of the 15th International Congress of Phonetic Sciences*, ed. by M. J. Solé, D. Recasens and J. Romero, 1053–1056. Barcelona, Spain: ICPhS Archive.
- Mollaei, Fatemeh, Douglas M. Shiller, Shari R. Baum, and Vincent L. Gracco. 2016. Sensorimotor control of vocal pitch and formant frequencies in Parkinson's disease. *Brain Research* 1646:269–277. <https://doi.org/10.1016/j.brainres.2016.06.013>
- Natke, Ulrich, Thomas M. Donath, and Karl Th. Kalveram. 2003. Control of voice fundamental frequency in speaking versus singing. *The Journal of the Acoustical Society of America* 113.3:1587–1593. <https://doi.org/10.1121/1.1543928>
- Ning, Li-Hsin. 2018. Vocal adaptation in perturbed auditory feedback for L2 learners of Mandarin. *Journal of Chinese Language Teaching* 15.3:1–20.
- Ning, Li-Hsin, Chilin Shih, and Torrey M. Loucks. 2014. Mandarin tone learning in L2 adults: A test of perceptual and sensorimotor contributions. *Speech Communication* 63–64:55–69. <https://doi.org/10.1016/j.specom.2014.05.001>
- Ning, Li-Hsin, Torrey M. Loucks, and Chilin Shih. 2015. The effects of language learning and vocal training on sensorimotor control of lexical tone. *Journal of Phonetics* 51:50–69. <https://doi.org/10.1016/j.wocn.2014.12.003>
- Oller, D. Kimbrough, and Rebecca E. Eilers. 1988. The role of audition in infant babbling. *Child Development* 59.2:441–449. <https://doi.org/10.2307/1130323>
- Parrella, Benjamin, and John Houde. 2019. Modeling the role of sensory feedback in speech motor control and learning. *Journal of Speech, Language, and Hearing Research* 62.8S:2963–2985. https://doi.org/10.1044/2019_JSLHR-S-CSMC7-18-0127
- Purcell, David W., and Kevin G. Munhall. 2006. Adaptive control of vowel formant frequency: Evidence from real-time formant manipulation. *The Journal of the Acoustical Society of America* 120.2:966–977. <https://doi.org/10.1121/1.2217714>

- Scheerer, Nichole E., and Jeffery A. Jones. 2012. The relationship between vocal accuracy and variability to the level of compensation to altered auditory feedback. *Neuroscience Letters* 529.2:128–132. <https://doi.org/10.1016/j.neulet.2012.09.012>
- Sparks, Garen, Dorothy E. Grant, Kathleen Millay, Delaina Walker-Baston, and Linda S. Hynan. 2002. The effect of fast speech rate on stuttering frequency during delayed auditory feedback. *Journal of Fluency Disorders* 27.3:187–201. [https://doi.org/10.1016/S0094-730X\(02\)00128-6](https://doi.org/10.1016/S0094-730X(02)00128-6)
- Terband, Hayo, Frits van Brenk, and Anniek van Doornik-van der Zee. 2014. Auditory feedback perturbation in children with developmental speech sound disorders. *Journal of Communication Disorders* 51:64–77. <https://doi.org/10.1016/j.jcomdis.2014.06.009>
- Van Borsel, John, Gert Reunes, and Nathalie Van den Bergh. 2003. Delayed auditory feedback in the treatment of stuttering: Clients as consumers. *International Journal of Language & Communication Disorders* 38.2:119–129. <https://doi.org/10.1080/1368282021000042902>
- Waldstein, Robin S. 1990. Effects of postlingual deafness on speech production: Implications for the role of auditory feedback. *The Journal of the Acoustical Society of America* 88.5:2099–2114. <https://doi.org/10.1121/1.400107>
- Zarate, Jean Mary, and Robert J. Zatorre. 2005. Neural substrates governing audiovocal integration for vocal pitch regulation in singing. *Annals of the New York Academy of Sciences* 1060.1:404–408. <https://doi.org/10.1196/annals.1360.058>
- Zarate, Jean Mary, and Robert J. Zatorre. 2008. Experience-dependent neural substrates involved in vocal pitch regulation during singing. *NeuroImage* 40.4:1871–1887. <https://doi.org/10.1016/j.neuroimage.2008.01.026>

Address for correspondence

Li-Hsin Ning
 Department of English
 National Taiwan Normal University
 Taipei, TAIWAN
lihsin@ntnu.edu.tw

Publication history

Date received: 20 July 2019
 Date revised: 22 September 2019
 Date accepted: 18 December 2019