Pitch-shift Responses to Auditory Perturbation Involve a Mix of Opposing and Following Responses: Evidence from Steady Vowel and Glissando Vocalizations

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

ABSTRACT

In this paper, we investigate whether opposing responses and following responses are equally likely to appear in the pitch-shift responses for steady vowels and glissando vocalizations. Two groups of participants were recruited: seven musicians and seven non-musicians. The results show that following responses were not a minority. Musicians and non-musicians produced a mix of opposing and following responses (almost equally likely) for steady vowels and glissando vocalizations. This result suggests that both mechanisms should be activated simultaneously in online voice pitch regulation.

Keywords: audio-vocal control, pitch-shift paradigm, opposing response, following response, glissando

1. INTRODUCTION

Auditory feedback plays an essential role in speech production. When unexpected changes occur in auditory feedback, speakers would make opposing (also called compensatory) responses that go in the opposite direction from the unexpected shifts to reduce the perceived errors. Vocal compensation can appear in steady vowel production [1], word production [2, 3], glissando vocalizations [4], and singing [5]. However, research has shown that opposing responses may not be the majority in the responses to auditory perturbation and that following responses (i.e., responses that go in the same direction as the direction of unexpected shifts) can occupy a considerable proportion of steady vowels and tone word (level tone) productions if we examine the pitch contours on a trial basis [6-8]. It is unclear whether both mechanisms (opposing and following) would be equally likely to appear in glissando vocalizations in which pitch trajectories are not stable. To investigate glissando vocalizations, it would also be interesting to explore how musicians and non-musicians utilize opposing and following mechanisms. If musicians are less affected by pitch perturbation and have reduced compensation [9], we would expect that they would disfavor following responses that generate a larger deviation from the intended pitch.

2. METHOD

2.1. Participants

Fourteen native speakers of Mandarin (20 ~ 30 years old; $M = 23$, $SD = 3.2$) were recruited for this study. Half of them were music majors at the time of the experiment. They played piano, cello, violin, flute, or tuba. The other half were non-musicians and had not played music as an amateur in the previous five years. All the participants reported no history of speech or hearing disorders. They passed a hearing screening test (20 dB bilaterally at 250, 500, 750, 1000, 2000, 3000, and 4000 Hz) prior to recording using a MAICO pure-tone audiometer (model MA 25).

2.2. Materials and Procedures

The experiment consisted of three production blocks: upward glissando, downward glissando, and steady vowel. In the upward or downward glissando, participants first listened to a model note, which started with a steady note for 500 ms, was transitioned into an upward or downward glide (100 cents/half second for 2 seconds), and then was held steady again for 500 ms. Participants were instructed to mimic the glissando pattern (without vibrato) using the vowel /a/ within their comfortable pitch range. In the steady-vowel condition, participants were asked to produce a sustained vowel /a/ (for roughly 3 seconds) after hearing the beep sound. Each recording block contained 30 repetitions, for a total of 90 trials (30 repetitions × 3 blocks = 90 trials) per participant.

In each vocalization, participants may have heard a single pitch-shift stimulus presented in the auditory feedback. The voice f_0 was shifted upward, shifted downward, or not shifted (as a control), which were equally likely to occur randomly in a recording block. The pitch-shift stimuli were all 200 ms long and fixed at ±100 cents. They appeared at a random time, 500 ~ 700 ms, after vocal onset. The intertrial delay was 1,000 ms. Participants were asked to *ignore* pitch-shift stimuli and maintain the intended pitch contour. The order of the three recording blocks was counterbalanced across participants.

2.3. Apparatus

The participants sat in a soundproof booth and wore AKG K240 headphones. We placed a standalone microphone, Audio Tech ATR20, in front of them. The voice signal from the microphone was processed (shifted in pitch) in near-real time (14 ~ 20 ms delay) through an Eventide Ultra-Harmonizer (model H7600), which was controlled by Max/MSP (v.8, Cycling 74). To mask bone-conducted auditory feedback, we amplified the voice signal's intensity with a 10-dB gain via a McLELLAND MAR-16P headphone amplifier when presented over the headphones. The vocalizations, pitch-shifted signals, and TTL pulses indicating the onset of pitch-shift stimuli were recorded using a WinDaq DI-720 acquisition device and sampled at 8 kHz per channel in WinDaq Pro.

2.4. Data Preprocessing

The signals recorded in WinDaq were imported into MATLAB (R2020a) and sorted based on the direction of pitch-shift stimuli (upward shift, downward shift, or no shift). Each vocalization was segmented into a 1.2-s long signal that included a 200-ms preshift period, a 200-ms shift period, and an 800-ms postshift period. We then wrote the segmented signals as sound files and estimated the pitch every 10 ms in PRAAT. The pitch values were transformed into cents using the formula cents=1200×log2(f_0/baseline), where the baseline indicated the mean pitch of the preshift period. The f_0 records were then imported back into MATLAB for further processing.

Each segmented vocalization was classified as one of the following four types by comparing it with the averaged pitch contour of the corresponding control trials: opposing response, following response, non-response, or error. If the response changed in the opposite direction to the pitch-shift stimulus, it was tagged as an "opposing" response. If the response changed in the same direction as the pitch-shift stimulus, it was tagged as a "following" response. If the response did not show a clear upward or downward deviation from the averaged control (i.e., less than ±2 SD of the preshift mean), it was tagged as a "non-response." If the response showed an erroneous pitch-tracking result, it was tagged as an "error."

After the classification, we calculated the percentage of opposing responses, following responses, non-responses, and errors under each condition (3 production blocks × 2 stimulus directions) for each participant. Difference waves were obtained for the upward glissando and downward glissando by subtracting the averaged pitch contour of the corresponding control trials from the averaged opposing or following pitch contours at every point (see Fig. 1). We conducted statistical analyses using generalized additive mixed models of the difference waves (cent values) for the upward and downward glissando conditions and the standardized pitch contours (cent values) for the steady-vowel condition.

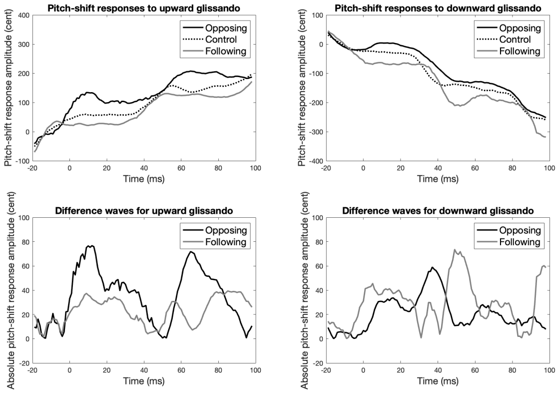

Figure 1: Illustration of averaged opposing responses (dark solid lines), averaged following responses (grey solid lines), and their difference waves (obtained by subtracting the averaged controls displayed in the dotted lines) for one participant in the upward glissando (the left panel) and downward glissando (the right panel).

3. RESULTS

3.1. Proportions of the opposing and following responses

Table 1 displays the mean percentages of opposing responses, following responses, nonresponses, and errors by musicality, production block, and pitch-shift stimulus direction. The results show that our participants produced a mix of opposing and following responses in glissando and steady vowel vocalizations. For musicians, *opposing* responses were more common than following responses in the steady vowel vocalizations whereas opposing and following responses were equally likely to appear in the glissando vocalizations. However, for non-musicians, *following* responses were more common than opposing responses when the pitch-shift stimuli went in an opposite direction from the intended overall pitch trajectories, such as downshift stimuli in the upward glissando (55%) and upshift stimuli in the downward glissando (47%). These findings suggest that following responses, as one response type in the face of auditory perturbation, were not neglectable, which implies that both mechanisms (opposing and following) could be activated simultaneously when we encounter unexpected perturbation.

MUSICALITY	Production block	Pitch-shift stimulus direction	Opposing response	Following response	Non-response	Error
Musician	Glissando up	Upshift	44%	42%	0%	14%
		Downshift	34%	47%	4%	14%
	Glissando down	Upshift	55%	31%	5%	9%
		Downshift	44%	43%	5%	8%
	Steady vowel	Upshift	65%	34%	0%	1%
		Downshift	62%	37%	1%	0%
Non-musician	Glissando up	Upshift	65%	18%	4%	13%
		Downshift	34%	55%	4%	7%
	Glissando down	Upshift	27%	47%	10%	16%
		Downshift	65%	23%	2%	10%
	Steady vowel	Upshift	55%	41%	2%	2%
		Downshift	39%	55%	1%	5%

Table 1: The mean percentages of opposing responses, following responses, nonresponses, and errors by musicality, production block, and pitch-shift stimulus direction.

3.2. Time-varying changes in pitch-shift responses

We used generalized additive mixed models (GAMMs) to account for the time-varying changes in pitch-shift responses. The `bam()` function in the `mgcv` package of R [10]. All the GAMMs involved a factor smooth that accounted for the nonlinear random effects (intercepts and slopes) derived from individual variability across time. This method fits a wiggly curve for each participant, using the same calibration for penalization across all participants. For details of how to create baseline models, conduct model comparisons, and set up parametric predictors (musicality, production block, stimulus direction), please refer to Sun and Shih [11] and Ning [8]. In what follows, we presented two sets of model fitting results: one for *opposing responses* and the other for *following responses*.

3.2.1. Opposing responses

Fig. 2 depicts musicality's effect on the time-varying opposing responses. The results show that non-musicians had *larger* opposing responses than musicians (constant difference) in the upshift of downward glissando (t=2.307, p<.05), the downshift of upward glissando (t=2.278, p<.05), the upshift of upward glissando (t=2.218, p<.05), and the upshift of the steady-vowel (t=–1.899, p<.05) condition. We observed the dynamic changes over time in the upshift of the steady-vowel (F(6.568, 7.694)=3.278, p<.01) condition, in which the divergence between musicians and non-musicians increased significantly from vocal onset to vocal offset.

Fig. 3 depicts the production block effect on the opposing responses. We selected the steady-vowel condition as the reference level. The results demonstrate that opposing responses in the glissando conditions were constantly *larger* than those in the steady-vowel condition. Significant constant height differences appeared in almost all combinations of musicality and stimulus direction: musicians' downshift in downward glissando (t=3.610, p<.001), musicians' upshift in downward glissando (t=0.995, p=.320), musicians' downshift in upward glissando (t=2.733, p<.01), musicians' upshift in upward glissando (t=2.085, p<.05), non-musicians' downshift in downward glissando (t=4.088, p<.001), non-musicians' upshift in downward glissando (t=2.998, p<.01), non-musicians' downshift in upward glissando (t=2.760, p<.01), and non-musicians' upshift in upward glissando (t=2.852, p<.01).

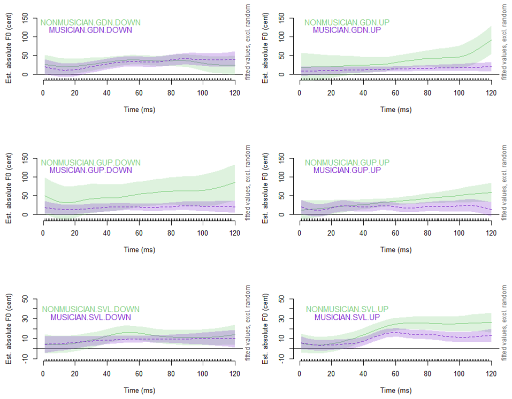

Figure 2: Musicality effect (musicians vs. non-musicians) regarding *opposing responses* at each level of the combination of production block (GDN, GUP, SVL) and stimulus direction (DOWN, UP). The purple curves represent musicians' opposing responses (in absolute values), and the green curves represent non-musicians' opposing responses (in absolute values). The pitch-shift stimuli appeared at 20 ~ 40 ms on the x-axis.

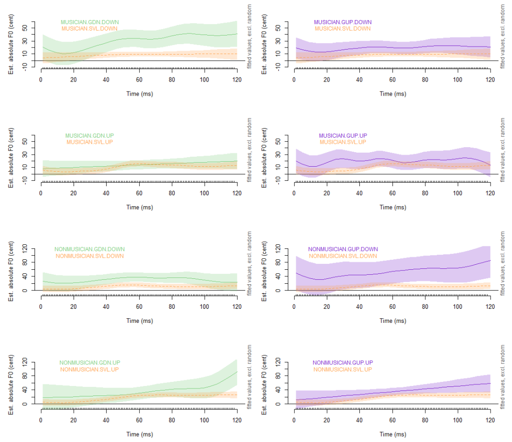

Figure 3: Production block effect (upward glissando, downward glissando, steady vowel) regarding *opposing responses* at each level of the combination of musicality (MUSICIAN, NONMUSICIAN) and stimulus direction (DOWN, UP). The purple curves represent opposing responses (in absolute values) for upward glissando; the green curves represent opposing responses (in absolute values) for downward glissando; the orange curves represent opposing responses (in absolute values) for the steady-vowel condition.

Fig. 4 depicts stimulus direction's effect on the opposing responses. Generally, opposing responses to downshifts were *larger* than opposing responses to upshifts in the glissando conditions whereas we found the opposite pattern (downshifts *smaller* than upshifts) in the steady-vowel condition.

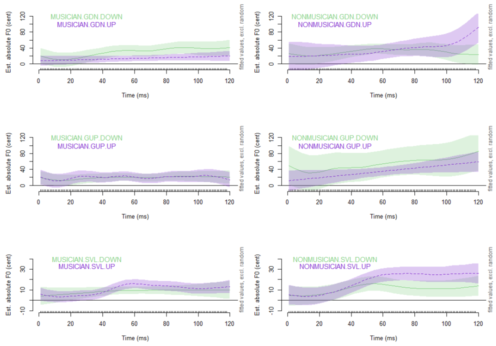

Figure 4: Stimulus direction effect (upshift vs. downshift) regarding *opposing responses* at each level of the combination of musicality (MUSICIAN, NONMUSICIAN) and production block (GDN, GUP, SVL). The purple curves represent opposing responses (in absolute values) for upshift stimuli; the green curves represent opposing responses (in absolute values) for downshift stimuli.

3.2.2. Following responses

Similar to opposing responses, non-musicians also demonstrated *larger* following responses than musicians (see Fig. 5). However, this musicality effect was only significant in the glissando vocalizations: the upshift of downward glissando (t=2.307, p<.05), the downshift of downward glissando (t=2.218, p<.05), and the downshift of upward glissando (t=2.278, p<.05).

Fig. 6 displays the production block effect on the following responses. As in the opposing responses, following responses in the glissando conditions were *larger* than those in the steady-vowel condition (all p<.05).

Finally, for the stimulus direction effect, we found a near-significant pattern in the non-musicians' steady-vowel productions, in which following upshifts led to *larger* responses than following downshifts (t=–1.681, p=.09).

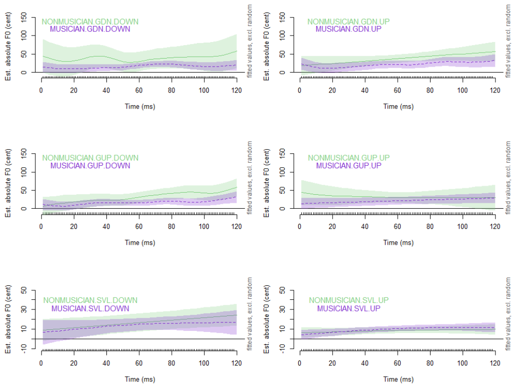

Figure 5: Musicality effect regarding *following responses* at each level of the combination of production block and stimulus direction. The purple curves represent musicians' following responses and the green curves represent non-musicians' following responses.

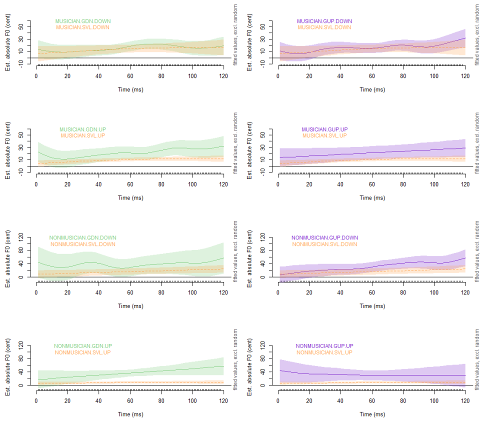

Figure 6: Production block effect regarding *following responses* at each level of the combination of musicality and stimulus direction. The purple curves represent following responses for upward glissando; the green curves represent following responses for downward glissando; the orange curves represent following responses for the steady-vowel condition.

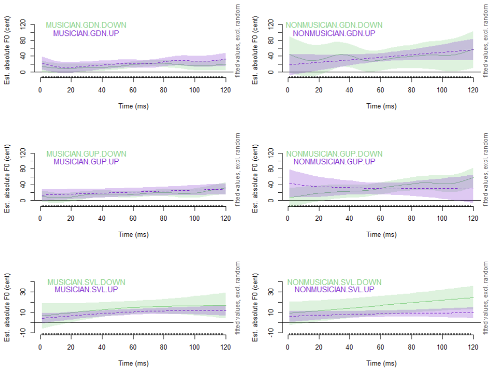

Figure 7: Stimulus direction effect regarding *following responses* at each level of the combination of musicality and production block. The purple curves represent following responses for upshift stimuli; the green curves represent following responses for downshift stimuli.

4. DISCUSSION AND CONCLUSION

Our results show that for musicians and non-musicians, following responses were almost as common as opposing responses when they encountered unexpected perturbations in auditory feedback. This finding suggests the opposing and following mechanisms could be activated simultaneously so that speakers may utilize one of the mechanisms from time to time. Following responses' availability should be independent of vocalization task (vowels and glissandos) and musical background. As the literature suggests, our musicians were also less affected (i.e., smaller responses) by perturbations than non-musicians were. The advantage in vocal pitch control should be associated with many years of musical training. We discovered larger (opposing and following) responses in the scenarios that required more rigid pitch control, such as in glissando vocalizations. Using generalized additive mixed models enables us to discover constant differences and time-varying changes in the pitch contours.

5. REFERENCES

[1] T. A. Burnett, M. B. Freeland, and C. R. Larson, “Voice F0 responses to manipulations in pitch feedback,” *Journal of the Acoustical Society of America,* vol. 103, no. 6, pp. 3153-3161, 1998.

[2] S. H. Chen, H. Liu, Y. Xu, and C. R. Larson, “Voice F0 responses to pitch-shifted voice feedback during English speech,” *Journal of the Acoustical Society of America,* vol. 121, no. 2, pp. 1157-1163, 2007.

[3] L.-H. Ning, “Pitch-shift responses as an online monitoring mechanism during level tone production,” *Journal of the Acoustical Society of America,* vol. 145, no. 4, pp. 2192-2197, 2019, doi: https://doi.org/10.1121/1.5096977.

[4] T. A. Burnett and C. R. Larson, “Early pitch-shift response is active in both steady and dynamic voice pitch control,” *Journal of the Acoustical Society of America,* vol. 112, no. 3, pp. 1058-1063, 2002.

[5] U. Natke, T. M. Donath, and K. T. Kalveram, “Control of voice fundamental frequency in speaking versus singing,” *Journal of the Acoustical Society of America,* vol. 113, no. 3, pp. 1587-1593, 2003.

[6] R. Behroozmand, O. Korzyukov, L. Sattler, and C. R. Larson, “Opposing and following vocal responses to pitch-shifted auditory feedback: Evidence for different mechanisms of voice pitch control,” *The Journal of the Acoustical Society of America,* vol. 132, no. 4, pp. 2468-2477, 2012, doi: 10.1121/1.4746984.

[7] M. K. Franken, D. J. Acheson, J. M. McQueen, P. Hagoort, and F. Eisner, “Opposing and following responses in sensorimotor speech control: Why responses go both ways,” (in eng), *Psychon Bull Rev,* vol. 25, no. 4, pp. 1458-1467, Aug 2018, doi: 10.3758/s13423-018-1494-x.

[8] L.-H. Ning, “Comparison of Involuntary and Volitional Responses to Pitch-Shifted Auditory Feedback: Evidence for Tone Speakers’ Flexibility to Switch Between Opposing and Following Responses,” *Journal of Speech, Language, and Hearing Research,* vol. 65, no. 6, pp. 2160-2186, 2022, doi: doi:10.1044/2022_JSLHR-21-00597.

[9] J. M. Zarate and R. J. Zatorre, “Experience-dependent neural substrates involved in vocal pitch regulation during singing,” *NeuroImage,* vol. 40, pp. 1871-1887, 2008.

[10] S. Wood, *Generalized Additive Models: An Introduction with R.* Boca Raton, FL: Chapman & Hall/CRC, 2006.

[11] Y. Sun and C. Shih, “Boundary-conditioned anticipatory tonal coarticulation in Standard Mandarin,” *Journal of Phonetics,* vol. 84, p. 101018, 2021, doi: https://doi.org/10.1016/j.wocn.2020.101018.

ACKNOWLEDGEMENTS

I would like to express my sincere appreciation to the research assistants, Hao-Yu Wu, Pei-Jiun Chen, and Hsin-Jou Chen, for their invaluable assistance with data collection. Their contributions were crucial to the success of this research project.