Proceedings of the 12th International Conference on Music Perception and Cognition and the 8th Triennial Conference of the European Society for the Cognitive Sciences of Music, July 23-28, 2012, Thessaloniki, Greece
Cambouropoulos E., Tsougras C., Mavromatis P., Pastiadis K. (Editors)

Effect of Augmented Auditory Feedback on Pitch Production Accuracy in Singing

Dustin Wang[1], Nan Yan[2], Manwa L. Ng[3]

Division of Speech and Hearing Sciences, the University of Hong Kong, Hong Kong
[1]wangdn03@hku.hk, [2]nyan@hku.hk, [3]manwa@hku.hk

ABSTRACT

The effect of augmented (accompanying) auditory feedback on pitch production accuracy during singing is controversial. Yet, the lack of control of vocal range as well as the different criteria of grouping participants into poor and normal pitch singers might have contributed to the contradictory findings reported in the literature. In the present study, 7 poor pitch singers as well as 11 controls who had no formal training of singing were recruited to perform in both a single-note pitch-matching task and a song-singing task. All participants are native speakers of a tonal language. Absolute and relative pitch accuracy were compared between speaker groups for the two tasks. Acoustic analysis was carried out using PRAAT and the stimuli were generated using a music notation software (MUSESCORE) to better control the tempo of presenting the stimuli and the accompaniment. The objective of the current study is to investigate the effect of augmented auditory feedback on pitch accuracy for both poor and good pitch singers and to compare the effect between two types of tasks. Data collection is still in progress, however, available data show that the effect of augmented feedback is positive for the moderately poor pitch singers but not the severely poor ones in the pitch-matching task, but its influence on the performance in the song-singing task is negative.

I. INTRODUCTION

Similar to language acquisition, singing is a common human ability that is acquired at an early age. People can sing in an artistic performance or just for the purpose of recreation. However, not everyone is able to acquire an acceptable level of singing skills. Criteria for good singing include accurate pitch and rhythm as well as good vocal quality, among which the grasp of pitch accuracy is considered the fundamental one. How can we help people who have trouble singing in tunes is an interesting research question.

Before we learn to produce pitches in singing, we must first learn to perceive different pitches. As the major route of pitch perception, auditory feedback has a unique role in pitch control for both professional and amateur singers. While most of the previous studies using manipulation of the auditory feedback concerned blocking the feedback with noise (Mürbe, Pabst, Hofmann, & Sundberg, 2002), few studies directly focused on the effect of augmented version of the auditory feedback, which refers to presenting the target melody simultaneously during the participant's production, coexistent with the singer's own auditory feedback. The rationale is that this accompaniment serves as a tutor to the singer and may help him/her to improve.

In a feedback blocking and augmentation experiment, Pfordresher and Brown (2007) recruited 79 participants to perform both the pitch matching task and familiar-song singing task. It was found that the augmentation of auditory feedback had a detrimental effect on the absolute accuracy for poor singers but improved the relative accuracy for both good and poor singers. However, some confounding factors were not well controlled in the study. Firstly, since there were both male and female participants, whose vocal ranges varied significantly, there was possible mismatch between the participant's vocal range and the fundamental frequency of the target stimuli. For instance, a female participant might sometimes be required to match the stimuli appropriate for males, and vice versa. Also, although augmented feedback was used, since its onset was the same as the singer´s production, it was impossible to tell whether the effect was due to the existence of the accompaniment or the singer using the accompaniment as the second stimuli where there was no gap before the production in the pitch matching setting. Lastly, the auditory feedback augmentation was not applied to the song-singing task.

In a more recent study by Wise and Sloboda (2008), it was shown that the accuracy of either single tone or sequence composed of two, three and five tones were significantly enhanced by the existence of augmented auditory feedback and the "tone-deaf" group benefited more than the "non-tone-deaf" group. In the task of singing a familiar song, augmented feedback significantly improved the performances of both groups. However, grouping of the participants was based on self-evaluation, and according to their performance in the pitch matching task, all participants were actually good singers when simpler stimuli were presented, with the average pitch deviation less than 0.5 semitones, a threshold justified by both self-judgment (Estis, Coblentz, & Moore, 2009) and the study of just-noticeable difference (JND) of singing voice (Hutchins & Peretz, 2011), which is also adopted in the current study. Additionally, the performance of both groups in the song-singing task was evaluated by professional judges, without using objective methods.

Seeing the contradicting results from the studies mentioned above, whether augmented auditory feedback has a positive influence on the pitch production accuracy of both poor pitch singers and good singers respectively requires further investigation. In the present study, grouping of the participants was more strictly controlled and the range of the stimuli was also adjusted according to each participant's comfortable pitch. Acoustical analysis was adopted to evaluate the effect of auditory feedback augmentation of untrained people who were either poor pitch singers or good singers on their performances in the pitch-matching task and the song-singing task.

II. METHOD

A. Participants

The present study adopted a 2 x 2 (feedback x group) mixed factorial design for both pitch-matching and song-singing tasks. The first within-subjects variable is the condition of auditory feedback, i.e., normal condition versus augmented condition. The second within-subjects variable is the type of the task. A natural groups design is used as a factor of group, i.e., a threshold of 0.5 semitones is adopted to distinguish the two groups by their performance in the pitch-matching task in normal condition. In order to guarantee a larger percentage of poor pitch singers among all participants, a pre-selection criteria was used which required that the participants had trouble staying in tunes when singing, as judged either by themselves or by others. Also, the participants had no formal vocal training. Up to now, a total of 20 participants have been recruited from the University of Hong Kong campus. The age of the participants ranged from 20 to 30 years. They were screened for pure pitch discrimination deficit of "congenital amusia" using the Montreal Battery of Evaluation of Amusia (MBEA) (Peretz, Champod, & Hyde, 2003), and hearing loss was screened out by the pure-tone audiometric test. Among the 18 qualified participants, 7 (5 females and 2 males) were classified as poor pitch singers (PS) and the rest (8 females and 3 males) good singers (GS). The participants were native Mandarin speakers, Cantonese speakers and bilinguals.

B. Equipment

Fundamental frequency (F0) was used as the main acoustic correlate of the perceived pitch and was calculated using PRAAT (version 5.2.44) (Boersma & Weenink, 2009), based on which the two outcome measures for pitch accuracy were calculated: absolute deviation, which was obtained by averaging all single tones in the pitch-matching task, and relative deviation, which was obtained by averaging all intervals within the phrases in the song-sing task. Both measures were independent of the direction of the deviation. Sub-Harmonic Summation was selected as the pitch extraction algorithm in PRAAT which has been proven to be more reliable over other algorithms (Keelan, Lai, & Zechner, 2011).

A multi-track recorder (TRAVERSO DAW, version 0.49.1) (Sijrier, Doebelin, & Levitt, 2009) was used for presenting the stimuli and recording the responses. Figure 1 shows the interface of the program. For the pitch-matching task, the first track provided the metronome, which ran at 60 bpm. The second and third tracks provided the stimuli and augmented feedback respectively. The fourth track contained bars to visually hint the participant with the onset and offset of the production. The fifth track recorded participant's production. For the song-singing task, one track provided the metronome running at 100 bpm and another track provided the augmented feedback. All stimuli as well as the augmented feedback were generated using MUSESCORE (version 1.2) (Schweer, 2012). Speech samples were obtained by using a high-quality microphone (Shure, SM58) in a sound-attenuated room (with the background noise less than 35 dB(A) measured by a sound level meter). All acoustic signals were digitized at 20 kHz and a quantization rate of 16 bits/sample.

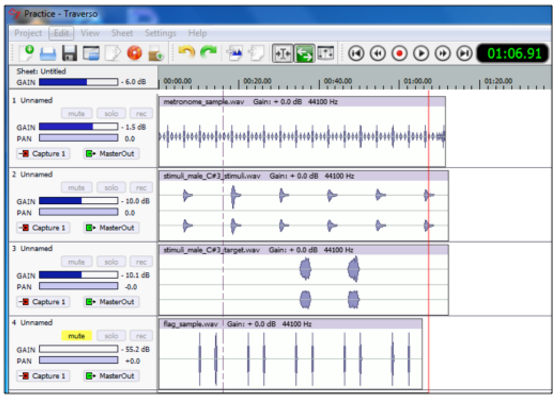

Figure 1. Recording program interface

C. Procedure

Participants were required to perform three experiment tasks, two screening tests and a questionnaire. At the beginning, pure-tone audiometric test and questionnaire were administered, followed by the first task, then MBEA test and the other two experiment tasks. The entire session took approximately one hour. The first task was the song-singing task in normal condition. The participants were asked to sing the song "Twinkle Twinkle Little Star" (See Figure 2) a cappella in their most comfortable pitch. The average F0 of the entire song was calculated and used as an indicator of the most comfortable pitch of the participant, based on which the choice of the stimuli in the following tasks was made.

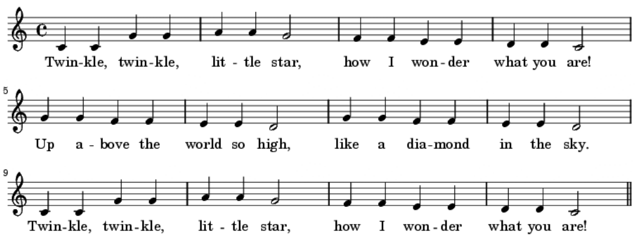

Figure 2. "Twinkle Twinkle Little Star"

The second task was the pitch-matching task. The stimuli were single tones and were presented via the headphones. The stimuli had the timbre of synthesized piano and there were a total of nine different tones in accordance with each participant's most comfortable pitch. The pitch of the tone in the middle was assigned with the most comfortable pitch, with four tones above and below it. All nine tones were equally distributed in the 12-note equal temperament scale with an interval of 0.5 semitones between adjacent tones. For example, if the participant's most comfortable pitch is D4, all other tones would be: A#3, B3, C4, C#4, D4, D#, E4, F4 and F#4. The participants were asked to match the stimuli by producing the syllable /da/ for 4 seconds, starting from 4 seconds after the onset of the stimuli. In the normal condition, the participants only heard their own voice during production, while in the augmented condition, the accompaniment was presented via the headphones during the middle part of the participants' production with its onset 2 seconds later than that of the production. The accompaniment carried the timbre of synthesized human voice: mezzo-soprano for female participants and baritone for male participants. The flow of the task is illustrated in Figure 3. The second task had one session in which there were a total of 9 (stimuli) x 2 (condition) x 2 (repetition) = 36 trials in a randomized order.

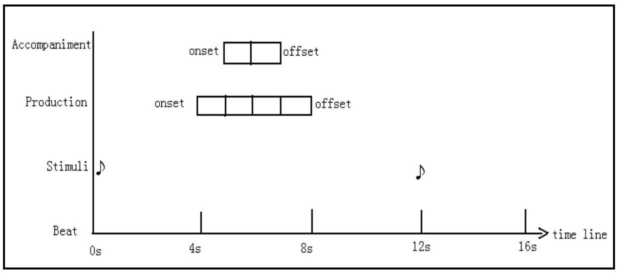

Figure 3. Flow of pitch-matching task

The third task was the song-singing task in augmented condition. The participants were asked to sing "Twinkle Twinkle Little Star" once again with the accompaniment presented through the headphones simultaneously with the production. The average pitch level of the accompaniment was also in accordance with the most comfortable pitch of the participant.

III. RESULTS

A. Pitch accuracy

Figure 4 shows the pitch accuracy of both GS and PS groups in the pitch-matching task in normal condition. The accuracy was the measured absolute deviation as mentioned before. Accordingly, the larger is the value, the lower is the accuracy. As seen in Figure 4, GS group is below the 0.5-semitone threshold, with a mean value of 0.26 semitones, and ranges from 0.12 semitones to 0.48 semitones. The PS group is above the threshold, with a mean value of 1.35 semitones, and ranges from 0.51 semitones to 3.41 semitones. Compared to GS group (SD=0.11 semitones), the PS group (SD=1.03 semitones) is much more varied due to an extreme case of P7 (3.41 semitones).

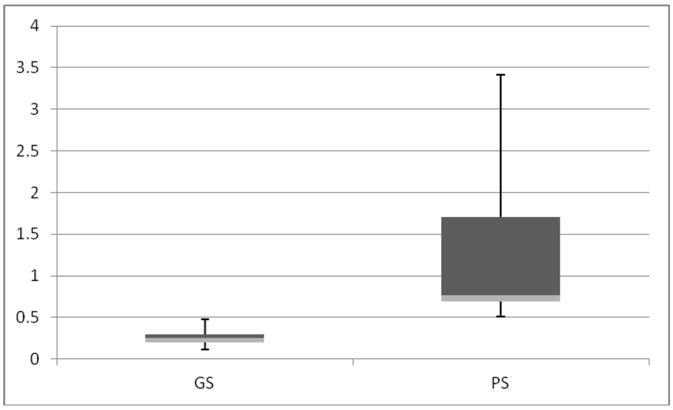

Figure 4. Absolute deviation in normal condition (unit: st)

As shown in Figure 5, compared to the normal condition, when auditory feedback is augmented, the mean deviation increases from 0.26 semitones (SD = 0.11 semitones) to 0.30 semitones (SD = 0.15 semitones) for the GS group and from 1.35 semitones (SD = 1.03 semitones) to 1.42 semitones (SD = 1.63 semitones) for the PS group. Also listed in the Figure are the data of all 18 participants. For the GS group, it can be seen that augmented feedback has little influence on the pitch accuracy. However, for the PS group, except for P7, who has extremely low pitch accuracy, the augmented feedback increases the pitch accuracy slightly. This is confirmed by using Wilcoxon Signed Ranks Test. When the extreme case is excluded, the PS group significantly improved their performance but not the GS group (PS: Z = -2.201, p = 0.028; GS: Z = -1.746, p = 0.081).

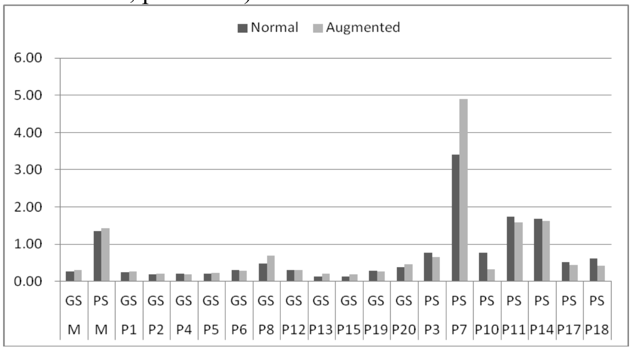

Figure 5. Absolute deviation in both conditions (unit: st)

Figure 6 shows the influence of augmented feedback on the relative deviation in the song-singing task. For both groups, the augmented feedback decreases the relative pitch accuracy. The deviation increases from 0.23 semitones (SD=0.08 semitones) to 0.28 semitones (SD=0.12 semitones) for the GS group, and from 0.28 semitones (SD=0.12 semitones) to 0.36 semitones (SD=0.1 semitones) for the PS group. However, both changes were not significant according to the Wilcoxon test (PS: Z = -1.682, p = 0.093; GS: Z = -1.329, p = 0.184).

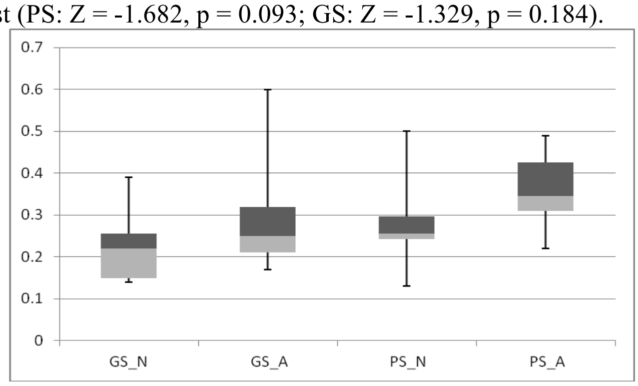

Figure 6. Relative deviation in both conditions (unit: st)

B. Pitch precision

Figure 7 shows the influence of augmented feedback on pitch precision. Pfordresher et al. (2010) used the variance of the participant's responses in the pitch-matching task as the index for pitch precision, in order to measure the consistency of the participant in singing the same target tone across multiple times. Since each tone in the pitch-matching task was only repeated twice in the current study, Pearson's correlation coefficient was calculated for each participant instead. The result shows that for the GS group, most participants benefited from augmented feedback and for the PS group, pitch precision of all participants were improved, especially those with relatively low precision (P7 and P14). The Wilcoxon test showed that the effect on both groups were significant (PS: Z = -2.366, p = 0.018; GS: Z = -2.041, p = 0.041).

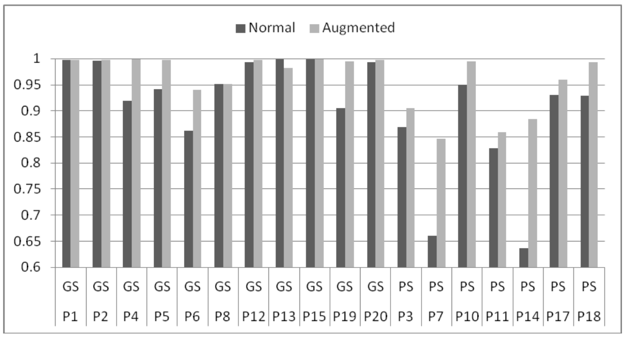

Figure 7. Pitch precision in both conditions

IV. DISCUSSION

Due to the limited number of participants, no solid conclusions can be drawn with confidence at the current stage. However, certain trends can be observed from the available data. In general, similar to Pfordresher and Brown (2007), the averaged pitch matching accuracy decreased when augmented feedback was applied, possibly due to a single case of extremely poor accuracy. It seems that poor pitch singers with moderate performance benefit from augmented feedback. A possible explanation for the seemingly discrepant result is that Pfordresher and Brown (2007) used a more strict threshold of 1 semitone in distinguishing poor and good pitch singers, while only a threshold of 0.5 semitones was used in the present study. Using such a loose subject selection criterion might have led to misgrouping extremely poor pitch singers as poor singers in their study. Taken this discrepancy into consideration, results from the two studies appear to be virtually consistent: moderately poor singers can benefit from augmented feedback but not the extremely poor ones.

Another difference between the current study and that of Pfordresher and Brown (2007) concerns the relative accuracy of pitch intervals which is adversely affected by augmented feedback in the current study. This may be due to the difference of the task conducted. Pfordresher and Brown (2007) used tone sequences composed of five tones in a pitch-matching task compared to the song-singing task in this study. It is possible that augmented feedback has more positive influence on the pitch-matching task than the song-singing task. The essential difference between the two tasks is that in the former one, stimuli are unfamiliar to the participants, so only short-term memory is involved, while in the latter one, the stimuli are familiar tunes stored in the long-term memory. Still, more experiment is needed to confirm this hypothesis.

The current result also shows that augmented feedback positively influences pitch precision. Although being two distinct indices, pitch accuracy and precision are correlated and inaccuracy might be a deeper deficit than imprecision in that inaccurate singers are very likely to be imprecise, while the chances for accurate singers to be precise and imprecise were similar (Pfordresher et al., 2010). This may suggest imprecision is more easily effected by augmented feedback than inaccuracy even for the extreme poor pitch singers (such as P7).

Since the study is still in progress, only after more data is available, especially for the PS group, can we have more confirmative results. In the future, it may be necessary to categorize the participants into more groups according to the pitch accuracy, for example, good singers (pitch deviation less than 0.5 semitones), moderately poor singers (pitch deviation greater than 0.5 semitones but less than 2 semitones) and extremely poor singers (pitch deviation greater than 2 semitones).

REFERENCES

Boersma, P., & Weenink, D. (2009). Praat: doing phonetics by computer (Version 5.1.13). Retrieved from http://www.praat.org

Estis, J. M., Coblentz, J. K., & Moore, R. E. (2009). Effects of increasing time delays on pitch-matching accuracy in trained singers and untrained individuals. *Journal of Voice, 23*(4), 439-445.

Hutchins, S., & Peretz, I. (2011). Perception and action in singing. In A. Green, E. Chapman, J. Kalaska & F. Lepore (Eds.), *Progress in Brain Research* (Vol. 191, pp. 103-118): Elsevier.

Keelan, E., Lai, C., & Zechner, K. (2011). *The importance of optimal parameter setting for pitch extraction*. Paper presented at the 160th Meeting Acoustical Society of America, Cancun, Mexico.

Mürbe, D., Pabst, F., Hofmann, G., & Sundberg, J. (2002). Significance of auditory and kinesthetic feedback to singers' pitch control. *Journal of Voice, 16*(1), 44-51.

Peretz, I., Champod, A. S., & Hyde, K. (2003). Varieties of musical disorders. *Annals of the New York Academy of Sciences, 999*(1), 58-75.

Pfordresher, P. Q., & Brown, S. (2007). Poor-pitch singing in the absence of "tone deafness". *Music Perception, 25*(2), 95.

Pfordresher, P. Q., Brown, S., Meier, K. M., Belyk, M., & Liotti, M. (2010). Imprecise singing is widespread. *J. Acoust. Soc. Am, 128*(4), 4.

Schweer, W. (2012). MuseScore (Version 1.2). Retrieved from http://musescore.org/

Sijrier, R., Doebelin, N., & Levitt, B. (2009). Traverso DAW (Version 0.49.0). Retrieved from http://traverso-daw.org/

Wise, K. J., & Sloboda, J. A. (2008). Establishing an empirical profile of self-defined " tone deafness": Perception, singing performance and self-assessment. *Musicae Scientiae, 12*(1), 3-26.