Proceedings of the 11th International Conference on Music Perception and Cognition (ICMPC11). Seattle, Washington, USA.
S.M. Demorest, S.J. Morrison, P.S. Campbell (Eds)

ALTERED AUDITORY FEEDBACK EFFECTS ON KEYBOARD AND SINGING PERFORMANCE

Peter Q. Pfordresher, Thomas Varco

Department of Psychology, University at Buffalo, State University of New York, USA

ABSTRACT

Past research that has investigated the effects of altered auditory feedback (AAF) on keyboard production suggests that fluency in action planning may be sensitive to both the timing and the contents of auditory feedback. Furthermore, the effect of AAF that manipulates feedback timing (synchronization of feedback with actions) typically disrupts the production of timing whereas manipulations of feedback contents (such as shifting the sequential relationship of feedback pitches relative to the action sequence) typically disrupts accuracy (e.g., Pfordresher, 2003). A limitation of research supporting this dissociation has been the exclusive use of keyboard production, whereas other action systems – vocalization in particular – may rely in different ways on auditory feedback (cf. Howell et al., 1983). We ran two studies that test whether the results found for keyboard production generalize to vocal production, here singing. Participants sang melodies that were learned and memorized through imitation, and then produced melodies repeatedly at a prescribed rate while hearing different AAF conditions that were designed to simulate the kinds of manipulations used in keyboard studies. Experiment 1 focused exclusively on singing whereas in Experiment 2 participants engaged in both singing and keyboard production tasks. Results overall suggest that the effect of AAF is consistent across effector systems, and analyses of individual differences suggest that the amount of disruption experienced during production with one effector system predicts sensitivity of the other effector system to similar manipulations of feedback. These results support the view that disruption from AAF is based on abstract, effector independent, response-effect associations.

1. INTRODUCTION

It has been known for some time that altering the effects of auditory feedback can severely disrupt the production of complex sequence. This effect was first found in the domain of speech production, which is disrupted when feedback is delayed by a fixed amount of time, particularly for delays around 200 ms (Black, 1951; Lee, 1950). Subsequently it was found that disruptive effects extended to the production of instrumental music (Havlicek, 1968), suggesting that the effect of altered auditory feedback (AAF) may extend across effector systems.

For some time, it was thought that the effects of AAF were limited to alterations that influence onset synchrony between actions and sounds. In several studies, Howell and colleagues demonstrated that the effect of DAF (which almost always results in asynchronies between actions and feedback) is not influenced by alterations of the anticipated sound category, such as modifying speech feedback to sound like a square wave tone (Howell & Archer, 1984). Furthermore, in the domain of music performance, Finney (1997) found that altering the pitch of auditory feedback in a quasi-random way, while maintaining synchrony, is not disruptive.

However, other research in keyboard performance suggests a broader role for auditory feedback. Specifically, Pfordresher (2003) found that alterations of feedback pitch that vary the sequential relationship between actions and sound (while maintaining synchrony) are in fact disruptive. Importantly, the disruptive effect of these alterations (here called *serial shifts*) is to disrupt accuracy, whereas the effect of asynchronous AAF is primarily on timing. There is thus a dissociation in the effects of asynchronous AAF and serially shifted AAF in keyboard production, with asynchronous feedback disrupting timing of production and serial shifts disrupting action planning (Pfordresher, 2006). We here test whether this dissociation is also found for vocal production, namely singing.

Why might one expect different effects across effector systems? Howell and colleagues (Howell, Powell, & Khan, 1983) suggested that mapping from actions to anticipated consequences may be more solidified for vocal production than keyboard production. They highlighted music performance as a specific example, pointing out that there is a great deal of flexibility between actions and the sounds that might result from those actions. By contrast, Pfordresher (2006) has suggested that perception/action links may function similarly across effector systems.

We therefore ran the experiments reported here simply to test whether the kind of dissociation reported in Pfordresher (2003) extends to vocal production. Experiment 1 directly tests this idea by incorporating feedback alterations like those used in keyboard tasks in a singing task. Experiment 2 further explores whether the degree of disruption experienced by during AAF of the voice while singing is similar to the degree of disruption experienced during keyboard performance.

2. EXPERIMENT 1

In Experiment 1 we endeavored to create AAF conditions in a singing tasks modeled on those used in earlier studies using keyboard production. In addition we included a constant frequency shift condition modeled on those commonly used in other studies of vocal production (e.g., Zarate & Zatorre, 2008). Participants learned and then reproduced short, novel melodies by singing. While singing they could hear feedback that was normal, frequency shifted, shifted to a random degree in regular intervals, or timed in such a way as to model asynchronous or serially shifted AAF.

2.1 Method

Participants. 10 students from the University at Buffalo participated in the experiment. The participants' mean age was 20 years, with a range of 18 to 23 years. Eight participants reported playing an instrument, 7 having at least one year of musical training, with an average of 5 years; 6 participants had formal vocal training. All participants reported normal hearing, were right-handed, and were native English speakers. Eight participants were male, two were female

Conditions. Four different monophonic, 8-note melodic sequences were used with a range of C2 to G2 for males, and C3 to G3 for females. Five feedback conditions were used in this experiment. One was a normal feedback control. Two AAF conditions were designed to model asynchronous and serially shifted feedback from keyboard studies, but are here referred to according to the delay amount used in the interest of accuracy. Because true sequential shifts are very difficult to implement with vocal performance, it was simulated by using a delay plug-in with a delay time equal to the inter-onset interval of the notes in the sequence. Specifically the *300 ms Delay* condition was designed to model an asynchronous delay, whereas the *600 ms Delay* condition was designed to model the serial shift (given the prescribed IOI of 600ms). Two further AAF conditions were used to test whether the restrictions on disruptive AAF conditions found for keyboard tasks also hold for singing. One was a condition that modeled the random feedback condition used in keyboard tasks (e.g., Finney, 1997; Pfordresher, 2005). In this condition, called *Random F0 shift*, feedback pitch was shifted to random levels within (within ±3 semitones) every 600 ms. The final condition was a *Constant F0 shift* which, as the title implies, involved a single step change in feedback pitch which was always +3 semitones. Participants heard only the altered ("wet") audio and none of their unaltered ("dry") audio in the headphones. Thus, as a participant imitated a note, they heard the previous note that they produced. The feedback conditions were randomly ordered such that each appeared once per sequence, for a total of 20 trials.

Apparatus. The experiment took place inside of a Whisper Room sound isolation room. The stimuli were generated using Yamaha's Vocaloid Leon software package and presented over Sennheiser HD 280 Pro headphones. Participants were recorded using a Shure PG58 microphone into a Lexicon Omega recording interface. Steinberg's Cubase LE software package was used to present stimuli and feedback alterations, and to record imitations. Two VST plug-ins, Cakewalk Delay and de la Mancha pitchfork, were used within Cubase to alter the audio in real-time.

Procedure. Participants began the experiment with a set of exercises intended to warm-up the voice. For each melodic sequence, there was a "learning phase" followed by five "imitation phases." In the learning phase, the participants were instructed to listen to the sequence as it was looped six times. They then sang along with the looped sequence six times. The participants could repeat the learning phase as much as needed until they had the sequence memorized. In each imitation phase, the participants heard the sequence one time along with a metronome set at 100 BPM (600 ms IOIs). The metronome continued for four beats after the sequence while the participants imitated the sequence repeatedly for 37 seconds (chosen to allow around 7-8 times through the sequence). After the participants imitated the sequence twice, a VST plug-in was activated to alter the feedback. In order to learn the structure of the experiment, participants had a sample trial that consisted of a learning phase and two imitation phases using a melodic sequence different from the experimental sequences. Participants were also presented with a flowchart of the experimental procedure for visual reinforcement.

Data analysis. The fundamental frequency of each produced pitch was analyzed with respect to its pitch and timing. For timing, we analyzed the inter-onset intervals (IOIs) from the onset of one syllable to the next. For pitch, we analyzed accuracy with respect to the absolute difference between produced and intended F0 for each individual note, as well as the absolute difference between produced and intended pitch intervals (pairwise differences, see Pfordresher & Brown, 2007 for further description). Whereas the former measure assesses absolute pitch in production, the latter assesses relative pitch. Though we report both measures of pitch accuracy, we consider interval accuracy to better reflect the errorfulness of production.

2.2 Results

Each measure of performance was analyzed using a single-factor within-subjects analysis of variance (ANOVA). Means and standard errors for each measure are shown in Table 1.

Feedback condition	**M IOI** (ms)	**Note error** (cents)	**Interval error** (cents)
Normal	603.6 (6)	79.1 (21)	81.3 (18)
300 ms Delay	**660.0** (11)	94.9 (25)	99.9 (21)
600 ms Delay	626.6 (6)	**112.6** (28)	**127.1** (26)
Random F0 shift	620.5 (13)	42.5 (13)	91.7 (13)
Constant F0 shift	620.8 (13)	**112.1** (24)	103.1 (20)

Table 1: Results from Experiment 1. Parentheses show between-subjects standard errors. Bolded numbers indicate significant differences from normal feedback according to Tukey's HSD post-hoc test ($\alpha = .05$).

There was a significant effect of feedback condition for each performance measures, Mean IOI F (4, 36) = 7.49, $p < .01$, Note error F (4, 36) = 4.36, $p <.01$, Interval error F (4, 36) = 4.66, $p < .01$. It is more important, however, to examine which AAF conditions differ from the normal feedback conditions, in order to determine whether conditions causing disruption for keyboard performance are similarly disruptive for vocal performance.

In general, the current results replicate the qualitative findings found for keyboard production. With respect to timing, mean IOIs increased significantly, relative to normal feedback, only when participants experienced the 300ms delay condition, designed to model asynchronous feedback. This is important in two respects. First, it replicates one half of the dissociation reported by Pfordresher (2003). Second, the fact performances were not significantly slower with delays of 600ms than with normal feedback indicates that our attempt to model the serial shift of feedback was successful. In more concrete terms, likely experienced the 600ms delay not as asynchronous, but as a discrepancy between the expected and perceived pitch of their voice.

Turning to measures of pitch accuracy, both note and interval measures show no significant effect of the 300ms delay on production but do show a significant increase in error for the 600 ms delay condition. In addition, note errors increased significantly when participants experienced the constant F0 shift whereas interval errors did not. This discrepancy relates to an important difference between these conditions. When experiencing a fixed F0 shift, participants typically "transpose" sung notes but do not necessarily make more "errors" in singing. Thus these data suggest that the 600ms delay, which models the effects of serially shifted feedback, selectively disrupts accuracy in production.

2.3 Discussion

Experiment 1 replicated the effects found in keyboard production when participants sang melodies. This despite the fact that wer were not able, practically speaking, to manipulate AAF for singing with the same degree of control as is possible for keyboard performance. Thus these data not only support the idea that the dissociation in the effects of asynchronous and serially shifted AAF is effector independent, they show that this dissociation is a highly robust effect.

A limitation of Experiment 1, however, is that only singing performance was recorded. It is unclear from these results whether the disruptive effect of AAF on singing is directly comparable to the effect of AAF on keyboard production for an individual. Experiment 2 was designed to address this question.

3. EXPERIMENT 2

Experiment 2 was designed to be identical to Experiment 1, except that each participant produced melodies by singing or by performing them on the keyboard, in two halves of the experiment. Furthermore, in order to maintain consistency across the performance tasks, the learning phase and AAF manipulations for keyboard trials were identical to singing trials. Because we were interested in individual differences in Experiment 2, more participants were sampled than in Experiment 1.

3.1 Method

Participants. Seventeen introductory psychology students participated in the experiment. The participants' mean age was 19 years, with a range of 18 to 25 years. Ten participants reported playing an instrument, 7 having at least one year of musical training, with an average of 5 years. Three participants reported piano training. All participants reported normal hearing. All were right handed. Nine participants were male, and 8 were female.

Conditions. The same set of stimuli and feedback conditions were used as in Experiment 1. The feedback conditions were randomly ordered such that each appeared once for each run through a sequence, for a total of 20 trials. Sequences were randomly assigned to be sung or played on the keyboard such that participants would not sing and play the same melody. Singing and keyboard trials were performed in separate halves of the experiment, with order of these tasks counterbalanced across participants.

Apparatus. The apparatus was the same as the previous experiment, with the addition of a Roland SC-55mkII sound generator controlled by an M-Audio Keystation 49e MIDI controller, which was used for keyboard production trials. Keyboard stimuli were generated using the MIDI controller with General MIDI software instruments and time- and duration-quantized in Cubase.

Procedure. The same structure for sample trials, learning phase, and imitation phase was used as the previous experiment. For this experiment, participants had a sample trial before both the vocal and keyboard segments, and each segment consisted of two sets of one learning phase and five imitation phases.

3.2 Results

Initially we expected to analyze the results for Experiment 2 in the same way as we had for Experiment 1. Unfortunately, due to a computer error participants performed sequences at a rate closer to 120 BPM (500ms IOIs) on average rather than the intended rate of 100 BPM (600ms IOIs). Because of this error, the 300 and 600 ms conditions cannot be interpreted as modeling "asynchronous" and "serially shifted" feedback, as had been intended. At the same time, preliminary analyses suggested that these conditions, and not the random or constant F0 shift conditions, disrupted production. Thus, rather than focus on the effects of feedback conditions we instead focus on individual differences in the disruptive effect of AAF across keyboard and singing tasks for the 300ms and 600ms delays.

We computed disruption for each participant and AAF condition which are differences between that condition and normal feedback. In order to best match production tasks with respect to pitch accuracy, we use interval errors for singing tasks, and converted interval errors into error rates by treating any interval error that was greater than 100 cents as an error. This metric better matches the error rate metric that we use to measure accuracy in keyboard production. Table 2 reports correlations across keyboard and singing tasks, broken down by measure of disruption and feedback condition. We focus on the two most critical feedback conditions, those associated most reliably with disruption of timing or accuracy.

AAF condition	IOI difference	Error difference
300 ms Delay	**0.54**	-0.25
600 ms Delay	**0.57**	**0.58**

Table 2: Correlations among disruption scores for keyboard and singing tasks. Bolded values indicate correlations that exceed the critical value of $r = 0.41$ (for df = 15, $\alpha = .05$, one-tailed).

In general correlations across production tasks were significant, offering further support for the notion that AAF disruption is effector independent. Of additional interest is the fact that the only non-significant correlation pertained to the effect of the 300ms delay on accuracy. Although, as mentioned before, these delays cannot be interpreted as modeling “asynchronous” versus “serially shifted” feedback it is nevertheless the case that the 300ms delay would not have resulted in mismatches between expected and perceived pitch contents. That is, this condition could still be considered to be “asynchronous”.

4. GENERAL DISCUSSION

The two experiments reported here offer support for the idea that the effects of AAF on production are not specific to a single effector system**.** We here focused on two effector systems most commonly associated with the communication of complex sequences, namely vocal and manual. People use these effector systems to communicate through music (investigated here) and language.

In Experiment 1 we verified that different qualitative effects of various types of AAF hold in vocal production just as in manual production of music. Experiment 2 further demonstrated that the degree to which an individual’s production is disrupted by AAF within one effector system predicts the disruption he or she would experience with a similar AAF manipulation in the other effector system.

An important theoretical implication of these results is that sensitivity to AAF is not necessarily based on fixed response-effect associations but instead may be based on a more general tendency to expect that a given movement pattern should be correlated with a concurrent perceived sequence. This conclusion is based on the logic, articulated by Howell et al. (1983), that response-effect associations are more reliable for the voice than they are in music production. Every time an individual adjusts their vocal folds to a certain length and tension, and forces air through that opening, a specific pitch can be anticipated with high reliability. The same is not true for the consequences of finger movements. Even concert pianists must associated myriad consequences for their finger movements. Despite these very different constraints on the response-effect configurations, highly similar consequences are found during AAF. Thus, it seems more likely to conclude that the disruptive effect of AAF is not due to disruption of response-effect associations learned by direct associative chaining, but instead by more abstract connections between the planned goals of our actions and their anticipated consequences (cf. Hommel et al., 2001).

5. ACKNOWLEDGEMENT

This research was supported in part by NSF grant BCS-0642592. We thank Rebecca O’Connor for assistance in conducting the experiments. She and John D. Kulpa assisted in analyzing the data.

6. REFERENCES

Black, J. W. (1951). The effect of dleayed side-tone upon vocal rate and intensity. *Journal of Speech and Hearing Disorders, 16*, 56-60.

Finney, S. W. (1997). Auditory feedback and musical keyboard performance. *Music Perception, 15*, 153-174.

Havlicek, L. (1968). Effects of delayed auditory feedback on musical performance. *Journal of Research in Music Education, 16*, 308-318.

Hommel, B., Müsseler, J., Aschersleben, G., & Prinz, W. . (2001). The theory of event coding (TEC): A framework for perception and action planning. *Behavioral and Brain Sciences*(24), 849-937.

Howell, P., Powell, D. J., & Khan, I. (1983). Amplitude contour of the delayed signal and interference in delayed auditory feedback tasks. *Journal of Experimental Psychology: Human Perception and Performance, 9*, 772-784.

Howell, P., & Archer, A. . (1984). Susceptibility to the effects of delayed auditory feedback. *Perception & Psychophysics, 36*, 296-302.

Lee, B. S. (1950). Effects of delayed speech feedback. *Journal of the Acoustical Society of America, 22*, 824-826.

Pfordresher, P. Q. (2003). Auditory feedback in music performance: Evidence for a dissociation of sequencing and timing. *Journal of Experimental Psychology: Human Perception and performance, 29*, 949-964.

Pfordresher, P. Q. (2005). Auditory Feedback in Music Performance: The Role of Melodic Structure and Musical Skill. *Journal of Experimental Psychology: Human Perception and Performance, 31*, 1331-1345.

Pfordresher, P. Q. (2006). Coordination of perception and action in music performance. *Advances in Cognitive Psychology. Special Issue: Music performance, 2*, 183-198.

Pfordresher, P. Q. (2008). Auditory Feedback in Music Performance: The Role of Transition-Based Similarity. *Journal of Experimental Psychology: Human Perception and Performance, 34*, 708-725.

Pfordresher, P. Q., & Brown, S. (2007). Poor-pitch singing in the absence of "tone deafness". *Music Perception, 25*, 95-115.

Zarate, J. M., & Zatorre, R. J. (2008). Experience-dependent neural substrates involved in vocal pitch regulation during singing.. *NeuroImage, 40*, 1871-1887.