FINNEY, STEVEN A., *Auditory Feedback and Musical Keyboard Performance* , Music Perception, 15:2 (1997:Winter) p.153

Music Perception
Winter 1997, Vol. 15, No. 2, 153–174

Auditory Feedback and Musical Keyboard Performance

STEVEN A. FINNEY
Brown University

In an investigation into the role of auditory feedback guidance in musical performance, musically experienced subjects performed on an electronic keyboard under altered feedback conditions that included pitch and timing manipulations, as well as absence of auditory feedback. The results largely replicated the data reported by Gates and Bradshaw (1974): performance in the absence of auditory feedback showed no impairment, whereas performance under delayed auditory feedback showed significant impairment. In an extension of the Gates and Bradshaw study, however, it was found that altered pitch feedback caused little or no impairment and that altering the pitches in the delayed auditory feedback condition significantly reduced the amount of delayed auditory feedback impairment. These results show that different components of auditory feedback (pitch and timing) have separable effects on musical performance and pose a problem for theories of auditory feedback effects that do not explicitly distinguish these components.

PERFORMANCE on most musical instruments involves the execution of a sequence of finger movements; the goal of these movements is to produce an intended sequence of sounds. Skilled musicians necessarily have some form of internal representation in which the motor plan and the expected sound are linked: for instance, a performer can detect and react to a finger placement or timing error when the auditory expectation is violated. Because sound is the underlying goal of the musical performance process, the auditory consequences (feedback) might also serve to guide motor performance in ways other than error detection.

Lashley (1951) claimed that skilled musical performance was too rapid to be guided by feedback, basing this conclusion on the speed of the neural mechanisms involved. However, little empirical work has directly addressed the relationship between musical instrument performance and auditory feedback. One way to investigate the role of feedback is to measure the behavioral consequences of feedback removal or alteration. An example of this

Address correspondence to Steven Finney, Department of Cognitive and Linguistic Sciences, Box 1978, Brown University, Providence, RI, 02912. (e-mail: Steven_Finney@brown.edu)

approach occurs in the speech literature, where extensive experimentation has investigated the effects of delayed auditory feedback (DAF). In a DAF speech experiment, subjects hear their own speech through headphones, delayed by a short time interval; when compared with normal speech, speech under DAF typically has a reduced rate, increased intensity (volume), and an increased number of articulatory errors (for early reviews of this literature, see Smith, 1962; Yates, 1963). Such a DAF manipulation has also been applied to music performance tasks. Havlicek (1968) measured performance errors during sight-reading on various acoustic instruments (including piano), with the sound heard by subjects delayed by approximately 200 ms. Havlicek reported more errors in a DAF condition than in a normal feedback condition, as well as increased intensity. Gates and Bradshaw (1974) provided additional evidence of a DAF impairment effect in musical performance. Subjects in their experiment learned a baroque piece from written notation to a criterion level of performance on an electronic organ and then played the piece "as fast as possible" from the written notation; the dependent measure was total elapsed time to perform the piece. Performance under DAF of 180 ms to both ears caused significant disruption (slower performance) when compared with the normal feedback condition. Gates and Bradshaw also reported that impairment similar to that produced by DAF occurred when subjects heard a different musical piece rather than their own performance. The results of Havlicek (1968) and Gates and Bradshaw (1974) can be interpreted as support for a sensory guidance hypothesis (i.e., closed-loop performance), where sensory feedback continuously guides sequencing of movement and where disruption of (necessary) feedback impairs performance.

The demonstration that DAF impairs musical performance is perhaps not surprising, given the well-documented impairment effect of DAF on speech. An additional result reported by Gates and Bradshaw (1974) is more unexpected. In an additional feedback condition (made possible by the use of an electronic instrument), subjects could not hear their performance at all. Under a sensory guidance hypothesis, it would be expected that performance in such a condition would also be impaired. However, performance in the no auditory feedback condition (as measured by elapsed performance time) did not differ significantly from the normal feedback condition. A similar result was reported by Banton (1995), who measured performance errors in a sight-reading task and found no significant impairment when auditory feedback was removed. Gates and Bradshaw (1974) argued that "auditory imagery" might be used to guide performance in the no-feedback condition, suggesting that actual sensory feedback per se is not necessary in this task, because auditory imagery can take its place. However, incorrect (temporal) feedback *can* disrupt the task, either by interfering with the ability to form an auditory image or by distracting subjects' attention (as suggested by Borden, 1979, for speech).

These explanations for the impairing effect of DAF do not specifically rely on the temporal character of DAF, but would potentially apply to any major auditory feedback alteration. Two results in the speech literature, however, give conflicting predictions about whether this would be true.

Howell, Powell, and Khan (1983) argued that DAF impairment in speech is caused by the presence of a temporally asynchronous signal and is independent of the content of the asynchronous signal (the "displaced rhythm hypothesis"). Two results supporting this view are that a delayed signal consisting of a modulated sawtooth wave (i.e., no speech information) caused impairment in speech similar to that caused by DAF and that intermittent square-wave gating of an undelayed speech signal also caused impairment. Howell et al. also claim that this displaced rhythm hypothesis should apply to DAF impairment in both speech and nonspeech tasks. Thus, it is specifically the timing manipulation that is the source of the DAF effect.

An opposing hypothesis can be derived from research on stuttering. In a number of studies, the speech of stutterers has been found to *improve* under DAF (see Van Riper, 1982, for review and references). Kalinowski, Armson, Roland-Mieszkowski, Stuart, and Gracco (1993) (see also Howell, El-Yaniv, & Powell, 1987) compared the performance of stutterers under a DAF condition to that under a pitch-altered condition in which the speech feedback heard by stutterers was raised by a half-octave. They found that both delay and pitch manipulations improved stutterers' performance, and to a similar extent. This result suggests that pitch and timing alterations may affect speech performance in the same way and provides preliminary support for the view that altered feedback per se (rather than some particular component of altered feedback, e.g., timing) is the relevant factor in DAF effects.

These two hypotheses have interesting implications for the role of feedback in musical performance. Auditory feedback in music contains multiple types of information. The pitch component (which is essential to musical meaning) contains information about hand and finger positioning. Timing, clearly, is also essential in musical performance, and the timing of sound signals gives information about the timing of the performer's movements. Although arbitrary pitch feedback manipulations are difficult to implement in speech, computer Musical Instrument Digital Interface (MIDI) systems provide the capability for flexible manipulation of auditory feedback in real time in the musical domain. In particular, computer control of a MIDI-equipped keyboard allows independent manipulation of the pitch and timing components of auditory feedback.

FINNEY, STEVEN A., ***Auditory Feedback and Musical Keyboard Performance*** , Music Perception, 15:2 (1997:Winter) p.153

Two specific hypotheses (motivated by the speech studies cited earlier) will be addressed in Experiment 1:

1. If feedback (or auditory imagery) guides performance, then any disruption of feedback should impair musical performance. In particular, major disruption of pitch might be expected to cause impairment similar to that of delay.
2. If DAF impairment is due strictly to the presence of temporally asynchronous signals and is independent of the identity of the signals (as suggested by Howell et al., 1983), then musical performance under a DAF condition in which pitches are altered should cause impairment similar to that caused by DAF alone.

Experiment 1 was also designed to further investigate the effect of absence of auditory feedback. Given the central role of sound in musical performance, the results of Gates and Bradshaw (1974) and Banton (1995), in which no impairment was found, are somewhat counterintuitive. It seemed worth determining if more detailed measurements would show impairment in this condition.

Experiment 1

In Experiment 1, experienced keyboard players performed two excerpts from the Two-Part Inventions of J. S. Bach under five different feedback conditions that included pitch and timing manipulations, as well as absence of auditory feedback; each subject performed in all conditions. Quantitative measures applied to the resulting performance data allowed relative impairment across the different feedback conditions to be determined.

METHOD

Subjects

Eleven keyboard players of various levels of experience were recruited from the greater Brown University community; the amount of formal training reported ranged from 6 to 15 years (mean, 10.9 years), and the number of hours per week of current musical activity ranged from 0 to 20 (mean, 6.7 hr). Ten of the 11 subjects reported piano as their primary keyboard instrument; the other subject was an organist. One subject reported having perfect pitch.

Materials

The first part of the experiment involved multiple performances of the first 6 measures of the Two-Part Invention in C-major by J. S. Bach (G. Schirmer edition); this excerpt contains 84 notes played by the right hand and 50 played by the left hand (Figure 1). As a replication, the procedure was then repeated with the first 17 measures of the D-minor

FINNEY, STEVEN A., ***Auditory Feedback and Musical Keyboard Performance*** , Music Perception, 15:2 (1997:Winter) p.153
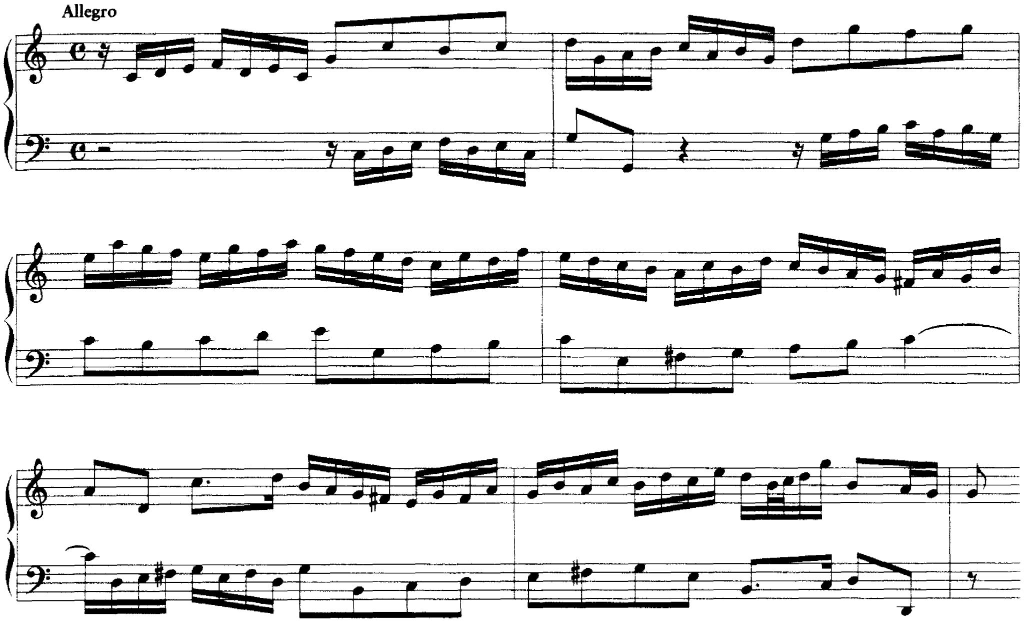
Fig. 1. Two-Part Invention in C-major by J. S. Bach: performed excerpt.

Invention (82 notes in the right hand, 71 in the left hand; Figure 2). The Bach Inventions were chosen because they are readily available, are two-voiced polyphonic (simplifying analysis), and are of a moderate level of difficulty. Some subjects, but not all, had played these pieces before.

Equipment

Subjects performed on a Yamaha DX-7II electronic keyboard (henceforth DX-7); this keyboard is velocity-sensitive (i.e., to simplify slightly, increased keystroke force causes increased volume). The DX-7 has unweighted keys and so does not have the "feel" of a piano; such a keyboard would be inappropriate for the study of expressive musical performance by skilled pianists, but should be adequate for the lower-level domain of study addressed here.

Local MIDI mode on the DX-7 was turned off so that no sound was produced in direct response to key presses. Key press information went to the MIDI output port, which was connected to a Silicon Graphics (SGI) Indigo computer via an Opcode MIDI Translator II interface. A custom program using the SGI MIDI library received and recorded the incoming signals; for each physical key press, the data included the identity of the key (i.e., the note), the time (to the nearest millisecond) that the key was pressed, the time that the key was released, and the velocity of the key press. (The resulting computer files were used for quantitative analysis; they could also be played back as auditory output). The computer also sent (possibly altered) MIDI signals back to the tone generator of the DX-7. In the "normal" feedback condition, the notes sent back to the keyboard were identical to the notes the subject played. In the "delay" condition, MIDI messages were sent in a manner that delayed each note by 250 ms (close to the delay time of maximum impairment found for musical performance by Gates, Bradshaw, & Nettleton, 1974). In the "silent" (no auditory feedback) condition, no MIDI output was sent, and the subject heard no pitch information. In the "pitch" condition, most input notes were mapped to different output notes, creating a near-random sounding scale (a pseudo-random mapping was chosen because it
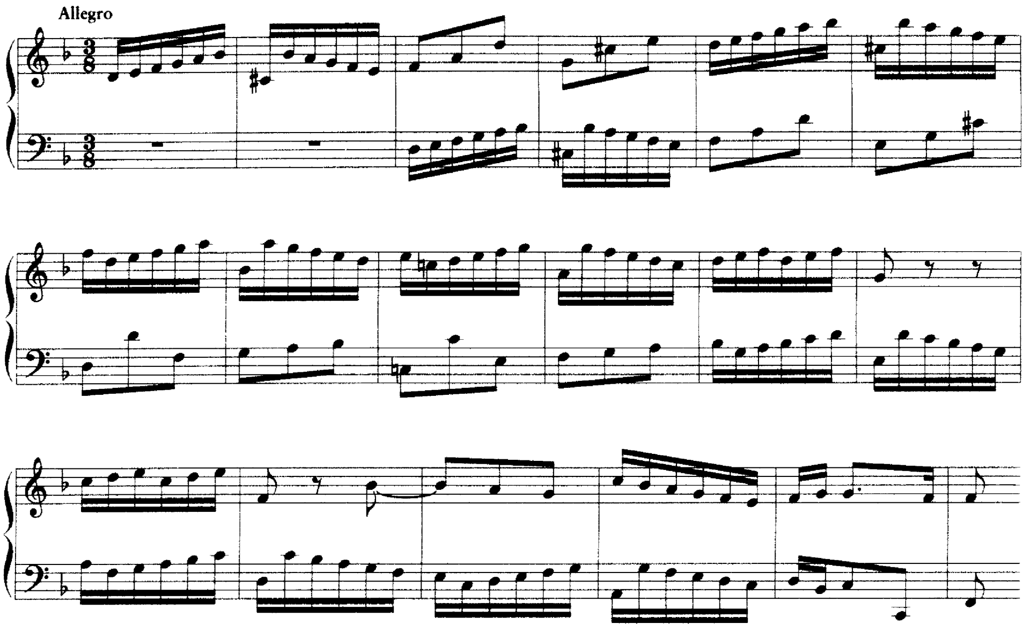
Fig. 2. Two-Part Invention in D-minor by J. S. Bach: performed excerpt.

seemed likely that a simple transposition would have no effect). The mapping was arbitrarily chosen and consisted of the following note changes: C → F♯ above, C♯ → B below, D → F above, E → G an octave above, F → B below, G → G♯ above, G♯ → D below, A → B♭ above, B♭ → C above. The notes D♯, F♯, and B were unchanged. In the "delay + pitch" condition, the output signal was delayed 250 ms, and the pitch was also mapped to a randomly chosen note within a musical fourth of the fingered note (e.g., a C was changed to a note within the range of a G below to the F above, with the mapping of *each keystroke* randomly determined within that range); the intent of this was to make the delayed information highly uninformative.[1]

The tone generator of the DX-7 responded to the received MIDI signals with a velocity-sensitive voice that resembled an electronic harpsichord (Clavinette, DX7-II Cartridge voice 48); increased key force resulted in increased volume and some increased brightness of timbre. A metronome was implemented by having a program on the computer send out a regularly timed MIDI signal to a second, short percussive voice. The speed of the system was such that the response with normal feedback was perceived as instantaneous; subjects listened to the output from their performance though Koss PRO/4AA headphones plugged into the keyboard. The DX-7 volume level was set to 6, which gave a sound level of about 80 dB SPL in the subject's headphones for a hard key press; in the silent condition, where there was no feedback, some mechanical key noise from the keyboard was audible.

---

[1] It should be noted that the pitch alteration used in the delay + pitch condition was not identical to that used in the pitch condition, invalidating direct comparison between the two. However, the overall effect of both was to make the performed Bach pieces sound vaguely like 20th century atonal music. Experiment 2 addresses the issue of different types of pitch alteration; the results do not show show any significant differences between the pitch and the random-pitch alterations.

FINNEY, STEVEN A., ***Auditory Feedback and Musical Keyboard Performance***, Music Perception, 15:2 (1997:Winter) p.153

Procedure

Subjects sat in front of a table that supported the DX-7 keyboard; the sheet music was placed on a music stand directly behind the keyboard. Subjects had unimpeded view of the keyboard and their hands. As a warmup, subjects first played scales to get accustomed to the keyboard; they then practiced the excerpt of the C-major Invention until they could play it comfortably and accurately. Subjects were given written instructions that asked them to play at an even tempo without expressive variation and to leave out ornaments such as trills. A metronome speed was then chosen jointly by the experimenter (the author) and the subject; the goal was to find a setting that approached the subject's maximum speed but was relatively comfortable and error-free. The same metronome speed was used for all trials of a given subject on a given piece, but the speed of the metronome varied widely across subjects (for the C-major Invention, a quarter note ranged from 32 per minute to 94 per minute; for the D-minor Invention, an eighth noted ranged from 64 per minute to 222 per minute). The amount of DAF remained fixed at 250 ms regardless of the subject's chosen tempo. A trial started with a metronome count of 8 beats; the metronome then stopped, and subjects were instructed to come in on the following beat maintaining this tempo. The motivation for this procedure was to provide a base tempo for all trials of a given subject; a metronome running throughout the performance was deemed unfeasible because of the disorienting effect it created in the delay condition. This procedure is similar to that used in the keyboard studies of MacKenzie and Van Eerd (1990).

Because of concern that subjects might get frustrated and disoriented under conditions of altered feedback, the normal feedback condition was always used as the initial condition and was run every three trials (leading to a total of four trials in this condition). Each of the other four conditions was run twice, with the order randomly selected for each subject. Before each trial, the experimenter correctly informed the subject about the type of feedback for that trial. Once the initial orientation, practice and metronome setting had taken place, the 12 trials typically could be completed in less than 10 min.

After this initial experiment, subjects were asked for verbal reports of their impressions and conscious strategies. Each subject then repeated the entire procedure with the D-minor Invention, using a different randomized sequence of conditions.

RESULTS

Many dependent measures can be extracted from the raw data (which contained detailed key-press timing and velocity information). The measures chosen were the number of note-choice (pitch) errors, total elapsed time per trial (as in Gates and Bradshaw, 1974), mean keystroke velocity, interhand coordination, and internote timing consistency.

Errors

"Note errors" were defined as errors in the sequencing of notes within the monophonic part of each hand; timing of the notes within the monophonic line (or between hands) was not considered as an error. To quantify this measure, the computer data files for each trial were split into separate files for each hand. The note data in each resulting file were then compared with the correct sequence of notes for that hand, and the UNIX "diff" program was used to find the minimum set of differences between the two

FINNEY, STEVEN A., ***Auditory Feedback and Musical Keyboard Performance*** , Music Perception, 15:2 (1997:Winter) p.153

sequences.[2] Each difference was then categorized as a deletion, addition, or substitution (as in Palmer & van de Sande, 1993), and each such individual discrepancy was counted as an error. The errors for the two hands of one trial were combined to give a total error count for that trial, and the values for the trials for a subject in a given condition were then averaged to give a data point for that subject. The mean values across subjects are shown in Figure 3.

An analysis of variance (ANOVA) was performed on the data from the C-major Invention, with feedback condition as a within-subjects factor. The omnibus ANOVA was highly significant [$F(4, 40) = 20.00, p < .001$]. Six comparisons were of particular interest: comparing each altered feedback condition with the normal feedback condition (four comparisons), as well as the comparison of altered pitch feedback with delayed feedback, and delayed feedback compared with delayed feedback combined with pitch alteration. These pairwise comparisons were evaluated by using Tukey's honestly significant difference (HSD) method with significance at the .05 level; this approach will be used throughout the paper.

As can be seen from Figure 3, for the C-major Invention, the delay condition resulted in more errors than any of the other conditions. Three of the six comparisons proved significant: normal vs delay, pitch vs delay, and delay vs delay + pitch. The normal condition did not differ significantly from silent, pitch, or delay + pitch.

Analysis of the data from the D-minor Invention gave similar results (see Figure 3). The omnibus ANOVA was highly significant [$F(4, 40) = 11.81, p < .001$]. Again, delay showed the most errors, and of the pairwise comparisons of interest, the same three were significant: normal vs delay, pitch vs delay, and delay vs delay + pitch. The other comparisons were not significant. Once again, the analysis did not reliably differentiate between the normal, silent, and pitch conditions.

For both the C-major and D-minor Inventions, the majority of the errors in all conditions consisted of note additions (for the C-major Invention, note-addition errors comprised between 65% (delay condition) and 79% (pitch condition) of the total errors; for the D-minor Invention, this ranged from 57% in the delay + pitch condition to 80% in the silent condition). Because the primary concern here is demonstrating differences between the effects of different classes of feedback conditions rather than detailed contextual error analysis (such as that done in musical planning studies such as

---

[2] The "diff" program is a standard UNIX utility for comparing two text files. By invoking it with a text file representation of a correct performance as the first argument and the subject's performance as the second, the "diff" output corresponds to meaningful error interpretations of the data (e.g., a report of an added text line corresponds to a note addition). It should be noted that comparing two long sequential files is a nontrivial computational problem that does not have a unique solution (Large, 1993). Thus, the error count that results from the use of "diff" will not necessarily precisely match that reached by other methods, although the differences should be small.

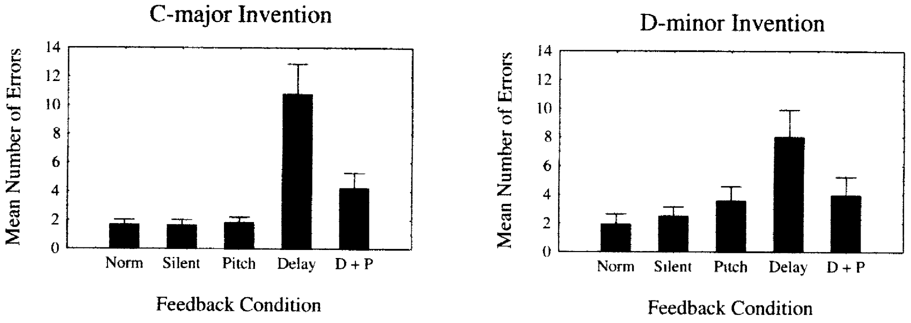

Fig. 3. Experiment 1: Mean number of note errors for the five feedback conditions (D+P = delay+pitch condition).

Palmer & van de Sande, 1993), further detailed analysis of error subtypes will not be pursued.

Performance Time

The total time to perform a piece (the measure used by Gates and Bradshaw, 1974) both quantifies "difficulty" to some extent and bears a direct relation to maintenance of the initial tempo, because slowing down in a condition will manifest itself in a longer performance time. Total performance times for the C-major Invention (number of seconds between first key press and final key release) are shown in Figure 4.

The omnibus ANOVA for the C-major data was significant [$F(4, 40) = 3.72$, $p = .01$], although none of the six planned comparisons was significant. The same analysis performed on the D-minor data (see Figure 4) also gave an omnibus ANOVA that was significant [$F(4, 40) = 3.89$, $p = .01$], although again no pairwise comparisons approached significance. Thus, the performance time data do not statistically distinguish between the pairwise conditions of interest, although a trend for performance to be slowest in the delay condition was apparent in both pieces.[3]

Velocity

In addition to timing information, the collected data also included a velocity measurement for each key press; this value ranged between 0 and

---

[3] Gates and Bradshaw (1974) found a robust effect for delay impairment using such an elapsed time measurement. In Experiment 1 here, the impairment, although visible in the data of Figure 4, is not statistically significant. The fact that a trend exists in the current data suggests that the source of this difference in statistical results may be a difference in experimental power. Gates and Bradshaw included more trials in each condition (four, compared with two here); the total time in each trial of the Gates and Bradshaw experiment experiment was longer (closer to 60 sec than the 25–30 sec here); and the Newman-Keuls test they used for pairwise comparisons is more liberal than the Tukey test used here. It also should be noted that Experiment 2 here *did* show a significant effect with performance time in the DAF condition.

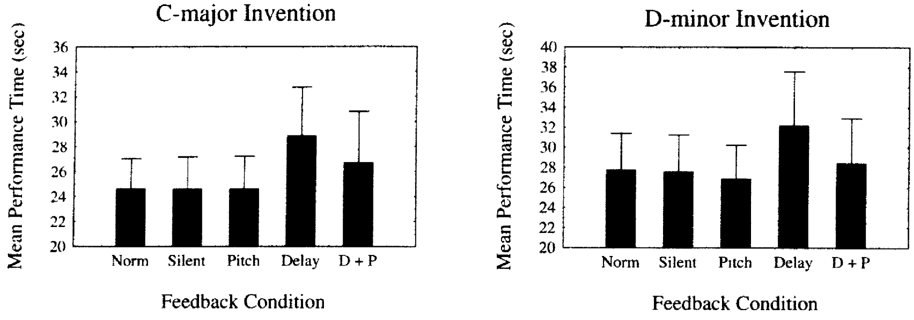

Fig. 4. Experiment 1: Mean performance time (in seconds) for the five feedback conditions (D+P = delay + pitch condition).

128 and used a metric internal to the MIDI implementation of the DX7-II keyboard. Although no detailed correspondence to standard force or velocity units is available, a key press of 50 MIDI units gave around 60–70 dB SPL in the headphones (depending on note pitch), and within the range displayed here, an additional velocity unit added 0.5–1.0 dB. The mean keystroke velocities for performance in the C-major Invention, as well as those for the D-minor Invention, are given in Figure 5. An omnibus ANOVA of the velocity data for the C-major Invention was highly significant [$F(4, 40)=14.92$, $p < .001$]. Of the six pairwise comparisons of interest, significant differences existed between the normal and delay condition and between the normal and pitch + delay condition. None of the other comparisons were significant; in particular, the mean values for delay and delay + pitch were almost identical. For the D-minor Invention, the omnibus ANOVA was significant [$F(4,40)=15.61$, $p < .001$]. As in the C-major Invention, pairwise comparison of normal versus delay was significant as was normal versus delay + pitch. No other differences were significant.

The finding of increased velocity in the delay conditions provides quantitative support for the observation in Havlicek (1968) that musical performance under DAF led to increased intensity and is consistent with the re-

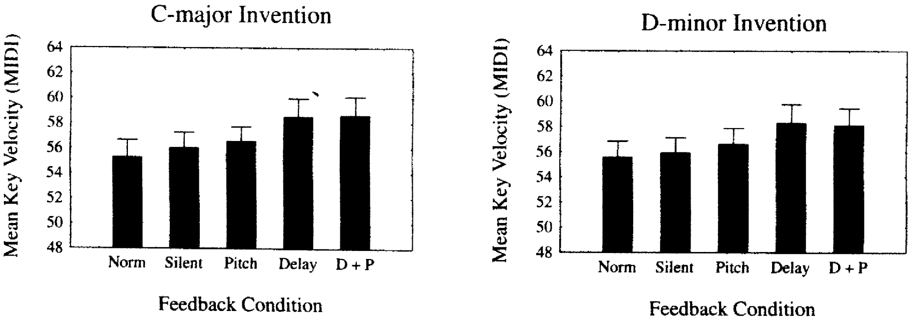

Fig. 5. Experiment 1: Mean key velocity (DX7-II specific units. D+P = delay+pitch condition).

FINNEY, STEVEN A., ***Auditory Feedback and Musical Keyboard Performance***, Music Perception, 15:2 (1997:Winter) p.153

ports of increased speech volume under DAF, and with the finding of Chase, Harvey, Standfast, Rapin, and Sutton (1961) that performance of a keytapping task under DAF led to an increase in key force. It should be noted that *both* the delay and the delay + pitch conditions show increased velocity in this experiment and that both were distinguished from normal feedback.

Interhand Coordination

Performance of polyphonic keyboard music requires coordination between the two hands. In the two Bach Inventions, certain pairs of notes in the right- and left-hand parts are notated as occurring simultanously (see Figures 1 and 2); in an unimpaired performance of this style of music (particularly given the instructions for nonexpressive interpretation), the key presses for these left and right hand notes should coincide. An interhand coordination measure quantified the extent to which this occurred by finding the average time difference between such keystroke pairs. To calculate this measurement, all note errors in the data files (additions, deletions, and substitutions) were first removed from consideration, as was one note on each side of such errors. Next, the error-free portions of the right-hand and left-hand files for a trial were compared, and the first 15 pairs of notated simultaneous notes were extracted. The difference in milliseconds between the key presses for each pair (as a positive number) was taken as a raw data value, and these 15 points were averaged to give a subject's score for a given performance. The mean values are given in Figure 6.

For the C-major Invention, the omnibus ANOVA was highly significant [$F(4, 40)=7.27$, $p < .001$]. The delay condition differed significantly from all other conditions, which did not differ significantly from each other. The same pattern occurred with the D-minor Invention data, with an omnibus ANOVA result of $F(4, 40)=20.71$, $p < .001$.

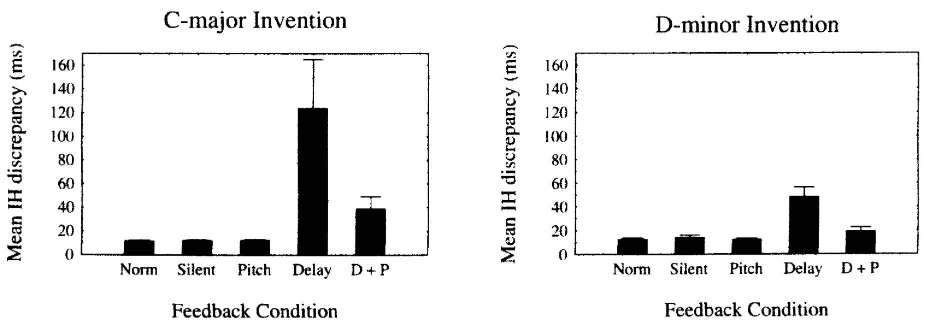

Fig. 6. Experiment 1: Mean interhand timing discrepancy (ms) for the five feedback conditions (D+P = delay+pitch condition).

FINNEY, STEVEN A., ***Auditory Feedback and Musical Keyboard Performance***, Music Perception, 15:2 (1997:Winter) p.153

Consistency Measurements

None of the measures just described have shown any difference between the normal and silent conditions. Two additional measurements were therefore applied to these two conditions and to the pitch condition. These measurements involved two aspects of *consistency* of timing, specifically variance of note duration (time between key press and the corresponding key release) and internote interval (i.e., time between one key press and the next key press). Such measures cannot be directly applied to music where the indicated note durations are of different lengths, such as the two Bach pieces. However, both of the musical excerpts had extended sequences of notes in the right hand that were equally timed (sixteenth notes in both cases). A 32-note right-hand section of the C-major Invention (bars 3 and 4; see Figure 1) was extracted from the data files, as well as a 42-note section from the D-minor Invention (bars 5–11; see Figure 2). For these extracts, only error-free versions were analyzed, because such measurements of timing variance would be highly affected by errors. Because most subjects had errors in the two delay conditions, these could not included in the analysis.

Note durations and internote intervals were calculated for the extracts of the two pieces, and the coefficient of variability was then calculated by dividing the standard deviation by the mean. An ANOVA was performed on the coefficient of variability data for both internote interval and note length for both the C-major and D-minor Invention extracts, giving a total of four one-way ANOVAs. No results approached significance; three of the calculated F's were less than 1.0, whereas the coefficient of variability for note duration for the D-minor Invention had a p value of .156 (F = 2.127). Thus, these timing components were not significantly affected by these feedback manipulations.

DISCUSSION

The above analyses consistently show that the DAF condition caused performance impairment. Pitch alteration did not significantly impair performance and was robustly superior to delay.

However, a significant concern remains regarding this comparison of pitch and delay. The delay manipulation was chosen to be maximally impairing, based on previous work of Gates, Bradshaw, and Nettleton (1974) investigating a range of delay intervals. The pitch manipulation is the first of its sort and may not be maximally impairing; in particular, the random-sounding nature of the feedback alteration may have contributed to subjects being able to ignore it. In order to strengthen the comparison between delay and pitch, investigation of other pitch manipulations that might be

FINNEY, STEVEN A., ***Auditory Feedback and Musical Keyboard Performance*** , Music Perception, 15:2 (1997:Winter) p.153

more comparable to the delay condition is necessary. Although it is not a priori obvious how to equate altered feedback conditions in the pitch and timing domains, subjects' verbal reports suggested two pitch manipulations that might be highly impairing: one in which the feedback was largely correct but there were occasional intermittent pitch alterations and one in which the pitches made some musical sense but were not the pitches of the piece the subject was performing. These two manipulations were tested in Experiment 2. In addition, as noted in Footnote 1, the fact that different pitch alterations were used in the pitch and delay + pitch conditions is a minor design flaw. Although nothing crucial hinges on this distinction, it nonetheless seemed worth determining if these two pitch alterations had different consequences.

Experiment 2

METHOD

Subjects

Nine keyboard players of various levels of experience were recruited from the greater Brown University community; none of the subjects had taken part in Experiment 1. The amount of formal training ranged from 6 to 16 years (mean, 10.4 years), and the number of hours per week of current musical activity ranged from 0 to 15 (mean, 5.3 hr). All subjects reported piano as their primary keyboard instrument. Two subjects reported having perfect pitch.

Materials

The same musical pieces (Bach C-major and D-minor Inventions) as in Experiment 1 were used.

Equipment

The equipment and keyboard settings were the same as Experiment 1, except that (due to equipment failure) Koss PRO/4X headphones were substituted for the Koss PRO/4AA headphones used in Experiment 1. Although the PRO/4X headphones produced a slightly higher sound level (3–4 dB), the consistent pattern of effects across the two experiments suggests that this equipment change did not affect the results in any important way.

Procedure

The experimental procedure was fundamentally the same as in Experiment 1, although some of the feedback manipulations used were different. There were two new altered pitch conditions. The "small" pitch condition was intended to simulate fingering errors by having occasional, small pitch alterations. Every other occurrence of the note E above middle C (usually played by the right hand) was raised a whole step; the same alternation occurred for E below middle C. Similarly, every other occurrence of A (in either hand) was lowered a half step. All other notes were sounded correctly.

FINNEY, STEVEN A., ***Auditory Feedback and Musical Keyboard Performance*** , Music Perception, 15:2 (1997:Winter) p.153

The "melodic" condition provided a sound output that made some musical sense but was not what the subject was expecting. This is not trivial, because the rhythm and fingered notes of the piece are fixed; furthermore, the procedure should be robust to possible errors by the subjects. The approach used was to chose a monophonic baroque piece (the Giga of the 4th Solo Violin Sonata in D-minor by J. S. Bach) and program the system so that every subsequent keystroke (regardless of hand or note) played the next note of this piece. The result did not have the harmonic structure of baroque music, but it did result in baroque-style melodic lines and a somewhat baroque-sounding result.[4]

In addition to these two new pitch alterations, the pitch condition of Experiment 1 was also used for comparison (although to avoid confusion with the other pitch alterations, it will now be referred to as the "large" pitch change condition). Also, the randomized pitch mapping of the delay + pitch condition of Experiment 1 was used, but without an associated delay; this will be referred to as the "random" condition. Finally, the DAF and normal conditions of Experiment 1 were also included.

The six feedback conditions thus consisted of "normal," "delay," and "large" ("pitch") conditions (as in Experiment 1), plus "small," "melodic," and "random" pitch alterations. Each piece had a total of 13 trials. The first, fifth, and ninth trials were normal feedback. The five altered feedback conditions were randomly assigned (per subject) to the first 5 remaining trials, and then a second trial of each was assigned to the remaining trials.

RESULTS

Errors

Note errors were quantified as in Experiment 1; the mean values are given in Figure 7. Inspection of the graphs shows a higher rate of errors in the delay condition than in the other conditions (for both Inventions). This is supported by the statistical analyses. The omnibus ANOVA for the C-major Invention was significant, [$F(5, 40) = 6.94$, $p < .001$]. Tukey's HSD test showed that the error rate in the delayed condition was higher than that in each of the other conditions, whereas no significant differences occurred between normal and any of the altered pitch conditions. The omnibus ANOVA for the D-minor data was also significant [$F(5, 40) = 6.05$, $p < .001$], and again the delay condition was significantly different from all other conditions, which did not differ among themselves. Furthermore, a one-way ANOVA restricted to the four altered-pitch conditions gave non-significant results ($F < 1.0$) for the data for both Inventions, supporting the hypothesis that there are no differences in effect between the various pitch-alteration conditions.

Performance Time

The mean values for total performance time are given in Figure 8. The performance time for the delay condition is higher than in the other condi-

---

[4] Due to limitations of the DX-7 keyboard (which allows only one instance of a note to be active at a time), the use of such a monophonic piece in dual-handed performance caused occasional clipping of notes. Informal replication with a few subjects using a different melodically coherent alteration that did not cause clipping (reversing the keyboard so that low notes were on the right and high notes were on the left) also did not show any noticeable signs of impairment.

FINNEY, STEVEN A., ***Auditory Feedback and Musical Keyboard Performance*** , Music Perception, 15:2 (1997:Winter) p.153

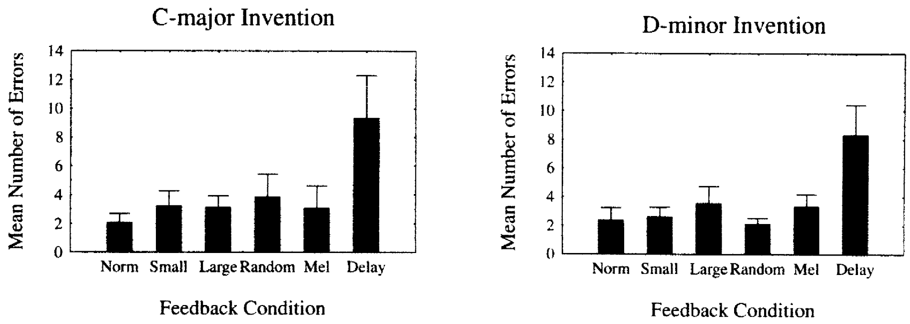

Fig. 7. Experiment 2: Mean number of note errors for the six feedback conditions.

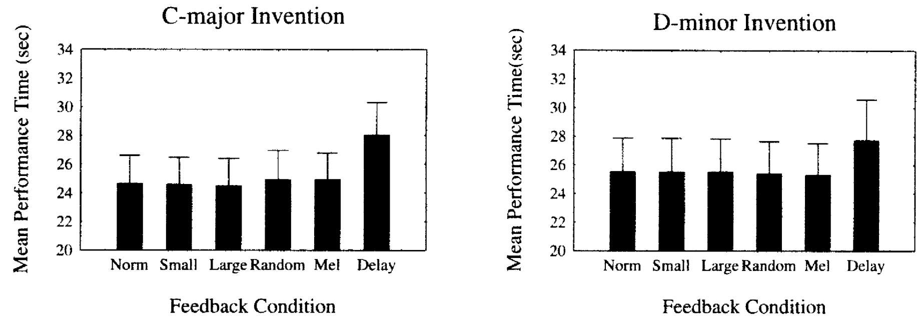

Fig. 8. Experiment 2: Mean performance time (in seconds) for the six feedback conditions.

tions, for both Inventions. This is verified by statistical analyses. The omnibus ANOVA for the C-major Invention was significant, [$F(5, 40) = 8.36$, $p < .001$]. Pairwise comparisons showed that the performance time in the delayed condition was higher than each of the other conditions, but no significant differences occurred between normal and any of the pitch conditions. The omnibus ANOVA for the D-minor data was also significant [$F(5, 40) = 5.02$, $p < .01$], and again the delay condition was significantly higher than all other conditions, which did not differ among themselves. A one-way ANOVA restricted to the four altered-pitch conditions gave nonsignificant results for both Inventions.

Velocity

The mean values for key velocity are given in Figure 9. For the C-major Invention, the omnibus ANOVA was significant [$F(5, 40) = 3.977$, $p < .01$]. The delay condition differed significantly from the normal condition by the Tukey HSD test; no other pairwise comparisons were significant. The omnibus ANOVA for the D-minor data was not significant [$F(5, 40) = 1.27$, $p > .25$]. Again, a one-way ANOVA restricted to the four altered-pitch conditions gave nonsignificant results for the data for both Inventions.

FINNEY, STEVEN A., ***Auditory Feedback and Musical Keyboard Performance*** , Music Perception, 15:2 (1997:Winter) p.153

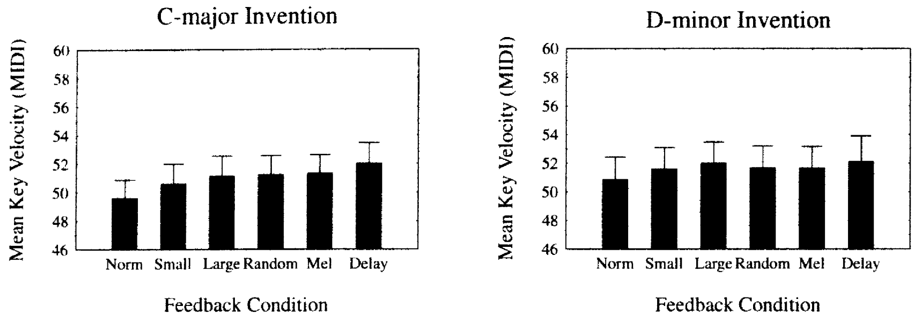

Fig. 9. Experiment 2: Mean key velocity (DX7-II specific units).

Interhand Coordination

Interhand coordination was quantified as in Experiment 1; the mean values are given in Figure 10. Inspection of the graphs shows a strikingly higher rate for the delay condition than the other conditions for both Inventions. This is supported by the statistical analyses. The omnibus ANOVA for the C-major Invention was significant, [$F(5, 40) = 19.80$, $p < .001$]. Tukey's HSD test showed that the error rate in the delayed condition was higher than each of the other conditions, whereas no significant differences occurred between any of the normal and pitch conditions. The omnibus ANOVA for the D-minor data was also significant [$F(5, 40) = 9.29$, $p < .001$], and again the delay condition was significantly different from all other conditions, which did not differ among themselves. Furthermore, a one-way ANOVA restricted to the four altered-pitch conditions gave nonsignificant results ($p > .20$) for the data for both Inventions.

Results: Summary

The results of the two experiments can be summarized as follows: No measurements differentiated the normal-feedback and no-feedback condi-

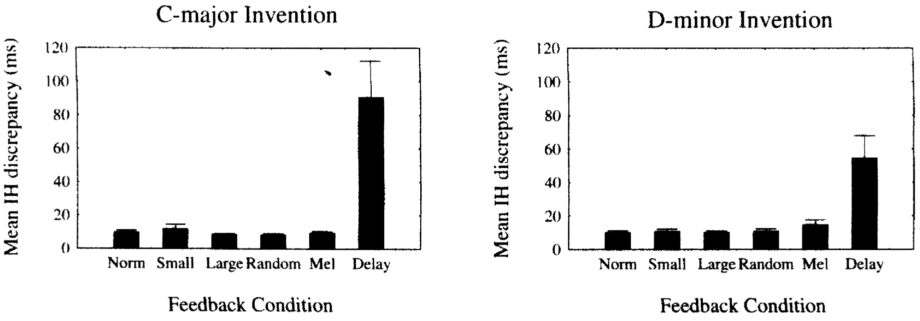

Fig. 10. Experiment 2: Mean interhand timing discrepancy (in milliseconds) for the six feedback conditions.

FINNEY, STEVEN A., ***Auditory Feedback and Musical Keyboard Performance***, Music Perception, 15:2 (1997:Winter) p.153

tions. All of the applicable measurements showed impairment in the DAF (delay) condition relative to the normal-feedback condition. These data confirm the results of Gates and Bradshaw (1974) even though (1) a different task was used (performance to tempo rather than "as fast as possible") and (2) more detailed measurements were utilized.

It should be noted that all subjects, regardless of skill level, showed impairment under the DAF condition. In Experiment 1, all 11 subjects showed decreased interhand coordination on the C-major Invention, and 10 of 11 subjects had more note errors. For the D-minor Invention, all 11 subjects were impaired on both measures. For Experiment 2, all 9 subjects were impaired on both measures on both Inventions. Thus, the presence of impairment under DAF is not dependent on skill level, at least for the range of subjects tested here. In addition, because the delay value was fixed at 250 ms whereas subjects' chosen tempos varied widely (over a range of 3:1 in Experiment 1, and 2.5:1 in Experiment 2), presence of impairment is not due to any particular relationship between the delay interval and performance tempo.

The additional conditions in these experiments provide important new data. First, feedback in which pitches were altered did not cause significant impairment; this result was consistent over four different pitch alterations. Second, altering the pitches in a DAF condition led to significantly improved performance compared with a pure DAF condition. In fact, parametric statistics did not demonstrate impairment in the delay + pitch condition, although it should be noted that (1) there were consistent trends toward delay + pitch impairment for both Inventions with multiple measures (see Figures 3, 4, and 6), (2) for the note-error data, a sign test for subjects suggests that subjects were impaired in the delay + pitch condition compared with the normal condition (for both the C-major and D-minor Inventions, 9 of 11 subjects showed worse performance with delay + pitch, significant with a sign test at $p = .027$), and (3) human listeners presented with randomized pairs of stimuli were able fairly reliably to detect impairment in performances in the delay + pitch condition compared with normal.

Discussion

FEEDBACK CONTROL OF MUSICAL PERFORMANCE

The overall question under consideration is the role of auditory feedback in musical performance. One possibility raised in the introduction was that musical performance is continuously guided by auditory feedback in a closed-loop manner; if such feedback is absent, auditory imagery can take its place. The impairing effect of DAF on performance would then be due to an interfering effect of altered feedback on necessary control mecha-

FINNEY, STEVEN A., ***Auditory Feedback and Musical Keyboard Performance*** , Music Perception, 15:2 (1997:Winter) p.153

nisms. The data provided on performance under altered-pitch conditions provides evidence against such hypotheses, as this form of feedback disruption did not cause impairment. In addition, the hypothesis that auditory imagery might guide performance when feedback is absent (Gates and Bradshaw, 1974) is compromised by the pitch results, because hearing an unexpected pitch sequence seems likely to interfere with an auditory image, yet no impairment was found. Further aspects of the difference between timing and pitch alterations will be discussed in the following section, but the main point here is that auditory feedback alterations do not necessary disrupt performance.

Additional evidence against closed-loop control comes from the fact that Experiment 1 replicated the finding (Gates and Bradshaw, 1974; Banton, 1995) that absence of auditory feedback does not cause impairment, even though a wider variety of measurements were used. However, certain important limitations on existing studies on auditory feedback absence should be noted:

1. All of the reported experiments have involved performance from sheet music. In such a situation, the motor task is defined by a visual stimulus, and audition is not logically necessary at all, because the subject must simply produce the correct movements in response to a visual code. Absence of auditory feedback might cause impairment in forms of performance where sheet music is not used.
2. None of the studies have investigated expressive aspects of performance, such as asking subjects to manipulate aspects of the *volume* of their performance (e.g., crescendos); such control might be very difficult without feedback.
3. Although pitch information was not available in the no-auditory-feedback conditions, there may have been some residual key-click noise, and visual, tactile, and proprioceptive information was unaltered. These forms of feedback may aid in performance. In particular, most keyboard experiments manipulating auditory feedback have allowed subjects to see their hands. Lee (1989) and Banton (1995) have demonstrated that performers' view of their hands plays an important role in musical keyboard performance. The availability of this visual feedback may help compensate for the absence of auditory feedback, and effects of auditory feedback might be demonstrated if such visual feedback were absent.[5]

---

[5] Lee (1989) briefly mentions an experiment in which view of the hands was blocked and auditory feedback was absent, but insufficient detail is provided with which to evaluate the experiment.

FINNEY, STEVEN A., ***Auditory Feedback and Musical Keyboard Performance*** , Music Perception, 15:2 (1997:Winter) p.153

4. One reason to hypothesize a link between motor patterns and sound expectations is the ability to correct errors, so error-correcting performance may differ between feedback and no-feedback conditions.

TIMING

The difference in effects between DAF and pitch alterations strongly suggests that the DAF impairment is in some way specific to the timing manipulation. This initially seems to provide support for the displaced rhythm hypothesis of Howell et al. (1983), which argues that it is simply the presence of an asynchronous signal that causes DAF impairment. However, the displaced rhythm hypothesis further claims that the informational content of the delayed signal is not relevant. The results in the delay + pitch condition are not consistent with this, as both the delay and the delay + pitch conditions in Experiment 1 alter the timing of the auditory responses to keystrokes in an identical manner, but the delay condition was significantly more impairing than the delay + pitch condition. The difference between the delay and delay + pitch conditions also argues against Havlicek's (1968) suggestion that DAF disruption results from a conflict between the delayed auditory feedback and the undelayed proprioceptive or tactile feedback, because the latter is similar in the two conditions.

However, tentative evidence exists for at least some impairment in the delay + pitch condition (and thus for some effect of delayed feedback per se). First, as described in the summary of results, there was a consistent (although nonsignificant) trend toward impairment in the delay + pitch condition. Second, both the delay and the delay + pitch conditions reliably showed increased key velocity. Although such increased key velocity cannot be clearly defined as "impairment," it is a consistent behavioral alteration that suggests that delay alone has an effect of some kind on performance. Such a reaction might be due either to an attempt to emphasize the rhythm in the absence of auditory confirmation or might be a stress reaction of some sort (something occasionally reported in the DAF literature, e.g., Yates, 1963; Havlicek, 1968).

Clearly, some more principled explanation for the relative effects of delay, delay + pitch, and pitch alterations is desirable. Explanations could come from one of two perspectives: (1) general considerations of human motor-sensory associations or (2) considerations of musical performance. The following paragraphs offer some speculations along these lines, although the current experiments and analyses do not allow resolution of the possibilities.

An explanation of the first type is that the difference between pitch and timing alterations is due to a hard-wired expectation of synchronous feed-

FINNEY, STEVEN A., ***Auditory Feedback and Musical Keyboard Performance***, Music Perception, 15:2 (1997:Winter) p.153

back. Various forms of synchronously timed feedback are ubiquitous in human interaction with the world (seeing, hearing, or feeling the effect of a physical action), and such expectations might be innate. Pitch information resulting from finger movements, however, is fairly specific to music; alteration of this sort of learned feedback may thus be less impairing. Why, then, the lessened impairment in the delay + pitch condition compared with delay? This could have to do with the extent to which motor and sensory events are perceived as associated. With consistent delay and correct pitches, the delayed events are correlated (although mismatched) with the motor events, and the mismatch is disturbing. When timing is off *and* pitch is off, however, the change of pitch may weaken the association between the motor and sound events so that the mismatched feedback is not perceived as closely associated with the motor performance and performance is not impaired. One interesting question this raises is what the effect would be of variably (e.g., randomly) timed delay, as such a variable delay might contribute to weakening the association between sound and performance.

Two preliminary musically based explanations were suggested by subjects' comments. One frequent subject report was that pitch alterations could be ignored, whereas delay could not. This alone still leaves the question of *why* such a difference exists, but one possibility is that musically incoherent feedback (such as large pitch alterations) can be ignored. This would explain the lack of impairment in the delay + pitch and pitch conditions of Experiment 1; however, it leaves unexplained the lack of impairment in the more musically coherent pitch conditions of Experiment 2.

Subjects also often reported trying to coordinate their fingering with the sound in the delay condition (e.g., "waiting for the notes to catch up"). In the case of a practiced polyphonic piece, subjects knew what fingered notes should correspond with what sounded notes. When a discrepancy was perceived (due to the delayed feedback), they might try to "correct" the error by fingering a note (in either the matching or the opposite hand). But correction is impossible, because the sound from the attempted correction will also be delayed, leading to a potential propagation of errors. However, because the pitches were all incorrect in the delay + pitch condition, there would be no motivation for engaging in corrective fingering, explaining the lessened impairment in this condition.

CONCLUSIONS

The research here has investigated the effect of altered feedback in musical keyboard performance, taking as a starting point some hypotheses generated from the speech literature. Such research has two somewhat sepa-

FINNEY, STEVEN A., ***Auditory Feedback and Musical Keyboard Performance***, Music Perception, 15:2 (1997:Winter) p.153

rate domains of application: (1) understanding the processes by which altered feedback (in particular, DAF) causes an impairing effect and (2) using the altered-feedback results to try to understand the role of feedback in normal musical performance.

With respect to the general understanding of DAF, the data provide evidence against two hypotheses provided in the speech literature. The DAF effects cannot be explained simply by an unelaborated theory of distracting effects of unexpected feedback (e.g., Borden, 1979) or by a theory that ignores the identity of the delayed units (Howell et al., 1983). Clearly, instrumental performance and speech have many differences with respect to altered feedback (two of the most salient being the existence of immediate bone-conducted feedback in speech, and the polyphonic nature of keyboard performance), and whether the same explanation should apply to both is an open question.

With respect to feedback and music performance, perhaps the main result is simply the demonstration that the appropriate question is not "what is the role of feedback in musical performance," because different components of feedback (pitch and timing information) may have different effects. This result could be linked to those results in music perception that suggest that pitch and timing information are processed somewhat independently (e.g., Palmer & Krumhansl, 1987), although such a hypothesis is tentative because of differences between performance and melodic perception and because of the different senses of timing involved in the two domains of study.

A number of directions for future research are possible. The study of units of motor planning in musical performance (Palmer & van de Sande, 1995) could be pursued by observing the effects of occasional feedback disruptions as affected by their placement within hypothesized planning units (see also Wing, 1977). The extent of impairment under various feedback conditions for musicians of different skill levels and backgrounds (of a wider range, and more carefully controlled than was possible here) might provide evidence for differences in the internal representation of motor patterns and expected sounds for skilled and unskilled musicians. Overall, the ability to manipulate feedback in real time with MIDI computer systems seems to provide a flexible system for many areas of investigation involving feedback, both in music psychology and in the wider area of serial behavior studies.[6]

---

[6] Thanks to Jim Anderson, David Ascher, Peter Eimas, Bruce Kay, Caroline Palmer, Bruno Repp, Alex Staller, Gary Strangman, Mike Tarr, Bill Warren, and Todd Winkler for their assistance, and to E. C. Carterette and two anonymous reviewers for their extensive and helpful comments.

FINNEY, STEVEN A., ***Auditory Feedback and Musical Keyboard Performance*** , Music Perception, 15:2 (1997:Winter) p.153

References

Banton, L. J. (1995). The role of visual and auditory feedback during the sight-reading of music. *Psychology of Music, 23,* 3–16.

Borden, G. J. (1979). An interpretation of research on feedback interruption in speech. *Brain and Language, 7,* 307–319.

Chase, R., Harvey, S., Standfast, S., Rapin, I., & Sutton, S. (1961). Studies on sensory feedback I: Effect of delayed auditory feedback on speech and keytapping. *Quarterly Journal of Experimental Psychology, 13,* 141–152.

Gates, A., & Bradshaw, J. L. (1974). Effects of auditory feedback on a musical performance task. *Perception and Psychophysics, 16,* 105–109.

Gates, A., Bradshaw, J. L., & Nettleton, N. C. (1974). Effects of different delayed auditory feedback intervals on a musical performance task. *Perception and Psychophysics, 15,* 21–25.

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

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., El-Yaniv, N., & Powell, D. J. (1987). Factors affecting fluency in stutterers. In H. F. M. Peters & W. Hulstijin (Eds.), *Speech motor dynamics in stuttering.* New York: Springer-Verlag.

Kalinowski, J., Armson, J., Roland-Mieszkowski, M., Stuart, A., & Gracco, V. (1993). Effects of alterations in auditory feedback and speech rate on stuttering frequency. *Language and Speech, 36,* 1–16.

Large, E. W. (1993). Dynamic programming for the analysis of serial behaviors. *Behavior Research Methods, Instruments, and Computers, 25,* 238–241.

Lashley, K. (1951). The problem of serial order in behavior. In L. A. Jeffress (Ed.), *Cerebral mechanisms of behavior.* New York: Wiley.

Lee, S-H. (1989). Using the personal computer to analyze piano performance. *Psychomusicology, 8,* 143–149.

MacKenzie, C. L., & Van Eerd, D. L. (1990). Rhythmic precision in the performance of piano scales. In M. Jeannerod (Ed.), *Attention and performance XIII.* Hillsdale, NJ: Lawrence Erlbaum.

Palmer, C., & Krumhansl, C. (1987). Independent temporal and pitch structures in determination of musical phrases. *Journal of Experimental Psychology: Human Perception and Performance, 13,* 116–126.

Palmer, C., & van de Sande, C. (1993). Units of knowledge in music performance. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19,* 457–470.

Palmer, C., & van de Sande, C. (1995). Range of planning in music performance. *Journal of Experimental Psychology: Human Perception and Performance, 21,* 947–962.

Smith, K. U. (1962). *Delayed sensory feedback and behavior.* Philadelphia: W. B. Saunders.

Van Riper, C. (1982). *The nature of stuttering.* Englewood Cliffs, NJ: Prentice-Hall.

Wing, A. M. (1977). Perturbations of auditory feedback delay and the timing of movement. *Journal of Experimental Psychology: Human Perception and Performance, 3.1,* 175–186.

Yates, A. (1963). Delayed auditory feedback. *Psychological Bulletin, 60,* 213–232.