Real-Time Visual Feedback in Music Pedagogy:

Do different visual representations have different effects on learning?

Alex Brandmeyer

Master’s Thesis Cognitive Psychology
Radboud University Nijmegen
Supervisor: Peter Desain
August 2006
Table of Contents

Introduction… 3
Methods… 6
Results… 14
Discussion… 19
Conclusion… 27
Figures… 29
References 45
Introduction

Computers are a ubiquitous presence in our lives. From home to work to school, they are an integral part of daily life for many people. Education is one area in particular where computers have been present for a long time. In Music Education, there exist commercially available programs which provide exercises and which perform automatic analysis of performances. However, the majority of music instruction is still based on the master-apprentice model, in which a student works with a more experienced teacher to improve their performance abilities (Welch 1985). During instruction, students are required to play material, or to imitate material played by their teacher. Following the student’s performance, the teacher gives the student verbal feedback on their performance. Is it possible that other types of feedback employing the analytical resources of computers could be utilized to improve a student’s performances, and if so, what type or types of feedback?

Welch describes the feedback as knowledge of results (KR). He proposes that the KR provided by verbal feedback has two disadvantages when compared with real-time visual feedback, as shown in figure 1:

1. the KR is delayed in time from the actual performance, so the student must recall completely from memory what they have done.

2. the verbal feedback can be sometimes abstract and difficult to understand, with musical metaphors being a typical means of describing performances examples.

While the teaching methods currently employed are obviously effective, it seems logical that the addition of real-time visual feedback systems to the current instruction situation could provide a form of clarification and reinforce the perception of different aspects of a musical performance during the critical period when that information is most useful to the musician.
The use of visual representations and cues to aid musicians is by no means a novel concept. The original music recording medium, the musical score, has been used to help many a musician compose or learn a new song. The conductor’s waving baton keeps the orchestra in time. Computers and visual displays of all types are used to analyze and produce sounds, providing vital information about different aspects of sounds for producers and researchers through different representations such as spectrograms and waveform images, as well as visual displays of midi data and effect parameters. The physical layout and coloring of piano keys, along with the appearances of many other instruments, present a visual image that reflects the underlying structure of the western music system, providing a visuospatial basis for making music with these instruments. Vision is by no means essential to musical practice, as exhibited by the long list of incredibly talented musical performers without sight. However, there are many aspects of visual perception and auditory perception that reinforce each other (Bregman 1990; Levitin, MacLean et al. 1999), and recent research has show that this is also the case in musical performance and perception (Vines, Krumhansl et al. 2005). This correspondence of our senses in the domain of music gives rise to forms and metaphors that can potentially be used in a visual feedback system for musicians.

There is an increasing body of research investigating the potential benefits of real-time visual feedback in the setting of music instruction (Welch, Howard et al. 1989; Rossiter and Howard 1996; Jordá 2003; Wilson, Thorpe et al. 2005). These studies have been based predominantly on the instruction of singing, but the idea naturally extends to all forms of music instruction. Although some of the visual feedback interfaces have included many different parameters for visualization, the primary parameters for visualization have tended to be pitch (F0) and spectral content. While the earliest study lacked strong methodology, making its result less conclusive, the more recent study by Wilson, Thorpe, et al. did conduct a quantitative assessment of the effects of visual feedback on singing performance, finding that performance accuracy improved with the use of visual feedback. Additionally, they found that different visual
representations had differing effects with students of different skill levels.

A recent review has been conducted on the existing research on visual feedback for musicians (Hoppe, Sadakata et al. 2006). These authors conclude that additional quantitative studies are needed to evaluate specifically how visual feedback affects performance accuracy and learning for musicians, and what types of visual feedback are most beneficial for specific tasks. Studies on the use of feedback with instruments other than voice are limited. Therefore, methodical investigation of the use of visual feedback for different musical instruments and with different visual representations may provide additional support for the notion that visualization of musical parameters can aid music students and their instructors.

This study investigates the role of visual feedback in a musical imitation task for percussionists. We asked conservatory level drum students to perform repeated imitations of several drum patterns, during some of which they received one of two types of visual feedback designed for this study: either holistic, or analytic. It is hypothesized that the performance accuracy and rate of improvement of performance accuracy will be higher across subjects in the visual feedback conditions than in control conditions. We also hypothesize that the information provided in the holistic visual feedback will lead to the highest performance accuracies and biggest improvements in accuracy out of the three feedback conditions.

Because of the close relationship between this study and music education practices, specifically for percussion, and because the author and the research team are by no means experts on drumming, we chose to work with a professional drum instructor from the Amsterdam Conservatory in the design of the task, as well as the selection and recording of the target materials for the task. The materials are based on 3 different “expressive” renditions of 2 basic beat patterns. Musical expression is an elusive concept that has been widely studied by musicologists and music psychologists (Lerdahl and Jackendoff 1983; Clarke 1985; Repp 1998; Timmers 2003; Windsor, Desain et al. 2006). It is closely related to concepts of “style” and “feeling,” and generally has to do with deviations in timing, tempo, and dynamics from the strict
interpretation provided by a musical score. These three “styles” or “expressions” were provided by the drum teacher, and were familiar to all but 2 of the subjects before the test. Additionally, ideas about what forms of visual feedback to use were discussed with him, with the basic concepts for the two forms of visual feedback coming from those conversations. It is hoped that this collaboration gives strength to the methodology.

**Methods**

**Subjects**

The participants in the study were 18 conservatory level drum students, 12 from the Koninlijk Conservatorium in the Hague, Netherlands, and 6 from the Muziek Conservatorium van Utrecht. They had an average age of 22.4 (St. Dev. = 5.5 years), and an average of 11.8 years of experience playing drums (St. Dev. = 4.36 years). The average amount of practicing time per week was 13.17 hrs. (St. Dev. = 9.57), and the average amount of time playing was 15.58 (St. Dev. = 7.88). Before the test began, the participant was provided a set of instructions describing the experimental task, the visual feedback, and the target materials, and was allowed to ask questions to the experimenter throughout the instruction period. The participant also saw examples of visual feedback, heard the target materials, and practiced the task using a beat pattern not included in the actual test. Once this was completed, the experiment began.

**Task**

The participants were asked to imitate the 6 target performances which they heard as precisely as possible. A within-subject design was used, and the experiment was divided into 3 sections, with a break in between them. In each section there was a different visual feedback condition: analytic, holistic, or control (no visual feedback). Additionally in each section, one expressive style was performed, with the 8th note performance always being played first, and the
16\textsuperscript{th} note performance coming second. Each performance was repeated for 5 trials, making a total of 10 trials in each section, and 30 trials in the whole test (see figure 2 for a schematic of the trial and experiment design). The order of the visual feedback conditions and the expressive styles were randomized and counterbalanced to make for a block design of 9 subjects (see table 1 for block design of the subjects).

In each trial, 4 bars of the current target performance were played over the speakers, and, in the \textit{analytic} and \textit{holistic} conditions, were used to generate real-time visualization of the target performance. There was then a pause for the participant, followed by a 1-bar metronome count-in, after which the imitation performance began and lasted for 8 bars. Throughout the target presentation and during the imitation performance, a metronome click was heard over the speakers by the participants at a level determined to be comfortable by each participant. Between each trial there was also a pause.

Following the experiment, the subjects were given a survey in order to gather information about their age, amount of drumming experience, and current practice habits, as well as qualitative information about their opinions of the different types of visual feedback and their familiarity with the target materials. The ease of use and the overall usefulness of the two types of feedback were both reported on a scale of 1 (not easy/not useful) to 4 (easy/very useful). This data was compiled and is reported below.

\textbf{Target Materials}

The target materials consisted of 2 beat patterns and 3 expressive styles, making for a total of 6 possible target performances. The beat patterns were 8\textsuperscript{th} note and 16\textsuperscript{th} note, and the expressive styles were on-the-beat, laid-back, and rushed (see figure 3 for scores). These patterns were chosen based on our consultation with a drum teacher at the Amsterdam Conservatory of Music, and were considered by him to be appropriate for practice by conservatory level drum students. The patterns all had a beat structure which repeated every half-bar. The materials were recorded in both MIDI and digital audio (44kHz, 16-bit stereo) formats by the same teacher using
the same drum kit and configuration as the subjects (see Technical Implementation below) at 82 beats per minute. During the recording, the teacher was given a metronome click over a pair of Sony MDR-600 headphones. Each target performance was recorded for 8 bars two times making 32 half-bar segments of each performance beat pattern. The drum teacher was satisfied that the quality of the performances and recordings were adequate for instructional use.

**Visual Feedback: Analytic**

*Analytic* visual feedback (see figure 4) consisted of a display similar to a musical staff consisting of 3 horizontal lines, and vertical lines marking the metronome positions on bars and quarter notes. At the beginning of each trial, the staff appears on the right of the screen and scrolls to the left. Each of the three drum voices is represented by a different shape (Kick Drum = Square, Snare Drum = Circle, Hi-Hat = Triangle). The Kick always appeared on the bottom line, the Snare in the middle, and the hi-hat on the top, in a similar way to common drum notation. Whenever a note is played, it appeared on the right of the screen and scrolled to the left. At any given point, you are able to see the last two bars of the performance. The horizontal position of the different notes represents their relationship in the time dimension. The size of each shape varied with the loudness of the note it represented, growing larger with an increase in loudness. Thus, different accent structures on the hi-hat produced unique visual patterns amongst the 6 performances, while the repeating structure of the bass and the snare drums was visually the same across the performances (see figure 4b).

During the presentation of target materials, the target pattern appears on screen in real-time along with the audio. During response periods, the current target pattern appears in the background as grey shapes, while the responses appear in the foreground as colored, transparent shapes. If the timing and the loudness of the target and response are perfectly matched, the grey shapes are perfectly covered by the colored response shape. If there are discrepancies between
either the timings or the amplitudes, then the amount of difference can be seen by the differences in the overlap of the targets and responses (see figure 4c).

The design of the analytic feedback was motivated by two sources. As mentioned, it bears some similarities to a musical score and to drum notation, so it utilizes visual representations already familiar to most musicians. Secondly, in our discussions about the experiment with the teacher, he suggested that this type of display would provide and intuitive means of displaying timing and dynamics information about a drum performance in real-time.

Visual Feedback: Holistic

The holistic feedback used 3 shapes representing each of the expressive styles to show the expressive quality of the current performance (see figure 5a). Each shape was defined by 8 2-dimensional vectors (see figure 5b). With each incoming note, an analysis was done of the current performance. The analysis returned 4 probabilities summing to 1: on-the-beat, laid-back, rushed, and beginner (for details see Analysis below). The first 3 probabilities were used to define a weighted combination of the 3 shapes (see figure 5c for examples), which was done by multiplying the vectors by the weighted probability, and then summing the vectors together for the 3 shapes. The beginner probability was mapped onto the shape’s size: as the beginner probability grew, the size of the shape shrunk. An exact reproduction of the target performance will produce the same shape.

During the target presentation, the shape of the target appears on screen. During the imitation, the target shape appears with a grey color in the background while the imitation shape appears in the foreground with a color and a transparency. As the 4 probabilities change with each incoming note, the shape on the screen changes as a weighted mixture of the 3 shapes plus the size changes from the beginner probability. Once again, the amount and differences in the overlap of the two shapes represents the success of the imitation.
The holistic feedback was motivated by the hypothesis that providing direct feedback on higher level features like “style” and “expression” would be more relevant to the task, and would lessen the cognitive load of the subjects. By using one simple shape, rather than many, as in the analytic feedback, it was believed that the strain on the working memory of the subjects would be reduced, allowing them to effectively perform the task and obtain the benefits of the visual feedback simultaneously.

**Technical System:**

Previously, much research on expressive musical performance has been done using MIDI data recorded from different sources. A good example is the Yamaha DiskKlavier grand piano, which is a traditional grand piano that has additional optical sensors that generate accurate MIDI data about the dynamic and dynamics of the key presses. In the setup of the current system, we wanted to use real, acoustic drums, as opposed to one of the widely available MIDI drum kits. The reason is that the “feeling” of playing on the MIDI drum pads differs significantly to the “feeling” of playing on an acoustic kit, and is disagreeable to most percussionists. For the purposes of this study we developed a solution of using contact microphones to generate peaks as the drums are struck. This is similar to many commercially available triggers that can be fitted onto a standard acoustic drum-head and used to trigger a MIDI drum module.

The system used to run the experiment consisted primarily of a PowerMac G5 and a Mapex 5-Piece fusion style drum set. The kick drum, snare drum, and lower hi-hat cymbal had piezzo contact microphones mounted specifically for use as triggers. The output of each trigger was connected to a MOTU 828mk2 USB2 audio interface attached to the PowerMac. A pair of Bruel and Kjaer condenser microphones and a separate MOTU 828 interfaces were used to make digital audio recordings of the target performances. During the task, audio was presented over studio monitor speakers through a Rotel amplifier, also via the MOTU 828mk2 interface. A second monitor for displaying instructions and visual feedback was connected to the PowerMac,
and placed on a table just behind and above the tom drums, in an easily viewable position at
about eye-level for the subjects seated at the drum set. Logic Express 7.1 was used as a sequencer
to present MIDI and audio stimuli, as well as to record the performance data of the subjects and
the target performances of the drum teacher. Max/MSP 4.5 (Puckette 1988) and the ~Bonk
external (Puckette, Apel et al. 1998) was used to detect peaks in the audio signal from the drum
triggers and generate MIDI data. Additionally, Max/MSP converted all target and response MIDI
data into XML data strings. Macromedia Flash 8 was used for presenting the real-time visual
feedback. XML Data created in Max/MSP was sent using the ~Flashserver external (Matthes
2006) to Flash in frames every 30ms, which was then used by Flash to generate the different
types of real-time visual feedback. A schematic is provided in Figure 6.

Analysis: Target Materials

Research on expressive musical performance typically uses MIDI data to calculate
different measures of the timing and dynamics of a given recording (citations). Typical measures
include inter-onset-intervals (the time between successive notes) and asynchronies (deviation
from a specified position or velocity, usually given by another note, or a “metronome” of some
kind), along with many other measures which utilize the pitch, timing, and velocity data supplied
in the innate structure of MIDI.

The MIDI recordings of the target performances were converted to text-based tables
containing the velocity (loudness information) and onset times of all notes in the performances
using POCO (Honing 1990). These text tables were then imported into the JMP statistical
package and analyzed in half bar sections corresponding to the repeating structure in the beat
patterns, each containing either 6 or 10 notes, depending on whether the pattern was 8th Note or
16th Note respectively (see figure 7a for a diagram of note positions).

As a starting point for our analysis, we defined several abstract measures, or “features”,
which were then expanded to cover the different permutations of the note positions in each half
bar section of the target performances (see figure 7b for a list of definitions for the abstract features). A total of 100 features for each of the different 8\textsuperscript{th} note performances and 241 for the different 16\textsuperscript{th} note performances were derived. The means and standard deviations of each of these features were calculated based on the 32 half bars sections of each of the 6 target performances recorded by the drum teacher. Using these means and standard deviations, a normal distribution was assumed, and a probability density function was calculated for each feature in each of the 6 possible targets (see figure 7c for formula). Then, a measure of the separation of the features between the 3 styles was taken for both the 8\textsuperscript{th} and the 16\textsuperscript{th} note groups, based on the difference between the sum of the integrals for each of the 3 distributions, and the maximum of the 3 integrals for any given $x$-value (see figure 7d for formula). The motive in measuring the separation of the probability density functions between the 3 styles was to identify features that had distinct values for each style, and thus, features that could be used to identify the expressive quality of the performances. See figure 8a for examples of the resulting density functions and their separation between styles. The features that had the highest separation values for both patterns intuitively corresponded with what was known in musical terms about the accent structure of the hi-hat notes, and it’s relationship to the dynamic and timing relationships in the recorded data. We made a selection of 8 features, 4 for timing and 4 for dynamics, having the highest separation between the 3 expressive styles (see figure 8a for examples of the selected features and their probability density functions). Due to the different number of notes in the 8\textsuperscript{th} and 16\textsuperscript{th} note patterns, the 8 selected features ended up being slightly different for each set.

The target probability measures focused on the expressive aspects of the performances, but did not say anything about the level of expertise of those performances. A performance could exhibit the proper expressive style, but might not be a skilled performance. So, a similar technique was utilized to compare the performance of an expert with that of a novice. For the 8\textsuperscript{th} note and 16\textsuperscript{th} note on-the-beat patterns, two beginning drummers (< 6 months experience) performed 16 bars, giving 32 half bar sections of beginner material. The means and standard
deviations were calculated for all of the possible features, from which the probability density functions were derived. The separation of the beginner density functions from those of the target on-the-beat (expert) patterns was measured, and a distinct set of 8 features with the highest separation was selected (see figure 8b for example features and plots of the probability density functions).

From the 8 selected style features and the 8 selected expert features, we created algorithms to calculate an overall probability that a given performance is on-the-beat, laid-back, rushed, and novice. These probabilities are based on the average of the normalized densities for each of the selected features. The 4 probabilities always sum to 1. These probabilities were calculated for all the half bars recorded by the teacher. The half bar with the highest probability for each style was selected and used as a loop in presenting the target patterns to the subjects.

The same probability measures were derived from the incoming subject performances and used to generate the 4 probabilities for the holistic feedback. Following the tests, the MIDI recordings of the subject performances were treated and analyzed in a similar way to the target recordings. MIDI recordings were imported into POCO and converted to text files containing information about the timing, velocity, subject, pattern, style, and trial number of the performances. These were imported into JMP and analyzed using the same techniques employed with the target materials. Half bars containing extra or missing notes were excluded from the analysis. A total of 8122 half bars out of 8640 recorded half bars (94%) were included in the analysis, with the rate being above 92% for all three feedback conditions, 91% for the two beat patterns, and 92% for the three different styles, indicating that there was no significant imbalance in the resulting data set.

**Analysis: Imitations**

To assess the quality of the imitations, we looked at several measures: the probability that a given half bar was the target (on-the-beat, rushed, or laid-back, depending on the expressive
style of the target performance in that trial), the probability that a given half bar was a \textit{beginner} performance, an average difference measure of the raw timing between imitation and target performance notes, and an average difference measure of the raw velocities between imitation and target performance notes. The target probability measure reflected an overall average of the features that most distinctly identified the expressive performances. The \textit{beginner} probability was an indication of how close a performance was to an expert as compared with a beginner, without any indication of the expressive quality of the performance. Finally, the raw difference measures for timing and velocity provide the most direct means of comparing the data of the target and imitation performances, but do not necessarily capture higher-level features such as the expressive or expertise level of the imitations.

Using the averaged results of the 4 above measures, effects of condition and trial number within each of the 3 blocks were examined to determine whether or not the different visual feedback conditions provided any significant benefits in regards to overall performance within a block, as well as to the rate of improvement across the 5 trials within a block. This was done by calculating a two-way ANOVA for each of the measures, and checking the effects, as well as the interaction, of feedback condition and trial number. When significant effects were found, a pairwise comparison was conducted in order to evaluate the relationships within the independent variables.

\textbf{Results}

\textbf{Quantitative Results: Targets}

The expert performances given by the teacher were analyzed using the same probability used to assess the imitations described in the analysis section. Ideally, the probability of the target performances being either \textit{on-the-beat}, \textit{laid-back}, or \textit{rushed} should be 1 for the performances of each style, and the probability of the target performances being \textit{beginner} should be 0. However,
due to the nature of the probability density functions and the normalized averaging method which formed the bases of the probability calculations, as well as to the fact that there were inherent variations between the 32 half bars recorded for each of the 6 target performances, this is never the case.

Overall, the average target probability of the 8th note target performances was 0.73 (St. Dev. = .11), while the average target probability of the 16th note target performances was 0.68 (St. Dev. = 0.10). The average beginner probability for the 8th note target performances was 0.13 (St. Dev. = 0.05), while for the 16th note target performances it was 0.14 (St. Dev. = 0.03). For the performances recorded by the beginning drummers, the beginner probability was 0.42 (St. Dev. = 0.05) for the 8th note pattern, and 0.38 (St. Dev. = 0.05) for the 16th note pattern. The target probability of the beginner performances was 0.23 (St. Dev. = 0.12) for the 8th note patterns, and 0.38 (St. Dev. = 0.07) for the 16th note patterns. These numbers give us a relative indication of the success of the imitations when looked at using the target and beginner probability measures. The remaining probabilities for each of these performances belonged to the two expressive styles that were non-targets.

**Quantitative Results: Imitations**

The imitations of the subjects were evaluated using the 4 measures previously described in the *Methods-Analysis* section above: the target probability, the beginner probability, an average measure of the raw timing deviations from the means of the corresponding target performances, and an average measure of the raw velocity deviations from the means of the corresponding target performances.

The target probability measure revealed the most significant positive effect of visual feedback condition, in favor of the holistic feedback. The averaged probabilities were significantly lower than the average target probabilities of the 8th and 16th note target performances, but were significantly higher than the target probability of the 8th note beginner
performances (they were very close to the average target probability of the 16th note beginner performance). The differences of the averages between the 3 conditions were not very large, but were found to be statistically significant. The overall average was highest for the holistic visual feedback condition ($M = 0.376$, St. Dev. = 0.1535), followed by the no-feedback condition ($M = 0.3486$, St. Dev. = 0.1571), and lastly for the analytic ($M = 0.3408$, St. Dev. = 0.1574) condition. Additionally, the imitations exhibited overall improvement from trial 1 to trial 5 in all conditions. A two-way ANOVA revealed a significant effect of trial [$F(4, 7556) = 7.44, p<.0001$], as well of condition [$F(2, 7556) = 4.1041, p = 0.0165$]. Planned pair-wise comparisons indicated that there was no significant difference between the analytic and no-feedback conditions, but there was a significant difference found between them and the holistic feedback condition (Tukey, $p = 0.0161$). No significant interaction between trial and condition was observed [$F(8, 7556) = 0.9016, p = 0.5139$], indicating that no one condition offered a significantly different benefit to the rate of improvement on the probability measure from the beginning to the end of the block. See figure 9 for charts of these results. Thus, when taking the target probability as a measure of the success of the imitations, the hypothesis that holistic feedback will lead to the highest rate of performance was confirmed. However, this is not the case for visual feedback in general, as the performance in the analytic feedback condition was not significantly different from that in the no-feedback condition. Additionally, our hypothesis that visual feedback would lead to significant differences in the rate of learning was not confirmed.

In the case of the beginner probability measure, a lower score indicates a performance that is closer to the expert performance than to the beginner performances. In all three conditions, the averages were within a range of 0.4%, suggesting that there was not a truly meaningful advantage of any feedback condition. Marginally significant effects of trial [$F(4, 7556) = 2.7809, p = .0253$], as well of condition [$F(2, 7556) = 3.5012, p = 0.0302$] were observed, but it is stressed here that the differences between conditions were of such a low percentage that they are not considered meaningful. A pair-wise comparison indicated that the differences between the
holistic and no-feedback conditions were marginally significant (Tukey, \( p = 0.0302 \)). The overall average was lowest for the no-feedback condition (\( M = 0.3400, \) St. Dev. = 0.068), followed by the analytic condition (\( M = 0.3413, \) St. Dev. = 0.061), and lastly for the holistic (\( M = 0.3441, \) St. Dev. = 0.067) condition. Additionally, there was a marginally greater performance in the 5th trial of the performances than the 3rd, as indicated by a pair-wise comparison (Tukey, \( p = 0.0253 \)), but with no clear trend from the beginning to the end of the trials in any condition, and no other significant differences between individual trials revealed. As indicated by a two-way ANOVA, no significant interaction between trial and condition was observed \( [F(8, 7556) = 1.7109, \ p = 0.0905] \). These results are shown in figure 10. This means that no one condition offered a significantly different benefit to the rate of improvement on the probability measure from the beginning to the end of the block. Thus, when using the beginner probability as a measure of performance, none of the hypotheses were confirmed. The relationship between this finding and the explicit instructions given to the subjects will be discussed below.

We also analyzed the imitation performances using a measure based on the means of the raw timing values of each note in the target performances. This was a measure of the averaged differences of the 6 or 10 notes in each half bar of the imitations from the mean timing values in the notes of the target performances. The overall means between the 3 conditions were within a range of 5 ms, indicating that the performance on this measure was very closely matched between the 3 conditions. The means, in order of smallest averaged difference over notes to the highest by condition, were no-feedback (\( M = 29ms, \) St. Dev. = 17ms), analytic (\( M = 31ms, \) St. Dev. = 18ms), and holistic (\( M = 34ms, \) St. Dev. = 19ms). Significant effects of trial \( [F(4, 7556) = 7.44, \ p < .0001] \), as well of condition \( [F(2, 7556) = 4.1041, \ p < .0001] \) were observed. There was a clear improvement from the first trial to the last, as indicated by a planned pair-wise comparison. A separate planned pair-wise comparison indicated a significant between the holistic condition and the other two, but not between the analytic and the no-feedback conditions (Tukey, \( p <
The interaction between Trial and Condition was not significant \( F(8, 7556) = 1.807, \ p < .0708 \). Results for the raw timing difference measure are shown in figure 11.

The final measure of performance quality which was examined was the average across the notes within a half bar of the raw velocity differences (on a scale of 0 to 1) between the imitations and the means of the target performances, similar to the raw timing measure. Between the three conditions, there was an average overall difference of less than 1%, indicating, again, that performance on this measure was closely matched between the three feedback conditions. The means, in order of smallest averaged difference over notes to the highest by condition, were no-feedback \((M = 0.179, \text{ St. Dev.} = 0.06)\), holistic \((M = 0.182, \text{ St. Dev.} = 0.05)\), and analytic \((M = 0.187, \text{ St. Dev.} = 0.06)\). A two-way ANOVA revealed no significant effects of trial \([F(4, 7556) = 1.4456, p = .2161]\), condition \([F(2, 7556) = 0.3254, p = .7222]\), or interaction between the two \([F(8, 7556) = 0.8697, p = .5412]\). These results are shown in figure 12. Thus, as viewed with the velocity measure, the subjects did not significantly improve across trials, or perform better in a particular condition.

**Qualitative Results**

Subjects generally reported a higher preference for the analytic feedback than the holistic feedback. As mentioned before, scores were reported on a scale of 1 to 4. Analytic feedback had a high rating for ease of understanding \((M = 3.6, \text{ St. Dev.} = 0.5)\), and a moderately high rating for usefulness \((M = 2.89, \text{ St. Dev.} = 0.75)\). Holistic feedback had a lower rating for ease of understanding \((M = 2.45, \text{ St. Dev.} = 1.04)\), as well as for ease of understanding \((M = 2.5, \text{ St. Dev.} = 0.86)\). 83% of the subjects preferred the analytic feedback to the holistic. Additionally, subjects reported a high rate of familiarity with the target patterns (89%) and styles (94%).

Written responses of the subjects generally reflected the ratings above. There were two interesting points that repeatedly came out of their written responses: 1) Subjects found that it was difficult to understand the relationship between their performances and the responses of the
holistic feedback. In particular, they were unable to understand the correlation between the size of the shape (beginner probability) and their performance. Additionally, many subjects reported that they felt they would have understood the holistic feedback if they had more time to practice with it. In general, subjects felt that they would have benefited from the feedback more if they had more time and exposure to the feedback with the different patterns and styles.

In their general comments, the subjects had very mixed opinions about the visual feedback. Some felt that the visual feedback was a distraction from what was essentially a task involving audition and motor skills. On the other hand, many subjects felt that the visual feedback was quite helpful, and that, if it were available in a school- or home-setting, they would enjoy being able to use the visual feedback for practicing. In fact, many subjects said that they enjoyed using the visual feedback, and that it added an additional challenge and motivation to performing the target materials which would not have been present in a task that did not utilize real-time visual feedback.

Discussion

The most interesting finding in this study is that the effectiveness of the visual feedback depends on which measure of performance quality you are looking at. This raises questions about what each measure actually indicates, as well as the best way to quantitatively evaluate the ‘quality’ of musical performances and musical expression. One motivation in choosing the stimuli that were used was to find a balance between materials that were not reduced to the point of losing their musicality, while being simple enough to facilitate quantitative analysis. Thus, the two beat patterns had a repeating, isochronous structure in the bass and the snare drums, were of a short duration (one half-note in total), and differed only in the length and number of hi-hat notes (4 8\textsuperscript{th} notes vs. 8 16\textsuperscript{th} notes). Even with this simple structure of 6 or 10 notes in a pattern, the
number of potential features/measures for evaluation was still very large (see the *Analysis* section).

This preponderance of potential features was the main motivation for calculating the probability density functions for the candidate features, and determining which features had the strongest separation in value ranges for the expressive styles, as well as between the target performances and the beginner performances that were recorded. Additionally, and as mentioned before, the features that had the highest resulting separations in expressive style intuitively corresponded with the differences in accent structure between the expressive variations. Similarly, the features that were specifically selected for calculating the *beginner* probability had to do with overall consistency in the timing and dynamics of the bass and snare drum notes. This consistency is something that you would expect from a drum teacher with several decades of drumming experience, but not from a novice drummer with only a few months of playing and no formal instruction. Therefore, based on the preliminary analysis of the target performances, and on the close correlation of the resulting features with what was intuitively known about the performances themselves, it was determined that the target probability measure and the beginner probability measure were adequate for use in generating the *holistic* feedback, as well as for use as measures of the quality of the imitation performances.

The additional calculation of the raw timing and velocity measures was done because of the fact that the probability measures were an abstraction from the data, and were therefore indirect measures. Another consideration was that the values resulting from the probability measures we derived were not perfect; that is to say that the target *on-the-beat* pattern did not have an *on-the-beat* probability of 1, the *beginner* performances did not have a *beginner* probability of 1 (in fact it was much lower), and so on. While it was felt the results of the analysis were adequate for the purposes of the *holistic* visual feedback, a difference measure of the basic performance parameters between the target and imitation performances was a more direct and traditional means of measuring the quality of the imitations. Calculating these ‘direct’ measures
of the raw timing and velocity differences between the targets and imitations provides a means to evaluate their correlation with the more indirect probability measures.

However, the relationship of the two types of visual feedback with the two probability measures and the two raw measures is not the same. Whereas with the holistic feedback we have a direct display of the target probabilities and the beginner probability, the analytic feedback provides indirect access to the features that these measures are based on. This is not to say that the timing and dynamics features that the probabilities calculations were based on (namely, the velocity and timing proportions of the hi-hat notes for the target probabilities, as well as those of the bass and snare drums for the beginner probability) were not apparent in the analytic feedback. To the contrary, the visual patterns of the hi-hat shapes, their relationship to the metrical grid which was displayed, and to the timing and velocity relationships between the snare and bass drum shapes were based on the same data from which these features were calculated. On the other hand, the holistic feedback was solely based on the probability measures, and didn’t display any other direct information about the timing or velocities of the performed notes. With these considerations in mind, we would expect the results of the different measures to reflect this in the visual feedback conditions.

Indeed, we see this most strongly in the results of the target probability measure. The holistic feedback condition resulted in significantly higher overall target probabilities than the other two conditions, with no significant difference between the analytic and no-feedback conditions. Even though the probability calculations were based on 16 different features in total, the subjects were able to successfully use the information provided by the feedback to attain a higher overall performance in that condition. Although there was no significant interaction between the condition and trial in the two-way ANOVA conducted on the target probabilities, the average starting target probability (trial 1) and ending target probability (trial 5) were the highest in the holistic condition, and exhibited the largest overall increase (4.7% in the holistic condition followed by 4% in the no-feedback condition), suggesting that the holistic feedback not only
provided benefits starting from the very first trial, but also did have a slight advantage when considering the rate of learning in the different feedback conditions.

However, the *holistic* feedback did not provide the same benefit for the overall performance in terms of the *beginner* probability. In fact, overall performance in regards to the *beginner* probability was the lowest of the three conditions for *holistic* feedback, even though the increase in *beginner* probability was directly visualized as a decrease in the size of the imitation shape during the imitation performances. Firstly, it must be emphasized that the differences between the conditions were so small (0.4%) that it cannot be truly said that one visual feedback condition improved performance on this measure in a convincing way. The significant difference between the *holistic* and *no-feedback* conditions is partly due to the very large size of the data set (8122 data points), which allows for even miniscule differences to become significant.

A second consideration regarding the *beginner* probability is that the instructions given to the subjects did not give any detailed information about the basis of the beginner probability. The instructions given to the subjects on the *holistic* feedback read as follows:

“If your performance resembles a style, the visual representation will show the shape belonging to that style. In addition, your level of expertise will be shown by the size of the shape: the shape will become larger with improvement, and will get smaller if for some reason your performance gets worse.”

Whereas the three expressive styles had been shown to the subjects in the form of musical scores, and had been presented as audio examples, no details about what constituted ‘expertise’ were given to them. This lack of clarity manifested itself in the form of qualitative feedback from the subjects, both written and verbal, that they did not understand the connection between what they were doing and what caused the size of the shape to increase or decrease. In order not to bias the results of the experiment, the experimenter did not provide any additional indication of what specific features or performance attributes determined the beginner probability. Therefore, in the *holistic condition*, it may make sense that the *beginner* probability was slightly less, as the
subjects may have in fact altered aspects of their performance in trying to discover what made the shapes change in size.

Although it is surprising on the surface that the feedback condition which led to the highest overall performance on a measure of the target style probability had lower overall ratings for helpfulness and ease of use by the subjects than the other visual feedback condition, consideration of the specific qualitative feedback given by subjects in conjunction with the ambiguous results on the beginner probability measure sheds some light on this discrepancy. If we consider the fact that the emphasis of the task was to perform two beat patterns in three different expressive styles, and that the focus of the instructions was making these differences in expression clear, then it makes sense that the feedback which provides a direct indication of the expressive style of the imitations would lead to the highest overall success on a measure of expressive style. The goal in making the holistic feedback was to create a form of visual feedback which represented information about the goal of a performance task, i.e. performing a drum pattern with a specific expressive style, without displaying any additional/unnecessary information. From this standpoint, the visual feedback was successful.

The analytic feedback, on the other hand, did not lead to higher results than the other conditions on any of the 4 measures. This is very interesting when considering that it was generally rated highly by subjects in terms of ease of understanding and helpfulness. One possible reason for the high ratings was that the analytic feedback employed visual analogies to a musical score (left to right progression of time), as well as cognitive metaphors relating size and loudness (big = loud, small = quiet) which would be intuitively obvious to most, and which we could speculate have a basis in our intuitive grasp of the physics of sound (big things make louder noises). As such, the subject may have more quickly grasped the direct connections between their performances and the real-time visual feedback that they were presented with.

However, one thing to consider here is the relationship between the analytic feedback and Cognitive Load Theory. Because of the large number of shape on the screen at any given time
during the analytic feedback condition (80 shapes in the 16th note imitation conditions), there was no way subjects could attend to all objects on screen. Even attending to the most recent half bar during imitations required the subjects to scan between 12 and 20 shapes. If the goal of the task is performing a repeating half bar pattern in a specific expressive style, then knowing if the previous half bar was in the correct style and nothing more might be more beneficial (and less distracting) than knowing the timing and dynamics details of all the notes in the previous half bar, or the previous 2 bars, as was displayed in the analytic feedback. In other words, requiring the subject to attend to detailed visual information during the musical performance, itself a complex motor task having significant demands on the subject’s attentional faculties, may in fact detract from the quality of the musical performance, due to limitations of attention.

Another striking finding is that, for both raw measures, the no-feedback condition imitations had the smallest differences from the target performances. In the case of the raw velocity measure, none of the differences were significant, while in the case of the raw timing measure, the difference between the averages of the three conditions had a range of 5 ms. While this is a very small range, it is arguable that this amount of timing difference is meaningful in the context of expressive music performance. Additionally, the subjects were instructed to imitate the target performances as “precisely” as possible, so the more timings and velocities of the performed notes resemble those of the target performances, the better the imitation. Here, we can note that the analytic and no-feedback conditions were not significantly different; so in terms of the raw measures, for analytic feedback, there was no significant difference from the no-feedback condition. Another way to put this is that, compared to no visual feedback, there was no benefit or cost of presenting analytic visual feedback, when looking at the raw differences between the target performances and the imitations.

With the holistic feedback, however, there was a significant difference in the raw timing measure from the other two feedback conditions. Why would this be, when the target probability measures used to generate the holistic feedback used 4 out of 8 features based on note timing (the
other 4 being based on the note dynamics)? The explanation is that the features used in the probability measure were based on proportional measures, and not on the raw differences themselves. Therefore, the actual dynamics values could be relatively louder or softer than the targets and still have the same proportional relationships. In the same manner, the timings could be shifted around the metronome click, and as long as the maintained the same relative asynchronies, they would end up having the same resulting target probability measures. At the same time, the raw measures of difference from the target performances could remain substantially higher.

This leads to another observation regarding the performance of the different expressive patterns. The results of the target performance analysis led to the selection of the features for calculating the three style probabilities. The features that were selected did not end up being the raw values of the component notes, but rather proportional measures of the timing and the dynamics of the hi-hats. This makes sense considering that the bass drum and snare drum pattern were the same for all the target performances. Additionally, the feature selection was based on the separation of the probability density functions for each of the potential features/measures that we calculated from the target performances. These probability density functions were in turn derived from the means and standard deviations of the individual target performances for each of those measures. What comes out of this consideration is that there was less deviation in the proportional measures of timing and dynamics on the hi-hats than there was in other measures, and therefore, more separation between the probability density functions for the three styles. If we think about the expressive performance of these simple drum patterns, then the most salient feature of the different styles is the relative values of the hi-hat notes, not the absolute values. As long as the laid-back pattern is played “loud-soft-loud-soft…”, and the timings are what they should be, then it is perceived as a laid-back performance by the performer and the listener. It is less important if the raw velocity values are [.8, .4, .8, .4 …] or [.9, .45, .9, .45 …], yet these two patterns will give rise to different velocity measures. In regards to the timing of accent playing for
percussionists, it has also been found that there are systematic microdeviations following loud and soft accents (Dahl, submitted). Therefore, in a similar way to the dynamics, it is the proportion of the timings of successive hi-hat notes that is most important to the perception of a performance being of a particular expressive style, as opposed to the raw timing within a half-bar window. If the goal of the task was to imitate the expressive style of the target performances as accurately as possible, then it seems that the most valid measures of the quality of the imitations are the features which are most strongly indicative of expression in the target performance. These are the very features from which the target probability measure was derived.

The relationship between the nature of real-time visual feedback and different measures of imitation quality has also come up in a separate experiment on real-time visual feedback (Hoppe, Brandmeyer, et al 2006). In this experiment subjects were required to imitate the expressive timing and dynamics of 12 different 4-note rhythms, each with 9 expressive variants. A novel form of real-time visual feedback using abstract shapes to represent the intervals, dynamics, and expressive timing of the 4 notes was utilized. The intervals were represented by the formation of a triangle by the individual shapes, while the dynamics were represented by the size/thickness of the shapes, and the expressive timing was represented by the outward bend of the shapes. The results of the analysis indicated that subjects in fact performed worse on a measure of expressive timing accuracy than a control group who did not receive visual feedback. However, the subjects who received real-time visual feedback performed better in terms of their accuracy on a measure of dynamics than the subjects who did not receive visual feedback. One way to interpret this result is that the feedback on dynamics may have been easier to interpret than the feedback on expressive timing. Additionally, attending to the visual feedback may have interfered more with the accuracy of the expressive timing than it did with the production of accurate dynamics in the imitations. While it is hard to separate these two possibilities, it is important to note that there is a congruity between this study and the present one in that the
effectiveness of the visual feedback depends on what measure is being examined, as well as what the emphasis of the task is.

A final point for discussion is the lack of significant results showing a difference in rate of learning across trials between the feedback conditions. Three out of the 4 measures of performance quality showed significant effects of trial, while in 2 of them (Target Probability and Raw Timing Difference) the $p$ value was less than .0001, with a clear trend of improvement from trial 1 to 5, indicating that the subjects did learn to imitate the target performances more accurately (in timing) and more expressively. It is possible that the short duration of the experimental conditions did not allow enough time for a significant learning effect to become apparent. In regards to the target probability and the raw timing measure, no ceiling effect was observed on the final trials, indicating that, with a longer set of trials, subjects may have continued to improve on those measures. Many of the subjects expressed that they felt they would have been able to make better use of the visual feedback if they had been able to spend more time with it. Indeed, it would be very interesting to see if there are different results in a study which tests the effects of the different types of visual feedback over several testing sessions, to see if long-term usage has distinct costs or benefits from the short-term usage explored in this study.

Conclusions

Real-time visual feedback for musicians is a research area that has seen an increased amount of interest in the past several years. While many studies have looked at the effects of visual-feedback on singing, there is not very much published data on its effects on performance with other instruments, or on what type of visual representations may be the most beneficial. This study examined the effects of two different real-time visual feedback representations on the quality of imitations of expressive percussion performances. It was found that, when looking at a probabilistic measure of expressivity, providing holistic visual feedback to subjects led to a
significant increase in overall performance compared to the analytic visual feedback, and to the control condition (no-feedback). No significant benefit to rate of learning was found, although the range of improvement was also highest for the holistic feedback condition. When looking at raw measures of timing, however, no advantage of providing visual feedback was found. No meaningful differences were found between conditions on a probabilistic measure of expertise, nor with a measure of raw velocity differences. While the lack of beneficial effects on 3 of the 4 measures may seem discouraging, it is believed that the initial measure, the probabilistic measure of expressivity, is the measure most closely related to the goal of the task: imitating the expressivity of the target performances. Therefore, it is concluded that real-time visual feedback that provides high-level quantitative information about the goal of the task (holistic feedback) can benefit learning to imitate expressive percussive performances. It is also concluded that real-time visual feedback which provides detailed information about an imitation (analytic feedback) does offer any significant benefits.

Due to the fact that no significant differences in rate of learning were found between conditions, it is proposed that further research along these lines be conducted with longer periods of use in a particular feedback condition, as well as with many sessions spread across a time-span of weeks. In this way, differences between short-term use (presented in the current study) and long-term use can be evaluated. It may be the case that long-term use of the real-time visual feedback systems is required for more significant effects on learning rate and overall performance quality to become apparent. Additionally, research on different types of musical tasks, different musical instruments, and different visual representations is still required to definitively answer the question of whether or not visual feedback could play an important role in music pedagogy.
A) Learning in traditional Master-Apprentice model

B) Learning with real-time visual feedback

Figure 1. Schematics from Welch 1985 illustrating advantages of real-time visual feedback. A) In the traditional model, the student receives a pitch model, performs an imitation, and receives knowledge of results (KR) in the form of verbal feedback. The KR is not aligned with the critical learning period (CP). B) The model incorporating real-time visual feedback provides KR in the form of a visual display which allows students to make immediate corrections, and to associate their actions with a result (changes in the visual display).
A) Trial Structure

<table>
<thead>
<tr>
<th></th>
<th>Target presentation</th>
<th>Imitation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pause</td>
<td>1 bar click</td>
<td>Pause</td>
</tr>
<tr>
<td>4 bars</td>
<td>Target performance</td>
<td>1 bar click</td>
</tr>
<tr>
<td></td>
<td></td>
<td>8 Bars Imitation</td>
</tr>
</tbody>
</table>

B) Experiment Structure

<table>
<thead>
<tr>
<th>Block 1</th>
<th>Block 2</th>
<th>Block 3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8th Note</td>
<td>8th Note</td>
<td>8th Note</td>
</tr>
<tr>
<td>Break</td>
<td>Break</td>
<td>Break</td>
</tr>
<tr>
<td>1 trial</td>
<td>2 trials</td>
<td>1 trial</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>4</td>
<td>Break</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>trials</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5</td>
</tr>
</tbody>
</table>

Figure 2. Schematic of A) task and B) experiment. A) When the subject was ready to begin a trial, they hit the snare drum. One bar of metronome was followed by 4 bars of the target audio presented over the speakers, as well as the display of the target in the visual feedback conditions. After the target had finished playing, the subject could hit the snare drum when they were ready to begin their imitation. They were given one bar of metronome for a count-in, after which they had to perform the target pattern for 8 bars. B) The experiment was divided into three sections. Each section consisted of one expressive style paired with one visual feedback condition in a randomized design (see Table 1 for Block Design structure and specific pairings). Within each section there were two parts consisting of 5 trials each. The first part used the 8th note pattern, and the second part used the 16th note pattern. Between 8th note and 16th note patterns, and between the three sections, the subjects were allowed to take a short break if they liked.
<table>
<thead>
<tr>
<th>Subject</th>
<th>Trial</th>
<th>Feedback</th>
<th>Expression</th>
<th>Pattern</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>C</td>
<td>OTB</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>C</td>
<td>OTB</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>H</td>
<td>LB</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>H</td>
<td>LB</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>A</td>
<td>Rash</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>A</td>
<td>Rash</td>
<td>16</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>H</td>
<td>OTB</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>H</td>
<td>OTB</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>A</td>
<td>LB</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>A</td>
<td>LB</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>C</td>
<td>Rash</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>C</td>
<td>Rash</td>
<td>16</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>A</td>
<td>OTB</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>A</td>
<td>OTB</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>C</td>
<td>LB</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>C</td>
<td>LB</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>H</td>
<td>Rash</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>H</td>
<td>Rash</td>
<td>16</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>C</td>
<td>LB</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>C</td>
<td>LB</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>H</td>
<td>Rash</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>H</td>
<td>Rash</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>A</td>
<td>OTB</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>A</td>
<td>OTB</td>
<td>16</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>H</td>
<td>LB</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>H</td>
<td>LB</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>A</td>
<td>Rash</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>A</td>
<td>Rash</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>C</td>
<td>OTB</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>C</td>
<td>OTB</td>
<td>16</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>A</td>
<td>LB</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>A</td>
<td>LB</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>C</td>
<td>Rash</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>C</td>
<td>Rash</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>H</td>
<td>OTB</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>H</td>
<td>OTB</td>
<td>16</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>C</td>
<td>Rash</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>C</td>
<td>Rash</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>H</td>
<td>OTB</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>H</td>
<td>OTB</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>A</td>
<td>LB</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>A</td>
<td>LB</td>
<td>16</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>H</td>
<td>Rash</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>H</td>
<td>Rash</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>A</td>
<td>OTB</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>A</td>
<td>OTB</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>C</td>
<td>LB</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>C</td>
<td>LB</td>
<td>16</td>
</tr>
<tr>
<td>9</td>
<td>1</td>
<td>A</td>
<td>Rash</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>A</td>
<td>Rash</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>C</td>
<td>OTB</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>C</td>
<td>OTB</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>H</td>
<td>LB</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>H</td>
<td>LB</td>
<td>16</td>
</tr>
</tbody>
</table>

Key:

C = Control/No-Feedback
A = Analytic
H = Holistic
OTB = on-the-beat
LB = laid-back
RSH = rushed
8 = 8th note
16 = 16th note
Table 1. Block design of subjects, materials, and conditions. The order of the feedback conditions and expressive styles were randomized into pairs in order to eliminate ordering effects. This design led to block size of 9 subjects, of which two complete blocks of subjects were tested.
<table>
<thead>
<tr>
<th></th>
<th>8th note</th>
<th>16th note</th>
</tr>
</thead>
<tbody>
<tr>
<td>On-the-beat</td>
<td><img src="image1" alt="On-the-beat 8th note" /></td>
<td><img src="image2" alt="On-the-beat 16th note" /></td>
</tr>
<tr>
<td>Laid-Back</td>
<td><img src="image3" alt="Laid-Back 8th note" /></td>
<td><img src="image4" alt="Laid-Back 16th note" /></td>
</tr>
<tr>
<td>Rushed</td>
<td><img src="image5" alt="Rushed 8th note" /></td>
<td><img src="image6" alt="Rushed 16th note" /></td>
</tr>
</tbody>
</table>

Figure 3. Scores of target performances. Three different styles of two different beat patterns were notated by the same drum teacher who provided the target performances. The different styles are defined by differing accent patterns on the hi-hat.
Figure 4. Analytic Feedback. A) The first panel shows one bar of a display of one of the performances (8th note laid-back). The bass drum is the pink square, the snare is the green circle, and the hi-hat is the blue triangle. This representation bears some analogy to a musical score. B) For each different accent pattern, the different repeating patterns can be easily discerned. C) For the imitation displays, the target appeared in the background as grey shapes, while the imitation notes appeared as colored transparent shapes on top of the targets. With the transparency effect, differences in timing and loudness (size) between the imitation and the target were easy to see. Early timing would show a given shape to the left, while late timing would show it to the right. In a similar manner, bigger shapes indicate a louder note while smaller shapes indicate a quiet note. Because timing and dynamics are directly mapped on to the shapes, the relationships between them visually display relationships such as inter-onset-interval and the proportion of the velocities.
of adjacent notes. These relationships form the basis of the direct measures used to evaluate the performances.
A) Holistic feedback target shapes

![Holistic Feedback Target Shapes]

B) Definitions for target shapes

- **On-the-beat**
  \[ \begin{bmatrix} [0, -1], [0.5, -0.5], [1, 0], [0.5, 0.5], [0, 1], [-0.5, 0.5], [-1, 0], [-0.5, -0.5] \end{bmatrix} \]

- **Laid-back**
  \[ \begin{bmatrix} [-0.5, -1], [0, -1], [0.5, 0], [1, 1], [0.5, 1], [0, 1], [-0.5, 0], [-1, -1] \end{bmatrix} \]

- **Rushed**
  \[ \begin{bmatrix} [0.5, -1], [1, -1], [0.5, 0], [0, 1], [-0.5, 1], [-1, 1], [-0.5, 0], [0, -1] \end{bmatrix} \]

C) Examples of imitation performances

![Imitation Performances]

Figure 5. Holistic Feedback. A) The Holistic feedback was based on the expressive styles we were using in the target performances. For each of the 3 styles, there was a simple target shape that was chosen to have some visual analogy to the style; the parallelograms we used for rushed and laid-back lean to the front or the back, while on-the-beat is represented by a rotated square. B) Each of the shapes was defined by a set of 8 2-dimensional vectors (points). C) During the imitations, the target shape was shown in the background as a grey shape, while the imitation was shown in front as a colored transparent shape, similar to the analytic feedback. After each note received by Flash, the 4 probabilities on-the-beat, laid-back, rushed, and beginner were all recalculated in the manner described in the “Methods:Analysis” section, and used to make a mixture of the three target shapes based on a weighted combination of their shapes, whose definitions were multiplied by their respective probabilities. Excluding the beginner probability from the weighted summation had the effect of negatively mapping it on to the size of the imitation shape. That is, as the beginner probability increased, the size of the imitation shape shrank.
Figure 6. Technical Setup. The system used three different environments to implement the experiment. Logic Express Audio 7.1 was used to record the stimuli and responses, and to sequence and playback the target materials in the structure specified in the experiment design. Max/MSP, along with two externals, was used to generate MIDI data from the drum microphones, and to communicate MIDI and XML data to Logic and Flash respectively. The ~bonk external (Puckette, Apel et al. 1998) was used to perform a peak detection in the signals from the 3 contact microphones, with the raw velocities from the peak detector being converted into MIDI velocities in Max. The MIDI from the drum kit and the target MIDI from Logic were both sent to an XML assembler which time stamped the MIDI data with millisecond accuracy, and then sent packets of XML to Flash every 30ms for rendering using the ~flashserver external (Matthes 2006). Incoming XML data in Flash was then parsed for either direct rendering (analytic feedback) or for real-time statistical analysis, the results of which were then rendered using different functions (holistic feedback).
A) Note positions within performances used to calculate features

![Note positions diagram](image)

B) Feature Definitions

- $n =$ half bar number
- $t_{ni} =$ time position within half bar of note $i$
- $v_{ni} =$ velocity between 0 and 1 of note $i$
- $s_{ni} =$ strict timing for note $i$
- $v_{ni+1} =$ velocity of note $i+1$
- $t_{ni+1} -$ $s_{ni} =$ $t_{ni}$ - $s_{ni}$
- $v_{ni} -$ $v_{ni+1} =$ velocity between notes $i$ and $i+1$
- $t_{ni+1} -$ $t_{ni} =$ time between notes $i$ and $i+1$
- $t_{ni+1} -$ $t_{ni} =$ time between notes $i$ and $i+1$
- $prop_{ji} =$ $t_{ni+1} -$ $t_{ni} =$ $t_{ni}$ / $t_{ni+1}$
- $dynamics(n)$ = $v_{ni} / v_{ni+1}$
- $asynchrony_i = timing(n+1) - timing(n)$
- $vasynchrony_i = dynamics(n+1) - dynamics(n)$

- $ioi_{4+std} = \sqrt{(ioi_{4+std} - M(ioi_{4+std}))^2} + \frac{(ioi_{4+std} - M(ioi_{4+std}))^2}{1}$

C) Probability Density Function Definition

- $d = \frac{e^{-(x-M)^2}}{std \sqrt{2\pi}}$

D) Separation Function Definition

- $Separation = \int_0^1 d_{on-the-beat} + \int_0^1 d_{laid-back} + \int_0^1 d_{rushed} - \int_0^1 \max(d_{on-the-beat}, d_{laid-back}, d_{rushed})$

Figure 7. Statistical Evaluation of Target Performances. A) Each note within a half bar section was assigned a position number to allow for the definition of features which measure the timing
and dynamics relationships between different notes. There were 6 note positions for the 8\textsuperscript{th} note patterns, and 10 note positions for the 16\textsuperscript{th} note patterns. B) Then, several abstract features expressing relationships and note values were defined. Using all different possible combinations of note positions, many specific features were then generated from the abstract definitions. For example, the abstract feature \textit{vel} \textsubscript{i} leads to either 6 or 10 features (depending on 8\textsuperscript{th} or 16\textsuperscript{th} note patterns), with \textit{i} taking the value of each possible note position. C) The function used to calculate the probability density function, also called the probability distribution, assumes a normal distribution, and takes as parameters \textit{M} (the mean of the values \textit{x} for all half bars in each pattern/style combination), \textit{std} (the Standard Deviation of the values \textit{x} for all half bars in each pattern/style combination), and the value \textit{x} itself. D) The separation measure takes the sum of the integrals of the 3 probability distributions for each style in each feature. It then calculates the difference between this sum and the integral of the Max() function on the 3 integrals. This is akin to taking the sum of the surface areas of each probability distribution, and subtracting the surface area of the 3 distributions minus any overlap in their surfaces. This gives an indication of how much overlap there is between their distributions relative to their total surface area.
Figure 8. Examples of probability distributions and separation measures. A) 2 dynamic measures, based on the formulas given in figure 7b (Veli8a = Veli_{23}, and Veli8b = Veli_{35}) and 2 timing measures (Prop8ab = Prop_{23}, and Prop8ad = Prop_{61}) used in the calculation of the target probability measure, are shown. These, along with 4 other measures (2 additional timing, and 2 additional dynamics measures), were selected because they had the highest separation between the 3 expressive styles out of the possible timing and dynamics features. The velocity proportion features for the successive hi-hat notes had the highest separation of any of the measures, corresponding with what is known in musical terms about the differences between the 3 expressive patterns; namely, the accent pattern played on the hi-hat. B) Similarly to figure 8a, 4 features selected for use in the beginner probability calculation are shown. What is interesting about these separations is that the target performances (shown in red) have a much tighter distribution than the beginner performances (shown in green). This makes sense, as we would expect the performance parameters of an expert drum teacher to be much more consistent than those of a novice drummer with no formal training. The parameters shown are, from left to right, the asynchrony of note 1 and 2 (bass drum and first hi-hat), the standard deviation of the inter-onset-intervals of the bass-to-snare and snare-to-bass drum successions, the proportion of those two intervals, and the raw velocity of the snare drum; the teacher tended to play the snare drum much more sharply, and thus with a higher velocity, than the novice drummers.
Figure 9. Target Probability Measure Results (on a scale of 0 to 1). A) Overall Means by Condition B) Overall Means by Trial within Condition
Figure 10. Beginner Probability Measure Results (on a scale of 0 to 1). A) Overall Means by Condition B) Overall Means by Trial within Condition

$p = 0.0302$

$p = 0.0253$
Figure 11. Raw Timing Difference Measure Results (in seconds). A) Overall Means by Condition
B) Overall Means by Trial within Condition

$p < 0.0001$
Figure 12. Raw Velocity Difference Measure Results (on a scale of 0 to 1). A) Overall Means by Condition B) Overall Means by Trial within Condition

\[ p = 0.7222 \]

\[ p = 0.2161 \]
References


Dahl, S.


