Psychological Research
DOI 10.1007/s00426-010-0291-6

ORIGINAL ARTICLE

Learning expressive percussion performance under different visual feedback conditions

Alex Brandmeyer · Renee Timmers · Makiko Sadakata · Peter Desain

Received: 10 December 2009 / Accepted: 29 May 2010

Abstract A study was conducted to test the effect of two different forms of real-time visual feedback on expressive percussion performance. Conservatory percussion students performed imitations of recorded teacher performances while receiving either high-level feedback on the expressive style of their performances, low-level feedback on the timing and dynamics of the performed notes, or no feedback. The high-level feedback was based on a Bayesian analysis of the performances, while the low-level feedback was based on the raw participant timing and dynamics data. Results indicated that neither form of feedback led to significantly smaller timing and dynamics errors. However, high-level feedback did lead to a higher proficiency in imitating the expressive style of the target performances, as indicated by a probabilistic measure of expressive style. We conclude that, while potentially disruptive to timing processes involved in music performance due to extraneous cognitive load, high-level visual feedback can improve participant imitations of expressive performance features.

Introduction

The impact of feedback on the performance of various tasks has been widely studied in the perceptual and motor sciences since the first half of the twentieth century (see Annett, 1969 for a review of pre-1970s literature). Also termed knowledge of results (KR) or extrinsic feedback, it appears both in research settings as well as in everyday life, and can take a wide variety of forms, including right/wrong indicators, test scores, and verbal commentary. More recent studies have demonstrated effects of various forms of feedback on the learning of complex motor tasks, such as athletic, linguistic, and musical performance (Escartí & Guzmán, 1999; Pennington, 1999; Rossiter, Howard, & DeCosta, 1996). With respect to the latter, a growing body of research has investigated the effects of real-time visual feedback (RTVFB) on pitch accuracy and voice quality in singing performance (see Hoppe, Sadakata, & Desain, 2006 for a review).

RTVFB on music performance was first proposed and investigated by Welch and colleagues (Welch, 1985; Welch, Howard, & Rush, 1989) in the context of singing with accurate pitch. It was noted that traditional verbal feedback on performance as a form of KR was subject to a time delay, thus reducing the effectiveness of the feedback (Annett, 1969; Evans, 1960). By providing a real-time visualization of pitch and/or other vocalization parameters, the time delay for the KR is removed. Findings in the above studies by Welch and colleagues, along with more recently conducted research (Thorpe, Callaghan, & Doorn, 1999; Welch, Himonides, Howard, & Brereton, 2004) have generally reported beneficial effects of RTVFB on performance accuracy.

Apart from singing, a study by Sadakata, Hoppe, Brandmeyer, Timmers, and Desain (2008) has looked at the effects of RTVFB on the expressive performance of simple rhythms. Musical expression refers to the micro-deviations in the timing and dynamics of musical notes from what is specified in a score (Palmer, 1997). The ability to perform music expressively is one of the skills which is recognized in accomplished performers

A. Brandmeyer (✉) · M. Sadakata · P. Desain
Donders Institute for Brain, Cognition, and Behaviour,
Radboud University Nijmegen, PO Box 9104,
6500 HE Nijmegen, The Netherlands
e-mail: a.brandmeyer@donders.ru.nl

R. Timmers
Department of Music, University of Sheffield,
34 Leavygreave Road, Sheffield S3 7RD, UK

Published online: 24 June 2010

(McPherson & Schubert, 2004), but which is sometimes neglected in music education practice (Person, 1993; Tait, 1992). This may be due to the difficulties inherent in trying to verbally describe specific aspects of musical performance (Hoffren, 1964; Welch, 1985), and to preconceptions about how expressive performance skills are acquired (Juslin, Friberg, Schoonderwaldt, & Karlsson, 2004). However, despite the difficulties involved in learning to perform expressively, expressive music performance constitutes a prime example of a highly refined motoric skill.

In the above study by Sadakata et al. (2008), amateur musicians were trained to imitate simple four-note patterns containing various expressive deviations from the musical scores that were provided. Half received RTVFB in the form of abstract shapes that visualized the timing of each note as curvature, and the dynamics as size. The other half received no RTVFB and served as a control. The participants also completed pre- and post-tests without any RTVFB before and after the training. Results indicated that the RTVFB was helpful for improving the accuracy of dynamic aspects of the performance, but was detrimental for the timing dimension, as indicated by smaller RMS timing error in the control group during both the training and post-test.

These results can be interpreted using Cognitive Load Theory (CLT) (Chandler & Sweller, 1991; Sweller, 1988, 1994; for a review see Paas, Renkl, & Sweller, 2003) which provides a framework for designing effective instructional materials based on the constraints of working memory. CLT identifies three types of cognitive load: intrinsic, extraneous, and germane. Intrinsic cognitive load is caused by the inherent difficulty of a given task, while extraneous cognitive load derives from the manner in which information is presented. Germane cognitive load refers to the working memory resources that are involved in learning new materials and skills in general, independent of a given task.

CLT emphasizes the limitations of working memory and attention during learning. In this regard, CLT is closely linked to research on working memory capacity and performance on tasks involving divided attention. Performance on a wide range of tasks suffers when working memory capacity has been reached, resulting from the retrieval of response tendencies that conflict with the current task (Engle, 2002). Similarly, performance in perceptual and memory retrieval tasks suffers as a result of divided attention when compared to conditions in which attention is directed (Corbetta, Miezin, Dobmeyer, Shulman, & Petersen, 1991; Craik, Govoni, Naveh-Benjamin, & Anderson, 1996). In this regard, the application of CLT to instructional materials serves to reduce extraneous load on working memory and executive attention processes during learning.

With respect to the RTVFB in the Sadakata et al. study, it may be the case that the visual complexity of the chosen representation led to a high extraneous cognitive load, due to the number of elements and parameters it contained. This is referred to in the CLT literature as high element interactivity (Sweller, 1994), and indicates that a given display or representation imposes a high load on working memory due to the processing of individual elements and their relationships to one another. Alternatively, the representation of timing may have been difficult to interpret, which also creates extraneous cognitive load. This in turn may have led the participants to focus more on dynamics than on timing.

In a study of singing, Wilson, Lee, Callaghan, and Thorpe (2008) also considered CLT with respect to different RTVFB representations, and reported that performance decreased relative to baseline during training with RTVFB, and then rebounded to a significantly higher level than baseline during the post-training assessment. The authors concluded that the decrease during training was due to a higher cognitive load on participants created by the RTVFB display.

Wilson et al. also reported that different visual representations were more or less effective depending on the musical skill level of the participant. While beginning students performed more accurately using a more detailed display representing pitch frequency contour, advanced students achieved better performance when using a simple display which represented categorical information about pitch (i.e., a keyboard display showing performed notes such as C, F#, etc.).

Research on music perception has shown that some musical features, such as rhythm, are also perceived categorically (Clark, 1987; Desain & Honing, 2003). Moreover, important work on categorical perception by Rosch (2002) has led her to propose that category systems "provide maximum information with the least cognitive effort". Thus, categorical feedback may provide the most information with the least cognitive load. However, while providing categorical feedback on pitch can be done using fundamental voice frequency, the specific sets of parameters that distinguish different categories of expressive performance, such as a "funky", "romantic", or "jazzy" performance, are not immediately given.

One approach in machine learning and perception that has been successful in complex domains is the use of probabilistic models based on Bayes' theorem. Specifically, Bayesian methods have been used successfully in tasks such as computer vision (Knill, Kerstern, & Yuille, 1996), handwriting recognition (Cheung, Yeung, & Chin, 1998) and music transcription (Cemgil, Desain, & Kappen, 2000), as well as classification of rhythm and key in music (Temperley, 2007). These methods are based on the identification of feature sets that distinguish between the categories of interest in a given domain, and that can then be

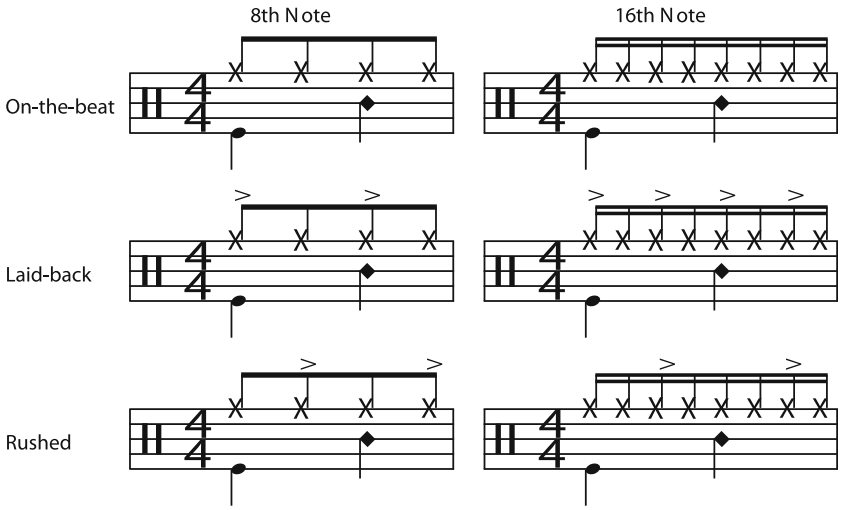

Fig. 1 Scores of target performances. Three different expressive styles for two different beat patterns were notated by the same drum teacher who performed the target materials. Each pattern contained only notes performed on the bass drum, snare drum, and hi-hat cymbal (notated from bottom to top, respectively). The different styles are defined by differing accent patterns on the hi-hat, indicated by the '>' symbols above a given note, as well as by expressive timing variations (not specified in the scores)

used as evidence in an application of Bayes' theorem. By identifying appropriate timing and dynamics features, this type approach can also be used to distinguish different categories of expressive musical performance, such as the various styles used in contemporary drumming (e.g. "laid-back", "rushed" or "on-the-beat"), or the performance characteristics of different musicians.

The present study aims to extend the above research findings on RTVFB with simple sung melodies and tapped rhythms to the domain of expressive percussion performance, and to test the effect of cognitive load on performance using two different RTVFB representations. An experiment was conducted in which advanced drum students imitated target performances by an instructor. Imitation paradigms have been used in previous studies of expressive musical performance (Clark, 1993; Repp, 2000), including those making use of RTVFB (Sadakata et al., 2008), as well as in studies on speech production (Kent, 1974; Repp & Williams, 1985, 1987). The imitations were performed in three different RTVFB conditions: low level, high level, and no feedback (control). The two RTVFB representations differ in the type of information they display, and in the number of visual elements used to provide feedback.

The first representation ("low level") displays the timing and dynamics error of each performed note, and is similar to the representation used by Sadakata et al. (2008) in that many visual elements are displayed on screen, giving it a high extraneous cognitive load. The second representation ("high level") displays categorical feedback about the expressive style and skill level of the performance, and uses only two visual elements to give feedback, thus reducing the extraneous cognitive load. It is based on the real-time output of a set of Bayesian classifiers of the expressive style and skill level of the imitation performances developed using the target performance materials from the experiment. While the primary focus of the task was on imitation of expressive performances, the classification of skill level ensured that the high-level feedback provided useful information on features of the performance not related to expressive style. The feature analysis and Bayesian formulation used in the high-level feedback are described in the Appendix.

It was expected that participant imitation performances would replicate specific timing and dynamics features used by the Bayesian classifiers to distinguish expressive style and skill level more accurately in the high-level condition than in the low-level or control conditions. Additionally, a lower overall root-mean-square (RMS) error for both timing and dynamics was expected in the high-level condition than in the low-level condition, due to a reduced cognitive load. Finally, an increased rate of improvement across trials was expected in the high-level feedback condition relative to the other two conditions.

Methods

Participants

The participants in the study were 18 conservatory-level percussion students, 12 from the Royal Music Conservatory in The Hague, Netherlands, and 6 from the Music Conservatory of Utrecht. They had an average age of 22.4, and an average of 11.8 years of experience playing drums. The average amount of practicing time per week was 13.2 h, while the average amount of total playing time was 15.6 h.

Materials

A percussion instructor from the Amsterdam Conservatory assisted in selecting two standard beat patterns for the experiment: *8th note* and *16th note*. In addition, he chose

three common expressive styles: *on-the-beat, laid-back*, and *rushed*. Notation of the materials was provided by the instructor, and is shown in Fig. 1. While expressive timing is not explicitly annotated in the musical scores, a lengthening of intervals containing accented notes consistent with previous observations (Semjen & Garcia-Colera, 1986; Dahl, 2000) was found (see Fig. 6). Using the experimental setup described below, the instructor recorded each pattern (repeated 36 times) in each of the three styles, making for a total of six different performances.

One novice percussionist (the first author) with less than 3 months experience playing drums (8 years of formal music instruction) also recorded 36 repetitions of both the 8th-note and 16th-note on-the-beat patterns. The repetitions of the instructor and the novice performances were then analyzed using the methods described in the Appendix, and a set of features was selected for use in generating the high-level feedback. From each of the instructor performances, one repetition was selected, looped for four bars, and presented as a target during the experimental trials. More details on the materials, including mean timing and dynamics profiles for the instructor performances, can be found in the Appendix.

Procedure

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 the experimenter questions throughout the instruction period. The participant also saw examples of visual feedback, heard the target materials, and practiced the task using a beat pattern (simple quarter-note) not included in the actual experiment. Once this was completed, the experiment began.

During each trial, the participants were first asked to listen carefully to the target performance. They were then asked to imitate the target performance as precisely as possible. A within-participant design was used, and the experiment was divided into three blocks, with short breaks between them. In each block there was a different visual feedback condition: low level, high level, or control (no RTVFB). Only one of the three expressive styles was performed in each block, with the 8th-note pattern always being played first, and the 16th-note pattern coming second. This means that, for individual participants, each expressive style was paired with one of the visual feedback conditions in a randomized, counterbalanced design requiring groups of nine participants.

In each trial, four bars of the current target performance were played over loudspeakers, and, in the RTVFB conditions, were used to generate a visualization of the target performance. There was then a pause for the participant, followed by a one-bar metronome count-in, after which the imitation performance began and lasted for eight bars. With the exception of the control condition, it was during this period that RTVFB was presented. Throughout the target presentation and during the imitation performance, a metronome click was heard over the speakers at a level determined to be comfortable by each participant. Participants were instructed to perform in time with the metronome. Each of the two beat patterns was repeated for five trials, making a total of ten trials in each section, and 30 trials in the whole test. Between trials there was also a pause.

Technical system

An Apple PowerMac G5 computer connected to a five-piece Mapex drum kit using piezo contact microphones placed on the drum heads of the bass and snare drums, and on the underside of the hi-hat cymbal was used to collect MIDI data via an Alesis D4 drum–machine interface. Two Brüel and Kjaer microphones recorded performance audio at 44.1 kHz 16-bit quality. The application Logic Express was used for presentation of target materials, and for the recording of the student performances. MIDI data were routed from MAX/MSP to Macromedia Flash 8, which then generated the different visual feedback displays. Target performances and the metronome signal were presented using a pair of loudspeakers placed in front of the participants. A latency check using a Tektronix THS710A oscilloscope along with a contact microphone and a light-sensitive diode revealed an average latency of 82 ms (SD = 15.8 ms) over 20 measurements between performed notes and on-screen visualization.

Visual feedback: high-level

The high-level feedback displayed categorical information about the expressive style and skill level of the imitation performance. This was done using the continuous probabilities generated by a set of Bayesian classifiers formulated using the target materials as training data. Eight performance features distinguishing between the three expressive style categories, and eight features that distinguished the instructor and novice performances were selected. Bayes' rule in combination with these feature sets was then used in real time with the performance data from the most recent half-bar repetition of the imitation to generate a set of four probabilities representing the three expressive style categories and the instructor/novice distinction. The details of the feature selection process and the formulation of the Bayesian classifiers can be found in the Appendix.

Simple, four-sided polygons were used to represent each of the three expressive style categories (see Fig. 2a–c). The

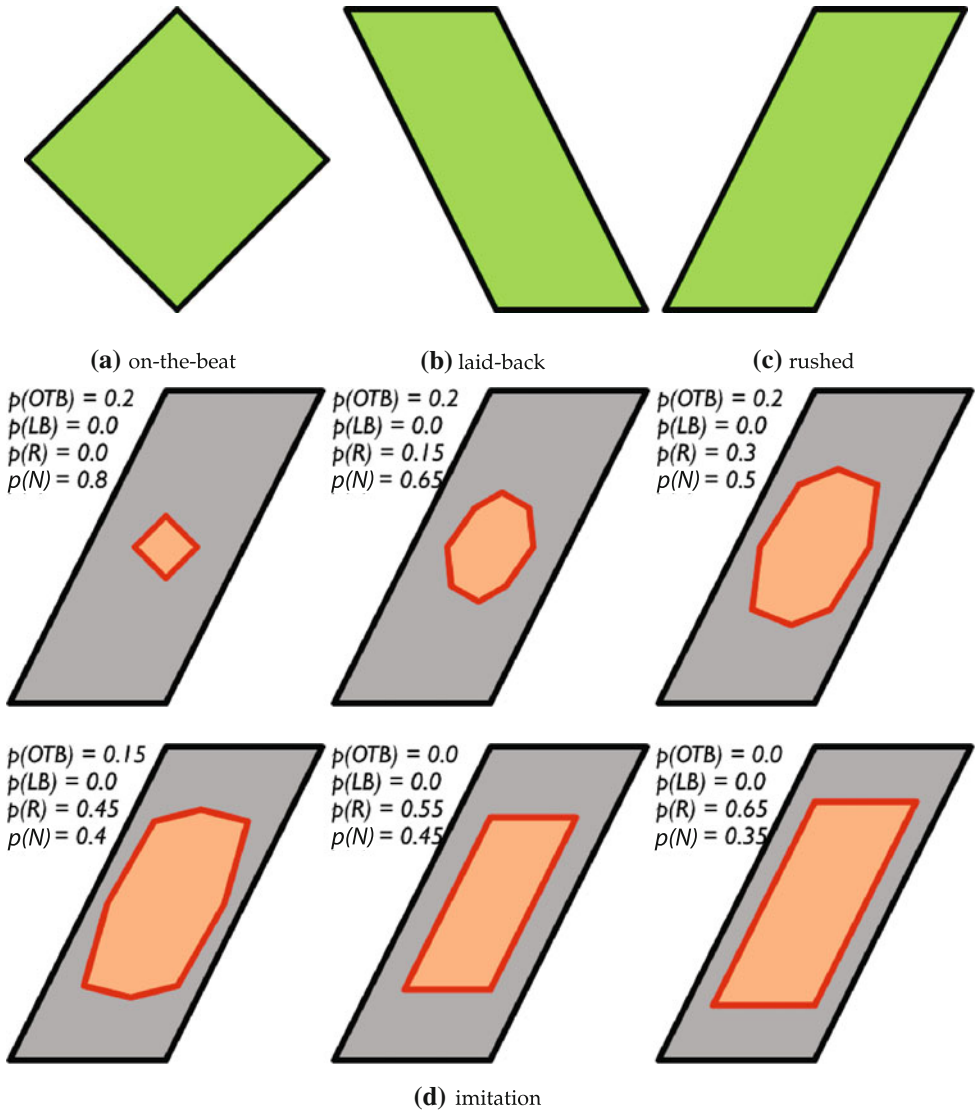

Fig. 2 High-level visual feedback. Panels **a–c** show the prototypical shapes used to represent each of the three expressive styles. These shapes were defined by eight points: one in each corner, and one at the midpoint of each line segment. Panel **d** shows an example sequence of the RTVFB during an imitation performance. Here, the target performance is represented by the *grey shape* presented in the background, while the ongoing imitation performance is presented as a *semi-transparent grey shape* in the foreground. With each incoming note, the four probabilities were recalculated using the most recent half-bar's data. The corresponding shape definitions of the three styles were then weighted by the current probabilities and summed to produce a mixture of the three shapes. In addition, the novice probability determined the overall size of the resulting shape. Thus, the feedback display was dynamically updated for each performed note. An increase in the probability corresponding to the style of the target performance lead to a *grey* imitation figure which more closely resembled the shape of the *grey* target figure, while a decrease in the novice probability led to an imitation figure whose size more resembled that of the target

three shapes were chosen because they were visually distinct from one another, while still being four-sided. They were defined by eight points: one in each corner, and one at the midpoint of each segment. This allowed for simple morphing between the three shapes. Before each imitation, the shape representing the style of the target was displayed while the target performance was presented. Then, during the imitations, the corresponding shape was presented in the background as a grey target which the participants were instructed to match. During the first half-bar, while the data needed to calculate the features were being collected, the imitation shape was not presented. After the first half-bar, the imitation shape faded from completely transparent to 60% opacity over the course of the next half-bar.

During the real-time calculation of the four probabilities (values between 0 and 1), lower- and upper-bounds of 0.15 and 0.7 were chosen. Probabilities between these bounds were re-scaled to a value between 0 and 1, while those

outside were set to either 0 or 1. Probabilities were continuously updated with each incoming note. These final values were used as weights on the eight vertices of the corresponding shape definitions for each category, and were combined into a foreground shape representing the imitation. This shape would morph between successive intervals as the probabilities were updated. Higher probabilities for a given style led to a shape which more resembled that of the corresponding style. In addition, the shape grew in size as the novice probability shrank, and vice versa. An example sequence of probabilities and the corresponding shapes can be seen in Fig. 2d.

In the sequence, the performance starts off in the on-the-beat style, and has a relatively high-novice probability, which reduces the size of the figure. As the performance progresses, the novice probability decreases, causing the shape to grow in size, while the performance moves to the rushed style, causing the form of the shape to become more like the target.

Visual feedback: low-level

Examples of the low-level feedback can be seen in Fig. 3. Panels a–f show the target patterns, while panel g is an example of the display during imitation. It drew on standard musical notation for percussion, with notes on the bass drum, snare drum, and hi-hat cymbal plotted from low to high, respectively, and with time proceeding from left to right, with grid lines placed at metrical time points corresponding with the quarter-note level. In these respects, it was considered to be easy for the participants to make use of. However, due to the number of elements displayed on screen at any given point (more than 20), it was expected to have a high element interactivity, thus giving it a high extraneous cognitive load.

Performed notes were displayed as three different shapes representing the voice of the note (square = bass drum, circle = snare drum, triangle = hi-hat cymbal). The size of the shapes varied with the dynamics of the performed note: the louder the performed note, the larger the shape. The association of visual size with loudness has been shown in research in audio-visual perception to be present in a large majority of participants (Walker, 1987), and has worked effectively as an audio-visual mapping in a previous study using RTVFB for musicians (Sadakata et al., 2008). The display showed the most recently performed two bars, and scrolled to the left continuously; new notes always appeared at 80% of the screen width on the right side.

During the presentation of the target performance for a given trial, the instructor's performance was displayed in real time. Then, during the imitation phase of a trial, the target performance would appear in a grey color in the background, while the imitation performance appeared in the foreground as a semi-transparent colored overlay. Extra notes were also displayed, while missed notes were absent from the imitation overlay. The more accurate the student was with the imitation, the more overlap there would be between the student's shapes and the instructor's shapes.

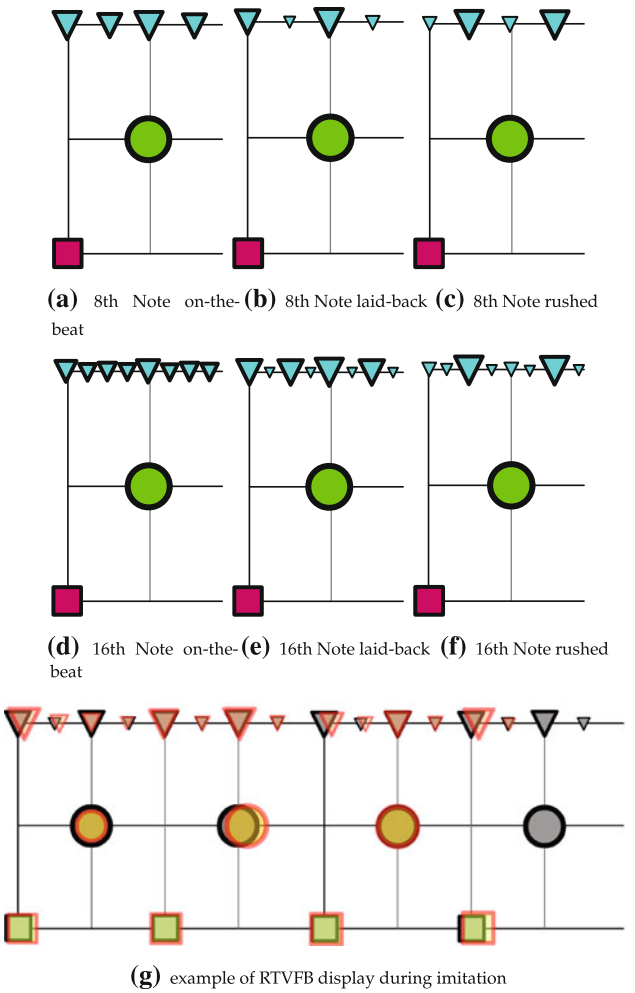

Fig. 3 Low-level feedback. The first six panels show one half-bar of the target display for each of the performances. In panel *g* is an example of the RTVFB display during imitation performances. In this instance, it can be seen that the last performed notes were played slightly later and louder than the target performance

Analysis

MIDI recordings of the participant performances were converted to text files and analyzed using the JMP 5.1 statistics package. Half bars containing extra or missing notes were excluded from the analysis. A total of 8,122 half bars out of 8,640 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.

Four performance measures were then averaged on a per trial basis for each participant. The average RMS timing error (in terms of seconds) and RMS dynamics error (a value between zero and one derived from the standard MIDI 0–127 range) for each half-bar repetition in the imitation performances were calculated using the values of the corresponding notes in the selected target performances. In addition, for a given imitation, one of the instructor performances served as a target performance, with the probability (between zero and one) generated by the corresponding classifier (the "target probability") serving as a measure of how well the expressive features of the instructor performance had been imitated, with higher values indicating greater success in imitation. The "novice probability" used in the generation of the high-level feedback was also taken as a performance measure, representing a set of performance features that distinguished the instructor performances from that of the novice percussionist. Here, a lower probability indicates a performance more like that of the instructor.

The RMS error measures and the probabilistic measures capture two distinct tendencies in the data. The RMS error measures reflect the absolute difference between a given imitation and the target. They are directly related to the low-level feedback, which displayed the raw performance data of the target and the imitation overlaid on one another, and can be considered as a measure of the precision of the imitation. The probabilistic measures make use of second-order features capturing the timing and dynamics profiles of the imitation. These features are based predominantly on proportional relationships of subsequent notes to one another. As such, a given feature in an imitation could be identical to that in the target performance even though the constituent notes were themselves slightly shifted in timing or dynamics.

A mixed effects model with the participant modeled as a random variable and trial taken as a continuous variable was used to calculate an ANOVA for each of the four performance measures, in order to see how the performance of the participants was influenced by the visual-feedback condition, beat pattern, expressive style, and trial number. Interactions between these variables were also checked for their influence on the performance measures.

Results

The main effects are plotted in Fig. 4, while ANOVA results are presented in Table 1, including test parameters and significance levels. Interaction effects for all four measures are presented at the end, following the main effects for each measure.

Target probability

In the analysis of the target probability measure, significant main effects of RTVFB condition, expressive style, beat pattern, and trial were found. For the three visual feedback conditions, the average target probability was highest in the high-level visual feedback condition (mean = 0.37, SE = 0.01), followed by the no-feedback control condition (mean = 0.35, SE = 0.009), then the low-level feedback condition (mean = 0.34, SE = 0.009). A planned pair-wise comparison (Tukey-HSD) revealed that the difference between the high-level feedback and the other two conditions was significant, but that the difference between low-level feedback and the no-feedback condition was not.

With respect to trial, the average target probability increased from trial 1 (mean = 0.33, SE = 0.013) to trial 5 (mean = 0.37, SE = 0.012), indicating that performances in general improved across trials. A regression line fit to the trial data had a slope of 0.008, confirming the positive trend across trials.

With regard to the three expressive styles and beat pattern of the performances, performance was highest for the *on-the-beat* performances (mean = 0.41, SE = 0.007), followed by the *rushed* (mean = 0.34, SE = 0.01), then the *laid-back* performances (mean = 0.30, SE = 0.01). A Tukey-HSD pair-wise comparison revealed that the differences between the three styles were all significant. Target probabilities were significantly higher for the 8th-note pattern (mean = 0.38, SE = 0.007) than the 16th-note pattern (mean = 0.32, SE = 0.008).

Novice probability

Analysis of the novice probability measure revealed a significant main effect of expressive style. With the novice probability, a lower number indicates a more skilled performance. Novice probability was lowest for the *on-the-beat* performances (mean = 0.33, SE = 0.006), followed by those performed with the *rushed* (mean = 0.34, SE = 0.006), and the *laid-back* (mean = 0.35, SE = 0.006) styles, similar to the pattern observed for the target probability. No significant effects were found with pattern, visual feedback condition, or trial.

RMS timing error

Significant effects of visual feedback condition, expressive style, beat pattern, and trial were found in the analysis of the average RMS timing error of the participants' imitation performances. Performances had the lowest average timing error in the control condition (mean = 29.7 ms, SE = 0.9 ms), followed by the low-level feedback condition (mean = 31.1 ms, SE = 0.9 ms), and lastly by the high-

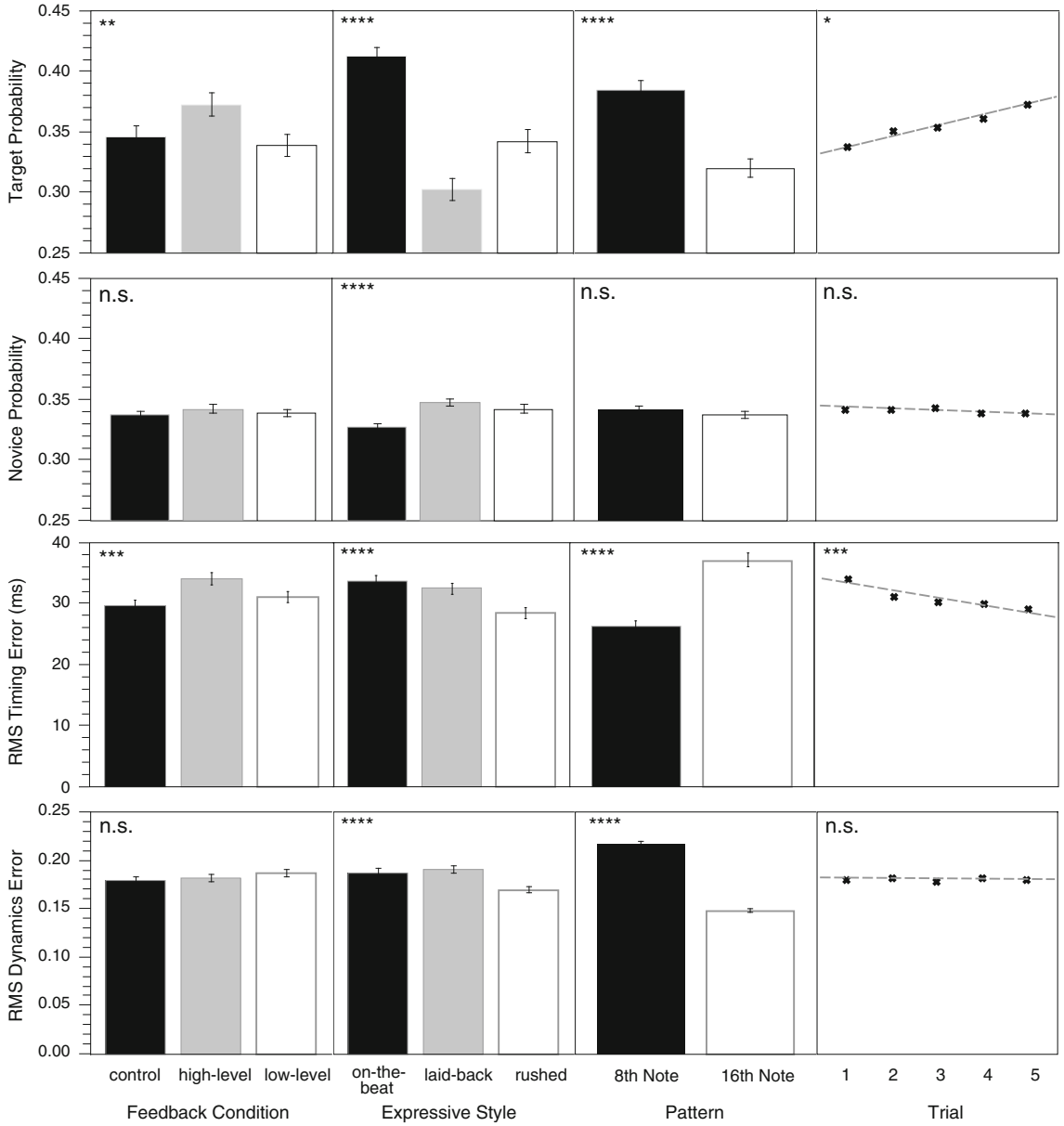

Fig. 4 Main effects for the four performance measures. Effects of RTFVB condition, expressive style, beat pattern, and trial were revealed using a mixed-effects ANOVA. *Asterisks* indicate level of significance: $*p < 0.05$, $**p < 0.01$, $***p < 0.001$, $****p < 0.0001$. For the target probability measure, a higher value indicates a better performance, while for the other three measures, a lower value indicates a better performance

level feedback condition (mean = 34.1 ms, SE = 0.9 ms), with the differences between conditions all being significant (planned pair-wise comparison, Tukey-HSD).

With regard to expressive style, timing error was lowest in the *rushed* performances (mean = 28.5 ms, SE = 1.1 ms), followed by the *laid-back* (mean = 32.6 ms, SE = 1.1 ms) and the *on-the-beat* (mean = 33.6 ms, SE = 1.1 ms) performances, with the difference between the rushed and the other two styles being significant (Tukey-HSD). Error was significantly lower in 8th-note performances (mean = 26.2 ms, SE = 1 ms) than the 16th-note performances (mean = 36.9, SE = 1 ms). A regression line fit to the trial data had a slope of −1.1 ms, indicating that the timing error decreased across trials.

RMS dynamics error

An ANOVA on the average RMS dynamics error showed significant effects of expressive style and beat pattern. The error was highest for the laid-back performances (mean = 0.191, SE = 0.003) followed by the on-the-beat (mean = 0.187, SE = 0.003) and the rushed performances

Table 1 Main effects and significant interactions for mixed model ANOVAs of performance measures

	DOF	Den.	*P* (target)		*P* (novice)		Timing error		Dynamics error	
			F ratio	*p*	*F* ratio	*p*	*F* ratio	*p*	*F* ratio	*p*
Condition	2	487	4.82	<0.01	0.83	n.s.	8.94	<0.001	2.28	n.s.
Style	2	487	46.92	<0.0001	13.50	<0.0001	12.86	<0.0001	18.59	<0.0001
Pattern	1	487	47.37	<0.0001	1.29	n.s.	148.89	<0.0001	514.31	<0.0001
Trial	1	487	5.98	<0.05	0.92	n.s.	12.65	<0.001	0.09	n.s.
Style × pattern	2	487	4.18	<0.05	9.97	<0.0001	7.22	<0.001	43.56	<0.0001
Trial × pattern	1	487	1.16	n.s.	0.06	n.s.	4.56	<0.05	0.03	n.s.
Con × Sty × Pat	4	487	1.47	n.s.	3.46	<0.01	3.33	<0.05	0.49	n.s.

The unit of observation has been set to the trial level, in order to model interactions between trial and condition. Three- and four-way interactions with no significant effects are omitted

(mean = 0.17, SE = 0.003). The rushed performance had a significantly lower dynamics error than the other two styles. For beat pattern, the dynamics error was significantly lower for the 16th-note pattern (mean = 0.148, SE = 0.002) than for the 8th-note pattern (mean = 0.217, SE = 0.002).

However, the effects of visual feedback and trial did not reach significance, indicating that no one visual feedback condition provided any greater benefits with regard to imitating the overall dynamic levels of the instructor performance. Additionally, the accuracy of the participant imitations with respect to overall dynamics did not improve across trials.

Interaction effects

A significant interaction of style and pattern was found for all four measures. While style and pattern were both under experimental control, they were not specifically related to the hypotheses tested by the experiment. Additionally, no systematic effects within or between measures were observed in the subsequent planned pair-wise comparisons, limiting any interpretation related to the difficulty of performing a given pattern or style.

For the RMS timing error, and additional interaction of trial and pattern was found. A significant decrease in timing error between the first trial (mean = 42.6 ms, SE = 0.001) and last three trials (mean range = 34.6–35.3 ms, SE = 0.001) of the 16th-note performances was revealed by a planned pair-wise comparison (Tukey-HSD). The timing error on all 8th-note trials was significantly lower than that of the 16th-note trials, but no significant effect existed within the 8th-note trials.

A three-way interaction of condition, style and pattern was found for both the novice probability and the RMS timing error. However, no systematic effects with respect to condition were observed within the interaction for either measure.

Discussion

The results of the present study are generally congruent with the previous research findings regarding musical performance under different RTVFB conditions. In line with the findings of Sadakata et al. (2008), but contrary to our hypothesis, RMS timing error was significantly higher in the two feedback conditions than in the control condition. This is also in line with the findings reported by Wilson et al. (2008) that participant performance was worse during training with VFB.

While we hypothesized that reducing the number of elements in the high-level VFB representation would reduce extraneous cognitive load, it may be that the initial use of a VFB display during music performance diverts attentional resources away from the primary task. In essence, a dual-task (perform and monitor VFB) is being compared with a single task (perform). The dual task can be interpreted as having a higher intrinsic cognitive load than the single task, leading to divided attention and a greater timing error. This interpretation is consistent with previous findings in the RTVFB literature. The high skill-level of the participants may also have led to relatively low timing error in the control condition.

However, the finding that the target probability measure increased during the high-level feedback condition resembles the findings of Rossiter et al. (1996). In their study, visual feedback on various measured parameters of singing (i.e., F0 amplitude, laryngeal closed/open quotient) was presented to participants. It was shown that, depending on which particular parameter of performance was used to generate visual feedback, that the visualized parameter alone showed significant increases, while non-visualized parameters showed little or no changes. In the present study, high-level feedback was based primarily on the target probability measure. It is worth noting that, while the participants were not explicitly told which features of the performance were used to generate the visual feedback,

they nonetheless performed with a higher target probability when provided with high-level visual feedback.

The disparity between the results for the timing error and the target probability may lie in how these measures were calculated. Whereas the RMS error measures are fixed to the absolute timing and dynamics parameters of the target materials, the target probability was based on a set of features capturing the profiles of the target materials using proportional measures. As such, many combinations of values in a succession of notes could result in the same relative proportions. A performance slightly shifted in absolute dynamics or timing could still capture the expressive aspects of the performance, thus resulting in a higher target probability and a higher RMS timing or dynamics error. One might argue that it is these higher-order features which better capture the "expression" of the performances than the RMS errors.

We evaluated this possibility by assessing how the participant performances relate to qualitative judgements of the imitations. Following the experiment, we presented a subset of 54 of the 8th-note participant performance recordings (18 from each style) to 3 professional percussion instructors. The performances were selected such that the low to high range for each of the four performance measures was evenly represented within the set. For each evaluation, the instructors were presented first with the target performance, and then with the participant imitation, and could listen as many times as needed. They then rated the quality of the imitation on a 7-point scale.

The ratings for each performance were then correlated with the four performance measures. For the target probability and novice probability, there were correlations of $r = 0.326$ and $r = -0.322$, respectively. The correlations with RMS timing error and RMS dynamics error were $r = -0.084$ and $r = -0.308$. All the correlations, with the exception of the RMS timing error (not significant), were significant at the $p < 0.0001$ level. Thus, while none of the correlations were above the 0.5 level, a stronger relationship with the quality ratings was found for the probabilistic measures and the RMS dynamics error than with the RMS timing error measure.

No effect of VFB condition was found for the novice probability measure. One potential explanation is that, unlike the target performances, the recordings of the novice were not presented to the participants, as the primary focus was on the learning and imitation of the expressive styles. The inclusion of the novice performance in the development of the high-level feedback served as a check to ensure that participants imitated all aspects of the instructor performance. The three styles performed by the teacher differed primarily with respect to the timing and dynamics of the hi-hat, whereas the snare and bass drum were played consistently across styles. However, the bass and snare drum were performed much less consistently by the novice, which led to a set of features distinguishing these aspects of the novice performances from those of the instructor. More details about these features are provided in the Appendix.

With respect to the effects of beat pattern on the four performance measures, no one clear interpretation is possible. While it would be natural to expect that 8th-note performances would generally be less difficult to perform than the 16th-note performances, leading to better overall results in the performance measures, this is not the case with respect to the RMS dynamics error. A potential explanation is that the RMS dynamics error of the bass and snare notes contributes disproportionately to the overall average, and that the additional hi-hat notes in the 16th-note pattern offset this, leading to a lower average error.

The significant effects of expressive performance style on the various performance measures also tell a mixed tale. While the on-the-beat performances were considered easier to play, due to the more complex accenting patterns for the laid-back and rushed performances, the RMS timing error was actually lowest for the rushed performance, which, given the offbeat accenting, would perhaps be considered the most difficult. However, with regard to the two probabilistic measures, performance was significantly better for the on-the-beat performances. In a sense, this is a microcosm of the overall finding that imitation of higher-order features relevant to the expressive style improved at the expense of temporal precision.

Conclusions

High-level real-time visual feedback at the categorical level can help to improve the low-level performance features upon which it is based. However, at least in the initial periods of use examined during this study, the high-level RTVFB may also increase the cognitive load placed on participants during the imitation task, as indicated by a significantly higher RMS timing error than in the other two RTVFB conditions. The design of the present experiment prohibits us from knowing whether or not the positive and detrimental effects of RTVFB persist in non-feedback conditions, or if longer-term familiarization and use of the RTVFB system leads to changes in the observed effects. These would both be interesting questions to pursue in follow-up research.

While traditionally applied to the design of instructional and educational materials such as textbooks or multimedia learning tools, Cognitive Load Theory shows promise as a tool in the design of interactive computer-based training systems for complex tasks such as music performance. The present results highlight both the benefits and the

difficulties that come along with the various manners in which knowledge of results can be given to learners.

This may be especially true in learning situations involving complex tasks such as expressive drum performance, which requires the combined use and sequencing of multiple effectors at short, precisely timed intervals, and with subtle manipulations in force in order to achieve the desired performance. As such, it places strong demands on cognitive and working memory resources, and does not leave much over for the processing of additional feedback information above and beyond normal sensory feedback mechanisms. Thus, any type of RTVFB designed for these type of tasks must minimize all forms of extraneous cognitive load, while also providing knowledge of results which is useful for improving various aspects of performance.

Acknowledgments The authors would like to acknowledge Gerhard Jeltes from the Amsterdam conservatory for his help choosing and recording the target materials, along with the Royal Conservatory in the Hague and the Music Conservatory of Utrecht for help organizing students as participants. This research was funded by the Technologiestichting STW.

Appendix

Applications of Bayes' rule to problems in machine learning and perception have achieved success in several different domains, including computer vision (Knill et al., 1996), handwriting recognition (Cheung et al., 1998), and music transcription (Cemgil et al., 2000). Although Bayes rule has not been explicitly applied to the classification of expressive style or skill level, the successes which have been achieved with it in pattern recognition for other domains suggest that it may be a fruitful approach.

A statistical analysis of the target materials was conducted, the results of which subsequently led to the development of a set of Bayesian classifiers. These classifiers were used on student performance data in the experimental portion of the study to identify which of the three expressive styles performed by the instructor it most resembled, and whether or not the participant performance most resembled the instructor or the novice performances. The resulting classification rates were used as the basis for the high-level RTVFB. Additionally, the classifiers were applied to the target materials, and the most prototypical repetition of each target performance was chosen to present to participants during the experiment.

Feature analysis

An analysis was conducted in order to find a set of features F which could be used to distinguish between N classes of interest; in our case, *on-the-beat, laid-back, rushed*, and *novice*. The timing data were represented in terms of millisecond values, while the MIDI velocity data was scaled to a 0–1 range representing dynamics. MIDI timing data were also included for the metronome ticks. Data were analyzed per half-bar repetition, with each containing six

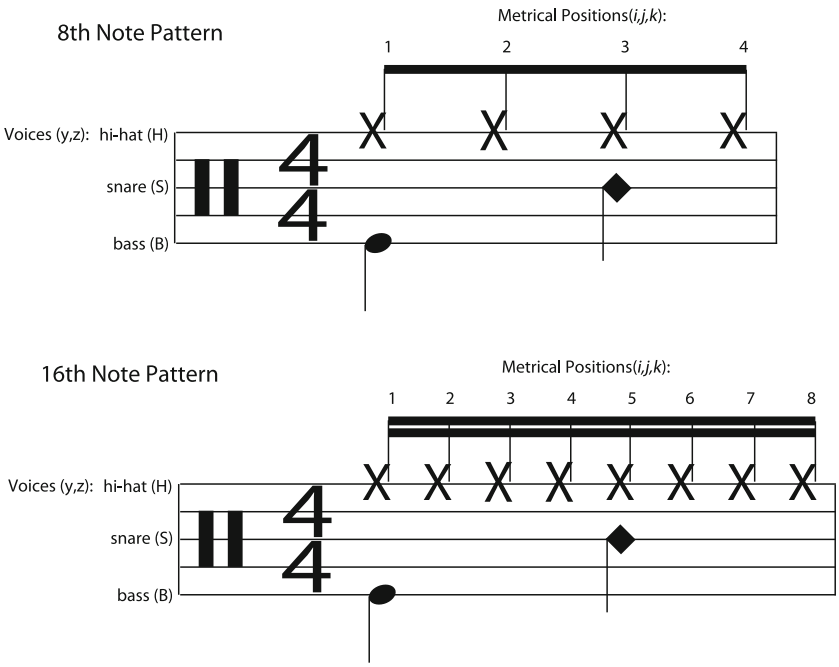

Fig. 5 Schematic of score positions used for analysis of performance features. Both patterns used in the study were organized schematically for the analysis procedure. Each metrical subdivision present in the pattern was given an index. The three different drum voices were also given unique indices

Table 2 Features used in the analysis of the instructor and novice performances

Basic features ($\Theta_{r,c}$)	
t_i^y	Timing of voice y at metrical position i
v_i^y	Velocity of voice y at metrical position i
T_i	Metronome/mechanical timing at metrical position i
Derived features	
$\Delta t_i^y = t_i^y - T_i$	Timing difference from metronome of voice y at metrical position i
$\Delta v_i^y = v_i^y - \bar{v}^y$	Velocity difference of voice y at position i from mean velocity of voice y
$\mathrm{IOI}_{i,j}^y = t_j^y - t_i^y$	Inter-onset interval of voice y at successive metrical positions i and j
$\mathrm{VI}_{i,j}^y = v_j^y - v_i^y$	Velocity interval of voice y at successive metrical positions i and j
$TP_{i,j,k}^y = \mathrm{IOI}_{i,j}^y/\mathrm{IOI}_{j,k}^y$	Relative proportion of successive inter-onset intervals of voice y at positions i, j, and k
$VP_{i,j,k}^y = \mathrm{VI}_{i,j}^y/\mathrm{VI}_{j,k}^y$	Relative proportion of successive velocity intervals of voice y at positions i, j, and k
$\mathrm{TA}_i^{y,z} = t_i^y - t_i^z$	Timing asynchrony of voice y and z at metrical position i
$VA_i^{y,z} = v_i^y - v_i^z$	Velocity difference of voice y and z at metrical position i
$\sigma(\mathrm{IOI})_{1..n}^y$	Standard deviation of the inter-onset interval for voice y over a given half bar segment
$\sigma(\mathrm{VI})_{1..n}^y$	Standard deviation of the velocity interval for voice y over a given half bar segment

Basic features were taken from the MIDI performance data. Derived features were calculated for all possible permutations in each beat pattern. The relative timing and velocity proportions, inter-onset interval, velocity interval, and the two standard deviation measures based on them were calculated either for the hi-hat, or for the bass and snare drum together. For features based on more than one note, the first note of the subsequent half bar's data was sometimes included

Table 3 Separation indexes of selected features for 8th- and 16th-note instructor performances with three expression types, and for novice versus instructor performances

8th note			16th note		
Feature	S	Type	Feature	S	Type
Expressive style features					
$\mathrm{VI}_{2,3}^H$	0.999	Dynamics	$\mathrm{VI}_{6,7}^H$	0.987	Dynamics
$\mathrm{VI}_{3,4}^H$	0.999	Dynamics	$\mathrm{VI}_{7,8}^H$	0.982	Dynamics
$\mathrm{VI}_{4,1}^H$	0.999	Dynamics	$\mathrm{VI}_{2,3}^H$	0.977	Dynamics
$\mathrm{VI}_{1,2}^H$	0.997	Dynamics	$\mathrm{VI}_{3,4}^H$	0.970	Dynamics
$TP_{4,1,2}^H$	0.762	Timing	$TP_{4,5,6}^H$	0.696	Timing
$TP_{1,2,3}^H$	0.725	Timing	$TP_{5,6,7}^H$	0.633	Timing
$TP_{2,3,4}^H$	0.655	Timing	$TP_{8,1,2}^H$	0.627	Timing
$TP_{3,4,1}^H$	0.604	Timing	$TP_{1,2,3}^H$	0.616	Timing
Skill-level features					
v_3^S	0.999	Dynamics	v_5^S	0.999	Dynamics
$\mathrm{VI}_{1,3}^{BS}$	0.999	Dynamics	$\mathrm{VI}_{1,5}^{BS}$	0.999	Dynamics
$\mathrm{VI}_{3,1}^{SB}$	0.999	Dynamics	$\mathrm{VI}_{5,1}^{SB}$	0.997	Dynamics
$VA_3^{S,H}$	0.994	Dynamics	$VA_5^{S,H}$	0.994	Dynamics
$\mathrm{TA}_3^{S,H}$	0.645	Timing	$\mathrm{TA}_5^{S,H}$	0.768	Timing
$\sigma(\mathrm{IOI})_{1...6}^{BS}$	0.625	Timing	$\sigma(\mathrm{IOI})_{1...8}^{BS}$	0.691	Timing
$TP_{1,3,1}^{BS}$	0.593	Timing	$TP_{1,5,1}^{BS}$	0.689	Timing
Δt_1^H	0.791	Timing	Δt_1^H	0.609	Timing

The features distinguishing between the expressive styles were all related to the differences in timing and dynamics between successive hi-hat notes. The features which were selected for the novice/instructor distinction were primarily based on the bass and snare drum notes. While the instructor played the bass and snare drums relatively consistently across performances, the novice performances were more variable in this regard

or ten notes for the *8th-note* or *16th-note* patterns, respectively (see Fig. 5 for a schematic diagram).

Given a set of R repetitions in class c_n , for each repetition r of a performance, there is a corresponding set of parameters $\Theta_{r,c}$ containing the raw data.

A set of M features $[f_1 \ldots f_m]$ was then defined (see Table 2). This set included typical musical performance measures such as inter-onset intervals, relative proportions of successive intervals, asynchronies, and measures of variance. A total of 90 features for the 8th-note performances, and 186 features for the 16th-note performances were defined. Each of these features was labeled as either a dynamics or a timing feature. Using these definitions, each parameter set $\Theta_{r,c}$ produced a set of values $[x_{r,c,1} \ldots x_{r,c,m}]$ corresponding to the individual features for a given repetition of a performance.

For each feature f_m within the feature set, the distribution $D_{c,m}$ of all corresponding values across repetitions within each individual class c_n was estimated using a Gaussian function, based on the observation that the majority of features exhibited unimodal distributions with low skew. This facilitated a probabilistic interpretation of a given value x_m with respect to the distribution of values for a specific feature f_m given class c_n. This was done using a probability density function $d_{c,m}(x)$ corresponding to each distribution $D_{c,m}$.

Subsequently, a comparison of the distributions between classes for each feature was done using an index of separability S. Given a specific feature f_m, this index estimates the amount of surface overlap between the distributions for all classes by integrating the max function for the N density functions across all possible values of x_m, and dividing by the total possible surface area of the distributions. The index is then normalized such that $S \in [0, 1]$.

$$S(f_m) = \frac{\int_{-\infty}^{\infty} \max(d_{1,m}(x), \ldots, d_{N,m}(x))\mathrm{d}x}{N-1} \quad (1)$$

Accordingly, S is a measure of the distinctness of a particular feature in the different classes of interest. From

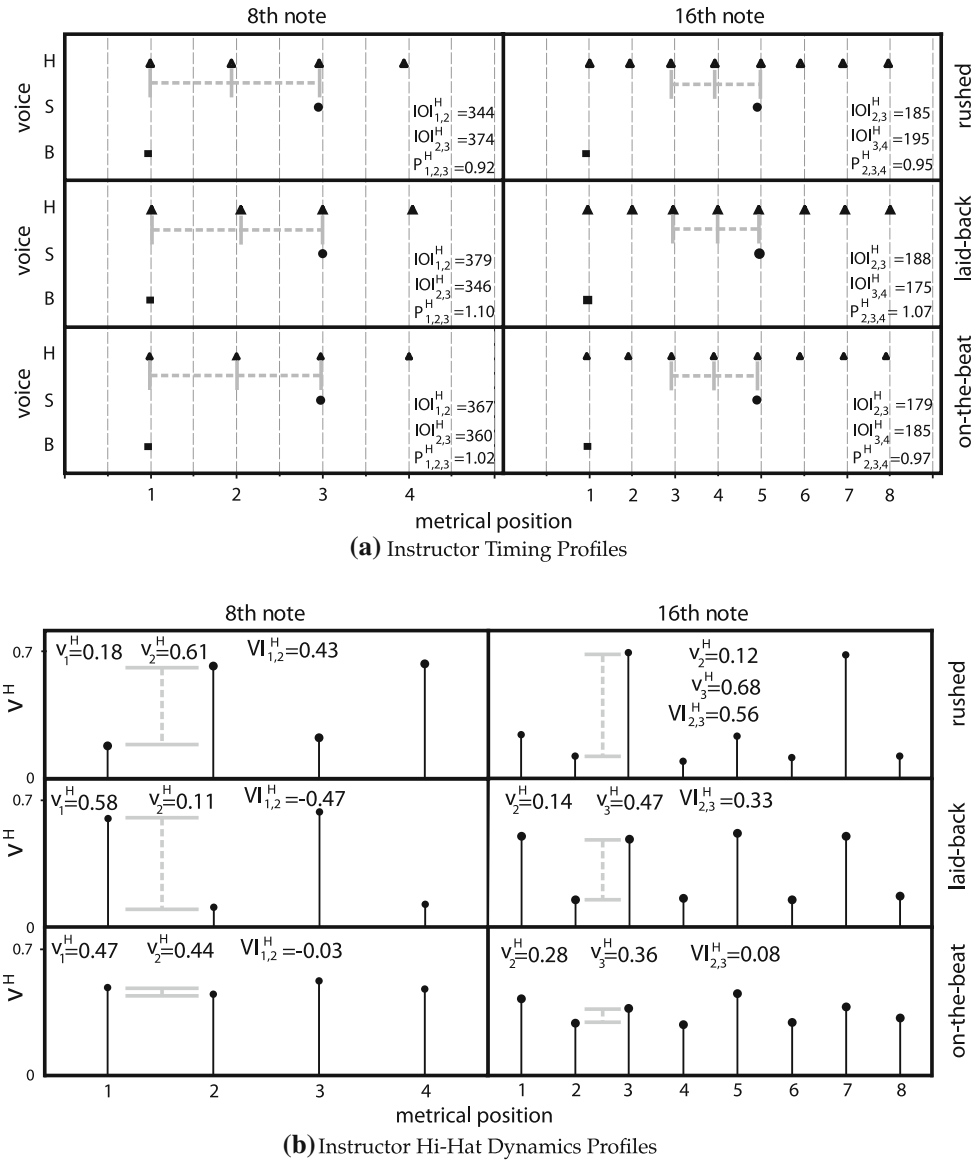

Fig. 6 Instructor performance profiles. Panel **a** shows timing profiles for the six instructor performances, while panel **b** shows dynamics profiles for the hi-hat notes in the six performances. In addition, some of the features selected in the analysis of the target performance are illustrated using the same notation as the feature definitions

the perspective of signal detection theory, S is similar to d' in that our classification task is a probabilistic decision process based on observations of events which are normally distributed and which makes use of an optimal criterion (Wickens, 2002). When d' is high for a given class of events (the "signal"), the probability that another event ("noise") will be mistakenly identified as belonging to that class is low. The main difference between the measures is that S distinguishes between an arbitrary number of classes, while d' distinguishes between two.

In our situation, features having a high separability index have a more distinct distribution of values for each class, meaning that these features are good candidates to use for classifying performances of the same material. The 16 features which were selected based on the results of the feature analysis are shown in Table 3. Additionally, the mean timing and dynamics profiles of the instructor performances along with a schematic illustration of a subset of the selected features are shown in Fig. 6.

Bayesian formulation

A subset F consisting of $L = 16$ features possessing the highest separation indexes, half of which were timing features, and the other half of which were dynamics features, was selected to form the basis of a classifier that specifies the likelihood that a performance belongs to a particular class c_n; in the current case, *on-the-beat*, *laid-back*, *rushed*, or *novice*. These measures were formulated using an application of Bayes rule, which is given by:

$$P(c|F) = \frac{P(F|c)P(c)}{P(F)} \quad (2)$$

We assume that the likelihood of all classes are equal, such that the $P(c)$ term is set to $\frac{1}{N}$. The probability of the feature set F given a class c_n is calculated in the following equation:

$$P(F|c_n) = \frac{1}{L}\sum_{1}^{L} d_{c,i}(x) \quad (3)$$

This is the average of the probability density functions for the individual features $[f_1 \ldots f_L]$ given a class c_n. While taking the product of the individual density functions is typically chosen, we made an ad hoc decision to use the average, as it led to more graded transitions in the resulting probabilities. This was more suitable for the generation of the high-level feedback in the experiment. The overall probability of a feature set F is the sum of the conditional probabilities $P(F|c_n)$ for each possible class c_n in the set of N classes.

$$P(F) = \sum_{i=1}^{N} P(F|c_i) \quad (4)$$

The results from Eqs. 3 and 4 provide the terms necessary for calculating Eq. 2. A total of N probabilities is calculated whose sum is equal to 1. The chance probability for each class is equal to $\frac{1}{N}$.

References

Annett, J. (1969). Feedback and human behavior. Penguin Education.

Cemgil, A. T., Desain, P., & Kappen, B. (2000). Rhythm quantization for transcription. *Computer Music Journal, 24*(2), 60–76.

Chandler, P., & Sweller, J. (1991). Cognitive load theory and the format of instruction. *Cognition and Instruction, 8*(4), 293–332.

Cheung, K.-W., Yeung, D.-Y., & Chin, R. (1998). A bayesian framework for deformable pattern recognition with application to handwriten character recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 20*(12), 1382–1388.

Clark, E. F. (1987). Categorical rhythm perception: an ecological perspective. In A. Gabrielsson (Ed.), *Action and perception in rhythm and music*. Royal Swedish Academy of Music.

Clark, E. F. (1993). Imitating and evaluating real and transformed musical performances. *Music Perception, 10*(3), 317–341.

Corbetta, M., Miezin, F. M., Dobmeyer, S., Shulman, G. L., & Petersen, S. E. (1991). Selective and divided attention during visual discriminations of shape, color, and speed: functional anatomy by positron emission tomography. *Journal of Neuroscience, 11*(8), 2383–2402.

Craik, F. I. M., Govoni, R., Naveh-Benjamin, M., & Anderson, N. D. (1996). The effects of divided attention on encoding and retrieval processes in human memory. *Journal of Experimental Psychology, 125*(2), 159–180.

Dahl, S. (2000). The playing of an accent—preliminary observations from temporal and kinematic analysis of percussionists. *Journal of New Music Research, 29*(3), 225–233.

Desain, P., & Honing, H. (2003). The formation of rhythmic categories and metric priming. *Perception, 32*(3), 341–365.

Engle, R. W. (2002). Working memory capacity as executive attention. *Current Directions in Psychological Science, 11*(1), 19–23.

Escartí, A., & Guzmán, J. F. (1999). Effects of feedback on self-efficacy, performance and choice in an athletic task. *Journal of Applied Sport Psychology, 11*(1), 83–96.

Evans, J. (1960). *An investigation of teaching machine variables using learning programs in symbolic logic*. Unpublished doctoral dissertation, University of Pittsburgh.

Hoffren, J. (1964). A test of musical expression. *Council for Research in Music Education, 2*, 32–35.

Hoppe, D., Sadakata, M., & Desain, P. (2006). Development of real-time visual feedback assistance in singing training: a review. *Journal of Computer Assisted Learning, 22*, 308–316.

Juslin, P., Friberg, A., Schoonderwaldt, E., & Karlsson, J. (2004). Feedback learning of musical expressivity. In A.Williamon (Ed.), *Musical excellence*. Oxford, England: Oxford University Press.

Kent, R. D. (1974). Auditory-motor formant tracking: a study of speech imitation. *Journal of Speech and Hearing Research, 17*, 203–222.

Knill, D. C., Kerstern, D., & Yuille, A. (1996). Introduction: a Bayesian formulation of visual perception. In D.C. Knill & W. Richards (Eds.), *Perception as Bayesian inference* (pp. 1–21). Cambridge: Cambridge University Press.

McPherson, G. E., & Schubert, E. (2004). Measuring performance enhancement in music. In A. Williamon (Ed.), Musical excellence. Oxford University Press.

Paas, F., Renkl, A., & Sweller, J. (2003). Cognitive load theory and instructional design: recent developments. *Educational Psychologist, 38*(1), 1–4.

Palmer, C. (1997). Music performance. *Annual Review of Psychology, 48*, 115–138.

Pennington, M. C. (1999). Computer-aided pronunciation pedagogy: promise, limitations, directions. *Computer Assisted Language Learning, 12*(5), 427–440.

Person, R. S. (1993). *The subjectivity of musical performance: an exploratory music-psychological real world enquiry into the determinants and education of musical reality*. Unpublished doctoral dissertation, University of Huddersfield, Huddersfield, UK.

Repp, B. H. (2000). Pattern typicality and dimensional interactions in pianists' imitation of expressive timing and dynamics. *Music Perception, 18*(2), 173–211.

Repp, B. H., & Williams, D. R. (1985). Categorical trends in vowel imitation: preliminary observations from a replication experiment. *Speech Communication, 4*, 105–120.

Repp, B. H., & Williams, D. R. (1987). Categorical tendencies in imitating self-produced isolated vowels. *Speech Communication, 6*, 1–14.

Rosch, E. (2002). Principles of categorization. In Levitin, D. J. (Ed.), *Cognitive psychology*. MIT Press.

Rossiter, D., Howard, D. M., & DeCosta, M. (1996). Voice development under training with and without the influence of real-time visually presented feedback. *Journal Acoustical Society of America, 99*(5), 3253–3256.

Sadakata, M., Hoppe, D., Brandmeyer, A., Timmers, R., & Desain, P. (2008). Real-time visual feedback for learning to perform short rhythms with expressive variations in timing and loudness. *Journal of New Music Research, 37*(3), 207–220.

Semjen, A., & Garcia-Colera, A. (1986). Planning and timing of finger tapping sequences with a stressed element. *Journal of Motor Behavior, 18*, 287–322.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science, 12*(2), 257–285.

Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction, 4*(4), 295–312.

Tait, M. (1992). Teaching strategies and styles. In R. Cowell (Ed.), *Handbook of research on music teaching and learning* (p. 525–534). New York:Schirmer.

Temperley, D. (2007). *Music and probability*. MIT Press.

Thorpe, C. W., Callaghan, J., & van Doorn, J. (1999). Visual feedback of acoustic voice features: New tools for the teaching of singing. *Australian Voice, 1*(5), 32–39.

Walker, R. (1987). The effects of culture, environment, age, and musical training on choices of visual metaphors for sound. *Percetpion and Psychophysics, 42*, 491–502.

Welch, G. F. (1985). A schema theory of how children learn to sing in tune. *Psychology of Music, 13*, 3–18.

Welch, G. F., Himonides, E., Howard, D. M., & Brereton, J. (2004). Voxed: Technology as a meaningful teaching aid in the singing studio. In Proceedings of the conference on interdisciplinary musicology (cim04).

Welch, G. F., Howard, D. M., & Rush, C. (1989). Real-time visual feedback in the development of vocal pitch accuracy in singing. *Psychology of Music, 17*, 146–157.

Wickens, T. D. (2002). *Elementary signal detection theory*. Oxford University Press.

Wilson, P. H., Lee, Callaghan, J., & Thorpe, C. W. (2008). Learning to sing in tune: does real-time visual feedback help? *Journal of Interdisciplinary Music Studies, 2*(1–2), 157–172.