The Role of Visual Feedback and Creative Exploration for the Improvement of Timing Accuracy in Performing Musical Ornaments

Renee Timmers
University of Sheffield, Sheffield, United Kingdom

Makiko Sadakata & Peter Desain
Radboud University Nijmegen, Nijmegen, The Netherlands

In developing a visual feedback system for a creative activity such as music performance, the objective is not just to reinforce one particular manner of performing. Instead, a desirable characteristic might be that the visual feedback enhances flexibility and originality, in addition to contributing to performance precision. In an experimental study focused on the timing of a musical ornament, we examined whether the instruction to explore ornament timing in training trials with or without visual feedback leads to improved temporal precision in imitating target performances of the ornament, and whether visual feedback enhances the performance diversity during training. The study uncovered distinct strategies of exploration of the performance of the musical ornament and highlighted the dynamics of exploration behavior during training. Visual feedback enhanced exploration of temporal characteristics and influenced imitation accuracy. This study opens up educational possibilities for the training of performance skills and provides direction for further investigation of creative processes in performance.

Received January 29, 2011, accepted March 22, 2012.

Key words: performance, visual feedback, training, ornamentation, creativity

Music performance is a skill that requires a high level of temporal control (e.g., Palmer, 1997) and one of the aims of performance training software is to assist in the acquisition of such temporal precision. Moreover, advanced performance of music also requires high levels of creativity, flexibility, and imagination (e.g., Keller & Appel, 2010; McPherson, 1995) – when performing with others and when developing a personal (and original) interpretation of music. These set further demands for performance training software as will be discussed below.

Creativity in performance of classical music is related to the interpretation of music as it is notated in a score. The creative potential of performers (in the sense of the ability to generate diverse, unusual, and original performances) may seem restricted by the musical notation compared to, for example, the potential variety and originality in improvised music. Even so, performing music from a score provides musicians with considerable expressive freedom: Although the notes are the same in any two performances from the same score, the timing of simultaneous and sequential events may be varied, as well as the intensity of events, their duration, and, for some instruments, their intonation. The combinatorial possibilities for expression are in principle endless.

To some extent, performers tend to explore only a particular area of the hypothetical performance space as suggested by found performance rules (e.g., Repp, 1992a, 1992b; Sundberg, Askenfelt, & Frydén, 1983; Sundberg, Friberg, & Frydén, 1991; Todd, 1985; 1992). Nevertheless, over the centuries performance styles change considerably and, between performers, performance strategies may differ widely (e.g., Fabian, 2003; Philip, 1992; Timmers, Ashley, Desain, & Heijink, 2000).

A special instance of performance expression relates to the interpretation of musical elements whose notation is underspecified. Although this concerns music notation in general, it concerns ornamental notes in particular. Ornaments are notated by special symbols that refer to the kind of ornament, but do not contain the explicit rhythmic prescription that is otherwise used in music notation. This concerns both the onset timing of ornaments and the time interval between the ornament onset and the onset of the subsequent note. Additionally, as with all notes, the intensity of the event relative to the intensity of other events is not prescribed.

Several conventions have been established by musicians related to the performance of ornaments. For example, it is common in performance treatises to distinguish

Music Perception volume 30, issue 2, pp. 71–88. issn 0730-7829, electronic issn 1533-8312. DOI: 10.1525/mp.2012.30.2.71

between ornaments that are accented and long, and conventionally performed "on the beat," and ornaments that are unaccented and short, and often performed "before the beat" (Donington, 1989; Neumann 1986). Indeed, recently, we showed that the timing of one-note ornaments clusters into two categories – even for the same musical fragment – that show distinct temporal positioning of the ornament and distinct durations of the ornament (Desain & Timmers, 2008).

Having established an interpretation of the music, performers show amazing consistency over repeated performances, which may even approach the limits of motor control and just-noticeable differences in perception (Desain & Timmers, 2008). However, before establishing an interpretation, performers often explore different possibilities. Although musicians perform music with expressive variations even when sight-reading (e.g., Sloboda, 1983), advanced performers do spend considerable time defining, shaping, and refining an interpretation, especially when preparing for a concert performance (Chaffin, Imreh, Lemieux, & Chen, 2003). This process often includes the exploration of alternatives (see also Davidson & King, 2004). It is this process of exploring, shaping, and refining that we interpret as a creative activity, that demands considerable resources of concentration, goal-setting, self-evaluation, and problem-solving ability (Chaffin & Lemieux, 2004).

When developing music performance visual feedback systems that aim to improve student performance skills, it is important to keep the need for flexibility, intentionality, and creativity in mind. Tools that provide automatic feedback have the tendency to reinforce norms, such as when presenting metronomic timing or equal temperament tuning as a reference. The implication is that there is one correct way to perform the music (e.g., Wilson, Lee, Callaghan, & Thorpe, 2008). Alternatively, performance characteristics may be visualized without an indication of a norm (as when using a spectrogram to show intensity and timbre) providing no explicit feedback (see also Wilson et al., 2008).

An example of a flexible system with explicit feedback is the system developed by Juslin and colleagues (e.g., Juslin, Friberg, Schoonderwaldt, & Karlsson, 2004). This system evaluates the emotional quality of a performance by comparing the expressive characteristics of a performance with optimal expressive characteristics per emotional category. If a performance is fast, loud, and staccato, it receives a high evaluation on the anger scale, while if the performance is slow, soft, and legato it is evaluated as sad. This system promotes contrasting performing styles that communicate different emotional intentions.

Similarly, within the Practice Space project, we have developed a number of visual feedback prototype systems that require music students to perform the same music with a variety of expressive interpretations (Brandmeyer, Timmers, Sadakata, & Desain, 2011; Sadakata, Hoppe, Brandmeyer, Timmers, & Desain, 2008). For example (in Brandmeyer et al., 2011), drum students performed standard drum patterns with three different interpretations (laid-back, rushed, or on the beat). While playing, visual feedback was given using a visual form whose shape indicated the proximity of the student performance to the target performances. Each target interpretation was represented by a specific shape. If the student performance was very similar to one of the target interpretations, the visual form would resemble the respective shape, while if the student performance was somewhere in between target interpretations, the visual form would have an intermediate shape. The feedback was updated in real-time. The performance space was determined by expert performances of the three target interpretations and a residual category derived from beginner performances of the material, but the space could be expanded by extrapolating interpretations or by adding additional target performances. It was shown to be useful for advanced students who improved their skills by imitating the target performances.

Here we report on the effectiveness of a visual feedback program that provides feedback on the timing of a one-note musical ornament (a grace note). In the development of the ornament visual feedback system, we took the idea of enhancing expressive diversity one step further and asked participants to explore grace note timing as widely as possible as part of "an exploration training." Our aim was to test the effect of visual feedback and the effect of the instruction to *explore* ways of timing the ornament on subsequent temporal precision in timing the ornament. Precision in timing the ornament was measured before and after the training and training was done with or without visual feedback. Many issues are involved in developing feedback systems for education, such as the demands and cognitive load of the learning environment (Chandler & Sweller, 1991), the intuitiveness of the visual feedback, the target group of music students, or the extraction of relevant performance features for training. These issues are not explicitly considered in this paper, but they have been in previous papers (Brandmeyer et al., 2011; Sadakata, Brandmeyer, Timmers, Miezlaiskyte, & Desain, 2008).

Previous research has demonstrated the potential effectiveness of a training procedure that uses a diversity of instances. For example, studies demonstrated enhanced skill acquisition after training using varied (e.g., variety of contexts, multiple speakers) rather than singular examples in motor skills acquisition (Magill & Hall, 1990) and speech perception (Logan, Lively, & Pisoni, 1991; Sadakata & McQueen, 2011).

Moreover, we take diversity in performance as a positive sign of skill. Rather than interpreting variation around a central tendency as a residual of unexplained noise, we measure the diversity in performances and interpret it as an indication of creative exploration, and as a possible contributor to skill enhancement. Although diversity is not sufficient for creativity, it is an important contributor to it: Tests of creativity often include an assessment of participants' ability to generate divergent examples (e.g., Runco, 1986; Torrance, 1974). Individuals who transcend conventions and transform established behavior (such as Picasso or Stravinsky) often create a large number of works varying in quality and originality (Gardner, 1993). We think that exploration and creating a diversity of works (or performances) is trainable, and may be facilitated by feedback on behavior. In the longer term, diversity training may even "make creativity" in a similar sense that "practice makes perfect" (e.g., Ericsson, Krampe, & Tesch-Römer, 1993), while acknowledging that several social as well as personal factors influence creativity and are a prerequisite for creative activity (Runco, 2004).

Additionally, we assume that flexibility forms a necessary phase within skill development, and that diversification may improve performance. These hypotheses are derived from the assumption that skill development shows analogies with a dynamic system that has phases of equilibrium and stability and transitional phases, resulting in non-linear changes. Within-participant instability may be indicative of multiple attractors for behavior and transitional phases between them (e.g., Kelso, Scholz, & Schöner, 1986). Seen from a developmental perspective, instability may be indicative of a transitional phase towards new behavior (Thelen & Smith, 1994; Thelen, Ulrich, & Niles, 1987). Instability may also be a necessary stage and, in a sense, facilitate the transition to a new phase in skill development.

These are large themes for an investigation that is limited in focus. Nevertheless, we see them as relevant and important for our understanding of music performance as a cognitive skill and accordingly for the development of a system for performance education. Diversification of performance behavior is a way of being p-creative (person-specific creativity, see Boden, 2004); Performers vary their production of and their perspective on music without necessarily considering the relationship of their performances to performances by others. This variation may additionally lead to changes and potentially to improvements in the underlying performance skill. Online visual feedback is understood to be an effective means to provide immediate knowledge of results (e.g., Welch 1985). Such extrinsic or enhanced feedback has been demonstrated to improve learning in different domains, e.g., learning of bimanual coordination tasks (Swinnen, Timothy, Verschueren, Serrien, & Bogaerds, 1987), learning of new vocabulary (Metcalfe & Kornell, 2007), or second language learning (Hirata, 2004), although feedback is not effective in all circumstances (Schmidt, 1991; Schmidt & Bjork, 1992). Additionally, automatized feedback may enhance motivation and effort, foster precision, and build confidence (Pennington, 1999). In the current study, we assume that visualization of performance characteristics may benefit performers in a number of ways: it provides a visual trace of the performance, which remains present for reflection after the performance. This may facilitate reflection on the performance and strategy formation to vary different aspects of the performance. It may also assist in providing knowledge of results for participants to see whether the intended variation was realized and what was done differently than anticipated.

The focus of the visual feedback is the timing of the melody notes surrounding and including the grace note. Building on previous research (Timmers, Ashley, Desain, Honing, & Windsor, 2002), we take the timing of the music without the grace note as reference to interpret the timing of the music with the grace note. The feedback shows this relationship between the timing of the melody notes with and without the grace note. The student is instructed to "explore" grace note performance in the training session. The effectiveness of the exploration training is examined by comparing the ability of students to imitate nine target performances of the music with the grace note before training and after training (within-subjects pretest posttest design). A between-subjects design was used to test the effect of visual feedback on the exploration training and the ability to imitate the target performances in the posttest. Success of imitation was measured for different aspects of the grace note timing separately to gather additional information about the aspects that performers are sensitive to and the aspects that the visual feedback reinforces.

Method

PARTICIPANTS

Twenty-four piano students of the Royal Conservatoire of The Hague participated in the experiment. Twelve were randomly assigned to the visual feedback group and the the remaining twelve were assigned to the control group. The mean age was 24.25 (*SD* = 4.00, min = 18, max = 30) for the visual feedback group (VFB group) and 26.75 (*SD* = 5.05, min = 20, max = 37) for the control group. Mean years of piano training was 17.58 (*SD* = 3.96, min = 12, max = 25) for the VFB group and 17.17 (*SD* = 6.39, min = 7, max = 30) for the control group.

MUSICAL MATERIAL

Two excerpts from the Theme of Beethoven's "Paisiello Variations" (WoO 70) were used as musical material. Each excerpt contains one grace note (see Figure 1). These excerpts are directly comparable, coming from the same piece of music. Nevertheless, we know from previous studies that the two grace notes afford different treatment (Desain & Honing, 1994; Desain & Timmers, 2008; Timmers et al., 2002; Windsor, Desain, Aarts, Heijink, & Timmers, 2001).

The two excerpts were performed in a moderate tempo (dotted eighth note is 60 bpm, 1 s in duration, which means that eighth notes are 333 ms in duration). The tempo was indicated by a mechanical piano performance of the first bar of the excerpt (illustrated as cross notes in Figure 1). The participants continued performing the excerpt after this mechanical introduction.

PROCEDURE

A pretest posttest design was used to examine the effectiveness of the exploration training in improving performance – specifically, the ability of participants to imitate nine target performances in the posttest compared to the pretest. Additionally, a between-subjects design was used to test the effect of visual feedback on this improvement from pretest to posttest. Participants of the visual feedback group received visual feedback during training, but not during the pretest and posttest. Figure 2 illustrates the visual feedback: the feedback shows the relative timing of the melody notes with the grace note (top line) compared to the melody notes without the grace notes. The feedback was online and given while the participant was performing: the timing of the melody notes with the grace note appeared while performing.

To establish a baseline performance of the excerpt, participants first performed the excerpt a few times without the grace note until they were satisfied with their performance. For participants of the visual feedback group, this last performance without the grace note was stored for future use. Next, participants did the pretest imitation task. In this task, participants heard nine target performances and were asked to imitate each target as precisely as possible. The order of target performances was randomized over participants. They imitated each target immediately following its presentation and had only one trial per imitation. The only difference between the nine targets was a difference in timing of the ornament and its immediately surrounding notes.

An exploration training followed the pretest imitation task. During this training session, the participants of the

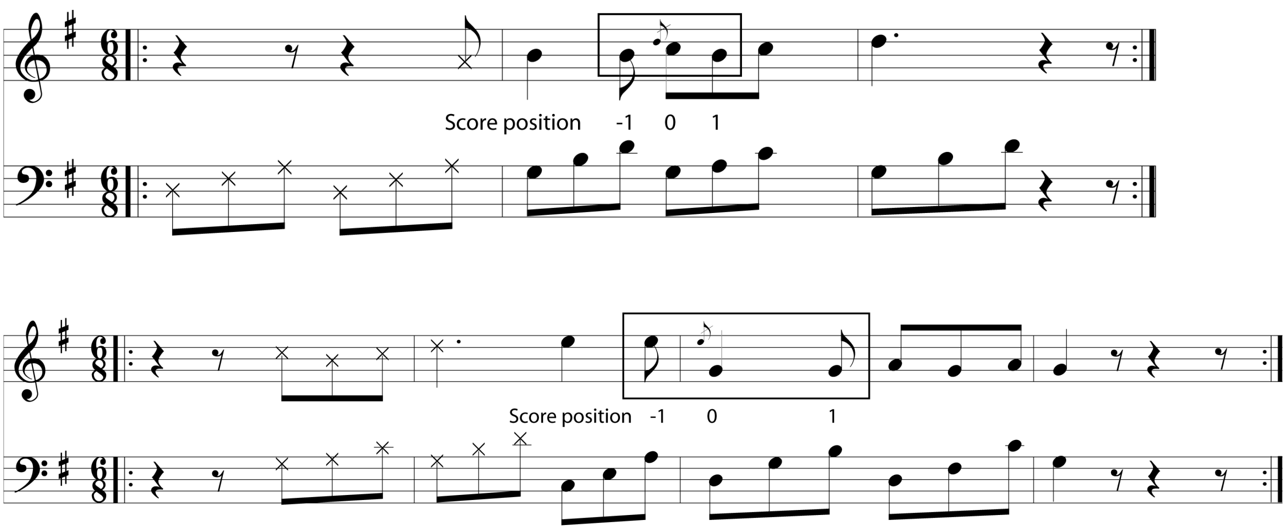

FIGURE 1. Scores of Excerpt 1 (top) and Excerpt 2 (bottom). The first bar of each excerpt is performed mechanically by a computer to indicate the tempo (notes indicated with crosses). The timing of the melody notes within the rectangle (the previous note, grace note, main note, and next note) is visualized in the visual feedback.

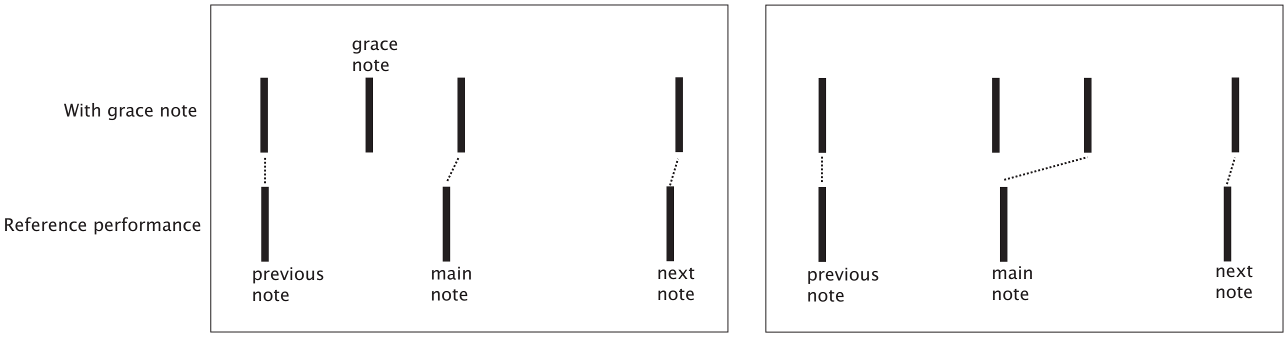

FIGURE 2. Illustration of two performances of the grace note as used in the instruction and the visual feedback. In the actual visual feedback, the previous, grace, main, and next note were indicated by different colors instead of verbal labels, corresponding to colors used in the scores of the excerpts (shown above the visual feedback), and no dashed lines were used to connect the notes.

experimental group received visual feedback on the timing of their performances. The training consisted of 36 trials. A trial always started with the mechanical introduction to indicate the tempo, followed by a performance of the rest of the excerpt by the participant, including the grace note. Participants were instructed to explore the performance of the grace note as widely as possible, to focus on the timing and duration of the grace note and the surrounding notes, and to look for extremes as well as in-between performances. The exploration training was followed by the posttest imitation task, which was a repetition of the pretest imitation procedure.

The same procedure was followed for Excerpts 1 and 2. The instructions, scores, and visual feedback (if applicable) were shown on a flat-screen placed on the MIDI-grand piano in front of the pianist (in the normal position for scores). The entire experiment was self-paced: Participants indicated their desire to move on to the next instruction screen or trial by pressing a specified piano key. The order of the two excerpts was counterbalanced across participants. The duration of the experiment was slightly over 30 min.

VISUAL FEEDBACK

Participants of the VFB group received real time visual feedback during the exploration training on the timing of the grace note and the melody notes surrounding the grace note (the previous note, main note, and next note). No visual feedback was provided during the pretest and the posttest. During the entire trial, the timing of the surrounding melody notes measured from the performance without the grace note was shown on the screen as reference. This included the note preceding the grace note, which was to be inserted (we refer to this note as the "previous note"), the note following the grace note to be inserted (referred to as the "main note" because it is the note that the grace note embellishes), and the note following the main note (referred to as the "next note").

The timing of the notes with the grace note was visualized while the pianist was performing. The onset of a note was indicated with a vertical colored line. The color indicated which note was performed. The horizontal position indicated the temporal position of the note measured from the onset of the note we call the "previous note." Temporal placement was from left to right: later onsets were further to the right (see Figure 2 for two examples). The feedback remained visible until the start of the next trial.

The visual feedback was explained by way of the two examples given in Figure 2. The control group was also given these two visual examples to explain the exploration task. This was to ensure that both groups received identical instructions for the exploration training.

TARGET PERFORMANCES

In the pretest and posttest, participants imitated nine target performances out of a total of 18 target performances. These target performances were constructed using a model of grace note timing (Timmers et al., 2002) that takes a performance without a grace note and adds a grace note according to specified timing parameters. The timing parameters concern the duration of the grace note (ioi_G) and the time-steal characteristics. These time-steal characteristics relate to the changes in the timing of the melody notes if a grace note is inserted (compared to the timing of the melody notes without the grace note) and to the temporal position of the grace note with respect to the original timing of the melody notes. Starting with the latter, the onset of the grace note can be before the original timing of the main note, in which case we state that it takes time from the previous note (it falls within the time originally allocated to the previous note). The time stolen from the previous note is positive ($p > 0$). Alternatively, the onset of the grace note can coincide with the original timing of the main note, in which case the time stolen from the previous note is nil ($p \approx 0$). This situation often leads to shortening of the main note interonset interval ($m > 0$). Instead of shortening the main note interonset interval, the main IOI may keep its duration and the subsequent note onset may be delayed by a period proportional to the grace note duration: The addition of the grace note leads to an insertion of time ($i > 0$) rather than a taking of time. In reality, a combination of these strategies may be used.

As explained, the three timing parameters (p, m, and i) are estimated by comparing the onset timing of notes in the performance with the grace note to the onset-timing of notes in the performance without the grace note. The timing differences are expressed as proportions of the grace note duration (ioi_G) and the three proportions sum to 1. In this way, p, m, and i indicate the proportion of the grace note that is timed by taking time from the previous note or main note or by inserting time. Note that a prerequisite of this procedure is that the base-tempo of the performances are very close. Otherwise, the timing parameters are additionally affected by tempo differences between the conditions.

The target performances for the experiment were constructed based on the observed variability within the performances of sixteen professional pianists as recorded and analyzed previously (Timmers et al., 2002). For example, the grace note of Excerpt 2 was on average longer in duration than the grace note of Excerpt 1. An original performance was used without a grace note by one of the pianists who showed a standard high quality performance of the music. The grace note was added in 18 different ways to create 18 target performances: a short, medium, or long grace note was performed before the original main note timing, at the original main note timing or somewhere in the middle, and this was done with some inserted or subtracted time to add some additional variation. See Table 1 for a summary of the settings. From these 18 versions, two sets of nine target performances were selected. A participant received one of the sets for Excerpt 1 and the other set for Excerpt 2. The order of the two excerpts was counter-balanced across participants.

TABLE 1. **Parameter Settings for the Two Sets of Nine Variants of Grace Note Timing.**

Number	Set	G_{ioi} Excerpt 1	G_{ioi} Excerpt 2	p	m	i
1	1	35	83	–0.1	0.9	0.2
2	2	62	113	–0.1	0.9	0.2
3	1	90	144	–0.1	0.9	0.2
4	2	35	83	0.1	1.1	–0.2
5	1	62	113	0.1	1.1	–0.2
6	2	90	144	0.1	1.1	–0.2
7	2	35	83	0.5	0.3	0.2
8	1	62	113	0.5	0.3	0.2
9	2	90	144	0.5	0.3	0.2
10	1	35	83	0.7	0.5	–0.2
11	2	62	113	0.7	0.5	–0.2
12	1	90	144	0.7	0.5	–0.2
13	1	35	83	0.9	–0.1	0.2
14	2	62	113	0.9	–0.1	0.2
15	1	90	144	0.9	–0.1	0.2
16	2	35	83	1.1	0.1	–0.2
17	1	62	113	1.1	0.1	–0.2
18	2	90	144	1.1	0.1	–0.2

Note: The length of short, medium and long grace notes are different for the two excerpts. The other parameter settings of p, m and i were shared.

Data Analysis

PERFORMANCE FEATURES

The analysis of the performances focused on the timing of the grace note and its surrounding melody notes to parallel the variation of timing parameters in the target performances. However, in the exploration training, participants did not restrict their exploration to timing. They also varied the dynamics of the grace note and its surrounding notes. Therefore, in the analysis of the exploration performances, force of keystrokes (measured as key-velocity in MIDI instruments) was also taken into account. Figure 3 gives a schematic of the timing (t) and key-velocity (v) parameters involved in the feature calculations. These relate to the performance of the previous, grace, main, or next note of the melody (capital letters) or accompaniment (small letters).

The first feature is the grace note duration or grace IOI. This is the interval between the onset of the grace note (t_G) and the main note (t_M):

$$ioi_G = t_M - t_G \tag{1}$$

The second feature is the proportion of the grace note stolen from the previous note (p). In contrast to our procedure in previous studies, due to tempo variations, the performances with the grace note could not directly be compared to the performances without the grace note. Therefore, two alternative definitions of stolen proportions were calculated: a stolen proportion based on the predicted timing of the melody notes given the durations of the initial notes of the excerpt, and a stolen proportion

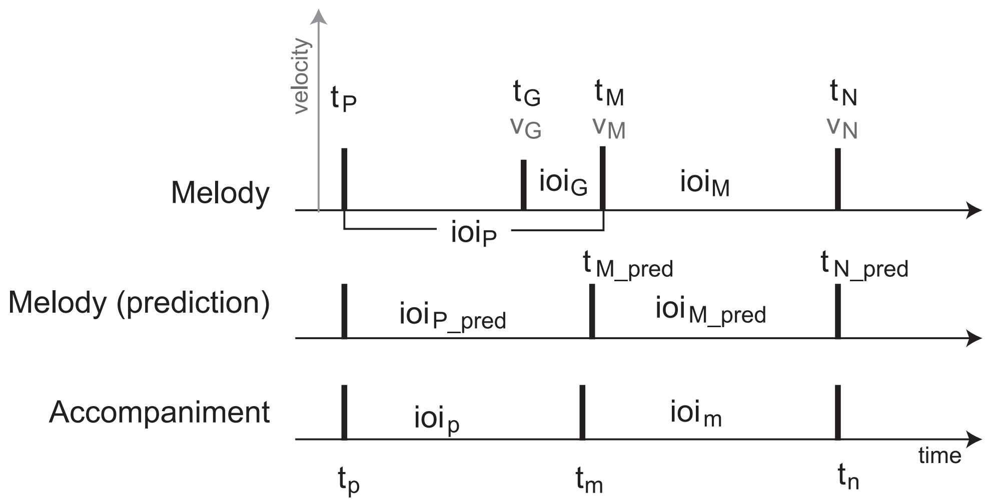

FIGURE 3. Schematic of feature parameters related to the onset timing (t), interonset interval (IOI), or key-velocity (v) of the previous (*P*), grace (*G*), main (*M*), and next (*N*) note belonging to the melody (top, capital letters) or accompaniment (bottom, small letters). Timing parameters are given in black and key-velocity parameters in grey. The predicted timing of the melody notes is related to the predicted onsets of the main and next note with respect to the previous note based on the tempo of the first notes of the performance.

based on the position of the accompaniment notes. The latter definition makes use of the practice that the accompaniment notes are timed "on time," (they keep tempo) while the melody notes are timed more flexibly (e.g., Hudson, 1994). While this type of rubato is not very common in modern performances of Western classical repertoire, it is intrinsic to ornament timing: the timing of the accompaniment notes remains relatively unaffected by the addition of an ornament.

The first proportion stolen from the previous note is calculated by comparing the onset timing of the grace note with the predicted onset timing of the main note:

$$p_1 = \frac{t_{M\_pred} - t_G}{ioi_G} \tag{2}$$

The second proportion stolen from the previous note is calculated by comparing the onset timing of the grace note with the onset timing of the accompaniment note:

$$p_2 = \frac{t_m - t_G}{ioi_G} \tag{3}$$

The first proportion stolen from the main note IOI was calculated by comparing the main note IOI to the predicted main note IOI (Eq. 4), while the second proportion stolen from the main note IOI compared the main note IOI to the main note IOI of the accompaniment (Eq. 5).

$$m_1 = \frac{ioi_{M\_pred} - ioi_M}{ioi_G} \tag{4}$$

$$m_2 = \frac{ioi_m - ioi_M}{ioi_G} \tag{5}$$

The first proportion inserted (i_1) was calculated by comparing the onset of the next note to the predicted onset of the next note (Eq. 6), while the second proportion inserted (i_2) compared the onset of the next note to the onset of the next note in the accompaniment (Eq. 7).

$$i_1 = \frac{t_N - t_{N\_pred}}{ioi_G} \tag{6}$$

$$i_2 = \frac{t_N - t_n}{ioi_G} \tag{7}$$

These pairs of features are associated but not overlapping and do not need to correlate. For example, if the addition of a grace note leads to the addition of time relative to the predicted time of the melody notes (i_1), this may not need to affect the relationship between the accompaniment and main note (i_2). The opposite may be true as well: no time is added with respect to the predicting onset timing (i_1 is close to 0). Nevertheless, the relative onset of the melody note following the main note is displaced with respect to its associated accompaniment note (affecting i_2).

In addition to these seven timing features, four key-velocity features were defined that characterize the dynamics of the grace note and surrounding melody notes. These features are only used in the analysis of the exploration training. The features measure the relative accenting (Δv) of the grace note (v_G in Eq. 8), main note

(v_M Eq. 9), or next note (v_N Eq. 10) compared to the average key-velocity ($\bar{v}$) of the initial melody notes (notes up to the grace note) and the difference in key-velocity of the grace note and the main note (Eq. 11).

$$\Delta v_G = v_G - \bar{v} \quad (8)$$

$$\Delta v_M = v_M - \bar{v} \quad (9)$$

$$\Delta v_N = v_N - \bar{v} \quad (10)$$

$$\Delta v_{GM} = v_G - v_M \quad (11)$$

OUTLIER ANALYSIS

Performances that were incomplete or did not contain the correct notes were excluded from further analysis. One of the participants performed the grace notes with an additional note in the pretests and the performances of this participant had to be excluded from the pre and posttest analyses. A minority of 4.4% of the pretest and posttest performances was excluded. For the training data, a minority of 5.8% of the 864 trials per excerpt (36 x 24) was excluded or not performed.

Additionally, an outlier analysis was run for each feature. If a feature value was larger than three standard deviations above the mean or smaller than three standard deviations below the mean, the value was replaced by the maximum of three standard deviations above or below the mean, respectively. This procedure tempered the influence of outliers without deleting data.

Results

The results section is split into five subsections. The first two focus on the imitation of the target performances in the pretests and posttests and examine the effects of feedback, training, and excerpt. The second two subsections are related to the data of the exploration training. Strategies of exploration are distinguished using a factor analysis. This is followed by a characterization of the exploration of each factor and an examination of the influence of visual feedback on exploration strategy. The final subsection examines the relationship between behavior in the exploration training and in the pre and posttests.

IMITATION OF THE TARGET PERFORMANCES

The ability of the performers to imitate the target performances was assessed per feature and per test using the correlation between the target feature values and participants' performance feature values. The correlation values were transformed using Fisher's transformation to obtain normally distributed profiles. Each participant imitated nine target performances per test, and therefore the correlations are based on nine data points (or sometimes eight data points if one of the performances was deleted). Using correlation as a measure of imitation accuracy has the benefit of indicating whether performers varied a feature in the instructed direction irrespective of the average performance manner. It is more informative than taking only the average error between the performed values and the target values because it indicates in what direction a feature was varied in relation to the instructed variation. Moreover, correlations have been used previously by other authors to assess similarity in timing profiles (e.g., Clarke, 1993; Repp, 2000).

Table 2 shows the mean (Fisher transformed) correlations between target and imitation across participants for the first test (first pretest) and the last test (second posttest). The table shows that the only feature that was accurately imitated by the majority of the participants was the grace IOI (ioi_G). A mixed-model ANOVA with test as within-subjects factor and feedback as between-subjects factor confirmed that the imitation accuracy improved from pretest to posttest for the grace note IOI, $F(1, 21) = 7.91$, $p = .01$, $r = .52$. There was no interaction between test and feedback, nor a main effect of feedback.

Four other features also showed a significant increase in correlations from pretest to posttest. First, p_2 showed a main effect of test, $F(1, 21) = 10.67$, $p = .004$, $r = .58$, a main effect of feedback, $F(1, 21) = 14.29$, $p = .001$, $r = .64$, and an interaction between feedback and test, $F(1, 21) = 9.87$, $p = .005$, $r = .57$. Despite the main effect of test, it was mainly the visual feedback group that improved between pre and posttest, having a particularly low correlation in the first pretest. The correlation in the last posttest was still low, especially for the feedback group.

Second, m_2 showed a significant main effect of test, $F(1, 21) = 11.62$, $p = .003$, $r = .60$, and a main effect of feedback, $F(1, 21) = 6.02$, $p = .023$, $r = .47$. Both groups improved imitation from pretest to posttest. However, the feedback group had overall a lower imitation score than the control group.

Third, i_2 showed a significant main effect of test, $F(1, 21) = 4.59$, $p = .044$, $r = .41$, and a significant interaction between test and feedback, $F(1, 21) = 4.42$, $p = .048$, $r = .42$. Despite the main effect of test, it was the visual feedback group that improved most between pre and posttest, reaching a high imitation score in the last posttest.

Finally, p_1 showed a main effect of feedback, $F(1, 21) = 4.57$, $p = .044$, $r = .42$, and a significant interaction between test and feedback, $F(1, 21) = 4.99$, $p = .037$, $r = .44$. The control group improved the imitation of this feature in the posttest compared to the pretest in contrast

TABLE 2. Mean Transformed Correlations Between Target Performances and Imitations for the First and Last Test of a Participants' Experimental Session.

	Control Group (N = 12)		Visual Feedback Group (N = 12)	
	Pretest	Posttest	Pretest	Posttest
ioi_G	0.47	0.84	0.65	0.91
p_1	−0.024	0.36	0.61	0.25
p_2	0.21	0.22	−0.59	0.023
m_1	0.059	0.15	0.22	−0.003
m_2	0.067	0.30	−0.40	0.074
i_1	−0.20	0.001	0.019	0.084
i_2	0.32	0.32	0.22	0.65

Note: Due to Fisher's transformation of the correlation values, values can become larger than 1 or smaller than −1.

TABLE 3. Frequencies of High Correlations (Untransformed) Between Target and Imitation (r > .60) per Feature and Test for the Control Group and the VBF group.

	Control Group (N = 12)		Visual Feedback Group (N = 12)	
	Pretest	Posttest	Pretest	Posttest
ioi_G	6	9	4	10
p_1	1	5	5	2
p_2	2	1		
m_1	2	2	3	1
m_2	1	4		
i_1	1		1	
i_2	2	2	2	5

to the feedback group, which showed a decrease in imitation scores. The initial imitation value (in the first pretest) for this feature of the feedback group performances was relatively high.

Generally, the average correlations were low for all features except grace note IOI, indicating that participants had great difficulty performing successful imitations. It might be that participants focused on the imitation of particular features (for example the grace note IOI), and neglected other features. Table 3 shows frequencies of high correlations between target and imitation ($r > .60$) per feature and test for the two groups. It shows that the number of high correlations was larger for the posttest and indeed was focused on particular features. The focus of high correlations was also different for the two groups, suggesting an effect of training as well as an influence of visual feedback on imitation accuracy: The feedback group showed in particular an improvement in the imitation of i_2, while the imitation of p_1 or m_2 was not reinforced. These particular differences between the two groups were not anticipated. As a posthoc explanation, we think that the feedback reinforced controlling the timing of the note following the main note in addition to the other more obvious notes (grace note and main note) by showing this note in the feedback. The control participants may have paid less attention to this aspect and would have focused on the timing of the grace and main note.

EFFECTS OF MUSICAL EXCERPT ON THE IMITATION OF THE TARGET PERFORMANCES

The above analysis examined the imitation of the target performances and the improvement from the first pretest to the last posttest of a participant, irrespective of which grace note and musical context were performed. However, for some features, musical excerpt with its corresponding grace note did have an effect on the correlations obtained. A mixed-model ANOVA was run with test (pre or posttest) and excerpt (Excerpt 1 or 2) as within-subjects factors and feedback as the between-subjects factor. Focusing on the effects of excerpt, a main effect of excerpt was found for p_1, $F(1, 20) = 18.61$, $p < .0001$, $r = .69$, and i_1, $F(1, 20) = 4.95$, $p = .038$, $r = .45$, and a significant interaction between excerpt and feedback was found for m_1, $F(1, 20) = 6.39$, $p = .02$, $r = .49$. Imitation was better for Excerpt 2 than Excerpt 1 for p_1 and i_1. For m_1, the effect of excerpt was different for the two groups: The control group showed higher correlations for Excerpt 1 than for Excerpt 2, while the VFB group showed higher correlations for Excerpt 2 than for Excerpt 1. The effects of test and feedback were already examined in the analyses above (when comparing between the first and last test of a participant) and were similar for the ANOVA reported here, except that test was now only marginally significant for grace note IOI (comparing the pretest and posttest of an excerpt, $F(1, 20) = 4.05$, $p = .058$, $r = .41$.

These effects of excerpt on imitation accuracy are for the most part as expected: The grace note and the main note are longer in Excerpt 2 than in Excerpt 1, which may facilitate the perception and production of variations in onset timing of the target notes (grace, main, and next note), which corresponds with the higher imitation accuracy of p_1 and i_1 for Excerpt 2. Additionally, the imitation of the main note IOI (m_1) was better for Excerpt 2 than for Excerpt 1 for the visual feedback group only. That is, relatively better as the imitation of this feature was not good in general, as observed earlier.

FACTORS OF EXPLORATION IN THE EXPLORATION TRAINING

The second set of analyses focuses on the training data and the strategies performers used to explore grace note performance, and possible influences of feedback and excerpt. Participants had 36 trials per musical excerpt to explore different ways of performing the grace note. To uncover general strategies of exploration, a principal component analysis was run on the eleven extracted features for each excerpt separately. Four dynamics features were included in addition to the seven timing features. The principal component analysis resulted in five components that had an eigenvalue greater than one. These five components were subjected to a varimax rotation to increase interpretability.

Table 4 shows the correlations between each rotated factor and each extracted feature. It also shows the percent explained variance per factor (bottom row of the table). For both excerpts, Factor 1 correlates strongly with m_2 and p_2 (negative correlation). Additionally, for Excerpt 1, Factor 1 correlates strongly with p_1 (negative correlation), while for Excerpt 2, Factor 1 correlates strongly with m_1. This factor captures variations in the stolen time characteristics: the grace note is timed within the "nominal" time of the previous IOI (it precedes the predicted onset of the main note and precedes the accompaniment main note) or it is timed by taking time from the main note – it shortens the main note IOI). For Excerpt 1, the variation in the onset timing of the grace note is done without affecting the main note IOI much, which is indicated by the limited correlation with m_1. This was is in line with a predominantly before the beat performance of the grace note in Excerpt 1 (see Timmers et al., 2002). In contrast, for Excerpt 2, the variations do affect the main note IOI, indicating that on-the-beat performance of the grace note in this excerpt is common.

Factor 2 correlates for both excerpts strongly with Δv_G and Δv_{GM}, and quite strongly with ioi_G. These correlations indicate that this factor captures variations in the accenting of the grace note.

Factor 3 correlates strongly with i_1 for both excerpts. Additionally, it correlates strongly with m_1 for Excerpt 1, and with p_1 for Excerpt 2. The signs of the correlations are opposite for i_1 and the other features. This indicates that this factor captures variations in the insertion of time with respect to the predicted onset timing of the next note. If more time is inserted, less time is stolen from the main note for Excerpt 1. For Excerpt 2, there is less time taken from the previous note if time is inserted (the grace note is timed late).

Factor 4 correlates for both excerpts most strongly with Δv_M and Δv_N, indicating that this factor captures variations in the accenting of the main note and the next note. The correlations with Δv_M are stronger than the correlations with Δv_N.

Finally, Factor 5 correlates strongly with i_2 for both excerpts. Additionally, it correlates with ioi_G for Excerpt 1, and with p_2 for Excerpt 2. This factor captures variations in the displacement of the next melody note compared to the next accompaniment note. For Excerpt 1, this displacement is often done in combination with a lengthening of the grace note IOI. Displacement between melody and accompaniment notes is a way to give additional time to the grace note while maintaining the rhythm in the accompaniment. For Excerpt 2, this displacement is done in contrast to a

TABLE 4. **Correlations Between Feature Values and Factors for Excerpts 1 and 2.**

	Excerpt 1					Excerpt 2				
	F1	F2	F3	F4	F5	F1	F2	F3	F4	F5
ioi_G	.00	**–.43**	.06	.08	**.69**	**.20**	**.61**	–.02	–.08	**–.26**
p_1	**–.82**	.05	**.37**	–.05	–.02	–.16	–.03	**–.94**	.02	–.03
p_2	**–.85**	.05	–.08	–.02	**–.39**	**–.63**	**–.15**	**–.12**	.09	**.68**
m_1	**.39**	–.05	**.84**	–.01	–.04	**.91**	.05	–.21	–.07	–.06
m_2	**.87**	–.06	**.21**	.04	**–.22**	**.85**	**.10**	**.36**	.00	**.23**
i_1	**.35**	.01	**–.93**	.04	.03	**–.63**	–.04	**.76**	.04	.03
i_2	.05	–.01	**–.10**	**.10**	**.93**	**–.17**	**.11**	**–.10**	–.08	**–.95**
Δv_G	.06	**–.89**	–.02	**–.36**	**.12**	–.03	**.89**	.04	**.30**	.02
Δv_M	–.06	**.13**	**–.10**	**–.91**	–.06	**–.11**	**–.18**	.05	**.89**	.05
Δv_N	.03	**–.11**	**.13**	**–.54**	–.09	.02	.08	–.04	**.76**	.06
Δv_{GM}	.10	**–.89**	.06	**.37**	**.16**	.05	**.85**	–.00	**–.45**	–.02
%var	22.2	16.5	16.4	15.1	14.5	22.4	18.1	15.1	15.4	13.6

Note: Significant correlations are indicated in bold (N = 823 for Excerpt 1 and N = 804 for Excerpt 2). Correlations larger than .10 are significant at .004 level, which corresponds to a p value of .05, taking multiple testing per factors into account. The bottom row shows the percent explained variance for each factor.

displacement between the grace note onset and the main accompaniment note, as indicated by the opposite signs of these correlations: the more the next melody note is delayed compared to the next accompaniment note, the less the grace note precedes the main accompaniment note. In other words, both the grace and the next melody note are timed later or earlier with respect to the accompaniment notes.

EXPLORATION OF FACTORS

An overview of the exploration of each factor is given per feedback group in Figure 4. It shows the contours of the distributions of the data per factor, group, and excerpt. The distributions for Factors 3–5 are close to a normal distribution with one clear peak in the middle, although the distributions are leptokurtic, especially the distributions of Factor 5 in both excerpts. The distributions of Factors 1 and 2 show two local maxima rather than one global maximum. These correspond with two opposite interpretations of the grace notes of Excerpts 1 and 2: for Factor 1, the grace notes were either performed by predominantly stealing time from the previous note or by predominantly taking time from the main note. For Factor 2, the grace notes were relatively accented compared to the main note, or the main notes were relatively accented. These two opposite interpretations and interpretations in between occur relatively often, leading to a broad and flat distribution.

The strongest difference in data distribution across feedback groups concerns Factors 1 and 2 for both excerpts: The feedback group shows an especially wide distribution for Factor 1 and a relatively narrow distribution for Factor 2, compared to the control group. This means that participants from the visual feedback group varied the time-steal characteristics of the grace note timing more than participants from the control group. In contrast, the control group participants varied the dynamic accenting of the grace note performances particularly strongly, more so than the visual feedback group.

Comparison of the distributions using independent samples Mann-Whitney *U* tests confirmed that the distributions of data are significantly different for the two groups for all factors of Excerpt 1 (standardized test statistics were $U = -8.71$, $U = 10.63$, $U = -6.37$, $U = -5.61$, $U = 6.89$ for Factors 1–5, respectively, and $N = 823$, $p = .000$ for all factors), and for Factors 1, 2, 3, and 5 of Excerpt 2 ($U = -14.21$, $U = 8.85$, $U = -6.58$, $U = -5.01$, respectively and $N = 804$, $p = .000$,), but not for Factor 4 ($U = 0.58$, $N = 804$, $p = .559$).

This analysis includes all training trials of individual participants. A statistically more stringent measure is to define one measure of exploration per participant; for example, the variance in factor values across training trials. The difference across groups in factor variance was tested and was only marginally significant for Factor 1, Excerpt 1 (independent samples Mann-Whitney *U* test, standardized test statistics was $U = 1.85$, $N = 24$, $p = .065$, mean $\sigma = 1.68$ and 2.25, for the control and feedback group, respectively), and for Factor 2, Excerpt 2 (independent samples Mann-Whitney *U* test, standardized test statistics was $U = -1.67$, $N = 24$, $p = .094$, mean $\sigma =$ 1.66 and 1.28, for the control and feedback group, respectively). These trends are in the expected direction and indicate that the visual feedback allowed participants to focus on the exploration of the timing of the grace note and surrounding notes rather than focusing on varying the accenting pattern.

The limited statistical significance in this comparison between the two groups using collated data for individual participants is a result of strong individual differences between participants. Table 5 highlights these variations in the tendency to explore the different factors by showing the mean, minimum, and maximum variances of data of individual participants per factor and excerpt. Some participants were conservative and show similar factor values over trials as reflected in small variances, while other participants explored the factors widely as reflected in variances that are two to four times the average variances and 20 to 30 times the smallest variances. The most extreme case is Factor 5, Excerpt 2, for which most participants showed very small variances. Only a few participants explored this factor, as can also be seen from the pointy distribution (high kurtosis) shown in Figure 4.

Participants differed in the amount of variation within dimensions, and also in their mean factor values over training performances. Figure 4 shows the mean factor values for individual participants as crosses at the bottom of each graph, and highlights the considerable

TABLE 5. **Amount of Variation per Factor, as Given by the Mean, Maximum, and Minimum Factor Variance, and the Average Size of Changes Across Trials, as Given by the Median Intertrial Absolute Difference in Factor Values.**

	Excerpt 1				Excerpt 2			
	Mean Var.	Min Var.	Max Var.	Median Change	Mean Var.	Min Var.	Max Var.	Median Change
F1	1.97	0.27	6.19	0.92	2.09	0.36	8.08	0.85
F2	1.30	0.61	3.63	0.93	1.47	0.43	3.42	0.96
F3	1.66	0.26	5.27	0.79	1.47	0.16	5.47	0.62
F4	1.43	0.44	2.97	1.05	1.25	0.39	3.13	0.94
F5	1.18	0.24	5.09	0.67	1.22	0.05	6.04	0.38

Note: Unit is *z*-scores.

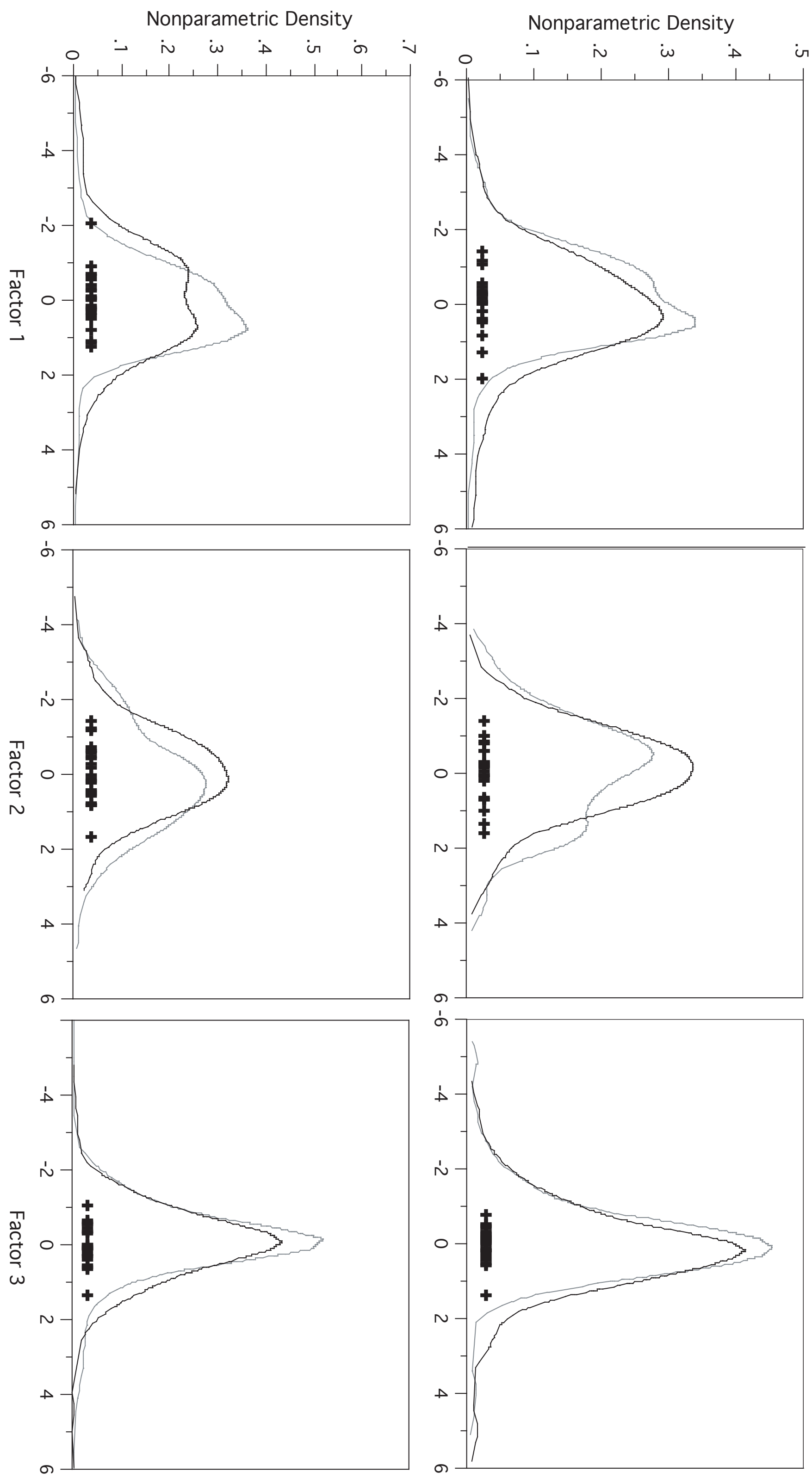

FIGURE 4. Continued.

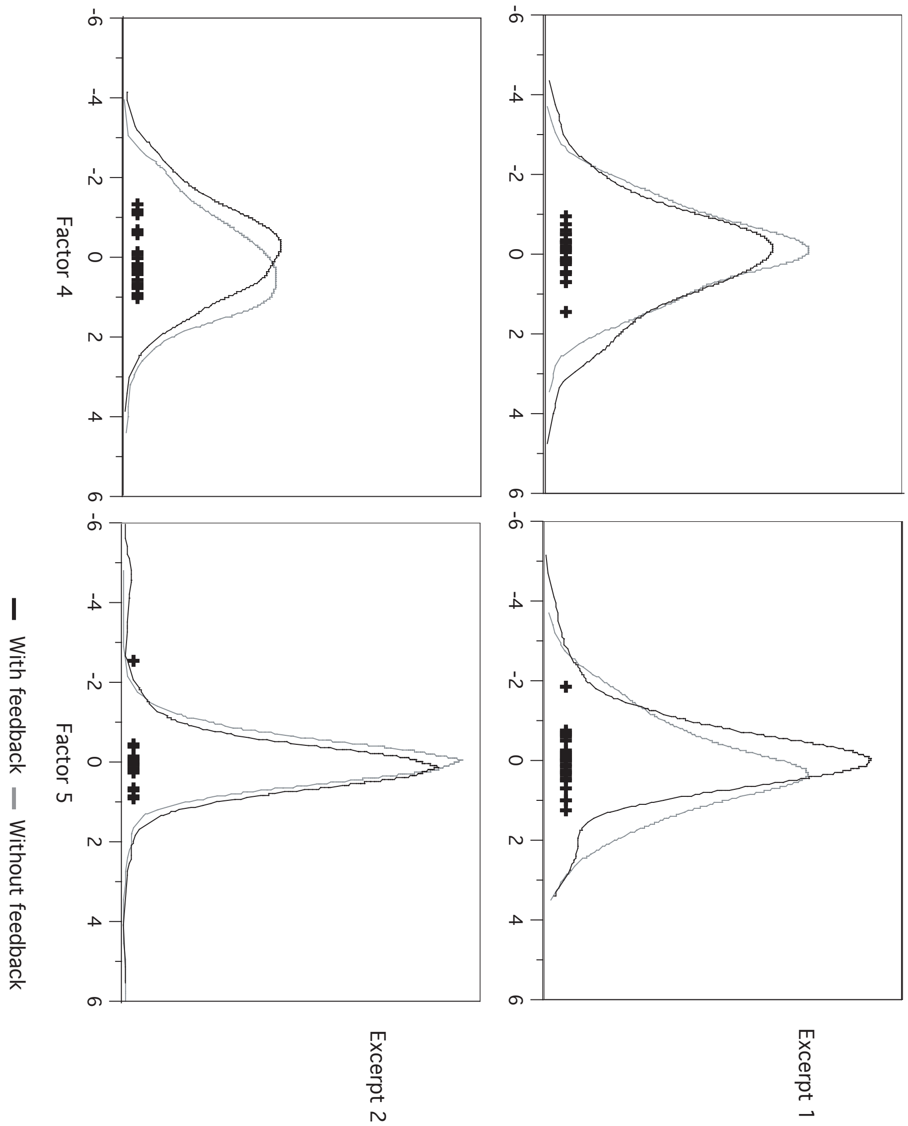

FIGURE 4. Nonparametric density of data distributions plotted by factor *z*-score for Factors 1-5. Distributions for the performances of the visual feedback group (black line) and control group (grey line) are plotted separately. Top graphs concern performances of Excerpt 1, and bottom graphs concern performances of Excerpt 2.

spread in means. The more extreme means correspond with contrasting interpretations of the music that were noted before. As will be seen below, these means were center points for individual pianists that were relatively stable.

Exploration of factors was further investigated by examining the (absolute) changes in factor values between successive training trials. This was done to get an impression of the degree to which participants explored the factors in a gradual manner or instead made abrupt changes in interpretation. The majority of absolute intertrial changes were small, with median values close to or below one standard deviation unit for all factors of both excerpts (see Table 5). These small intertrial changes were alternated with larger changes in interpretation to explore more widely. Participants showing wide exploration of factors used larger intertrial changes in factor values, as highlighted by a positive correlation between median intertrial changes values and standard deviation values of factors ($r = .69$, $r = .65$, $r = .82$, $r = .81$, and $r = .87$ for Factors 1–5 of Excerpt 1, and $r = .82$, $r = .87$, $r = .74$, $r = .78$, and $r = .91$ for Factors 1–5 of Excerpt 2).

Interestingly, there was also a positive association between absolute factor values and absolute intertrial changes for the respective factor. This means that outer factor values were reached by larger changes in interpretations, while smaller differences in interpretation were more strongly associated with central factor values ($r = .46$, $r = .30$, $r = .56$, $r = .38$, and $r = .46$ for Factors 1-5 of Excerpt 1, and $r = .47$, $r = .37$, $r = .50$, $r = .35$, and $r = .56$ for Factors 1–5 of Excerpt 2). As can be expected, corresponding correlations are stronger when distances to mean values of individual performers are correlated with absolute intertrial difference values ($r = .53$, $r = .48$, $r = .58$, $r = .45$, and $r = .56$ for Factors 1–5 of Excerpt 1, and $r = .53$, $r = .49$, $r = .57$, $r = .50$, and $r = .65$ for Factors 1–5 of Excerpt 2). This highlights regression towards the mean: the mean values for individual participants form a relatively stable center that is associated with smaller changes in performances, while outer values are associated with larger changes and therefore less stability.

Nevertheless, these correlations are still not very high, which is related to the previously mentioned predominance of smaller absolute differences in general. Interestingly, participants seemed to have two modes of performance: a repetition mode characterized by successive smaller difference values or an exploration mode characterized by successive larger difference values. Table 6 shows the probability of a smaller intertrial change and a larger intertrial change (> 1 standard deviation) as well as the first order probabilities of smaller changes following smaller or larger intertrial changes. As can be seen, the chance that a larger change follows a larger change is considerably higher than the probability of a larger change following a smaller change.

RELATIONSHIP BETWEEN EXPLORATION AND IMITATION

Finally, the relationship was investigated between participants' tendency to explore and their ability to imitate the target performances independent of the influence of feedback. Participants' imitation correlations per feature obtained in the pretest and posttest were correlated with their total factor variance during training (summed over the five factors). The results of both excerpts were included ($N = 46$ or 47, due to missing data in the imitation scores). Significant correlations included correlations with the imitation of features in the posttest: The exploration of factors was correlated with the imitation of ioi_G in the posttest ($r = .29$, $p = .05$) and the imitation of i_2 in the posttest ($r = .38$, $p = .009$). This suggests a positive relationship between the tendency to explore these particular features during training and the ability to imitate these features afterwards. Note that these are two of the three features that were relatively well imitated (see Table 3).

Summary and Discussion

We examined the effectiveness of a visual feedback system that does not reinforce a particular norm of playing, but that aims to enhance performance and train expression by promoting performance diversity. Performers' ability to imitate target performances improved with exploration training as indicated by higher correlations in the posttest than the pretest for several features. Generally, however, the correlations between target and imitation features were low, indicating that imitation of the features was a difficult task for even advanced piano students. The improvement in imitation accuracy was in particular observed for the imitation of the grace note IOI. Visual feedback was found to influence the quality of the imitations of particular features: It had both a positive and a negative effect. This qualitative effect of visual feedback on the imitation accuracy of particular features was also reflected in a different distribution of high correlations across the two groups in the posttest. As explained in the results section, we attribute this qualitative difference to the fact that the feedback emphasizes controlling the timing of the note following the main note in addition to controlling the timing of the other more obvious notes (grace note and main note) by showing this note in the feedback. The control participants may have paid less attention to this aspect and would have focused on the grace and main note timing.

TABLE 6. Probabilities of Smaller and Larger Intertrial changes (> 1 SD), as well as First Order Probabilities of Smaller or Larger Changes Following Smaller or Larger Intertrial Changes.

	Overall probability		Probability following a small change		Probability following a large change	
	Small change	Large change	Small change	Large change	Small change	Large change
Excerpt 1						
FB0	.58	.42	.68	.32	.45	.55
FB1	.53	.47	.60	.44	.40	.56
Excerpt 2						
FB0	.62	.38	.73	.27	.45	.55
FB1	.61	.39	.68	.32	.50	.50

Note: Probabilities are calculated across the five factors.

An influence of visual feedback was also found in a difference in distributions of data across exploration dimensions: Participants of the visual feedback group varied the time-steal characteristics of the ornament performance relatively extensively, while the participants of the control group varied the accenting pattern of the grace note relatively widely. This suggests that without visual feedback, the accenting pattern of the grace note is a relatively salient dimension. The visual feedback facilitates a stronger focus on the timing characteristics of the ornament performance as instructed. Thus, visual feedback could potentially be used to guide learners' attention towards an important dimension that could otherwise be overlooked. It is noteworthy in this respect that advanced performers have been found to control onset timing more deliberately in the communication of temporal structure than less advanced performers who primarily varied dynamics and articulation (Sloboda, 1983). It is also noteworthy that variations in timing and dynamics may be perceptually confused in the sense that a dynamic accent may be perceived as a lengthening, or a lengthening may have the perceptual effect of a volume increase. This suggests that without feedback, participants may have been unaware that they were not varying timing.

The results have highlighted characteristics of the dynamics of exploration (Thelen & Smith, 1994): it highlighted how the performances of participants fluctuated around relatively stable centers functioning as "attractor" zones. These center points of performance characteristics did not necessarily overlap for different participants, but could represent "opposite" interpretations. For example, participants may vary around an "on-the-beat" interpretation of the grace note, while others varied around a "before-the-beat" interpretation. Small-scale changes in performance characteristics were very frequent, resulting in a fine-scale "sampling" of the performance space. Some performers restricted the exploration to these fine changes and remained close to their central zones, while other participants showed less stability and their performances varied much in timing and dynamics characteristics. These latter performers tended to perform better in the imitation task following the exploration training. Thus, the exploration task did seem to benefit imitation skills, given the improvement in correlation for the imitation of the grace note IOI in particular, and a positive association between factor variance and imitation ability of participants in the posttest. Note, however, that no comparison was made with a condition without training.

The comparison between factor variances of individual participants emphasized personal (p-creativity) rather than historical creativity (h-creativity; Boden, 2004) by indicating how much participants deviated with respect to their own central tendencies. We compared the degrees to which participants explored the performance space as defined by the five factors of exploration and found that this was correlated with imitation performance in the posttest. Nevertheless, in the process of exploration, participants abandoned performance constraints and as a group they were, in our opinion, h-creative. Participants varied the different dimensions of the grace note timing more independently than expected: there was no association between grace accenting and the position of the grace note or between grace accenting and its duration. This contrasts with previous empirical findings as well as theoretical notions about grace note performance; accented graces are long and on the beat, while graces before the beat can only have a short duration (Donington, 1989; Neumann, 1978, 1986). Similarly, insertion of grace notes by inserting time was one of the explored tendencies. This is not an acknowledged way of performing grace notes in performance treatises, nor was it one of the performance strategies found previously (see again Donington, 1989; Neumann, 1978, 1986). In general, we were rather surprised with the inventiveness of our participants. They went far beyond the target interpretations presented in the pretest and the posttest. For example, the shortest grace IOI was less than 10 ms, and the longest 776 ms with a lower quartile of 47 ms, and higher quartile of 147 ms for Excerpt 1. The median of 73 ms was slightly longer than the mean grace IOI used in the target performances of Excerpt 1. To be truly creative, the explored performances should also be aesthetically valuable (Gardner, 1993). While this may be questionable for some of the performances, exploration did seem to have the additional benefit of contributing to performance precision.

Participants seemed to explore all of the options that were available: the proportion between time taken from the previous note and the main note, the relative accenting of notes, the insertion of pauses, and asynchrony with respect to the accompaniment. Although these exploratory dimensions may seem particular to ornament timing, they are actually expressive strategies that can be used in performance more generally; varying the relative rhythmic proportions of subsequent notes (as in the practice of notes inégales, Moelants, 2011), changing the duration contrast between rhythmic values (Sundberg et al., 1991), inserting a micropause (Sundberg et al., 1991), accenting using duration or intensity (Sloboda, 1983), or varying the melody while keeping tempo in the accompaniment (tempo rubato; Donington, 1989; Hudson, 1994), which results in varying asynchrony locally (Ashley, 2002; Timmers et al., 2000) for prominent use of this within a different musical genre).

These strategies are of a different and a more local nature than the expressive strategies seminally summarized by Clarke (1988), who emphasized patterning across a phrase. They may tap into the diversity of small-scale patterns rather than to the more frequently investigated and more common patterns across the phrase as identified by Repp (1992a).

For educational purposes, it is important to explicitly refer to these kinds of strategies and to be more aware of their applicability. Providing training in these strategies in a well-defined setting such as ornament performance might be a way of enhancing their usability. Building on research on music performance of ornaments as well as larger musical structures, an automatic evaluation and feedback system is now within reach that assists music students in enhancing their technical skills as well as motivating their exploratory tendencies (see also Brandmeyer et al., 2011). Such a system does not need to be normative, but can enhance flexibility and motivate strategies to explore different aspects of performance, which may actually additionally enhance performance precision.

Efficacy of diversity in training materials is acknowledged in learning of various skills (Logan, Lively, & Pisoni, 1991; Magill & Hall, 1990; Paas & van Marriënboer, 1994; Roller, Cohen, Kimball, & Bloomberg, 2001; Sadakata & McQueen, 2011). The advantage of processing a larger variety of training materials is the need to develop a more general and robust representation of a target task rather than reinforcing specialization, which limits the potential for transfer of learning (Logan et al., 1991). Exploring various ways to realize the same piece certainly increases diversity in training materials as does asking participants to explore something deliberately promotes an active form of learning (e.g., Dale, 1954).

We have illustrated an approach towards the investigation of the exploration of a conceptual space. Within the general conceptual space – as defined by the performances as a whole – individual participants explored particular areas (smaller or larger), generally making only small intertrial changes, but interlacing these with wider explorations. Despite the focus on a small-scale phenomenon (the timing of an ornament), we hope to have highlighted a number of important avenues in music research to investigate further: the ability to measure the exploration of a conceptual space in performance, the importance of creativity and intentional diversity in skill development, and implications of viewing skill development as a dynamic system.

Author Note

This research was funded by the Netherlands Society of Technical Sciences (STW 6301). We would like to acknowledge the support of Bart van Oort and Kees Tazelaar of the Royal Conservatory of The Hague in organizing piano students as participants and facilities for making the piano recordings. Last but not least, we are grateful to Alex Brandmeyer for his realization of the real-time visual feedback system.

Correspondence concerning this article should be addressed to Renee Timmers, 34 Leavygreave Road, Sheffield S3 7RD, United Kingdom. E-MAIL: r.timmers@sheffield.ac.uk

References

ASHLEY, R. (2002). Do[n't] change a hair for me: The art of jazz rubato. *Music Perception, 19,* 311–332.

BODEN, M. A. (2004). *The creative mind. Myths and mechanisms* (2nd ed.). London, UK: Routledge.

BRANDMEYER, A., TIMMERS, R., SADAKATA, M., & DESAIN, P. (2011). Learning expressive percussion performance under different visual feedback conditions. *Psychological Research, 75,* 107–121.

CHAFFIN, R., IMREH, G., LEMIEUX, A. F., & CHEN, C. (2003). "Seeing the big picture": Piano practice as expert problem solving. *Music Perception, 20,* 465–490.

CHAFFIN, R., & LEMIEUX, A. F. (2004). General perspectives on achieving musical excellence. In A. Williamon (Ed.), *Musical excellence: Strategies and techniques to enhance performance* (pp. 19–39). Oxford, UK: Oxford University Press.

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

CLARKE, E. F. (1988). Generative principle in music performance. In J. A. Sloboda (Ed.), *Generative processes in music: The psychology of performance, improvisation and composition* (pp. 1–26). Oxford, UK: Clarendon Press.

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

DALE, E. (1954). *Audio-visual methods in teaching.* New York: Holt, Rinehart, and Winston.

DAVIDSON, J. W., & KING, E. C. (2004). Strategies for ensemble practice. In A. Williamon (Ed.), *Musical excellence: Strategies and techniques to enhance performance* (pp. 105–122). Oxford, UK: Oxford University Press.

DESAIN, P., & HONING, H. (1994). Does expressive timing in music performance scale proportionally with tempo? *Psychological Research 56*, 285–292.

DESAIN, P., & TIMMERS, R. (2008). Stealing time: How grace notes can be added. In K. Miyazaki, Y. Hiraga, M. Adachi, Y. Nakajima,

& M. Tsuzaki (Eds.) *International Conference on Music Perception and Cognition 10 Abstract cards* (p. 27). Sapporo, Japan: Hokkaido University.

Donington, R. (1989). *The interpretation of early music*. London, UK: Faber and Faber.

Ericsson, K. A., Krampe, R. T., & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review, 100*, 363–406.

Fabian, D. (2003). *Bach performance practice 1945–1975: A comprehensive review of recordings and literature*. Aldershot, UK: Ashgate.

Gardner, H. (1993). *Creating minds: An anatomy of creativity seen through the lives of Freud, Einstein, Picasso, Stravinsky, Eliot, Graham and Gandhi.* New York: Basic Books.

Hirata, Y. (2004). Computer assisted pronunciation training for native English speakers learning Japanese pitch and durational contrasts. *Computer Assisted Language Learning, 17*, 357–376.

Hudson, R. (1994). *Stolen time: The history of tempo rubato*. Oxford, UK: Clarendon Press.

Juslin, P., Friberg, A., Schoonderwaldt, E., & Karlsson, J. (2004). Feedback learning of musical expressivity. In A. Williamon (Ed.), *Musical excellence: Strategies and techniques to enhance performance* (pp. 247–270). Oxford, UK: Oxford University Press.

Keller, P. E., & Appel, M. (2010). Individual differences, auditory imagery, and the coordination of body movements and sounds in musical ensembles. *Music Perception, 28,* 27–46.

Kelso, J. A., Scholz, J. P., & Schöner, G. (1986). Non-equilibrium phase transitions in coordinated biological motion: Critical fluctuations. *Physics Letters A, 118,* 279–284.

Logan, J. S., Lively, S. E., & Pisoni, D. B. (1991). Training Japanese listeners to identify English /r/ and /l/: A first report. *Journal of Acoustical Society of America, 89,* 874–886.

Metcalfe, J., & Kornell, N. (2007). Principles of cognitive science in education: The effects of generation, errors, and feedback. *Psychonomic Bulletin and Review, 14,* 225–239.

Magill, R. A., & Hall, K. G. (1990). A review of the contextual interference effect in motor skill acquisition. *Human Movement Science, 9,* 241–289.

McPherson, G. E. (1995). Cognitive strategies and skill acquisition in musical performance. *Bulletin of the Council for Research in Music Education, 133,* 64–71.

Moelants, D. (2011). The performance of notes inégales: The influence of tempo, musical structure, and individual performance style on expressive timing. *Music Perception, 28,* 449–460.

Neumann, F. (1978). *Ornamentation in Baroque and post-Baroque music: With special emphasis on J. S. Bach.* Chicago, IL: University of Chicago Press.

Neumann, F. (1986). *Ornamentation and Improvisation in Mozart.* Princeton, NJ: Princeton University Press.

Paas, F. G. W. C., & van Marriënboer, J. J. G. (1994). Variability of worked examples and transfer of geometrical problem-solving skills: A cognitive-load approach. *Journal of Educational Psychology, 86,* 122–133.

Palmer, C. (1997). Music performance. *Annual Review of Psychology, 48,* 115–38.

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

Philip, R. (1992). *Early recordings and musical style: Changing tastes in instrumental performance: 1900–1950.* Cambridge, UK: Cambridge University Press.

Repp, B. H. (1992a). Diversity and commonality in music performance – An analysis of timing microstructure in Schumann's Traumerei. *Journal of the Acoustical Society of America, 92,* 2546–2568.

Repp, B. H. (1992b). A constraint on the expressive timing of a melodic gesture: vidence from performance and aesthetic judgment. *Music Perception, 10,* 221–242.

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

Roller, C. A., Cohen. H. S., Kimball, K. T., & Bloomberg, J. J. (2001). Variable practice with lenses improves visuo-motor-plasticity. *Cognitive Brain Research, 12,* 341–352.

Runco, M. A. (1986). Flexibility and originality in children's divergent thinking. *Journal of Psychology, 120,*345–352.

Runco, M. A. (2004). Creativity. *Annual Review of Psychology, 55,* 657–687.

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

Sadakata, M., Brandmeyer, A., Timmers, R., Miezlaiskyte, A., & Desain, P. (2008). Longitudinal study of rhythm tapping with visual feedback. In K. Miyazaki, Y. Hiraga, M. Adachi, Y. Nakajima, & M. Tsuzaki (Eds.) *International Conference on Music Perception and Cognition 10, Abstract cards* (p. 228). Sapporo, Japan: Hokkaido University.

Sadakata, M., & McQueen, J. M. (2011). The role of variability in non-native perceptual learning of a Japanese geminate-singleton fricative contrast. In P. Cosi, R. De Mori, G. Di Fabbrizio, & R. Pieraccini (Eds.), *Proceedings of Interspeech*, (pp. 873–876). Florence, Italy: International Speech Communication Association.

Schmidt, R. A. (1991). Frequent augmented feedback can degrade learning: Evidence and inter- pretations. In G. E. Stelmach & J. Requin (Eds.), *Tutorials in Motor Neuroscience* (pp. 59–75). Dordrecht: Kluwer.

Schmidt, R. A., & Bjork, R. A. (1992). Common principles in three paradigms suggest new concepts for training. *Psychological Science, 3,* 207–217.

Sloboda, J. A. (1983). The communication of musical metre in piano performance. *Quarterly Journal of Experimental Psychology, 35,* 377–396.

Sundberg, J., Askenfelt, A., & Frydén, L. (1983). Musical performance: A synthesis-by-rule approach. *Computer Music Journal, 7*, 37–43.

Sundberg, J., Friberg A., & Frydén, L. (1991). Common secrets of musicians and listeners: An analysis-by-synthesis study of musical performance. In P. Howell, R. West, & I. Cross (Eds.), *Representing musical structure* (pp. 161–197). London, UK: Academic Press.

Swinnen, S. P., Timothy, D. L., Verschueren, S., Serrien, D. J., & Bogaerds, H. (1987). Interlib coordination: Learning and transfer under different feedback conditions. *Human Movement Science, 16*, 749–785.

Thelen, E., & Smith, L. B. (1994). *A dynamic systems approach to the development of cognition and action.* Cambridge, MA: MIT Press.

Thelen, E., Ulrich, B. D., & Niles, D. (1987). Bilateral coordination in human infants: Stepping on a split-belt treadmill. *Journal of Experimental Psychology: Human Perception and Performance, 13,* 405–410.

Torrance, E. P. (1974). *Torrance tests of creative thinking: Norms and technical manual.* Bensenville, IL: Scholastic Testing Services.

Timmers, R., Ashley, R., Desain, P., & Heijink, H. (2000). The influence of musical context on tempo rubato. *Journal of New Music Research, 29,* 131–158.

Timmers, R., Ashley, R., Desain, P., Honing, H., & Windsor, W. L. (2002). Timing of ornaments in the theme of Beethoven's Paisiello Variations: Empirical data and a model. *Music Perception, 20,* 3–33.

Todd, N. P. (1985). A model of expressive timing in music. *Music Perception, 3*, 33–58.

Todd, N. P. (1992). The dynamics of dynamics: A model of musical expression. *Journal of the Acoustical Society of America, 91*, 3540–3550.

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

Windsor, W. L., Desain, P., Aarts., Heijink, H., & Timmers, R. (2001). The timing of grace notes in skilled musical performance at different tempi: A case study. *Psychology of Music, 29*, 149–169.

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