

Bruno Gingras, Caroline Palmer, Peter N. Schubert, and Stephen McAdams, Influence of melodic emphasis, texture, salience, and performer individuality on performance errors, *Psychology of Music* (44(4), pp.847-863). Copyright © 2016, © SAGE Publications. DOI: 10.1177/0305735615594491

Influence of melodic emphasis, texture, salience, and performer individuality on performance errors

Running head: Factors influencing performance errors

Bruno Gingras¹, Caroline Palmer², Peter N. Schubert³, and Stephen McAdams³

¹ Institute of Psychology, University of Innsbruck, Austria

² Department of Psychology, McGill University, Canada

³ Schulich School of Music, McGill University, Canada

Research Article

Corresponding author:

Bruno Gingras

Institute of Psychology, University of Innsbruck, Innrain 52f, A-6020 Innsbruck, Austria

Email: brunogingras@gmail.com

Telephone: +43-0-512-507-56052

Abstract

We investigated the influence of melodic emphasis, musical texture, musical salience, and performer individuality on the distribution and frequency of errors in keyboard performance. Eight performers recorded different interpretations of two short Baroque organ pieces of contrasting texture (homophonic versus polyphonic style). Melodic emphasis affected the distribution of performance errors according to the location of the part intended as melody, whereas musical texture influenced the type of errors. The effect of musical salience was examined by inviting sixteen performers to record a Bach organ fugue. Error rates were lower for notes belonging to recurring musical motives than for non-motivic passages, and for outer voices compared to inner voices. These results are consistent with error detection studies showing that errors in inner voices or unfamiliar melodies are less perceptually salient, suggesting that the error likelihood is inversely related to a note's degree of perceptual and musical salience. Across all three pieces, error patterns were more consistent in within-performer comparisons than between-performer comparisons of recordings of the same piece, implying that error patterns are indicative of individual differences in the interpretation of musical structures.

Keywords: performance errors, melodic emphasis, texture, salience, individuality, organ

Introduction

Music performance is one of the most challenging time-based activities in which humans routinely engage, involving complex motor coordination (Moore, 1992; Wilson, 1992), synchronization and coordination of musical gestures in a temporal context (Pfordresher, Palmer, & Jungers, 2007; Repp, 1999), memorization of complex sequences of events (Palmer, 2005), and in the case of score-based music, sight-reading or memorization of a score (see Parncutt & McPherson, 2002 for a survey of these issues). Not surprisingly, even skilled performances contain various types of performance errors (Repp, 1996a). These errors may be ascribed to several causes: among the most commonly mentioned are the technical requirements of the piece, score reading or memorization issues, a lack of concentration or preparation, or a stress-induced performance degradation (Palmer & Van de Sande, 1993, 1995; Repp, 1996a; Wan & Huon, 2005).

For the last several decades, speech production errors have been studied as a way to understand the mechanisms involved in sentence production (Dell, 1985; Garrett, 1975). The analysis of typing errors has also afforded a useful experimental paradigm to model human performance in activities involving fine motor coordination in the production of sequentially ordered events (Rumelhart & Norman, 1982; Shaffer, 1976). Similarly, the study of performance errors may lead to a better comprehension of the cognitive processes involved in music performance. More specifically, the distribution and relative frequency of errors may provide clues about a performer's mental representation of the musical structure of a piece, while revealing relationships between intention and performance (Palmer & Van de Sande, 1993, 1995; Repp, 1996a; Shaffer, 1976).

Most studies of errors in music performance have been conducted either with piano music from the Romantic and Classical eras (Repp, 1996a) or with short stimuli newly composed or adapted specifically for experimental purposes (Palmer & Van de Sande, 1993, 1995). In contrast, the present study focused on the Baroque organ repertoire, with the aim of assessing whether previous findings could be extended to another keyboard instrument and to a different repertoire. We also addressed related issues that had previously received little attention, such as examining whether error rates are lower for musically salient elements (recurring motives and themes) and evaluating the consistency and individuality of performers' error patterns. Because we employed recordings of three entire pieces that were also used to study expressive strategies (Gingras, Schubert, & McAdams, 2010; Gingras, Lagrandeur-Ponce, Giordano, & McAdams, 2011), error production was analyzed in an ecological context, thus complementing earlier studies in which performance errors were elicited (Palmer & Van de Sande, 1993, 1995).

Several aspects of musical structure have been shown to influence error patterns. For instance, in multivoiced music, errors occur more frequently in inner voices than in outer voices (Palmer & Van de Sande, 1993; Repp, 1996a).¹ Furthermore, musical texture (homophonic versus polyphonic music) has been found to affect the type of errors (Palmer & Van de Sande, 1993), with more harmonically related errors occurring in homophonic pieces, in which across-voice associations are emphasized, than in polyphonic pieces, which favor within-voice associations. Interestingly, in error-detection tasks, sensitivity to errors was lower for errors in inner voices and for harmonically related errors; in addition, sensitivity to harmonically related errors was greater in polyphonic than in homophonic textures (Palmer & Holleran, 1994).

¹ Following Palmer & Holleran (1994), we use the term “multivoiced” music to refer to music composed for several parts or voices; the terms “homophonic” and “polyphonic” are reserved for specific musical textures.

Finally, the number of notated onsets to be performed simultaneously (or *onset density*) has been shown to influence error rates, with higher rates per note for score events with a higher onset density (Repp, 1996a). These findings indicate that both the production and perception of performance errors are influenced by structural and textural considerations, suggesting that both performers' and listeners' conceptual representations of the music are shaped by the musical texture.

Regarding the effect of melodic emphasis on error distribution, Palmer & Van de Sande (1993) reported that errors were less likely to occur in the voice intended as melody than in nonmelody voices, and that the error pattern varied according to the performer's interpretative goal. In addition, errors were found to be less frequent in the highest voice regardless of the interpretative goal. Again, this relationship is mirrored in perception studies reporting that listeners are generally more sensitive to changes in the highest voice (Dewitt & Samuel, 1990; Palmer & Holleran, 1994), an effect that has been documented at a pre-attentive level in electrophysiological studies (Fujioka, Trainor, Ross, Kakigi, & Pantev, 2005). However, the lower error rates in the melody voice observed in Palmer & Van de Sande, Experiment 1 (1993) could be explained by the fact that they examined three-voice textures in which the hand controlling the melody played only one voice whereas the alternate hand controlled two voices. The authors addressed this possible confound in Experiment 2, which replicated the melody advantage whether or not the hand controlling the melody also controlled an additional voice. Here, we circumvented this issue by using a four-part piece containing two voices in each hand, thus keeping the assignment of voices to hand fixed across conditions.

One aspect that has been unexplored so far is whether error rates are lower for musically salient elements such as recurring musical motives or themes. Given the effects of musical

structure and melodic emphasis on error rates (Palmer & Van de Sande, 1993), performers could be expected to make fewer errors when playing motivic passages than non-motivic passages; likewise, listeners should be more versions of familiar than unfamiliar melodies. Here, we tested the effect of musical salience (motivic vs. non-motivic notes) by analyzing performances of an organ fugue in which recurring motives are clearly delineated.

Finally, we sought to determine whether individual performers exhibited consistent, idiosyncratic error patterns in repeated performances of the same piece. High-level pianists have been shown to be extremely consistent regarding patterns of timing, articulation, and dynamics (Palmer, 1989; Repp, 1992, 1996b, 1996c; Widmer & Goebel, 2004). Although Repp (1996a) reported on performers' consistency with respect to error production, an exhaustive statistical analysis was not included. This hypothesis was amenable to a more rigorous testing in the present study, because our database included repeated performances of all three pieces.

PERFORMANCE ERRORS: DIFFERENT LEVELS OF OBSERVATION

Performance errors may be observed at several levels (Repp, 1996a). For our purposes, the following stages may be differentiated: the input level (reading of the score by the performer), the memory retrieval level (the performer's recall of the items to be produced), the kinematic level (motion of the performer's hands and fingers), and the mechanical level (the generation of the sound by the instrument). In this study, we analyzed errors registered in terms of keypress events recorded in MIDI (Musical Instrument Digital Interface) format. One advantage of this approach is that errors can be defined objectively and unambiguously (Repp, 1996a). On the other hand, the perceptibility of errors is not taken directly into account at this level. However, it

seems plausible to posit a link between the distribution of performance errors and their perceptibility, as evidenced by earlier results (Palmer & Holleran, 1994).

CODING OF PERFORMANCE ERRORS

Although various definitions and categorizations of performance errors have been proposed, one commonality is that errors are broadly understood as deviations from the written score (Large, 1993; Palmer & Van de Sande, 1993; Repp, 1996a). However, not all of these deviations should be defined as errors, because the performer enjoys a certain degree of artistic license (Palmer & Van de Sande, 1993). For this reason, most studies have focused on errors that can be clearly identified on a categorical basis relative to the score (Repp, 1996a), such as pitch substitutions (playing a note with the wrong pitch), deletions (failure to play a note that is in the score), and intrusions or additions (adding notes that are not in the score). To these categories, we also added timing errors, which were defined as notes whose onsets were located more than 150 milliseconds before or after their expected onset time.² The expected onset time corresponds to the average onset time of other nominally synchronous note onsets as indicated on the score, or to the onset time predicted by the local tempo in the case of unaccompanied note onsets.

In the context of this study, we distinguished between *score-based errors*, which comprised pitch substitutions, deletions (including “added ties” – repeated notes in the score that were not re-attacked in performance), and timing errors, on the one hand, and *non-score-based errors*, which include all performance notes that were added to the information in the score, such

² Expressive onset asynchronies in keyboard performance, even exaggerated ones, are typically smaller than 100 ms (Goebel, 2001; Repp, 1996c). We confirmed through listening that errors identified as timing errors indeed sounded like performance errors.

as intrusions and repetitions (re-attacked notes in performance that were not repeated in the score), on the other hand.³ This distinction is important because score-based errors can be assigned to a specific note, allowing a characterization by voice, position, and limb assignment, whereas non-score-based errors cannot always be unambiguously assigned to a context. The bulk of this article focuses on errors linked to specific score notes, and on the contextual effects that can be observed from the distributions of these errors.

Errors were coded in a parsimonious manner; that is, in cases where an error could be analyzed as one error or as two distinct errors, the coding that minimized the number of errors was chosen (**Palmer & Van de Sande, 1993**). Furthermore, we used an error detection mechanism that was completely objective and computer-monitored, thereby ensuring that the criteria for error detection were explicit and identical across performances.

Method

MUSICAL MATERIALS

Three organ pieces were used for this study: *Premier Agnus*, a polyphonic piece by Nicolas de Grigny (1672-1703), *Wachet auf, ruft uns die Stimme* (SSWV 534), a homophonic chorale setting by Samuel Scheidt (1587-1654), and the fugue in D minor (BWV 538), better known as the “Dorian” fugue, by Johann Sebastian Bach (1685-1750). The scores of the *Premier Agnus* and of *Wachet auf*, as well as the first few measures of the Dorian fugue, are included in the Appendix.

³ “Untied” notes (Repp, 1996a) are treated as repetitions.

PARTICIPANTS

All participants were professional organists from the Montreal area. All participants signed a consent form and received financial compensation for their participation in the study, which was approved and reviewed by the Research Ethics Board of McGill University (Montreal, Canada).

Eight organists (two female, six male; aged 23-30 years) performed the *Premier Agnus* piece. All participants identified themselves as right-handers. They had received organ instruction for a mean duration of 10 years (range = 7-13 years) and held or had held a position as church organist for an average of 8 years (range = 1-21 years).

Eight organists (two female, six male; aged 19-30 years) performed the *Wachet auf* piece, including seven organists who recorded the *Premier Agnus*. All participants identified themselves as right-handers. They had received organ instruction for a mean duration of 9 years (range = 3-13 years) and held or had held a position as church organist for an average of 8 years (range = 1-21 years).

Sixteen organists (two female, fourteen male; aged 24-59 years) recorded the Dorian fugue. Three of them also performed the *Premier Agnus* and *Wachet auf*. Fourteen identified themselves as right-handers, one as left-hander, and one as ambidextrous. They had received organ instruction for a mean duration of 10 years (range = 4-25 years) and held or had held a position as church organist for an average of 18 years (range = 4-39 years).

PROCEDURE

Notated scores for the *Premier Agnus* and the *Wachet auf*, which are short, non-technically demanding pieces, were given to the organists 20 minutes before the recording session began, to give them time to practice. None of the performers were familiar with these pieces. For each piece, organists were asked to record different interpretations. Two recordings were made for each interpretation to allow for a measure of consistency. Both pieces were played only on the keyboard (manuals, played with the hands), and the pedalboard was not used.

For the *Premier Agnus*, three different interpretations were recorded. In one interpretation, organists were asked to emphasize the soprano part, in another, the alto part, and in a third one, the tenor part. Two recordings were made for each interpretation. The order of the instructions was randomized according to a Latin square design. For *Wachet auf*, two different interpretations were recorded. Performers were asked to record two expressive renditions of the piece, followed by two inexpressive renditions, for which they were instructed to play without adding any expressiveness beyond what was notated in the score and as mechanically as possible (Palmer, 1989).

A third piece, the Dorian fugue, was communicated to performers several weeks in advance. Most organists were familiar with this piece. Organists were given 20 minutes to practice, after which they made two recordings of the piece. No directives were given regarding the interpretation. Use of the pedals and manuals was necessary for this piece.

All performances were recorded on the Casavant organ of the Church of St-Andrew & St-Paul in Montreal, Canada. This five-manual organ (5 keyboards and a pedalboard) was built in 1931, and the console was restored in 2000, at which time a MIDI system was installed by Solid

State Organ Systems. The scanning rate of the MIDI system was estimated at 750 Hz (1.33 ms), the on and off points being determined by key-bottom contact.⁴ All performers used the same registration for each piece. The audio signal was recorded through two Boehringer ECM 8000 omnidirectional microphones. The audio and MIDI signals were sent to a PC computer through an RME Fireface audio interface (Audio AG, Haimhausen, Germany). Audio and MIDI data were then recorded using Cakewalk's SONAR software (Cakewalk, Inc., Boston, MA, USA) and stored on a hard disk.

DATA ANALYSIS

Performance notes obtained from the MIDI data were matched to score notes, using a note-by-note score-performance matching algorithm developed by Gingras & McAdams (2011). The error analysis was part of the matching process and thus completely automated. The automated analysis provided by the matching algorithm has been shown to correspond very closely to human analysis, with discrepancies observed on fewer than 0.1% of the notes in a large corpus of performances of the *Premier Agnus* and of *Wachet auf* analyzed both by hand and using the algorithm (see Table 3 in Gingras & McAdams, 2011). Hand and voice assignments of score notes were determined by the first author, a music theorist and church organist.

⁴ Information provided by Mark Gilliam, Sales manager of Solid State Organ Systems.

Results

OVERVIEW AND DESCRIPTIVE STATISTICS

Error frequencies and percentages. The frequencies and percentages of the different error types are summarized in Table 1, which also lists the total number of score notes and performed notes for each piece. Global score error rates were highest in *Wachet auf* and lowest for the Dorian fugue. The differences in score-based error rates between the three pieces appear to be due mostly to discrepancies in the rate of added ties, corresponding to notes that were repeated in the score but not re-attacked in performances. The proportion of repeated notes was much higher for *Wachet auf* (55 repeated notes out of 369 total score notes, or 14.9%) than in the *Premier Agnus* (21 of 320 notes, or 6.6%) or the Dorian fugue (58 of 2,701 notes, or 2.1%), which probably explains the higher rates of added ties in the former. One explanation for the relatively high incidence of added ties in organ performance is that note onsets on the organ are not as salient as on the piano, because the organ sound is continuous, and performers are perhaps less mindful of re-striking repeated notes. We also observed that the frequency of non-score-based errors was higher for *Wachet auf*, a piece characterized by a generally homophonic texture, than for the *Premier Agnus*, whose texture is much more polyphonic, although the level of difficulty appeared to be comparable for both pieces (full scores are provided in the Appendix). Error rates were generally comparable to those reported by Repp (1996a); omission rates were lower in the present study, but the omissions reported by Repp probably included added ties as well (this category was not explicitly defined in Repp's study).

[Insert Table 1 here]

Order of recording. Because performers made several recordings of each piece, the order in which the repeated recordings occurred could be a confounding factor when comparing error rates across interpretations. To examine this effect, repeated-measures analyses of variance were conducted on the total error frequency (combined score and non-score-based errors) by performance for each piece, with order of recording as a within-subject factor. The results showed no significant effect of order of recording (*Premier Agnus*, $F(5, 35) = 1.31, p = .30$, Greenhouse-Geisser $\epsilon = 0.42$; *Wachet auf*, $F(3, 21) = 1.49, p = .25, \epsilon = 0.82$; Dorian fugue, $F(1, 15) = 0.78, p = .39$), indicating that error rates were not demonstrably influenced by the order of recording. Order of recording will thus not be considered in subsequent analyses.

Onset density. For all three pieces, the mean error frequency (combining score and non-score-based errors) per score event, normalized for the number of notated onsets per score event, was weakly but positively correlated with onset density, with coefficients of 0.17 ($df = 146, p = .045$), 0.24 ($df = 143, p = .004$), and 0.08 ($df = 1382, p = .003$) for the *Premier Agnus*, *Wachet auf*, and the Dorian fugue, respectively.

EFFECTS OF MELODIC EMPHASIS, NOTE POSITION, AND MUSICAL SALIENCE

Only score-based errors were used for the analysis of effects of melodic emphasis, note position and musical salience on error rates, because they could be unambiguously assigned to a specific note in the score and therefore to a specific voice or limb, unlike most non-score-based errors. The effects of note position analyzed in this article include voice assignment (soprano, alto, tenor, or bass) and position (outer versus inner voices), as well as limb assignment (foot

pedals, left hand, or right hand). Additionally, in the case of the Dorian fugue, we examined the effect of musical salience (notes belonging to recurring motives versus notes that do not). Separate analyses will be presented for all three pieces.⁵

In all cases, error rates were modeled using a logistic regression model, in which the dependent variable represented the probability of error per score note, in other words, the probability that a given note is wrongly played by the performer, rather than a continuous error proportion or rate across the entire melodic sequence. Because onset density was shown to influence error rate, it was included as a covariate in order to take its effect into account.⁶

Melodic emphasis. Organists recorded three different interpretations of the *Premier Agnus*, emphasizing a different voice (either the soprano, alto, or tenor part) in each interpretation. We hypothesized that error rates would be lower for the emphasized voice than for the non-emphasized ones on the basis of earlier studies (Palmer & Van de Sande, 1993). A repeated-measures logistic regression on error rate per voice, with interpretation as a fixed factor and onset density as a covariate, showed no main effect of voice or interpretation, but a significant interaction between them, $\chi^2(6) = 85.3, p < .001$. This result indicates that although global error rates did not vary significantly between voices or interpretations, organists made fewer errors for the notes belonging to the voice that they were trying to emphasize (Figure 1a). A similar interpretation could be made for the logistic regression analysis on error rate per hand,

5 Although all three pieces were nominally four-voice pieces, the last chord of the *Premier Agnus* and a few short passages in the Dorian fugue contain additional voices. These voices, which comprise a very small fraction of the total number of score notes, were not included in the analyses by voice subsequently presented.

6 Onset density was indeed a significant covariate in all the logistic regression models presented subsequently.

which showed no main effect of hand or interpretation, but a significant interaction between these factors, $\chi^2(2) = 34.5, p < .001$. On the other hand, the logistic regression on error rate per voice position showed a main effect of voice position, $\chi^2(1) = 4.23, p = .040$, and a significant interaction between voice position and interpretation, $\chi^2(2) = 13.6, p = .001$, indicating that whereas error rates were generally lower for outer voices, this effect was modulated by the interpretation (Figure 1b). Except for the fact that we did not observe a lower error rate for the highest voice across all conditions, these results are very similar to those reported by Palmer & Van de Sande (1993).

a)

[Insert Figure 1a here]

b)

[Insert Figure 1b here]

To investigate whether these effects were specifically related to melodic emphasis but not to other types of expressive manipulations, performers recorded two different interpretations of *Wachet auf*: an expressive interpretation of the piece, followed by an inexpressive one. Because these instructions did not imply specific contrasts in melodic emphasis, it was hypothesized that the distribution of errors would not be significantly affected by the type of interpretation. A typical error distribution pattern, with lower rates in the highest voice and in outer voices, was

thus expected (Palmer & Van de Sande, 1993). A repeated-measures logistic regression on error rate per voice, with interpretation as a fixed factor and onset density as a covariate, showed a main effect of voice, $\chi^2(3) = 41.2, p < .001$, no effect of interpretation, and no significant interaction (Figure 2a). The same statistical analysis was conducted on error rates by hand, showing no main effect or interaction, and by voice position, showing a significant effect of voice position, $\chi^2(1) = 20.1, p < .001$, and no other effect (Figure 2b). Overall, error rates were lower for the soprano voice (which contained the melody of this chorale setting) and for the outer voices (soprano and bass), thus essentially replicating earlier findings by Palmer & Van de Sande (1993), but the interpretation did not significantly affect the distribution of error rates.

a)

[Insert Figure 2a here]

b)

[Insert Figure 2b here]

Effect of musical salience. The effect of musical salience was examined using the recordings of the Dorian fugue. Because no directives were given regarding the interpretation, a typical error distribution pattern, with lower rates in the highest voice and in outer voices, was again expected (Palmer & Van de Sande, 1993). The Dorian fugue is a complex piece, both in terms of length and motivic richness, and it requires performers to use the pedals in addition to

the keyboard. Moreover, because this piece is a fugue, motivic material is distributed among all the voices. The potential interplay between voice position, limb assignment, and musical salience on error rates was thus subjected to a detailed analysis. Five main motives were considered: the fugue subject, the first and second countersubjects, and two short recurring motives derived from the first countersubject, which saturate the fugue (see Appendix; brackets indicate motivic material). All other sections of the piece were considered non-motivic for the purpose of this analysis.

[Insert Figure 3 here]

As expected, error rates were lower for the highest voice (soprano) and for outer voices (Figure 3). Error rates were also significantly lower for motivic notes than for non-motivic ones. Finally, error rates were higher for the left hand than for the right hand or the pedal. Given that the majority of motives occur in outer voices in the Dorian fugue, presumably because the composer sought to ensure their perceptual salience (Huron, 1989; Huron & Fantini, 1989) and that all pedal notes belong to an outer voice in this piece,⁷ the effects of voice position (and, by extension, those related to specific voices), musical salience, and limb assignment are interdependent to a certain extent. Thus, a rigorous statistical treatment of these effects should consider the combined effects of voice position and musical salience, while excluding the pedal part from analyses considering interactions between voice position and limb assignment (note that multicollinearity cannot be assessed between categorical factors such as voice assignment,

⁷ Notes in the pedal part sound one octave lower than written on the score, because they are played on 16' stops.

limb assignment, or voice position). A repeated-measures logistic regression on error rate by voice position (inner/outer) and musical salience (motivic/non-motivic notes), excluding the pedal part showed significant effects of voice position, $\chi^2(1) = 75.3, p < .001$, and musical salience $\chi^2(1) = 11.7, p < .001$.

An analysis combining the effects of voice position, musical salience, and hand assignment (left/right, excluding pedal notes) in a single model yielded a more complex picture, with main effects of voice position and musical salience (but no effect of hand assignment) and significant interactions between hand assignment and position, as well as hand assignment and musical salience (Table 2). Whereas error rates for motivic notes in outer voices were comparable for both hands, they were markedly higher in the left hand for non-motivic notes belonging to inner voices (Figure 4).

[Insert Table 2 here]

[Insert Figure 4 here]

EFFECTS OF MUSICAL TEXTURE: HOMOPHONIC VERSUS POLYPHONIC PIECE

Palmer and Van de Sande (1993) had previously shown that the proportion of harmonically related errors was higher for homophonic pieces than for polyphonic pieces. Here, we analyzed the effect of musical texture on two error types, namely pitch substitutions (replacing a score note by a note with the wrong pitch) and intrusions (playing additional notes

not indicated in the score), by evaluating the type of errors produced in performances of a mostly homophonic piece (*Wachet auf*) and of a polyphonic piece (*Premier Agnus*). These two pieces are of similar length, with a mostly four-voice texture throughout (the average number of active voices per score event, or *voice density* was 3.98 for both pieces), thus providing an adequate basis for comparison.

Pitch and intrusion errors were categorized into three types: errors related only to the harmonic context, errors related only to the melodic context, and errors that were both harmonically and melodically related. An error was defined as harmonically related if its pitch was equivalent, via octave transposition, to that of another score note present in the same score event. An error was defined as melodically related if another note with the exact same pitch was found in the score events immediately preceding or following the onset of the wrong note. Following Palmer & Van de Sande (1993), chance estimates were computed for harmonic relatedness, corresponding to the average number of pitch classes per score event divided by the total number of possible pitch classes (12); equal probability was assumed for all pitch classes. Statistical analyses were conducted both on the aggregate data (chi-square test) and on individual performers (two-tailed Wilcoxon paired-sample exact tests) to test for differences between proportions and chance estimates.

Table 3 shows that the proportion of melodically related errors was greater in the polyphonic piece (*Premier Agnus*) than in the homophonic piece (*Wachet auf*), whereas the proportion of harmonically related errors followed an inverse trend. A chi-square test on the aggregate data showed a significant effect of texture on the relative proportions of error types, $\chi^2(3) = 8.49, p = .037$. Analyses by performer reveal that the proportion of harmonically related errors (including errors that were both harmonically and melodically related) differed

significantly from the chance estimate for the *Premier Agnus* ($T = 1, p = .031$), but not for *Wachet auf* ($T = 16, p = .85$). These results indicate that the proportion of harmonically related errors was greater in a homophonic texture (*Wachet auf*) than in a polyphonic texture (*Premier Agnus*), and that performers made fewer harmonically related errors than expected by chance in a polyphonic texture. From these observations, which reproduce those of Palmer & Van de Sande (1993), it may be inferred that performers were more sensitive to vertical, within-chord associations in the homophonic texture, while paying more attention to horizontal, within-voice associations in the polyphonic texture.

[Insert Table 3 here]

CONSISTENCY AND INDIVIDUALITY OF ERROR PATTERNS

To evaluate the consistency and individuality of performers' error patterns, all pairs of performances were compared by tabulating the frequency of co-occurrence of errors in the same score event in different performances. Because this comparison was conducted on an event-by-event basis instead of a note-by-note basis, both score and non-score-based errors were included in this analysis.⁸ Jaccard coefficients were computed as a measure of the degree of concordance between the error patterns of each pair of performances (Legendre & Legendre, 1998). The Jaccard coefficient, whose range is bounded between 0 and 1, represents the proportion of events for which both performances contained an error divided by the total number of events for which

⁸ Four performances (out of 48) of the *Premier Agnus* did not contain a single error and were therefore omitted from this analysis.

at least one performance contained an error. It is particularly well suited to measure the concordance between error patterns because it does not take into account events for which neither performance contained an error. For all three pieces, Jaccard coefficients, although generally modest, were significantly higher for comparisons between pairs of performances played by the same performer than between performances played by different performers (Table 4). These analyses indicate that the event-by-event error patterns of two performances of the same piece by the same performer were generally more similar than those of two performances by different performers. The significantly lower coefficients observed for between-performers comparisons also suggest that performer idiosyncrasies may be better predictors of the location of performance errors than difficulties associated with particular events in the score, at least in the case of these three pieces.

[Insert Table 4 here]

Discussion

Our analyses of error patterns in recordings of Baroque organ pieces indicated that performers' error patterns are modulated to a large extent by the local musical context, such as the position or musical salience of a note or group of notes, as well as the global musical texture, such as the degree of polyphony of a piece. For the most part, these results are congruent with earlier findings: performers tend to make fewer errors in the highest voice, as well as in the outer voices of a multivoiced piece, and they make more harmonically related errors in a homophonic texture than in a polyphonic one (Palmer & Van de Sande, 1993). In addition, we have shown

that error rates were lower for motivic notes than for non-motivic ones. Furthermore, error patterns were more consistent in within-performer than across-performer comparisons, indicating that individualistic interpretations of musical structure (rather than simply information in the notated composition) drive performance and impact the production of errors. We discuss each of these factors in turn.

Whereas Palmer & Van de Sande (1993) reported lower error rates in the highest voice regardless of the position of the main melody, suggesting an articulatory advantage for outer right-hand fingers, we did not observe lower error rates for the highest voice or for the right hand in all conditions. In the case of the *Premier Agnus*, error rates by voice and hand varied according to the position of the emphasized voice, and no main effect of voice or hand emerged. For *Wachet auf*, although error rates were lower for the highest voice, which contains the chorale melody and is played by the right hand, global error rates did not differ significantly between hands. In the case of the Dorian fugue, error rates were lower for the highest voice, as well as globally higher for the left hand than for the right hand. However, a more refined analysis revealed that error rates in both hands were comparable for perceptually and/or musically salient notes (such as notes belonging to an outer voice or to a recurring motive), but were noticeably higher in the left hand for less salient notes.

This discrepancy between our findings and those of earlier studies regarding hand and voice assignment effects could be explained by differences in the musical stimuli (number of voices controlled by each hand), the skill level of the performers, or the experimental instructions: this study used natural performance tempi, whereas Palmer & Van de Sande (1993) elicited errors by asking performers to use faster tempi. The differential effects of voice position and musical salience by hand assignment observed for the Dorian fugue suggest that the right-

hand advantage can probably best be explained by a combination of hand-dominance effects and attentional processes. In a series of articles, Peters (1981, 1985) reported that right-handers typically performed bimanual tasks better when the right hand took the “figure” and the left hand took the “ground” of a dual movement, and that subjects’ performance could be influenced by directing their attentional processes. If we assume that performers directed more attentional resources towards perceptually or musically salient notes, this model would fit nicely with our observations on the Dorian fugue. Indeed, there was no clear right-hand advantage in terms of error rates for salient notes, whereas the left hand was at a clear disadvantage for less salient notes. However, a thorough study of the effects of hand assignment and handedness on error rate would entail a comparison of the performances of left-handed and right-handed keyboardists of equivalent skill level and equivalent familiarity with the musical pieces; this project was beyond the scope of the present study.⁹

As mentioned above, organists made fewer errors in the outer voices, and they made more harmonically related errors in a homophonic texture than in a polyphonic one. These findings, which replicate previous findings (Palmer & Van de Sande, 1993), are consistent with measures of listeners’ sensitivity to altered pitches in performance, which is higher for errors in the outer voices and especially in the highest voice, and for harmonically unrelated pitch errors than for related ones (Palmer & Holleran, 1994). Moreover, our finding that error rates were lower for motivic notes than for non-motivic ones is in line with Dewitt and Samuel’s (1990) observation that listeners are more proficient at detecting changes in familiar than in unfamiliar melodies. These complementary observations regarding the production and detection of performance errors

⁹ Note that the performances of the 14 right-handed organists were grouped together with those of one ambidextrous and one left-handed organist for the analyses of the Dorian fugue. Palmer & Van de Sande (1993) did not report on the handedness of their participants.

suggest that performers' and listeners' mental representations of the score are well-matched in terms of the relative perceptual and musical salience of structural note categories. These relationships may be encapsulated by the following statement: the likelihood of a note, or group of notes, being wrongly played is inversely correlated with its degree of perceptual and musical salience.

Our experimental manipulation of melodic emphasis showed that organists made fewer errors in a given voice when it was emphasized than when it was not. This indicates that their mental representations of a musical score are flexible, and suggests that interpretations of the same piece that highlight different musical features lead to distinct conceptualizations of the performance in terms of the relative salience of musical elements, as reflected by characteristic error patterns. On the other hand, interpretations of the same piece that differed only in their level of expressivity had no significant effect on the distribution of errors, implying that only interpretative goals that specifically attempt to manipulate the relative salience of musical elements affect error patterns.

Another aspect of the complementarity between production and perception may be found in the interaction between hand assignment and perceptual salience. As reported earlier, listeners are more sensitive to errors in the highest voice, normally played by the right hand, and performers' error rates for this voice are usually lower than for other voices. Furthermore, a large proportion of the Western musical repertoire ascribes greater importance to the highest voice, which often contains prominent melodic material, whereas other voices take an accompanimental role (Palmer & Van de Sande, 1993). The relationships identified between hand assignment, relative salience, and error rates in the Dorian fugue further point to a clear right-hand advantage, at least for right-handed performers. Whether by design or by accident, the frequency mapping of

the keyboard takes into account both cognitive-motor and perceptual constraints of right-handed musicians by locating the higher, more salient frequencies on the right side; indeed, whereas naïve left-handers have been found to prefer keyboards with an inverted frequency mapping, right-handers prefer the normal configuration regardless of their musical experience (Laeng & Park, 1999).

Although performance errors were influenced in large part by the musical structure, we have shown that they are also, to some extent, performer-specific. Error patterns from different performances of the same piece by the same organist were more similar than those of recordings by different organists. Although performance errors are not normally considered as part of the expression of a musician's individuality, these findings suggest that error patterns, like timing, articulation, or intensity change patterns, are shaped by a performer's unique conception of a score and of its musical realization. In fact, the analogies with timing patterns can be pursued further. Both the production and perception of temporal patterns are affected by structural considerations (Repp, 1998). Furthermore, temporal patterns are considered one of the hallmarks of a performer's artistic individuality (Repp, 1992). As we have demonstrated, similar relationships hold true for errors regarding the influence of musical structure, the complementarity between the production and perception of errors, and the individuality of error patterns.

In sum, error patterns in music performance are shaped by a rich nexus of relationships between musical structure, cognitive-motor determinants of performance, perceptual and psychoacoustic constraints, and considerations linked to performers' expressive goals. Although performance errors may be viewed as unwelcome by-products of music production activities,

their study is as relevant to the understanding of the cognitive processes involved in music performance as that of more celebrated aspects of musical artistry.

Ethics

This work was approved by the Research Ethics Board of McGill University, Montreal, Canada (REB Certificate 67-1104).

Acknowledgements

Rhonda Amsel provided helpful suggestions regarding the statistical analyses. We thank Julien Boissinot, Darryl Cameron, and Bennett K. Smith for their technical assistance, Peter Holmes for permission to use McGill University's sound recording equipment, Jonas Braasch and Nils Peters for their advice regarding sound recording, the musical authorities of the Church of St-Andrew & St-Paul (Montreal) for permission to use their Casavant organ, and the organists whose performances were recorded for this project.

Funding

This work was supported by a doctoral fellowship from the Social Sciences and Humanities Research Council of Canada (SSHRC 752-2003-2178) and a student award from the Centre for Interdisciplinary Research in Music Media and Technology (CIRMMT) to Bruno Gingras, as well as grants from the Natural Sciences and Engineering Research Council (RGPIN 298173-2010; RGPIN 312774-2010) and Canada Research Chairs awarded to Caroline Palmer and to Stephen McAdams.

References

- Dewitt, L. A., & Samuel, A. G. (1990). The role of knowledge-based expectations in music perception - evidence from musical restoration. *Journal of Experimental Psychology-General*, 119(2), 123-144.
- Fujioka, T., Trainor, L. J., Ross, B., Kakigi, R., & Pantev, C. (2005). Automatic encoding of polyphonic melodies in musicians and nonmusicians. *Journal of Cognitive Neuroscience*, 17(10), 1578-1592.
- Goebel, W. (2001). Melody lead in piano performance: Expressive device or artifact? *Journal of the Acoustical Society of America*, 110(1), 563-572.
- Huron, D. (1989). Voice denumerability in polyphonic music of homogeneous timbres. *Music Perception*, 6(4), 361-382.
- Huron, D., & Fantini, D. A. (1989). The avoidance of inner-voice entries - perceptual evidence and musical practice. *Music Perception*, 7(1), 43-47.
- Laeng, B., & Park, A. (1999). Handedness effects on playing a normal or reversed keyboard. *Laterality*, 4(4), 363-377.
- Large, E. W. (1993). Dynamic programming for the analysis of serial behaviors. *Behavior Research Methods Instruments & Computers*, 25(2), 238-241.
- Moore, G. P. (1992). Piano trills. *Music Perception*, 9(3), 351-360.
- Palmer, C. (1989). Mapping musical thought to musical performance. *Journal of Experimental Psychology-Human Perception and Performance*, 15(12), 331-346.
- Palmer, C. (2005). Sequence memory in music performance. *Current Directions in Psychological Science*, 14(5), 247-250.
- Palmer, C., & Holleran, S. (1994). Harmonic, melodic, and frequency height influences in the perception of multivoiced music. *Perception & Psychophysics*, 56(3), 301-312.
- Palmer, C., & Van de Sande, C. (1993). Units of knowledge in music performance. *Journal of Experimental Psychology-Learning Memory and Cognition*, 19(2), 457-470.

- Palmer, C., & Van de Sande, C. (1995). Range of planning in music performance. *Journal of Experimental Psychology-Human Perception and Performance*, 21(5), 947-962.
- Parncutt, R., & McPherson, G. (2002). *The science & psychology of music performance : creative strategies for teaching and learning*. Oxford ; New York: Oxford University Press.
- Peters, M. (1981). Handedness - coordination of within-hand and between-hand alternating movements. *American Journal of Psychology*, 94(4), 633-643.
- Peters, M. (1985). Constraints in the coordination of bimanual movements and their expression in skilled and unskilled subjects. . *Quarterly Journal of Experimental Psychology Section a-Human Experimental Psychology*, 37(2), 171-196.
- Pfordresher, P. Q., Palmer, C., & Jungers, M. K. (2007). Speed, accuracy, and serial order in sequence production. *Cognitive Science*, 31(1), 63-98.
- Repp, B. H. (1992). Diversity and commonality in music performance - an analysis of timing microstructure in Schumann's "Träumerei". *Journal of the Acoustical Society of America*, 92(5), 2546-2568.
- Repp, B. H. (1996a). The art of inaccuracy: Why pianists' errors are difficult to hear. *Music Perception*, 14(2), 161-183.
- Repp, B. H. (1996b). The dynamics of expressive piano performance: Schumann's "Traumerei" revisited. *Journal of the Acoustical Society of America*, 100(1), 641-650.
- Repp, B. H. (1996c). Patterns of note onset asynchronies in expressive piano performance. *Journal of the Acoustical Society of America*, 100(6), 3917-3931.
- Repp, B. H. (1998). Variations on a theme by Chopin: Relations between perception and production of timing in music. *Journal of Experimental Psychology-Human Perception and Performance*, 24(3), 791-811.
- Repp, B. H. (1999). Control of expressive and metronomic timing in pianists. *Journal of Motor Behavior*, 31(2), 145-164.
- Rumelhart, D. E., & Norman, D. A. (1982). Simulating a skilled typist - a study of skilled cognitive-motor performance. *Cognitive Science*, 6(1), 1-36.
- Shaffer, L. H. (1976). Intention and performance. *Psychological Review*, 83(5), 375-393.

- Wan, C. Y., & Huon, G. F. (2005). Performance degradation under pressure in music: An examination of attentional processes. *Psychology of Music*, 33(2), 155.
- Widmer, G., & Goebel, W. (2004). Computational models of expressive music performance: The state of the art. *Journal of New Music Research*, 33(3), 203-216.
- Wilson, F. R. (1992). Digitizing digital dexterity: A novel application for MIDI recordings of keyboard performance. *Psychomusicology*, 11(2), 79-95.

Table 1. Error frequencies and percentages.

Piece	Score	Score-based errors					Non-score-based errors		
	notes	Deletion errors		Pitch	Timing	Total	Insertion errors		Total
	(played		errors	errors	score			non-	
	notes)				errors			score	
									errors
		Omis-	Added				Intru-	Repe-	
		sions	ties*				sions	titions	
Premier	15,360	35	98	37	38	208	56	16	72
Agnus	(15,231)	0.23%	0.64%	0.24%	0.25%	1.35%	0.37%	0.11%	0.48%
Wachet	11,808	17	222	41	25	305	96	67	163
auf	(11,607)	0.14%	1.88%	0.35%	0.21%	2.58%	0.83%	0.58%	1.40%
Dorian	86,432	116	75	189	156	534	380	132	512
fugue [†]	(92,257)	0.13%	0.09%	0.22%	0.18%	0.62%	0.41%	0.14%	0.55%

Note. Frequencies and percentages are computed on the aggregate data of all performances of a given piece (48, 32, and 32 performances were recorded for the *Premier Agnus*, *Wachet auf*, and the Dorian fugue, respectively). Score-based errors are expressed as percentages of all score notes, non-score-based errors as percentages of total notes played. The total number of notes played is indicated in parentheses.

* Added ties correspond to repeated notes in the score that were not re-attacked in performance, resulting in a missing note onset (see section *Coding of performance errors* in the Introduction).

[†] The number of notes played far exceeds the number of score notes for the Dorian fugue, given that the performances were heavily ornamented.

Table 2. Repeated-measures logistic regression on error rates for the Dorian fugue (with onset density as covariate).

Source	df	χ^2	p
Voice position	1	110.9	< .001
Motivcity	1	8.58	.003
Hand	1	1.27	.260
Voice position × Musical salience	1	0.44	.507
Voice position × Hand	1	6.33	.012
Musical salience × Hand	1	14.4	< .001
Voice position × Musical salience × Hand	1	3.49	.062

Table 3. Effect of musical texture on the type of pitch and intrusion errors.

	Only harmonically related	Only melodically related	Harmonically & melodically related	Harmonically & melodically unrelated	Chance estimates
<i>Premier Agnus</i>	8	34	8	43	3.2/12
(polyphonic)	(8.6%)	(36.6%)	(8.6%)	(46.2%)	(26.6%)
<i>Wachet auf</i>	25	36	22	54	3.1/12
(homophonic)	(18.2%)	(26.3%)	(16.1%)	(39.4%)	(25.9%)

Note. Error frequencies are given for each error type, with percentages (relative to the total number of pitch and intrusion errors) in parentheses. Chance estimates provided for the proportion of harmonically related errors.

Table 4. Mean Jaccard coefficients for error patterns between all pairs of performances for all three pieces.

	<i>Premier Agnus</i> (df = 147)			<i>Wachet auf</i> (df = 150)			Dorian fugue (df = 1383)		
	pairs	mean	SD	pairs	mean	SD	pairs	mean	SD
Within	117	0.12	0.15	48	0.28	0.14	16	0.14	0.12
Between	1005	0.05	0.09	448	0.11	0.09	480	0.03	0.02
Two-tailed	W = 83289, z = 6.03,			W = 19157, z = 8.21,			W = 7356, z = 6.07,		
Wilcoxon tests	p < .001			p < .001			p < .001		

Note. Jaccard coefficients were calculated on an event-by-event basis between all pairs of performances for all three pieces (degrees of freedom given in parentheses). For each piece, the mean coefficient was computed within and between performers. Two-tailed Wilcoxon rank-sum tests were conducted to assess whether the within-performer coefficients were significantly higher than the between-performer coefficients.

Figure captions:

Figure 1. Effect of melodic emphasis on error rate for the *Premier Agnus*. Mean error rates (in %) averaged across performers. Error bars represent standard errors of the mean. a) Error rates by voice. b) Error rates by hand and voice position.

Figure 2. Effect of interpretation on error rate for *Wachet auf*. Mean error rates (in %) averaged across performers. Error bars represent standard errors of the mean. a) Error rates by voice. b) Error rates by hand and voice position.

Figure 3. Error rates for different structural note categories for the Dorian fugue. Mean error rate (in %) for all categories, averaged across performers. Error bars represent standard errors of the mean.

Figure 4. Effects of voice position, musical salience, and hand assignment on error rates for the Dorian fugue. Mean error rates (in %) for all categories, averaged across performers. Error bars represent standard errors of the mean.

Figure 1

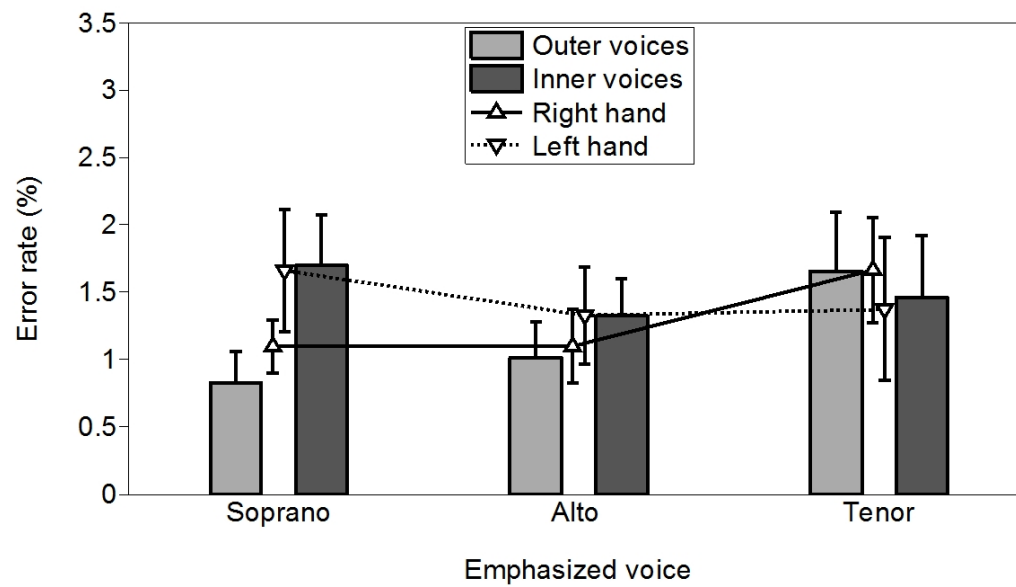
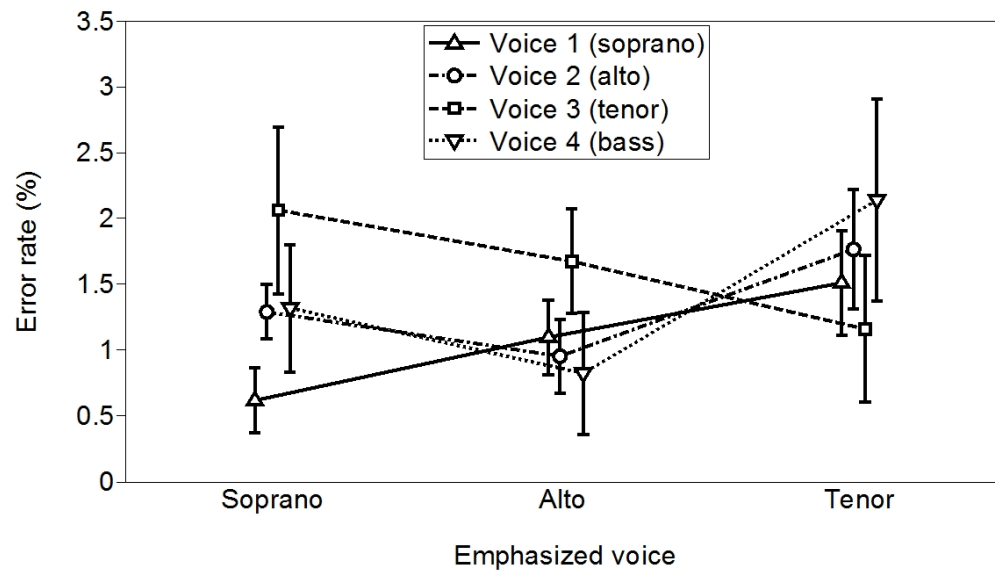
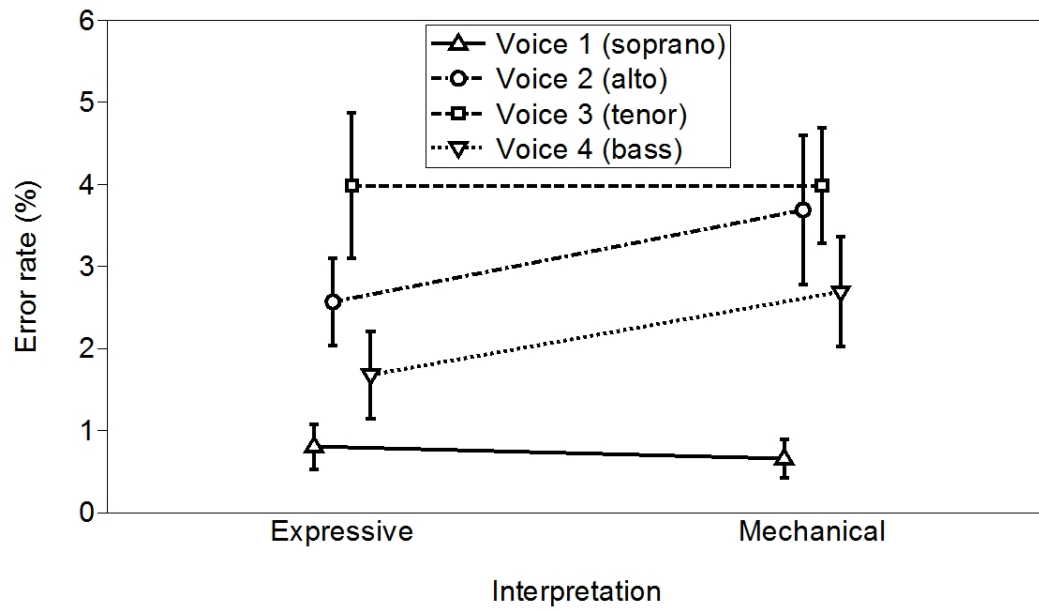


Figure 2



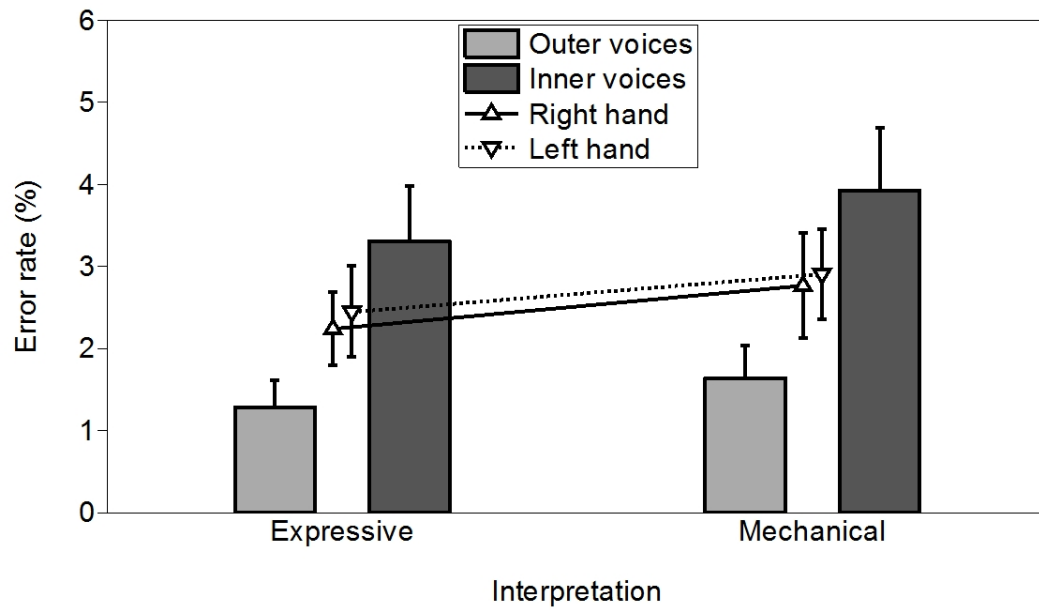


Figure 3

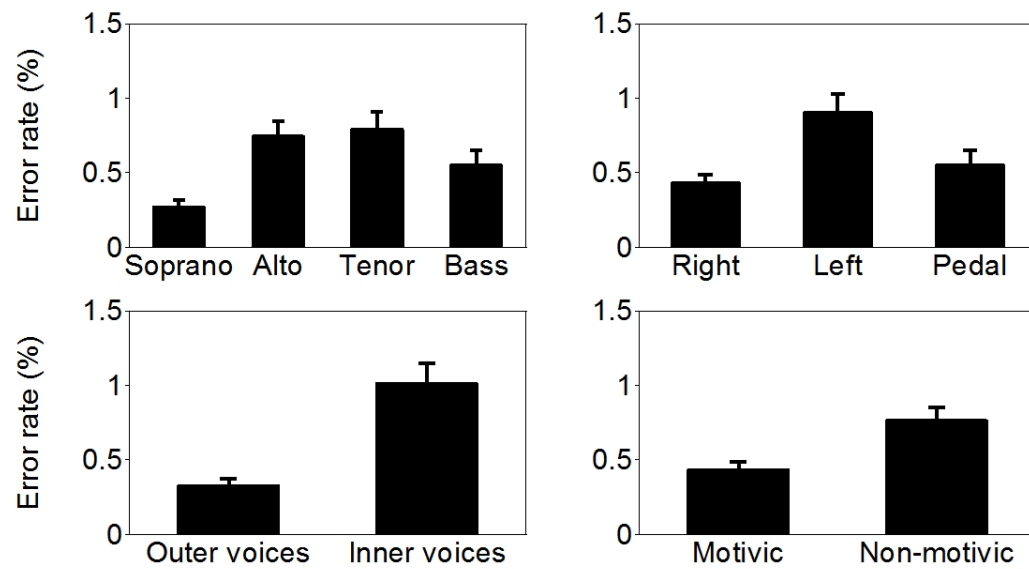
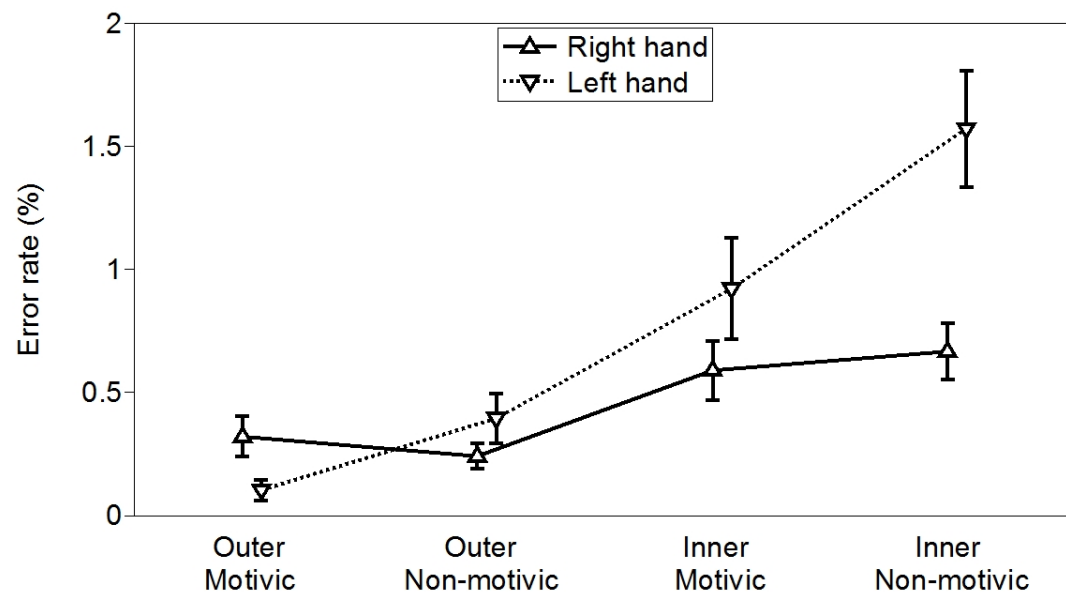


Figure 4



Appendi

x: musical scores

a) Nicolas de Grigny, *Premier Agnus*

The image displays a musical score for a piece by Samuel Scheidt, identified as 'Wachet auf, ruft uns die Stimme' (SSWV 534). The score is written for a single melodic line and a basso continuo line, both in a key signature of one flat (B-flat) and a common time signature (C). The notation is arranged in five systems, each with a treble clef for the upper voice and a bass clef for the lower voice. The first system contains measures 1 through 3. The second system, starting with a measure number '4' above the first measure, contains measures 4 through 8. The third system, starting with a measure number '9' above the first measure, contains measures 9 through 13. The fourth system, starting with a measure number '14' above the first measure, contains measures 14 through 18. The fifth system, starting with a measure number '19' above the first measure, contains measures 19 through 20, which concludes with a double bar line. The music features a variety of rhythmic values, including eighth, sixteenth, and thirty-second notes, as well as rests and accidentals (sharps and flats).

b) Samuel Scheidt, *Wachet auf, ruft uns die Stimme*, SSWV 534

In Organo Pleno

4

7

10

14

c) J.S. Bach, Fugue in D minor (“Dorian”), BWV 538, measures 1-29

This musical score is written for piano in G major, 4/4 time. It consists of four systems of staves, each with a grand staff (treble and bass clef) and a separate bass line.

- System 1 (Measures 1-8):** The right hand plays the **Subject**, a melodic line starting on G4. The left hand has whole rests. A trill (tr) is marked on the final note of the subject in measure 8.
- System 2 (Measures 9-15):** The right hand continues with the **First countersubject**, a descending eighth-note scale. The left hand has whole rests.
- System 3 (Measures 16-22):** The right hand introduces **Motive 1** (measures 16-17) and **Motive 2** (measures 18-22). The left hand plays the **Second countersubject**, a descending eighth-note scale. A trill (tr) is marked on the final note of the second countersubject in measure 22.
- System 4 (Measures 23-29):** The right hand continues with the **Second countersubject**. The left hand has whole rests.