Expressive strategies and performer-listener communication in organ performance

Bruno Gingras

Schulich School of Music

McGill University, Montreal

April 2008

A thesis submitted to McGill University in partial fulfilment of the requirements of the degree of Doctor of Philosophy

Abstract

This dissertation investigated expressive strategies and performer-listener communication in organ performance. Four core issues were explored: (a) the communication of voice emphasis; (b) the communication of artistic individuality; (c) the influence of musical structure on error patterns; and (d) the relationship between performers' interpretive choices and their analyses of the formal structure of a piece.

Performances were recorded on an organ equipped with a MIDI (Musical Instrument Digital Interface) console, allowing precise measurements of performance parameters. Performances were then matched to scores using an algorithm relying both on structural and temporal information, which I developed in the context of this project.

Two experiments investigated the communication of voice-specific emphasis in organ performance. The modification of articulation patterns was the most consistent strategy used by performers to emphasize a voice. Listeners who were themselves organists were more sensitive to differences between performers and interpretations than non-organists; however, musical structure was a major factor in the perception of voice prominence.

The perception of artistic individuality in organ performance was examined by inviting participants to sort different interpretations of a chorale setting by several performers. Most participants performed above chance level. The performance of musicians and non-musicians was comparable. Sorting accuracy was lower for mechanical interpretations than for expressive ones, demonstrating an effect of expressive intent. In addition, sorting accuracy was significantly higher for prize-winning performers than for non-winners.

Analyses of error patterns in organ performance showed that the likelihood of a note being wrongly played was inversely correlated with its degree of perceptual salience and musical significance or familiarity. Furthermore, individual performers exhibited consistent and idiosyncratic error patterns.

An exploration of the relationship between analysis and performance revealed that large tempo variations coincided with major formal subdivisions. Moreover, the degree of agreement on a formal subdivision was correlated with the magnitude of the concomitant tempo deviation.

By uniting music-theoretical analyses of three organ pieces, the systematic study of performance practice, the scientific investigation of the behavior of organists and listeners using methodologies from cognitive psychology, and computational methods for score-performance matching, this thesis proposes a new integrative framework for music performance research.

Résumé

Cette thèse étudie les stratégies expressives et la communication entre interprète et auditeur dans la musique d'orgue. Quatre thèmes principaux sont abordés: (a) la communication de l'accentuation des voix; (b) la communication de l'individualité artistique; (c) l'influence de la structure musicale sur les schémas d'erreurs; (d) les rapports entre les choix interprétatifs des organistes et leur analyse formelle d'une pièce.

Les enregistrements ont été réalisés sur un orgue muni d'une console MIDI (Musical Instrument Digital Interface), qui permet de mesurer précisément les paramètres expressifs. Les données MIDI ont ensuite été appariées à la partition au moyen d'un algorithme que j'ai développé dans le cadre de cette étude, et qui utilise à la fois l'information structurelle et temporelle.

Deux expériences explorent la communication de l'accentuation d'une voix à l'orgue. La modification des patrons d'articulation s'avère la stratégie utilisée le plus couramment pour faire ressortir une voix. Les auditeurs qui sont eux-mêmes organistes sont plus sensibles aux différences entre interprètes et interprétations que les non-organistes; cependant, la structure musicale représente un facteur important dans la perception de l'accentuation.

La perception de l'individualité artistique à l'orgue est examinée au moyen d'une expérience de catégorisation auditive d'une série d'interprétations d'un choral. La plupart des participants ont obtenu des taux de réussite supérieurs au hasard. Les résultats des musiciens et des non-musiciens sont comparables. Par contre, les interprètes ayant gagné des prix sont identifiés plus aisément que ceux qui n'ont pas été primés. En outre, les interprétations mécaniques sont moins bien classifiées que les interprétations expressives.

L'analyse de la répartition des erreurs montre que la probabilité qu'une note soit jouée de façon erronée est inversement corrélée avec son importance perceptuelle et musicale. D'autre part, les schémas d'erreurs sont spécifiques et particuliers à chaque interprète.

L'examen des rapports entre analyse et interprétation révèle que les variations de tempo plus marquées coïncident avec les principales démarcations formelles. De plus, pour une démarcation donnée, l'ampleur de ces variations est reliée au degré de concordance entre interprètes.

En combinant l'analyse musico-théorique de trois pièces d'orgue, l'exploration systématique des pratiques d'interprétation, l'investigation du comportement des organistes et des auditeurs par le biais d'une approche cognitiviste, et l'utilisation de techniques automatisées d'appariement à la partition, cette thèse présente un nouveau modèle intégratif pour la recherche en interprétation.

Acknowledgments

I would like to thank my supervisors, Stephen McAdams and Peter Schubert, who encouraged me to pursue this interdisciplinary research project, which combines my interests in music theory, music cognition, and music performance. Professor McAdams provided an inspiring laboratory environment and gave me the necessary resources to conduct the experiments described in this thesis. I also credit him for persuading me to learn the MATLAB programming language and write the score-performance matcher. Professor Schubert supported me through various projects, dispensed wise guidance, and made sure that my research remained firmly rooted on solid music-theoretical foundations.

This thesis began as a course project under the supervision of Caroline Palmer, whose ideas spurred various research questions and inspired a number of experiments. I am heavily indebted to her work for many of the themes developed in this dissertation.

I wish to express my sincere appreciation to the organists whose performances were recorded for this project. I could not have conducted this research project without their collaboration. I would also like to express my gratitude to the musical authorities of the Church of St-Andrew & St-Paul for permission to use their Casavant organ, which was at the time the only one in Montreal equipped with a MIDI console.

Julien Boissinot, Darryl Cameron, and Bennett Smith provided technical assistance with computers and sound recording material on several occasions. In addition, Bennett Smith designed the interface for the experiment on the perception of voice emphasis. I am grateful to Peter Holmes for letting me borrow McGill University's sound recording equipment numerous times. Jonas Braasch and Nils Peters offered valuable advice regarding the recording of organ performances. Rhonda Amsel provided helpful suggestions regarding the statistical analysis of the data on performance errors and on the communication of voice emphasis. Tamara Lagrandeur-Ponce and Bruno Giordano contributed their ideas and helped with the data collection and analysis for the study on the communication of artistic individuality.

I am pleased to acknowledge the generous financial support of the Social Sciences and Humanities Research Council of Canada and of the Centre for Interdisciplinary Research in Music Media and Technology (CIRMMT), which granted me fellowships that enabled me to complete my dissertation. This research was also supported by a grant from the Natural Sciences and Engineering Research Council and a Canada Research Chair awarded to Stephen McAdams.

Finally, I would like to thank my family and Manuela for their unwavering emotional support, particularly in the finishing stages of this dissertation.

Contributions of authors

The following chapters are based on manuscripts prepared for submission to peer-reviewed journals:

1. Gingras, B., McAdams, S., & Schubert, P. N. The communication of voice emphasis in organ performance. Manuscript prepared for submission to *Music Perception* (Chapter 2).
2. Gingras, B., Lagrandeur-Ponce, T., Giordano, B. L., & McAdams, S. The communication of artistic individuality in organ performance. Manuscript prepared for submission to *Perception* (Chapter 3).
3. Gingras, B., McAdams, S., Palmer, C., & Schubert, P. N. Performance error frequencies are inversely proportional to perceptual salience and musical significance. Manuscript prepared for submission to *Music Perception* (Chapter 4).
4. Gingras, B., McAdams, S., & Schubert, P. N. The performer as analyst: A case study of J.S. Bach's "Dorian" fugue (BWV 538). Manuscript prepared for submission to *Journal of New Music Research* (Chapter 5).
5. Gingras, B., & McAdams, S. Improved score-performance matching using both structural and temporal information from MIDI recordings. Manuscript prepared for submission to *Computer Music Journal* (Chapter 6).

I was responsible for designing and carrying out the experiments, conducting the data analysis, and preparing the manuscripts for all of the above-mentioned chapters. Professor Stephen McAdams provided the necessary funding and laboratory equipment, and generally contributed guidance in the conception and interpretation of the experiments and data. Professor Peter Schubert contributed to the experimental design and the interpretation of the data collected in the studies presented in Chapters 2, 4, and 5, as well as providing general supervision and guidance. Professor Caroline Palmer was involved in the design of the experiments described in Chapter 4 and provided critical suggestions regarding the data analysis. Tamara Lagrandeur-Ponce helped with the data collection and analysis in the sorting task experiment described in Chapter 3, whereas Bruno Giordano assisted in the design of the interface used in the sorting task and in the statistical analysis of the data collected in this experiment.

Table of contents

List of figures

List of tables

Chapter 1. Introduction

Scholarly writing on music performance has increased enormously in recent years (Gabrielsson, 2003). However, experimental research on music performance has been carried out for the most part by scholars whose main area of expertise lies outside music, whereas the study of performance by music theorists and musicologists has generally remained non-empirical and subjective, centering on the analytical, pedagogical, socio-cultural and philosophical implications of performance practice (Berry, 1989; Cook, 2007; Cook, Johnson, & Zender, 1999; Davies, 2001; Rink, 1995b, 2002). Music theorists, musicologists, as well as performers, could benefit immensely by reclaiming the field of empirical performance research, where a combination of experimental methodology, quantitative analysis, and musical expertise stands to yield fruitful insights. By allowing an objective characterization of performance parameters, the use of experimental methods opens up new fields of inquiries in performance research and sets the stage for a more rigorous analysis of topics of interest to musicologists and theorists.

Quantitative research on music performance has so far largely focused on the piano, and more specifically on classical and Romantic repertoire (Gabrielsson, 2003). Although a few empirical studies have explored violin (De Poli, Roda, & Vidolin, 1998), guitar (Askenfelt & Jansson, 1992; Heijink & Meulenbroek, 2002), and clarinet performance (Vines, Krumhansl, Wanderley, & Levitin, 2006), other instruments have been largely neglected. However, while the piano can justifiably be seen as a model instrument for performance research, due to its relative ease of use in a laboratory setting, its widespread practice among the general population, and the large amount of music written for this instrument, it remains to be seen whether observations relating to piano performance are applicable to other instruments. In particular, it is interesting to consider the case of keyboard instruments such as the organ or harpsichord, for which it is virtually impossible to differentiate individual notes on the basis of intensity or timbre (ignoring registration effects or the use of the swell and crescendo pedals on the organ). Although organ music is an important part of the Western musical tradition, very few empirical studies on organ performance have been published so far (Jerkert, 2004; Nielsen, 1999). Because the organist has little control over local timbre variations or note intensity, timing becomes the main expressive parameter by which the performer must convey most, if not all, of the musical expressivity on this instrument. Organ performance thus presents a uniquely restrictive paradigm for a case study of music performance.

The development of MIDI (Musical Digital Instrument Interface) technology (Roads, 1996), although initially intended for performers and composers, has greatly benefited piano performance research as well (Goebl & Bresin, 2003; Palmer, 1989). However, until now, no empirical study on organ performance using MIDI technology has been published. Having established a fruitful collaboration with the Church of St-Andrew & St-Paul, which hosts one of the largest organs in the Montreal area, and the only pipe organ equipped with a MIDI console that incorporates a replay feature, I was in a privileged position to conduct such a study.

My dissertation investigates expressive strategies and performer-listener communication in organ performance. More specifically, it explores four core issues: (a) the communication of voice emphasis; (b) the communication of artistic individuality; (c) the influence of musical structure on error patterns; and (d) the relationship between performers' interpretive choices and their analyses of the formal structure of a piece. This research project unites music-theoretical analyses of three organ pieces, the systematic study of performance practice on an instrument that has been ignored by the music performance research community, the scientific study of the behavior of organists and of listeners using methodologies from cognitive psychology, and computational methods for analyzing MIDI representations of the performances with respect to the original score. As such, this thesis achieves a unique synthesis of approaches borrowed from several disciplines, thus proposing a new integrative paradigm for research on expressive strategies and performer-listener communication in organ music. This paradigm could be applied to other musical instruments, and several tools developed over the course of this project, such as the matching algorithm and the experimental interfaces developed to investigate performer-listener communication, constitute significant innovations from which other studies on music performance will likely benefit. Finally, by reaching out to performers, music theorists, as well as experimental scientists, this study attempts to bridge the intercultural gap between art and science.

COMMUNICATION IN MUSIC PERFORMANCE: A REVIEW

The communication of expressive intention in music performance is a complex issue, which involves both the controlled use of expressive strategies by the performer as a means to convey a specific interpretation and the recognition of this expressive intent by the listener. The expressive content of a musical performance is multifaceted: according to Clarke (2002, p.190), "the sounds of a performance have the potential to convey a wealth of information to a listener, ranging from physical characteristics related to the space in which the performance is taking place and the nature of the instrument, to less palpable properties such as the performance ideology of the performer". Among the elements thought to be communicated in music performance are moods and emotions (Juslin, 2001), markers of a performer's artistic individuality (Sloboda, 2000), and aspects related to the structural content of a piece (Palmer, 1997). In many cases, the communication of a specific interpretation of the musical structure requires the performer to use expressive strategies in an attempt to direct listeners' attention to local elements such as motives and themes, or to more general features such as musical parts (or voices) in a polyphonic texture. While performance errors may be viewed as unwelcome by-products of music production activities, several studies have shown that error patterns are shaped by considerations linked to performers' expressive goals (Palmer & Van de Sande, 1993, 1995; Repp, 1996a); consequently, their investigation is also deeply relevant to the understanding of communication processes in music performance.

The following paragraphs will briefly review the literature addressing these topics and introduce the main issues examined in this dissertation.

The communication of voice emphasis

A substantial body of research has been conducted in order to identify and characterize the expressive strategies used by pianists to emphasize a given voice or melody in a polyphonic texture (Goebl, 2001; Palmer, 1989, 1996; Repp, 1996b). These studies have shown that the notes of the principal melody are played somewhat louder, and also 20 to 30 milliseconds earlier, than nominally simultaneous notes in other voices. This asynchrony between melody note onsets and note onsets in the remaining voices has been termed “melody lead.” While Palmer (1996) claims that pianists intentionally play the melody notes somewhat earlier, other researchers such as Repp (1996b) and Goebl (2001) have suggested that melody lead may be an artifact caused by the fact that, when a note is played louder, its key is pressed faster and strikes the hammer earlier than another key which is struck at the same time but softly.

Although the organ keyboard action may have superficially similar properties to the piano, a pipe valve is either open or closed, meaning that dynamic differentiation is impossible on the organ. In this context, organists may have to use expressive strategies which do not entail dynamic differentiation as a means to separate voices (Goebl, 2001). A logical hypothesis is that, since note intensity is constant, the parameter of articulation (offset-to-onset intervals) may become more important for distinguishing parts in a polyphonic setting for organ than it is on the piano.

An investigation of the expressive strategies used by organists to emphasize individual voices could also help shed light on the melody lead phenomenon. Indeed, if it could be shown that organists play notes in the emphasized voice 20-30 ms earlier than nominally simultaneous events in other voices, this would be a strong argument in favor of the hypothesis that melody lead can be used as an independent expressive device in the absence of dynamic differentiation. On the other hand, a lack of sizeable onset asynchronies in organ performance would imply that melody leads are indeed strongly linked to dynamic differentiation between voices.

The issue of the communication of voice emphasis in music performance may also be addressed by studying listeners' perception of voice prominence in performances of polyphonic music. However, we must first determine whether listeners can recognize and follow individual voices in a polyphonic texture, especially when these voices are not differentiated in timbre. In a study on the perception of polyphonic organ music, Huron (1989) found that the error rate in estimating the number of voices increased sharply when there were more than three voices, suggesting that listeners have difficulties following more than three concurrent parts. Moreover, Huron observed that voice entries were more difficult to detect in inner voices than in outer voices. This sensitivity differential in the perception of outer voices and inner voices has been replicated in several other studies, which confirmed that listeners were more sensitive to changes in the outer voices and especially in the highest voice (Dewitt & Samuel, 1990; Palmer & Holleran, 1994). Furthermore, this effect was recently documented at a pre-attentive level in electrophysiological studies (Fujioka, Trainor, Ross, Kakigi, &

Pantev, 2005). In the realm of psychoacoustics, a study on stream segregation in complex auditory sequences showed that temporal irregularities were detected more easily in outer subsequences than in inner ones (Brochard, Drake, Botte, & McAdams, 1999).

The communication of melodic emphasis in piano performance has been investigated specifically by Palmer (1996), who reported that whereas pianists could recognize the performer's emphasized melody both when intensity and timing cues were present and when only timing cues were present (in modified recordings), non-pianists could only recognize the emphasized melody in the presence of intensity and timing cues. This study suggested that onset asynchronies were, in themselves, sufficient to convey a sense of melodic emphasis, but only for listeners who had keyboard expertise. However, in an experiment comparing the role of asynchrony versus intensity in the perception of voice prominence in piano music, Goebl and Parncutt (2002) found that the effects of asynchrony were marginal, and that intensity differentiation was the major perceptual cue used by listeners. Little is known about the perception of voice emphasis on other keyboard instruments.

The communication of artistic individuality

Although a large body of research has been devoted to the study of communication of expressive intent in music performance, issues relating to the communication and perception of artistic individuality in music performance have been only tangentially addressed in music cognition research. Nevertheless, the more general problem of the recognition of individuals based on their actions or utterances has motivated a substantial body of research in various related fields.

Studies on the recognition of individuals based on their body movements, in which participants viewed point-light depictions of themselves, their friends or strangers performing various actions, have shown that subjects' visual sensitivity to their own motion was highest (Loula, Prasad, Harber, & Shiffrar, 2005). Subjects performed above chance when asked to identify their friends' actions, but not those of strangers. Moreover, actors were recognized more easily when performing expressive actions, such as boxing or dancing, than less expressive actions such as walking.

In the field of speaker recognition, researchers have established the prominent role of features such as fundamental frequency, formant mean, and speech rhythm, in the recognition of an individual's voice (Brown, 1981; Holmgren, 1967; Van Dommelen, 1990; Voiers, 1964). Later work has identified voice-selective areas in the human auditory cortex which could be responsible for speaker recognition (Belin, Zatorre, Lafaille, Ahad, & Pike, 2000). Building upon the well-established role of prosodic cues in speech perception, Palmer and her colleagues examined the role of musical prosodic cues (such as variations in amplitude and relative duration) in a discrimination task between familiar and novel performances of the same piece (Palmer, Jungers, & Jusczyk, 2001). Their results, which show that not only adult musicians and non-musicians, but also 10-month-old infants were able to identify correctly the familiar performances, provide evidence that prosodic features of music performances can be stored in memory.

Research on communication in expressive music performance has shown that both musicians and non-musicians can distinguish between different levels of expressiveness in performances of the same piece (Kendall & Carterette, 1990), and that they can recognize the emotions that performers intended to communicate (Juslin, 2000). More recently, Keller and colleagues reported that pianists were able to recognize their own performances reliably and were better at synchronizing themselves with their own pre-recorded performances in a piano duet than with performances from other pianists (Keller, Knoblich, & Repp, 2007). Focusing on the perception of similarity between musical performances, Timmers (2005) found that models based on absolute values of tempo and loudness were better predictors of perceptual distances between performances than models based on normalized variations, and that models based on local tempo features fared better than global models.

Although these studies, as well as several others, bear direct relevance on the issue of music performer identification by human listeners, no published study has focused explicitly on this topic, with the exception of Benadon (2003). Indeed, Stamatatos & Widmer's (2005) claim that their learning ensemble, which achieved a 70% recognition rate on piano performances of 22 pianists playing two pieces by Chopin, displayed a level of accuracy "unlikely to be matched by human listeners" has not yet been empirically verified.

Error patterns in music performance

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 synchronic, across-voice associations are emphasized, than in polyphonic pieces, which favor diachronic, 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). 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.

One aspect which has not been empirically examined so far is whether these effects would extend to piece-specific elements such as motives or themes. Performers could be expected to make fewer errors when playing motivic notes than non-motivic notes; likewise, listeners would be expected to be more sensitive to errors in motivic passages, especially if a motive or theme is familiar or easily recognizable. Additionally, a number of related issues have received little or no attention, such as the effects of hand assignment and structural salience on error rate, and the consistency and individuality of performers' error patterns. Finally,

[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.

the studies mentioned here were conducted on piano performance, using excerpts from the Romantic and Classical eras (Repp, 1996a) or short stimuli newly composed or adapted specifically for experimental purposes (Palmer & Van de Sande, 1993, 1995). It remains to be seen whether their findings could be extended to the performance of organ music from the Baroque repertoire.

Relationships between analysis and performance

Relationships between music-theoretical analysis and performance have been extensively treated in scholarly literature (Berry, 1989; Cone, 1968; Narmour, 1988; Rink, 1995b, 2002; Schmalfeldt, 1985). Whereas scholars such as Berry and Narmour intimated that performers should be acquainted with the theoretical and analytical methodology proposed by theorists, these studies were met, perhaps understandably, with little interest from performers. Indeed, these authors conveyed a view that simultaneously relegated the performers to a role of simple practitioners who should heed advice from the theorist regarding the structure of the pieces they are performing, while putting structural concerns to the forefront of performance issues (Cook, 1999). More recently, however, Rink (1995a) and Lester (1995) have advocated a different view, one that gives value to the performers' analytical insights about a piece. Lester even went so far as to reverse the paradigm accepted by scholars by proposing that analysts work from performances instead of working from the score. Leonard Meyer already hinted at such a view in 1973, when he wrote that, while performance is the actualization of an analytical act, this analysis may very well be intuitive and unsystematic: "For what a performer *does* is to make the relationships and patterns potential in the composer's score clear to the mind and ear of the experienced listener" (Meyer, 1973, p. 29).

Empirical investigations of piano performance have established that performers tend to slow down at sectional boundaries or formal subdivisions of a piece (Clarke, 1985; Gabrielsson, 1987; Palmer, 1989; Repp, 1990; Shaffer, 1981). This expressive device has been termed *phrase-final lengthening*. Moreover, it has been shown that the magnitude of the ritardando corresponds to the hierarchical importance of the boundary, with larger tempo variations associated with the major formal subdivisions of the piece (Repp, 1992; Shaffer & Todd, 1987; Todd, 1985). Several scholars proposed that these tempo fluctuations are a means of conveying information about the grouping structure of a piece to the listener, a model known as the *musical communication hypothesis* (Clarke, 1985, 1988; Palmer, 1989, 1996; Repp, 1992, 1995). Clarke (1989) reported that listeners were sensitive to minute changes in timing (as little as 20 ms for inter-onset intervals between 100 and 400 ms). Palmer (1989) demonstrated that tempo fluctuations were, at least in part, under the performers' voluntary control, since they were smaller in mechanical performances than in expressive performances of the same piece, and they could be modified according to the performers' interpretation of the piece. Penel and Drake (1998) refined these findings by showing that performers had more control over higher-level timing patterns, which involve phrases or larger sections of a piece, than over local timing patterns, which consist of rhythmic groupings comprising a few notes. More recently, Penel and Drake (2004) demonstrated that phrase-final lengthening could be accounted for partly by perceptual and motor constraints, and partly by the musical communication model. While further research is necessary to fully elucidate the role of phrase-final lengthening in expressive performance, there is sufficient evidence to posit a clear relationship between the timing variations applied by performers and the formal structure of the piece. However, the relationships between analysis and performance could be investigated in a more direct manner by inviting performers to record a piece for which they would be asked to provide their own written analyses, and to compare their performances to their analyses.

METHODOLOGY

This research project is based on two distinct series of experiments, one of which is centered on expressive strategies in organ performance and the other on the communication between performer and listener. The following paragraphs summarize the aims and experimental procedures associated with each series.

Expressive strategies in organ performance

In the first series of experiments, skilled organists who were either enrolled in or had already completed a degree in organ performance were invited to perform on the Casavant organ of the Church of St-Andrew & St-Paul in Montreal, which is equipped with a MIDI console. Performances were recorded under two different sets of conditions:

1. "Experimental" conditions in which organists were asked to follow specific interpretive guidelines, such as:

1.1 Emphasizing a specific voice (respectively the soprano, alto, and tenor parts) in performances of the *Premier Agnus* from the Mass of the *Premier Livre d'orgue* (1699) by Nicolas de Grigny (1672-1703);

1.2 Performing musically expressive and mechanical (that is, not adding any expressiveness beyond what is notated in the musical score) renditions of a chorale setting of *Wachet auf, ruft uns die Stimme* (SSWV 534) from the *Görlitzer Tabulaturbuch* (1650) by Samuel Scheidt (1587-1654);

2. A "recital-like" setting in which organists were invited to perform the Fugue in D minor (BWV 538), also known as the "Dorian" fugue, by J.S. Bach (1685-1750) as they would in a concert situation.

Performer-listener communication in organ performance

In the second series of experiments, listeners were invited to listen to recordings of the performances obtained in the first series. Two experiments were carried out in Stephen McAdams' Music Perception and Cognition Laboratory at the Schulich School of Music, McGill University, Montreal. The first one investigated the perception of voice prominence in polyphonic organ music, using the recordings of the *Premier Agnus*. This experiment used an innovative interface that allowed a continuous monitoring of the relative prominence of the voices over the course of a performance. The second one explored the perception of artistic individuality in organ performance by means of a sorting task in which listeners were asked to group together excerpts from the recordings of *Wachet auf* which they thought had been played by the same organist.

Data analysis

Analysis of the recorded performances. For each performance, MIDI and audio data were recorded. The MIDI data were then matched to the scores using a new score-performance matching algorithm which was written specifically for this research project. This matcher, which is described in detail in Chapter 6, constitutes a significant improvement over earlier algorithms since it takes into consideration not only the structural information, but also the temporal information available in the MIDI data (Heijink, Desain, Honing, & Windsor, 2000).

Statistical methods. Quantitative data obtained from the matched performances, as well as behavioral data obtained from the perceptual experiments, were analyzed using both traditional statistical methods, such as analyses of variance (ANOVA) and regression analyses, and more advanced methods, such as multidimensional scaling analyses (Borg & Groenen, 1997).

THESIS OUTLINE

Each of the four principal topics explored in this dissertation was given a chapter of its own. In addition, the description and evaluation of the score-performance matching algorithm was given a separate chapter. The following paragraphs present a brief outline of the dissertation.

Chapter 2 describes two experiments which respectively explore the production and perception of voice emphasis. The first one examines the expressive strategies used by organists to emphasize a voice, using the data from the performances of the *Premier Agnus*. Three parameters are analyzed: note onset asynchronies, local tempo variation, and articulation. The second experiment investigates the perception of voice prominence by asking participants to listen to selected recordings of the *Premier Agnus* and rate the relative prominence of the upper voices by means of a continuous response method.

Chapter 3 investigates the communication of artistic individuality by means of a sorting task in which listeners are asked to group together excerpts from the recordings of *Wachet auf* which they think have been played by the same organist. The first objective of this study is to determine whether participants could perform above chance in this perceptual task. A second objective is to identify the acoustical parameters used by listeners to discriminate between performers. Furthermore, since performers have been asked to record expressive and mechanical interpretations of *Wachet auf*, this study also seeks to assess the effect of expressive intent on the ability of listeners to identify performers. Finally, effects related to listeners' musical expertise and performers' level of accomplishment are examined.

Chapter 4 is concerned with the influence of musical structure (motivic versus non-motivic passages), texture (homophonic versus polyphonic style), expressive intent, conditions of preparation (quick-study versus prepared piece), and level of accomplishment (prize-winning performers versus non-winners) on the distribution and frequency of errors in organ performance. This study also addresses related issues such as the combined effects of hand assignment and structural salience on error rate and the degree of consistency and individuality of performers' error patterns. Recordings of all three pieces were used for this study:

the *Premier Agnus* and *Wachet auf* were used for the quick-study condition, while the Dorian fugue was used for the prepared condition.

Chapter 5 aims to clarify the relationship between the performer's view of the piece as an analyst and as a performer, by examining whether performers whose written analyses substantially differed also emphasized distinct formal aspects in their performances of the Dorian fugue. This project seeks to describe more accurately the link between interpretative choices and musical structure from a music-theoretical perspective. Furthermore, this study explores a stylistic repertoire that has been relatively neglected in the literature on performance research, which has generally focused on Classical and Romantic piano literature.

Chapter 6 introduces the score-performance matching algorithm used to match the MIDI recordings of the performances obtained for this project to the scores of the three pieces chosen for this study. This matcher relies on both structural and temporal information, allowing it to generate an accurate match even for heavily ornamented performances. A detailed description of the matching procedure is given, as well as a quantitative assessment of the accuracy of the algorithm. This chapter also introduces a heuristic for the identification of ornaments and errors that is based on perceptual principles, and which could theoretically be amenable to empirical study.

Finally, Chapter 7 summarizes the main findings presented in this thesis and suggests avenues for further research.

REFERENCES

Askenfelt, A., & Jansson, E. V. (1992). On vibration sensation and finger touch in stringed-instrument playing. *Music Perception, 9*(3), 311-349.

Belin, P., Zatorre, R. J., Lafaille, P., Ahad, P., & Pike, B. (2000). Voice-selective areas in human auditory cortex. *Nature, 403*(6767), 309-312.

Benadon, F. (2003). Spectrographic and calligraphic cues in the identification of jazz saxophonists. In R. Kopiez, A. C. Lehmann, I. Wolther & C. Wolf (Eds.), *Proceedings of the 5th Triennial ESCOM Conference* (pp. 245-249). Hanover, Germany.

Berry, W. (1989). *Musical structure and performance*. New Haven: Yale University Press.

Borg, I., & Groenen, P. J. F. (1997). *Modern multidimensional scaling: Theory and applications*. New York: Springer.

Brochard, R., Drake, C., Botte, M. C., & McAdams, S. (1999). Perceptual organization of complex auditory sequences: Effect of number of simultaneous subsequences and frequency separation. *Journal of Experimental Psychology-Human Perception and Performance, 25*(6), 1742-1759.

Brown, R. (1981). An experimental study of the relative importance of acoustic parameters for auditory speaker recognition. *Language and Speech, 24*(4), 295-310.

Clarke, E. F. (1985). Structure and expression in rhythmic performance. In P. Howell, I. Cross & R. West (Eds.), *Musical structure and cognition* (pp. 209-236). London: Academic Press.

Clarke, E. F. (1988). Generative principles in music performance. In *Generative processes in music: The psychology of performance, improvisation and composition* (pp. 1-26).

Clarke, E. F. (1989). The perception of expressive timing in music. *Psychological Research-Psychologische Forschung, 51*(1), 2-9.

Clarke, E. F. (2002). Listening to performance. In J. Rink (Ed.), *Musical performance: A guide to understanding* (pp. 185-196). United Kingdom: Cambridge University Press.

Cone, E. T. (1968). *Musical form and musical performance* (1st ed.). New York: W. W. Norton.

Cook, N. (1999). Analysing performance and performing analysis. In *Rethinking music* (pp. 239-261). United Kingdom: Oxford University Press.

Cook, N. (2007). *Music, performance, meaning: selected essays*. Aldershot, England; Burlington, VT: Ashgate.

Cook, N., Johnson, P., & Zender, H. (1999). *Theory into practice: Composition, performance and the listening experience*. Leuven, Belgium: Leuven University Press.

Davies, S. (2001). *Musical works and performances: A philosophical exploration*. Oxford, England; New York: Clarendon Press; Oxford University Press.

De Poli, G., Roda, A., & Vidolin, A. (1998). Note-by-note analysis of the influence of expressive intentions and musical structure in violin performance. *Journal of New Music Research, 27*(3), 293-321.

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.

Gabrielsson, A. (1987). Once again: The theme from Mozart's piano sonata in A major (K. 331). In A. Gabrielsson (Ed.), *Action and perception in rhythm and music* (pp. 81-104). Stockholm: Royal Swedish Academy of Music.

Gabrielsson, A. (2003). Music performance research at the millenium. *Psychology of Music, 31*(3), 221-272.

Goebl, W. (2001). Melody lead in piano performance: Expressive device or artifact? *Journal of the Acoustical Society of America, 110*(1), 563-572.

Goebl, W., & Bresin, R. (2003). Measurement and reproduction accuracy of computer-controlled grand pianos. *Journal of the Acoustical Society of America, 114*(4), 2273-2283.

Goebl, W., & Parncutt, R. (2002). The influence of relative intensity on the perception of onset asynchronies. In C. Stevens, D. Burnham, G. E. McPherson, E. Schubert & J. Renwick (Eds.), *Proceedings of the 7th International Conference on Music Perception and Cognition*. Sydney, Australia: Adelaide: Causal Productions.

Heijink, H., Desain, P., Honing, H., & Windsor, L. (2000). Make me a match: An evaluation of different approaches to score-performance matching. *Computer Music Journal, 24*(1), 43-56.

Heijink, H., & Meulenbroek, R. G. J. (2002). On the complexity of classical guitar playing: Functional adaptations to task constraints. *Journal of Motor Behavior, 34*(4), 339-351.

Holmgren, G. L. (1967). Physical and psychological correlates of speaker recognition. *Journal of Speech and Hearing Research, 10*(1), 57-66.

Huron, D. (1989). Voice denumerability in polyphonic music of homogeneous timbres. *Music Perception, 6*(4), 361-382.

Jerkert, J. (2004, June). *Musical articulation in the organ.* Paper presented at the Joint Baltic-Nordic Acoustics Meeting, Mariehamn, Finland.

Juslin, P. N. (2000). Cue utilization in communication of emotion in music performance: Relating performance to perception. *Journal of Experimental Psychology-Human Perception and Performance, 26*(6), 1797-1812.

Juslin, P. N. (2001). Communicating emotion in music performance: A review and a theoretical framework. In P. N. Juslin & J. Sloboda (Eds.), *Music and emotion: Theory and research* (pp. 309-337). New York: Oxford University Press.

Keller, P. E., Knoblich, G., & Repp, B. H. (2007). Pianists duet better when they play with themselves: On the possible role of action simulation in synchronization. *Consciousness and Cognition, 16*(1), 102-111.

Kendall, R. A., & Carterette, E. C. (1990). The communication of musical expression. *Music Perception, 8*(2), 129-164.

Lester, J. (1995). Performance and analysis: Interaction and interpretation. In *The practice of performance: Studies in musical interpretation* (pp. 197-216). United Kingdom: Cambridge University.

Loula, F., Prasad, S., Harber, K., & Shiffrar, M. (2005). Recognizing people from their movement. *Journal of Experimental Psychology-Human Perception and Performance, 31*(1), 210-220.

Meyer, L. B. (1973). *Explaining music: Essays and explorations*. Berkeley: University of California Press.

Narmour, E. (1988). On the relationship of analytical theory to performance and interpretations. In E. Narmour & R. A. Solie (Eds.), *Explorations in music, the arts, and ideas* (pp. 317-340). Stuyvesant, NY: Pendragon.

Nielsen, S. G. (1999). Learning strategies in instrumental music practice. *British Journal of Music Education, 16*(3), 275-291.

Palmer, C. (1989). Mapping musical thought to musical performance. *Journal of Experimental Psychology-Human Perception and Performance, 15*(12), 331-346.

Palmer, C. (1996). On the assignment of structure in music performance. *Music Perception, 14*(1), 23-56.

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

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., Jungers, M. K., & Jusczyk, P. W. (2001). Episodic memory for musical prosody. *Journal of Memory and Language, 45*(4), 526-545.

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.

Penel, A., & Drake, C. (1998). Sources of timing variations in music performance: A psychological segmentation model. *Psychological Research-Psychologische Forschung, 61*(1), 12-32.

Penel, A., & Drake, C. (2004). Timing variations in music performance: Musical communication, perceptual compensation, and/or motor control? *Perception & Psychophysics, 66*(4), 545-562.

Repp, B. H. (1990). Patterns of expressive timing in performances of a Beethoven minuet by 19 famous pianists. *Journal of the Acoustical Society of America, 88*(2), 622-641.

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. (1995). Expressive timing in Schumann's "Träumerei" - an analysis of performances by graduate student pianists. *Journal of the Acoustical Society of America, 98*(5), 2413-2427.

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). Patterns of note onset asynchronies in expressive piano performance. *Journal of the Acoustical Society of America, 100*(6), 3917-3931.

Rink, J. (1995a). Playing in time: Rhythm, metre and tempo in Brahms's Fantasien op. 116. In J. Rink (Ed.), *The practice of performance: Studies in musical interpretation* (pp. 254-282). United Kingdom: Cambridge University.

Rink, J. (1995b). *The practice of performance: Studies in musical interpretation*. Cambridge; New York: Cambridge University Press.

Rink, J. (2002). *Musical performance: A guide to understanding*. Cambridge, U.K.; New York: Cambridge University Press.

Roads, C. (1996). *The computer music tutorial*. Cambridge, Mass.: MIT Press.

Schmalfeldt, J. (1985). On the relation of analysis to performance: Beethoven's bagatelles op. 126, nos. 2 and 5. *Journal of Music Theory, 29*(1), 1-31.

Shaffer, L. H. (1981). Performances of Chopin, Bach, and Bartok: Studies in motor programming. *Cognitive Psychology, 13*(3), 326-376.

Shaffer, L. H., & Todd, N. P. M. (1987). The interpretive component in musical performance. In A. Gabrielsson (Ed.), *Action and perception in rhythm and music* (pp. 139-152). Stockholm: Royal Swedish Academy of Music.

Sloboda, J. A. (2000). Individual differences in music performance. *Trends in Cognitive Sciences, 4*(10), 397-403.

Stamatatos, E., & Widmer, G. (2005). Automatic identification of music performers with learning ensembles. *Artificial Intelligence, 165*(1), 37-56.

Timmers, R. (2005). Predicting the similarity between expressive performance of music from measurements of tempo and dynamics. *Journal of the Acoustical Society of America, 117*(1), 391-399.

Todd, N. (1985). A model of expressive timing in tonal music. *Music Perception, 3*(1), 33-58.

Van Dommelen, W. A. (1990). Acoustic parameters in human speaker recognition. *Language and Speech, 33*(3), 259-272.

Vines, B. W., Krumhansl, C. L., Wanderley, M. M., & Levitin, D. J. (2006). Cross-modal interactions in the perception of musical performance. *Cognition, 101*(1), 80-113.

Voiers, W. D. (1964). Perceptual bases of speaker identity. *Journal of the Acoustical Society of America, 36*(6), 1065-1073.

Chapter 2. The communication of voice emphasis in organ performance

Studies have shown that pianists emphasize a voice or melody in a polyphonic texture by playing its notes somewhat louder, and also 20 to 30 milliseconds earlier, than nominally simultaneous notes in other voices. However, little is known about the communication of voice emphasis on other keyboard instruments. This chapter describes two experiments which explore respectively the production and perception of voice emphasis in organ performance. The first one examines the expressive strategies used by organists to emphasize a voice in performances of a short Baroque polyphonic piece. Three parameters are analyzed: onset asynchrony, local tempo variation, and articulation. The second experiment investigates the perception of voice prominence by asking participants to listen to selected recordings collected in the first experiment and rate the relative prominence of the upper voices by means of a continuous response method.

This chapter is based on the following research article:

Gingras, B., McAdams, S., & Schubert, P. N. The communication of voice emphasis in organ performance. Manuscript prepared for submission to *Music Perception*.

ABSTRACT

Two experiments investigated the communication of voice-specific emphasis in organ performance. In Experiment 1, eight organists recorded three interpretations of a short Baroque polyphonic piece, each emphasizing a different voice, on an organ equipped with a MIDI console. Three parameters were analyzed: note onset asynchronies, local tempo variation, and articulation. Onset asynchronies were much smaller than those observed in piano performance, and were generally too small to be perceptible. Variations in the spread of local tempo deviations were observed across voices, but no systematic attempt to differentiate between voices according to a melodic interpretation could be detected. The modification of articulation patterns was found to be the most widespread and consistent strategy used by organists to emphasize a voice. Specifically, a voice was generally played in a more detached manner when it was emphasized than when it was not. In Experiment 2, 30 musicians (10 organists and 20 non-organists) listened to a selection of the recordings collected in Experiment 1 and rated the relative prominence of the upper voices using a continuous response method. Besides highlighting the importance of structural elements in the musical score such as salient passages in specific voices, results indicate that organists were more sensitive to differences between performers and interpretations than non-organists and that the communication of voice emphasis is not as efficient in organ performance as in piano performance.

INTRODUCTION

The communication of expressive intention in music performance is a complex issue, which involves both the controlled use of expressive strategies by the performer as a means to convey a specific interpretation and the recognition of this expressive intent by the listener. The expressive content of a musical performance is multifaceted. Among the elements generally thought to be communicated in music performance are moods and emotions (Juslin, 2001), markers of a performer's artistic individuality (Sloboda, 2000; see also Chapter 3), and aspects related to the structural content of a piece (Palmer, 1997). In many cases, the communication of a specific interpretation of the musical structure requires the performer to use expressive strategies in an attempt to direct the listener's attention to local elements such as motives and themes or to more general features such as musical parts (or voices) in a polyphonic texture. A substantial body of research has been conducted on piano performance in order to identify and characterize those expressive strategies, showing that pianists emphasize a melody or voice by playing its notes louder and earlier than nominally simultaneous notes in other voices (Goebl, 2001; Palmer, 1989, 1996; Repp, 1996b).

However, although the piano can justifiably be seen as a model instrument for such experiments, due to its relative ease of use in a laboratory setting, its widespread practice among the general population, and the large amount of music written for this instrument, it remains to be seen whether these findings may be applicable to other instruments. In particular, it is interesting to consider the case of keyboard instruments such as the organ or harpsichord, for which it is virtually impossible to differentiate individual notes on the basis of intensity (ignoring registration effects or the use of the swell and crescendo pedals on the organ). The first experiment described in this paper addressed this issue by analyzing the expressive strategies used by organists to emphasize specific voices in a polyphonic organ piece.

The issue of communication in music performance may also be addressed from the listener's viewpoint by asking how successful listeners are at recognizing the expressive intent that the performer attempted to convey. The second experiment presented in this article sought to answer this question by asking listeners to rate the relative prominence of the voices for the performances recorded in Experiment 1.

Emphasizing specific parts in polyphonic keyboard performance

Musical expressivity in piano performance is essentially conveyed by manipulating three parameters: the inter-onset interval between successive notes (local variations of tempo such as *rubato* and *accelerando*), the intensity of the notes (dynamics), and the offset-to-onset intervals (articulation effects, such as *legato* and *staccato*). Regarding the expressive strategies used by pianists to emphasize a given voice or melody in a polyphonic texture, several studies have shown that the notes of the principal melody are played somewhat louder, and also 20 to 30 milliseconds earlier, than nominally simultaneous notes in other voices (Goebl, 2001; Palmer, 1989, 1996; Repp, 1996b). This onset asynchrony between the melody notes and notes in the remaining voices has been termed "melody lead." While Palmer (1996) claims that pianists intentionally play the melody notes somewhat earlier, other researchers such as Repp (1996b) and Goebl (2001) have suggested that melody lead may be an artifact due to the fact that when a note is played louder, its key is pressed faster and strikes the hammer earlier than another key that is struck at the same time but softly.

Although the organ keyboard action may have superficially similar properties to the piano, a pipe valve is either open or closed, meaning that dynamic differentiation is impossible on the organ. In this context, organists may have to use expressive strategies which do not entail dynamic differentiation as a means to separate voices (Goebl, 2001). A logical hypothesis is that, because note intensity is constant, the parameter of articulation (offset-to-onset intervals) may become more important for distinguishing parts in a polyphonic setting for organ than it is on the piano.

Studying the expressive strategies used by organists to emphasize individual voices could also help shed light on the melody lead phenomenon. Indeed, if it can be shown that organists play notes in the emphasized voice 20-30 ms earlier than nominally simultaneous events in other voices, even though this strategy cannot help differentiate between voices on the basis of intensity, this would be a strong argument in favor of the hypothesis that melody lead can be used as an independent expressive device in the absence of dynamic differentiation. On the other hand, a lack of substantial melody leads in organ performance would imply that melody leads are indeed tied to dynamic differentiation between voices.

The perception of voice prominence in polyphonic textures

Prior to addressing issues related to the perception of voice emphasis in polyphonic texture, it must be determined whether listeners can recognize and follow individual voices in a polyphonic texture, especially when these voices are not differentiated in timbre. In a study on the perception of polyphonic organ music, Huron (1989) found that the error rate in estimating the number of voices increased sharply when there were more than three voices, suggesting that listeners have difficulties tracking more than three concurrent parts. Moreover, he observed that voice entries were more difficult to detect in inner voices than in outer voices. This sensitivity differential in the perception of outer voices and inner voices has been replicated in several other studies, showing that listeners were more sensitive to changes in the outer voices and especially in the highest voice (Dewitt & Samuel, 1990; Palmer & Holleran, 1994). Furthermore, this effect was recently documented at a pre-attentive level in electrophysiological studies (Fujioka, Trainor, Ross, Kakigi, & Pantev, 2005). In the realm of psychoacoustics, a study on stream segregation in complex auditory sequences showed that temporal irregularities were detected more easily in outer subsequences than in inner ones (Brochard, Drake, Botte, & McAdams, 1999).

The communication of melodic emphasis in piano performance has been studied by Palmer (1996), who reported that whereas pianists could recognize the performer's emphasized melody both when intensity and timing cues were present and when only timing cues were present (in modified recordings), non-pianists could only recognize the emphasized melody in the presence of intensity and timing cues. This study suggested that onset asynchronies were, in themselves, sufficient to convey a sense of melodic emphasis, but only for listeners who had keyboard expertise. However, in an experiment comparing the role of asynchrony versus intensity in the perception of voice prominence in piano music, Goebl and Parncutt (2002) found that the effects of asynchrony were marginal, and that intensity differentiation was the major perceptual cue used by listeners. Little is known about the perception of voice emphasis on other keyboard instruments.

EXPERIMENT 1: PRODUCTION OF VOICE EMPHASIS

In order to identify the expressive strategies used by organists to emphasize a specific voice, organists were invited to record different interpretations of a polyphonic organ piece in which they emphasized different voices. The *Premier Agnus*, from the *Premier livre d'orgue* (1699) by Nicolas de Grigny (1672-1703), was chosen for this experiment as being representative of the Baroque organ repertoire (Figure 2.1; trills, mordents, and grace notes were removed from the original score). This relatively short piece can be played without the use of the pedals. As is typical of the Baroque contrapuntal writing style, the piece contains four distinct melodic lines (parts or voices): these are, from the highest to the lowest, the soprano, alto, tenor, and bass parts. In contrast to the Classical and Romantic piano repertoire, this piece has no obvious principal melodic line and thus lends itself well to multiple interpretations. Another motivation behind the choice of this particular piece is the fact that the four voices are active throughout the piece, and the melodic and rhythmic content of the three upper voices is relatively similar (the bass voice is, however, markedly different). Finally, this piece has no obvious recurring thematic material, which made it ideally suited for a study on the communication of voice emphasis; otherwise, performers, as well as listeners, might have been sensitive to the recurrence of familiar motives, which could have been a potentially confounding factor.

Performances were recorded on an organ equipped with a MIDI console, which allows precise measurement of performance parameters. Three parameters were analyzed from the MIDI data: note onset asynchronies, local tempo variations, and articulation.

Method

Participants

Eight skilled organists (O1, O2,…, O8), two female and six male, all right-handed, participated in the experiment. They were professional organists from the Montreal area, or organ students at McGill University in Montreal. Their average age was 27 years (the youngest was 23; the oldest 30). They had received organ instruction for a mean duration of 10 years (minimum 7, maximum 13). All of them held or had held a position as church organist for an average of 8 years (minimum 1; maximum 21). Three of them had previously won prizes in provincial or national organ competitions. All organists had also played piano for an average of 16 years (minimum 5; maximum 27), though most of them claimed to have played the piano only “sometimes” or “rarely” during the two years preceding the experiment. Six of them had already played on the Casavant organ used for the recording session. None of them were familiar with the piece. Organists were paid $20 for their participation.

Figure 2.1. Nicolas de Grigny, *Premier Agnus*. Score prepared with computer software. Ornaments such as trills, mordents, and appoggiaturas were removed from the original score.

Materials and apparatus

Organists performed the *Premier Agnus* by Grigny using the score shown in Figure 2.1. The performances were recorded on the Casavant organ of the Church of St-Andrew & St-Paul in Montreal. This five-manual organ (five keyboards and a pedal-board) 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.[1] This MIDI system did not include a facility for key velocity measurement. For the experiment, the stops used were the Spitz Principal 8’, the Spitz Principal 4’, and the Fifteenth 2’ on the “Great” manual.

The audio signal was recorded through two omnidirectional microphones Boehringer ECM 8000. The microphones were located 1.20 m behind the organ bench, at a height of 1.70 m, and were placed 60 cm apart. The audio and MIDI signals were sent to a PC computer through a MOTU audio interface. Audio and MIDI data were then recorded using Cakewalk’s SONAR software and stored on a hard disk.

Procedure

The score was given to the organists 20 minutes before the recording session began in order to give them time to practice. They were instructed to record three different interpretations of the piece. In one interpretation, they strove to emphasize or bring out, the soprano part, in another, the alto part, and in a third one, the tenor part. For each of the three instructions, two recordings were made (the organists were allowed to do three recordings and choose the two most satisfactory). The order of the instructions was randomized according to a Latin square diagram.

[1] Information provided by Mark Gilliam, Sales manager of Solid State Organ Systems.

Data analysis

A unique identifier was assigned to each note attack indicated in the score for a total of 320 notes, of which 91 were labeled as belonging to the soprano voice (the uppermost voice, identified as Voice 1), 92 to the alto voice (Voice 2), 97 to the tenor voice (Voice 3), 38 to the bass voice (Voice 4), and 2 to additional inner voices in the last chord of the piece (Voices 5 and 6). Similarly, a unique identifier was assigned to all nominally simultaneous note onsets (two or more notes attacked together) present in the score.

Note onsets and offsets were extracted from the MIDI data of the performances and matched to the score. Note onset values are dependent on the precise location at which they are measured; the measurements reported in the present study correspond to the key-bottom contact, as is the case with electronic keyboards (Goebl, 2001, p. 564).[2] Wrong notes were marked as pitch errors (or substitutions), omissions (including "added ties" – repeated notes in the score that were not re-attacked in performance), and timing errors, intrusions and repetitions (re-attacked notes in performance that were not repeated in the score).[3] For all performances, the rate of errors, defined as the proportion of wrong notes or missing notes relative to the total number of score notes, was very low, especially considering that the subjects were unfamiliar with the piece and had 20 minutes to

[2] Goebl & Bresin (2003) analyzed in detail the measurement accuracy of a computer-controlled grand piano. To the author's knowledge, no such study is available for an organ equipped with a MIDI console.

[3] "Untied" notes (Repp, 1996a) were treated as repetitions. Timing errors are not mentioned in Repp (1996a). Such mistimed attacks are clearly heard as errors, rather than as expressive mannerisms when listening to the recordings. The largest reported expressive asynchronies in piano performance rarely exceed 100 ms, especially in the right hand (Goebl, 2001; Repp, 1996b).

rehearse it before the recordings: 1.11% (of n_{total} = 15,360), comparable to the error rates observed by Repp(1996a), Palmer (1996), and Goebl (2001).

Results

For each of the three expressive parameters analyzed (onset asynchrony, local tempo variations, and articulation), comparisons of group means across all voice/emphasis combinations are presented, followed by comparisons of the note-by-note patterns between performances. These two approaches are seen as complementary, as the first examines global statistical tendencies whereas the second provides a measure of similarity between performances.

Note onset asynchrony

Note onset asynchrony, or chord asynchrony, is defined as the difference in onset time between note onsets that are notated in the musical score as synchronous (Palmer, 1989). Several measures of onset asynchrony have been constructed. Rasch (1979) proposed to use the root mean square, or standard deviation of the onset times of nominally simultaneous notes. Palmer (1989, 1996) used the difference in onset times between the notated melody and the mean onset of the remaining voices, while Repp (1996b) presented a measure of asynchrony in which the lag time for each individual note in a chord was obtained by subtracting from its onset time that of the highest note in the chord. Goebl's (2001) melody lead, defined as the difference in onset time between the melody and each other voice in a chord, conceptually mirrors Repp's lag time, save for the distinction between "highest note" (Repp) and "melody" (Goebl).

The choice of the highest note as a reference note for the computation of asynchronies seemed inappropriate for this particular organ piece, because the main melody was not necessarily located in the uppermost voice. Asynchronies were thus calculated for each note as the difference between its onset time and the mean onset of the remaining notes in the chord, with a positive asynchrony referring to a lead, as described in Palmer (1989). One potential disadvantage of using this definition is that the sum of those differences, when computed for all the notes, will necessarily equal zero. Consequently, the asynchronies computed for all voices are not independent variables. Analyses were thus conducted separately for each voice.

As shown in Figure 2.2, mean asynchronies for each voice were very small, averaging at most a few milliseconds for all voice/interpretation combinations. Chord asynchronies, measured using Rasch's (1979) definition, averaged 9 ms. In comparison, Palmer (1989) reported chord asynchronies of 18 ms for musical performances at the piano. Furthermore, the total number of large asynchronies was relatively low: only 16.8%, or roughly one-sixth, of all nominally simultaneous note pairs were performed with asynchronies larger than 20 ms (2,051 of 12,227 note pairs).

Mixed-model repeated-measures analyses of variance (ANOVA) were conducted separately on the mean asynchronies for each voice, with emphasized voice as within-subject factor. Main effects of voice emphasis were found for the soprano, $F(2, 14) = 5.58$, $p < .05$, alto, $F(2, 14) = 11.38$, $p < .01$, and tenor, $F(2, 14) = 12.92$, $p < .001$, but not for the bass. These results indicate that interpretation affected asynchronies for the upper voices: as can be seen in Figure 2.2, larger positive asynchronies were observed for a voice when it was emphasized. However, the melody lead, measured using Palmer's (1989) definition and treating the emphasized voice as melody for each interpretation, was negligible: the mean melody lead, computed across all performances, averaged 2.0 ± 0.6 ms. In fact, only the melody lead for the tenor was significantly greater than zero (one-tailed *t* tests, Bonferroni-corrected, $p < .05$). In comparison, Palmer (1989, 1996) and Goebl (2001) reported average melody leads of 20-30 ms.

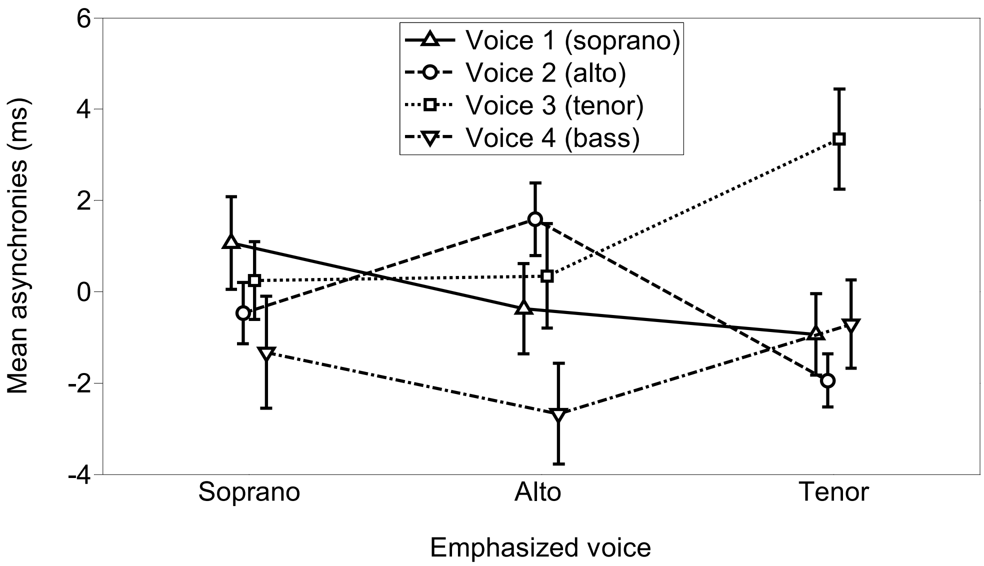


Figure 2.2. Mean onset asynchronies for all voice/emphasis combinations (excluding Voices 5 and 6). Values averaged across organists. Error bars represent standard errors of the mean.

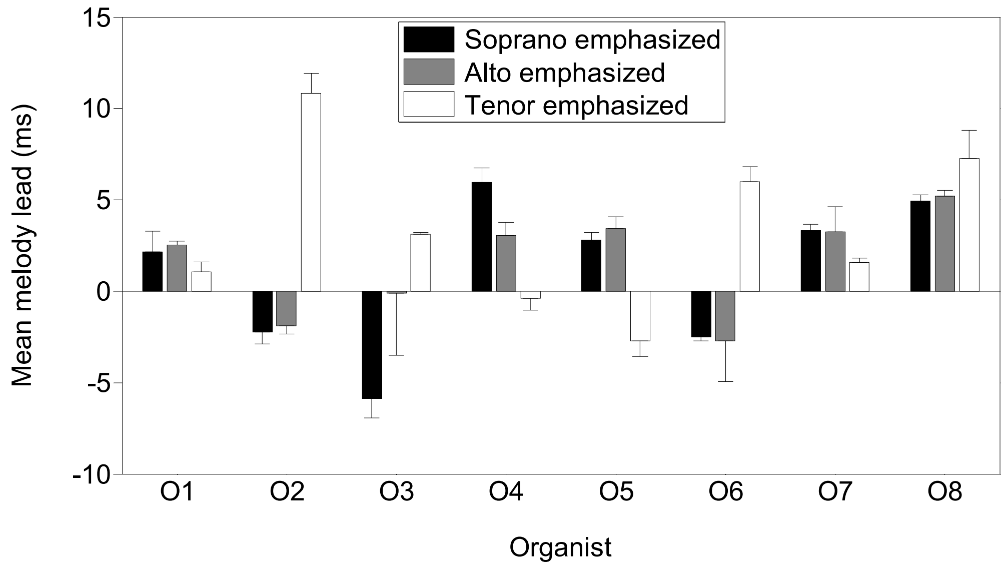


Figure 2.3. Mean melody leads by emphasized voice. Values computed as the differences between onset times of notes in the emphasized voice and the mean onset times of nominally simultaneous notes in the remaining voices, for each organist. Each bar represents the average across two performances. Error bars represent standard errors of the mean.

An examination of the individual organists' profiles (Figure 2.3) reveals that only Organist 2 had a mean melody lead larger than 10 ms, when emphasizing the tenor part. The mean melody leads of several organists were negative, indicating that the emphasized voice actually trailed the other voices. Although the melody leads observed here were much smaller than those reported in piano performance studies, it is interesting to note that the three organists who showed consistently positive melody leads across all instructions (O1, O7, and O8) were the only participants who claimed to have played the piano "frequently" in the two years preceding the experiment. A mixed-model repeated-measures ANOVA conducted on the mean melody lead with emphasized voice as a within-subject factor showed no significant effect of voice emphasis, $F(2, 14) = 1.68$, $p = .22$, indicating that mean melody leads did not vary significantly according to which voice was emphasized.

In order to compare patterns of asynchronies between performances, note-by-note correlations were computed between all pairs of performances for every note for which an onset asynchrony value could be determined (Table 2.1).[4] The mean correlation coefficient for all pairwise comparisons between the 48 performances was relatively low, as only 22.2% of all pairwise correlations were highly significant ($p < .01$) (Table 2.1a). The group comparisons show that organists had more consistent patterns of asynchronies within their own performances than with those of other performers (Table 2.1b, left column), an observation which replicates Repp's (1996b) findings. The within-organist correlations (Table 2.1b, left column, first row), were much lower on average than the intra-subject correlations reported in both Palmer (1989) and Repp (1996b), suggesting that asynchrony patterns may be used less systematically by organists than by pianists. Asynchrony patterns of performances recorded under the same instruction were not more similar than those of performances recorded under different instructions (Table 2.1b, middle column). However, within the performances of individual organists, the mean correlation coefficient for pairs of performances following a given instruction was significantly larger than the mean correlation coefficient with other performances following a different instruction by the same organist (Table 2.1b, right column). Taken together, these results

[4] Given that each of the eight organists recorded the piece six times, a total of 48 performances was recorded, yielding 1,128 different pairs of performances [(48×47)/2].

indicate that, although some organists may have systematically modified their asynchrony patterns in accordance with the instructions, there was no common strategy among organists.[5]

Table 2.1. Mean correlations coefficients for the onset asynchrony between all pairs of performances.

a)

All performances			
pairs	mean	*SD*	%**
1,128	0. 07	0.12	22.2

b)

	Organists				Voice emphasis				Emphasis within organists			
	pairs	mean	*SD*	%**	pairs	mean	*SD*	%**	pairs	mean	*SD*	%**
Within	120	0.29	0.10	86.7	360	0.06	0.12	19.2	24	0.34	0.10	95.8
Between	1,008	0.04	0.10	14.5	768	0.07	0.13	23.6	96	0.28	0.10	84.4
H_1:$\mu_{within} > \mu_{between}$	U = 115,268, $p < .001$				U = 153,663, $p = .69$				U = 1,511, $p < .01$			

Note. Correlations were calculated on a note-by-note basis for all notes that were part of a chord ($df_{max} = 265$; this number may be reduced for some pairs due to missing notes). (a) Mean correlation coefficient averaged across all pairs of performances. (b) For each comparison group, the mean correlation coefficient was computed within and between groups. One-tailed Mann-Whitney tests were conducted to assess whether the intra-group correlations were significantly higher than the inter-group correlations. %**: percentage of highly significant correlations ($p < .01$). *SD*: standard deviation.

[5] If organists shared a common strategy, the within-instructions mean correlation coefficient would be expected to be significantly higher than the between-instructions coefficient (Table 2.1b, middle column).

A dissimilarity matrix, computed from the correlation matrix summarized by Table 2.1, was used to generate a three-dimensional multidimensional scaling representation of the distance between performances on the basis of their asynchrony profiles (Figure 2.4). A strong correlation was found between coordinates on the first dimension and the differential between mean asynchronies of the uppermost voices (soprano and alto) and of the lower voices (tenor and bass), $r(46) = 0.94$, $p < .001$. Organists who were prone to lead with the left hand, such as O2, O3, and O6 (see Figure 2.3), had the lowest coordinates on this dimension, while organists who led with the right hand (O4 and O5) had the highest coordinates. High values on the second dimension appeared to be linked to the presence of some large asynchronies associated with specific events in the score; this was the case for O6 and O7. The graph shows that whereas the performances of individual organists were generally grouped together, performances emphasizing the same voice did not show a tendency to be clustered together.

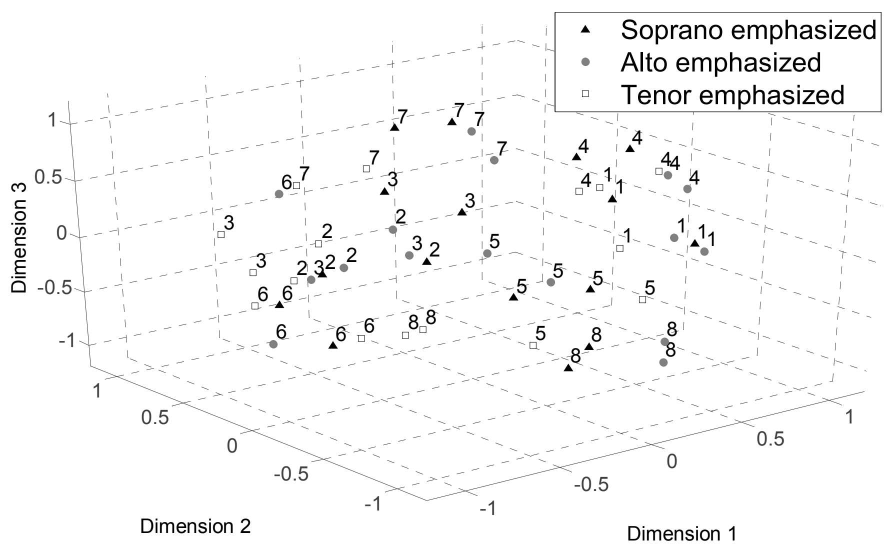


Figure 2.4. Multidimensional scaling of the distances between all performances, based on the note-by-note onset asynchrony correlations computed between all pairs of performances (monotonic regression; Kruskal stress-I = 0.20; RSQ = 0.63). Numbers identify individual organists. Each symbol with its accompanying number identifies a single performance.

On the one hand, the results reported here support the hypothesis that organists may use onset asynchrony as an expressive parameter for specific voice emphasis: onsets in the emphasized voice occurred a few milliseconds earlier on average than those of nominally simultaneous notes, and the location of the emphasized voice influenced asynchronies in the upper voices. On the other hand, these asynchronies were much smaller than those observed in piano performance, and most of them did not differ significantly from zero. The minimum difference in onset times for listeners to be able to discriminate between onsets is generally considered to be at least 10 to 20 ms (Hirsh, 1959), which suggests that most asynchronies observed in this experiment were likely to be too small to be detected.

As Repp (1996b) and Goebl (2001) have shown, melody leads in piano performance appear to be correlated with, and may in fact be caused by, dynamic differentiation between voices. These findings may account for the lack of large expressive melody leads or chord asynchronies in organ performance, given that dynamic differentiation between simultaneous note-events is not possible on this instrument.[6] Therefore, the present study appears to validate Repp's and Goebl's explanations of the melody lead in piano performance as a velocity artifact. The slight tendency for the emphasized voice to lead could be a residual of the organists' training as pianists, since the emphasized voice or melody tends to be played louder than the accompanying voices on the piano, and the sound production mechanism may be activated earlier due to faster finger speed (Goebl, 2001; Palmer, 1996; Repp, 1996b). Indeed, as previously mentioned, organists who claimed to play the piano frequently exhibited small but consistently positive melody leads.

In contrast to piano tones, which are characterized by a short rise time followed by a decay (Palmer & Brown, 1991), organ tones typically reach peak amplitude 50 to 100 ms after note onset and maintain a quasi-constant intensity while the key is pressed (Braasch & Ahrens, 2000). Thus, onset asynchronies on the order of those observed in this experiment may not affect the acoustic signal

[6] It is assumed here that all simultaneous notes are played with the same combination of stops, as was the case in this experiment. The use of the crescendo pedal, while allowing dynamic differentiation over time, cannot be used to differentiate the dynamic levels of simultaneous note onsets as can be done on the piano.

to a great extent in organ performance. Furthermore, since organ pipes may be located several meters away from the console and be quite distant from each other, the sound production mechanism of the organ itself may create large asynchronies. This causes differential delays both in the transmission from the console to the pipes and in the time required from the sound to travel back from the pipes to the organist or to the audience. Therefore, the asynchronies observed at key-bottom contact should not be equated with those perceived when listening to the sound output. An organist using note onset asynchrony as an expressive device would have to take into account those delays, which can create asynchronies that are probably much larger than those measured at key-bottom contact. Taken together, these observations suggest that onset asynchrony might not be an efficient expressive device in organ performance. However, a more exhaustive study of the use of onset asynchrony as an expressive strategy in organ performance should sample a larger musical repertoire.

Local tempo variations

In a study on piano performance, Palmer (1989) reported that the amount of deviation from the mean tempo of a performance was generally more important in musical performances than in non-musical ones, indicating that local tempo deviations were an important aspect of musical expressivity. However, data from a later study (Palmer, 1996) suggest that, in piano performance at least, local tempo deviations do not seem to play an important role as an expressive parameter used to contrast different melodic interpretations. In order to assess whether tempo variations play a role in the differentiation between principal voice (melody) and secondary voices in organ performance, a comparison of means for the overall amount of deviation from the mean tempo for each voice across different interpretations was undertaken, followed by a comparison of the note-by-note local tempo patterns for each performance.

A commonly used measure of the amount of tempo deviation is the standard deviation of the local tempo, expressed in percentage of the mean tempo, which gives a measure of overall spread (Bengtsson & Gabrielsson, 1983; Gabrielsson, 1987; Palmer, 1996). In this study, the mean tempo for each performance was defined as the amount of time from the average onset time of the initial chord and the average onset time of the final chord, divided by the number of half-notes in between those two chords; the half-note was chosen as unit since the piece is in cut time (2/2 meter). For each note n, the local tempo was determined by computing the difference in onset time between n and the next note belonging to the same voice $n+1$, and dividing the value by the ratio of the nominal duration of n to that of a half note. Local tempi for notes followed by a rest in the same voice were not determined. Finally, the local deviation from the mean tempo was expressed as a percentage of the mean tempo.

Figure 2.5 shows the standard deviation of the local tempo for each voice/emphasis combination, averaged across all organists. While the standard deviations of the local tempo for Voices 1 and 2 were virtually identical to each other across all interpretations, Voices 3 and 4 showed markedly lower values, indicating that the organists played these voices with smaller tempo variations. The comparatively lower values observed for Voice 4, which contains mostly half-notes, may also reflect the fact that local tempo deviations are not proportional to note duration. Indeed, a correlation of -0.20 ($p < .001$, $n = 14,422$) was observed between nominal note duration (i.e., quarter note, half-note, etc…) and absolute percentage of deviation from mean tempo. Furthermore, a correlation of -0.85 ($p < .05$, $n = 6$) was found between the standard deviation of the local tempo and nominal note duration[7], indicating that the spread of the local tempo variations was in fact almost inversely proportional to nominal note duration when expressed as a percentage of deviation from the mean tempo. These observations provide a plausible explanation for the lower values of standard deviation of local tempo observed for Voice 4.

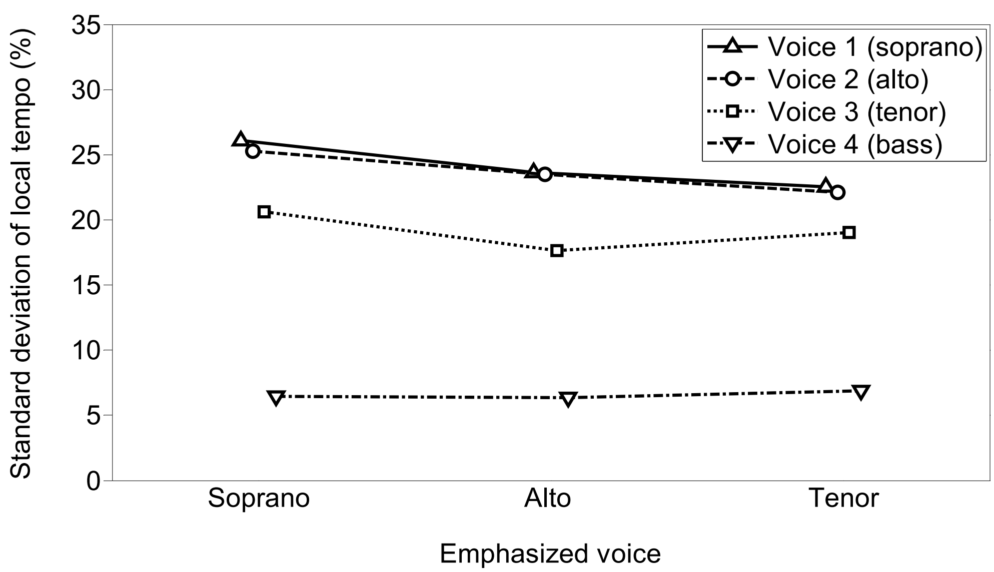


Figure 2.5. Standard deviation of the local tempo, averaged across organists. Values expressed as percentage of the mean tempo for all voice/emphasis combinations (excluding Voices 5 and 6).

[7] The precise repartition per category was as follows: 189 sixteenths, 6,647 eighths, 3,424 quarters, 468 dotted quarters, 3,363 half-notes, and 331 dotted half-notes, for a total of 14,422 notes.

A mixed-model repeated-measures ANOVA conducted on the standard deviation of local tempo in each performance, with voice emphasis and voice (1-4) as within-subject factors, showed a significant effect of voice, $F(3, 21) = 259.29$, $p < .001$, as well as a significant effect of voice emphasis, $F(2, 14) = 6.24$, $p < .05$. Post-hoc tests (Tukey-HSD) confirmed that the standard deviation of the local tempo was larger when the soprano was emphasized than in the other conditions. There was no significant interaction between voice and emphasis. Since the distinct rhythmic content of Voice 4 probably accounted for its smaller standard deviation of local tempo, a mixed-model repeated-measures ANOVA was also conducted on the standard deviation of local tempo for the upper three voices only, with voice emphasis and voice (1-3) as within-subject factors. Again, significant effects of voice, $F(2, 14) = 27.05$, $p < .001$, and emphasis, $F(2, 14) = 7.98$, $p < .01$, were observed. These results indicate that there were differences in the amount of local tempo variation for each voice, with the voices played by the right hand (Voices 1 and 2) performed with larger tempo variations than the voices belonging to the left hand (Voices 3 and 4) across all instructions. Furthermore, a greater amount of tempo variation was applied when the soprano was emphasized. These results are consistent with previous observations regarding right-handed keyboardists' tendency to prefer to use rubato in the right hand (Peters, 1985). However, the present data provide no clear indication that organists modulated local tempo variations in order to emphasize a given voice.

Local tempo patterns were also compared on a note-by-note basis by computing correlations for every note for which local tempo could be determined between all pairs of performances (Table 2.2).[8] The correlation coefficients for rubato patterns were much higher than those observed for asynchrony patterns. Indeed, all 1128 pairwise correlations between the 48 performances were highly significant ($p < .01$), suggesting a strong general agreement among organists (Table 2.2a). As with asynchrony patterns, the group comparisons show that organists exhibited idiosyncratic tempo patterns that differentiated their performances from those of other performers (Table 2.2b, left column). The within-organist correlations (Table 2.2b, left column, first row) were comparable to the intra-subject correlations reported in Palmer (1989). Although the temporal patterns of performances emphasizing the same voice were not significantly more correlated than those of performances emphasizing different voices (Table 2.2b, middle column), the mean correlation for pairs of performances recorded by the same organist and emphasizing the same voice was significantly larger than the mean correlation with other performances by the same organist emphasizing a different voice (Table 2.2b, right column). As with asynchrony patterns, these results indicate that while organists (or at least some of them) systematically modified their local tempo patterns in accordance with the voice emphasized, there was no common strategy used by different organists to emphasize a specific voice by means of variations in local tempo patterns.

[8] The percentage of deviation from the mean tempo was used for these correlations.

Table 2.2. Mean correlation coefficients for the local tempo patterns between each pair of performances.

a)

All performances			
pairs	mean	*SD*	%**
1,128	0. 69	0.16	100.0

b)

	Organists				Voice emphasis				Emphasis within organists			
	pairs	mean	*SD*	%**	pairs	mean	*SD*	%**	pairs	mean	*SD*	%**
Within	120	0.84	0.07	100.0	360	0.69	0.16	100.0	24	0.88	0.06	100.0
Between	1,008	0.67	0.16	100.0	768	0.69	0.16	100.0	96	0.84	0.07	100.0
$H_1{:}\mu_{within} > \mu_{between}$	U = 105,727, $p < .001$				U = 140,123, $p = .36$				U = 1,556, $p < 0.01$			

Note. Correlations were calculated on a note-by-note basis for all notes for which the local tempo could be computed ($df_{max} = 308$; this number may be reduced for some pairs due to missing notes). (a) Mean correlation coefficient averaged across all pairs of performances. (b) For each comparison group, the mean correlation coefficient was computed within and between groups. One-tailed Mann-Whitney tests were conducted to assess whether the intra-group correlations were significantly higher than the inter-group correlations. %**: percentage of highly significant correlations ($p < .01$). *SD*: standard deviation.

A dissimilarity matrix, computed from the correlation matrix summarized in Table 2.2, was used to generate a multidimensional scaling representation of the distance between performances on the basis of their local tempo profiles. A one-dimensional solution (not shown), provided a good fit (monotonic regression, Stress-I = 0.17, *RSQ* = 0.96). The main clustering was observed between the performances of Organist 2, who played the piece using *notes inégales*, and those of other organists, who did not.[9] However, there was no tendency for performances following a given instruction to be grouped together.

These observations suggest that changes in either the amount of tempo variation or the note-by-note local tempo patterns play only a minor role in the differentiation between principal voice and secondary voices in polyphonic organ music. By and large, these results corroborate Palmer's (1996) observations regarding variations in the range of local tempo patterns across different melodic interpretations, although no significant effect of voice or emphasis was reported in that study, in contrast to what was observed here. Further studies, perhaps involving a greater number of performers and a larger variety of musical excerpts, would be necessary in order to describe precisely the changes in temporal patterns that may be employed by some organists to differentiate between melodic interpretations.

This study also highlights the need for developing a measure of deviation from the mean tempo that could be used to compare the amount of tempo variation in melodies or voices that contain different rhythmic material. Many researchers use the percentage of deviation from a performer's mean tempo when comparing across performers (Gabrielsson 1987, Palmer 1989). However, the strong correlation observed between nominal note duration and the standard deviation of the local tempo deviation suggests that a more refined measurement

[9] The term "*notes inégales*" refers to a rubato style typical of French Baroque music (and thus appropriate for the piece performed in this experiment), in which eighth-notes on weak beats are shortened, whereas eighth-notes on strong beats are lengthened.

of deviation from the mean tempo should be developed if valid comparisons between melodies or voices are to be made.

Articulation

Articulation refers to the amount of overlap between two consecutive note events belonging to the same voice. When the offset of note *n* occurs after the onset of note *n+1*, the articulation is defined as legato, and the overlap is positive. When the offset of note *n* precedes the onset of note *n+1*, the articulation is defined as staccato, and the overlap is negative. The offset of a note was defined as the time at which a key was released (as measured by the MIDI system) and the onset was the time at which a key was pressed. When the same key was struck twice in succession, regardless of whether the consecutive note-events belonged to the same voice or to two different voices, the amount of overlap was not computed, because the performer must physically release the key in order to play it again, necessarily causing a negative overlap (Palmer, 1989); there are 16 such instances in the score.

As with onset asynchronies and local tempo patterns, articulation is an important expressive dimension of music performance. Palmer (1989) observed that, in piano performance, melody notes were performed in a more legato manner (that is, with a larger mean overlap) in musical performances than in unmusical performances of the same piece. A subsequent study of the effect of melody interpretation on articulation reported significant effects of both intended melody (larger overlaps for lower melody interpretations) and voice (larger overlaps for upper voices across different melody interpretations), but no significant interaction between intended melody and voice (Palmer, 1996).

As Figure 2.6 shows, the situation is somewhat different for organ performance: whereas Voice 4 was played more staccato than the other voices across all instructions, the mean overlap of each of the upper voices was lower (greater negative values) when it was emphasized than when it was not. In other words, a voice was played in a more detached manner when emphasized.

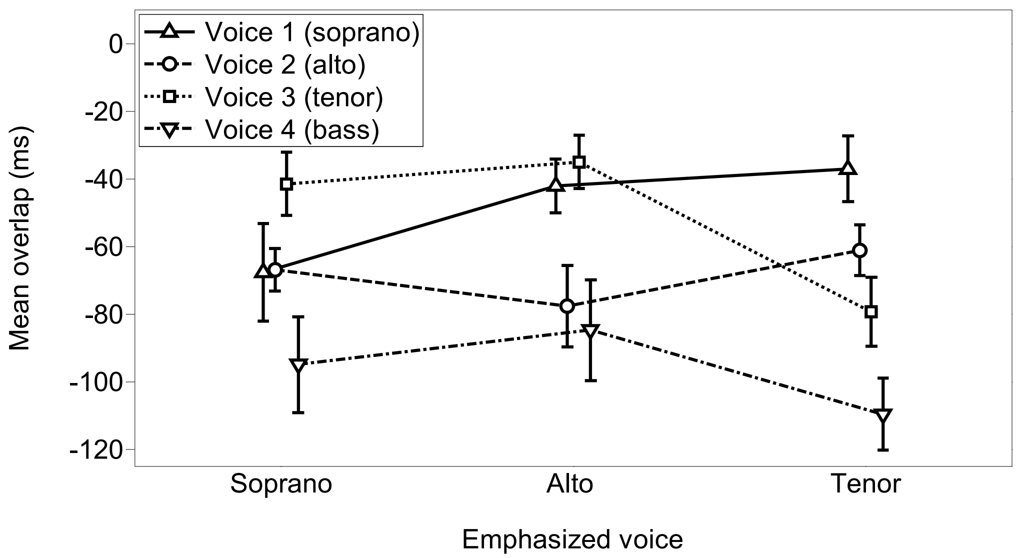


Figure 2.6. Mean overlap for all voice/emphasis combinations (excluding Voices 5 and 6). Values given in milliseconds and averaged across organists. Error bars represent standard errors of the mean.

A mixed-model repeated-measures ANOVA conducted on the overlap with voice emphasis and voice (1-4) as within-subject factors showed a significant effect of voice, $F(3, 21) = 25.68$, $p < .001$, as well as a significant interaction between emphasis and voice, $F(6, 42) = 5.56$, $p < .001$. No other significant main effect or interaction was observed.[10] The presence of an interaction between emphasis and voice, combined with the absence of a main effect of voice emphasis, indicates that although the mean overlap averaged across all voices did not differ significantly with respect to voice emphasis, it varied for specific voices with respect to the voice emphasis.

Articulation patterns were also compared by computing correlations for overlap on a note-by-note basis for each note for which the amount of overlap could be determined between all pairs of performances (Table 2.3). Although the correlation coefficients observed for articulation patterns were lower than those recorded for local tempo patterns, a large proportion (83.1%) of all pairwise correlations was highly significant (Table 2.3a). As with local tempo patterns, this indicates a fairly strong agreement between organists. The comparisons again showed that organists exhibited idiosyncratic articulation patterns that differentiated their performances from those of other performers (Table 2.3b, left column). While lower than the intra-subject correlations for overlap reported in Palmer (1989), the within-organist correlations (Table 2.3b, left column, first row) were nevertheless fairly high, with 90.8% of highly significant correlations. In contrast to what was observed with asynchrony and local tempo patterns, performances emphasizing the same voice were significantly more similar to each other than to performances emphasizing different voices, indicating that there was

[10] Since the different rhythmic content of Voice 4 may explain its lower overlap values, a mixed-model repeated-measures ANOVA was conducted on the overlap for the upper three voices with voice emphasis and voice (1-3) as within-subject factors. Again, a significant effect of voice, $F(2, 14) = 12.32$, $p < .001$, and a significant interaction between emphasis and voice, $F(4, 28) = 15.83$, $p < .001$, were observed. No other significant main effect or interaction was observed.

a systematic shift across organists in the articulation patterns according to the voice emphasized (Table 2.3b, middle column). Not surprisingly, the same phenomenon was also observed when comparing performances by an individual organist emphasizing the same voice with other performances emphasizing different voices (Table 2.3b, right column). These results indicate that, not only did organists systematically alter their articulation patterns with respect to voice emphasis, but also, and more importantly, that different organists followed a common strategy in their use of articulation patterns to emphasize a specific voice.

A dissimilarity matrix, computed from the correlation matrix summarized in Table 2.3, was used to generate a multidimensional scaling solution representing the distance between performances on the basis of their articulation profiles; a two-dimensional solution provided a good fit (Figure 2.7). The first dimension was related to the contrast in articulation between hands: a strong correlation was found between coordinates on the first dimension and the differential between mean overlap of the uppermost voices (soprano and alto) and of the lower voices (tenor and bass), $r(46) = .91$, $p < .001$. Performances in which the left hand was more detached than the right hand can be found on the left side of the graph, and performances where the right hand was more detached than the left are located on the right side. Coordinates on the second dimension were correlated with the mean overlap differential between the upper voices (alto and soprano), $r(46) = .74$, $p < .001$. Performances found in the upper part of the graph showed little or no contrast in articulation between the two upper voices, whereas the alto was played significantly more detached than the soprano in the performances located in the lower part. Similarly to what was observed for asynchrony and local tempo patterns, there was a tendency for performances recorded by the same organist to be clustered together. However, a more prominent tendency was for performances emphasizing the tenor part to be grouped on the left side of the graph, while performances emphasizing the soprano or alto voices were mostly located on the right side. Furthermore, within each organist's performances, performances emphasizing the tenor were likely to be located to the left of performances emphasizing the alto or soprano. This reflects the fact that many organists shared a common strategy regarding articulation patterns, as discussed above. Whereas some organists, such as O3, exhibited extreme systematic contrast in articulation patterns between different interpretations, other performers such as O6 did not show any systematic trend. As can be seen in Figure 2.7, the majority of organists did not differentiate much between the soprano- and alto-emphasizing performances; the main contrast was between the tenor-emphasizing performances and those emphasizing one of the upper voices. Since the upper voices were played by the right hand while the tenor voice was mostly under the control of the left hand, this suggests a within/between-hands effect on the ability to contrast voices on the basis of articulation patterns.

Table 2.3. Mean correlation coefficients for the articulation patterns between each pair of performances.

a)

All performances			
pairs	mean	*SD*	%**
1,128	0. 30	0.15	83.1

b)

	Organists				Voice emphasis				Emphasis within organists			
	pairs	mean	*SD*	%**	pairs	mean	*SD*	%**	pairs	mean	*SD*	%**
Within	120	0.54	0.19	90.8	360	0.34	0.14	92.7	24	0.68	0.11	100.0
Between	1,008	0.27	0.12	82.1	768	0.28	0.15	78.5	96	0.50	0.19	88.5
H_1:$\mu_{within} > \mu_{between}$	U = 106,435, $p < .001$				U = 171,311, $p < .001$				U = 1,857, $p < .001$			

Note. Correlations were calculated on a note-by-note basis for all notes for which overlap could be computed (df_{max} = 287; this number may be reduced for some pairs due to missing notes). (a) Mean correlation coefficient averaged across all pairs of performances. (b) For each comparison group, the mean correlation coefficient was computed within and between groups. One-tailed Mann-Whitney tests were conducted to assess whether the intra-group correlations were significantly higher than the inter-group correlations. %**: percentage of highly significant correlations ($p < .01$). *SD*: standard deviation.

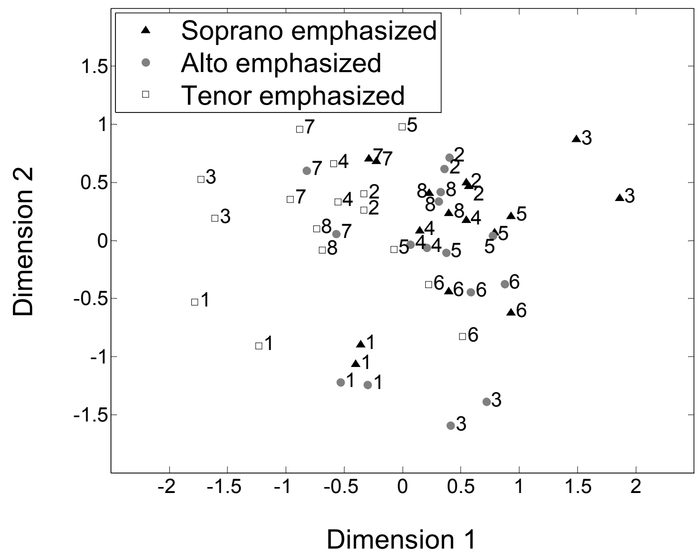


Figure 2.7. Multidimensional scaling of the distances between all performances, based on the note-by-note overlap correlation coefficients computed between all pairs of performances (monotonic regression; Kruskal stress-I = 0.23; RSQ = 0.76). Numbers identify individual organists. Each symbol with its accompanying number identifies a single performance.

The observation that the emphasized voice is played more staccato than the secondary voices may be somewhat unexpected, considering that pianists play the melody notes more legato in musical performances compared to unmusical ones (Palmer, 1989**).** However, while the piano is a percussive instrument, the organ is essentially a wind instrument controlled by a keyboard, which implies that each instrument may favor a different articulation strategy. Because the organ sound is continuous, short rests created by a more staccato articulation may emphasize the next note attack, thereby lending more emphasis to that note. Indeed, Drake and Palmer (1993) reported that the largest negative overlaps (longer silences between notes) preceded notes in a strong metrical position;

likewise, in complex musical structures, events before and on melodic jumps were played more staccato. These observations suggest that a large negative overlap may indeed function as a kind of accent that increases the salience of the next note. It is very likely that these types of accents are used more consistently by organists than by pianists: because the organ does not allow variations in intensity, organists must be able to convey all elements of expressive performance by exclusively manipulating parameters related to inter-onset or offset-to-onset timing. In that regard, it would be interesting to compare articulation strategies employed by organists with those used by performers of wind instruments that allow only limited dynamic differentiation, such as the recorder.

Discussion

This study sought to identify the expressive means used by organists to emphasize a specific voice in a polyphonic organ piece. Three parameters were analyzed: note onset asynchrony, local tempo variations, and articulation (note overlap). Although significant differences in onset asynchronies were observed across voice/emphasis combinations, it is unclear how these differences could be perceptible given their small scale. Moreover, comparisons with piano performance studies suggest that these differences may be a residual of the organists' training as pianists rather than a conscious expressive strategy. Variations in the spread of local tempo deviations were observed across voices and interpretations, but there was no interaction between voice and emphasis which would indicate an attempt to differentiate between voices according to melodic emphasis. Variations in the amount of overlap appear to be the most widespread and consistent strategy used by organists to emphasize a voice, at least in the experiment described here. Specifically, a voice was played in a more detached manner when it was emphasized than when it was not.

As mentioned by other researchers (Goebl, 2001; Palmer, 1996), the choice of repertoire, with its associated performance styles and typical textures, may also affect performers' use of expressive parameters. For instance, Romantic music would typically be performed with larger onset asynchronies than other musical styles (cf. Methuen-Campbell, 1992). Thus, the lack of large asynchronies observed in this experiment might also be related, at least in part, to the style of the musical excerpt that was performed. This question can only be answered by sampling a larger repertoire of musical styles.

Regarding performance issues, this experiment also demonstrated that most performers showed a well-developed aptitude to immediately modify their interpretation of an unfamiliar musical excerpt following specific instructions. It may be that their task was actually made easier by the fact that the score had not been practiced and overlearned yet; indeed, this interpretative flexibility seems to decrease once the performer has settled on a particular reading of the piece (Palmer, 1996).

Although the experiment did not specifically address the issue of fingering, two of the organists mentioned consciously changing their fingerings when emphasizing different voices. It is likely that other organists may have modified their fingerings, whether consciously or not. Further studies would be necessary to clarify whether performers systematically alter their fingering patterns according to which voice is emphasized and to determine the role of motor patterns in changing melodic interpretations.

Although this study has shown which expressive parameters were manipulated by organists to emphasize a specific voice, it has not addressed the question of whether these manipulations were successful, that is, whether listeners could actually identify which voice was being emphasized. Experiment 2 aimed to answer this question.

EXPERIMENT 2: PERCEPTION OF VOICE EMPHASIS

The perception of voice emphasis in polyphonic organ music was investigated by inviting participants to listen to a representative selection of the recordings collected in Experiment 1 and rate the relative prominence of the three upper voices. The aim of this experiment was not only to assess the efficiency of the performers' expressive strategies, but also to evaluate the relative contribution of the musical structure of the piece and of the expressive intent of the performer in the formation of a percept of relative voice prominence. Thus, listeners rated the relative prominence of the voices using a continuous response method, which allowed us to probe their response to specific musical events in the piece. In addition, a completely "deadpan", computer-controlled performance of the piece was recorded on the same organ as an experimental control in an attempt to discriminate further between effects related to musical structure and effects of expressive performance. Finally, as mentioned previously, listener instrumental expertise has been shown to influence the perception of melodic emphasis

(Palmer, 1996); for this reason, both organists and non-organists were recruited for this experiment in order to take this effect into account.

Recordings were selected on the basis of the analysis of the data obtained in Experiment 1. Given that articulation was identified as the main expressive parameter used to emphasize different voices, organists were selected mainly according to the degree of contrast in articulation between their different interpretations. The performances of Organist 3 exhibited a strong contrast between the interpretations emphasizing the soprano, alto, and tenor parts, as shown in Figure 2.8 (see also the multidimensional scaling representation in Figure 2.7). Organist 4, who differentiated mostly between the tenor-emphasizing performances and the soprano- and alto-emphasizing ones, was categorized as a moderate contrast performer, whereas Organist 6, whose interpretations could not be clearly differentiated on the basis of articulation, was identified as a weak contrast performer. It was hypothesized that differences in the perceptual prominence of the voices would be greater between interpretations of Organist 3 than between those of Organists 4 and 6.

Method

Participants

Since the experiment required an explicit understanding of the structure of polyphonic music, only listeners with university-level musical training were selected. Two groups of participants were recruited: 20 non-organists (music students having completed at least one year of undergraduate studies), recruited from the McGill and University of Montreal campuses, and 10 organists from the

Montreal area, who were either enrolled in or had previously completed a degree in organ performance. None of the organists whose recordings were selected for this experiment were invited. Participants were given $10 as compensation for their time. The mean age of the participants was 24 years for the non-organists (range: 20 to 33 years) and 25 years for the organists (range: 20 to 31 years).

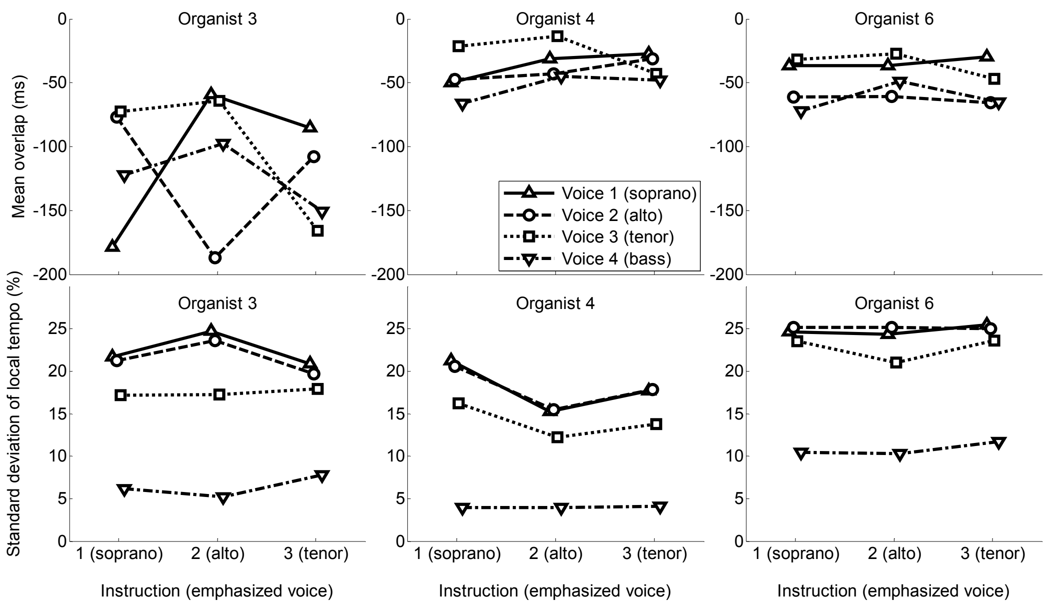


Figure 2.8. Mean overlap (in milliseconds) and standard deviation of local tempo for the organists whose recordings were selected for Experiment 2. Values expressed as percentage of the mean tempo for all voice/emphasis combinations (excluding Voices 5 and 6).

Materials

The main phase of the experiment employed 10 performances. The performances of Organists 3, 4, and 6 from Experiment 1 were selected for this study. Three performances, each emphasizing a different voice, were chosen for each organist. Since performers were asked to record two versions for each interpretation, the recording with the fewest number of errors was selected. A mechanical, computer-controlled performance recorded on the same organ and using the same registration was added as an experimental control. The tempo selected for this performance was the average tempo of all the performances recorded in Experiment 1; note-to-note overlap values and onset asynchronies were set to 0 ms for all notes and no local tempo deviations were implemented. Two recordings from Organist 1, each emphasizing a different voice, were used in the trial phase of the experiment.

Procedure

Participants were asked to rate the relative prominence of the three upper voices (soprano, alto, tenor), ignoring the bass part, while listening to recordings of the performances. The computer interface, programmed into PsiExp (Smith, 1995), consisted of a screen with a triangle whose vertices were marked “SOP”, “ALT”, “TEN”, for soprano, alto, and tenor, respectively (Figure 2.9). A cursor, located at the center of the triangle at the beginning of the performance, could be moved around the triangle simply by moving the mouse (it was not necessary to click). The relative font size of the letters in the “SOP”, “ALT”, and “TEN” markings varied as the cursor was moved in the triangle, indicating the relative prominence of the respective voices. When participants felt that a voice was becoming prominent, they could move the cursor toward the vertex corresponding to that voice. Participants were warned that some performances may contain errors, and they were asked to not take them into account as much as possible.

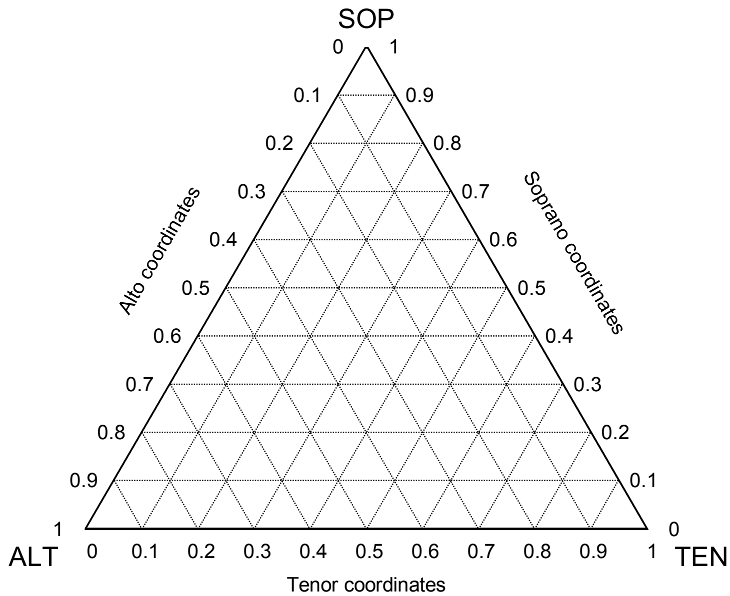


Figure 2.9. The triangle used by listeners to rate the relative prominence of the upper voices and its system of ternary coordinates.

The experiment was divided into two parts. The first part began with a silent trial run, during which the experimenter was there to answer any questions, followed by two recordings of the *Premier Agnus* by Organist 1. These performances were used as a practice run during which participants were familiarized with the use of the interface. The data from this experimental phase was not analyzed. In the second phase of the experiment, participants heard ten recordings of the *Premier Agnus*: three each from Organists 3, 4, and 6, and a recording of a mechanical performance. The order of the performances was randomized. Participants were provided with a score of the piece.

The experiment took place in a sound-attenuated booth on an Apple McIntosh G5 computer. Participants wore Sennheiser HD 280 Pro headphones (diotic listening). The loudness level was set at 70 dB. All participants first passed an audiogram to ensure that they had normal hearing. After having familiarized themselves with the experimental interface and completed the training phase, they proceeded with the main phase of the experiment. Once the experiment was completed, participants filled out a questionnaire. The entire experiment lasted approximately 1 hour. For each participant, a log file that recorded the coordinates of the cursor in the triangle continuously over time was produced for each performance.

Results

Coordinates for each voice were obtained by mapping the position of the cursor in the triangle used to evaluate the relative prominence of the upper voices onto a system of ternary coordinates, as shown in Figure 2.9. Coordinates in this system have the following properties: they are bound between 0 and 1 for each individual axis (or voice in our case), and the sum of the coordinates on all three axes for any point in the triangle is equal to 1.

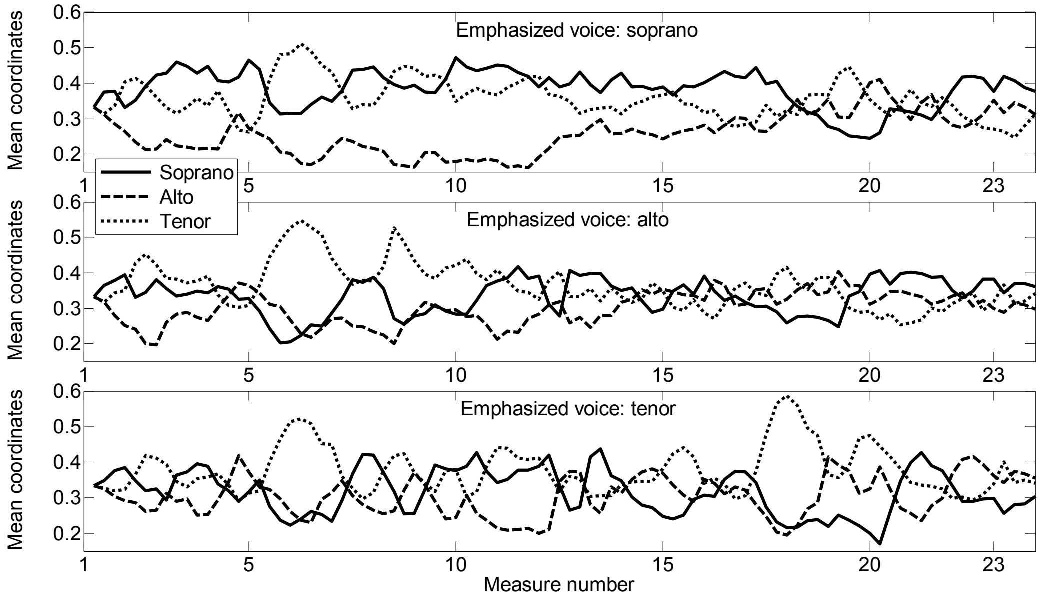
Since the tempi varied between different performances of the piece, the prominence rating profiles needed to be aligned temporally in order to compare profiles across performances. We used the matched score of the MIDI data of the performances to establish a correspondence between MIDI events and score events in the piece. Coordinates were then averaged over each quarter note of the score.

The data collected in this experiment poses several analytical challenges: on the one hand, continuous ratings are not easily amenable to traditional statistical analysis; moreover, the values collected for all the voices are strongly interdependent since they all sum up to 1 for any point in time. For these reasons, statistical analyses will be conducted separately for each voice, and on the mean coordinates averaged over entire performances; the analysis of the continuous ratings will remain descriptive.

Continuous voice prominence ratings

Figure 2.10 shows the continuous ratings averaged over all performances which emphasized a specific voice; separate graphs are given for organists and non-organists. A number of peaks were observed consistently across all interpretations; for instance, the tenor part (dotted curve) reached a high point around measures 6-7 and a secondary one around m. 9, whereas the soprano (solid curve) reached a local climax around m. 8. Although the overall contour of the voice prominence profiles was relatively constant across all interpretations for non-organists, some peaks were specific to an interpretation for the organists: note for instance the high peak in the tenor coordinates in m. 18 in the interpretations emphasizing the tenor voice. The general contour similarity and the presence of invariant peaks suggests that the listeners' perception of prominence was determined in large part by the musical structure.

a) Organists



b) Non-organists

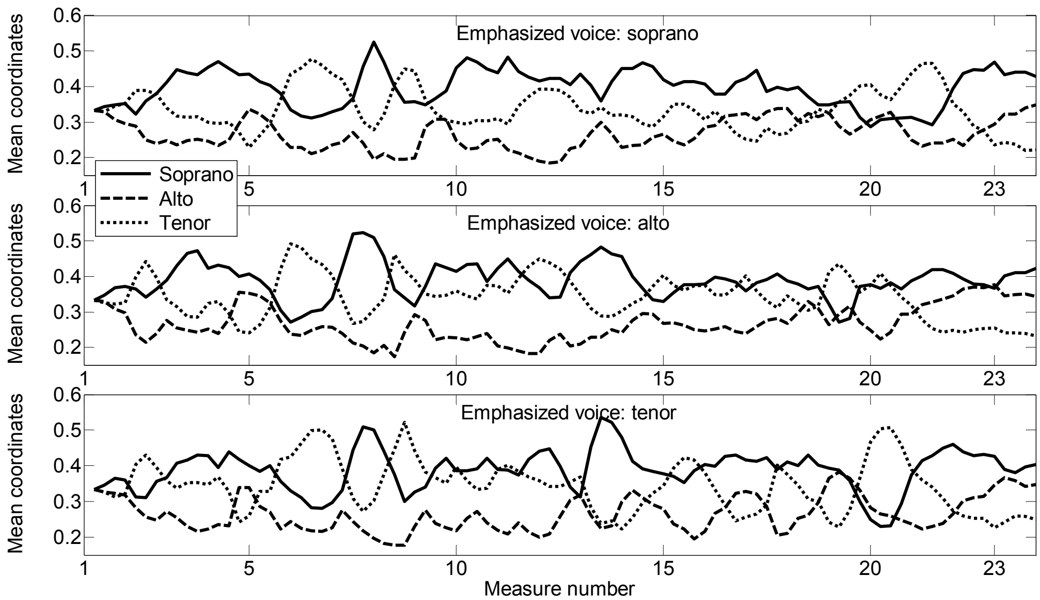


Figure 2.10. Coordinates for the relative prominence of the soprano, alto, and tenor voices of the *Premier Agnus* averaged over all performances emphasizing a specific voice. a) Organists; b) Non-organists.

In order to focus on the role of musical structure in the perception of voice prominence, we analyzed the prominence profiles for the mechanical performance, where, presumably, the only factors influencing the participants' ratings were related to music-structural considerations (Figure 2.11). Major peaks were observed for both organists and non-organists around m. 6 (tenor), m. 9 (tenor), mm. 17-18 (tenor), m. 19 (alto), and m. 20 (tenor). In addition, we observed clear peaks for the soprano around mm. 13 and 21 in the organists' ratings. An examination of the score reveals that most of these peaks correspond to instances where a melodic passage in the voice rated as prominent is scarcely interrupted by melodic activity in other voices (see for instance, the tenor part in m. 5 and 19, and the alto part in m. 18), thereby generating a figure/ground contrast between an active voice and others which take up an accompanimental role (Figure 2.1). Series of onsets in one voice closely spaced in time also seemed to attract attention; thus, the sixteenth-note runs in the tenor in mm. 8 and 17 were associated with local peaks in the participants' ratings. Finally, the peaks in the soprano voice observed for organists correspond to voice entries after a rest (m. 12 and 20 in the soprano). If we assume that those peaks are indeed related to the musical structures described here, and there is no reason to do otherwise given that the performance was completely mechanical, we may conclude that there was a delay equivalent to approximately one measure before the listeners' response to a particular musical feature of the score reached its maximal value. If we now surmise that this delay was more or less invariant across performances, the musical features mentioned above could also account for the most important peaks observed in the prominence profiles for the expressive performances

(Figure 2.10). Indeed, these peaks also correspond to passages in the score where one voice is structurally salient: the tenor is the most active voice in m. 5, while the soprano is active in a high register in m. 7, and the tenor enjoys a run of sixteenth-notes in m. 8; finally, it is possible that organists were sensitive to interpretation-specific contrasts that emphasized the run of sixteenth-notes in m.17 in the tenor.

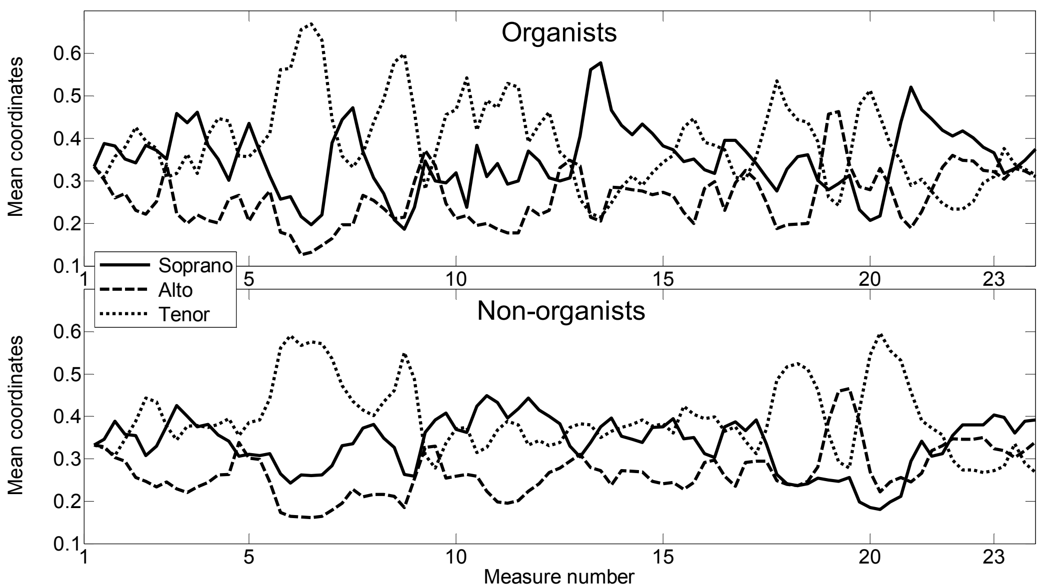


Figure 2.11. Coordinates for the relative prominence of the soprano, alto, and tenor voices of the *Premier Agnus* for the mechanical performance.

Comparison of the mean coordinates across performances

An examination of the mean coordinates averaged over entire performances show that while organists were sensitive to differences between performers and voice emphasis, the mean ratings for non-organists did not vary to a great extent regardless of performer or expressive intent (excluding the mechanical performance): the soprano was nearly always the most prominent voice, and the alto was the least prominent (Figure 2.12).

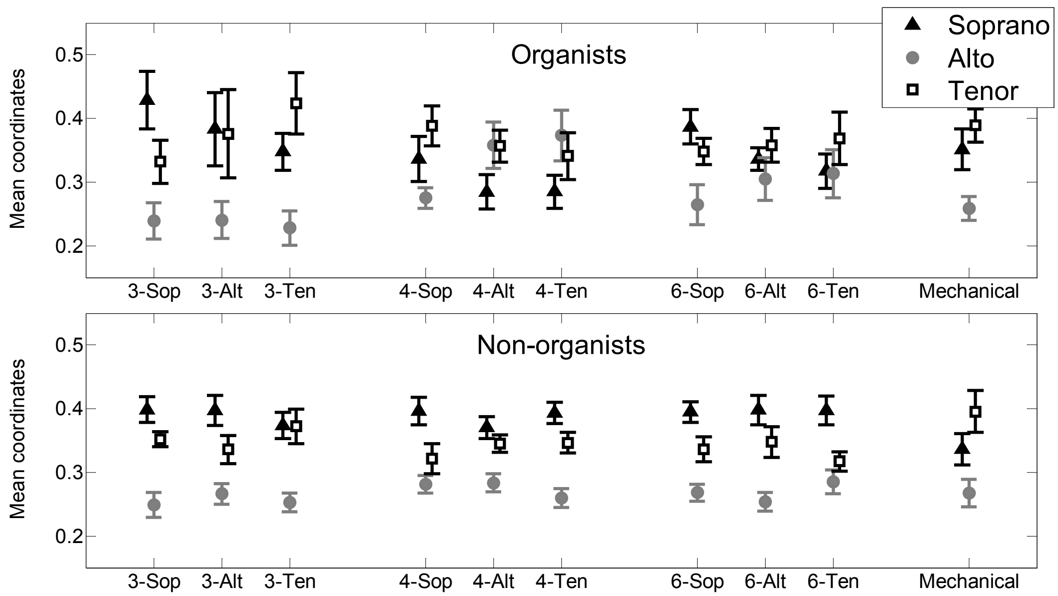


Figure 2.12. Mean coordinates for the relative prominence of the soprano, alto, and tenor voices averaged over entire performances of the *Premier Agnus*. Numbers refer to individual organists; *Sop*, *Alt*, *Ten*: interpretations emphasizing the soprano, alto, and tenor voices, respectively. Error bars represent standard errors of the mean.

Mixed-model repeated-measures ANOVAs on the mean coordinates for the expressive performances were conducted separately for each voice with performer and emphasis as within-subject factors and musical training (organists versus non-organists) as a between-subjects factor. Main effects of performer, $F(2, 56) = 4.57$, $p < .05$, emphasis, $F(2, 56) = 3.62$, $p < .05$, and musical training, $F(1, 28) = 4.57$, $p < .05$, were observed for the soprano, as well as an interaction between musical training and performer, $F(2, 56) = 3.88$, $p < .05$. For the alto voice, a significant effect of performer, $F(2, 56) = 11.94$, $p < .001$, and an interaction between musical training and performer, $F(2, 56) = 5.52$, $p < .01$, were reported. No effect or interaction reached significance for the tenor voice. Although these analyses indicate effects of performer and of expressive intent on the perception of voice prominence, their interpretation is made more difficult because of the presence of interactions between the within-subject factors and the between-subjects factor (musical training). In order to investigate these interactions, mixed-model repeated-measures ANOVAs were conducted separately for each voice and for each group of participants (organists and non-organists) with performer and emphasis as within-subject factors. The results are summarized in Table 2.4. Significant effects of performer were observed for the soprano and alto voices for the organists, as well as a marginally significant effect of emphasis for the soprano part. Post-hoc tests (Tukey HSD) confirmed that the coordinates for the soprano were significantly higher for the performances of Organist 3 than for those of Organist 4, and that the coordinates for the alto were significantly lower for the performances of Organist 3 than for those of Organist 4. No interaction reached significance. These results do not provide clear evidence in favor of our hypothesis that a greater contrast would be observed between the performances of Organist 3 than those of Organist 4 or 6. Indeed, there were differences between performers, but the effects of voice emphasis remained marginal. No effects reached significance for the non-organists.

Table 2.4. Mixed-model repeated-measures analyses of variance on the mean coordinates by voice for the expressive performances, for organists and non-organists.

	Performer	Emphasis	Performer × Emphasis
Organists			
Soprano	$F(2, 18) = 4.47^{*}$ $p = 0.03$	$F(2, 18) = 3.01$ $p = 0.07$	$F(4, 36) = 0.08$ $p = 0.99$
Alto	$F(2, 18) = 8.28^{**}$ $p = 0.003$	$F(2, 18) = 2.08$ $p = 0.15$	$F(4, 36) = 0.89$ $p = 0.48$
Tenor	$F(2, 18) = 0.24$ $p = 0.79$	$F(2, 18) = 0.29$ $p = 0.75$	$F(4, 36) = 0.96$ $p = 0.44$
Non-organists			
Soprano	$F(2, 38) = 0.22$ $p = 0.80$	$F(2, 38) = 0.20$, $p =$ 0.82	$F(4, 76) = 0.48$ $p = 0.75$
Alto	$F(2, 38) = 1.24$ $p = 0.30$	$F(2, 38) = 0.01$ $p = 0.99$	$F(4, 76) = 1.11$ $p = 0.36$
Tenor	$F(2, 38) = 0.89$ $p = 0.42$	$F(2, 38) = 0.18$ $p = 0.84$	$F(4, 76) = 0.99$ $p = 0.42$

Note. $^{*}p < .05$. $^{**}p < .01$.

In light of the fact that non-organists rated the soprano as most prominent in nearly all of the recordings made by human performers, it is interesting to note that they rated the tenor as most prominent in the mechanical performance, in agreement with the organists. Because the mechanical performance was recorded on the same instrument, using the same registration, and at a tempo that corresponded to the average tempo of performances recorded in Experiment 1, it seems unlikely that this effect can be explained by low-level differences in the acoustical signal. Indeed, only one participant (out of 30) mentioned that one of the performances sounded "like it was played by a computer", suggesting that most participants did not notice or, at the very least, were not disturbed by the mechanical character of this performance.[11]

Discussion

The results reported here regarding the perception of voice emphasis in polyphonic organ music lead to several questions. First, the difference in voice prominence between the expressive performances and the mechanical one for the non-organists needs to be accounted for. In light of earlier research which has shown that listeners were most sensitive to changes in the outer voice, and especially in the highest voice, the fact that the soprano was perceived by non-organists as most prominent for nearly all expressive performances was perhaps not unexpected (Brochard et al., 1999; Dewitt & Samuel, 1990; Palmer & Holleran, 1994). However, given the prominence profiles observed for the expressive performances, the soprano would also have been expected to be more prominent in the mechanical performance, which presumably had a "neutral" character in terms of relative salience of the voices. Second, statistical analyses suggested clear differences between the sensitivity of organists and non-organists to different interpretations. Again, these results are consistent with earlier findings regarding the role of listeners' instrumental expertise in the perception of voice emphasis (Palmer, 1996). Yet, the fact that non-organists exhibited a markedly different profile for the mechanical performance suggests that they were sensitive, at least to some extent, to differences between interpretations: although the expressive strategies employed by performers to convey voice emphasis had little

[11] It is likely that most participants were not aware of the possibility to record computer-controlled performances on an organ equipped with a MIDI console.

effect on their perception of voice emphasis, the lack of *any* expressive strategy caused a significant shift in their perception.

These observations suggest that the application of any expressive strategy, regardless of its expressive intent, might have enhanced the relative salience of the soprano voice in comparison to a deadpan rendition. Indeed, if the increased sensitivity to pitch changes and local tempo variations in the outer voices also applies to other musical parameters such as articulation, the presence of articulatory or timing differences between voices could be expected to increase the relative prominence of the outer voices, regardless of the exact nature of these contrasts. This might explain why non-organists rated the soprano voice as more salient in nearly all of the expressive performances. On the other hand, in the absence of expressive contrast between voices, musical features of the score could be expected to play a larger role in the perception of voice prominence. Indeed, we have observed that peaks in the relative prominence of a voice often corresponded to passages where this voice was structurally salient in the score. In the *Premier Agnus*, the tenor voice has a greater number of these passages than the other voices, which would explain why it was perceived as more prominent in the mechanical performance.

However, this model does not account for the differences between organists and non-organists. Given that expressive strategies in organ performance rely to a large extent on timing and articulation contrasts, which may be more perceptually subtle than intensity contrasts, it may be that the recognition of a performer's intentions depends to a certain extent on the explicit knowledge of the different expressive strategies employed by organists. Indeed, as mentioned previously, non-keyboardists were unable to recognize a pianist's expressive intent when only timing cues were available (Palmer, 1996). Thus, although they may perceive differences between interpretations, non-organists may have a relatively undifferentiated understanding of the expressive goals associated with a given strategy: whereas a louder note is unambiguously acoustically emphasized, the intent associated with a staccato articulation may be less definite. On the other hand, the purpose of such expressive strategies may be clearer for practicing organists, who are presumably well acquainted with performance issues related to their instrument. Yet, even organists were not particularly successful at recognizing performers' expressive intentions in the present experiment, although their voice prominence profiles indicate that they differentiated between performers. It may be that it is simply more difficult to produce contrasts between voices on the organ (assuming identical registration for all voices) than on the piano, given the expressive capabilities associated with this instrument.

GENERAL DISCUSSION

It seems plausible to propose that, by creating articulatory contrasts between voices, organists were attempting to create a figure/ground separation in which the non-emphasized voices, played legato, receded into the background, while the note onsets of the emphasized voice, preceded by longer gaps, became more salient. In that view, the detached quality of the emphasized voice would also cause it to stand out from the other voices and call itself to the attention of the listener. The previously quoted study by Drake & Palmer (1993), which suggests that performers may emphasize a metrically strong note, or a melodic feature such as a jump or turn, by playing the preceding note with a staccato articulation, indirectly supports this hypothesis.

However, results from Experiment 2 argue that, in a musical context, voice emphasis through articulatory contrasts is much more difficult to detect than emphasis brought about by dynamic differentiation between voices, as is the case with the piano. Furthermore, in contrast to intensity levels, which represent ecologically relevant differences in acoustic energy, articulatory contrasts do not appear to be objectively valenced; thus, as the differences in the performance of organists and non-organists suggests, the recognition of a performer's intentions may require familiarity with the performance practices associated with a specific instrument.

According to theories of auditory stream segregation (Bregman, 1990; Bregman, Ahad, Crum, & O'Reilly, 2000), a stream is generally more easily perceptually segregated when the offset-to-onset intervals between its constituent tones are minimal, that is, when silent gaps between tones are short or nonexistent. Musically speaking, this suggests that voice segregation would be favored when notes are articulated in a legato manner. However, most studies concerned with auditory stream segregation discuss parameters involved in the perception of one versus two streams in a sequence of alternating high and low tones (the fusion/fission paradigm). To the authors' knowledge, no published study has discussed the role of offset-to-onset intervals on a stream's relative prominence, in a situation where two or more continuous streams are clearly differentiated by frequency, and where stream segregation through timbral or loudness differentiation is impossible. The generalizability of the results presented here should be assessed by applying the methodology outlined in the present study to other pieces, as well as other musical genres. Furthermore, in order to evaluate listeners' abilities to detect contrasts in articulation in a more general context, experiments involving two or more frequency-differentiated streams of either pure tones or periodic sounds, with a mixture of synchronous and asynchronous onsets, and a variable length of offset-to-onset intervals, should be conducted.

The relative salience of a voice in a polyphonic instrumental texture is clearly a complex phenomenon, which is influenced by the listener's familiarity with the instrument, the musical features of the score, the position of the voice, and by performance factors, such as variations in local tempo or in articulation. For the most part, the present study has focused on global changes in expressive parameters, measured by variations in mean values across voice/emphasis combinations, or in the degree of similarity between performances. From a musicological standpoint, it would be interesting to identify which particular notes were affected the most by these expressive changes when comparing one interpretation to another. Further analyses might also attempt to determine what effects, if any, were induced by these expressive changes on the perception of voice prominence by analyzing the continuous prominence ratings of listeners. Such an analysis would allow for a closer examination of the links between performance issues, perceptual constraints, and music-theoretical models.

ACKNOWLEDGMENTS

This research was supported by fellowships from the Social Sciences and Humanities Research Council of Canada and from the Centre for Interdisciplinary Research in Music Media and Technology (CIRMMT) to Bruno Gingras, as well as a grant from the Natural Sciences and Engineering Research Council and a Canada Research Chair awarded to Stephen McAdams. The design of this study benefited from the comments and suggestions of Caroline Palmer. Bennett Smith programmed the computer interface for the experiment on the perception of voice emphasis. The authors also wish to thank Julien Boissinot and Darryl Cameron for their technical assistance, Peter Holmes for permission to use McGill University's sound recording equipment, Jonas Braasch for his advice regarding sound recording, the musical authorities of the Church of St-Andrew & St-Paul (Montreal) for permission to use their Casavant organ, the organists and listeners for their participation, and Al Bregman, Werner Goebl, and Marcelo Wanderley for their helpful comments and suggestions.

REFERENCES

Bengtsson, I., & Gabrielsson, A. (1983). Analysis and synthesis of musical rhythm. In J. Sundberg (Ed.), *Studies of music performance* (pp. 27-60). Stockholm: Royal Swedish Academy of Music.

Braasch, J., & Ahrens, C. (2000). Attack transients of free reed pipes in comparison to striking reed pipes and diapason pipes. *Acustica, 86*, 662-670.

Bregman, A. S. (1990). *Auditory scene analysis: The perceptual organization of sound*. Cambridge, Mass.: MIT Press.

Bregman, A. S., Ahad, P. A., Crum, P. A. C., & O'Reilly, J. (2000). Effects of time intervals and tone durations on auditory stream segregation. *Perception & Psychophysics, 62*(3), 626-636.

Brochard, R., Drake, C., Botte, M. C., & McAdams, S. (1999). Perceptual organization of complex auditory sequences: Effect of number of simultaneous subsequences and frequency separation. *Journal of Experimental Psychology-Human Perception and Performance, 25*(6), 1742-1759.

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.

Drake, C., & Palmer, C. (1993). Accent structures in music performance. *Music Perception, 10*(3), 343-378.

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.

Gabrielsson, A. (1987). Once again: The theme from Mozart's piano sonata in A major (K. 331). In A. Gabrielsson (Ed.), *Action and perception in rhythm and music* (pp. 81-104). Stockholm: Royal Swedish Academy of Music.

Goebl, W. (2001). Melody lead in piano performance: Expressive device or artifact? *Journal of the Acoustical Society of America, 110*(1), 563-572.

Goebl, W., & Bresin, R. (2003). Measurement and reproduction accuracy of computer-controlled grand pianos. *Journal of the Acoustical Society of America, 114*(4), 2273-2283.

Goebl, W., & Parncutt, R. (2002). The influence of relative intensity on the perception of onset asynchronies. In C. Stevens, D. Burnham, G. E. McPherson, E. Schubert & J. Renwick (Eds.), *Proceedings of the 7th International Conference on Music Perception and Cognition*. Sydney, Australia: Adelaide: Causal Productions.

Hirsh, I. J. (1959). Auditory perception of temporal order. *Journal of the Acoustical Society of America, 31*(6), 759–767.

Huron, D. (1989). Voice denumerability in polyphonic music of homogeneous timbres. *Music Perception, 6*(4), 361-382.

Juslin, P. N. (2001). Communicating emotion in music performance: A review and a theoretical framework. In P. N. Juslin & J. Sloboda (Eds.), *Music and emotion: Theory and research* (pp. 309-337). New York: Oxford University Press.

Methuen-Campbell, J. (1992). Chopin in performance. In J. Samson (Ed.), *The Cambridge companion to Chopin*. Cambridge, England: Cambridge University Press.

Palmer, C. (1989). Mapping musical thought to musical performance. *Journal of Experimental Psychology-Human Perception and Performance, 15*(12), 331-346.

Palmer, C. (1996). On the assignment of structure in music performance. *Music Perception, 14*(1), 23-56.

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

Palmer, C., & Brown, J. C. (1991). Investigations in the amplitude of sounded piano tones. *Journal of the Acoustical Society of America, 90*(1), 60-66.

Palmer, C., & Holleran, S. (1994). Harmonic, melodic, and frequency height influences in the perception of multivoiced music. *Perception & Psychophysics, 56*(3), 301-312.

Peters, M. (1985). Performance of a rubato-like task: When two things cannot be done at the same time. *Music Perception, 2*(4), 471-482.

Rasch, R. A. (1979). Synchronization in performed ensemble music. *Acustica, 43*(2), 121-131.

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). Patterns of note onset asynchronies in expressive piano performance. *Journal of the Acoustical Society of America, 100*(6), 3917-3931.

Sloboda, J. A. (2000). Individual differences in music performance. *Trends in Cognitive Sciences, 4*(10), 397-403.

Smith, B. K. (1995). *PsiExp: An environment for psychoacoustic experimentation using the IRCAM musical workstation.* Paper presented at the Society for Music Perception and Cognition, University of California, Berkeley.

Chapter 3. The communication of artistic individuality in organ performance

Although a large body of research has been devoted to the study of communication of expressive intent in music performance, issues relating to the communication and perception of artistic individuality in music performance have been only tangentially addressed in music cognition research. Chapter 3 investigates the communication of artistic individuality by means of a sorting task in which listeners are asked to group together excerpts which they think have been played by the same performer. The first objective of this study is to determine whether participants could perform above chance in this perceptual task. A second objective is to identify the acoustical parameters used by listeners to discriminate between performers. Furthermore, since performers have been asked to record expressive and mechanical interpretations of the chorale setting, this study also seeks to assess the effect of expressive intent on the ability of listeners to identify performers. Finally, effects related to listeners' musical expertise and performers' level of accomplishment are examined.

This chapter is based on the following research article:

Gingras, B., Lagrandeur-Ponce, T., Giordano, B. L., & McAdams, S. The communication of artistic individuality in organ performance. Manuscript prepared for submission to *Perception*.

ABSTRACT

The effects of listener expertise, performer expertise, and expressive intent on the communication of artistic individuality in organ performance were investigated. Six organists, three of whom were prize-winners at national competitions, each recorded two "mechanical" and two expressive interpretations of a chorale setting by Samuel Scheidt (1587-1654). In a subsequent sorting task, 20 non-musicians and 20 musicians listened to these interpretations and grouped together recordings they thought had been played by the same performer. Twenty-eight participants (70%) performed significantly above chance level, demonstrating that most listeners can identify specific performers even on an instrument with a limited range of expressive parameters such as the organ. There was no significant difference in sorting accuracy between musicians and non-musicians. Mean tempo and articulation were found to be the most important dimensions along which listeners differentiated the excerpts. Participants' sorting accuracy was lower for mechanical interpretations than for expressive ones, showing an effect of expressive intent. Sorting accuracy was significantly higher for prize-winning performers than for non-winners, suggesting that the performers' ability to convey a sense of artistic individuality was linked to their level of expertise. Moreover, sorting accuracy was generally better for performers who exhibited either greater consistency or distinctiveness in their recordings.

INTRODUCTION

Certain musicians need only to play a few notes to be unequivocally recognized (Benadon, 2003). They are able to quickly convey a sense of musical individuality through unique and distinctive characteristics of their performance style. However, it is often difficult to identify exactly what musical features allow for such quick and accurate recognition. While issues relating to the communication and perception of artistic individuality in music performance have been only tangentially addressed in music cognition research, the more general problem of the recognition of individuals based on their actions or utterances has motivated a substantial body of research in various related fields.

Studies on the recognition of individuals based on their body movements, in which participants viewed point-light depictions of themselves, their friends or strangers performing various actions, have shown that subjects' visual sensitivity to their own motion was highest (Loula, Prasad, Harber, & Shiffrar, 2005). Subjects performed above chance when asked to identify their friends' actions, but not those of strangers. Moreover, actors were recognized more easily when performing expressive actions, such as boxing or dancing, than expressive actions such as walking.

In the field of speaker recognition, researchers have established the prominent role of features such as fundamental frequency, formant mean, and speech rhythm in the recognition of an individual's voice (Brown, 1981; Holmgren, 1967; Van Dommelen, 1990; Voiers, 1964). Later work has identified voice-selective areas in the human auditory cortex which could be responsible for speaker recognition (Belin, Zatorre, Lafaille, Ahad, & Pike, 2000). Building upon the well-established role of prosodic cues in speech perception, Palmer and her colleagues examined the role of musical prosodic cues (such as variations in amplitude and relative duration) in a discrimination task between familiar and novel performances of the same piece (Palmer, Jungers, & Jusczyk, 2001). Their results, which show that not only adult musicians and non-musicians, but also 10-month-old infants were able to identify correctly the familiar performances, provide evidence that prosodic features of music performances can be stored in memory.

Research on communication in expressive music performance has shown that both musicians and non-musicians can distinguish among different levels of expressiveness in performances of the same piece (Kendall & Carterette, 1990), and that they can recognize the emotions that performers intended to communicate (Juslin, 2000). More recently, Keller and colleagues reported that pianists were able to recognize their own performances reliably and were better at synchronizing themselves with their own pre-recorded performances in a piano duet than with performances from other pianists (Keller, Knoblich, & Repp, 2007). Focusing on the perception of similarity between musical performances, Timmers (2005) found that models based on absolute values of tempo and loudness were better predictors of perceptual distances between performances than models based on normalized variations, and that models based on local tempo features fared better than global models.

Artificial intelligence experts have also attempted to create computational models that could recognize music performers. For instance, Stamatatos &

Widmer (2005) programmed a learning ensemble that achieved a 70% recognition rate, using a database of piano performances of 22 pianists playing two pieces by Chopin. The authors noted that their model displayed a level of accuracy "unlikely to be matched by human listeners".

Although these studies, as well as several others, bear direct relevance on the issue of music performer identification by human listeners, no published study has focused explicitly on this topic, with the exception of Benadon (2003). The present study sought to fill that lacuna and expand on previous research by specifically asking listeners to listen to a set of performances of the same organ piece and to group together interpretations recorded by the same performer.

The first objective of this study, which motivated the choice of organ music, was to determine whether participants could perform above chance in such a sorting task when listening to an instrument that allows only for limited timbral and dynamic differentiation. To address this question, all organists were asked to record the piece on the same instrument and using the same registration, thus severely restricting the range of acoustic cues available to listeners.

The second objective was to identify the acoustical parameters used by listeners to discriminate between performers. Given the absence of timbral and dynamic differentiation in organ performance, it was logically hypothesized that tempo and articulation (the degree of overlap between two successive notes) would be the most relevant parameters for this perceptual task.

A third objective was to explore issues related to the listeners' musical expertise and familiarity with a given instrument. Familiarity with an instrument could help a listener in focusing on the appropriate acoustical features of a performance. Indeed, Palmer (1996) reported that only listeners with keyboard experience could recognize the intended interpretation of a pianist when listening to recordings for which the intensity cues were removed. A recent study compared the neurophysiological responses to music in instrumentalists with different listening biographies, showing that instrumental expertise and listening biography entailed different patterns of neural activation (Margulis, Mlsna, Uppunda, Parrish, & Wong, 2007). On the other hand, Palmer and colleagues reported that non-musicians were as proficient as musicians in distinguishing familiar from novel performances of the same piece (Palmer et al., 2001), and Timmers (2005) found that predictive models of perceptual similarity between performances were highly similar for non-musicians and musicians. These studies suggest that the effect of listener expertise could be task-dependent. To address this issue, two groups of listeners, musicians (non-organists) and non-musicians with limited exposure to organ music, were invited to listen to the performances.[1]

Another goal of this study was to assess the effect of expressive intent on the ability of listeners to identify a performer. Since expressive actions have been shown to elicit stronger perceptions of individuality than more prosaic activities (Loula et al., 2005), we surmised that a performer's artistic individuality would be conveyed more clearly when performing a piece in an expressive manner, rather than in a "mechanical" or "deadpan" rendition. In order to test this hypothesis, performers were asked to record expressive and mechanical interpretations of the piece.

[1] It was unfortunately not possible to constitute a third group of listeners who were themselves organists, due to the limited availability of organists.

Finally, we sought to explore a potential link between the performers' level of expertise and the ability of listeners to recognize their performances. Performers were thus divided into two groups: those having previously won one or more prizes at national organ competitions and those who were non-prize winners. We hypothesized that prize-winning performers would be easier to identify, either because their superior technical proficiency would result in more controlled and consistent performances, or because their artistic individuality could have, in itself, led to their success in competitions. Furthermore, we predicted that performers whose recordings sounded very different from each other would be more difficult to identify than performers whose renditions were quite similar to each other.

METHOD

First phase: obtaining the organ performances

Musical Materials. The piece chosen for this experiment was Samuel Scheidt's (1587-1654) chorale setting of *Wachet auf, ruft uns die Stimme* (SSWV 534). This piece was selected for the following reasons: first, it is typical of the Baroque organ repertoire; second, it is a relatively easy piece that performers could learn in a short amount of time, and finally, its brevity made it possible to record several performances without tiring the performers.

Performers. Eight professional organists from the Montreal area were invited to record this piece. Their mean age was 26 years (range: 19 to 30 years). All participants identified themselves as right-handers. They had received organ instruction for a mean duration of 9 years (range = 3-13 years) and had 4 to 21 years of experience playing the organ. All of them held or had held a position as church organist for an average of 8 years (range = 1-21 years).Three of them had previously won one or more prizes in national organ competitions. None of the performers was familiar with the piece prior to the experiment.

Procedure. The musical score was given to the performers 20 minutes before the recording session began, in order to give them time to practice. They were asked to record two expressive interpretations of the piece, followed by two mechanical 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).

Performances were recorded on the Casavant organ of the Church of St-Andrew & St-Paul in Montreal, which is equipped with a MIDI console (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.[2] For the experiment, the stops used were the Spitz Principal 8', the Spitz Principal 4', and the Fifteenth 2' on the "Great" manual. All performers used the same registration.

The audio signal was recorded through two Boehringer ECM 8000 omnidirectional microphones. The microphones were located 1.20 m behind the organ bench, at a height of 1.70 m, and were placed 60 cm apart. The audio and MIDI signals were sent to a PC computer through a MOTU audio interface. Audio and MIDI data were then recorded using Cakewalk's SONAR software and

[2] Information provided by Mark Gilliam, Sales manager of Solid State Organ Systems.

stored on a hard disk. The MIDI data were matched to the score of the piece using an algorithm developed in MATLAB for this project (see Chapter 6).

Second phase: listening experiment

Musical Materials. The first musical phrase of Scheidt's setting of *Wachet auf* was used in the main phase of the listening experiment (Figure 3.1). This phrase represented a syntactically coherent musical unit, ending with a perfect cadence in the dominant key. All four recordings made by each performer were used. In addition, an identical duplicate of the first expressive recording of each performer was added as an experimental control. In order to keep a reasonable number of excerpts and to increase the difficulty of the sorting task by reducing the range of variation between excerpts, the recordings of the performers with the fastest and slowest global tempi (both non-prize winners) were not used in the listening experiment. Thus, there were five excerpts for each of the six remaining performers, for a total of 30 excerpts ranging in duration from 10 to 14 seconds. For the training phase, a similar musical phrase taken from the same piece was used. Three recordings made by two performers were used for this phase, for a total of six excerpts ranging in duration from 9 to 12 seconds (see Figure 3.1).

Participants. Twenty non-musicians with less than 2 years of musical training and limited exposure to organ music (no regular church attendance) and 20 musicians (music students having completed at least one year of undergraduate studies) participated in the listening experiment. They were recruited from the McGill University psychology subject pool or from the McGill community. Those who registered via the subject pool received academic credits, while others were given $10 as compensation for their time. The mean age of the participants was 21.6 years for the musicians (*SD* = 1.9 years), and 22.1 years for the non-musicians (*SD* = 2.8 years).

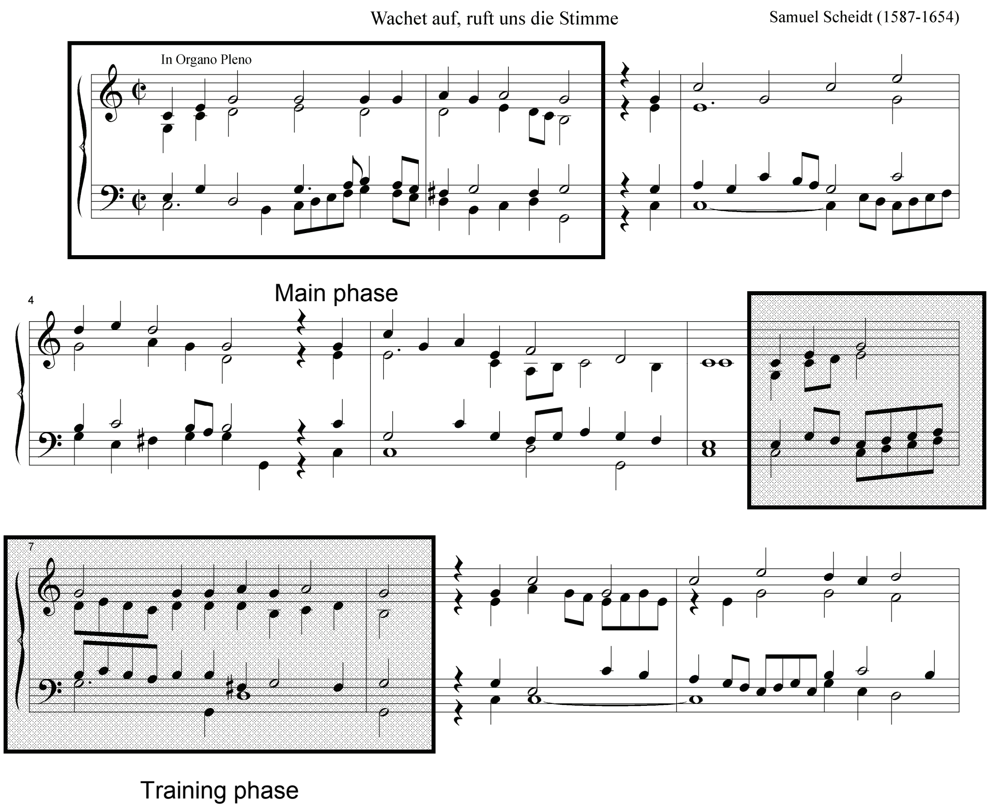


Figure 3.1. Excerpts from Samuel Scheidt's chorale setting of *Wachet auf, ruft uns die Stimme* used for the training phase (grayed box) and main phase (non-colored box) of the listening experiment.

Procedure. The experimental interface, programmed into MATLAB (adapted from Giordano, McAdams, & McDonnell, 2007), consisted of a screen in which all the excerpts were identified by squares numbered from 1 to 30, and the six performers were represented by empty boxes labeled A to F, in which the squares could be placed. The numbering of the excerpts was randomized for each listener. Participants could not assign an excerpt to a performer before having listened to it. They were free to move squares in and out of boxes and could listen to an excerpt or to the contents of a box as many times as they pleased.

The experiment took place in a sound-attenuated booth on an Apple McIntosh G5 computer. Participants wore Sennheiser HD 280 Pro headphones (diotic listening). The loudness level was set at 70 dB. All participants first passed an audiogram to ensure that they had normal hearing. After having familiarized themselves with the experimental interface and completed the training phase, they proceeded with the main phase of the experiment. Participants were instructed to listen to the 30 excerpts and group together those that were played by the same performer. The sorting task was constrained: participants were told that the excerpts had been recorded by six different performers and that each performer had recorded the piece five times. However, they were not made aware that performers had recorded mechanical and expressive versions of the piece and that some excerpts were identical duplicates. Once the experiment was completed, participants filled out a questionnaire. The entire experiment lasted approximately 1 hour.

For each participant, a co-occurrence matrix indicating which excerpts were grouped together (that is, assigned to the same performer) was produced. A log file containing data on each participant's sorting process and listening activity was also recorded.

RESULTS

Characterization of the musical features of the excerpts

In order to compare the excerpts on the basis of their musical features, an analysis was conducted on the following parameters: global tempo (expressed as mean quarter note duration), local tempo variation (expressed as standard deviation of the local tempo), articulation (expressed as mean overlap), and onset asynchrony, which essentially comprise the range of expressive parameters that are controlled by the performer in Baroque organ music (excluding registration effects which were controlled for in this experiment). Table 3.1 lists the mean values for these parameters, for each performer (identified by the letters A to F). Since the purpose of this analysis was to compare the excerpts both on the basis of their respective performers and of their expressive intent, mixed-model analyses of variance (ANOVA) were conducted for each of the aforementioned parameters, with performer as a random factor and expressive intent as a fixed factor, on the 24 excerpts that were used in the main phase of the listening experiment (Table 3.2). Post-hoc tests (Tukey HSD) were conducted to identify the parameters on the basis of which individual performers could be significantly differentiated.

Table 3.1. Mean values for the expressive parameters, averaged for each performer.

Performer	A*	B	C	D*	E	F*
Global tempo: mean quarter note duration (ms)	637 (5)	716 (41)	723 (14)	839 (11)	832 (63)	656 (23)
Mean standard deviation of local tempo (ms)	54	42	52	67	59	41
Articulation: mean overlap (ms)	-63 (61)	-106 (40)	-44 (10)	-160 (37)	-109 (34)	-161 (9)
Mean onset asynchrony (ms)	9.0 (0.9)	10.7 (1.2)	7.8 (0.4)	8.5 (2.6)	8.5 (0.8)	7.2 (0.4)

Note. Prize-winners are indicated with an asterisk. Standard deviations are given in parentheses.

Table 3.2. Analyses of variance for the expressive parameters of the excerpts used in the main phase of the listening experiment.

Expressive parameters	Factors			Post-hoc tests
	Performer	Expressive intent	Performer × Expressive intent	Comparison by performer
Global tempo (mean quarter note duration)	$F(5, 12) = 59.88$***	$F(1, 5) = 2.22$	$F(5, 12) = 3.91$*	E D B C A F
Mean standard deviation of local tempo	$F(5, 12) = 5.39$**	$F(1, 5) = 8.35$*	$F(5, 12) = 3.25$*	D C A E F B
Articulation (mean overlap)	$F(5, 12) = 44.57$***	$F(1, 5) = 0.07$	$F(5, 12) = 20.18$***	A C E B F D
Onset asynchrony	$F(5, 12) = 5.54$**	$F(1, 5) = 0.56$	$F(5, 12) = 3.11$*	B D E A F C

Note. For post-hoc tests, performers whose means did not differ significantly are grouped together. * $p < .05$; ** $p < .01$; *** $p < .001$.

Local tempo variation. The local tempo was computed for each quarter note for all excerpts. The standard deviation of the local tempo was used as a measure of the degree of local tempo variation. Local tempo variations were significantly smaller for the mechanical excerpts than for the expressive ones. These results are congruent with previous studies reporting that expressive performances typically exhibit more pronounced local tempo variations than mechanical performances (Palmer, 1989). As with global tempo, different performers varied with respect to the amount of local tempo variation they used.

Overlap. Mean overlap was defined as the time interval between two consecutive notes, and calculated as the offset of note event n minus the onset of note event $n+1$. A positive overlap indicates a *legato* articulation, while a negative value represents a detached or *staccato* articulation. Significant differences in the amount of overlap were found between performers, but no effect of expressive intent was observed. Post-hoc tests confirmed that performers could be divided into three distinct groups on the basis of the mean overlap, with D and F using a very detached articulation, while A and C played quasi-*legato*. The highly significant interaction between performer and expressive intent reflects the fact that some organists performed the mechanical excerpts in a more *staccato* fashion than the expressive ones, while others did the exact opposite. In contrast, Palmer (1989) reported that pianists played unexpressive excerpts in a more detached way than expressive ones.

Onset asynchrony. Onset asynchrony was measured as the standard deviation of the difference in onset times between notes of a chord (Palmer, 1989; Rasch, 1979). As with other expressive parameters analyzed here, the degree of synchronization differed significantly among performers. However, unlike results reported for piano performance (Palmer, 1989), asynchronies were not larger in expressive performances than in mechanical ones. Given the lack of dynamic differentiation on the organ, these results are perhaps not unexpected, in light of more recent studies suggesting that onset asynchrony is related to dynamic differentiation between voices (Goebl, 2001; Repp, 1996; see also Chapter 2). It should also be noted that asynchronies across all excerpts averaged 9 ms (*SD* = 2 ms), which is noticeably less than the asynchronies of 15 to 20 ms which are typically observed in piano performance (Palmer, 1989). It is therefore unlikely that excerpts from different performers could have been segregated on the basis of differences in onset asynchrony, since the reported threshold for detecting onset asynchronies is around 20 ms (Hirsh, 1959).

From these analyses, it may be concluded that the main difference between expressive and mechanical excerpts lies in the amount of local tempo variation. Furthermore, different performers could be statistically distinguished on the basis of global tempo, amount of local tempo variation, mean overlap, and degree of synchronization, although the latter may not have been a perceptually relevant parameter given the small size of the asynchronies observed here.[3]

General assessment of the listeners' sorting accuracy

A measure of the listeners' sorting accuracy can be obtained by comparing their partitioning of the excerpts with the correct partition, which corresponds in this case to a partition in which all the excerpts recorded by the same performer

[3] Although the analyses presented here refer only to the first phrase of Scheidt's chorale setting, similar results were obtained for entire performances of the piece.

are grouped together. The adjusted Rand index (Hubert & Arabie, 1985) is a widely used statistical tool to measure the degree of agreement between two partitions. A comparison between each participant's grouping and the correct partition yielded positive adjusted Rand index values (indicating better than chance sorting accuracy) for 20 musicians (100% of the participants) and 18 non-musicians (90%). A more stringent criterion would be to assess whether a participant's sorting accuracy was significantly better than chance, corresponding to a probability of less than 5% ($p < .05$) of obtaining an adjusted Rand index value this high or higher by chance. Using this criterion, 15 musicians (75%) and 13 non-musicians (65%) performed significantly above chance, as estimated by bootstrap methods (Efron & Tibshirani, 1993). Although musicians fared slightly better than non-musicians, no significant difference was observed between the sorting accuracy of the two groups [$t(38) = 1.24$, $p = .22$].

To assess the sorting accuracy of the group of participants as a whole, a K-means cluster analysis was conducted on the aggregate partitioning data from all participants. Since participants had to assign excerpts to six groups, a solution was computed for six clusters. The adjusted Rand index of the solution was 0.49 ($p < .001$, bootstrap estimation method), indicating that the partitioning structure recovered from the aggregate data was a reasonably close approximation of the correct partition.

Representing the listeners' perceptual space

The co-occurrence matrix, which tabulates the relative frequency with which two excerpts were grouped together by participants, can be used to build a model of the listeners' perceptual space, by assuming that excerpts that are often grouped together are closer to each other than excerpts that are not (Arabie & Boorman, 1973). A multidimensional scaling (MDS) analysis was thus conducted on each participant's co-occurrence matrix, in order to uncover the main dimensions of the listeners' perceptual space. The INDSCAL procedure (Carroll & Chang, 1970), which models not only the perceptual space of the participants as a group, but also the weights that each participant gave to the dimensions of the MDS space, was used to interpret the spaces of individual listeners.

Fit of the MDS solution. In order to determine the best dimensional fit for the MDS solution, fit-by-dimensionality analyses were conducted, taking into account both the stress measure (Kruskal stress-I) and the proportion of variance explained (RSQ). Since the INDSCAL procedure only provides solutions with two or more dimensions, MDS solutions were first computed for one to five dimensions using a non-metric Euclidean distance model on the aggregate data for the entire group of participants. INDSCAL solutions were then computed for two to five dimensions (Figure 3.2). These analyses revealed that a two-dimensional representation (shown in Figure 3.3) provided an adequate fit; no increase in the proportion of variance explained was observed for solutions with a higher dimensionality using the INDSCAL procedure.

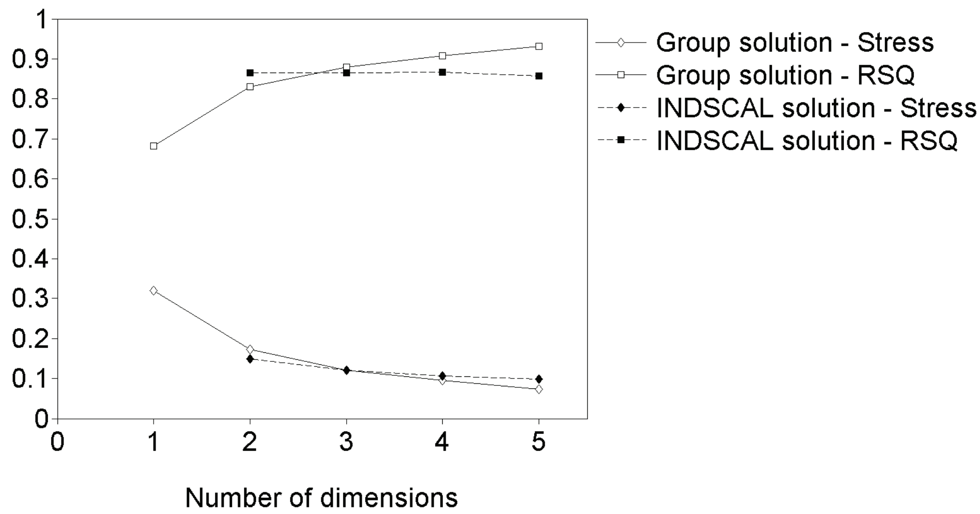


Figure 3.2. Fit-by-dimension plots for both group and INDSCAL multidimensional scaling solutions. *Stress*: Kruskal stress-I. *RSQ*: proportion of variance explained.

Interpretation of the dimensions. Regression analyses were conducted on the musical features of each excerpt to construct a statistical model for interpreting the dimensions of the MDS solution (Kruskal & Wish, 1978). Only musical parameters that were significantly correlated with at least one of the dimensions were included in the regression analyses. Significant correlations were found between coordinates on the first dimension (abscissa) and mean quarter note duration ($r = 0.89$, $p < .001$), as well as standard deviation of the local tempo variation ($r = 0.39$, $p < .05$). A forward stepwise multiple regression of these two parameters on the first dimension showed an excellent fit with mean quarter note duration as sole predictor ($R^2 = 0.79$, $F = 102.94$, $p < .001$). Only mean overlap was found to be significantly correlated with coordinates on the ordinal axis ($r = -0.78$, $p < .001$), and this parameter explained 60.1% of the total variance on that dimension ($F = 31.12$, $p < .001$).

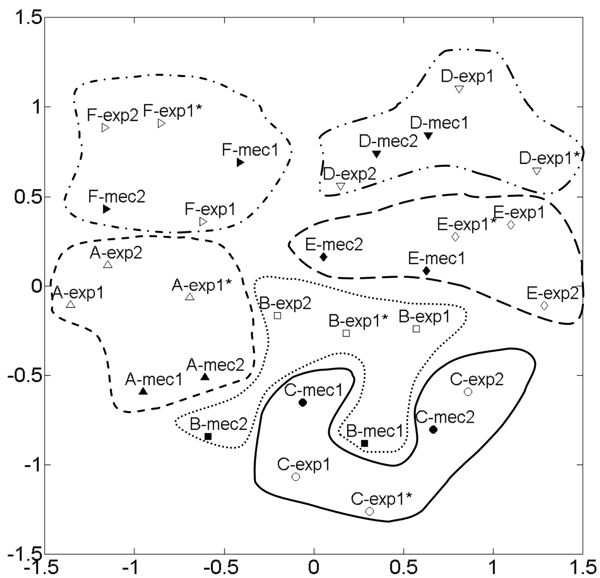


Figure 3.3. Multidimensional scaling of the perceptual distances between excerpts, based on the co-occurrence matrix from the sorting task. Letters *A* to *F* identify individual performers; *exp* (open symbols) refers to expressive excerpts and *mec* (filled symbols) to mechanical ones. Numbers refer to the order of recording; asterisks indicate duplicate excerpts. Excerpts from the same performer are circled together. The MDS solution was generated using the INDSCAL procedure (monotonic regression; Kruskal stress-I = 0.15; RSQ = 0.87).

These observations suggest that global tempo and articulation were the most important parameters used by listeners to discriminate between performers. The graphical representation (Figure 3.3) shows that performers who chose faster tempi are grouped on the left (A and F), whereas performers who employed more deliberate tempi are found on the right (D and E). Performers who played with a quasi-*legato* articulation are found in the lower section on the graph (C), while performers using a more detached articulation are located in the upper portion (D and F). These results, which underscore the importance of absolute tempo as a major component of the perceptual representation of the distance between performances, are in agreement with Timmers' (2005) findings, which established a preference for perceptual models based on absolute values. However, in contrast to Timmers, we did not find that a local tempo model was a better fit than a global one, although a very good fit was also obtained when using absolute differences in local tempo (computed on a quarter-note basis) to predict coordinates on the first dimension ($R^2 = 0.75$, $F = 81.66$, $p < .001$).

Individual weights. An inspection of the individual stress values (minimum: 0.06; maximum: 0.22) and RSQ values (minimum: 0.71; maximum: 0.98) confirms that the model provided a reasonably good fit for all participants. Musicians (mean weights for the first dimension: 0.70, *SD* = 0.03; for the second dimension: 0.69, *SD* = 0.03), and non-musicians (mean weights for the first dimension: 0.70, *SD* = 0.02; for the second dimension: 0.68, *SD* = 0.01) ascribed nearly identical importance to both dimensions. These results indicate that tempo and articulation were of equal perceptual relevance in the sorting task for both musicians and non-musicians and that differences between the weights of individual participants were negligible, as evidenced by the small standard deviations.

Effect of expressive intent

To analyze the effect of the performers' expressive intent on the listeners' ability to sort excerpts correctly, the participants' partitions must be decomposed by comparing the performer's identity and the expressive intent of excerpts that were grouped together. Such analyses typically involve comparisons of pairs of excerpts (Daws, 1996; Miller, 1969). The proportion of pairs of excerpts grouped together (observed pairs) to the total number of possible pairs was thus computed for each of the following types of pairs of excerpts (Figure 3.4):

a) One mechanical and one expressive excerpt from different performers

b) Two mechanical excerpts from different performers

c) Two expressive excerpts from different performers

d) One mechanical and one expressive excerpt from the same performer

e) Two mechanical excerpts from the same performer

f) Two expressive excerpts from the same performer

g) Two identical expressive excerpts from the same performer (duplicates).

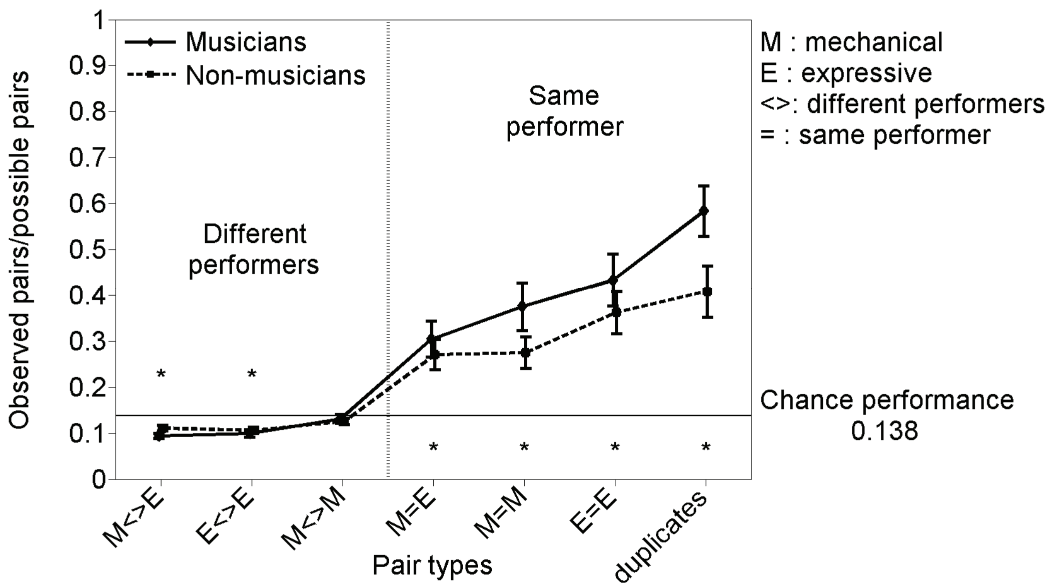


Figure 3.4. Proportion of observed pairs compared to the total number of possible pairs for all pair types. Error bars indicate standard errors of the mean. Asterisks indicate values that are significantly different from chance performance (for both musicians and non-musicians) as determined by two-tailed one-sample *t* tests (Bonferroni-corrected $p < .02$ in all cases).

Results show that the proportion of observed pairs was significantly above chance for all types of correct pairs (representing excerpts from the same performer), with expressive-expressive pairings occurring more frequently than pairs comprising at least one mechanical excerpt. For wrong pairs (corresponding to excerpts played by different performers), the proportion of observed pairs involving at least one expressive excerpt was significantly below chance, while mechanical-mechanical pairings occurred at a rate close to that expected by chance.[4] Taken together, these results indicate not only that participants exhibited a positive bias towards pairs composed of excerpts from the same performer, but also that they grouped expressive excerpts from the same performer more often than mechanical ones. Conversely, participants exhibited a negative bias towards pairs composed of excerpts from different performers that included at least one expressive excerpt, but did not discriminate against pairs composed of mechanical excerpts from different performers.

A repeated-measures logistic regression analysis was conducted on the proportion of pairs of excerpts grouped together with the following factors: participant's musical training (musician or non-musician), and, for each pair of

[4] Listeners had to sort 30 excerpts into six groups of five excerpts. For 30 items, a total of 435 different pairs can be generated (30! / (28! × 2!)). A partition of these 30 items into six groups of five items contains 60 pairs (6 × (5! / (3! × 2!))). The probability for a given pair of appearing in a given partition is thus equivalent to 60 / 435, or p = .138. Since there are many more possible wrong pairs (375) than correct pairs (60), the proportion of wrong pairs which include an expressive excerpt may not appear to be significantly below chance level in Figure 3.4, although it actually is.

excerpts, performer identity (same for both excerpts/different for each excerpt) and expressive intention (mechanical/mechanical, mechanical/expressive, and expressive/expressive).[5] The full model shows a significant effect of performer identity, $\chi^2(1) = 25.91$, $p < .001$, indicating that participants favored pairs composed of excerpts played by the same performer, an effect of expressive intention, $\chi^2(2) = 19.57$, $p < .001$, which reflects the differences in sorting accuracy observed between expressive and mechanical excerpts, and an interaction between expressive intention and performer identity, $\chi^2(2) = 15.17$, $p < .001$, which indicates that while pairs of expressive excerpts from the same performer were more likely to be grouped together than pairs of mechanical excerpts, the reverse was observed with pairs from different performers. Again, no significant effect of musical training was observed. A separate model was built for each level of the expressive intention factor, showing a significant effect of musical training only for the expressive pairs played by the same performer, $\chi^2(1) = 3.92$, $p < .05$. This effect seems largely explainable by the lower accuracy of the non-musicians on the duplicate pairs (see Figure 3.4).

Effect of performer expertise

In order to determine whether the performers' level of expertise had an effect on the listeners' sorting accuracy, a repeated-measures analysis of variance was conducted on the proportion of correct pairs (that is, pairs of excerpts recorded by the same performer), with performer expertise (prize-winner or non-winner) as a within-subject factor, and musical training as a between-subjects

[5] Duplicate pairs were included with the expressive-expressive pairings for this analysis.

factor. A significant effect of performer expertise was observed, $F(1, 38) = 11.97$, $p < .01$, indicating that participants were more accurate at sorting out the recordings of prize-winning performers than those of non-prize winners (Figure 3.5).

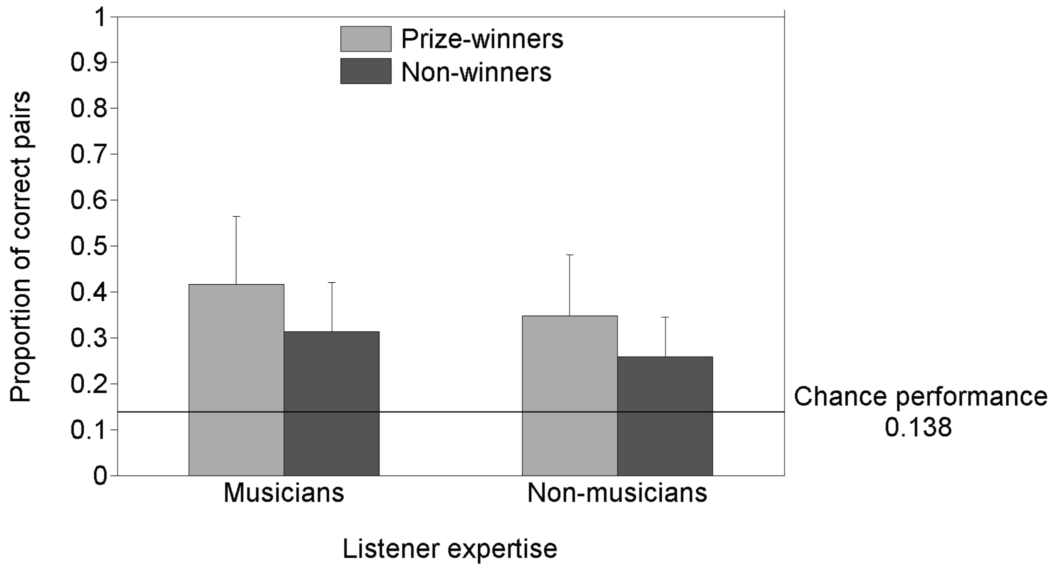


Figure 3.5. Proportion of correct pairs (excerpts recorded by the same performer) for prize-winning performers versus non-prize winners. Error bars indicate standard errors of the mean.

Predicting sorting accuracy for individual performers

We also sought to predict the sorting accuracy for individual performers, based on the musical features of the excerpts. One hypothesis, mentioned previously, would be that sorting accuracy is related to consistency: performers whose recordings sounded similar to each other would be easier to group together than performers whose recordings sounded quite different. An examination of the expressive parameters showed that prize-winners were generally more consistent regarding mean global tempo (as indicated by the size of the standard deviation, see Table 3.1), but not articulation. However, these parameters only relate to global aspects of the performance, and do not take into account local patterns. Local tempo variations constitute a central aspect of musical expressivity: high-level performers tend to exhibit very well-defined and idiosyncratic temporal profiles (Repp, 1990, 1992), while listeners are extremely sensitive to small changes in expressive timing (Clarke, 1989). As previously noted, coordinates on the first dimension of the MDS solution could be predicted almost equally well by absolute differences in local tempo rather than mean quarter note duration, implying that local tempo variations could have been a major component of the listeners' perceptual space. Moreover, the fact that participants had more difficulties in sorting out mechanical excerpts than expressive ones points to an important role for local tempo variations, since the only statistically significant difference between mechanical and expressive excerpts was that the former exhibited smaller local tempo variations on average.

The degree of similarity between local temporal patterns of different excerpts was evaluated by computing local tempo correlations on a quarter-note basis. These correlations were then averaged across all recordings from the same performer, yielding a measure of consistency. A high correlation indicated that an performer's temporal patterns were very consistent across recordings. Excerpts from each performer were also compared with excerpts from other performers, and the average correlation coefficients were used to provide a measure of distinctiveness: a low correlation coefficient indicated that an performer's temporal patterns were very different from those of other performers. As shown in

Table 3.3, sorting accuracy was higher for performers who were either very consistent (indicated by high correlations with their own excerpts), or very distinctive (indicated by low correlations with excerpts from other performers). These traits were exhibited most clearly in the prize-winners (performers A, D, and F), who were also sorted the most successfully by participants. Thus, A's excerpts all followed very similar temporal patterns, while F was not especially consistent, but exhibited a very distinct temporal pattern. On the other hand, B was by far the least consistent performer, as well as the most poorly recognized. Performers E and especially C were nearly as consistent as some of the prize-winning performers, and listeners' sorting accuracy for their excerpts was closer to that observed for prize-winners. However, a repeated-measures analysis of variance on the proportion of correct pairs which excluded performer B's data confirmed a robust effect of performer expertise, F (1, 38) = 9.12, $p < .01$.

This analysis shows that at least two main factors were involved in determining how well performers' artistic individuality was conveyed to listeners: first, the consistency of their performances, as reflected by the within-performer local tempo correlations, and second, the distinctiveness of their interpretations, as reflected by the between-performers correlations.

Table 3.3. Mean local tempo correlation coefficients and proportions of pairs correctly grouped, for each performer.

Performer	A*	B	C	D*	E	F*
Mean correlation with own excerpts	0.89 (0.05)	0.43 (0.29)	0.80 (0.08)	0.76 (0.09)	0.64 (0.18)	0.59 (0.12)
Mean correlation with other performers' excerpts	0.46 (0.23)	0.35 (0.29)	0.41 (0.23)	0.24 (0.33)	0.35 (0.29)	0.19 (0.28)
Proportion of correct pairs	35.3%	21.8%	33.0%	37.0%	31.0%	42.5%

Note. Mean correlation with own excerpts: mean correlation coefficients between excerpts played by the same performer. *Mean correlation with excerpts from other performers*: mean correlation coefficients between excerpts played by an performer and excerpts from other performers. *Proportions of correct pairs:* proportion of pairs of excerpts played by the same performer that were correctly grouped together by listeners. Prize-winning performers are marked with an asterisk. Standard deviations are given in parentheses.

Predicting sorting accuracy for individual listeners

While no significant difference was observed in the sorting accuracy of musicians and non-musicians as a group, large differences were observed between individual participants. In order to identify the factors responsible for these differences, each participant's log file was examined. Listening activity, defined by the total number of times a participant listened to each excerpt, was a logical candidate to invoke for individual differences in sorting accuracy, since it not only varied greatly between participants (who could listen to the excerpts as many times as they wanted), but was also expected to influence performance in the sorting task. Indeed, listening activity was found to be significantly correlated with sorting accuracy (musicians: $r(18) = 0.46$, $p < .05$; non-musicians, $r(18) =$

0.47, $p < .05$). Musicians listened to more excerpts than non-musicians on average (mean number of excerpts listened to for musicians: 207.6; for non-musicians: 178.6). Although this difference did not reach significance, it may account for the musicians' slightly higher accuracy. As can be seen in Figure 3.6, the slope of the linear regression for the proportion of correct pairs versus listening activity was indeed very similar for both groups.

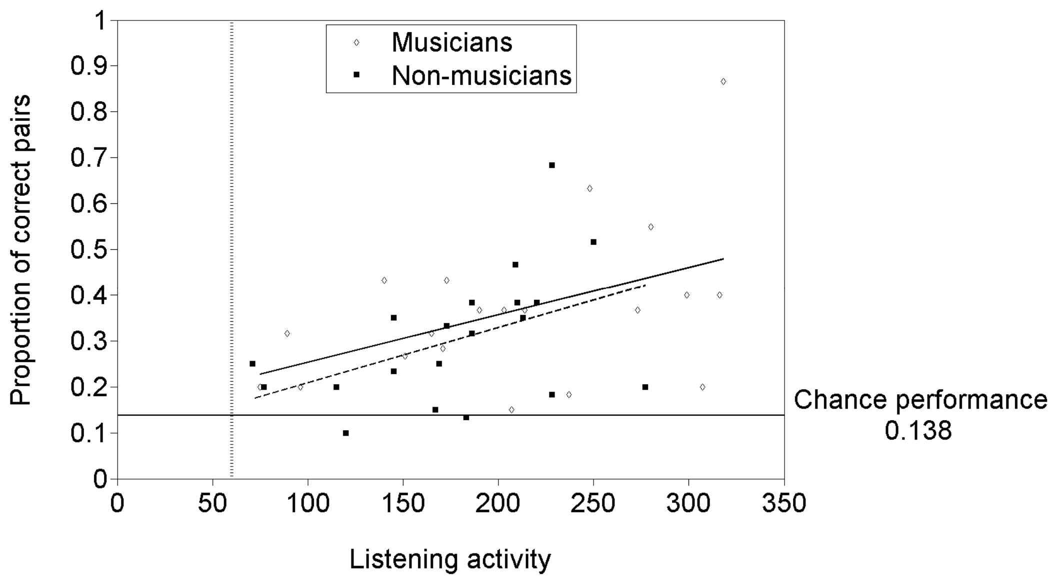


Figure 3.6. Proportion of correct pairs versus listening activity for musicians and non-musicians. Regression lines are indicated for musicians (solid line) and non-musicians (dashed line). Participants minimally had to listen to each of the 30 excerpts twice in order to complete the sorting task (vertical dotted line).

In order to model the participants' performance in the sorting task, a repeated-measures analysis of covariance on the proportion of correct pairs was conducted, with listening activity as a covariate, musical training as a between-subjects factor, and performer expertise as a within-subject factor, with each performer as a separate factor nested within performer expertise. Significant effects were observed for performer expertise, $F(1, 194) = 15.23$, $p < .001$, performer, $F(4, 194) = 2.73$, $p < .05$, and listening activity, $F(1, 37) = 9.91$, $p < .01$.

DISCUSSION

The present study is, to our knowledge, the first published research linking together the effects of listener expertise, performer expertise, and expressive intent on listeners' ability to successfully group together recordings of the same piece that were played by the same performer. Most listeners, whether musicians or non-musicians, were able to perform significantly better than chance in this task, even though these excerpts could only be differentiated over a limited number of acoustic dimensions. This suggests that sufficient information to identify a performer's individual style could be projected in a short (10- to 14-second) recording and in the absence or intensity or timbral cues. An MDS analysis showed that both musicians and non-musicians discriminated between performers mainly on the basis of tempo and articulation and that individual differences in the dimension weights were negligible, implying that all participants shared a common set of perceptual cues. These results, which are not unexpected in light of the small number of acoustic cues that were available to listeners, are in agreement with Timmers' (2005) findings that musicians and non-musicians draw on similar perceptual models when asked to assess the degree of similarity between performances.

As suggested by these observations, one possible strategy for completing the task would be simply to group together excerpts that sound similar on the basis of tempo and/or articulation. However, as some participants noted in their comments, absolute tempo was not always a reliable cue because some performers could exhibit a fairly wide range of tempi across their recordings. A more elaborate strategy might be to build a psychological representation of the performers' musical identities based on the available acoustic cues in order to sort out the excerpts. This strategy might have been used by some participants who employed adjectives to describe the performers' musical personalities. Since emotions (Juslin, 2000) and even person-related semantic dimensions such as male-female (Watt & Ash, 1998; Watt & Quinn, 2007) have been shown to be reliably transmitted through music, it is not unreasonable to suppose that some aspects of the performers' personalities could be conveyed as well. The present study did not, however, explicitly seek to identify the cognitive strategies employed by participants in the sorting task, and further research will be necessary in order to address this issue.

Expressive intent affected sorting accuracy: expressive interpretations from the same performer were more likely to be grouped together than mechanical ones, and expressive performances from different performers were less likely to be grouped together than mechanical ones. These observations provide evidence that performer individuality was conveyed more efficiently through expressive recordings, thus corroborating earlier findings on movement-based recognition (Loula et al., 2005). Since expressive intent was found to be linked to the magnitude of local tempo variations, it may be surmised that artistic individuality was conveyed, at least in part, through expressive variations in local tempo patterns.[6] Indeed, an analysis of local tempo patterns revealed that performers who exhibited superior consistency across their performances or who employed distinctive expressive patterns were sorted more successfully by listeners. Moreover, the performances of prize-winning performers were sorted more successfully than those of non-winners, and prize-winners were generally either more consistent or distinctive than non-prize winners. These findings imply that both superior consistency and the use of distinctive expressive features could be closely linked with the projection of a well-defined musical personality (Sloboda, 2000). This leads to the intriguing suggestion that success in performance competitions, and by extension peer recognition and critical acclaim, could be related to the degree of perceived artistic individuality as well as to the level of technical competence. It should be noted, however, that extreme individuality or distinctiveness may not always be preferred. Thus, statistically average human faces are generally perceived as more attractive than less typical ones (Langlois & Roggman, 1990), and conventionality is sometimes favored over individuality in music performance (Repp, 1997).

Although this study has shed some light on the phenomenon of artistic individuality in music performance, it also leaves several questions unanswered. For instance, the notion of a performer's individual "stylistic space" is an important concept that remains to be explored. Indeed, while this study has provided evidence that a performer's individual style could be recognized across

[6] It is worth noting in this context that a Baroque chorale setting may not be as conducive to the expression of a performer's musical individuality as, for instance, a Romantic piece could be, since large tempo variations are not typically part of this style.

varying levels of expressivity on several recordings of the same piece, it remains to be seen whether listeners could recognize an unfamiliar performer's style across different pieces or even different genres, and whether they could outperform computational models such as Stamatatos & Widmer's learning ensemble (2005) in such a task. The fact that computational approaches have achieved high recognition rates suggests that some musical characteristics or acoustical cues associated with a performer's specific style remain more or less invariant across various pieces and genres, potentially enabling listeners to recognize it. Possible associations between specific musical features or acoustical parameters of the performances and perceived personality traits should also be investigated, following Juslin's (2000) work on the communication of emotion in music performance. Finally, the results presented here point to interesting links between musical competence, aesthetic preferences, and the communication of artistic individuality, which warrant further inquiry.

ACKNOWLEDGMENTS

This research was supported by fellowships from the Social Sciences and Humanities Research Council of Canada and from the Centre for Interdisciplinary Research in Music Media and Technology (CIRMMT) to Bruno Gingras, as well as a grant from the Natural Sciences and Engineering Research Council and a Canada Research Chair awarded to Stephen McAdams. The design of this study benefited from the comments and suggestions of Caroline Palmer. We thank Julien Boissinot, Darryl Cameron, and Bennett Smith for their technical assistance, Peter Holmes for permission to use McGill University's sound recording equipment, Jonas Braasch for his 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 as well as the subjects in the sorting task experiment for their participation.

REFERENCES

Arabie, P., & Boorman, S. A. (1973). Multidimensional scaling of measures of distance between partitions. *Journal of Mathematical Psychology, 10*(2), 148-203.

Belin, P., Zatorre, R. J., Lafaille, P., Ahad, P., & Pike, B. (2000). Voice-selective areas in human auditory cortex. *Nature, 403*(6767), 309-312.

Benadon, F. (2003). Spectrographic and calligraphic cues in the identification of jazz saxophonists. In R. Kopiez, A. C. Lehmann, I. Wolther & C. Wolf (Eds.), *Proceedings of the 5th Triennial ESCOM Conference* (pp. 245-249). Hanover, Germany.

Brown, R. (1981). An experimental study of the relative importance of acoustic parameters for auditory speaker recognition. *Language and Speech, 24*(4), 295-310.

Carroll, J. D., & Chang, J. J. (1970). Analysis of individual differences in multidimensional scaling via an N-way generalization of Eckart-Young decomposition. *Psychometrika, 35*(3), 283-319.

Clarke, E. F. (1989). The perception of expressive timing in music. *Psychological Research-Psychologische Forschung, 51*(1), 2-9.

Daws, J. T. (1996). The analysis of free-sorting data: Beyond pairwise cooccurrences. *Journal of Classification, 13*(1), 57-80.

Efron, B., & Tibshirani, R. (1993). *An introduction to the bootstrap*. New York: Chapman & Hall.

Giordano, B. L., McAdams, S., & McDonnell, J. (2007). Acoustical and conceptual information for the perception of animate and inanimate sound sources. In *Proceedings of ICAD-07 (13th Meeting of the International Conference on Auditory Display)* (pp. 173-181). Montreal, Canada.

Goebl, W. (2001). Melody lead in piano performance: Expressive device or artifact? *Journal of the Acoustical Society of America, 110*(1), 563-572.

Hirsh, I. J. (1959). Auditory perception of temporal order. *Journal of the Acoustical Society of America, 31*(6), 759–767.

Holmgren, G. L. (1967). Physical and psychological correlates of speaker recognition. *Journal of Speech and Hearing Research, 10*(1), 57-66.

Hubert, L., & Arabie, P. (1985). Comparing partitions. *Journal of Classification, 2*(2-3), 193-218.

Juslin, P. N. (2000). Cue utilization in communication of emotion in music performance: Relating performance to perception. *Journal of Experimental Psychology-Human Perception and Performance, 26*(6), 1797-1812.

Keller, P. E., Knoblich, G., & Repp, B. H. (2007). Pianists duet better when they play with themselves: On the possible role of action simulation in synchronization. *Consciousness and Cognition, 16*(1), 102-111.

Kendall, R. A., & Carterette, E. C. (1990). The communication of musical expression. *Music Perception, 8*(2), 129-164.

Kruskal, J. B., & Wish, M. (1978). *Multidimensional scaling*. Beverly Hills, Calif.: Sage Publications.

Langlois, J. H., & Roggman, L. A. (1990). Attractive faces are only average. *Psychological Science, 1*(2), 115-121.

Loula, F., Prasad, S., Harber, K., & Shiffrar, M. (2005). Recognizing people from their movement. *Journal of Experimental Psychology-Human Perception and Performance, 31*(1), 210-220.

Margulis, E. H., Mlsna, L. M., Uppunda, A. K., Parrish, T. B., & Wong, P. C. (2007). Selective neurophysiologic responses to music in instrumentalists with different listening biographies. *Human Brain Mapping* (Published online: 2007 Dec. 10).

Miller, G. A. (1969). A psychological method to investigate verbal concepts. *Journal of Mathematical Psychology, 6*(2), 169-191.

Palmer, C. (1989). Mapping musical thought to musical performance. *Journal of Experimental Psychology-Human Perception and Performance, 15*(12), 331-346.

Palmer, C. (1996). On the assignment of structure in music performance. *Music Perception, 14*(1), 23-56.

Palmer, C., Jungers, M. K., & Jusczyk, P. W. (2001). Episodic memory for musical prosody. *Journal of Memory and Language, 45*(4), 526-545.

Rasch, R. A. (1979). Synchronization in performed ensemble music. *Acustica, 43*(2), 121-131.

Repp, B. H. (1990). Patterns of expressive timing in performances of a Beethoven minuet by 19 famous pianists. *Journal of the Acoustical Society of America, 88*(2), 622-641.

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. (1996). Patterns of note onset asynchronies in expressive piano performance. *Journal of the Acoustical Society of America, 100*(6), 3917-3931.

Repp, B. H. (1997). The aesthetic quality of a quantitatively average music performance: Two preliminary experiments. *Music Perception, 14*(4), 419-444.

Sloboda, J. A. (2000). Individual differences in music performance. *Trends in Cognitive Sciences, 4*(10), 397-403.

Stamatatos, E., & Widmer, G. (2005). Automatic identification of music performers with learning ensembles. *Artificial Intelligence, 165*(1), 37-56.

Timmers, R. (2005). Predicting the similarity between expressive performance of music from measurements of tempo and dynamics. *Journal of the Acoustical Society of America, 117*(1), 391-399.

Van Dommelen, W. A. (1990). Acoustic parameters in human speaker recognition. *Language and Speech, 33*(3), 259-272.

Voiers, W. D. (1964). Perceptual bases of speaker identity. *Journal of the Acoustical Society of America, 36*(6), 1065-1073.

Watt, R., & Ash, R. (1998). A psychological investigation of meaning in music. *Musicae Scientiae, 2*(1), 33-54.

Watt, R., & Quinn, S. (2007). Some robust higher-level percepts for music. *Perception, 36*(12), 1834-1848.

Chapter 4. Effects of musical structure, expressive intent, performer's preparation, and expertise on error patterns in organ performance

Several aspects of musical structure have been shown to influence error patterns in music performance. For instance, errors have been found to occur more frequently in inner voices than in outer voices in performances of polyphonic music. In addition, errors are less likely to occur in the voice intended as melody than in nonmelody voices, and error patterns are influenced by performers' interpretative goals. One aspect that has not been empirically examined so far is whether these effects extend to piece-specific elements such as motives or themes. Additionally, a number of related issues have received little or no attention, such as the effects of hand assignment and structural salience on error rate, and the consistency and individuality of performers' error patterns. Chapter 4 is concerned with the influence of musical structure (motivic versus non-motivic passages), texture (homophonic versus polyphonic style), expressive intent, conditions of preparation (quick-study versus prepared piece), and level of accomplishment (prize-winning performers versus non-winners) on the distribution and frequency of errors in organ performance.

This chapter is based on the following research article:

Gingras, B., McAdams, S., Palmer, C., & Schubert, P. N. Performance error frequencies are inversely proportional to perceptual salience and musical significance. Manuscript prepared for submission to *Music Perception*.

ABSTRACT

We compared the influence of musical structure (motivic versus non-motivic passages), texture (homophonic versus polyphonic style), expressive intent, conditions of preparation (quick study versus prepared piece), and level of accomplishment (prize-winning performers versus non-winners) on the distribution and frequency of errors in organ performance. In the quick-study condition, eight organists recorded different interpretations of two short Baroque pieces of contrasting texture, Grigny's *Premier Agnus* and Scheidt's *Wachet auf, ruft uns die Stimme*. In the prepared condition, sixteen organists made two recordings of J.S. Bach's organ fugue in D minor (BWV 538). Results show that error rates were positively correlated with onset density, and were generally lower for motivic notes and for notes belonging to outer voices. Expressive intent affected the distribution of errors: performers made fewer errors for the notes belonging to the voice that they were trying to emphasize. Musical texture influenced the type of errors: a greater proportion of pitch and intrusion errors were harmonically appropriate in a homophonic texture than in a polyphonic one. Individual performers exhibited consistent and idiosyncratic error patterns. Finally, while no significant relationship was found between level of accomplishment and error rate in the quick-study condition, prize-winners made significantly fewer errors than non-winners in the prepared condition.

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 high-level performances contain various types of performance errors (Repp, 1996a). Whether they are perceivable or not, such errors are often a cause of concern for performers (Repp, 1996a); indeed, the amount and conspicuousness of errors may be regarded as one of the determinants of the aesthetic quality of a performance. 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). In a domain perhaps more closely related to music performance, a useful experimental paradigm to model human performance in activities that involve fine motor coordination in the production of sequentially ordered events has been afforded by the analysis of typing errors (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).

The present article is concerned with the influence of musical structure (motivic versus non-motivic passages), texture (homophonic versus polyphonic style), expressive intent, conditions of preparation (quick study versus prepared piece), and level of accomplishment (prize-winning performers versus non-winners) on the distribution and frequency of errors in organ performance. Three 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 chorale setting of homophonic character by Samuel Scheidt (1587-1654), and the organ fugue in D minor (BWV 538), better known as the "Dorian" fugue, by Johann Sebastian Bach (1685-1750). The first two pieces were used for the quick-study condition, while the last piece was used for the prepared condition. Since the database compiled for this research consisted of recordings of complete pieces by professional organists, which were also used to study expressive strategies in organ performance, error production was analyzed in an ecological context, thus complementing earlier studies in which performance errors were elicited (Palmer & Van de Sande, 1993, 1995). Furthermore, most studies on errors in music performance were conducted either on piano music from the Romantic and Classical eras (Repp, 1996a) or on short stimuli newly composed or adapted specifically for experimental purposes (Palmer & Van de Sande, 1993, 1995). One of the goals of this study was to assess whether previous findings in piano performance could be extended to other keyboard instruments, as well as to a different repertoire. The present study also sought to address related issues that had previously received little or no attention, such as assessing the combined effects of hand assignment and structural salience on error rate, and evaluating the consistency and individuality of performers' error patterns. Finally, building on previous research on the production and perception of errors in music performance, we propose a theoretical model accounting for the effects of musical structure and expressive intent on error production.

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 synchronic, across-voice associations are emphasized, than in polyphonic pieces, which favor diachronic, 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). These

[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.

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. One aspect that has not been empirically examined so far is whether these effects would extend to piece-specific elements such as motives or themes. Performers could be expected to make less errors when playing motivic notes than non-motivic notes; likewise, listeners would be expected to be more sensitive to errors in motivic passages, especially if a motive or theme is familiar or easily recognizable. The latter hypothesis is supported by observations from DeWitt & Samuel (1990) who showed that listeners discriminated better between original and modified versions of familiar melodies than of unfamiliar ones. We tested the former by analyzing performances of the Dorian fugue, in which recurring thematic passages are clearly delineated.

Regarding the effect of the performer's expressive intent 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. However, errors were found to be less frequent in the highest voice regardless of the interpretative goal (Palmer & Van de Sande, 1993). 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 recently been documented at a pre-attentive level in electrophysiological studies (Fujioka, Trainor, Ross, Kakigi, & Pantev, 2005). However, these observations are marred by enculturation effects: since the main melody occurs more often in the upper voice in the Western musical repertoire, performers and listeners may both be predisposed to pay more attention to the highest voice (Palmer & Holleran, 1994; Palmer & Van de Sande, 1993). Moreover, earlier studies showing differential error rates examined three-voice textures (Palmer & Van de Sande, 1993), which could be a potential confounding factor since the left hand played two voices in most of these excerpts. One way to avoid such confounds is to analyze the distribution of performance errors for a piece in which the onset density is similar for each hand and for which the thematic material is distributed more or less equally among all parts. Fugues, in which the thematic material (subject and countersubject) are successively introduced in the different parts, constitute an ideally suited genre to carry out such an analysis. All pieces analyzed in this article contained two parts in each hand, and one (the Dorian fugue) included a pedal part, allowing for a more extensive analysis of potential relationships between melodic emphasis, pitch height, and limb assignment.

Error patterns are also dependent on the performer's level of competence: relationships between the frequency and distribution of errors and the level of musical competence, as well as the amount of practice, were evinced from studies on skill acquisition in music performance (Drake & Palmer, 2000; Palmer & Drake, 1997). It has been proposed that one of the main differences between expert and amateur performers lies in practice efficiency and in the use of metacognitive strategies (Hallam, 1997, 2001). If this hypothesis also holds true for professional performers, and if reduced error rate is one of the outcomes of efficient practice, we would expect to see a larger difference between the error rates of prize-winning organists compared to non-winners for a well-prepared performance than in a quick-study situation. This hypothesis was tested by comparing the error rates of prize-winners and non-winners for the *Premier Agnus* and the *Wachet auf*, which were performed in a quick-study condition, and for the Dorian fugue, which was a prepared piece.

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 & Goebl, 2004). Similar results for organ performance are reported in Chapter 2. 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, since the database included repeated performances of all 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 visual perception and cognition of the score by the performer, the kinematic level (motion of the performer), the mechanical level (the generation of the sound by the instrument), and finally the perception of the performance by the listener. In this study, we analyzed errors observed at the mechanical level, that is, errors registered in terms of key-depression events recorded in MIDI (Musical Instrument Digital Interface) format. One advantage of such an 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). Indeed, if we assume that errors that are less noticeable are more likely to occur, the distribution of errors may indirectly reveal something about their perceptibility.

The 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). It should be noted, however, that not all of these deviations should be defined as errors, since 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 (Repp, 1996a), such as pitch errors (playing a note with the wrong pitch), omissions (failure to play a note that is in the score), and intrusions (playing extraneous notes that are not in the score). To these categories, we also added timing errors; however, since expressive timing is one of the main artistic licenses used in music performance, only large timing deviations (more than 150 milliseconds) were counted as errors[2].

[2] Expressive onset asynchronies in keyboard performance, even exaggerated ones, are typically smaller than 100 ms (Goebl, 2001; Repp, 1996c).

In the context of this study, a distinction was made between *score errors*, which comprise pitch errors (also called substitutions), omissions (including "added ties" – repeated notes in the score that were not re-attacked in performance), and timing errors, and *non-score errors*, which include all performance notes that are extraneous to the score, such as intrusions and repetitions (re-attacked notes in performance that were not repeated in the score).[3] This distinction is important because score errors can be assigned to a specific note, allowing a characterization by voice, position, and limb assignment, whereas non-score errors cannot easily be 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 distribution 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 pieces were selected for this study. In the quick-study condition, organists recorded a short French Baroque polyphonic piece, the *Premier Agnus* by Nicolas de Grigny and a short German Baroque homophonic piece, a chorale

[3] "Untied" notes (Repp, 1996a) are treated as repetitions.

setting of *Wachet auf, ruft uns die Stimme* (SSWV 534) by Samuel Scheidt. In the prepared condition, performers recorded the organ fugue in D minor (BWV 538), also known as the "Dorian" fugue, by Johann Sebastian Bach. 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.

Performers

All performers were professional organists from the Montreal area, or organ students at McGill University in Montreal.

For the *Premier Agnus*, eight organists (two female, six male; aged 23-30 years) participated in the study. All participants identified themselves as right-handers. They had received organ instruction for a mean duration of 10 years (range = 7-13 years) and had 8 to 21 years of experience playing the organ. All of them held or had held a position as church organist for an average of 8 years (range = 1-21 years). Three of them had previously won one or more prizes at national or international organ competitions.

For *Wachet auf*, eight organists (two female, six male; aged 19-30 years) participated in the study. All participants identified themselves as right-handers. They had received organ instruction for a mean duration of 9 years (range = 3-13 years) and had 4 to 21 years of experience playing the organ. All of them held or had held a position as church organist for an average of 8 years (range = 1-21 years). Three of them had previously won one or more prizes at national or international organ competitions.

Sixteen organists (two female, fourteen male; aged 24-59 years) recorded the Dorian fugue. 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 had 8 to 47 years of experience playing the organ. All of them held or had held a position as church organist for an average of 18 years (range = 4-39 years). Nine of them had previously won one or more prizes at national or international organ competitions.

Procedure

In the quick-study condition, scores were given to the organists 20 minutes before the recording session began, in order to give them time to practice on the organ. None of the performers were familiar with the 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 manuals (that is, the pedal was not used). Organists were paid $20 for their participation.

For the polyphonic piece (*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 the homophonic piece (*Wachet auf*), two different interpretations were recorded. Performers were asked to record two expressive renditions of the piece, followed by two mechanical 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).

In the prepared condition (Dorian fugue), organists were given 20 minutes to practice, after which they made two recordings of the piece. The choice of the piece was communicated to performers several weeks in advance. Most organists were familiar with this piece. No directives were given regarding the interpretation. Use of the pedal and of the manuals was necessary for this piece. Organists were paid $30 for their participation.

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 pedal-board) 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. For the pieces in the quick-study condition, the stops used were the Spitz Principal 8', the Spitz Principal 4', and the Fifteenth 2' on the Great manual. For the prepared condition, the registration was as follows: Open Diapason 8', Violin Diapason 8', Octave 4', and Fifteenth 2' on the Great manual; Diapason 8', Hohlflute 8', Oboe 8', Octave 4', Mixture 2' IV on the Swell manual; Bassoon 16', Open Diapason 8', Principal 4'on the Choir manual; Open Diapason 16', Principal 16', Principal 8', Choral Bass 4' on the pedal. The Swell was coupled to the Great, while the Choir was coupled to the pedal.

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

The audio signal was recorded through two Boehringer ECM 8000 omnidirectional microphones. The audio and MIDI signals were sent to a PC computer through a MOTU audio interface. Audio and MIDI data were then recorded using Cakewalk's SONAR software and stored on a hard disk.

Data analysis

Performance notes obtained from the MIDI data were matched to score notes, using an algorithm developed in MATLAB for this project (see Chapter 6). The error analysis was part of the matching process and thus completely automated. Hand and voice assignments of score notes were determined by the first author, a music theorist and church organist.

RESULTS

General observations

Error frequencies and percentages. The frequencies and percentages of the different error types are summarized in Table 4.1, which also lists the total number of score notes and notes actually played for each piece. The frequency and percentage of added ties, which refer to notes that were repeated in the score but not re-attacked in performances, should also be considered in proportion to the number of repeated notes in the score of each piece (Table 4.2). Global score error rates were highest in *Wachet auf* and lowest for the Dorian fugue. These differences appear to be linked to discrepancies in the rate of added ties and in the proportion of repeated notes in the score. One possible 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, since the organ sound is continuous, and performers are perhaps less mindful of re-striking repeated notes. We also observed in the quick-study condition that the frequency of non-score errors was higher for the homophonic piece (*Wachet auf*) than for the polyphonic piece (*Premier Agnus*), although the two pieces were of comparable levels of difficulty. 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).

Table 4.1. Error frequencies and percentages.

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

† The number of notes played far exceeds the number of score notes for the Dorian fugue since the performances were heavily ornamented.

Table 4.2. Frequencies and percentages of added ties.

Piece	Total repeated notes in score	Frequency of added ties	Percentage of added ties
Premier Agnus	1,008 (6.56%)	98	9.72%
Wachet auf	1,760 (14.91%)	222	12.61%
Dorian fugue	1,856 (2.15%)	75	4.04%

Note. Frequencies and percentages are computed on the aggregate data of all performances of a given piece. Repeated notes are expressed as percentages of all score notes and added ties as percentages of all repeated notes.

Order of recording. Since performers made several recordings of each piece, the order of recording could potentially be a confounding factor for statistical analyses involving comparisons of error rates across interpretations, especially in the quick-study condition where the error rate might be hypothesized to decrease as participants became more familiar with the pieces. In order to examine this effect, repeated-measures analyses of variance were conducted on the total error frequency (combined score and non-score 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, either in the quick-study condition (*Premier Agnus*, $F(5, 35) = 1.31$, $p = .30$, Greenhouse-Geisser epsilon = 0.42; *Wachet auf*, $F(3, 21) = 1.49$, $p = .25$, Greenhouse-Geisser epsilon = 0.82) or in the prepared condition (Dorian fugue, $F(1, 15) = 0.78$, $p = .39$), suggesting that, at the time of recording, performers had achieved a stable error rate that was not demonstrably influenced by the order of recording. Order of recording will thus not be considered in subsequent analyses.

Onset density. Among general factors affecting error frequency, it seemed likely that the number of score notes played simultaneously (or *onset density*) would have an effect, with higher error rates per note for score events with a higher onset density (Repp, 1996a). Indeed, for all three pieces, the mean error frequency (combining score and non-score errors) per score event, normalized for the number of notes per score event, was weakly but positively correlated with onset density, with coefficients of 0.17 ($df = 146$, $p < .05$), 0.24 ($df = 143$, $p < .01$), and 0.08 ($df = 1382$, $p < .01$) for the *Premier Agnus, Wachet auf*, and the Dorian fugue, respectively.

Effects of note position and saliency

Only score errors were used for the analysis of effects of note position on error rates by voice and hand, because they could be unambiguously assigned to a specific note in the score and therefore to a specific voice or hand, unlike most non-score errors. The effects of note position analyzed in this article include voice and hand (or limb) assignment, as well as voice position (outer versus inner voices) for all three pieces, and motivicity (notes belonging to recurring thematic or motivic material versus notes that do not) in the case of the Dorian fugue. Separate analyses will be presented for all three pieces, followed by a brief discussion synthesizing the results.[5]

[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

Statistical considerations. The analyses presented in this section involve comparisons of error rates for different structural categories of notes (for instance, notes belonging to outer voices versus notes belonging to inner voices). Logistic regression models, which predict the error rate for each score note according to its structural characteristics, were applied to these analyses. Individual effects associated with each performer were also modeled. In addition, since onset density was shown to influence error rate, it was included as a covariate in order to take its effect into account, although onset densities were similar for most structural categories considered here (Table 4.3). Therefore, repeated-measures logistic regression analyses, with onset density as a covariate, were conducted for all comparisons involving error rates for different structural categories.

Table 4.3. Mean onset densities for different structural categories of notes.

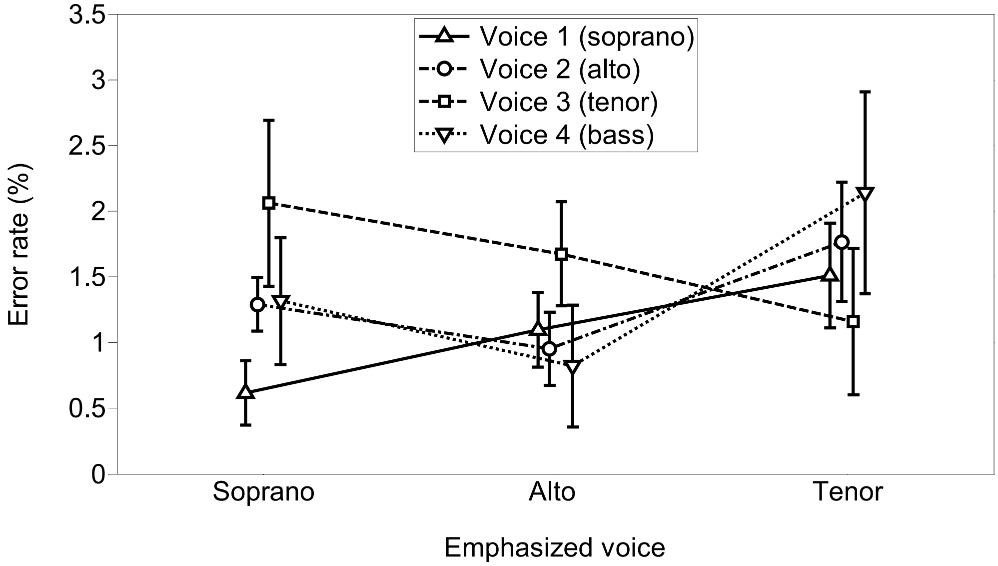
Structural category	*Premier Agnus*	*Wachet auf*	Dorian fugue†
Soprano	2.65	3.35	2.30
Alto	2.59	3.05	2.28
Tenor	2.45	3.04	2.29
Bass	3.39	3.17	2.66
Right hand	2.65	3.23	2.35
Left hand	2.71	3.05	2.28
Outer voices	2.86	3.26	2.30
Inner voices	2.54	3.03	2.44

† The onset density for the bass voice is equivalent to the onset density for the pedal. Mean onset density for motivic notes: 2.26; for non-motivic notes: 2.43.

comprise a very small fraction of the total number of score notes, were not included in the analyses by voice subsequently presented.

Premier Agnus. Three different interpretations were recorded for the *Premier Agnus*: in one interpretation, organists were asked to emphasize the soprano part, in another, the alto part, and in a third one, the tenor part. Following earlier studies which reported that errors were less likely to occur in the voice intended as melody than in nonmelody voices (Palmer & Van de Sande, 1993), it was hypothesized that error rates would be lower for the emphasized voice than for the non-emphasized ones. 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 voice and interpretation, $\chi^2(6) = 85.27$, $p < .001$. This result indicates that while the global error rate 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 4.1). A similar interpretation could be made for the logistic regression analysis on error rate per hand, which showed no main effect of hand or interpretation, but a significant interaction between these factors, $\chi^2(2) = 33.66$, $p < .001$. On the other hand, the logistic regression on error rate per voice position showed a main effect of voice position, $\chi^2(1) = 4.23$, $p < .05$, and a significant interaction between voice position and interpretation, $\chi^2(2) = 13.59$, $p < .01$, indicating that while error rates were generally lower for outer voices, this effect was modulated by the interpretation. 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) Error rates by voice



b) Error rates by hand and voice position

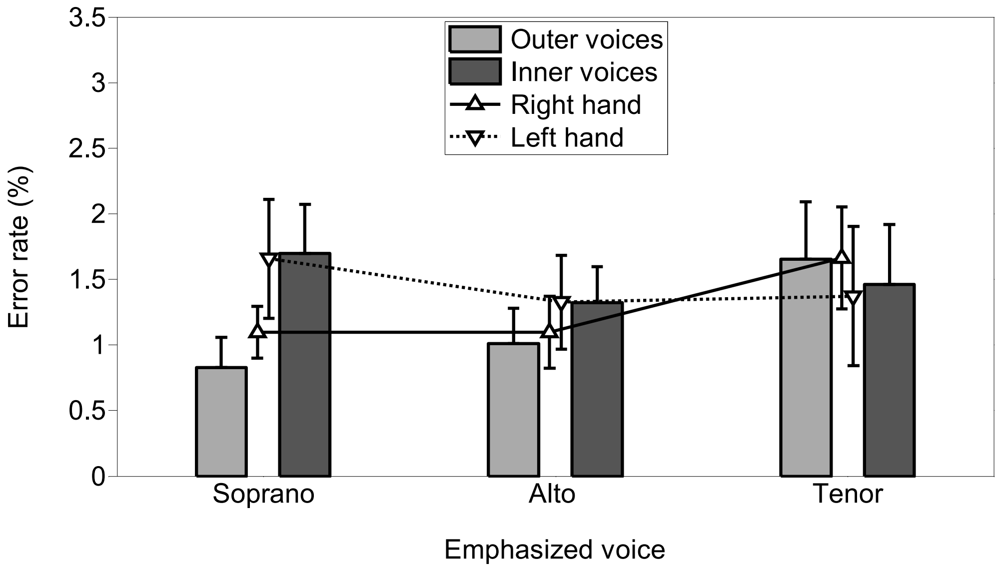
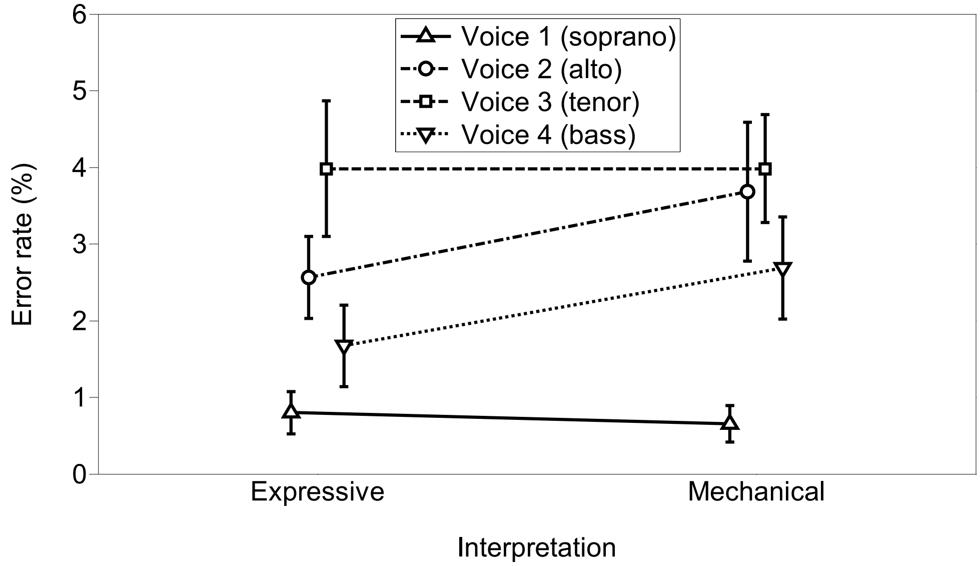


Figure 4.1. Effect of voice emphasis on error rate for the *Premier Agnus*. Mean error rates (in %) averaged across performers. Error bars represent standard errors of the mean. a) Error rates by voice. b) Error rates by hand and voice position.

Wachet auf. Performers recorded two different interpretations of *Wachet auf*: an expressive interpretation of the piece, followed by an unexpressive or mechanical one. In contrast to the instructions provided for the *Premier Agnus*, these instructions did not imply specific contrasts in melodic emphasis, and it was consequently 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, $\chi^2(3) = 41.20$, $p < .001$, no effect of interpretation, and no significant interaction. This analysis indicates that there was a significant difference in error rates between voices, and that interpretation did not influence the distribution of errors between voices (Figure 4.2). Using the same statistical model, logistic regression analyses were conducted on error rates by hand, showing no main effect or interaction, and by voice position, showing a significant effect of voice position, $\chi^2(1) = 20.05$, $p < .001$, and no other effect. 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).

a) Error rates by voice



b) Error rates by hand and voice position

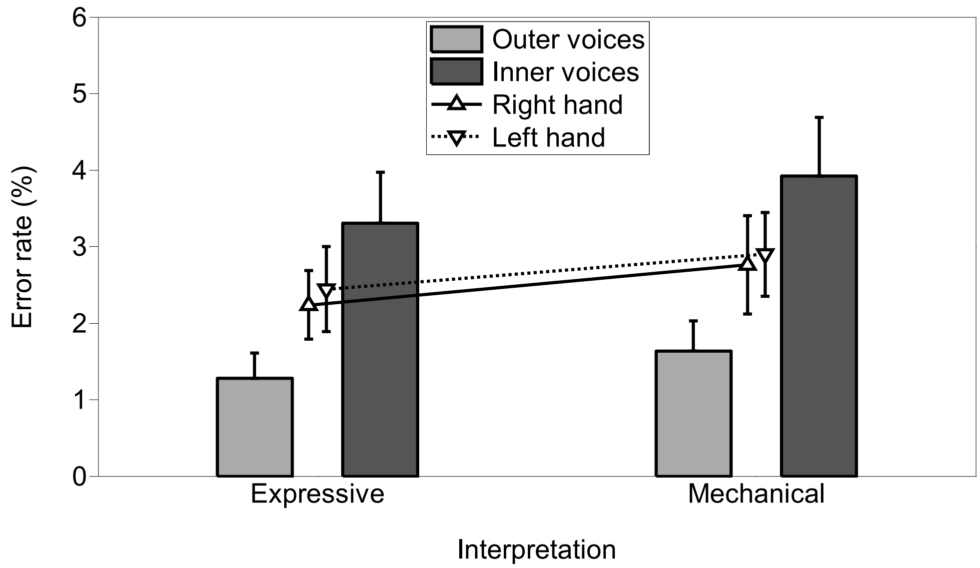


Figure 4.2. Effect of interpretation on error rate for *Wachet auf*. Mean error rates (in %) averaged across performers. Error bars represent standard errors of the mean. a) Error rates by voice. b) Error rates by hand and voice position.

Dorian fugue. Performers made two recordings of the Dorian fugue. Unlike pieces in the quick-study condition, no directives were given regarding the interpretation. The distribution of errors could therefore be expected to follow the pattern observed in earlier studies with lower error rates for the highest voice and for outer voices (Palmer & Van de Sande, 1993). However, in comparison with the quick-study pieces, the Dorian fugue is a much more complex piece, both in terms of length and motivic richness, and it requires performers to use the pedal. The potential interplay between voice position, limb assignment, and motivicity on error rates was therefore subjected to a detailed analysis. Since this piece is a fugue, motivic material is distributed among all the voices. 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). Following previous observations on differential error rates for melody versus nonmelody voices, the error rate was expected to be lower for motivic notes than for non-motivic ones.

Separate repeated-measures logistic regression analyses on error rates were conducted for voice, limb assignment, voice position, and motivicity, with onset density as covariate for all cases, showing significant effects of voice, $\chi^2(3) = 33.76$, $p < .001$, limb assignment, $\chi^2(2) = 33.25$, $p < .001$, voice position, $\chi^2(1) = 107.76$, $p < .001$, and motivicity, $\chi^2(1) = 31.46$, $p < .001$. As expected, error rates were lower for the highest voice and for outer voices (Figure 4.3). Error rates were also significantly lower for motivic notes than for non-motivic ones, thus confirming our hypothesis. Finally, error rates were higher for the left hand than for the right hand or the pedal.

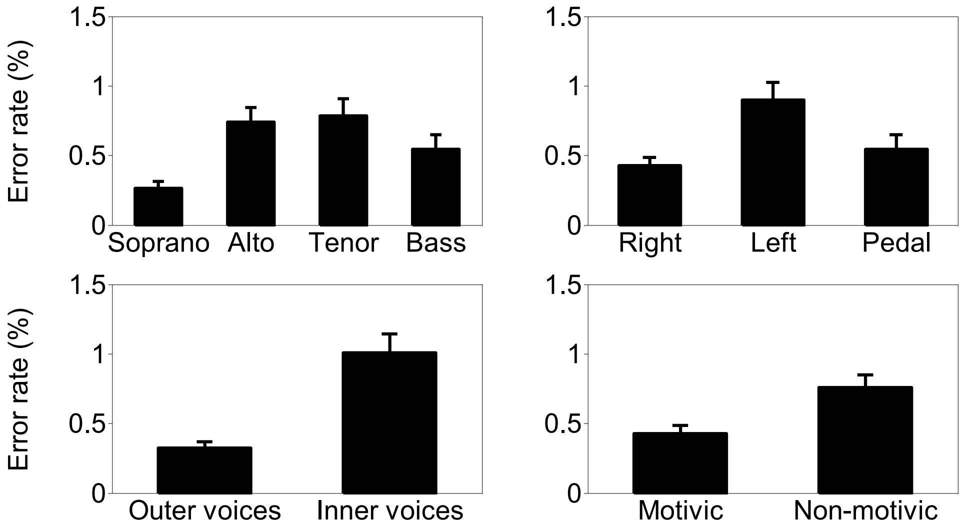


Figure 4.3. Error rates for different structural note categories for the Dorian fugue. Mean error rate (in %) for all categories, averaged across performers. Error bars represent standard errors of the mean.

However, all these comparisons implicitly assume independence between these effects, which is not the case in this piece. First, 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). Second, all pedal notes belong to an outer voice in this piece.[6] The effects of voice position (and, by extension, those related to specific voices), motivicity and limb assignment are thus interdependent to a certain extent. A more rigorous statistical treatment of these effects would consider the combined effects of voice position and motivicity and would exclude the pedal part from analyses considering interactions between voice position and limb effects. A repeated-measures logistic

[6] Notes in the pedal part sound one octave lower than written on the score since they are played on 16’ stops.

regression on error rate by voice position and motivicity confirms that the effects of voice position ($\chi^2(1) = 75.27$, $p < .001$) and motivicity ($\chi^2(1) = 11.68$, $p < .001$) were less pronounced when considered together than in isolation, as shown by a comparison with the chi-square values reported in the previous paragraph.

An analysis that combined the effects of voice position, motivicity, and hand assignment (excluding pedal notes) in a single model predictably yielded a more complex picture, with main effects of voice position and motivicity (but no effect of hand) and significant interactions between hand and position, as well as hand and motivicity (Table 4.4). While 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 4.4).

Table 4.4. Repeated-measures logistic regression on error rates for the Dorian fugue (with onset density as covariate).

Source	*df*	χ^2	*p*
Voice position	1	110.92	< .001
Motivicity	1	8.58	< .01
Hand	1	1.27	.26
Voice position × Motivicity	1	0.44	.51
Voice position × Hand	1	6.33	.01
Motivicity × Hand	1	14.38	< .001
Voice position × Motivicity × Hand	1	3.49	.06

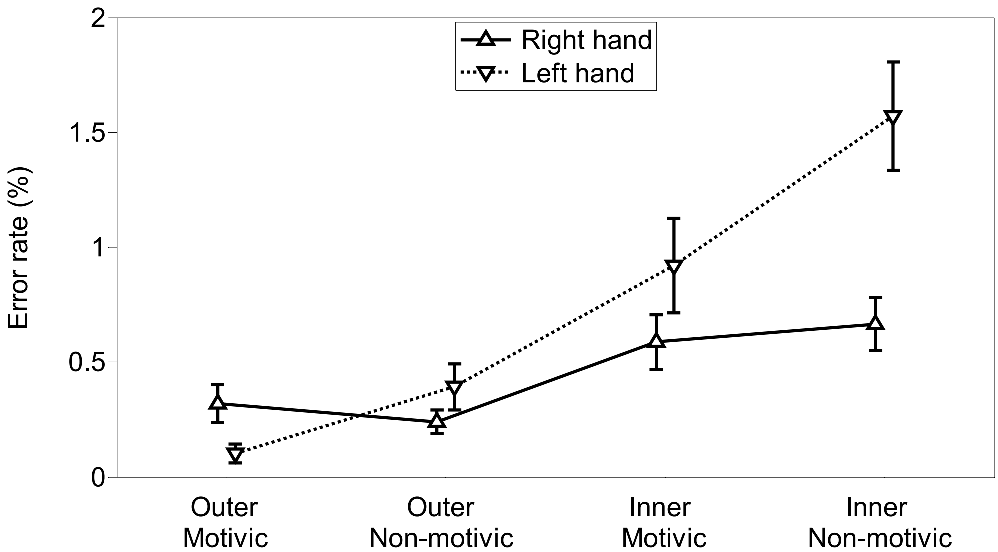


Figure 4.4. Effects of voice position, motivicity, and hand assignment on error rates for the Dorian fugue. Mean error rates (in %) for all categories, averaged across performers. Error bars represent standard errors of the mean.

Discussion. The results reported here generally replicate, over a large database of "ecological" performances and on a different instrument, earlier findings regarding keyboardists' tendencies to make fewer errors in the voice emphasized or intended as main melody, and in the highest voice as well as in outer voices. However, while 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 or 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 structurally salient notes (such as notes belonging to a recurring motive or to an outer voice), 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 skill level of the performers or in the experimental setup: this study used ecological performances, while Palmer & Van de Sande (1993) elicited errors by asking performers to use faster tempi. However, the differential effects of voice position and motivicity by hand assignment observed for the Dorian fugue suggest that the right-hand advantage can be probably best 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 structurally salient notes, this model would fit nicely with our observations on the Dorian fugue. Indeed, it seems that there was no clear right-hand advantage in terms of error rates for salient notes, while the left hand was at a clear disadvantage for less salient notes. It should be noted that 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; such a project was beyond the scope of the present article.[7]

Effects of musical texture

Palmer and Van de Sande (1993) had previously shown that musical texture influenced the type of errors: the proportion of harmonically related errors was higher for homophonic pieces than for polyphonic pieces. In this study, we analyzed the effect of musical texture on two error types, namely pitch errors (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 quick-study performances of a mostly homophonic piece (*Wachet auf*) and of a polyphonic piece (*Premier Agnus*). These two pieces are of equivalent levels of difficulty and similar length, with a mostly four-voice texture throughout (average number of active voices per score event, or *voice density*: 3.98 for both pieces), thus providing an adequate basis for comparison.

Empirical evaluation of the texture of a piece. Since onset and offset asynchrony are considered a hallmark of contrapuntal writing (Huron, 1993; Wright & Bregman, 1987), one way to compare the textures of two multivoiced pieces is to evaluate the number of concurrent rhythmic streams per active score event, with each stream corresponding to a note (or group of notes) whose onset and/or offset are not synchronous with those of other notes present in the same

[7] Note that the performances of the fourteen 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.

score event.[8] The number of concurrent rhythmic streams is thus bounded by definition between 1 and the total number of notes present in each active score event, with a low number of rhythmic streams corresponding to a homophonic texture.[9] The mean number of rhythmic streams, normalized for the duration of each score event, was estimated at 2.82 for the *Premier Agnus*, and 2.19 for *Wachet auf.* A one-tailed Mann-Whitney test on the number of concurrent rhythmic streams per active score event confirmed that there are significantly more streams per score event in the *Premier Agnus* than in the *Wachet auf* (U (148, 145) = 13,209, $p < .001$), even though the voice density of both pieces is similar, thus providing an indirect confirmation of the music-theoretical intuition that this piece is more polyphonic in character.

Analysis of error types. Pitch and intrusion errors were categorized in 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

[8] A score event is defined by a change in the texture of the piece brought about by the onset or offset or one of more notes. An active score event is a score event in which at least one voice is active.

[9] Note that this definition purposely avoids any reference to the *pitch content* of a piece, which makes it theoretically applicable to any multivoiced texture, regardless of its compositional style.

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 4.5 shows that the proportion of melodically related errors was greater in the polyphonic piece (*Premier Agnus*) than in the homophonic piece (*Wachet auf*), while 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, $\chi^2(3) = 8.49$, $p < .05$. 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* ($T = 1$, $p < .05$), but not for the *Wachet auf* ($T = 16$, $p = .85$). These results indicate that texture influenced the type of errors: the proportion of harmonically related errors was greater in a homophonic texture (*Wachet auf*) than in a polyphonic texture (*Premier Agnus*), and performers made less 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.

Table 4.5. Effect of musical texture on the type of pitch and intrusion errors.

	Harmonically-only related	Melodically-only related	Harmonically & melodically related	Harmonically & melodically unrelated	Chance estimates
Premier Agnus (Polyphonic)	8 (8.6%)	34 (36.6%)	8 (8.6%)	43 (46.2%)	3.2:12 (26.6%)
Wachet auf (Homophonic)	25 (18.2%)	36 (26.3%)	22 (16.1%)	54 (39.4%)	3.1:12 (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.

Effect of performer expertise

Hallam (1997) suggested that one of the main differences between expert and amateur performers lies in practice efficiency. If a reduction in error rate is one of the outcomes of efficient practice, we would expect to see a larger difference between the error rates of prize-winning organists versus non-winners for a well-prepared performance than in a quick-study situation. This hypothesis was tested by comparing the error rates of prize-winners and non-winners for the *Premier Agnus* and the *Wachet auf*, which were performed in a quick-study condition, and for the Dorian fugue, which was a prepared piece.

Repeated-measures analyses of variance were conducted on the total number of errors per performance with level of accomplishment (prize-winners versus non-prize winners) as a between-subjects factor, for all three pieces. Although level of accomplishment had no significant effect on error rate in quick-study conditions, $F(1, 6) = 0.43$, $p = .54$ for *Wachet auf* and $F(1, 6) = 0.54$, $p = .49$ for the *Premier Agnus*, prize-winners made significantly fewer errors than non-winners in the prepared condition, $F(1, 14) = 5.43$, $p < .05$ for the Dorian fugue. There are several potential explanations for this result. One is that prize-winners make better use of their practice time than non-winners, as previously suggested. Another is that performance degradation under stress may be lower for prize-winners than for non-winners (see Wan & Huon, 2005, for a discussion of performance degradation); self-expectations were possibly higher for the prepared piece than in the quick-study condition, for which performers had only 20 minutes to prepare the piece. Finally, it is worth noting that in most performance competitions, contestants presumably are awarded competition prizes on the basis of the quality of their prepared performances; because sight-reading or quick-study abilities are rarely directly evaluated in competitions, it should not necessarily be assumed that prize-winners perform better than non-winners in these conditions.

Consistency and individuality of error patterns

As previously mentioned, high-level performers exhibit a high degree of consistency in their use of temporal patterns, as well as in their patterns of articulation, of variation in intensity, and of onset asynchronies (Palmer, 1989; Repp, 1992, 1996b, 1996c; Widmer & Goebl, 2004). In order to test whether this was also the case for performance errors, all pairs of performances were compared by tabulating the frequency of the co-occurrence of errors in the same score event in different performances; both score and non-score errors were included in this analysis.[10] Phi coefficients were computed as a measure of the degree of concordance between the error patterns of each pair of performances. For all three pieces, the majority of phi coefficients for within-performer comparisons were highly significant (Table 4.6). In addition, phi coefficients were significantly higher for comparisons between pairs of performances played by the same performer than between performances played by different performers. These analyses demonstrate that performers exhibited both consistency and individuality in their error patterns, although the coefficients were not as high as those reported for tempo, articulation, or onset asynchrony patterns.

Table 4.6. Mean phi coefficients for error patterns between all pairs of performances for all three pieces.

	Premier Agnus (*df* = 147)				*Wachet auf* (*df* = 150)				Dorian fugue (*df* = 1382)			
	pairs	mean	*SD*	%**	pairs	mean	*SD*	%**	pairs	mean	*SD*	%**
Within	103	0.22	0.21	53.4	48	0.39	0.16	83.3	16	0.25	0.17	100.0
Between	843	0.10	0.17	25.7	448	0.15	0.14	32.8	480	0.04	0.05	26.0
H_1: $\mu_{within} > \mu_{between}$	U = 59,134.5, $p < .001$				U = 18,498, $p < .001$				U = 7,339, $p < .001$			

Note. Phi 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. One-tailed Mann-Whitney tests were conducted to assess

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

whether the within-performer coefficients were significantly higher than the between-performer coefficients. %**: percentage of highly significant coefficients ($p < .01$).

DISCUSSION

Several results presented in this article indicate that performers' error patterns are modulated to a large extent by the local musical context, such as the position or musical relevance 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.

As mentioned previously, listeners' sensitivity has been shown to be 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). Furthermore, listeners are more proficient at detecting changes in a familiar melody than in an unfamiliar one (Dewitt & Samuel, 1990). These complementary observations regarding the production and detection of performance errors suggest that the performers' and listeners' mental representations of the score, in terms of the relative perceptual and musical salience of structural note categories, are well-matched. 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 salience and musical significance or familiarity.

Performers' mental representations of a musical score are flexible: when asked to play different interpretations of the same piece in which they emphasize specific voices, performers made fewer errors in a given voice when it was emphasized than when it was not. This 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 performance 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, while 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. In light of these observations, it is worth mentioning that, whether by design or by accident, the frequency mapping of the keyboard takes into account both cognitive-motor and perceptual constraints, given the predominance of right-handers in the population; indeed, whereas naïve left-handers have been found to prefer reverse keyboards, right-handers prefer the normal configuration regardless of their musical experience (Laeng & Park, 1999).

Although performance errors are clearly determined in large part by the musical structure, we have shown that they are also, to some extent, performer-specific. Indeed, error patterns of performances of the same piece played by the same organist were more similar than those of recordings by different organists, indicating that individual performers exhibited both consistency and individuality in their error patterns. While 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 influenced by structural considerations (Repp, 1998), as shown by performers' final-phrase lengthening tendencies and listeners' context-dependent ability to detect temporal changes, and 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 consistent and idiosyncratic error patterns of individual performers.

As this discussion has exposed, 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.

ACKNOWLEDGMENTS

This research was supported by fellowships from the Social Sciences and Humanities Research Council of Canada and from the Centre for Interdisciplinary Research in Music Media and Technology (CIRMMT) to Bruno Gingras, as well as a grant from the Natural Sciences and Engineering Research Council and a Canada Research Chair awarded to Stephen McAdams. Rhonda Amsel provided helpful suggestions regarding the statistical analysis. We thank Julien Boissinot, Darryl Cameron, and Bennett 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.

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.

Drake, C., & Palmer, C. (2000). Skill acquisition in music performance: Relations between planning and temporal control. *Cognition, 74*(1), 1-32.

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.

Goebl, W. (2001). Melody lead in piano performance: Expressive device or artifact? *Journal of the Acoustical Society of America, 110*(1), 563-572.

Hallam, S. (1997). Approaches to instrumental music practice of experts and novices: Implications for education. In *Does practice make perfect? Current theory and research on instrumental music practice* (pp. 89). Norway: Norges Musikkhogskole.

Hallam, S. (2001). The development of metacognition in musicians: Implications for education. *British Journal of Music Education, 18*(1), 27-39.

Huron, D. (1989). Voice denumerability in polyphonic music of homogeneous timbres. *Music Perception, 6*(4), 361-382.

Huron, D. (1993). Note-onset asynchrony in J. S. Bach's two-part inventions. *Music Perception, 10*(4), 435-444.

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., & Drake, C. (1997). Monitoring and planning capacities in the acquisition of music performance skills. *Canadian Journal of Experimental Psychology-Revue Canadienne De Psychologie Experimentale, 51*(4), 369-384.

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 "Träumerei" 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), 155-172.

Widmer, G., & Goebl, 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.

Wright, J. K., & Bregman, A. S. (1987). Auditory stream segregation and the control of dissonance in polyphonic music. *Contemporary music review, 2*(1), 63-92.

APPENDIX: MUSICAL SCORES

a) Nicolas de Grigny, *Premier Agnus*

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

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

Chapter 5. The performer as analyst: A case study of J.S. Bach's "Dorian" fugue (BWV 538)

Chapter 5 aims to clarify the relationship between the performer's view of the piece as an analyst and as a performer, by examining whether performers whose written analyses substantially differed also emphasized distinct formal aspects in their performances of the Dorian fugue. This project seeks to describe more accurately the link between interpretative choices and musical structure from a music-theoretical perspective. Furthermore, this study explores a stylistic repertoire that has been relatively neglected in the literature on performance research, which has generally focused on Classical and Romantic piano literature.

This chapter is based on the following research article:

Gingras, B., McAdams, S., & Schubert, P. N. The performer as analyst: A case study of J.S. Bach's "Dorian" fugue (BWV 538). Manuscript prepared for submission to *Journal of New Music Research*.

ABSTRACT

This study sought to compare the performer's output as analyst and as performer. Sixteen professional organists were invited to perform J.S. Bach's organ fugue in D minor (BWV 538), also known as the "Dorian" fugue. Each performer recorded the fugue twice on an organ equipped with a MIDI console, which allowed precise measurement of performance parameters. Immediately after their performances, organists were invited to submit their own analyses of the piece by indicating its main formal subdivisions. A comparison of the written analyses indicated that, despite a fair amount of individual variation, performers generally agreed on the main structural boundaries of the piece. An analysis of the temporal profiles of the performances revealed that the largest tempo variations coincided with these structural boundaries. A multidimensional scaling analysis established that performers' temporal profiles varied across two main dimensions: one was related to the relative salience of the temporal variations associated with formal subdivisions, and another reflected the relative magnitude of the rallentandos corresponding to the multiple recurrences of a canonic episode in the piece. Although a significant correlation was found between the performers' degree of agreement on a formal subdivision and the average magnitude of the concomitant tempo deviation, no such correlation could be found within individual performers, suggesting that written analysis may not be the optimal strategy to determine the performer's analytical reading of a piece.

INTRODUCTION

Several studies have brought to the fore the relationship between music-theoretical analysis and performance (Berry, 1989; Cone, 1968; Narmour, 1988; Rink, 1995b, 2002; Schmalfeldt, 1985). Whereas scholars such as Berry and Narmour intimated that performers should be acquainted with the theoretical and analytical methodology proposed by theorists, these studies were met, perhaps understandably, with little interest from performers. Indeed, these authors conveyed a view that simultaneously relegated the performers to a role of simple practitioners who should heed advice from the theorist regarding the structure of the pieces they are performing, while putting structural concerns to the forefront of performance issues (Cook, 1999). More recently, however, Rink (1995a) and Lester (1995) have advocated a different view, one that gives value to the performers' analytical insights about a piece. Lester even went so far as to reverse the paradigm accepted by scholars by proposing that analysts should work from performances instead of working from the score. Leonard Meyer had already hinted at such a view in 1973 when he wrote that, although performance is the actualization of an analytical act, this analysis may very well be intuitive and unsystematic: "For what a performer *does* is to make the relationships and patterns potential in the composer's score clear to the mind and ear of the experienced listener" (Meyer, 1973, p. 29).

However, probing the analytical insights of the performer may prove to be a difficult task for several reasons. First, the analyst and the performer are rarely the same person; moreover, they seldom share the same language, in spite of

Schmalfeldt's (1985) compelling illustration of such an ideal situation. Second, as noted by Rothstein (1995), music-theoretical analysis and music performance have different goals, and it would be ill-advised to subsume one activity under the other. Third, investigating the performer's analytical insights as they are projected in performance necessarily entails a comprehensive exploration of the expressive dimensions of a performance, in order to determine which aspects of the musical structure were expressed and how they were conveyed.

The present study attempted to partially circumvent these problems by inviting performers to record a piece for which they were asked to provide their own written analysis and to compare their performances to their analyses. For this purpose, sixteen professional organists were invited to perform J.S. Bach's organ fugue in D minor (BWV 538), also known as the "Dorian" fugue, on an organ equipped with a MIDI console, after which they were invited to provide their written analysis of the piece by indicating its main formal subdivisions. This study intended to shed new light on the complex relationship between performance and analysis by giving preeminence to the actualized music rather than to score-based analytical readings, thus following Lester's advice to seek "ways in which analysis can be enhanced by explicitly taking note of performances, indeed by accounting them as part of the analytical premise" (Lester, 1995, p. 199). More precisely, it aimed to clarify the relationship between the performer's view of the piece as an analyst and as a performer by examining whether performers whose written analyses substantially differed also emphasized distinct formal aspects in their performances. To be sure, most performers' ability to report their analytical understanding of the piece in a written medium may not equal their capacity to express it in performance. However, by limiting the scope of the written analysis to the identification of large-scale formal subdivisions and comparing this to the performance, we hoped to gain substantial insights into the performers' formal conceptualizations of the piece. Furthermore, this study sought to explore a stylistic repertoire that has been relatively neglected in the literature on performance research, which has generally focused on Classical and Romantic piano literature.

An acknowledged masterpiece, the Dorian Fugue is one of Bach's most accomplished works for the organ (Figure 5.1). The New Grove Dictionary of Music and Musicians includes it among Bach's finest fugal works (Caldwell, 2007), whereas the eminent organ scholar Peter Williams mentions the "exceptional series of imitative episodes" that runs throughout the fugue, claiming that it "produces some of the most carefully argued four-part harmony in the organ repertoire" (Williams, 2003, p. 68-70). The piece is especially noteworthy for its pervasive motivic unity: indeed, most of the melodic material of the fugue, including the episodes, is derived from the first 16 measures of this 222-measure piece.

Figure 5.1. J.S. Bach, Fugue in D minor, BWV 538 ("Dorian" fugue), measures 1-29. Only the first appearance of the subject and of each countersubject is indicated. Grayed areas correspond to codettas.

Tempo variations as a marker of structural organization in performance

A large body of literature on performance research has established that performers tend to slow down at sectional boundaries or formal subdivisions of a piece (Clarke, 1985; Gabrielsson, 1987; Palmer, 1989; Repp, 1990; Shaffer, 1981). This expressive device has been termed *phrase-final lengthening*. Moreover, it has been shown that the magnitude of the ritardando corresponds to the hierarchical importance of the boundary, with larger tempo variations associated with the major formal subdivisions of the piece (Repp, 1992; Shaffer & Todd, 1987; Todd, 1985). Several scholars proposed that these tempo fluctuations are a means of conveying information about the grouping structure of a piece to the listener, a model known as the *musical communication hypothesis* (Clarke, 1985, 1988; Palmer, 1989, 1996; Repp, 1992, 1995). Clarke (1989) reported that listeners were sensitive to minute changes in timing (as little as 20 ms for inter-onset intervals between 100 and 400 ms). Palmer (1989) demonstrated that tempo fluctuations were, at least in part, under the performers' voluntary control, since they were smaller in mechanical performances than in expressive performances of the same piece, and they could be modified according to the performers' interpretation of the piece. Penel and Drake (1998) refined these findings by showing that performers had more control over higher-level timing patterns, which involve phrases or larger sections of a piece, than over local timing patterns, which consist of rhythmic groupings comprising only a few notes. More recently, Penel and Drake (2004) demonstrated that phrase-final lengthening could be accounted for partly by perceptual and motor constraints and partly by the musical communication model.

While further research is necessary to fully elucidate the role of phrase-final lengthening in expressive performance, there is sufficient evidence to posit a clear relationship between the timing variations applied by performers and the formal structure of the piece. Furthermore, it may be surmised, following Palmer's (1989) observations, that different interpretations of a piece would be characterized by different timing patterns. The present study, which was based on these assumptions, focused on the relationship between the temporal patterns employed by performers and their analytical readings of the Dorian fugue. The use of MIDI (Musical Instrument Digital Interface) technology, which has enabled the quantitative analysis of performance parameters, allowed an objective description of the interpretive details associated with each performance.

METHOD

Participants

Sixteen professional organists (two female, fourteen male; aged 24-59 years) were invited to participate in the experiment. All performers were professional organists from the Montreal area or organ students at McGill University in Montreal. Fourteen identified themselves as right-handers, one as a left-hander, and one as ambidextrous. They had received organ instruction for a mean duration of 10 years (range = 4-25 years) and had 8 to 47 years of experience playing the organ. All of them held or had held a position as church organist for an average of 18 years (range = 4-39 years). Nine of them had previously won one or more prizes at national or international organ competitions.

Procedure

The choice of the piece was communicated to performers at least four weeks in advance. Most organists were familiar with this piece. No directives were given regarding the interpretation. Before the recording session began, organists were given 20 minutes to practice, after which they made two recordings of the piece. Immediately after their performances, the organists were invited to fill out a questionnaire and submit their own analyses of the piece, indicating its main formal subdivisions. Organists were paid $30 for their participation. The entire experiment lasted approximately one hour.

Performances were recorded on the Casavant organ of the Church of St-Andrew & St-Paul in Montreal, Canada. This five-manual organ (five keyboards and a pedal-board) 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.[1] The following registration, which was established in consultation with the performers, was used for all recordings: Open Diapason 8', Violin Diapason 8', Octave 4', and Fifteenth 2' on the Great manual; Diapason 8', Hohlflute 8', Oboe 8', Octave 4', Mixture 2' IV on the Swell manual; Bassoon 16', Open Diapason 8', Principal 4'on the Choir manual; Open Diapason 16', Principal 16', Principal 8', Choral Bass 4' on the pedal. The Swell was coupled to the Great, while the Choir was coupled to the pedal.

The audio signal was recorded through two Boehringer ECM 8000 omnidirectional microphones. The audio and MIDI signals were sent to a PC computer through a MOTU audio interface. Audio and MIDI data were then recorded using Cakewalk's SONAR software and stored on a hard disk.

[1] Information provided by Mark Gilliam, Sales manager of Solid State Organ Systems.

Data analysis

The MIDI data from the performances was matched to a symbolic representation of the score, using a new matching algorithm that was specifically designed for this project (Chapter 6). This matcher allows a precise note-to-note mapping of a performance note to a score note. Furthermore, it identifies errors and recognizes ornaments. The use of automated methods was necessary since the score of this fugue contains 2701 notes.

RESULTS

Analytical readings of the Dorian fugue in the literature

Table 5.1 presents a detailed overview of the formal structure of the Dorian fugue. The main sections, as proposed by Williams (2003, p. 68-70), are indicated in Roman numerals, while recurring episodes are identified by letters, and cadences by the abbreviations PAC (for *perfect authentic cadence*) and IAC (for *imperfect authentic cadence*). Williams notes that "each middle entry is preceded by a strong perfect cadence" (p. 70); he also lists the fugue's recurring canonic episodes (identified as "Episode A" in Table 5.1), some of which produce striking verticalities which have been said to "defy harmonic analysis" (Bullivant, 1971, p. 104), as one of its unusual features (see Figure 5.2 for an example). These episodes, whose material is derived from the codetta of the exposition (see Figure 5.1), appear no less than 13 times in the fugue, each recurrence using different intervals of imitation. In addition to the association between cadences and subject entries noted by Williams, which underscores the role of cadences as sectional articulators, the exhaustive development of a motivic core presented in the opening measures, as well as the increasingly contrapuntally dense recurrences of the canonic episodes, all correspond neatly to Lester's model of heightening levels of activity in Bach's compositional process (Lester, 2001).

According to some scholars, the Dorian fugue contains a clear example of a counter-exposition: thus, Walker (2008) notes that "the four entries of alto (bar 43), soprano (57), tenor (71) and bass or pedal (81) can be said, by virtue of their entering in the same order as in the exposition but with exchanged starting notes, to constitute a counter-exposition"; a similar observation had already been made by Prout (1891, p. 148). Although his analysis does not explicitly identify a counter-exposition, we may assume that Wiliams does not consider the entries in mm. 43, 57, 71, and 81 as middle entries; in any case, these entries are not preceded by perfect authentic cadences.

Figure 5.2. Statement of the canonic episode in mm. 88-92 of the Dorian fugue. Note the dissonant character of the verticalities boxed in m. 90 and 91. Grayed areas correspond to the motive derived from the codetta.

Table 5.1. Overview of the formal structure of the Dorian fugue.

Section	Measure number	Structural function	Cadence
	1	Subject entry, alto (D minor)	
	8	Subject entry, soprano (A minor)	
	9		IAC D minor
	15	Codetta	
	18	Subject entry, tenor (D minor)	**PAC D minor**
	25	Codetta	
	29	Subject entry, pedal (A minor)	
	36	End of exposition; Episode A	
	43	Subject entry, alto (A minor)	
I	49	Episode A (derived from the codetta)	
	58	Subject entry, soprano (D minor)	IAC D minor
	64	Episode B (chromatic sequence)	
	67	Episode A	
	71	Subject entry, tenor (A minor)	
	78	Episode A	
	81	Subject entry, pedal (D minor)	IAC D minor
	88	Episode A	IAC D minor
	92	Episode C (derived from Episode A)	
	95	Episode A	
	101	Subject entry, stretto between soprano and pedal (F major)	**PAC F major**
	108	Episode C’	
	111	Episode A	
	115	Subject entry, tenor (C major)	**PAC C major**
	124	Episode A	
II	130	Subject entry, stretto between alto and tenor (G minor)	**PAC G minor**
	138	Episode A	
	146	Subject entry, tenor (B flat major)	**PAC B flat major**
	152	Episode D (ascending chromatic)	
	156	Episode A	
	160	Episode E (scalar passages in contrary motion)	
	162	Episode A	
	167	Subject entry, stretto between pedal and alto (D minor)	**PAC D minor**
	175	Episode B	
	178	Episode A (with pedal trill)	
	188	Subject entry, soprano (A minor)	**PAC A minor**
	194	Episode D’ (descending chromatic)	
III	197	Episode E	
	203	Subject entry, stretto between soprano and pedal (D minor)	**PAC D minor**
	204		IAC D minor
	211	Episode A	**PAC D minor**
	219	Dominant pedal in D minor; homophonic texture	
	222		**PAC D minor**

Note. Sections labelled following Williams’ analysis (2003, p. 68). Episodes are identified by letters. *IAC*: imperfect authentic cadence. *PAC*: perfect authentic cadence.

Performers' written analyses

On average, performers identified 7 boundaries (range: 3 to 16). A total of 21 different subdivisions were identified. Each of these boundaries was selected on average by 34% of the performers, with a percentage of agreement ranging from 93.8% (15 of 16 performers identifying a given measure as a boundary) to 6.3% (only one performer identifying a given measure as a boundary).[2] As can be seen in Figure 5.3, the four subject entries in stretto, on mm. 101, 130, 167, and 203 received the greatest agreement as structural boundaries; we note that m. 101 and 167 correspond to the beginning of sections I and II in Williams' reading of the piece. Approximately half of the performers also identified boundaries at mm. 36 (which corresponds to the end of the exposition), 81 (which corresponds to the last subject entry of the counter-exposition according to Walker), and 188. A number of formal subdivisions were mentioned only by one or two performers: these generally corresponded to the beginning of episodic sections (m. 64, 88, 138, 162, 211) or to subject entries which were not preceded by cadences (m. 43 and 71).

Comparing analysis and performance

General overview of the performances. Since each organist recorded two performances, a total of 32 performances were analyzed. Global tempi ranged from 41 to 61 beats per minute (BPM), with a mean global tempo of 52 BPM (the

[2] Boundaries marked within a range of two measures were considered to be the same; such variability was observed only for two boundaries (m. 57-58 and m. 203-204), these markings were conflated together to measure 58 and 204 respectively. All other formal subdivisions were assigned to the same measure by all performers who indicated them.

half note was taken as the beat since the piece is written in cut time). In comparison, Jerkert (2004) found tempi ranging from 52 to 64 BPM in CD recordings of the Dorian fugue from four internationally known organists. The error rate (wrong notes or missing notes) was very low: the mean error rate (wrong notes and missing notes) across all performances was 0.44%, and 31 of the 32 performances had less than 1% of errors. Performances were heavily ornamented: 7.6% of all performance notes were identified as ornamental, for an average of 18 ornaments per performance (mostly trills).

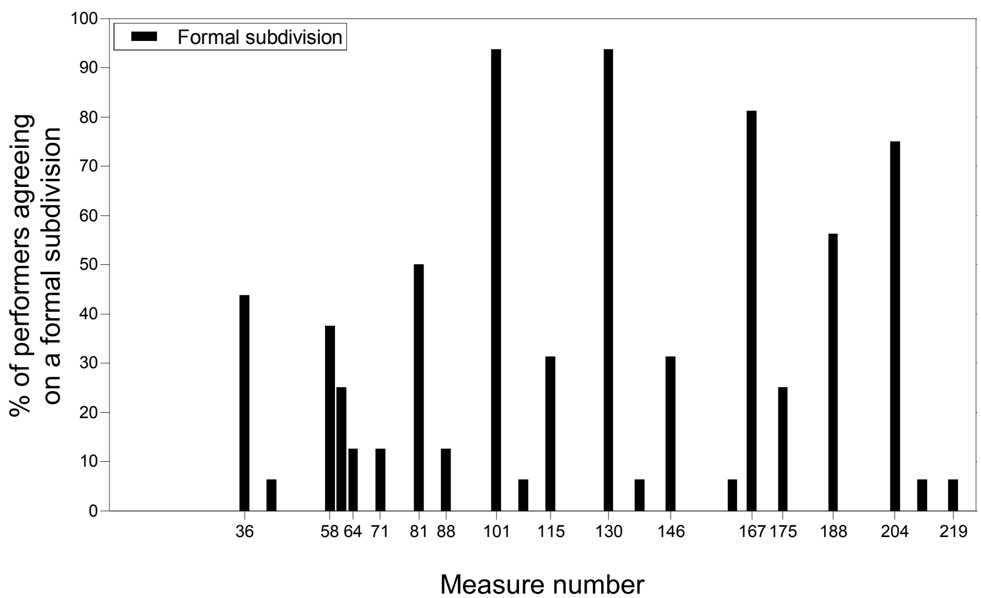


Figure 5.3. Performers' identifications of formal subdivisions in the Dorian fugue.

Analysis of the temporal profiles of the performances. For each performance, the local tempo was computed for each quarter note. The quarter note was chosen as a unit since note onsets can be found on practically each quarter note beat throughout the piece, except for the first 8 measures. Temporal profiles were thus obtained for each performance. High correlations were observed between the temporal profiles; the mean correlation between all pairs of performances was 0.65 (*SD* = 0.10, *df* = 887), with higher correlations for performances played by the same performer (mean correlation: 0.84, *SD* = 0.10) than for performances played by different performers (mean correlation: 0.64, *SD* = 0.09). These results indicate that there was a high degree of similarity among the temporal profiles of different performers. In order to examine general tendencies across performances, a "typical" temporal profile was generated by averaging local tempo values for each quarter note over all 32 performances.

For the most part, the most important rallentandos, characterized by a sharp decrease in the tempo, coincided with authentic cadences (indicated by dotted lines in Figure 5.5). On the other hand, a number of important rallentandos corresponded to features which may not be considered by music theorists as main formal subdivisions of the piece (although some performers identified them as such), such as the recurrences of Episode A in mm. 78 and 138 or the dominant pedal in m. 219. The important rallentando observed at m. 196 could be related to the performers' phrasing of the scalar passages of episode E. However, considering that both hands have to skip an octave at the very beginning of m. 196 (the only passage in the fugue which presents such a difficulty), it is likely due in part to motor constraints (Figure 5.4).

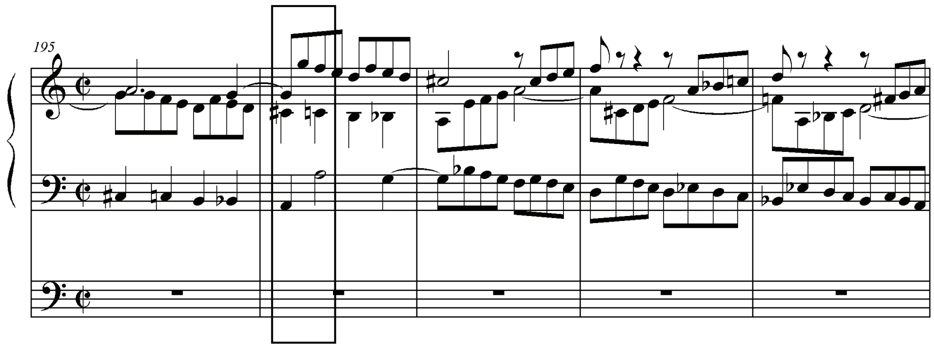


Figure 5.4. Dorian fugue, mm. 195-199. The boxed area corresponds to the octave skip in both hands.

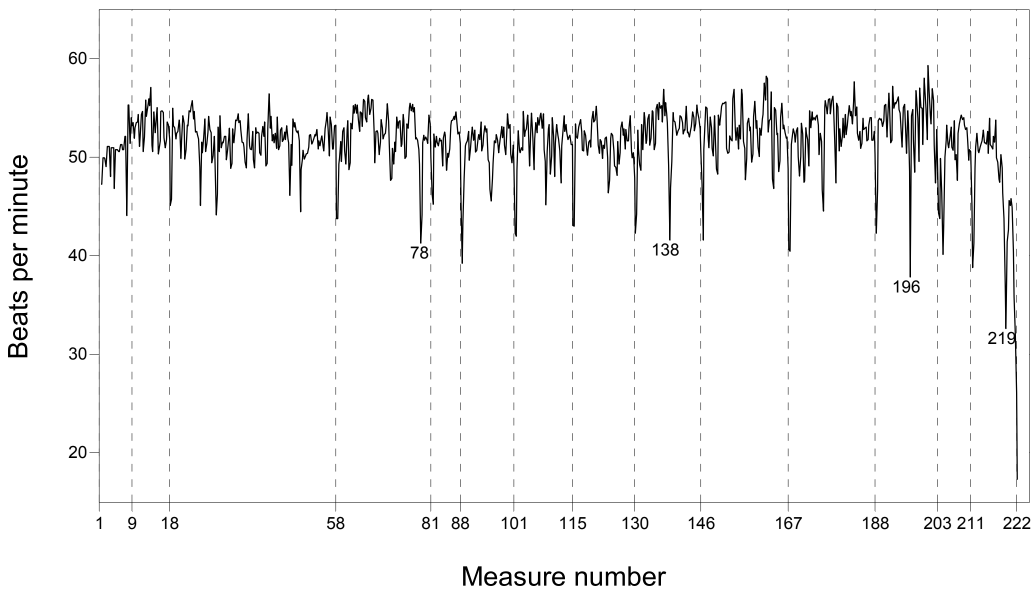


Figure 5.5. Average tempo profile for the performances of the Dorian fugue. Cadences are indicated by dotted lines (the cadence in m. 204 is not shown). Large temporal deviations that do not correspond to cadences are indicated by their measure number.

In order to compare the relative importance of the rallentandos across different locations in the piece, we evaluated the magnitude of each rallentando as the relative difference in tempo between the inflexion points in the tempo curve, that is, from the time the tempo began to slow down to where it begins to accelerate again. Thus, for each performance, rallentandos were identified by their beginning point and ending point at the quarter-note level. Since the beginning points and ending points of rallentando patterns did not necessarily coincide exactly for different performances, we chose to consider timing patterns at the level of the measure; this allowed for a more straightforward comparison between performances, while providing a one-to-one mapping with the measure numbers identified in the formal analyses. The largest rallentando for a given measure was defined as the rallentando with the largest tempo differential whose ending point was located within that measure.

Figure 5.6 represents the average size of the largest rallentando observed for each measure across all performances, expressed in percentage of the initial tempo (the tempo at the first inflexion point of the tempo curve). Again, we observe that the largest rallentandos coincided with structural points such as cadences, although mm. 78, 138, 196, and 219 were also characterized by important tempo variations as previously seen.

A direct comparison between the performers' analyses and their temporal profiles shows that most of the formal subdivisions identified by performers were associated with important tempo variations (Figure 5.7). In fact, 14 of the 20 largest tempo variations identified corresponded to formal subdivisions identified by the organists, and two other (m. 203 and m. 163) were one measure away from formal boundaries identified by performers. Most of the formal subdivisions that were not characterized by important rallentandos (m. 36, 43, 61, 64, 71, 108) were also not named by a large number of performers. Incidentally, we note that, except for m. 36, none of these subdivisions coincided with a cadence or with a statement of Episode A, while 17 of the 20 largest tempo variations corresponded either to cadences or to statements of Episode A. A significant correlation was found between the proportion of performers who agreed on a formal subdivision and the magnitude of the tempo variation associated with this formal subdivision, $r_s(19) = 0.43$, $p < .05$, indicating that the more agreed-upon subdivisions, which were presumably the most structurally important ones in the minds of the majority of performers, were characterized by larger tempo variations.

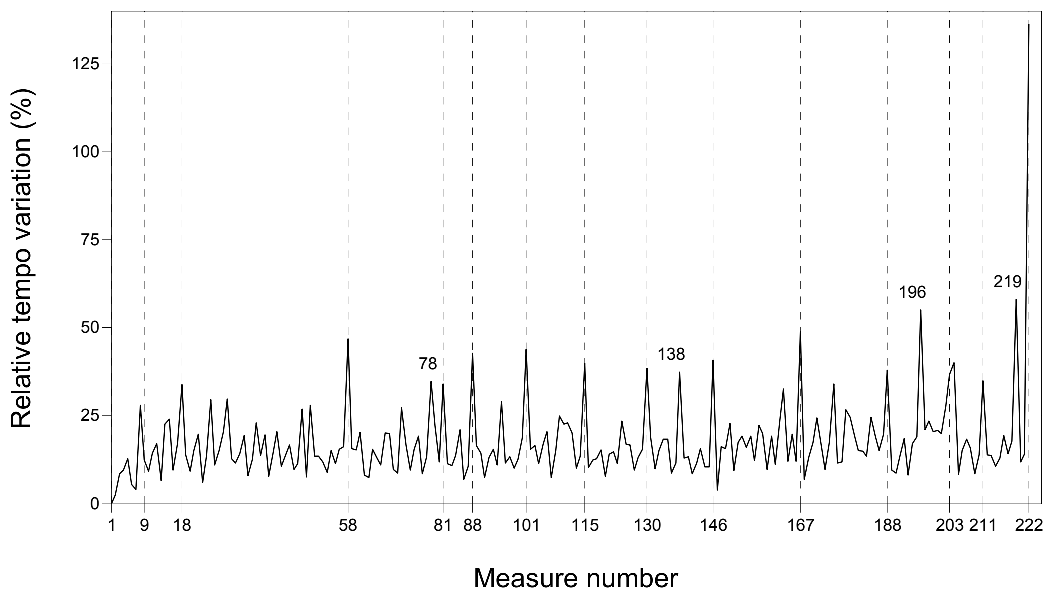


Figure 5.6. Average rallentando profile for the performances of the Dorian fugue. Cadences are indicated by dotted lines (the cadence in m. 204 is not shown). Large temporal deviations that do not correspond to cadences are indicated by their measure number.

However, it is worth noting that a few of the larger rallentandos were *not* associated with a formal subdivision identified by the performers. For instance, measure 18 corresponds to a subject entry in the tenor, which is preceded by a strong authentic cadence in D minor. Even though performers were clearly reluctant to identify this as a formal subdivision in their written analyses, since it is located halfway through the exposition and only 18 measures into the piece, they emphasized this subject entry by a relatively large rallentando. As mentioned above, the large ritardando observed at m. 196 may correspond to a technical difficulty related to parallel octave skips in both hands; nonetheless, this upward registral shift may also have structural implications, which implies that the sudden tempo change may be brought about both by motor considerations and by an expressive intent on the part of performers.

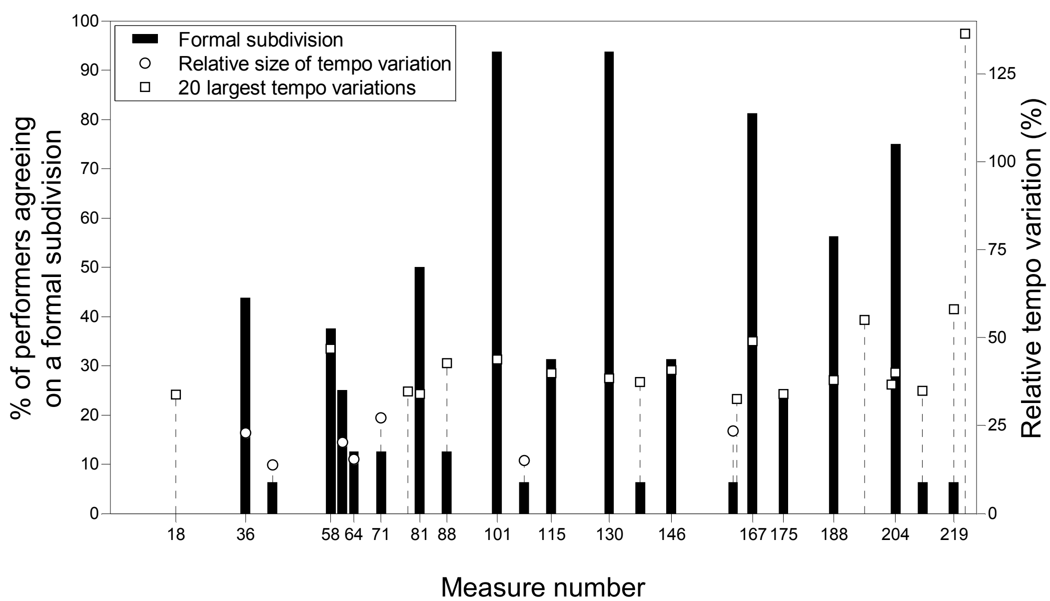


Figure 5.7. Comparison between the rallentando profiles and the formal subdivisions identified by performers. The relative size of the tempo variation associated with each formal subdivision is indicated by an open circle. The 20 largest tempo variations (including those which do not correspond to formal subdivisions) are indicated by open squares.

Analysis of individual performers' temporal profiles. In order to explore the temporal profiles of individual performers, a measure of similarity between the measure-by-measure rallentando patterns was obtained by computing correlations between all pairs of performances. A multidimensional scaling representation of the distance between the rallentando patterns of the performances was then conducted on the dissimilarity matrix obtained from the correlation coefficients. A two-dimensional solution (Figure 5.9) provided a reasonably good fit (stress-I = 0.23, $RSQ = 0.76$), as confirmed by a scree plot analysis (Figure 5.8). The dimensions were not significantly correlated with global tempo, $r(30) = -0.16$, $p = .37$, nor with the average magnitude of the tempo variation, $r(30) = 0.17$, $p = .36$, suggesting that disparities along these dimensions might be best explained by differences in local temporal patterns.

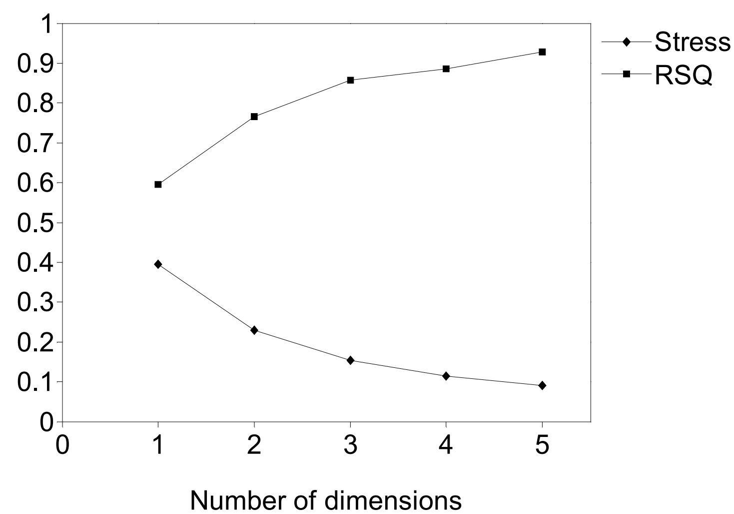


Figure 5.8. Fit-by-dimension plots for the multidimensional scaling representation of the rallentando profiles. *Stress*: Kruskal stress-I. *RSQ*: proportion of variance explained.

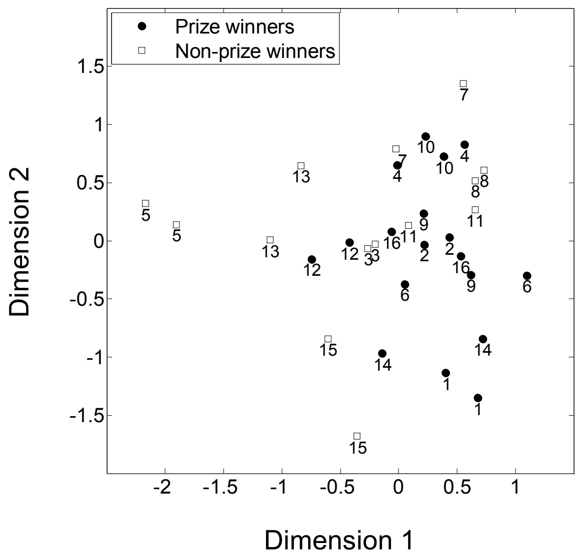


Figure 5.9. Multidimensional scaling of the distances between all performances, based on the correlations among rallentando profiles computed between all pairs of performances (monotonic regression; Kruskal stress-I = 0.23; RSQ = 0.76). Numbers identify individual organists. Each symbol with its accompanying number identifies a single performance.

A visual comparison of the rallentando profiles suggested that performances located on the left side of the graph did not exhibit a consistent association between large rallentandos and formal subdivisions, in contrast to performances found on the right side (Figure 5.10 contrasts the rallentando profiles of organists 5 and 8, both non-prize winners whose performances exhibited similar global tempi). To investigate this finding, the logarithm of the ratio of the average rallentandos for all measures identified as formal subdivisions by the performers to those of all measures which were not identified as such (excluding measures 1 and 222) was computed for each performance and regressed onto the first dimension, yielding a correlation of 0.68 ($df = 30$, $p <$

.001). This result indicates that the contrast between the magnitude of tempo variations associated with points identified as structurally important and those that were not increased with coordinates on the first dimension. In other words, the temporal profiles of performances with high coordinates on the first dimension (right side of Figure 5.9) reflected the formal subdivisions to a greater extent than those with lower coordinates (left side of Figure 5.9). Furthermore, a mixed-model repeated-measures ANOVA conducted on the rallentando profile of each performer, with level of accomplishment (prize-winners versus non-prize winners) as a between-subjects factor indicated a significant effect of the level of accomplishment on the coordinates on the first dimension, $F(1, 14) = 6.11$, $p < .05$. This corresponds to a tendency for performances of prize-winning organists to be located on the right side of Figure 5.9.

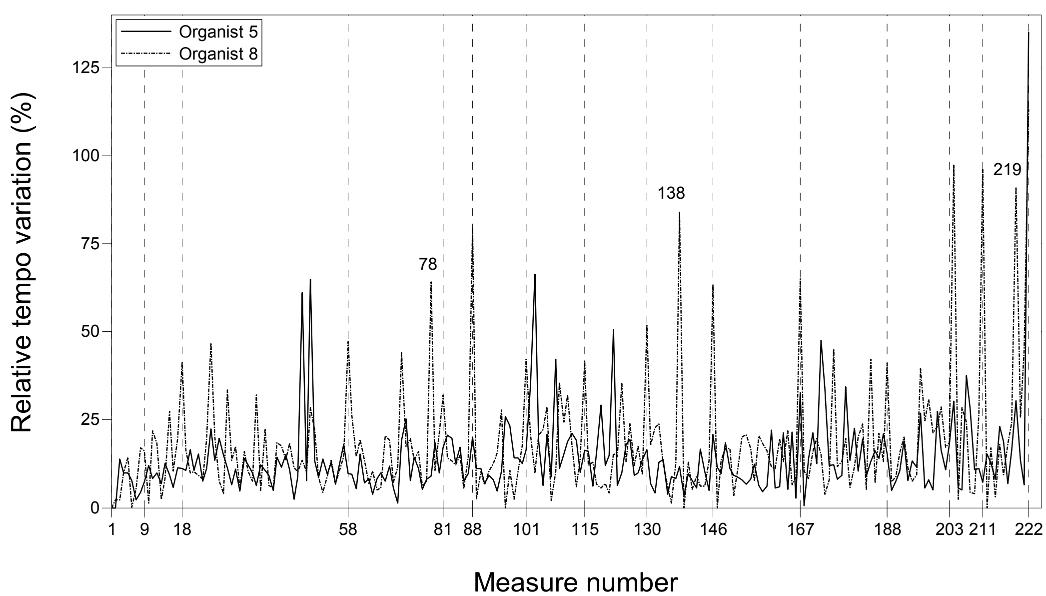


Figure 5.10. Comparison of the rallentando profiles for the performances of Organists 5 and 8. Profiles were averaged over two performances. The mean tempo was 49 BPM for Organist 5 and 45 BPM for Organist 8.

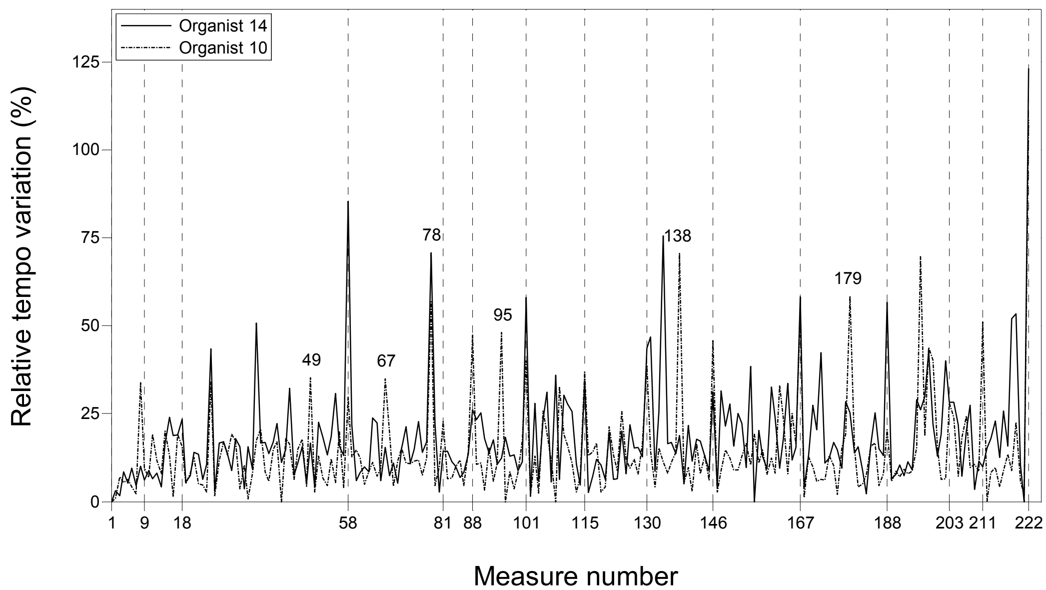


Figure 5.11. Comparison of the rallentando profiles for the performances of Organists 10 and 14. Profiles were averaged over two performances. The mean tempo was 51 BPM for both organists. Peaks corresponding to recurrences of Episode A are identified by their measure numbers (the peak at m. 179 was assumed to correspond to the beginning of Episode A in m. 178).

An examination of the individual rallentando profiles revealed that performances in the upper portion of the multidimensional scaling graph (Figure 5.9) exhibited more pronounced rallentandos associated with the recurrences of Episode A (Figure 5.11 contrasts the rallentando profiles of organists 10 and 14, both prize-winners whose performances exhibited similar global tempi). In order to quantify this observation, the logarithm of the ratio of the magnitude of the rallentandos for all measures corresponding to a recurrence of Episode A or to the codettas in the exposition (see Table 5.1) to all other measures (excluding measures 1 and 222) was computed for each performance. A correlation of 0.78 (*df* = 30, $p < .001$) was found between the logarithm of this ratio and the coordinates on the second dimension, indicating that the ratio was larger for performances found in the upper half of Figure 5.9. In other words, the rallentando profiles of performances in the upper graph were characterized by larger rallentandos associated with the recurrences of Episode A than those found in the lower portion of the graph.

Comparing individual analyses with tempo profiles. A further question that we sought to address in this study was the extent to which analytical readings of the piece were related to the temporal profiles for individual performers. Given that performers were free to interpret or analyze the piece as they wished, it was difficult to assess directly whether a performer who identified a structural boundary emphasized it to a greater extent in his or her performances than a performer who did not. Nevertheless, this relationship could be examined indirectly by comparing the temporal deviations of performers who labeled a specific measure as a formal subdivision to those of performers who did not. In order to conduct meaningful comparisons, these analyses were conducted only on formal subdivisions for which there was a substantial degree of disagreement (i.e., between 20% and 80% of performers indicated a subdivision), so that a minimum of four performers either did or did not identify a given measure as a formal subdivision. These subdivisions corresponded to mm. 36, 58, 61, 81, 115, 146, 175, 188, and 204 (see Figure 5.3). Separate t-tests were conducted for each of the subdivisions listed above; uncorrected p values were superior to .40 for all subdivisions, indicating that no significant difference was found in the average size of the rallentandos between the performers who analyzed a section as a boundary and those who did not.

DISCUSSION

The results presented here illustrate that there was a good agreement between the formal subdivisions indicated by organists in their written analyses and the temporal profiles observed in their performances. Cadences and recurrences of Episode A were highlighted by large variations in tempo, whereas other formal elements identified by performers, mostly those that did not correspond to cadences or to statements of Episode A, were not emphasized by means of temporal variations.

The application of multidimensional scaling analysis techniques revealed that, although the temporal profiles of different performers were fairly similar, individual interpretations of the piece could be contrasted on the basis of their rallentando profiles. Two main dimensions emerged, one relating to the relative salience of tempo variations associated with formal subdivisions (when contrasted with tempo variations not associated with formal subdivisions) and one relating to the magnitude of the rallentandos corresponding to the recurrences of Episode A. Assuming that the role of local tempo variations is, at least in part, to communicate a specific structural reading of the piece, we may say that the first dimension identified here corresponds to a signal-to-noise ratio in the communication of structure through timing variations, the "signal" being the temporal variations corresponding to structural events and the "noise" the fluctuations that are not associated with formal subdivisions. On the other hand, the second dimension corresponds more specifically to an interpretive choice on the part of performers, with some organists choosing to emphasize the statements of Episode A through the use of rallentandos to a larger extent than others.

The present study did not establish an unequivocal correlation between individual organists' written analyses and the temporal profiles of their performances, even though a significant correlation was found between the level of agreement on a formal subdivision and the local tempo variations associated with this subdivision averaged across all performances. This may be because performers viewed the written analysis as a separate task from the performance. Indeed, although we have shown that the temporal profiles were clearly informed by the structure of the piece, it does not necessary follow that each performer's written analysis of the piece corresponds to his or her performance. It is likely that most performers felt compelled to indicate formal subdivisions that corresponded to what they were taught in music analysis courses, rather than what they felt was specific to the Dorian fugue. A case in point is the contrast between the importance given to measure 36, which corresponds to the end of the exposition (traditionally seen as an important formal subdivision in fugal forms), in the written assessments, and the absence of an important tempo variation associated with this measure in most performers' temporal profiles. Conversely, most performers refrained from labeling recurrences of episodes as important formal subdivisions, presumably because episodes are generally not considered to be structural boundaries in traditional fugal analysis; yet, several performers clearly emphasized the return of Episode A through important tempo variations in their performances. Indeed, music-theoretical analysis is often seen as a rigorous and prescriptive exercise, where there is little margin for individuality, and performers may have felt compelled to produce an analysis that conformed to academic standards. On the other hand, although performance may well be regulated by expectations and norms, it represents a more convenient vehicle for the expression of individual interpretations. To simplify, we may say that whereas performers sought to analyze a particular piece, in this case the Dorian fugue, in conformity to a "formal archetype" of the fugue in their written analyses, they strove to highlight the unique and striking features of this piece in their performances.

Although one goal of the present study was to gain insight into the performers' individual interpretations of the formal structure of the piece, it appears that the methodology used here encouraged conformity to an academic model of analysis. The relationship between analysis and performance should perhaps be investigated by means of a different strategy: for instance, by asking performers to indicate formal subdivisions while listening to a recording of the piece, unwanted associations with written analysis, and its concomitant norms and expectations, could be avoided.[3] Indeed, an in-depth investigation of the relationship between analysis and performance should aim to obtain a performer's representation of a piece's structural hierarchy, which is unmediated by verbal processes, with the intent of comparing this representation to its actual musical realization.

While methodological improvements may be required, we believe that the experimental procedure outlined in this article represents a fruitful paradigm for the investigation of the relationships between analysis and performance, which

[3] See Cook (1999) for a discussion of the role of the verbal and written tradition in the relationship between music analysis and performance.

could potentially be applied to the study of other expressive parameters, such as articulation and dynamics, as well as other levels of musical structure, for instance phrases, themes, or motives, and finally to other musical genres.

ACKNOWLEDGMENTS

This research was supported by fellowships from the Social Sciences and Humanities Research Council of Canada and from the Centre for Interdisciplinary Research in Music Media and Technology (CIRMMT) to Bruno Gingras, as well as a grant from the Natural Sciences and Engineering Research Council and a Canada Research Chair awarded to Stephen McAdams. We thank Bennett Smith for his technical assistance, Peter Holmes for permission to use McGill University's sound recording equipment, Nils Peters for his 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.

REFERENCES

Berry, W. (1989). *Musical structure and performance*. New Haven: Yale University Press.

Bullivant, R. (1971). *Fugue*. London: Hutchinson.

Caldwell, J. (2007). Keyboard music to c1750, I.4. The period of J.S. Bach. *Grove music online*, ed. L. Macy. Retrieved April 6, 2008, from <http://www.grovemusic.com>

Clarke, E. F. (1985). Structure and expression in rhythmic performance. In P. Howell, I. Cross & R. West (Eds.), *Musical structure and cognition* (pp. 209-236). London: Academic Press.

Clarke, E. F. (1988). Generative principles in music performance. In J. Sloboda (Ed.), *Generative processes in music: The psychology of performance, improvisation and composition* (pp. 1-26). Oxford: Clarendon Press.

Clarke, E. F. (1989). The perception of expressive timing in music. *Psychological Research-Psychologische Forschung, 51*(1), 2-9.

Cone, E. T. (1968). *Musical form and musical performance* (1st ed.). New York: W. W. Norton.

Cook, N. (1999). Analysing performance and performing analysis. In N. Cook & M. Everist (Eds.), *Rethinking music* (pp. 239-261). United Kingdom: Oxford University Press.

Gabrielsson, A. (1987). Once again: The theme from Mozart's piano sonata in A major (K. 331). In A. Gabrielsson (Ed.), *Action and perception in rhythm and music* (pp. 81-104). Stockholm: Royal Swedish Academy of Music.

Jerkert, J. (2004, June). *Musical articulation in the organ.* Paper presented at the Joint Baltic-Nordic Acoustics Meeting, Mariehamn, Finland.

Lester, J. (1995). Performance and analysis: Interaction and interpretation. In J. Rink (Ed.), *The practice of performance: Studies in musical interpretation* (pp. 197-216). United Kingdom: Cambridge University.

Lester, J. (2001). Heightening levels of activity and J.S. Bach's parallel-section constructions. *Journal of the American Musicological Society, 54*(1), 49-96.

Meyer, L. B. (1973). *Explaining music; essays and explorations*. Berkeley: University of California Press.

Narmour, E. (1988). On the relationship of analytical theory to performance and interpretations. In E. Narmour & R. A. Solie (Eds.), *Explorations in music, the arts, and ideas* (pp. 317-340). Stuyvesant, NY: Pendragon.

Palmer, C. (1989). Mapping musical thought to musical performance. *Journal of Experimental Psychology-Human Perception and Performance, 15*(12), 331-346.

Palmer, C. (1996). On the assignment of structure in music performance. *Music Perception, 14*(1), 23-56.

Penel, A., & Drake, C. (1998). Sources of timing variations in music performance: A psychological segmentation model. *Psychological Research-Psychologische Forschung, 61*(1), 12-32.

Penel, A., & Drake, C. (2004). Timing variations in music performance: Musical communication, perceptual compensation, and/or motor control? *Perception & Psychophysics, 66*(4), 545-562.

Prout, E. (1891). Fugal structure. *Proceedings of the Musical Association, 18*(1), 135-159.

Repp, B. H. (1990). Patterns of expressive timing in performances of a Beethoven minuet by 19 famous pianists. *Journal of the Acoustical Society of America, 88*(2), 622-641.

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. (1995). Expressive timing in Schumann's "Träumerei" - an analysis of performances by graduate student pianists. *Journal of the Acoustical Society of America, 98*(5), 2413-2427.

Rink, J. (1995a). Playing in time: Rhythm, metre and tempo in Brahms's Fantasien op. 116. In J. Rink (Ed.), *The practice of performance: Studies in musical interpretation* (pp. 254-282). United Kingdom: Cambridge University.

Rink, J. (1995b). *The practice of performance: Studies in musical interpretation*. Cambridge; New York: Cambridge University Press.

Rink, J. (2002). *Musical performance: A guide to understanding*. Cambridge, U.K.; New York: Cambridge University Press.

Rothstein, W. (1995). Analysis and the act of performance. In J. Rink (Ed.), *The practice of performance: Studies in musical interpretation* (pp. 217-240). United Kingdom: Cambridge University.

Schmalfeldt, J. (1985). On the relation of analysis to performance: Beethoven's bagatelles op. 126, nos. 2 and 5. *Journal of Music Theory, 29*(1), 1-31.

Shaffer, L. H. (1981). Performances of Chopin, Bach, and Bartok: Studies in motor programming. *Cognitive Psychology, 13*(3), 326-376.

Shaffer, L. H., & Todd, N. P. M. (1987). The interpretive component in musical performance. In A. Gabrielsson (Ed.), *Action and perception in rhythm and music* (pp. 139-152). Stockholm: Royal Swedish Academy of Music.

Todd, N. (1985). A model of expressive timing in tonal music. *Music Perception, 3*(1), 33-58.

Walker, P. (2008). Counter-exposition. *Grove music online*, ed. L. Macy. Retrieved April 6, 2008, from <http://www.grovemusic.com>

Williams, P. F. (2003). *The organ music of J. S. Bach*. Cambridge: Cambridge University Press.

Chapter 6. Improved score-performance matching using both structural and temporal information from MIDI recordings

Although score-performance matching can be done reliably by hand, such a procedure becomes unwieldy for analyzing large databases of performances or performances of longer pieces. In fact, the amount of work involved in the completion of the hand matches of the performances of Grigny's *Premier Agnus* and of Scheidt's *Wachet auf* recorded in the context of this research project was a primary motivation in the design of the score-performance matching algorithm which is introduced in Chapter 6. This matcher relies on both structural and temporal information, allowing it to generate an accurate match even for heavily ornamented performances. A detailed description of the matching procedure is given, as well as a quantitative assessment of the accuracy of the algorithm. This chapter also introduces a heuristic for the identification of ornaments and errors that is based on perceptual principles, and which could theoretically be amenable to empirical study.

This chapter is based on the following research article:

Gingras, B., & McAdams, S. Improved score-performance matching using both structural and temporal information from MIDI recordings. Manuscript prepared for submission to *Computer Music Journal*.

ABSTRACT

Automated score-performance matching is a complex problem due to the use of expressive timing by performers and the presence of notes that are unspecified in the score, such as performance errors and ornaments. Automated matchers typically use performance data extracted from MIDI recordings. For the most part, these algorithms use structural information, such as pitch and chronological succession, but do not use timing information. As a result, most matchers cannot deal satisfactorily with ornamented performances or performances that exhibit extreme variations in tempo. The matcher presented here relies both on structural and temporal information, allowing it to generate an accurate match even for heavily ornamented performances. A comparison with hand-made score-performance matches on a corpus of 80 MIDI recordings of organ performances of two pieces, which were used as ground truth data for this purpose, shows that the matcher achieved an accuracy rate of 99.98%. This constitutes a significant improvement over matchers previously described in the literature. We also propose a heuristic for the identification of ornaments and errors that is based on perceptual principles, and which could theoretically be amenable to empirical study. Finally, this matcher is designed to accommodate multi-channel MIDI recordings of performances from keyboard instruments with multiple manuals, such as organ or harpsichord. This feature makes it a potentially valuable tool for the investigation of ensemble performances of MIDI instruments.

INTRODUCTION

Music performance has been characterized as a component of a communication system in which composers code musical ideas in notation, performers transduce this notation into an acoustical signal, and listeners recode the acoustical signal into musical ideas (Kendall & Carterette, 1990). This model applies particularly to score-based music performance, which characterizes a significant proportion of classical Western musical practice. The score, written by a composer, generally specifies the pitches and durations of the notes to be played by the performer in an unambiguous manner, while conveying less specific information about articulation, dynamics and ornamentation (Large, 1993; Palmer, 1997). Depending on the repertoire, the performer has more or less freedom in deciding how to interpret the score, but pitches and nominal note durations are generally less subject to variation than other musical parameters, given that they can be categorically defined. Since the score provides an explicit benchmark with which the performance can be compared, score-based music performance has constituted the focus of research in music performance (Palmer, 1997).

In order to study score-based music performance quantitatively on a note-by-note basis, the researcher needs to determine the corresponding score note for every performance note, a process called *score-performance matching*. Although score-performance matching can be done reliably by hand (Repp, 1996a), such a procedure becomes unwieldy for analyzing large databases of performances or performances of longer pieces. Fortunately, algorithms that automate this procedure have been developed. Such algorithms are called *matchers*. Automated matchers typically compare a representation of the performance (either audio or MIDI recording) to a symbolic representation of the score and try to seek the best match between both. In the last two decades, several such matchers have been developed (Heijink, Windsor, & Desain, 2000b; Large, 1993; Puckette & Lippe, 1992). An important distinction should be made between matching algorithms whose main purpose is that of real-time accompaniment, often called *score following* (Dannenberg, 1984; Puckette & Lippe, 1992), and algorithms that are designed to find the best possible match for a performance, which we will call *offline matchers* (Heijink et al., 2000b; Large, 1993; Raphael, 2006). While the former are mostly concerned with efficiency and real-time responsiveness and are used in performance settings, the latter seek accuracy and are mainly used for research purposes (Heijink, Desain, Honing, & Windsor, 2000a).

The MIDI protocol does not provide an exact representation of the performance; MIDI records quantifiable data such as note onsets, note offsets, pitch, and velocity, but ignores other aspects such as timbre and spectral content. On the other hand, extracting performance information directly from the audio recording is a method that retains all sonic aspects of the performance and which can be used with non-MIDI instruments. However, until recently, direct matching of an audio recording of a performance to a score of a polyphonic piece has proven a challenging task, although researchers have addressed this problem (Dixon, 2005; Raphael, 2006). Altogether, for performance research focusing on timing, tempo, and articulation, MIDI does convey most, if not all, of the relevant information, and remains far easier to process than audio recordings, especially for polyphonic music and long performances. The present article will concern itself solely with MIDI recordings of keyboard performances.

Some authors have treated the problem of matching a performance to a score as a typical sequence-alignment problem (Large, 1993) and have sought to adapt solutions from other disciplines, such as nucleic acids or amino acid sequencing in molecular biology (Gotoh, 1982; Needleman & Wunsch, 1970). Thus, a number of matching algorithms define the best alignment between two sequences *A* and *B* as the one for which the editing distance (usually defined as the number of changes such as deletions, additions, or substitutions) between *A* and *B* is the shortest (Mongeau & Sankoff, 1990). In cases where the performance closely matches the score, this model is generally adequate. However, even for expert performances, there is rarely a perfect one-to-one match between score and performance (Repp, 1996a). Discrepancies between score and performance can be attributed to three main factors: 1) performance errors, 2) temporal deviations brought about by expressive timing in performance, and 3) underspecification of scores (Heijink et al., 2000a).

A performance error can be defined in a very general way as an unintended deviation from the written score that occurs in performance (Palmer & Van de Sande, 1993). Most researchers have only considered errors that correspond to deletions (failure to play notes indicated in the score), additions (insertion of extraneous notes not indicated in the score), or substitutions (pitch errors or "wrong notes") (Repp, 1996a). Some researchers also take into account other error types which may be defined as "timing errors", or, to be more precise, chronological shifts between the succession of notes indicated in the score and that which was performed (Palmer & Van de Sande, 1993, 1995). This type of error should not be confused with temporal shifts caused by expressive timing (see below), although the boundary between them is necessarily subjective.

Since most matchers rely solely on a comparison between the chronological succession of notes and chords in the score and in the performance (Heijink et al., 2000b; Large, 1993), expressive timing in performance may affect the matching process by disrupting the order of the notes. For instance, a situation in which notes that should be played synchronously according to the score (for instance, notes belonging to the same chord) are played asynchronously in performance can lead to wrong note assignments in the score-to-performance matching process. Such asynchronies are common occurrences in piano performance (Goebl, 2001; Palmer, 1989, 1996; Repp, 1996b).

Finally, scores generally indicate ornaments by means of symbols, which do not specify the exact timing of the ornaments, nor the number of notes that comprise them in the case of complex ornaments such as trills (Dannenberg & Mukaino, 1988). In addition, in certain musical genres, such as the Baroque repertoire, performers routinely add ornaments that are not specified in the score. This *underspecification* of the musical scores represents another obstacle for matchers in ornamented pieces, because editing distance models assume an exact one-to-one mapping at the level of individual notes between score and performance (Pardo & Birmingham, 2001).

Indeed, in the case of performances that exhibit extreme expressive timing or heavy ornamentation, the analogy between score-performance matching and typical sequence-alignment problems does not apply: a performance may contain several additional notes not indicated in the score, and the order in which the notes are played in the performance may differ from the order in which they are notated. In this case, the score should be treated as a template which provides a more or less specific framework and indicates the key structural points, leaving several aspects of the performance, such as ornamentation and expressive timing, to be freely determined by the performer (Pardo & Birmingham, 2001).

Several authors have proposed using timing information to increase the accuracy of the score-performance matching process (Desain & Honing, 1992; Puckette & Lippe, 1992; Raphael, 2006). Hoshishiba and colleagues presented a matcher that uses temporal information (Hoshishiba, Horiguchi, & Fujinaga, 1996); however, the detailed implementation of this matcher was not described. Vantomme (1995) developed a score follower that gives precedence to temporal information over pitch information, unlike most algorithms described in the literature.

Conversely, very few researchers have tackled issues related to the identification of ornaments. Dannenberg & Mukaino (1988) proposed an algorithm which can cope with specific ornaments, such as trills and glissandi, by relying on the fact that notes composing these ornaments usually have a much shorter duration than score notes, as long as these ornaments are indicated in the score. However, an algorithm which could handle all types of ornaments, regardless of whether they are specified in the score or not, would have a wider applicability to all kinds of musical situations.

Among the best-known offline matchers are those developed by Honing (1990), Large (1993) and Heijink and colleagues (Desain, Honing, & Heijink, 1997; Heijink, 1996; Heijink et al., 2000b). The *strict matcher* (Honing, 1990) takes the notated order of the notes in the score as a strict temporal constraint on the performance; the performance is processed note-by-note, and only one possible interpretation is considered at any point in time, which results in a high sensitivity to performance errors. In contrast, the matcher developed by Large (1993), which will be henceforth referred to as the *Large matcher*, is somewhat more robust since it divides the performance into clusters (notes played together) before trying to match it to the score and uses complete knowledge of the performance and of the score to find the globally optimal match. Furthermore, this matcher considers many possible alternative solutions at any point in time, and can analyze some performance errors, such as insertions, deletions, and substitutions. Indeed, it was used in the context of research on errors in piano performance (Palmer & Van de Sande, 1993).

In spite of their usefulness, these matchers present several limitations. The most important one is that they use only pitch and note order to find the optimal score-performance match, not taking into account voice structure or timing information. As a result, these algorithms cannot deal satisfactorily with ornamented performances or performances that exhibit extreme expressive timing such that the chronological succession of notes does not correspond to that indicated in the score. In an attempt to solve some of these problems, Heijink and colleagues (Desain et al., 1997; Heijink, 1996) proposed a structure matcher, which takes into account the voice information present in the score by assigning each score note to a voice. This matcher is able to cope with extreme expressive timing resulting in deviation in the chronological succession of notes. However, the solution adopted by these authors is somewhat extreme in that parallel events in different voices are considered to be temporally independent, a model which does not seem to accurately represent common musical practice.

Other problems encountered with the offline matchers discussed here involve a sensitivity to errors, and particularly errors involving repeated notes (Heijink et al., 2000b, p. 549). In addition, all MIDI-based offline matchers described in the literature were designed for the analysis of piano performance, and cannot handle MIDI recordings of instruments with multiple manuals, such as the organ or harpsichord. Finally, most existing algorithms are designed to find a solution that maximizes the number of matched performance notes, regardless of the perceptual relevance of such an approach. However, a definition of the best match which is based solely on the number of matched notes is problematic, as it may ignore relevant structural and temporal information (Heijink et al., 2000b, p. 552).

In an attempt to solve these issues, we developed a matcher that relies both on structural information and on a temporal representation of the performance, which is obtained by sequentially tracking local tempo changes on a note-by-note basis and mapping performance events to the corresponding score events. This allows the matcher to generate an accurate match even for heavily ornamented performances. The best match is defined as the one that maximizes the number of matched performance notes, while minimizing the structural and temporal inconsistencies in the individual voices. Furthermore, this matcher is designed to accommodate multi-channel MIDI recordings. Finally, we propose a very general approach to the identification of ornaments. The first section of this article describes the algorithm used by the matcher, whereas the second section reports on the efficiency of this implementation. A final section discusses current limitations and possible improvements.

DESCRIPTION OF THE MATCHER

The matcher described here follows a three-step process; we will thus refer to it as the "three-step matcher". Before discussing each step in detail, we will outline an overview of this process. The first step, which corresponds to a structural matching algorithm, is similar to the algorithm described by Large (1993) in that it decomposes the performance into note clusters and establishes a preliminary match by relying solely on structural information such as pitch and note onset. The second step uses results from the first step, as well as temporal information, to construct a "temporal match" in which the onsets of score events are matched to corresponding performance clusters. Finally, the third step combines information from the first two steps to find the best note-by-note correspondence between score and performance. Unmatched performance notes are identified as ornaments or errors at this stage. At each step, several possible alternatives are considered.

Symbolic representation of the score

As described by Schwarz, Orio, and Schnell (2004), the score is parsed into a time-ordered sequence of *score events*, where each score event corresponds to a change in the polyphonic texture (one or more note onsets or offsets). Each score note is thus bound in time by its *onset event* and its *offset event*. Score notes are also defined by their pitch, voice, and MIDI channel. In addition, the matcher keeps track of embellishment markings in the score; this information is used for the identification of ornaments.

The use of voice information improves the quality of the match for polyphonic scores containing more than one voice, as it allows for a more refined representation of the musical structure of the score (Desain et al., 1997); likewise, notes that were played on different manuals on a MIDI-controlled organ, for instance, can be differentiated by taking into account the MIDI channel information. In contrast to the structure matcher (Desain et al., 1997), the temporal sequence of score events supersedes the voice information associated with each note; thus, the different voices are conceived as temporally related, so that notes in different voices that share the same onset event are expected to have quasi-synchronous onsets, as is normally the case with common-practice music performance.

First step: structural matching

In the first step, performance notes are initially grouped into clusters according to the proximity of their onsets in time. Notes that are played quasi-synchronously are assumed to belong to the same event (Schwarz et al., 2004). The three-step matcher initially groups together notes whose onsets can be found within a span of 40 milliseconds (this *maximum inter-onset interval* corresponds approximately to the maximal onset asynchronies observed in professional music performance; see Rasch, 1979), and whose onset times are closer to each other than to those of any other notes. This initial parsing is used to estimate the average onset time distance between adjacent clusters. This value is then used to generate a more refined parsing which adjusts the size of the maximum inter-onset interval according to the average onset time distance. One advantage of this two-step parsing is that it is more flexible than the procedure used by matchers that use a fixed maximum inter-onset interval for the parsing of performance notes into clusters (Honing, 1990; Large, 1993). Moreover, while the parsing of the performance notes into clusters is a critical step in the strict matcher and the Large matcher, it does not determine the final results for the three-step matcher, since an erroneous parsing can be corrected in subsequent steps.

Once the second parsing is completed, structural comparisons between the content of each performance cluster and each score event are conducted on the basis of three criteria: pitch similarity, number of onsets, and MIDI channel congruence (that is, whether corresponding notes were played in corresponding MIDI channels for multi-channel MIDI recordings). Structural ratings are then computed for each performance cluster/score cluster combination, and a table containing these ratings is built (Table 6.1). It is normally unnecessary to compute values for the entire table, because it is unlikely that actual score event/performance cluster pairings will be located far from the main diagonal going from the top left to the bottom right part of the table. Such calculations are computationally expensive and time-consuming, especially for performances containing hundreds or thousands of events. On the other hand, if the matcher does not consider all possible solutions, there is a risk that the optimal solution will be missed. Therefore, there must be a trade-off between computational efficiency and finding the best solution. The three-step matcher uses a measure of structural discrepancy to evaluate how many score event/performance cluster pairings should be computed. This *discrepancy index* is based on the ratio of the number of performance clusters to the number of score clusters, and of the number of performance onsets to the number of score onsets. When these ratios deviate significantly from a value of one, it suggests that the performance is heavily ornamented and/or that it contains several errors.

Score events

Performance clusters	1	2	3	4	5	6	7	8
1	100	0	0	25	0	0	0	15.625
2	0	100	0	0	0	81.25	56.25	25
3	0	0	100	0	37.5	0	0	25
4	25	0	0	100	0	25	0	50
5	0	0	100	0	37.5	0	0	25
6	0	0	62.5	25	50	0	0	0
7	0	0	37.5	0	100	0	0	0
8	0	81.25	0	25	0	100	50	25
9	0	56.25	0	0	0	50	100	0
10	15.625	25	25	50	0	25	0	100

Table 6.1. Structural ratings for performance clusters / score events pairings. Highlighted cells correspond to perfectly matched pairings. Note that more than one performance cluster may be perfectly matched to the same score event.

The structural ratings obtained at this stage are then used to generate a *structural pre-match*, which takes into account the chronological succession of events (but not the timing information). This structural pre-match includes only unique events (defined as events that are found only once in a span corresponding to approximately twenty events) that are perfectly matched. The purpose of the structural pre-match is not to create a complete mapping of the performance, but rather to establish a set of *landmark events* that will be used in the following steps (see McAdams, Vines, Vieillard, Smith, & Reynolds, 2004, for a discussion of landmark registration techniques). This step may prove to be crucial in instances where substantial sections of the score were omitted in performance (such as when several chords or even entire measures were skipped in performance), or when a performance is heavily ornamented.

Scores that comprise a greater number of unique events will be conducive to good structural matches, whereas pieces that have a small number of recurrent events, or that contain many similar events, tend to generate poor matches, regardless of the discrepancy index value between performance and score. More generally, we may say that a score that contains several identical events will cause more difficulties for the matching algorithm than a score with a large diversity of events, where almost each event is unique in the whole piece. This, of course, becomes increasingly relevant when the identical events are proximal in the score. The problem of repeated notes, as well as the larger issue of event similarity was mentioned by both Heijink et al. (2000b) and Large (1993), but they did not propose a coherent approach to this problem. The three-step matcher tackles this issue by computing an *event diversity index*, based on Shannon's diversity index (1948), and uses this information to estimate the number of solutions that should be considered in the following steps (temporal matching and note-by-note matching), so that a greater number of solutions are computed for scores that contain many similar or identical events.

Finally, the quality of the fit observed between the performance clusters and the score events in the structural pre-match is also used by the matcher to estimate the number of solutions that should be computed in subsequent steps. A performance with no errors or ornaments and a moderate amount of expressive timing will give a better *structural fit* than one that is either error-filled or that uses expressive timing deviations which creates asynchronies between hands, such that the note order in performance differs from that indicated in the score. Although very crude, this measure of fit provides a good assessment of the difficulty involved in matching a specific performance to a given score. Thus, the matcher takes into account the discrepancy between the number of performance clusters and score events, the structural fit between score and performance, as well as the event diversity index to determine the number of solutions to be computed. This approach has the advantage of tailoring the computational needs to the difficulty of the matching task.

Second step: temporal matching

The temporal matching is probably the feature that most significantly differentiates the three-step matcher from the majority of offline matchers described in the literature, and it proves to be crucial in determining the quality of the final match. During this step, the matcher initially uses information from the structural pre-match computed in the first step to predict the onset time for each score event, using onset times of landmark events as a starting point, and proceeding in a sequential way (that is, one score event at a time). The probable onset time of each event is estimated using a local tempo model which attributes a greater weight to events closely following or preceding the current event than to events which are more distant in time (Vantomme, 1995).

A delicate issue associated with temporal matching is determining the size of the temporal window for which performance-cluster candidates corresponding to a given score event should be considered. Temporal deviations in performance may be due to motor noise (Desain & Honing, 1993) or abrupt changes in tempo such as *ritardandos* or *accelerandos*; however, it may also be that a score event was omitted in performance. An erroneous interpretation in such situations may lead the temporal matcher completely astray and negatively affect the quality of the match. Vantomme (1995) used a "window of belief" to estimate the maximum tolerance in onset time deviation, resorting to pitch information only when the deviation for an expected event was greater than this tolerance threshold. Conversely, the three-step matcher evaluates the *event rating* of performance-cluster candidates both as a function of their structural rating obtained in the structural matching step and of a temporal rating which is based on the distance between the predicted onset time and the mean onset time of the notes belonging to the performance cluster. The relative weight ascribed to the structural rating depends on the general structural fit between score and performance, so that the temporal component becomes primordial in the case of poorly matched performances.

Moreover, the temporal matcher follows an iterative process, optimizing the quality of the match over several cycles: at each step, several solutions are considered, and only the ones with the highest ratings are selected. This step-by-step procedure increases the robustness of the matching process by making it less susceptible to errors brought about by local temporal deviations or score/performance mismatches. During the initial cycles, onset times of score events are predicted for both forward (proceeding from the first score event to the last) and backward (proceeding from the last score event to the first) passes. Solutions are obtained by pairing the forward and backward matches that show the highest agreement between onset times and retaining only the onset times which are common to both matches. The resulting match is then passed on to the next cycle, and onset times are computed for both forward and backward passes using information from the previous cycle until a stable solution is reached. Then, a new series of cycles is conducted, taking the match with the highest global event rating as the basis for the following cycle until a stable solution is reached (no distinction is made between backward and forward passes at this stage).

Third step: note-by-note matching

The third step consists of a specific note-by-note matching that uses information from the two previous steps and takes into account both voice and MIDI channel assignment for each note. As its name implies, the main difference between this note-by-note matching step and the previous steps is that performance notes are considered individually instead of being grouped into clusters. It is during this final step that errors and ornaments are identified.

During this step, a temporal fit between individual notes and score events is first estimated by computing *onset difference ratings* as a function of the time difference between the onsets of performance notes and the predicted onsets of score events obtained from the temporal matching step. All performance notes whose onsets occur within 250 ms of a predicted event onset are considered as possible candidates for a match; in addition, a minimum of three score events are considered for any given performance note, regardless of the onset time difference.

The note-by-note matcher then proceeds to match performance notes to score notes in a sequential way, from the first event of the piece to the last. As with the temporal matcher, several solutions are considered at each stage. For each score event, a *match rating* is computed between every score note *s* belonging to this event and each candidate performance note *p*. This match rating is based on the onset difference rating and a *pitch-distance rating*, calculated from the pitch interval (in semitones) between *s* and *p*. The note-by-note matcher preserves the order of the notes in a given voice: thus, to be considered as a potential match for a score note in voice *v*, the onset of *p* must occur later than the onset of the last matched note in *v*. This order constraint is based on the observation that notes belonging to a melodic line are not likely to be played in a different order from that indicated in the score (Desain et al., 1997). Moreover, only performance notes which are played in the appropriate MIDI channel may be considered as candidates; for instance, a note played on the pedal on a MIDI organ cannot be considered as a potential match for a score note meant to be played on the manuals, even if it matches the pitch of that note.

In most cases, the matching process is unambiguous: only one performance note *p* fits all the requirements in terms of onset time, pitch, and MIDI channel, to be matched to a given score note *s*. However, in cases where performance errors, expressive timing deviations, or ornaments introduce deviations from the score, a selection procedure must take place to find the optimal fit between score and performance. In such instances, the note-by-note matcher prioritizes exact pitch matches; thus, in a situation where only one of the candidate performance notes has the same pitch as *s*, this note receives the highest possible rating regardless of its onset time difference. If there is no such exact pitch match, the candidates are ranked according to their match rating. Before assigning a performance note *p* to *s*, the matcher verifies that *p* would not be a better match for a neighbouring score note; if this is the case, it moves on to the next best candidate and repeats the same process. If all of the candidates are better matches for other score notes than for *s*, *s* is left unmatched.

Once the entire piece has been matched, the best solution is selected as the one that maximizes the global match rating. Since the match ratings take into account structural as well as temporal information, the best solution is not necessarily the one which matches the highest number of notes. A solution that matches fewer notes but preserves the structural and temporal coherence of the piece to a greater extent may be favoured over one that matches more notes but ends up distorting the temporal structure.

Identification of performance errors and ornaments

The final phase of the matching procedure consists of the identification and categorization of performance errors and ornaments. As described in Chapter 4, the matcher identifies two general types of errors: *score errors* and *non-score errors.* Score errors comprise pitch errors (also called substitutions), omissions (including “added ties” – repeated notes in the score that were not re-attacked in performance), and timing errors, whereas non-score errors include all performance notes that are extraneous to the score, such as intrusions and repetitions (re-attacked notes in performance that were not repeated in the score).[1] The matcher codes errors in a parsimonious manner; that is, in cases where an error could be analyzed as one error or as two distinct errors, the matcher prefers a solution that minimizes the number of errors (Palmer & Van de Sande, 1993).

The distinction between score errors and non-score errors is relevant to the identification of ornaments. Indeed, whereas the interpretation of score errors is generally unambiguous since a score error represents, by definition, the omission or misplaying of a single score note, all non-score errors correspond to unmatched performance notes, which may be theoretically interpreted as ornaments. The problem of ornament identification can thus be recast as an interpretation of the status of unmatched performance notes. The approach privileged here is to assume that, by default, all unmatched performance notes are non-score errors, unless there is substantial evidence that one or more of these notes represent an ornament. In practice, for each unmatched performance note *u*, the matcher evaluates the likelihood that it belongs to an ornament; if this *ornamental rating* is superior to a threshold value, *u* is treated is an ornamental note; otherwise, it is categorized as a non-score error. However, in order to implement this procedure, a general definition of what a performance ornament is needs to be developed. In the following paragraphs, we will introduce some rules and present their implementation in the matching algorithm.

[1] "Untied" notes (Repp, 1996a) are treated as repetitions.

Formal definition of performance ornaments. Musically speaking, ornaments are often referred to as embellishments of a score note. In other words, each ornament can be said to be hierarchically subordinated to a score note in a representation of the musical structure (Lerdahl & Jackendoff, 1983; Schenker, 1987). In the musical realization of a score, this subordination is reflected in the fact that the ornamental notes must occupy the temporal and registral space of the score note that they intend to embellish: a trill occurring in measure 29 cannot normally be associated with a note in measure 14. However, although this concept of *score anchoring* is a necessary condition for a note to be considered an ornament of a score note, it is not a sufficient one: non-score performance errors may also occupy the temporal and registral space of a score note. Another fundamental property of ornamental notes is their *intentionality*: in contrast to random errors, ornaments generally form characteristic melodic figures, which may or may not represent typical patterns such as trills or mordents. This intentionality may be captured by well-formedness rules, elaborated in Gestalt principles.

To be perceived as part of a single ornamental figure, the individual notes that constitute an ornament should be organized temporally and perceptually so as to form a single-stream percept (Bregman, 1990). According to the proximity principle, notes whose onsets and/or pitches are close to each other will tend to be perceived as being connected to each other. Moreover, the percept of a continuous, single melodic line is enhanced if the offset of a note is close to the onset of the following note, so that there are no interruptions in the melodic activity, and if there is a limited overlap between successive notes (Huron, 2001, pp. 12-13). The belongingness principle may also be applied to the case of ornamental notes that are separated from the score note they are embellishing by a large pitch interval, but which belong to the same chord or harmony, as is the case with certain appoggiaturas.

Implementation in the matcher. The matcher first determines, for each score note *s*, whether there are unmatched performance notes that occupy the temporal and registral space of *s*. The temporal space occupied by *s* is bound by the onset of the immediately preceding note in the same voice and the onset of the following note in the same voice, while its registral space is bound by the pitches of score notes that sound together with *s*.[2] If there are performance notes which fit these criteria, they may be considered as potential embellishments to *s*. These notes then receive ornamental ratings, which are determined according to the rules of proximity and belongingness outlined above. Ratings are also influenced by the number of notes involved in the potential embellishment: because unmatched performance notes are more likely to be heard as errors if they occur in isolation rather than forming a coherent group, the matcher assumes that the likelihood of a group of unmatched performance notes being an ornament anchored to *s* increases with the size of the group. Furthermore, ratings take score indications into account: unmatched performance notes are more likely to be treated as embellishments to *s* if there is an indication in the score that *s* should be ornamented in performance.

[2] Note that, according to this definition, the registral space of a monophonic melody is unbound.

The evaluation of potential candidates is an iterative process. Ratings are first computed for all unmatched performance notes associated with a score note *s*; notes whose ornamental ratings are below the threshold value are treated as errors and excluded from the list of potential candidates. However, since the exclusion of a note may affect the ratings of the remaining notes, ornamental ratings are computed again for all remaining notes, until a stable configuration is reached where either all the candidates have ornamental ratings above the threshold value, or no viable candidates are left. A final selection process excludes groups of unmatched performance notes whose mean ornamental rating is below a minimal threshold.

In some instances, an ornament could be potentially anchored to two or more score notes. In these cases, an additional selection step is undertaken to assign the ornament to a single score note. This step uses a hierarchical forced-choice procedure which first prioritizes ornament-score note couplings that contain the greatest number of notes (thus minimizing the number of unmatched performance notes treated as errors), then couplings that maximize the temporal-registral fit between score note and ornament, and, as a last resort, couplings that maximize the mean ornamental rating of the embellishment.

Finally, ornaments are classified into appoggiaturas, mordents, trills, scalar patterns, and “unidentified ornaments”. Since the approach outlined here does not rely on the recognition of specific patterns, the matcher may recognize that certain groups of unmatched performance notes possess all the characteristics of an ornament (such as pitch and time proximity, as well as melodic continuity), even if they do not form a typical ornamental pattern.

Comparison with other offline matchers

To conclude this section, a summary of the principal features of the three-step matcher is provided in Table 6.2, along with a comparison with a few well-known offline matchers. Besides the use of temporal information, one of the main differences between the three-step matcher and other matchers is that it processes performances first at the level of clusters before moving down to the note level; it thus combines the advantages of both approaches, taking into account both voice structure and the grouping of score notes into events.

Table 6.2. Comparison between the three-step matcher and other matchers.

	Strict matcher (Honing, 1990)	Large matcher (Large, 1993)	Structure matcher (Desain et al., 1997)	Three-step matcher
Processing unit	Note	Cluster / event	Note	Cluster / event (steps 1 & 2); note (step 3)
Uses voice information	No	No	Yes	Yes
Uses temporal information	No	No	No	Yes
Solutions considered	One	Several	Several	Several
Definition of best solution	Most matched notes	Most matched notes	Most matched notes, preserves voice structure	Best structural / temporal fit for events (steps 1 & 2) and for notes (step 3)

ASSESSING THE ACCURACY OF THE MATCHER

In order to evaluate the accuracy of the matching algorithm, it is necessary to compare its solutions to those obtained using an independent reliable process. Score-performance matches realized by hand by the first author (a music theorist) on a corpus of 80 MIDI recordings of organ performances were used as ground truth data for this purpose. These recordings consisted of 48 performances of the *Premier Agnus* by Nicolas de Grigny (1672-1703) and 32 performances of *Wachet auf, ruft uns die Stimme* by Samuel Scheidt (1587-1654), for a total of 27,168 score notes. It should be noted that these matches, which we will refer to as *hand matches* (Heijink et al., 2000b), were completed before the programming of the three-step matcher was undertaken (Gingras, 2006); in fact, the amount of work involved in the completion of these hand matches was a primary motivation in the design of this matcher.

In addition, we sought to assess the improvement in matching accuracy brought about by taking into account the temporal information from the MIDI recordings. One way to evaluate this effect would be to compare two matching algorithms that are identical in all respects, except that one uses temporal information and the other does not. To that end, we implemented a version of the three-step matcher that does *not* take into account temporal information (the second step of the matching procedure uses only the chronological succession of the score events) but is otherwise identical to the original algorithm, and compared the results obtained by this implementation to the hand matches.

The three-step matcher was also used to match 32 performances of the Fugue in D minor (BWV 538), also known as the “Dorian” fugue, by J.S. Bach (1685-1750), for a total of 86,432 score notes. However, given the length of the piece, the task of matching the 32 performances by hand would have been prohibitively time-consuming; thus, only a comparison between the matches produced by the temporal and non-temporal implementations of the three-step matcher is presented here.

Method

The scores for the *Premier Agnus* and *Wachet auf* were entered by hand; voice information was included. The score of the Dorian fugue was prepared from a MIDI file obtained from an Internet archive ("Classical music archives", 1994); the MIDI data were hand-edited for errors so that it would match exactly the score of the piece. Voice information was added by hand. Scores were then set up in a format suitable for the matcher.

The matcher was implemented in the MATLAB programming language, and run under Windows XP on a Gateway laptop computer. On this configuration, the time required to match a single performance ranged from 10 to 20 seconds for the *Premier Agnus* and the *Wachet auf*, and from 15 to 35 minutes for the Dorian fugue.

Results

Comparison with hand matches. For each performance of *Premier Agnus* and of *Wachet auf*, the solutions provided by both versions of the three-step matcher were compared to the hand matches, and discrepancies between matches were identified. For each implementation, the percentage of discrepancies with the human matches to the total amount of score notes was computed. In order to provide a benchmark with previous offline matchers, the results are presented alongside those reported in Heijink et al. (2000b), who compared revised implementation of the strict matcher (Honing, 1990), a revised implementation of the Large matcher (Large, 1993), and an implementation of the structure matcher (Desain et al., 1997) to hand matches of piano performances. Excerpts from the Étude in C minor, Op. 10, No. 12, and the Fantaisie Impromptu, Op. 66, both by Fryderyk Chopin (1810-1849), were used for this purpose (Figure 6.1). Since the present article was not based on the same pieces, no direct comparison with the results reported by Heijink et al. (2000b) will be attempted here.

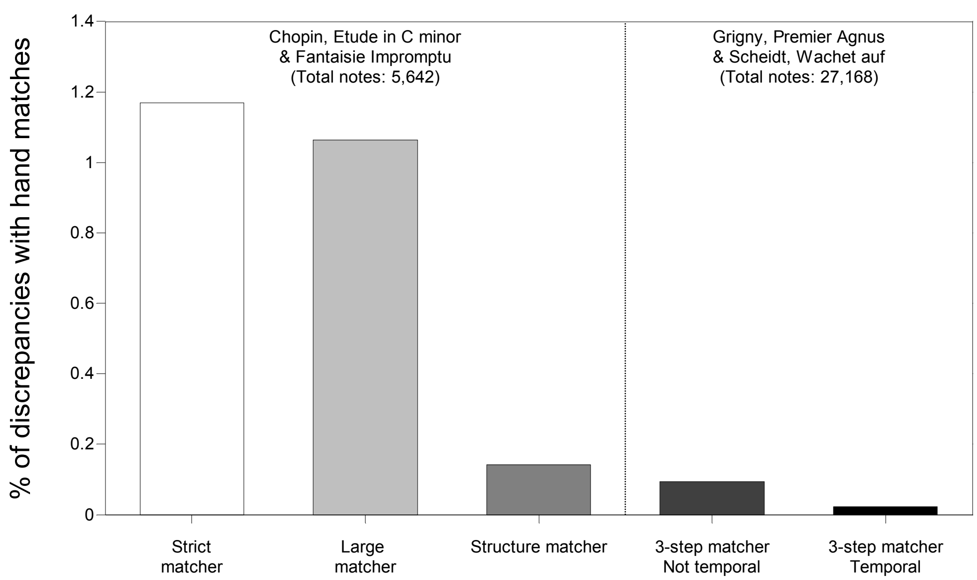


Figure 6.1. Comparison of the discrepancy rate between hand matches and solutions generated by automatic matchers. The results for the strict matcher, the Large matcher, and the structure matcher were obtained from Heijink et al. (2000b).

We note that, whereas 25 discrepancies (out of 27,168 notes) were observed between the hand matches and the solutions obtained using the non-temporal version of the three-step matcher, only 6 discrepancies were identified between the hand matches and those produced by the temporal version of the matcher, a fourfold improvement. This result clearly demonstrates that the use of temporal information substantially improved the matching accuracy.

Analysis of the discrepancies. An inspection of the discrepancies revealed that most of the disagreements between the non-temporal matches and the hand matches of the *Premier Agnus* and the *Wachet auf* involved repeated notes and timing errors. As mentioned previously, repeated notes pose a challenge to offline matchers that do not use temporal information. Likewise, timing errors cannot be properly resolved in the absence of temporal information. However, these discrepancies disappeared when comparing the temporal matches to the hand matches; in fact, after examining the six remaining discrepancies, the first author favours the matcher's interpretation in three of those six cases.

Discrepancies were further analyzed by categorizing them into three groups: Type 1 discrepancies refer to performance notes matched to a different score note in both matches; Type 2 discrepancies correspond to performance notes unmatched in one solution and matched to a score note in the other solution; and Type 3 discrepancies designate performance notes matched to the same score note in both solutions, but that are identified as score errors in one case and not in the other. The distribution of the discrepancies observed between the different matching methods tested here is summarized in Table 6.3. Comparisons between the solutions produced by the temporal and non-temporal implementations of the matcher for the performances of the Dorian fugue are also included.

Table 6.3. Distribution of the discrepancies observed between different matching methods.

		Premier Agnus (15360 notes)	*Wachet auf* (11808 notes)	Dorian fugue (86432 notes)
Hand matches/ temporal matcher				
	Type 1	0	0	
	Type 2	1	0	N/A
	Type 3	2	3	
	Total	3 (0.020%)	3 (0.025%)	
Hand matches / non-temporal matcher				
	Type 1	0	3	
	Type 2	1	4	N/A
	Type 3	13	4	
	Total	14 (0.091%)	11 (0.093%)	
Non-temporal matcher / temporal matcher				
	Type 1	0	3	295
	Type 2	0	2	49
	Type 3	11	5	95
	Total	11 (0.072%)	9 (0.077%)	439 (0.508%)

Note. Percentages refer to the proportion of discrepancies relative to the total number of score notes analyzed for each piece.

Whereas the majority of the discrepancies observed between the temporal and non-temporal implementations for the *Premier Agnus* and the *Wachet auf* belonged to Type 3, most of the discrepancies for the Dorian fugue were classified as Type 1. It should be noted that, in contrast to the recordings of the *Premier Agnus* and of the *Wachet auf* which contained very few ornaments, the performances of the Dorian fugue were heavily ornamented: the temporal implementation of the matcher identified 7.5% of all performance notes as ornamental. Upon close inspection of the matches generated by the temporal version, the authors found themselves in perfect agreement with the solutions provided by the matcher in practically every case. It is especially noteworthy that the matcher could successfully discriminate between ornaments and non-score errors. However, the non-temporal implementation was not nearly as successful, as the presence of ornaments specifically hampered the accuracy of the matches in the sections which were most lavishly embellished. Thus, it is likely that the abundant ornamentation affected the non-temporal implementation to a greater extent than the temporal one. Indeed, 244 (55.6%) of the 439 discrepancies observed for the Dorian fugue involved a note identified as ornamental by one or both implementations. Moreover, nearly all discrepancies involving an ornament (242 of 244) were classified as Type 1, which correspond to mismatched score notes. These results suggest that the use of timing information in automated matching procedures is especially important in the case of ornamented performances.

DISCUSSION

We have presented an offline score-to-performance matching algorithm that relies both on structural and temporal information, allowing it to generate an accurate match even for heavily ornamented performances. A comparison with score-performance hand matches on a corpus of 80 MIDI recordings of organ performances showed a near-perfect agreement between the solutions found by the matcher and the hand matches. Indeed, if the hand matches are treated as ground truth data, our algorithm achieved an accuracy of 99.98%, which corresponds to approximately 1 mismatched note for every 4,500 score notes. This constitutes a significant improvement over offline matchers previously described in the literature, whose best reported success rate was estimated at 99.8%, or approximately 1 mismatch for every 500 notes (Heijink et al., 2000b). As noted by Heijink et al. (2000b, p. 551), the highest possible matching accuracy is required in the context of music performance research, which is the typical domain of application of offline matchers. Thus, we believe that the improvements presented here are non-negligible and make this matcher suitable for large-scale performance studies.

In addition to its increased accuracy, this matcher is designed to accommodate multi-channel MIDI recordings of performances from keyboard instruments with multiple manuals, such as organ or harpsichord; it was actually used to study performances of complex organ pieces, such as J.S. Bach's "Dorian" fugue, in the context of performance research (see Chapters 4 and 5). This feature makes it a potentially valuable tool for the investigation of ensemble performances of MIDI instruments.

We have also proposed a heuristic for the identification of ornaments and errors that is based on perceptual principles, and which could theoretically be amenable to empirical study. It is worth noting that the approach described here does not rely on the recognition of specific patterns, in contrast to the technique pioneered by Dannenberg and Mukaino (1988); instead, it proceeds from a very general definition of performance ornaments to the identification of typical embellishment figures.

As this description of the ornament identification heuristic suggests, the accuracy of automatic matching algorithms could greatly benefit from implementing a model of basic perceptual principles of music cognition. Indeed, as noted by Desain et al. (1997), the fact that human listeners have no difficulty in matching scores to performances implies that modeling perceptual processes might help in resolving remaining challenges associated with score-performance matching. As an example, we may note that the matcher does not take into account scale and chord structure in its current implementation. For instance, a series of notes which constitute an E major arpeggio are all part of the same harmony; they will be perceived as more similar to each other by a human listener familiar with this musical style than other notes which do not belong to the E major chord. Applying this to the analysis of performance errors, a B might be a more likely substitution error for a G# in the context of an E major arpeggio than an A#, even though the pitch interval between G# and A# is smaller than that between B and G#. However, our algorithm is insensitive to the notion of harmonic context; moreover, the pitch distance rating used by the matcher is a simple measure of the interval in semitones between two notes.

The implementation of a hierarchical pitch space model such as that proposed by Lerdahl (2001) might allow the matcher to arrive at more accurate solutions for tonal excerpts. Although this model is style-specific and could prove irrelevant, if not detrimental, to the processing of atonal music or music from non-Western styles, we nevertheless believe that the efficiency of matching algorithms would greatly benefit from the integration of concepts such as scale and chord structure, and perhaps of notions such as consonance and dissonance. While pointing out the limitations of current algorithms, these suggestions underline the importance of issues related to the representation of musical similarity and to the larger question of the modeling of musical intelligence in the development of more effective matching paradigms.

ACKNOWLEDGMENTS

This research was supported by fellowships from the Social Sciences and Humanities Research Council of Canada and from the Centre for Interdisciplinary Research in Music Media and Technology (CIRMMT) to Bruno Gingras, as well as a grant from the Natural Sciences and Engineering Research Council and a Canada Research Chair awarded to Stephen McAdams.

REFERENCES

Bregman, A. S. (1990). *Auditory scene analysis: The perceptual organization of sound*. Cambridge, Mass.: MIT Press.

Classical music archives. (1994). Retrieved January 5, 2007, from http://www.classicalarchives.com/

Dannenberg, R. B. (1984). An on-line algorithm for real-time accompaniment. In *Proceedings of the 1988 International Computer Music Conference* (pp. 243-249). San Francisco: ICMA.

Dannenberg, R. B., & Mukaino, H. (1988). New techniques for enhanced quality of computer accompaniment. In *Proceedings of the 1988 International Computer Music Conference* (pp. 243-249). San Francisco: ICMA.

Desain, P., & Honing, H. (1992). *Music, mind, and machine: Studies in computer music, music cognition, and artificial intelligence*. Amsterdam: Thesis Publishers.

Desain, P., & Honing, H. (1993). Tempo curves considered harmful. *Contemporary music review, 7*(2), 123-138.

Desain, P., Honing, H., & Heijink, H. (1997). Robust score-performance matching: Taking advantage of structural information. In *Proceedings of the 1997 International Computer Music Conference* (pp. 337-340). San Francisco: ICMA.

Dixon, S. (2005). MATCH: A music alignment tool chest. In *6th International Conference on Music Information Retrieval* (pp. 492-497). London, UK.

Gingras, B. (2006). Emphasizing voices in polyphonic organ music: Issues of expressive performance on an instrument with fixed-tone intensity. In M. Baroni, A. R. Addessi, R. Caterina & M. Costa (Eds.), *Proceedings of the 9th International Conference on Music Perception and Cognition* (pp. 1712-1720). Bologna, Italy.

Goebl, W. (2001). Melody lead in piano performance: Expressive device or artifact? *Journal of the Acoustical Society of America, 110*(1), 563-572.

Gotoh, O. (1982). An improved algorithm for matching biological sequences. *Journal of Molecular Biology, 162*(3), 705-708.

Heijink, H. (1996). *Matching scores and performances.* Unpublished master's thesis, Nijmegen University.

Heijink, H., Desain, P., Honing, H., & Windsor, L. (2000a). Make me a match: An evaluation of different approaches to score-performance matching. *Computer Music Journal, 24*(1), 43-56.

Heijink, H., Windsor, L., & Desain, P. (2000b). Data processing in music performance research: Using structural information to improve score-performance matching. *Behavior Research Methods Instruments & Computers, 32*(4), 546-554.

Honing, H. (1990). POCO: An environment for analyzing, modifying, and generating expression in music. In *Proceedings of the International Computer Music Conference* (pp. 364-368). San Francisco: ICMA.

Hoshishiba, T., Horiguchi, S., & Fujinaga, I. (1996). Study of expression and individuality in music performance using normative data derived from MIDI recordings of piano music. In *Proceedings of the 4th International Conference on Music Perception and Cognition* (pp. 465-470). Montreal: McGill University, Faculty of Music.

Huron, D. (2001). Tone and voice: A derivation of the rules of voice-leading from perceptual principles. *Music Perception, 19*(1), 1-64.

Kendall, R. A., & Carterette, E. C. (1990). The communication of musical expression. *Music Perception, 8*(2), 129-164.

Large, E. W. (1993). Dynamic programming for the analysis of serial behaviors. *Behavior Research Methods Instruments & Computers, 25*(2), 238-241.

Lerdahl, F. (2001). *Tonal pitch space*. Oxford; New York: Oxford University Press.

Lerdahl, F., & Jackendoff, R. (1983). *A generative theory of tonal music*. Cambridge, Mass.: MIT Press.

McAdams, S., Vines, B. W., Vieillard, S., Smith, B. K., & Reynolds, R. (2004). Influences of large-scale form on continuous ratings in response to a contemporary piece in a live concert setting. *Music Perception, 22*(2), 297-350.

Mongeau, M., & Sankoff, D. (1990). Comparison of musical sequences. *Computers and the Humanities, 24*(3), 161-175.

Needleman, S. B., & Wunsch, C. D. (1970). A general method applicable to search for similarities in amino acid sequence of 2 proteins. *Journal of Molecular Biology, 48*(3), 443-453.

Palmer, C. (1989). Mapping musical thought to musical performance. *Journal of Experimental Psychology-Human Perception and Performance, 15*(12), 331-346.

Palmer, C. (1996). On the assignment of structure in music performance. *Music Perception, 14*(1), 23-56.

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

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.

Pardo, B., & Birmingham, W. (2001). Following a musical performance from a partially specified score, *Multimedia Technology Applications Conference*. Irvine, California.

Puckette, M., & Lippe, C. (1992). Score following in practice. In *Proceedings of the 1992 International Computer Music Conference* (pp. 182-185). San Francisco: ICMA.

Raphael, C. (2006). Aligning music scores with symbolic scores using a hybrid graphical model. *Machine Learning, 65*, 389-409.

Rasch, R. A. (1979). Synchronization in performed ensemble music. *Acustica, 43*(2), 121-131.

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). Patterns of note onset asynchronies in expressive piano performance. *Journal of the Acoustical Society of America, 100*(6), 3917-3931.

Schenker, H. (1987). *Counterpoint: A translation of Kontrapunkt* (J. Rothgeb, Trans.). New York, London: Schirmer Books; Collier Macmillan.

Schwarz, D., Orio, N., & Schnell, N. (2004). Robust polyphonic Midi score following with Hidden Markov Models. In *Proceedings of the International Computer Music Conference (ICMC)* (pp. 442-445). Miami, Florida: International Computer Music Association.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal, 27*, 379-423, 623-656.

Vantomme, J. D. (1995). Score following by temporal pattern. *Computer Music Journal, 19*(3), 50-59.

Chapter 7. Conclusions

This dissertation investigated expressive strategies and performer-listener communication in organ performance. Four core issues were explored: the communication of voice emphasis (Chapter 2), the communication of artistic individuality (Chapter 3), the influence of musical structure on error patterns (Chapter 4), and the relationship between performers' interpretive choices and their analyses of the formal structure of a piece (Chapter 5).

Two series of experiments were conducted: the first of which involved the analysis of recordings of organ pieces by skilled performers, whereas the second sought to obtain behavioral measurements of the listeners' perception of specific aspects of these performances, such as voice emphasis and artistic individuality.

All performances were recorded on an organ equipped with a MIDI console. The use of MIDI technology allowed an accurate analysis of performance parameters such as tempo, articulation, and onset asynchrony. The MIDI data were matched to the scores using a new score-performance matching algorithm written specifically for this research project which is described in Chapter 6.

Three organ pieces from the Baroque period were chosen for this project. The *Premier Agnus* by Nicolas de Grigny (1672-1703) was used to study the communication of voice emphasis, while the investigation of the communication of artistic individuality was conducted using the chorale setting of *Wachet auf, ruft uns die Stimme* by Samuel Scheidt (1587-1654). The exploration of the relationship between performers' interpretive choices and their analytical decisions was based on a comparison of the performers' recordings and of their written analyses of the Fugue in D minor (BWV 538), also known as the Dorian fugue, by J.S. Bach (1685-1750). Data from the performances of all three pieces were used for the study on error patterns.

A number of intriguing findings on expressive strategies and communication in organ performance have emerged from the collection of studies presented in this thesis. Firstly, I have shown in Chapter 2 that articulation was the main expressive parameter used by organists to emphasize a voice in polyphonic organ music. Indeed, the modification of articulation patterns was found to be the most widespread and consistent strategy used by organists to emphasize a voice. However, behavioral data suggest that structural elements in the musical score play a more important role in the perception of voice prominence than expressive cues in performance. Indeed, invariant peaks of relative perceptual prominence corresponding to salient passages in specific voices were observed across interpretations and performers. Furthermore, although listeners who were themselves organists were more sensitive to differences between performers and interpretations than non-organists, the performers' intentions were for the most part not recognized.

Conversely, the results presented in Chapter 3 indicate that the communication of artistic individuality can be achieved even on an instrument with a limited range of expressive parameters such as the organ. The majority of participants performed significantly above chance in a sorting task in which they were asked to group together performances they thought had been played by the same performer. Furthermore, there was no significant difference between the performance of musicians and non-musicians. Mean tempo and articulation were found to be the most important dimensions along which listeners differentiated the excerpts. It is noteworthy that whereas contrasts in articulation were apparently inefficient in communicating voice emphasis (Chapter 2), they constituted one of the main features used by listeners to discriminate between performers. This implies that although listeners can perceive differences in articulation between performances, they may not be able to relate them to a specific expressive intent. One of the most provocative findings of this study was that sorting accuracy was found to be significantly higher for prize-winning performers than for non-winners, suggesting that the performers' ability to convey a sense of artistic individuality was linked to their level of expertise. Moreover, sorting accuracy was generally higher for performers who exhibited either greater consistency or distinctiveness in their recordings. These observations point to interesting links between the performers' level of accomplishment and their ability to convey a sense of artistic individuality, which warrant further inquiry.

The investigation of error patterns in organ performance (Chapter 4) revealed that the pattern of performance errors was closely associated with the musical structure and with the performers' expressive intentions. Thus, error rates were lower for motivic notes than for non-motivic notes, and fewer errors were committed in a voice when it was emphasized than when it was not. 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 salience and musical significance or familiarity. In addition, error patterns were found to be performer-specific: individual performers exhibited consistent and idiosyncratic error patterns.

The exploration of structure-performance relationships in performances of the Dorian fugue by professional organists (Chapter 5) revealed that most major tempo variations coincided with formal features such as cadences and subject entries. Nevertheless, a number of large tempo deviations were associated with particular features of the piece that are not highlighted in traditional music-theoretical analysis. such as the successive recurrences of a canonic episode that reappears several times over the course of the fugue. Furthermore, individual performers' interpretative choices did not necessarily correspond to their written analyses.

While the results presented in Chapter 6 are not specifically related to the study of expressive strategies in organ performance, I believe that the innovations in the realm of score-performance matching that are introduced in this chapter have set the stage for new work in the analysis of musical ornamentation and performance errors that would not have been possible in such a rigorous and automated fashion in the past. Moreover, the approach used by the matcher for the identification of ornaments and errors is based on perceptual principles and could theoretically be amenable to empirical study.

In conclusion, score-based music performance involves several aspects which are interrelated to a large extent: the performer's understanding and conception of the structure of the piece, the interpretative choices involved in its realization, and the expressive means used to convey the chosen interpretation. The performer's expressive intentions may focus both on local elements (such as bringing out a specific melody or motive) and on large-scale issues (such as conveying the form of the piece through tempo variations). In addition, the expressive means used by the performer must be considered in relation not only to his or her interpretive goals, but also in light of the possibilities and limitations of the instrument, the structure and character of the piece, as well as the general performance traditions and prescriptions associated with the style or period to which the piece belongs. Indeed, whereas certain expressive features, such as the means used to emphasize a voice, appear to be instrument-specific (Chapter 2), others, such as time-contour profiles, may be similar across different instruments (Chapter 5). Furthermore, expressive intentions and interpretative choices, both on a local and on a large-scale level, are largely determined by a performer's artistic individuality. Artistic individuality is manifested at every level of the performance: idiosyncratic patterns are found at the level of the note-by-note articulation and onset asynchrony patterns (Chapters 2 and 3), but also in large-scale tempo variations (Chapter 5), and even in error patterns (Chapter 4).

As noted by several scholars, the empirical analysis of music performance data may be more meaningful when considered in the context of a communication process (Gabrielsson, 2003; Kendall & Carterette, 1990). This thesis presents an integrative framework for music performance research that analyzes the phenomenon of communication in music performance from several different angles: the expressive means used by the performer to express an intention, the perception of those intentions by the listener, as well as the music-theoretical analysis of the pieces. By juxtaposing these complementary viewpoints, this dissertation proposes both an inclusive experimental paradigm and a more holistic approach to music performance research.

Future research projects involve an extension of my doctoral research to harpsichord performance, and a study of the perceptual determinants of artistic individuality and aesthetic appeal in classical piano performance. Like the organ, the harpsichord affords very limited possibilities regarding dynamic differentiation of individual notes. However, it remains to be seen whether the expressive strategies observed in organ performance are also used in harpsichord performance. Links between artistic individuality and aesthetic appreciation are strongly suggested by the results presented in Chapter 3, and definitely warrant further investigation. I envision this as a fertile research undertaking which could lead to fruitful collaborations and potential educational applications, while creating sustained interest in the musical community.

REFERENCES

Gabrielsson, A. (2003). Music performance research at the millenium. *Psychology of Music, 31*(3), 221-272.

Kendall, R. A., & Carterette, E. C. (1990). The communication of musical expression. *Music Perception, 8*(2), 129-164.

Bibliography

Arabie, P., & Boorman, S. A. (1973). Multidimensional scaling of measures of distance between partitions. *Journal of Mathematical Psychology, 10*(2), 148-203.

Askenfelt, A., & Jansson, E. V. (1992). On vibration sensation and finger touch in stringed-instrument playing. *Music Perception, 9*(3), 311-349.

Belin, P., Zatorre, R. J., Lafaille, P., Ahad, P., & Pike, B. (2000). Voice-selective areas in human auditory cortex. *Nature, 403*(6767), 309-312.

Benadon, F. (2003). Spectrographic and calligraphic cues in the identification of jazz saxophonists. In R. Kopiez, A. C. Lehmann, I. Wolther & C. Wolf (Eds.), *Proceedings of the 5th Triennial ESCOM Conference* (pp. 245-249). Hanover, Germany.

Bengtsson, I., & Gabrielsson, A. (1983). Analysis and synthesis of musical rhythm. In J. Sundberg (Ed.), *Studies of music performance* (pp. 27-60). Stockholm: Royal Swedish Academy of Music.

Berry, W. (1989). *Musical structure and performance*. New Haven: Yale University Press.

Borg, I., & Groenen, P. J. F. (1997). *Modern multidimensional scaling: Theory and applications*. New York: Springer.

Braasch, J., & Ahrens, C. (2000). Attack transients of free reed pipes in comparison to striking reed pipes and diapason pipes. *Acustica, 86*, 662-670.

Bregman, A. S. (1990). *Auditory scene analysis: The perceptual organization of sound*. Cambridge, Mass.: MIT Press.

Bregman, A. S., Ahad, P. A., Crum, P. A. C., & O'Reilly, J. (2000). Effects of time intervals and tone durations on auditory stream segregation. *Perception & Psychophysics, 62*(3), 626-636.

Brochard, R., Drake, C., Botte, M. C., & McAdams, S. (1999). Perceptual organization of complex auditory sequences: Effect of number of simultaneous subsequences and frequency separation. *Journal of Experimental Psychology-Human Perception and Performance, 25*(6), 1742-1759.

Brown, R. (1981). An experimental study of the relative importance of acoustic parameters for auditory speaker recognition. *Language and Speech, 24*(4), 295-310.

Bullivant, R. (1971). *Fugue*. London: Hutchinson.

Caldwell, J. (2007). Keyboard music to c1750, I.4. The period of J.S. Bach. *Grove music online*, ed. L. Macy. Retrieved April 6, 2008, from <http://www.grovemusic.com>

Carroll, J. D., & Chang, J. J. (1970). Analysis of individual differences in multidimensional scaling via an N-way generalization of Eckart-Young decomposition. *Psychometrika, 35*(3), 283-319.

Clarke, E. F. (1985). Structure and expression in rhythmic performance. In P. Howell, I. Cross & R. West (Eds.), *Musical structure and cognition* (pp. 209-236). London: Academic Press.

Clarke, E. F. (1988). Generative principles in music performance. In J. Sloboda (Ed.), *Generative processes in music: The psychology of performance, improvisation and composition* (pp. 1-26).

Clarke, E. F. (1989). The perception of expressive timing in music. *Psychological Research-Psychologische Forschung, 51*(1), 2-9.

Clarke, E. F. (2002). Listening to performance. In J. Rink (Ed.), *Musical performance: A guide to understanding* (pp. 185-196). United Kingdom: Cambridge University Press.

Classical music archives. (1994). Retrieved January 5, 2007, from http://www.classicalarchives.com/

Cone, E. T. (1968). *Musical form and musical performance* (1st ed.). New York: W. W. Norton.

Cook, N. (1999). Analysing performance and performing analysis. In N. Cook & M. Everist (Eds.), *Rethinking music* (pp. 239-261). United Kingdom: Oxford University Press.

Cook, N. (2007). *Music, performance, meaning: selected essays*. Aldershot, England; Burlington, VT: Ashgate.

Cook, N., Johnson, P., & Zender, H. (1999). *Theory into practice: Composition, performance and the listening experience*. Leuven, Belgium: Leuven University Press.

Dannenberg, R. B. (1984). An on-line algorithm for real-time accompaniment. In *Proceedings of the 1988 International Computer Music Conference* (pp. 243-249). San Francisco: ICMA.

Dannenberg, R. B., & Mukaino, H. (1988). New techniques for enhanced quality of computer accompaniment. In *Proceedings of the 1988 International Computer Music Conference* (pp. 243-249). San Francisco: ICMA.

Davies, S. (2001). *Musical works and performances: A philosophical exploration*. Oxford, England; New York: Clarendon Press; Oxford University Press.

Daws, J. T. (1996). The analysis of free-sorting data: Beyond pairwise cooccurrences. *Journal of Classification, 13*(1), 57-80.

De Poli, G., Roda, A., & Vidolin, A. (1998). Note-by-note analysis of the influence of expressive intentions and musical structure in violin performance. *Journal of New Music Research, 27*(3), 293-321.

Desain, P., & Honing, H. (1992). *Music, mind, and machine: Studies in computer music, music cognition, and artificial intelligence*. Amsterdam: Thesis Publishers.

Desain, P., & Honing, H. (1993). Tempo curves considered harmful. *Contemporary music review, 7*(2), 123-138.

Desain, P., Honing, H., & Heijink, H. (1997). Robust score-performance matching: Taking advantage of structural information. In *Proceedings of the 1997 International Computer Music Conference* (pp. 337-340). San Francisco: ICMA.

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.

Dixon, S. (2005). MATCH: A music alignment tool chest. In *6th International Conference on Music Information Retrieval* (pp. 492-497). London, UK.

Drake, C., & Palmer, C. (1993). Accent structures in music performance. *Music Perception, 10*(3), 343-378.

Drake, C., & Palmer, C. (2000). Skill acquisition in music performance: Relations between planning and temporal control. *Cognition, 74*(1), 1-32.

Efron, B., & Tibshirani, R. (1993). *An introduction to the bootstrap*. New York: Chapman & Hall.

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.

Gabrielsson, A. (1987). Once again: The theme from Mozart's piano sonata in A major (K. 331). In A. Gabrielsson (Ed.), *Action and perception in rhythm and music* (pp. 81-104). Stockholm: Royal Swedish Academy of Music.

Gabrielsson, A. (2003). Music performance research at the millenium. *Psychology of Music, 31*(3), 221-272.

Gingras, B. (2006). Emphasizing voices in polyphonic organ music: Issues of expressive performance on an instrument with fixed-tone intensity. In M. Baroni, A. R. Addessi, R. Caterina & M. Costa (Eds.), *Proceedings of the 9th International Conference on Music Perception and Cognition* (pp. 1712-1720). Bologna, Italy.

Giordano, B. L., McAdams, S., & McDonnell, J. (2007). Acoustical and conceptual information for the perception of animate and inanimate sound sources. In *Proceedings of ICAD-07 (13th Meeting of the International Conference on Auditory Display)* (pp. 173-181). Montreal, Canada.

Goebl, W. (2001). Melody lead in piano performance: Expressive device or artifact? *Journal of the Acoustical Society of America, 110*(1), 563-572.

Goebl, W., & Bresin, R. (2003). Measurement and reproduction accuracy of computer-controlled grand pianos. *Journal of the Acoustical Society of America, 114*(4), 2273-2283.

Goebl, W., & Parncutt, R. (2002). The influence of relative intensity on the perception of onset asynchronies. In C. Stevens, D. Burnham, G. E. McPherson, E. Schubert & J. Renwick (Eds.), *Proceedings of the 7th International Conference on Music Perception and Cognition*. Sydney, Australia: Adelaide: Causal Productions.

Gotoh, O. (1982). An improved algorithm for matching biological sequences. *Journal of Molecular Biology, 162*(3), 705-708.

Hallam, S. (1997). Approaches to instrumental music practice of experts and novices: Implications for education. In *Does practice make perfect? Current theory and research on instrumental music practice* (pp. 89). Norway: Norges Musikkhogskole.

Hallam, S. (2001). The development of metacognition in musicians: Implications for education. *British Journal of Music Education, 18*(1), 27-39.

Heijink, H. (1996). *Matching scores and performances.* Unpublished master's thesis, Nijmegen University.

Heijink, H., Desain, P., Honing, H., & Windsor, L. (2000). Make me a match: An evaluation of different approaches to score-performance matching. *Computer Music Journal, 24*(1), 43-56.

Heijink, H., & Meulenbroek, R. G. J. (2002). On the complexity of classical guitar playing: Functional adaptations to task constraints. *Journal of Motor Behavior, 34*(4), 339-351.

Heijink, H., Windsor, L., & Desain, P. (2000b). Data processing in music performance research: Using structural information to improve score-performance matching. *Behavior Research Methods Instruments & Computers, 32*(4), 546-554.

Hirsh, I. J. (1959). Auditory perception of temporal order. *Journal of the Acoustical Society of America, 31*(6), 759–767.

Holmgren, G. L. (1967). Physical and psychological correlates of speaker recognition. *Journal of Speech and Hearing Research, 10*(1), 57-66.

Honing, H. (1990). POCO: An environment for analyzing, modifying, and generating expression in music. In *Proceedings of the International Computer Music Conference* (pp. 364-368). San Francisco: ICMA.

Hoshishiba, T., Horiguchi, S., & Fujinaga, I. (1996). Study of expression and individuality in music performance using normative data derived from MIDI recordings of piano music. In *Proceedings of the 4th International Conference on Music Perception and Cognition* (pp. 465-470). Montreal: McGill University, Faculty of Music.

Hubert, L., & Arabie, P. (1985). Comparing partitions. *Journal of Classification, 2*(2-3), 193-218.

Huron, D. (1989). Voice denumerability in polyphonic music of homogeneous timbres. *Music Perception, 6*(4), 361-382.

Huron, D. (1993). Note-onset asynchrony in J.S. Bach's two-part inventions. *Music Perception, 10*(4), 435-444.

Huron, D. (2001). Tone and voice: A derivation of the rules of voice-leading from perceptual principles. *Music Perception, 19*(1), 1-64.

Huron, D., & Fantini, D. A. (1989). The avoidance of inner-voice entries - perceptual evidence and musical practice. *Music Perception, 7*(1), 43-47.

Jerkert, J. (2004, June). *Musical articulation in the organ.* Paper presented at the Joint Baltic-Nordic Acoustics Meeting, Mariehamn, Finland.

Juslin, P. N. (2000). Cue utilization in communication of emotion in music performance: Relating performance to perception. *Journal of Experimental Psychology-Human Perception and Performance, 26*(6), 1797-1812.

Juslin, P. N. (2001). Communicating emotion in music performance: A review and a theoretical framework. In P. N. Juslin & J. Sloboda (Eds.), *Music and emotion: Theory and research* (pp. 309-337). New York: Oxford University Press.

Keller, P. E., Knoblich, G., & Repp, B. H. (2007). Pianists duet better when they play with themselves: On the possible role of action simulation in synchronization. *Consciousness and Cognition, 16*(1), 102-111.

Kendall, R. A., & Carterette, E. C. (1990). The communication of musical expression. *Music Perception, 8*(2), 129-164.

Kruskal, J. B., & Wish, M. (1978). *Multidimensional scaling*. Beverly Hills, Calif.: Sage Publications.

Laeng, B., & Park, A. (1999). Handedness effects on playing a normal or reversed keyboard. *Laterality, 4*(4), 363-377.

Langlois, J. H., & Roggman, L. A. (1990). Attractive faces are only average. *Psychological Science, 1*(2), 115-121.

Large, E. W. (1993). Dynamic programming for the analysis of serial behaviors. *Behavior Research Methods Instruments & Computers, 25*(2), 238-241.

Lerdahl, F. (2001). *Tonal pitch space*. Oxford; New York: Oxford University Press.

Lerdahl, F., & Jackendoff, R. (1983). *A generative theory of tonal music*. Cambridge, Mass.: MIT Press.

Lester, J. (1995). Performance and analysis: Interaction and interpretation. In J. Rink (Ed.), *The practice of performance: Studies in musical interpretation* (pp. 197-216). United Kingdom: Cambridge University.

Lester, J. (2001). Heightening levels of activity and J.S. Bach's parallel-section constructions. *Journal of the American Musicological Society, 54*(1), 49-96.

Loula, F., Prasad, S., Harber, K., & Shiffrar, M. (2005). Recognizing people from their movement. *Journal of Experimental Psychology-Human Perception and Performance, 31*(1), 210-220.

Margulis, E. H., Mlsna, L. M., Uppunda, A. K., Parrish, T. B., & Wong, P. C. (2007). Selective neurophysiologic responses to music in instrumentalists with different listening biographies. *Human Brain Mapping* (Published online: 2007 Dec. 10).

McAdams, S., Vines, B. W., Vieillard, S., Smith, B. K., & Reynolds, R. (2004). Influences of large-scale form on continuous ratings in response to a contemporary piece in a live concert setting. *Music Perception, 22*(2), 297-350.

Methuen-Campbell, J. (1992). Chopin in performance. In J. Samson (Ed.), *The Cambridge companion to Chopin*. Cambridge, England: Cambridge University Press.

Meyer, L. B. (1973). *Explaining music: Essays and explorations*. Berkeley: University of California Press.

Miller, G. A. (1969). A psychological method to investigate verbal concepts. *Journal of Mathematical Psychology, 6*(2), 169-191.

Mongeau, M., & Sankoff, D. (1990). Comparison of musical sequences. *Computers and the Humanities, 24*(3), 161-175.

Moore, G. P. (1992). Piano trills. *Music Perception, 9*(3), 351-360.

Narmour, E. (1988). On the relationship of analytical theory to performance and interpretations. In E. Narmour & R. A. Solie (Eds.), *Explorations in music, the arts, and ideas* (pp. 317-340). Stuyvesant, NY: Pendragon.

Needleman, S. B., & Wunsch, C. D. (1970). A general method applicable to search for similarities in amino acid sequence of 2 proteins. *Journal of Molecular Biology, 48*(3), 443-453.

Nielsen, S. G. (1999). Learning strategies in instrumental music practice. *British Journal of Music Education, 16*(3), 275-291.

Palmer, C. (1989). Mapping musical thought to musical performance. *Journal of Experimental Psychology-Human Perception and Performance, 15*(12), 331-346.

Palmer, C. (1996). On the assignment of structure in music performance. *Music Perception, 14*(1), 23-56.

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

Palmer, C. (2005). Sequence memory in music performance. *Current Directions in Psychological Science, 14*(5), 247-250.

Palmer, C., & Brown, J. C. (1991). Investigations in the amplitude of sounded piano tones. *Journal of the Acoustical Society of America, 90*(1), 60-66.

Palmer, C., & Drake, C. (1997). Monitoring and planning capacities in the acquisition of music performance skills. *Canadian Journal of Experimental Psychology-Revue Canadienne De Psychologie Experimentale, 51*(4), 369-384.

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.

Palmer, C., Jungers, M. K., & Jusczyk, P. W. (2001). Episodic memory for musical prosody. *Journal of Memory and Language, 45*(4), 526-545.

Pardo, B., & Birmingham, W. (2001). Following a musical performance from a partially specified score, *Multimedia Technology Applications Conference*. Irvine, California.

Parncutt, R., & McPherson, G. (2002). *The science & psychology of music performance: Creative strategies for teaching and learning*. Oxford; New York: Oxford University Press.

Penel, A., & Drake, C. (1998). Sources of timing variations in music performance: A psychological segmentation model. *Psychological Research-Psychologische Forschung, 61*(1), 12-32.

Penel, A., & Drake, C. (2004). Timing variations in music performance: Musical communication, perceptual compensation, and/or motor control? *Perception & Psychophysics, 66*(4), 545-562.

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.

Peters, M. (1985). Performance of a rubato-like task: When two things cannot be done at the same time. *Music Perception, 2*(4), 471-482.

Pfordresher, P. Q., Palmer, C., & Jungers, M. K. (2007). Speed, accuracy, and serial order in sequence production. *Cognitive Science, 31*(1), 63-98.

Prout, E. (1891). Fugal structure. *Proceedings of the Musical Association, 18*(1), 135-159.

Puckette, M., & Lippe, C. (1992). Score following in practice. In *Proceedings of the 1992 International Computer Music Conference* (pp. 182-185). San Francisco: ICMA.

Raphael, C. (2006). Aligning music scores with symbolic scores using a hybrid graphical model. *Machine Learning, 65*, 389-409.

Rasch, R. A. (1979). Synchronization in performed ensemble music. *Acustica, 43*(2), 121-131.

Repp, B. H. (1990). Patterns of expressive timing in performances of a Beethoven minuet by 19 famous pianists. *Journal of the Acoustical Society of America, 88*(2), 622-641.

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. (1995). Expressive timing in Schumann's "Träumerei" - an analysis of performances by graduate student pianists. *Journal of the Acoustical Society of America, 98*(5), 2413-2427.

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. (1997). The aesthetic quality of a quantitatively average music performance: Two preliminary experiments. *Music Perception, 14*(4), 419-444.

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.

Rink, J. (1995a). Playing in time: Rhythm, metre and tempo in Brahms's Fantasien op. 116. In J. Rink (Ed.), *The practice of performance: Studies in musical interpretation* (pp. 254-282). United Kingdom: Cambridge University.

Rink, J. (1995b). *The practice of performance: Studies in musical interpretation*. Cambridge; New York: Cambridge University Press.

Rink, J. (2002). *Musical performance: A guide to understanding*. Cambridge, U.K.; New York: Cambridge University Press.

Roads, C. (1996). *The computer music tutorial*. Cambridge, Mass.: MIT Press.

Rothstein, W. (1995). Analysis and the act of performance. In J. Rink (Ed.), *The practice of performance: Studies in musical interpretation* (pp. 217-240). United Kingdom: Cambridge University.

Rumelhart, D. E., & Norman, D. A. (1982). Simulating a skilled typist - a study of skilled cognitive-motor performance. *Cognitive Science, 6*(1), 1-36.

Schenker, H. (1987). *Counterpoint: A translation of Kontrapunkt* (J. Rothgeb, Trans.). New York, London: Schirmer Books; Collier Macmillan.

Schmalfeldt, J. (1985). On the relation of analysis to performance: Beethoven's bagatelles op. 126, nos. 2 and 5. *Journal of Music Theory, 29*(1), 1-31.

Schwarz, D., Orio, N., & Schnell, N. (2004). Robust polyphonic Midi score following with Hidden Markov Models. In *Proceedings of the International Computer Music Conference (ICMC)* (pp. 442-445). Miami, Florida: International Computer Music Association.

Shaffer, L. H. (1976). Intention and performance. *Psychological Review, 83*(5), 375-393.

Shaffer, L. H. (1981). Performances of Chopin, Bach, and Bartok: Studies in motor programming. *Cognitive Psychology, 13*(3), 326-376.

Shaffer, L. H., & Todd, N. P. M. (1987). The interpretive component in musical performance. In A. Gabrielsson (Ed.), *Action and perception in rhythm and music* (pp. 139-152). Stockholm: Royal Swedish Academy of Music.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal, 27*, 379-423, 623-656.

Sloboda, J. A. (2000). Individual differences in music performance. *Trends in Cognitive Sciences, 4*(10), 397-403.

Smith, B. K. (1995). *PsiExp: An environment for psychoacoustic experimentation using the IRCAM musical workstation.* Paper presented at the Society for Music Perception and Cognition, University of California, Berkeley.

Stamatatos, E., & Widmer, G. (2005). Automatic identification of music performers with learning ensembles. *Artificial Intelligence, 165*(1), 37-56.

Timmers, R. (2005). Predicting the similarity between expressive performance of music from measurements of tempo and dynamics. *Journal of the Acoustical Society of America, 117*(1), 391-399.

Todd, N. (1985). A model of expressive timing in tonal music. *Music Perception, 3*(1), 33-58.

Van Dommelen, W. A. (1990). Acoustic parameters in human speaker recognition. *Language and Speech, 33*(3), 259-272.

Vantomme, J. D. (1995). Score following by temporal pattern. *Computer Music Journal, 19*(3), 50-59.

Vines, B. W., Krumhansl, C. L., Wanderley, M. M., & Levitin, D. J. (2006). Cross-modal interactions in the perception of musical performance. *Cognition, 101*(1), 80-113.

Voiers, W. D. (1964). Perceptual bases of speaker identity. *Journal of the Acoustical Society of America, 36*(6), 1065-1073.

Walker, P. (2008). Counter-exposition. *Grove music online*, ed. L. Macy. Retrieved April 6, 2008, from <http://www.grovemusic.com>

Wan, C. Y., & Huon, G. F. (2005). Performance degradation under pressure in music: An examination of attentional processes. *Psychology of Music, 33*(2), 155-172.

Watt, R., & Ash, R. (1998). A psychological investigation of meaning in music. *Musicae Scientiae, 2*(1), 33-54.

Watt, R., & Quinn, S. (2007). Some robust higher-level percepts for music. *Perception, 36*(12), 1834-1848.

Widmer, G., & Goebl, W. (2004). Computational models of expressive music performance: The state of the art. *Journal of New Music Research, 33*(3), 203-216.

Williams, P. F. (2003). *The organ music of J. S. Bach*. Cambridge: Cambridge University Press.

Wilson, F. R. (1992). Digitizing digital dexterity: A novel application for MIDI recordings of keyboard performance. *Psychomusicology, 11*(2), 79-95.

Wright, J. K., & Bregman, A. S. (1987). Auditory stream segregation and the control of dissonance in polyphonic music. *Contemporary music review, 2*(1), 63-92.

Appendix: Certificates of ethical acceptability

McGill

Research Ethics Board Office
McGill University
845 Sherbrooke Street West
James Administration Bldg., rm 429
Montreal, QC H3A 2T5

Tel: (514) 398-6831
Fax: (514) 398-4853
Ethics website: www.mcgill.ca/rgo/ethics/human

Research Ethics Board II
Certificate of Ethical Acceptability of Research Involving Humans

REB File #: 67-1104

Project Title: Emphasizing voices in polyphonic organ performance: issues of expressive performance on an instrument with fixed tone intensity

Applicant's Name: Bruno Gingras **Department**: Music

Status: Ph.D. student

Granting Agency and Title (if applicable): N/A

This project was reviewed on Nov. 22. 2004 by

Expedited Review ✓
Full Review ____

Eleanor U. Stubley, Nov. 22, 2004
Signature/Date

Eleanor Stubley, Ph.D.
Acting Chair, REB II

Approval Period: Nov. 22. 2004 **to** Nov. 21. 2005

All research involving human subjects requires review on an annual basis. An Annual Report/Request for Renewal form should be submitted at least one month before the above expiry date. If a project has been completed or terminated for any reason before the expiry date, a Final Report form must be submitted. Should any modification or other unanticipated development occur before the next required review, the REB must be informed and any modification can't be initiated until approval is received. This project was reviewed and approved in accordance with the requirements of the McGill University Policy on the Ethical Conduct of Research Involving Human Subjects and with the Tri-Council Policy Statement on the Ethical Conduct of Research Involving Human Subjects.

McGill University

ETHICS REVIEW
RENEWAL REQUEST/FINAL REPORT

Continuing review of human subjects research requires, at a minimum, the submission of an annual status report to the REB. This form must be completed to request renewal of ethics approval. If a renewal is not received before the expiry date, the project is considered no longer approved and no further research activity may be conducted. When a project has been completed, this form can also be used as a Final Report, which is required to properly close a file. To avoid expired approvals and, in the case of funded projects, the freezing of funds, this form should be returned 3-4 weeks before the current approval expires.

REB File #: 67-1104
Project Title: Emphasizing Voices in Polyphonic Organ Performance: Issues of Expressive Performance on an Instrument with Fixed Tone Intensity
Principal Investigator: Bruno Gingras
Department/Phone/Email: Faculty of Music
(514) 398-4535 ext. 00288
bruno.gingras@mail.mcgill.ca
Faculty Supervisor (for student PI): Prof. Peter Schubert

1. Were there any significant changes made to this research project that have any ethical implications? _X__Yes ____No
If yes, describe these changes and append any relevant documents that have been revised.
The only significant change is that I will ask people to listen to recordings of the organ performances and try to determine which voice was emphasized in a specific performance. Cf. appended Consent form for listeners.

2. Are there any ethical concerns that arose during the course of this research? ___ Yes _X__ No. If yes, please describe.

3. Have any subjects experienced any adverse events in connection with this research project? ___ Yes _X__ No
If yes, please describe.

4. _X__ This is a request for renewal of ethics approval.

5. ____ This project is no longer active and ethics approval is no longer required.

6. List all current funding sources for this project and the corresponding project titles **if not exactly the same** as the project title above. Indicate the Principal Investigator of the award if not yourself.

Principal Investigator Signature: Bruno Gingras Date: Nov. 18 '05
Faculty Supervisor Signature: [signature] Date: Nov. 17 '05
(for student PI)

Submit to Lynda McNeil, Research Ethics Officer, James Administration Bldg., rm 419, fax: 398-4644 tel:398-6831

(version October 2002)

For Administrative Use **REB:** ___AGR ___EDU ___REB-I _✓_REB-II

___ The closing report of this terminated project has been reviewed and accepted

_✓_ The continuing review for this project has been reviewed and approved

_✓_ Expedited Review ___ Full Review

Signature of REB Chair or designate: [signature] Date: 11/25/05

Approval Period: Nov 25, 2005 to Nov. 24, 2006

Submit to Lynda McNeil, Research Ethics Officer, James Administration Bldg., rm 419, fax: 398-4644 tel:398-6831

(version October 2002)

McGill University

ETHICS REVIEW
RENEWAL REQUEST/FINAL REPORT

Continuing review of human subjects research requires, at a minimum, the submission of an annual status report to the REB. This form must be completed to request renewal of ethics approval. If a renewal is not received before the expiry date, the project is considered no longer approved and no further research activity may be conducted. When a project has been completed, this form can also be used as a Final Report, which is required to properly close a file. To avoid expired approvals and, in the case of funded projects, the freezing of funds, this form should be returned 3-4 weeks before the current approval expires.

REB File #: 67-1104
Project Title: Emphasizing Voices in Polyphonic Organ Performance: Issues of Expressive Performance on an Instrument with Fixed Tone Intensity
Principal Investigator: Bruno Gingras
Department/Phone/Email: Faculty of Music
(514) 398-4535 ext. 00288
bruno.gingras@mail.mcgill.ca
Faculty Supervisor (for student PI): Prof. Peter Schubert and Stephen McAdams

1. Were there any significant changes made to this research project that have any ethical implications? _X__Yes ____No
If yes, describe these changes and append any relevant documents that have been revised.
The only significant change is that I will ask performers to play on a harpsichord equipped with a MIDI (Musical Instrument Digital Interface) console. The experiments will be very similar to those already undertaken with organists. Cf. appended Sollicitation form for harpsichordists.

2. Are there any ethical concerns that arose during the course of this research? ___ Yes _X__ No. If yes, please describe.

3. Have any subjects experienced any adverse events in connection with this research project? ___ Yes _X__ No
If yes, please describe.

4. _X__ This is a request for renewal of ethics approval.

5. ____ This project is no longer active and ethics approval is no longer required.

6. List all current funding sources for this project and the corresponding project titles **if not exactly the same** as the project title above. Indicate the Principal Investigator of the award if not yourself.

Prof. McAdams, Canada Research Chair in Music Perception and Cognition

Principal Investigator Signature: Bruno Gingras Date: 18/10/2006

Faculty Supervisor Signature: ____________ Date: 18/10/2006
(for student PI)

Submit to Lynda McNeil, Research Ethics Officer, James Administration Bldg., rm 419, fax: 398-4644 tel:398-6831

(version October 2002)

For Administrative Use **REB:** ___AGR ___EDU ___REB-I ✓ REB-II

___ The closing report of this terminated project has been reviewed and accepted

✓ The continuing review for this project has been reviewed and approved

✓ Expedited Review ___ Full Review

Signature of REB Chair or designate: L McNeil Date: Oct. 27, 2006

Approval Period: Nov. 24, 2006 to Nov. 23, 2007

Submit to Lynda McNeil, Research Ethics Officer, James Administration Bldg., rm 419, fax: 398-4644 tel:398-6831

(version October 2002)

McGill University

ETHICS REVIEW
RENEWAL REQUEST/FINAL REPORT

Continuing review of human subject research requires, at a minimum, the submission of an annual status report to the REB. This form must be completed to request renewal of ethics approval. If a renewal is not received before the expiry date, the project is considered no longer approved and no further research activity may be conducted. When a project has been completed, this form can also be used as a Final Report, which is required to properly close a file. To avoid expired approvals and, in the case of funded projects, the freezing of funds, this form should be returned 3-4 weeks before the current approval expires.

REB File #:	67-1104
Project Title:	Emphasizing Voices in Polyphonic Organ Performance: Issues of Expressive Performance on an Instrument with Fixed Tone Intensity
Principal Investigator:	Bruno Gingras
Department/Phone/Email:	Schulich School of Music (514) 849-2935 bruno.gingras@mail.mcgill.ca
Faculty Supervisor (for student PI):	Profs. Peter Schubert and Stephen McAdams

1. Were there any significant changes made to this research project that have any ethical implications? ___Yes __X__No
If yes, describe these changes and append any relevant documents that have been revised.

2. Are there any ethical concerns that arose during the course of this research? ___ Yes _X__ No. If yes, please describe.

3. Have any subjects experienced any adverse events in connection with this research project? ___ Yes _X__ No
If yes, please describe.

4. _X__ This is a request for renewal of ethics approval.

5. ____ This project is no longer active and ethics approval is no longer required.

6. List all current funding sources for this project and the corresponding project titles **if not exactly the same** as the project title above. Indicate the Principal Investigator of the award if not yourself.

Stephen McAdams, Canada Research Chair in Music Perception and Cognition
Stephen McAdams, NSERC Discovery Grant, Perception of source, event and object behaviors in complex environments

Principal Investigator Signature: Bruno Gingras Date: 25/10/2007

Faculty Supervisor Signature: [signature] Date: 25/10/2007
(for student PI)

For Administrative Use **REB:** ____REB-I ✓REB-II ____REB-III

____ The closing report of this terminated project has been reviewed and accepted

✓ The continuing review for this project has been reviewed and approved

✓ Expedited Review ____ Full Review

Signature of REB Chair or designate: L McNeil Date: Nov. 6, 2007

Approval Period: Nov. 24, 2007 to Nov. 23, 2008

Submit to Lynda McNeil, Research Ethics Officer, James Administration Bldg., rm 419, fax: 398-4644 tel:398-6831

(version 06/07)