INSTRUMENTAL DIFFERENCES
IN CHARACTERISTICS OF
EXPRESSIVE MUSICAL PERFORMANCE

DISSERTATION

Presented in Partial Fulfillment of the Requirements for
the Degree Doctor of Philosophy in the Graduate
School of The Ohio State University

By

Timothy M. Walker, M.A.

*****

The Ohio State University
2004

Dissertation Committee:

Professor Mari Jones, Adviser

Professor Neal Johnson

Professor Jay Myung

Approved by

Adviser

Psychology Graduate Program
Performed music contains systematic deviations from the notated values of timing, dynamics, and articulation, which can communicate expressive content to a listener. The choice of musical instrument can be an important factor in shaping a performance, and may be a significant component in the communicative process. However, the differences in expressive variability across instruments have not been adequately studied, largely due to the practical difficulties of obtaining quantitative performance data from non-keyboard instruments. A methodology is proposed in which performances are recorded as digital audio and later transcribed into MIDI data for analysis. This methodology allows an arbitrary number and type of instrument to be studied. Details of the MIDI transcription process are provided, along with some algorithmic tools for comparing a performance file to a notated score.

An empirical study was conducted of performances of two different pieces of music by performers of four different types of instruments (clarinet, guitar, piano, and trumpet). Each performer played each piece twice; once with normal expression and once with no intended expressive variability. Performances were analyzed as functions of instrument and expressive level, with dependent measures of note error frequencies, timing devia-
tions, articulation, and dynamics. Significant effects were obtained for all dependent measures, indicating that performance variability is affected both by the physical instrument and by the performer’s expressive intent.

In addition, a perceptual study was conducted in which listeners were asked to identify the original performance instrument from the MIDI transcriptions of the performances, presented in a neutral timbre. Participants were not able to identify the instruments at a better-than-chance level, but there are indications that some of the instrument-specific features of the performances were in fact perceptually salient. The results indicate that listeners may be sensitive to specific cues contained in expressive variability, but that they may lack the explicit instrumental knowledge necessary to categorize performances accurately.

The studies are discussed together within the context of a process of musical communication between performers and listeners. The importance of instrumental effects and the technical considerations of representational formats for musical data are emphasized. These factors tend to interact with respect to the pragmatics of performing quantitative studies of performance and have numerous implications for future research opportunities.
Dedicated to the memory of my mother, Diane Walker
ACKNOWLEDGMENTS

I would like to thank Caroline Palmer, both for her help with this document and for the years of research experience in the Music Cognition Lab which prepared me for this dissertation.

I am also very grateful to my committee: Mari Jones, who has graciously accepted me as her advisee and provided me with expert guidance; Jay Myung, who has helped me with the analyses and who has always been a source of inspiration; and Neal Johnson, who has been continually supportive throughout my graduate career.

Thank you to Davy Temperley for rewriting the mftext code and posting his changes.

Thank you to all my family and friends who have provided encouragement and help throughout the years of working on and thinking about this research.

Finally, a very special thanks to my wife, Debbie, for her invaluable help on this project and with whom anything seems possible.
VITA

July 8, 1968. . . . . . . . . . . . . . . . . . . . . . . Born - Pittsburgh, PA

1990 . . . . . . . . . . . . . . . . . . . . . . . . . . . . B.S. Physics & Psychology, Carnegie Mellon University

1994 . . . . . . . . . . . . . . . . . . . . . . . . . . . . M.A. Psychology, The Ohio State University

1992 - 1994: . . . . . . . . . . . . . . . . . . . . . . Research Assistant, The Ohio State University

1994 - 1997: . . . . . . . . . . . . . . . . . . . . . . Teaching Assistant, The Ohio State University

1997 - 1998. . . . . . . . . . . . . . . . . . . . . . . Full-Time Instructor, Capital University

1998 - 2001. . . . . . . . . . . . . . . . . . . . . . . Assistant Professor, Capital University

2001 - Present . . . . . . . . . . . . . . . . . . . . . . Assistant Professor, Ohio Dominican University

PUBLICATIONS

Research Publication


FIELDS OF STUDY

Major Field: Psychology
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract</td>
<td>ii</td>
</tr>
<tr>
<td>Dedication</td>
<td>iv</td>
</tr>
<tr>
<td>Acknowledgments</td>
<td>v</td>
</tr>
<tr>
<td>Vita</td>
<td>vi</td>
</tr>
<tr>
<td>List of Tables</td>
<td>x</td>
</tr>
<tr>
<td>List of Figures</td>
<td>xii</td>
</tr>
</tbody>
</table>

## Chapters:

1. Introduction
   - Expressive Musical Performance               | 1
   - Methodology                                  | 3
   - Overview                                     | 4

2. Theoretical Motivation
   - Communication                                | 7
   - Instrument Characteristics                   | 10
   - Semiological Interpretations                  | 15
   - The Motor Theory of Speech Perception         | 18
   - Empirical Studies of Perception and Performance | 20
   - Previous Work                                 | 26
   - Summary                                       | 28
3. Musical Performances .................................................30
   Performance Experiment ...........................................34
   Method .............................................................37
   Participants .......................................................37
   Materials .........................................................37
   Equipment .........................................................38
   Procedure .........................................................39
   Data Transformation .............................................40
   Algorithmic Identification of Errors. .........................43
   Summary ........................................................54

4. Performance Results and Discussion ..........................56
   Error Analyses ....................................................58
   Melody 1 ..........................................................61
   Ständchen .........................................................61
   Timing Analyses ..................................................62
   Melody 1 ..........................................................62
   Ständchen .........................................................66
   Articulation Analyses ...........................................68
   Melody 1 ..........................................................69
   Ständchen .........................................................70
   Dynamics Analyses ..............................................71
   Melody 1 ..........................................................72
   Ständchen .........................................................75
   Discussion .........................................................77
   Summary ........................................................80

5. Perceptual Data ....................................................81
   Perception Experiment ...........................................83
   Method ..........................................................84
   Participants .......................................................84
   Materials .........................................................85
   Procedure .........................................................86
   Results ..........................................................87
   Melody 1 ..........................................................87
   Ständchen .........................................................93
   Discussion .........................................................98
   Summary ........................................................101
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>6. General Discussion</td>
<td>103</td>
</tr>
<tr>
<td>Sources of Variability</td>
<td>105</td>
</tr>
<tr>
<td>Perceptual Data</td>
<td>108</td>
</tr>
<tr>
<td>Methodology</td>
<td>110</td>
</tr>
<tr>
<td>Summary</td>
<td>113</td>
</tr>
<tr>
<td>Appendices:</td>
<td></td>
</tr>
<tr>
<td>A. Notation for Melody 1</td>
<td>115</td>
</tr>
<tr>
<td>B. Notation for Ständchen</td>
<td>116</td>
</tr>
<tr>
<td>C. Audio-to-Score Settings in Logic Audio</td>
<td>117</td>
</tr>
<tr>
<td>List of References</td>
<td>118</td>
</tr>
</tbody>
</table>
LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Levels of constraints in musical performance (e.g. MacKenzie &amp; Van Eerd, 1990)</td>
</tr>
<tr>
<td>4.1</td>
<td>Error rates for Melody 1 and Ständchen by instrument</td>
</tr>
<tr>
<td>4.2</td>
<td>Melody 1 IOIs by instrument</td>
</tr>
<tr>
<td>4.3</td>
<td>Melody 1 durations by instrument</td>
</tr>
<tr>
<td>4.4</td>
<td>Ständchen timing by instrument</td>
</tr>
<tr>
<td>4.5</td>
<td>Melody 1 articulation by instrument</td>
</tr>
<tr>
<td>4.6</td>
<td>Melody 1 velocities by instrument</td>
</tr>
<tr>
<td>5.1</td>
<td>Melody 1 number of responses of each instrument type by performer</td>
</tr>
<tr>
<td>5.2</td>
<td>Melody 1 average IOIs by response category</td>
</tr>
<tr>
<td>5.3</td>
<td>Melody 1 average total errors by response category</td>
</tr>
<tr>
<td>5.4</td>
<td>Melody 1 goodness ratings by instrument</td>
</tr>
<tr>
<td>5.5</td>
<td>Melody 1 goodness ratings by performer and expression type</td>
</tr>
<tr>
<td>5.6</td>
<td>Ständchen responses by performance instrument</td>
</tr>
<tr>
<td>5.7</td>
<td>Ständchen responses by instrument - mechanical performances only</td>
</tr>
<tr>
<td>5.8</td>
<td>Ständchen number of responses of each instrument type by performer</td>
</tr>
<tr>
<td>5.9</td>
<td>Ständchen average IOIs by response category</td>
</tr>
<tr>
<td>5.10</td>
<td>Ständchen average velocities by response category</td>
</tr>
</tbody>
</table>
5.11 Ständchen goodness ratings by instrument. ..............................97
5.12 Ständchen goodness ratings by performer and expression type. ..........99
**LIST OF FIGURES**

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1</td>
<td>Sample output from PerfComp</td>
<td>52</td>
</tr>
<tr>
<td>3.2</td>
<td>Sample verbose log from PerfComp</td>
<td>53</td>
</tr>
<tr>
<td>4.1</td>
<td>Melody 1 timing by instrument, expression type, and location</td>
<td>63</td>
</tr>
<tr>
<td>4.2</td>
<td>Melody 1 timing by instrument and location</td>
<td>64</td>
</tr>
<tr>
<td>4.3</td>
<td>Melody 1 timing by expression type and location</td>
<td>65</td>
</tr>
<tr>
<td>4.4</td>
<td>Ständchen timing by instrument and location</td>
<td>67</td>
</tr>
<tr>
<td>4.5</td>
<td>Melody 1 articulation by instrument and location</td>
<td>69</td>
</tr>
<tr>
<td>4.6</td>
<td>Ständchen articulation by instrument and location</td>
<td>71</td>
</tr>
<tr>
<td>4.7</td>
<td>Melody 1 velocities by instrument, expression type, and location</td>
<td>73</td>
</tr>
<tr>
<td>4.8</td>
<td>Melody 1 velocities by instrument and location</td>
<td>74</td>
</tr>
<tr>
<td>4.9</td>
<td>Melody 1 velocities by expression type and location</td>
<td>74</td>
</tr>
<tr>
<td>4.10</td>
<td>Ständchen velocities by instrument and location</td>
<td>76</td>
</tr>
<tr>
<td>4.11</td>
<td>Ständchen velocities by expression type and location</td>
<td>77</td>
</tr>
</tbody>
</table>
CHAPTER 1

INTRODUCTION

Expressive Musical Performance

This study is concerned primarily with the process of and influences on expressive musical performance; the nuances that lend unique character beyond the information that is present in the musical score alone. The idea that music has expressive characteristics has a long history; composers and theorists have maintained various theories of the nature of musical expression, a tradition that has been continued in the psychological models of the 20th century (e.g. Gjerdingen, 1988; Lerdahl & Jackendoff, 1983; Meyer, 1956; Narmour, 1990; Repp, 1993; Seashore, 1938).

Music contains abstract properties, which expressive performance may serve to accentuate. Musical meaning has often been attributed to compositional attributes such as harmonic progression, melodic movement, and voice leading (Gjerdingen, 1988; Meyer, 1956; Narmour, 1990; Sloboda, 1985), and there may be important correlates to expressive performance to be found in a compositional score. Additionally, expressive performance is often thought to reflect unscored musical qualities, including emotion, kinematics, and aesthetic judgement, as well as more structural features (Balkwill &
Thompson, 1999; Clynes, 1977; Feldman, Epstein, & Richards, 1992; Gabrielsson, 1988; Gabrielsson & Juslin, 1996; Nakamura, 1987; Palmer, 1996b; Repp, 1993; Raffman, 1993; Sundberg & Verrillo, 1980; Todd, 1992, 1995). Performance expression is defined here as systematic variation of event timing, dynamics, and articulation, and is often a manifestation of a performer’s interpretation of the musical structure present in a composed score (Clarke, 1988; Drake & Palmer, 1993; Palmer, 1989a; Sloboda, 1983). Elements of an expressive performance may contain correlates to musical structures such as phrasing (Palmer, 1989a, 1996a; Repp, 1992; Todd, 1992) and meter (Clarke, 1993; Palmer & Kelly, 1992; Shaffer, 1984; Shaffer, Clarke, & Todd, 1985; Sloboda, 1983).

The complicated set of possible causal factors conspire to make performed music a very difficult subject for empirical study. Interestingly, much of the seminal theory of music cognition is based in music theory, which is not particularly descriptive of performance characteristics. For example, many important theories of music perception (e.g. Lerdahl & Jackendoff, 1983; Meyer, 1956; Narmour, 1990) are essentially “score-based” theories, which describe how listeners infer structural and/or emotional properties from the music. These theories appear to assume that the listener extracts the score (i.e. the intended notes) in the face of performance variability. In other words, performance-specific characteristics are largely dismissed as surface features, while the structural properties of the score itself contain the “deep structure.” There is undoubtedly a great deal of value in these kinds of analyses, but it seems as if something is missed as well.
Methodology

The goal of this dissertation is to treat expressive performances as musically-important events on their own. In other words, deviations in performance variables from their notated values will be treated not just as indications to listeners of interesting elements in the score, but as relevant musical attributes emanating from the performer and his or her instrument. An obvious requirement is that the performance variables must first be recorded and represented, and the pragmatics involved in this process turn out to have rather profound effects on the types of studies that are typically conducted in music performance research.

The most common format for recording and manipulating musical events is the MIDI (Musical Instrument Digital Interface) format, which provides, among other things, information about the pitch values, beginnings, ends, and intensities of musical notes. As a real-time data collection mechanism, however, MIDI is not nearly as accurate or convenient for non-keyboard instruments as it is for keyboards. For models of performance and perception that are largely score-based, the limitations of MIDI are not particularly problematic; the focus, after all, is not really on the musical instrument. However, an empirical study of the properties of instrumental performance requires a satisfactory and comparable method of obtaining data from a variety of musical instruments.

As a result, a large component of this project involves formulating a methodology for acquiring and properly representing musical performance data from non-keyboard instruments. The general approach is to record performances as digital audio, and to then later transcribe the performances as MIDI data. In this way, the advantages of MIDI as a
representational format are retained, without the associated disadvantages of real-time MIDI recording.

The process of expressive performance is viewed within a context of communication from a performer to a listener (e.g. Kendall & Carterette, 1990). Therefore, once the performance data has been appropriately translated it is presented to a group of listeners in a perception task. The listeners should be able to respond differentially to the expressive attributes that are relevant to particular instruments. The communicative model for this context differs from traditional models of language communication in that it doesn’t assume a veridical transfer of semantic content from the source to the receiver. Rather, it supposes that performers and listeners each operate, independently, on the performance itself (e.g. Nattiez, 1990).

**Overview**

Chapter 2 of this dissertation outlines the motivating theory and empirical background for considering expressive performances on a variety of instruments, including communicative theories of music, the issues associated with MIDI performance data, and some relevant methodologies. Chapter 3 discusses in detail the methodology used in this study for obtaining and quantifying musical performances, including the challenges of and strategies for representing performance data for analysis. Chapter 4 presents the statistical analyses and results from the performance data and discusses the empirical findings. Chapter 5 describes the perceptual experiment based on the performance data, including
the results and discussion. Chapter 6 discusses the findings of the study as a whole and summarizes its conclusions and implications.
CHAPTER 2
THEORETICAL MOTIVATION

This chapter provides an overview of the theoretical ideas and empirical findings which were elemental in developing the ideas of the dissertation. It starts by describing some theories of music performance and perception as they relate to a process of musical communication. This is continued with a discussion of the influences that a specific instrument can have on communication of expressive performance. An instrument-specific model of musical communication is then suggested, using a bidirectional approach borrowed from semiological theory which presents the acoustic output from a performance as an artifact for perceptual processing. Some possible implications of this theoretical approach are discussed, including some theories of language processing which link performance and perception in an analogous fashion. A brief review of empirical literature of music performance and perception is provided, with discussion of the implications that methodology has on the ability to study instrument-specific differences. Central to this discussion is the pervasive use of MIDI as a recording format in empirical studies of music performance. Finally, an earlier study of instrument-specific performance is dis-
cussed in some detail, including a critique of its methodology and a rationale for the methodology in the current study.

Communication

If one goal of expressive music performance is to instantiate properties of music, then expressive features should be perceptually salient to listeners. For example, tonal relationships can be emphasized or (in the case of improvisation) modified by the performer (Järvinen, 1995; Pressing, 1987). Expressive timing in performance is an important vehicle by which tones are selectively emphasized. For example, music oftentimes contains multiple melodic lines including one called the melodic voice, which is structurally more important than the others. When notes from different melodic lines occur at the same point in the score, the melodic voice is often performed with a temporal lead, both in orchestral (Rasch, 1979) and solo performances (Palmer, 1989a). Thus, one function of musical expression may be to perceptually highlight certain musical events, in order to focus the listener’s attention (Drake & Palmer, 1993; Palmer, 1996a; Repp, 1992; Widmer, 2002). In other words, the process of musical expression does not exist in a vacuum; it is expressed to a listener.

One way in which the process of conveying musical properties to an audience can be characterized is as a communication of musical knowledge from composer to listener (Gabrielsson, 1988; Gabrielsson & Juslin, 1996; Kendall & Carterette, 1990; Nakamura, 1987; Repp, 1993; Senju & Ohgushi, 1987). Communication, both in language and in artistic domains, requires that a speaker or performer (the communicator) map some
abstract intention (the idea to be communicated) onto a perceptible sequence of actions, which may then be interpreted by a listener or observer. Furthermore, the communicator must successfully perform the appropriate actions. In other words, the message must be composed, performed and perceived in order for communication to succeed.

Music, as well as speech, can be thought of as a process of communication. A composer of a musical score has certain artistic intentions, which he or she attempts to fulfill through the careful use of melody, rhythm, harmonization, etc. The composition must then be performed, in order for the composer’s message to reach the ears of a listener. The performer(s) must interpret and present the compositional score, through phrasing, pacing, articulation, etc., in such a way that its meaning is preserved. A piece of music may not possess the same kind of semantic meaning as does speech but, to the extent to which musical knowledge is shared between composer, performer, and listener, an analogous notion of communication exists. Interestingly, although many current theories appeal to abstract musical structures as ways of explaining expressive variability, they do not completely describe the process by which the variability is realized; there is not a truly predictive model for musical expression.

Of primary interest herein are the sources and media of musical communication: who communicates with whom, and what mechanisms are used? Composers, performers, and listeners all play important roles in the communicative process. Many influential music theories (e.g. Lerdahl & Jackendoff, 1983; Meyer, 1956; Narmour, 1990), however, have focused upon interpretation of a perceived score by listeners, with little attention paid to performers and their instruments. Lerdahl and Jackendoff (1983), for example, argue
that musical structures can be more or less “well-formed” and/or “preferred.” These structures are based almost entirely on musical abstractions (e.g. metrical position, rhythmic cohesion, melodic and harmonic function) that do not address the performance nuances which are experienced by a listener in real-time appreciation of the music. Contributions of timbre, rubato, dynamics, articulation, or other largely performance-specific aspects are considered part of the surface structure. As such, they can have some effect upon the formation of deep structure; Lerdahl and Jackendoff propose relatively well-defined rules whereby superficial acoustic cues help reinforce metrical, rhythmic, and other structural constructs. However, performance characteristics are not themselves intrinsic to deep structure in this theory. This score-based approach of music theory appears to be incomplete, given the marked effects that performance characteristics have on music perception; performed music as a linear, acoustic event contains more information than is in the score alone (Meyer, 1996; Palmer, 1996a; Sloboda, 1983).

A score can be thought of as a “quantized” representation of music, in that each note has a clearly-defined length (e.g. eighth-notes, quarter-notes) and categorical or relational instructions for articulation, dynamics, and tempo. A performance, however, contains continually varying dimensions of relative onsets, articulation, intensities, tempo, and timbral properties. Score-based approaches of music theory, by definition, assume that listeners extract a quantized score from a real performance.

The process of building an internal, quantized, representation from a variable acoustic input seems to be evocative of language perception, in which discrete internal linguistic categories are thought to arise from a process of distinguishing between seemingly ambig-
uous acoustic objects (Burton, Baum, & Blumstein, 1989; Cutting & Rosner, 1974; Pisoni & Tash, 1974). There may in fact be many similarities in the acoustic perceptual processes for music and speech. However, units of language have semantic meanings that exist in abstraction, in a way that individual tones in music simply do not. Thus, in language perception a listener's task can be characterized as a search for semantic meaning, guided by syntactic knowledge and cues of inflection. A distinction is therefore often drawn in linguistics between the syntax and the semantics of an utterance (e.g. Bock, 1986; Boland, 1993; Chomsky, 1965, 1971, 1981; Chomsky & Halle, 1968; Ferreira & Henderson, 1990; Masson, 1995).

However, music is not speech; it is an art form, and expressive variability in a performed piece of music can bring a hierarchically uninteresting score to life. The distinction between syntax and semantics for language does not appear to apply to music, or at least not in the same fashion, so it may be misguided to relegate attributes of expressive musical performance to a mere surface level. The variability itself in a performed piece of music may be elemental to whatever communicative or expressive properties the music might possess. Therefore, it may be inappropriate to build exact parallels with speech into theories of musical perception.

**Instrument Characteristics**

Kendall and Carterette (1990) have proposed a somewhat more performance-oriented theory of musical communication. They define a communicative chain of the form

Composer -> performer -> listener,
and show that levels of musical expression can be communicated to listeners through the performer’s use of amplitude and, particularly, temporal cues.

In Kendall and Carterette’s study, a professional concert pianist performed a musical passage with three different levels of expression: mechanical (no expression); musical (normal expression); and exaggerated expression. These performances constituted a model of expressive performance to be used for comparison. Four different instrumentalists (violin, trumpet, clarinet, and oboe) then listened to the model performances and created their own set of recordings of the same musical passage with the same three levels of expression. In a series of experiments, listeners were asked to match the performances of each of the four instruments with the model (piano) performances on the basis of level of expression, and also to rate the overall perceived amount of expressive content within each performance. In addition to the acoustic recordings, synthesized versions of the recordings were created which maintained the original timing profiles, and were presented to listeners subject to the same tasks. Both musicians and non-musicians participated. A good degree of correspondence was obtained between listeners’ ratings and the intended levels of performed expression, and listeners were generally able to match each instrument with the proper model of expression, with better results for musicians and for natural (i.e. non-synthesized) performances.

Kendall and Carterette claimed that information is transferred from the composer and performer to the listener by means of a shared musical knowledge. Performance characteristics, however, may play a larger role in the communicative process than the authors assumed. Although they emphasized the fact that the listeners were able to identify
expressive levels reasonably well for each instrument, there were in fact some differences in perceptual judgments across instruments. For example, violin performances were more likely than other instruments to have the expressive level misidentified, and trumpet performances displayed a different pattern of event timing than the other instruments.

These results (and results from similar studies, such as Gabrielsson & Juslin, 1996 and Nakamura, 1987) suggest that the unique qualities of each performance, particularly the instrument on which the piece is performed, may systematically alter the content of communication. However, only one performer was employed per instrument in Kendall and Carterette’s study, which precludes a discrimination between instrument differences and performer differences. Furthermore, all of the performers were instructed to imitate a piano performance, which Kendall and Carterette used as a model of appropriate levels of expression. Thus, the judgements of musical expression may have been confounded by performer-specific (and possibly piano-specific) qualities of the expressive performances.

An interesting interpretation of Kendall and Carterette’s study is that a performer’s abstract intentions to produce an expressive performance may be mediated in production by the particular instrument that is used. A general level of expression may be communicable across instruments, as the authors assert; however, aspects which differ as a function of the instrument are also important to an understanding of musical communication. Differences in expressive performance characteristics (e.g. patterns of timing, dynamics, etc.) may arise due to the physical configuration of an instrument, the acoustic properties of the instrument, the stylistic background of the performer of the instrument, and a host of other
potential factors (Baily, 1977, 1985). These differences would presumably affect the nature of musical communication on each instrument.

Instead of viewing the performer and his or her instrument as merely conduits of musical expression, then, it may be more accurate to consider each element in the chain of communication as transformational. In this light, a modification to the chain of musical communication is offered, of the form

Composer -> performer(instrument) -> listener.

Two major theoretical changes are suggested. First, the performance is now considered as a function of the particular instrument used, rather than an instantiation of abstract, “pure” compositional intentions. Second, a greater emphasis is placed upon the links in the chain; the systematic transformations of variables of expression introduced by the performer and his or her instrument are an integral part of musical communication. This model provides a greater attention to the particular influences of a musical instrument, and allows for a communicative chain to be described in greater detail, as a function of the specific instrument in a performance.

The characteristics of a particular musical instrument are viewed as a significant factor in the interpretive process within this conception of musical communication; the intentions of a composer are filtered through different constraints of musical instruments, both from human performance (Rosenbaum, 1991) and from the physical configuration of the instrument (Askenfelt & Jansson, 1992; Baily, 1977, 1985; Nettheim, 1979; Parn curt, Sloboda, Clarke, Raekallio, & Desain, 1997; Sayegh, 1991). Although physical constraints (e.g. the spacing and action of keys or strings on a musical instrument) may affect the per-
formance, the effects should not necessarily be regarded as undesirable limitations or boundaries. The “shaping” of a musical piece by the instruments upon which it is played may be a significant source of its aesthetic quality; musical nuances in the surface, although not existing within the notational nomenclature, are nonetheless often a defining feature of the aesthetics of music (Raffman, 1993). Because part of being a skilled music performer entails being familiar with one’s instrument, the performer’s conception of artistic playing may integrate compensations for physical constraints, such that an acceptable performance cannot in fact be achieved without a consideration of the physical medium (Askenfelt & Jansson, 1992; Baily, 1977, 1985; Coffman, 1990; Dahl, 2000; Ducasse, 2003; Kay, Turvey, & Meijer, 2003; MacKenzie & Van Eerd, 1990; Meyer & Palmer, 2003; Nettelheim, 1979; Parncutt, Sloboda, Clarke, Raekallio, & Desain, 1997; Sudnow, 2002; Sundberg, 2000).

MacKenzie and Van Eerd (1990) identify constraints believed to be associated with different levels of music performance: “high” (e.g. musical structure and interpretation); “low” (e.g. biomechanical and anatomical); and “physical” (e.g. instrument acoustics and mechanics) levels. The authors found that the rhythmic precision with which pianists could perform simple sequences (musical scales) was inversely related to the speed of playing and fingerering, which they attribute to physical and physiological constraints of performing. Although MacKenzie and Van Eerd’s study addressed piano performance only, the division of levels is a useful way in which to think about different constraints across instruments; different instruments necessarily entail different physical constraints,
which may affect higher-level interpretive aspects of performance. A modification of MacKenzie and VanEerd’s levels, extended to multiple instruments, is shown in Table 2.1.

<table>
<thead>
<tr>
<th>Level</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>High (Cognitive)</td>
<td>Musical Structure</td>
</tr>
<tr>
<td></td>
<td>Interpretation</td>
</tr>
<tr>
<td></td>
<td>Expression/Mood</td>
</tr>
<tr>
<td>Low (Motor Control)</td>
<td>Sensorimotor Loop Times</td>
</tr>
<tr>
<td></td>
<td>Biomechanical</td>
</tr>
<tr>
<td></td>
<td>Anatomy</td>
</tr>
<tr>
<td>Physical (External)</td>
<td>Piano Action</td>
</tr>
<tr>
<td></td>
<td>Key Layout/Spacing</td>
</tr>
<tr>
<td></td>
<td>Guitar Action</td>
</tr>
<tr>
<td></td>
<td>Fretboard Size</td>
</tr>
</tbody>
</table>

Table 2.1: Levels of constraints in musical performance (e.g. MacKenzie & Van Eerd, 1990)

**Semiological Interpretations**

Although the performer may have a very specific interpretation of a piece of music, a listener is not directly privy to this interpretation; the listener has direct access only to the acoustics of the performance, not the intentionality which produced the acoustics. Therefore, listeners must extract the intentions of the composer and the performer from systematic variation in the signal (Nattiez, 1990). As such, it may be misleading to refer to direct communication from composer to performer to listener. A semiological view is suggested
here instead: in order for two members of the chain to communicate, one member (e.g. the composer) produces an identifiable, physical output, called a trace (e.g. the score). The other member (e.g. the performer) may receive this trace, and can analyze it and attempt to extract some meaning from it. However, the crucial point is that any meaning that is extracted belongs inherently to the receiver; it is not a direct transmission of the producer’s message (Eco, 1976; Peirce, 1935). This state of affairs can be represented in the most general sense by the following form:

Producer -> \textit{Trace} <- Receiver,

where, for example, the producer may be a composer, the trace may be a score, and the receiver may be a performer.

The directionalities of the links in the above model illustrate the limits on accessibility of information. The producer has certain intentions that he or she wishes to communicate. The instantiation of these intentions is the physical trace which is produced (e.g. the utterance, score, or performance). The receiver does not have direct access to the producer’s intentions, but does have access to the trace; thus the receiver must operate upon the trace in order to infer meaning. The trace symbolizes meaning for both the producer and the receiver, but the two meanings are the results of two different processes (a creative process by the producer and an interpretive process by the receiver), and are therefore not necessarily identical. In this way a form of a communicative chain is established, but in a non-direct fashion: the abstract ideas to be communicated are always mediated by a physical trace.
In the musical domain, at least two of these semiological “chains” can be established. The first is from composer to performer, and the second is from performer to listener:

Composer -> Score <- Performer(instrument) -> Performance <- Listener

The backwards arrows are intended to illustrate the fact that performers must attend to the written score and listeners must attend to the acoustic performance in order to extract meaningful information. The processes of creating a score by a composer and of interpreting a score by a performer are necessarily very different operations, and are explicitly identified by the semiological model. Nattiez (1990) refers to these operations as, respectively, poietic (i.e. creative) and esthesic (i.e. interpretive) processes. The musical representation above does not capture every possible relationship; indeed, there are a number of possible sub-component semiological chains (Peirce, 1935). For example, both composers and listeners may have instrument-specific knowledge that influences the way they compose and perceive music, respectively. Additionally, tasks such as improvisation, in which a single person fulfills multiple roles of composer, performer, and listener, offer interesting interpretations of the semiological model.

The semiological approach is important to psychological and musical theories because it brings to the forefront issues that aren’t as obvious in standard unidirectional models of communication. This model, for example, is not as subject to the composer-to-listener score-analysis bias that is present in many other theories. More emphasis is on the performer, who is involved in both of the links established above and is responsible for both interpretive and creative processes.
The Motor Theory of Speech Perception

To be sure, a semiological model such as discussed above does not at all preclude the capability of communication from (for example) composer to audience. It is merely the case that the communicative route contains a series of interpretive steps. Furthermore, the character of these interpretations should be familiar to the listener, as they are natural and presumably unavoidable consequences of physically (i.e. acoustically) realizing the composition. As an analogous process, consider the motor theory of speech perception (Liberman, Cooper, Harris, & MacNeilage, 1963; Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967), which similarly considers communication as mediated by physical articulators. According to this theory, the perception of speech entails a sympathetic production process by the perceiver; speech is understood by way of understanding the processes which must have created it.

In its original conception, the motor theory was applied only to speech. In a more general sense, however, the motor theory states that perception of an acoustic event is a function of the properties of the devices which produced it. This is essentially the claim of an instrument-specific theory of musical communication. Indeed, some perception research has shown a tendency for listeners to attend to physical-motion correlates in music (Feldman, Epstein, & Richards, 1992; Friberg, Sundberg, & Fryden, 2000; Repp, 1993; Sundberg, 2000; Todd, O’Boyle, & Lee, 1999), and instrument-specific differences have been suggested for notational audition in musicians (Brodsky, Henik, Rubinstein, & Zorman, 2003). Furthermore, Haueisen & Knösche (2001) presented recordings of piano
performances to musician listeners and found evidence of involuntary finger movement
motor activity in pianists but not in the other musicians.

The motor theory claims that perceivers are able to interpret speech by virtue of
being speakers themselves; they attend to the perceptual correlates in speech production. In terms of the semiological model presented earlier, the motor theory states that the
esthesic (interpretive) processes of a listener result in a veridical interpretation of the
acoustic signal precisely because the listener also possesses the same kinds of poietic (cre-
ative) processes as the speaker. In other words, spoken messages can be effectively com-
municated because the poietic and esthesic processes are both understood; they share a set
of governing rules.

One implication for musical communication is that listeners may be better able to
interpret musical structure if they possess knowledge about the poietic processes by which
music is created (i.e. if the listeners possess music performance experience). While it is
certainly true that non-musicians are able to perceive aspects of musical expression, there
may be a perceptual advantage by listeners who are also music performers, perhaps both
in a general sense and as a result of (instrument-specific) knowledge about performance
characteristics of a particular instrument. For example, Kendall and Carterette (1990)
found greater consistency in ratings of performed levels of musical expression by musi-
cians than by non-musicians, and Palmer (1996b) found that pianists’ identifications of
melody were influenced more by chord asynchronies in piano performances than nonpia-
nist listeners’ identifications.
Empirical Studies of Perception and Performance

Semiology and the motor theory of speech perception provide useful theoretical groundwork for performance-based analysis of musical expression. In particular, the concept of operating on a physical trace is a central idea to this dissertation. Rather than conjecturing as to the nature of underlying semantic intentions, the approach herein treats expressive deviation as an empirical artifact worthy of study in its own right. The goal is to quantify musical expression and to identify characteristic features of particular instrumental performances. Because the relevant trace is an instrumental musical performance, the two entities who operate on the trace are the performer and the listener.

Empirical studies of music perception need to include the relevant deviations of timing, accent, timbre, articulation, etc. if they wish to address the characteristics of performed music. For example, empirical studies have asked listeners to report the perceived emotional content of performances (Balkwill & Thompson, 1999; Schellenberg, Krysciak, & Campbell, 2000; Senju & Ohgushi, 1987; Sloboda & Lehmann, 2001); the perceived appropriateness of expression as a function of context (Timmers, 2003); the subjective tempo of an expressive performance (Repp, 1994); the perception of dynamics in expressive performance (Nakamura, 1987); the subjective quality of the final ritard (Sundberg & Verillo, 1980), and the perception of note lengthening (Repp, 1992). These are perceptual studies which use real performances as stimuli, as opposed to stimuli generated from the equivalent of a musical score, in order to study performance characteristics.

As with any experimental study, researchers of performed expression must make some choices as to the features they will explicitly control and measure. They can be very
precise and manipulate a single variable at a time, such as note length (e.g. Repp, 1992) and therefore be fairly confident in the particular attributes to which the participant is responding, and in what way they are responding. However, precision is a double-edged sword, and over-controlling musical expression runs the risk of missing the bigger picture. Indeed, the appeal of musical expression to many listeners is its ephemeral quality, and it is not clear that an expressive performance is reducible to the sum of its quantifiable parts. As Kendall and Carterette (1990) put it:

One of the problems in studying musical expression is that its chief domain lies in implicit procedures. Evidence for this can be seen in the difficulty of verbalizing about the generation and perception of expressiveness. It is no accident that most performance instruction is conducted one-on-one in an interactive, modeling environment. Despite its largely covert nature, something is being imparted by the performer, as is evident from the large number of multiple recordings and performances of the same musical notation. In our study, expressiveness is the intended message generated by the performer and directed at the listener. When and if the intended message is received, communication has occurred. The performer’s message is a synchronous modulation of the composed states that serve as the carrier. In the process of message parsing and recoding, the listener may
impose meanings unintended by either the composer or performer. Indeed, it is in the nature of the listener always to seek meaning. Even if the composer’s message is unparsable, the performer-generated message may be grasped by the listener. In this case, the composer’s work serves as little more than a framework for the dialogue between performer and listener. (p. 135).

There are a few interesting features of Kendall and Carterette’s study. First, they did not attempt to control any particular dimension of performances; they simply asked their performers to play with varying amounts of expression, and then used those performances as the experimental stimuli. Second, they characterize their study as predominantly perceptual in nature, despite the fact that they employ performers specifically for the study (and in fact do some quantitative analysis of the performances). This is significant because it illustrates the guiding principles of their underlying theory, namely that perception is intimately concerned with the performance itself and therefore also with all of the concomitant sources of the performance. The communicative relationship between performer and perceiver, therefore, is a useful construct for studying musical expression in general.

This suggests a slightly different interpretation of the distinction between studying perception and performance of musical expression. As Kendall and Carterette (1990) rightly note, expression is difficult to quantify and therefore correspondingly difficult to control and study experimentally. Philosophically, musical expression really only exists at
all within the context of it being interpreted as an art form by a human listener. From this perspective, it makes sense to think of the listener’s judgements of expression as valid measurements of performance characteristics. Therefore, a perceptual study of performed music can also simultaneously be considered as a study of music performance; the listener him/herself is the data collection device.

However, one would certainly not want to use perceptual reports as the sole mechanism for evaluating performance, for a variety of reasons. First, as a matter of efficiency and statistical power, there typically is not very much quantitative data in listeners’ ratings. Compared with the obvious wealth of information that must be contained within a single musical performance, a single checkmark along a 7-point goodness scale can seem a bit underwhelming, valuable though it might be. Second, there is the problem of level of abstraction. Listeners’ ratings are the outputs of a host of interpretations, perceptual systems, and biases, and it may not be entirely clear how they arrive at their responses. In other words, a listener’s ratings are based on only one side (the perceiver’s) of the physical trace. Third, and most important, the performance itself ought to be amenable to direct study as an acoustic artifact (i.e. from the performer’s side of the physical trace). The acoustic correlates of musical expression may not be obvious a priori, and may not be independent of a long list of other variables (such as musical style, instrument, performer’s training, etc.), but they nonetheless should be valid objects of analysis.

Unfortunately, the complexity of the acoustic array of a musical performance creates a daunting task for an empirical researcher, and is a very real practical barrier to comprehensive quantitative analysis. The use of MIDI provides a common format for musical
performance at a reasonable level of precision, and is widely used in performance studies. However, the characteristics of MIDI can be confining, in that it is amenable to a fairly narrowly-defined methodology. For example, MIDI captures acoustic musical qualities such as timbre, vibrato, slurs, and crescendo in, at best, a rudimentary fashion.

As in any endeavor, the tools which are readily available tend to define the way the work gets done, and music research is no different. There is no doubt that the ability to use MIDI data in a performance analysis greatly simplifies the process of data acquisition and formatting, often to the point of being a necessity in practical terms of being able to get the study done at all. Therefore, data which is not easily represented in MIDI tends not to be analyzed quantitatively. Probably the most important implication of this is on the use of instruments in studies of music performance. MIDI was developed for keyboard instruments and, although the format has been extended, retains many of its idiosyncratic keyboard properties. It is far easier to capture MIDI data from keyboard performances than from non-keyboard instruments since the keyboard can be fitted with sensors on each key to directly measure finger-presses, rather than resorting to a decomposition of the acoustic signal. For this reason, nearly every quantitative analysis of musical performance is done on the piano or electric keyboard (e.g. Bresin & Battel, 2000; Busse, 2002; Clarke, 1993; Desain & Honing, 1994; Drake, Dowling, & Palmer, 1991; Drake & Palmer, 1991, 1993; MacKenzie & Van Eerd, 1990; MacKenzie, Van Eerd, Graham, Huron, & Wills, 1986; Madison, 2000; Palmer, 1989a,1989b, 1996a, 1996b; Palmer & van de Sande, 1993, 1995; Repp, 1992, 1997, 2000; Repp, Windsor, & Desain, 2002; Shaffer, 1984; Shaffer, Clarke, & Todd, 1985; Sloboda, 1983; Sloboda & Lehmann, 2001; Tim-
mers, Ashley, Desain, & Heijink, 2000; Timmers, Ashley, Desain, Honing, & Windsor, 2002; Todd, 1992).

The implications of the piano-centric nature of the performance literature could be quite significant. The unique physical characteristics of an instrument appear to affect the nature of performance (Askenfelt & Jansson, 1992; Baily, 1977, 1985; Iyer, 2002; Kay, Turvey, & Meijer, 2003; MacKenzie & Van Eerd, 1990; Nettheim, 1979; Parnutt, Sloboda, Clarke, Raekallio, & Desain, 1997; Sayegh, 1991), yet nearly all that is known about patterns of timing, errors, and dynamics is derived from performances on the piano. It seems inappropriate to attempt to extrapolate these results to music performance in general. Even Kendall and Carterette (1990), one of the few studies to use multiple instruments in their design, use piano performances as the model for all of their expressive performances, probably to simplify determinations of note onsets. Importantly, Kendall and Carterette (1990) also find differences in perceived expression for performances on different instruments. This seems to emphasize the necessity of conducting truly quantitative studies of musical performances on instruments other than the piano.

Studies do exist which gather note information other than via MIDI pianos and keyboards, but they tend to be resource intensive. For example, Ellis (1991) used specialized wind controllers to gather timing data from saxophonists. Halsband, Binkofski & Camp (1994) used infrared cameras to gather movement information from LEDs attached to the wrists and fingers of pianists, and Dahl (2000) recorded movements of LEDs attached to the drumsticks of percussionists. Ashley (2002), Collier and Collier (2002), and Friberg and Sundstrom (2002) inspected digital waveforms and spectograms to determine the tem-
poral onsets of commercial Jazz recordings. Nakamura (1987) used a high-pass filter to capture the relative intensities of recordings, and Gabrielsson and Juslin (1996) interpreted by eye the timing, dynamics, and timbre of digital recordings. Technology to assist the process is improving, but these types of studies still entail very intensive stages of data collection and formatting, and so they remain rare in the literature.

**Previous Work**

It seems that a complete theory of musical communication might include influences of performers’ interpretations of the composition’s musical structure, physical constraints of the particular instrument used, and acoustic properties which are perceived by listeners. Empirical studies of these factors are sparse. Only in the past couple of decades has it become feasible to study music performance in a detailed quantitative fashion, particularly with the advent of MIDI, but MIDI’s piano-specific bias has weighted the available literature. Given the paucity of cross-instrument performance data, it becomes difficult to extrapolate beyond piano performances to music performance in general; crucial information pertaining to instrument-specific characteristics has not yet been obtained.

In order to test the idea that the physical constraints of each particular instrument play a significant role in musical communication, Walker (1996) conducted experiments in which the instrument-specific physical constraints and the acoustic timbre of the music were independently varied. In this way, the acoustic and physical properties of instruments could be separated. Performers played musical passages on each of two different instruments: a piano keyboard and a guitar (the study was composed of three groups of
musicians: pianists; guitarists; and musicians who played both instruments). Each instrument functioned as a MIDI output device, allowing each musical performance to be realized acoustically as either a piano timbre or a guitar timbre independently of the physical device which actually produced the MIDI output. There were few effects of timbre, except that pitch bending (a property unique to guitar) was employed to a greater extent in performances using the guitar timbre. The most important factor seemed to be the performance instrument. Performers on different instruments displayed different patterns of key choice in improvisation; pianists tended to choose a tonic of C, which has a major scale composed of all white keys (thus minimizing physical demands), and guitarists tended to choose A-major, G-major, and A-minor, which each contain many open (unfretted) strings in standard guitar tuning (minimizing physical demands). Furthermore, the relative frequency of pitch-class usage differed as a function of performance instrument, such that pianists tended to conform more to diatonicity than did guitarists. As a whole, it appeared as if the physical constraints of the performer’s instrument influenced aspects of the performance.

A perceptual experiment was also conducted, which tested listeners’ sensitivity to characteristics of the physical instrument by having them identify the original instrument on which the performance was recorded (regardless of the timbre with which it was sounded). Half of the MIDI recordings were played in a timbre consistent with the physical instrument (e.g. piano performance played with a piano timbre), and half with an inconsistent timbre (e.g. piano performance played with a guitar timbre). Listeners performed significantly better than chance at identifying the physical instrument that pro-
duced the performances. Listeners also rated how good each performance was, and gave higher ratings for performances in which the timbre and the physical instrument were consistent (listeners also tended to prefer piano performances in general). The findings from these two experiments suggest that instrument-specific characteristics are significant factors in music performance; performers display different attributes across instruments, and listeners appear to be sensitive to the differences.

There are, however, some limitations to the findings of Walker (1996). First, a range of two instruments is not sufficient to truly generalize to instrument-specific performance differences. Second, the naturalness of guitar performances was limited by the MIDI tracking equipment, perhaps interfering with performers’ ability to produce expressive performances. The current study involves a more comprehensive analysis of instrument-specific characteristics in expressive performance by extending the set of instruments to four (clarinet, guitar, piano, and trumpet) and by recording the performances as digital audio rather than as MIDI. The acquisition, representation, and analyses of the performance data is described in Chapters 3 and 4. Additionally, a perception task involving the identification of the performance instrument, similar to the perception experiment of Walker (1996), is described in Chapter 5.

Summary

Current theories of expressive music performance and perception provide a structure for describing a communicative process from composer to listener. However, many of these theories seem to be incomplete, by failing to emphasize the important interpretive
transformations that the performer and his or her instrument might contribute to this process. Therefore, expressive performance is proposed to be a function of the performer’s interpretation and properties of the musical instrument, involving both the performer and listener as active participants. Unfortunately, existing methodologies aren’t conducive to empirical study of performances on non-keyboard instruments, so comparative data that would distinguish instrument-specific effects is scarce. For this reason, an experimental methodology is used in this study which allows performances on multiple instruments to be compared quantitatively to one another, and for the instrumental performances to be presented to listeners as perceptual stimuli.
CHAPTER 3

MUSICAL PERFORMANCES

This chapter details a methodology for quantitatively studying music performances across a variety of instruments. First, the literature of empirical performance studies is reviewed, focussing on the methods for collecting and representing dependent variables. The methodology of the current study is then presented, including a section documenting the process for extracting MIDI data from digital audio. Finally, a computer algorithm is presented which compares a performance file to a quantized score file for the purpose of identifying and categorizing note errors in the performance.

The previous chapter discusses the pervasive use of MIDI (Musical Instrument Digital Interface) as a vehicle for representing musical performances, usually on a keyboard instrument. This is an example of empirical methodology being driven by the pragmatics involved in collecting and analyzing data from complex human behavior. The standardization of methods of data collection and analysis is often referred to as a research paradigm; it doesn’t qualify as a scientific paradigm in the sense that Kuhn (1962) meant, but the term is a helpful reminder of the influence, both good and bad, that such standardization has on a research field. Clearly, research paradigms can provide focus to a body of
research, help to clarify problems, lead to well-defined paths of empirical study, and unite researchers in a common and recognizable interest. However, they can also unintendedly lead to an overly-narrow definition of research relevancy, such that the results obtained from the paradigm may not be appropriately generalizable. In the beginning of chapter 5 of *The Musical Mind: The cognitive psychology of music* (1985), for example, John Sloboda discusses quite frankly how research paradigms can also degrade into explorations of minutiae, losing sight of the original stated goals.

In the case of investigating instrument-specific performance differences, the standard paradigm of capturing data from a MIDI device is not effective. An attempt was made in Walker (1996) to obtain MIDI data from non-keyboard instruments in real time, but available interfaces interfered differentially with the instruments in the study, as discussed in Chapter 2. The process of data collection should be as transparent as possible to the participants; it is paramount that the measurement devices in a study not bias the measurements themselves. Unfortunately, MIDI performances on non-keyboard instruments suffer from both lower precision and lower naturality.

With a goal of studying the detailed structure of expressive musical performance, the researcher must decide what features are relevant, and how to quantify and measure them. Typically, empirical studies have used three fundamental kinds of data: errors, patterns of timing, and patterns of dynamics. Note error types and locations are defined by pitches played by the performer other than those notated in the score, and can consist of substitutions, additions, or deletions of notes. Note errors can also be categorized by the nature of a wrong note; for example, a substitution or an addition may or may not be drawn from the
musical key of the notated piece (this distinction obviously does not apply to deleted
notes, since they were not played at all). Patterns of timing are defined as millisecond-
level deviations from the beat locations implied by a strictly-quantified interpretation of
the score, and patterns of dynamics are defined as deviations in perceived or measured
loudness of notes, usually measured as intensity of the acoustic onset.

For example, empirical studies have investigated timing and dynamics patterns as a
function of meter and phrasing (Drake & Palmer, 1993; Shaffer, 1984; Shaffer, Clarke, &
Todd, 1985; Sloboda, 1983); timing and dynamics patterns of expressive imitations (Repp,
2000); timing patterns as a function of rhythmic structure (Busse, 2002; Clarke, 1993);
timing patterns of melody leading (Palmer, 1996b; Timmers, Ashley, Desain, & Heijink,
2000); timing of grace notes (Timmers, Ashley, Desain, Honing, & Windsor, 2002); error
rates and timing as a function of phrasing (Palmer, 1989b; Palmer & van de Sande, 1995)
and auditory feedback (Pfordresher, 2003); error types (Palmer & van de Sande, 1993) and
timing patterns (MacKenzie, Van Eerd, Graham, Huron, & Wills, 1986) as a function of
harmonic and melodic structure; developmental differences in error rates (Drake, Dowling,
& Palmer, 1991); the relationship between rhythmic timing and global tempo (Desain
& Honing, 1994; Repp, Windsor, & Desain, 2002); timing relationships between multiple
performers (Ellis, 1991; Rasch, 1979); timing and dynamics patterns in scales (MacKen-
ze & Van Eerd, 1990); and the perceived emotional content as related to timing and
Interestingly, although performers can modulate attributes of their performances, they
seem not to be able to totally strip expressive elements from their performance; systematic
variability of timing and dynamics patterns, tied to the structure of the piece, is evident even when performers are instructed to play mechanically (Drake & Palmer, 1993; Kendall & Carterette, 1990; Palmer, 1989a, 1996b; Seashore, 1938).

These measures, of course, cannot comprehensively describe a musical performance, but they are used as the predominant dependent variables for empirical studies of music performance. In particular, patterns of timing deviations are used more than any other variable as a quantitative measure of performance. Interestingly, other performance variables, such as articulation and timbre, are not well represented in the literature, even though they are commonly considered to be important vehicles of musical expression. For timbre, at least, there is an obvious explanation; it is extraordinarily difficult to quantify. A few studies (e.g. Gabrielsson & Juslin, 1988) have attempted to examine differences in performed timbre, but only by way of qualitative descriptors. Articulation, on the other hand, is easily obtainable from MIDI data. Most studies of performance have looked only at the onsets of events, thus defining note durations by interonset intervals (IOIs) rather than by the time between onset and offset. Since articulation is presumed to be a dimension of expressive performance (Clarke, 1988; Gabrielsson, 1988; Sundberg, 2000), performance studies should include articulation data as well as IOIs (for examples, see Bresin & Battel, 2000; Madison, 2000; and Repp, 1997).

The current study attempts to retain the essential natural properties of performances on a variety of instruments while also maintaining quantitative and analytic rigor. As a result, a number of difficult technical issues have had to be addressed. The approach was to use MIDI data only as a way of representing performances for analysis, and not as a
recording format. Performances were recorded as digital audio and later translated into MIDI, which then allowed the extraction of errors, timing deviations, articulation, and dynamics as dependent variables. Since MIDI translation was a critical and intensive step, it and related issues are discussed in detail below.

Performance Experiment

This experiment is a quantitative investigation of instrument-specific characteristics in expressive music performance. Recordings of the same musical pieces by performers on four targeted instruments (clarinet, guitar, piano, and trumpet) were collected for analysis in terms of performance differences across instruments. The instruments were chosen in an attempt to represent several different categories of musical instruments (percussive string, plucked string, brass, and woodwind). Additionally, this set balances single-voiced (monophonic) wind instruments vs. multi-voiced (polyphonic) non-wind instruments: piano and guitar are both capable of producing simultaneous pitches, whereas trumpet and clarinet generally produce only one pitch at a time.

The method is similar to the performance experiment from Walker (1996), with a few major differences: each participant in this study played each piece of music twice, once with normal expression and once with no intended expression (defined as a mechanical, or “flat” performance); four instruments were used, rather than two; the recordings were made directly from each instrument’s acoustical output, rather than from a MIDI device (thereby allowing a naturalistic use of expressive cues); and the design was between-subject (each participant played one instrument only).
Although MIDI was not used as a recording format, it was used to represent the data for analysis. Pitch-to-MIDI software was used in order to transform the digital audio recordings to standard MIDI files. Each note then possessed clearly defined values of onset and offset times (in milliseconds), pitch values, and MIDI key velocities, which represent relative intensities. These values could then be transformed into measures of event timing, pitch, articulation, and dynamics.

Empirical studies have analyzed expressive patterns of performance variables in a variety of contexts, but comparisons across instruments have been lacking. Although Kendall and Carterette (1990) compared performances of the same music on different instruments, they had only one performer play each instrument; thus performer-specific factors could not be distinguished from instrument-specific factors. Also, their use of piano performances as models for all other performances confounds an analysis of instrument-specific performance differences. This experiment separates instrument-specific from performer-specific factors by comparing eight performers on each of four instruments: clarinet, guitar, piano, and trumpet. In addition, expressive and mechanical performances by the same performers on each instrument were collected; comparisons of performances with the presence or absence of expressive intention allows identification of instrument-specific cues of expression.

Walker’s (1996) study demonstrated that instrument-specific characteristics can shape music performance. Performers on piano and guitar displayed different patterns of key choice and relative frequencies of scalar degree in improvisation, for example. However, several aspects of that study were problematic, many of which arise from the use of
MIDI as a performance tool. MIDI is primarily defined in terms of piano keyboards, and displays some performance limitations when used on other instruments. For example, there were slight (10-20 ms) delays in note production when MIDI was employed on the guitar, as well as some mistakes in MIDI pitch tracking. This can result in an intrusion upon the naturalness of the performance, essential for a study of musical expression. Furthermore, the use of MIDI in performance limits the number of instruments that can be employed, since each instrument requires a different (and usually expensive) real-time MIDI-tracking device.

The inclusion of several musical instruments is desirable for a scientific assessment of instrument-specific influences on performance expression. Therefore, all recordings in the current study were made acoustically and recorded either onto digital audio tape (DAT) or direct to a computer disk for later conversion (via non-real-time pitch-to-MIDI software) to MIDI data for analysis. Three major improvements on Walker (1996) result from this design: 1) performers were able to use their own instruments and hear the same natural timbre to which they are accustomed, thus improving on the naturalness of the performances; 2) multiple instruments can easily be compared; and 3) the accuracy of MIDI data from non-keyboard instruments is improved. The only significant restriction on this design is that the performance timbre cannot be defined independently of the physical instrument. However, most of the significant findings from Walker (1996) were effects of physical instrument, not of sounded timbre. Therefore, the restriction on sounded timbre does not appear to be a major limitation to the methodology of the current study.
Method

Participants

Participants were experienced musicians who play one of four instruments: piano, guitar, trumpet, or clarinet. There were 32 participants total, 8 for each of the four instruments. The participants ranged in age between 17 and 52 (mean=30.5), and had between 5 and 39 years of experience playing their instrument (mean=18.5). They were recruited from the Columbus, OH area, and were typically either actively-performing local musicians or graduate students in a school of music. Each performer was paid $10 for participating in the study.

Materials

The musical stimuli consisted of two Western tonal single-voiced melodies, shown in Appendices A and B. The melody labelled “Melody 1” (Appendix A) was composed specifically for this experiment, and has some carefully chosen features. For example, it has two descending scalar runs, a fairly extensive range of pitches, and three pitch jumps of at least one octave. The intent was to provide a stimulus which would differentially challenge the defining characteristics of each instrument without placing an inordinate amount of difficulty on any subset of instruments. Scalar runs require a change in hand position (piano and possibly guitar), string (guitar), blowing (trumpet), or reeding (clarinet), depending upon the instrument, and large pitch jumps require either a trajectory movement (piano and guitar) or physical demand on breath (trumpet and clarinet). Since the demands of the four instruments manifest themselves in different ways, stimuli such as
Melody 1 should be good mechanisms for revealing performance differences as a function of physical instrument constraints.¹

The melody labelled “Ständchen” (Appendix B) was drawn from the classical musical literature, and was composed by Franz Schubert. It was chosen because of its potential for expressive content. It has a relatively simple yet interesting repeating rhythmic triplet pattern which lends itself to interpretation. While Melody 1 contains more technical challenges, it is also more mechanical in nature. By contrast, Ständchen was expected to display more deviation from the mechanical performance to the expressive performance than Melody 1. If expressive performances are realized differently on different instruments, this should be apparent in interpretations of Ständchen.

Thus, the two melodies represent different possible dimensions in the space of instrument-specific characteristics. Although they by no means are representative of all of the possible structural or stylistic influences on expressive performance, they at least provide two different vehicles for examining differences between instruments.

**Equipment**

All performances were recorded using a Shure BG-4 unidirectional condenser microphone arranged on a microphone stand close to the performer. The signal from the microphone was fed (using balanced cables) through a mixer into either a digital audio

¹. The notation as shown was presented to piano performers. Clarinet and trumpet are transposing instruments, so performers on those instruments had each piece notated one full step higher so that the sounded pitches would be the same for every instrument. Additionally, notation for guitar performances included blank tablature for guitarists to mark fingering.
tape (DAT) machine or directly into a computer running a direct-to-disk recording software package. Performances which were recorded to DAT were later transferred to computer and saved as files. All analog-to-digital conversion was performed at CD quality (16-bit sampling at 44.1 kHz sampling rate). Subsequent Pitch-to-MIDI conversion was accomplished by using Emagic Logic Audio Gold 4 sequencing software running on a Macintosh G4 computer.

Procedure

Each participant filled out a brief questionnaire which asked about their demographic information and musical background, read a form describing the nature of the experiment, and signed to indicate voluntary consent. The participant was then given a sheet of instructions to read, and the experimenter answered any questions about the task. The participant was given one of the notated musical pieces (half of them received Melody 1 first, and half received Ständchen first) and was asked to practice the piece until they felt comfortable enough to record a performance of it. They were instructed that their performance should have an “appropriate amount” of expression, subject to their interpretation of the piece. Participants were told that they could add whatever markings they felt were helpful to the score in order to facilitate an acceptable performance. During this first practice session, the experimenter set recording levels at a point to be maintained throughout the recording session.

When the participant was ready, the experimenter recorded a performance of the musical piece. The participant was then asked if the performance was acceptable as a final
recording; if not, the piece was re-recorded until the performer was satisfied. Judgements of acceptability of each performance were left entirely to the participant. The participant was next asked to perform the same piece of music mechanically (i.e. with no intended musical expression) and a second recording was made, again subject to the participant’s approval. Once the mechanical recording was completed, the participant was asked to mark an indication of performed tempo on the score. A score for the next piece of music (Melody 1 or Ständchen) was then given to the participant, and two recordings (expressive and mechanical) were made using the same procedure as with the first piece. After all the recordings were collected, the participant was given a debriefing sheet which explained the goals of the study and provided a space for general comments.

Data Transformation

In order to conduct comparative quantitative analyses, the performances needed to be converted into MIDI events with discrete onsets, offsets, and intensities. A two-step process was employed for MIDI conversion: an analysis by software followed by manual fine-tuning and correction. The digital audio files from each performer were imported into tracks within Emagic Logic Audio Gold, and the Audio to Score function was used to create corresponding MIDI events from the recordings. The resulting MIDI transcription is a first approximation to the audio performance, but is not perfectly accurate; onsets are frequently misplaced, individual notes (particularly those with vibrato) are sometimes scored

---

2. There were actually two more sets of recordings for each performer, involving an improvisation task. These recordings were not included in this study, however.
as multiple short notes, incidental noises (i.e. sounds of keys, frets, breathing, etc.) can be scored as spurious notes, and performed notes are occasionally misidentified or even dropped. Despite these imperfections, however, the software-produced MIDI sequences are valuable translations from the digital audio files, in that they provide an objective set of markers defining each performance. With some trial and error, software settings were obtained differentially for each instrument type which seemed to minimize transcription errors and thus provide the best basis for objective, accurate data. These settings are given in Appendix C.

In order to avoid bias during correction of the transcriptions, a neutral assistant provided all decision-making regarding changes to MIDI events. The experimenter played the MIDI output (in a synthesized timbre) and the digital audio simultaneously to separate loudspeakers. The assistant could request repeated hearings of sections or individual notes, and would identify perceived mismatches between the two outputs. The experimenter would then modify in small increments, at the assistant’s direction, attributes of an event such as its onset, etc., until the assistant was satisfied that the outputs matched satisfactorily.

The overarching principle in the manual correction phase was to change as little as possible; when in doubt, the software-produced output was left alone. Events were only changed when the transcription clearly seemed to be in error, and as much of the original information in a changed event was maintained as possible. For example, key velocities (i.e. intensities) were maintained for events which had altered onsets, and onsets were maintained for events which had altered pitches. Key velocities, in general, were not typi-
ally modified. On the rare occasion in which a note had to be added, the new note was initially assigned a velocity value equal to the velocity of the preceding note, and then modified as deemed appropriate. When a single note event was incorrectly represented as several short notes, a new note was defined which had the onset and velocity of the first note in the series, and the offset of the last note in the series. Note offsets, in particular, are difficult to identify unambiguously, which is further compounded by the fact that different instruments have radically different patterns of decay. For many analyses this is not a practical concern, since timing is often measured in terms of interonset intervals, which are defined solely by note onsets. However, the goal was to have as complete a representation as possible, and so offsets were treated with as much care and deliberation as were onsets.

After all performances had been adequately transcripted, the MIDI tracks were exported from Logic and saved as individual MIDI files. Each MIDI file was then translated to text equivalents in order to extract analyzable data. One row of data was written for each performed note, with seven columns representing: MIDI channel number; onset time (all times represented in milliseconds); offset time; duration (i.e. offset minus onset); MIDI pitch value (a number between 0 and 127 representing by semitones the musical note and octave which was played, 60 being middle C); key velocity (a measure of loudness as a number between 0 and 127 representing the intensity of each note); and interonset interval (IOI, the difference between the onset of the current note and the onset of the next note; defined as zero for the last note in the file).
Algorithmic Identification of Errors

With the data arranged as columns of numbers, performances can be analyzed in terms of any of the columns. However, variability in the performance needs to be considered in relation to the patterning of notes in the musical score (e.g. Bora, Tufan, & Bilgen, 2000; Dixon, 2001; Eck, 2001; Heijink, Desain, Honing, & Windsor, 2000; Müller & Mazzola, 2003). Therefore, a C++ computer program was written which takes as input two files in the text format described above: a performance file and a comparison file which is created as a strictly-quantized representation of the performed score. The program, called PerfComp, then provides a heuristic-based analysis of the performance. The main difficulty is that incorrect notes can produce an ambiguous interpretation of the performance; a note which does not match the score might truly have been played incorrectly or it might be a note from the score but played at the wrong time. The advantages of using a computer algorithm for error categorization, rather than manually coding them, are that the algorithm provides both a much more efficient process and an objective, consistent method.

PerfComp considers a few alternatives for each mismatch and decides whether to categorize the event as a substitution (i.e. a wrong note), an addition (i.e. an extra note), or a deletion (i.e. a note from the score was not played). The fundamental mechanism is to search for a nearby matching note in the comparison file, which would indicate the possibility that the performance has additions or deletions. However, the presence of a matching note does not guarantee that the performed note was not a substitution, and PerfComp
looks at both global and local properties of the performance file in order to reach an optimal interpretation.

At the global level, PerfComp allows for a bias towards either additions or deletions. It accomplishes this by pre-processing both files and counting the number of notes. The difference in number of notes in the two files is stored in a variable called `offset`, as defined in the following snippet of code (`pfile` and `cfile` are objects representing, respectively, the performance file and the comparison file, and `GetNumNotes()` is a public function which returns a count of the number of notes contained in its object):

```cpp
offset = pfile->GetNumNotes()
    - cfile->GetNumNotes();
```

If there are more notes in the performance file (i.e. if `offset > 0`), PerfComp will look for additions before deletions. Conversely, if there are more notes in the comparison file (i.e. if `offset < 0`), PerfComp will look for deletions before additions. Furthermore, `offset` is dynamically adjusted as errors are encountered, to represent a running count of the discrepancy between remaining notes in the performance file and the comparison file. Therefore, the expectations of the program may change as data is processed. A user-supplied parameter, called `minLook`, can (for non-zero values) ensure that a minimum number of notes will be searched in both directions (i.e. for both additions and deletions) for possible matches, regardless of the value of `offset`.

Pre-processing the files also allows PerfComp to interpret note onsets in terms of beats in the musical score; a user-supplied parameter gives the number of musical beats in the score, an overall tempo is determined from the first and last onset in the file, and each
note in the performance file is assigned a floating-point number defining its fractional beat position relative to its corresponding note in the comparison file. Additionally, PerfComp calculates the overall octave (register) in which the performance was played, and later takes that into account when comparing notes from the performance and comparison files. Therefore, playing the piece an octave up or down from the notation won't result in all the notes being flagged as incorrect. However, the octave of individual notes is tracked, so a single note played an octave up or down will be flagged as incorrect.

After the files have been read and pre-processed, PerfComp works through the files from beginning to end, comparing one note at a time from the performance file to the comparison file and keeping a running track of the current position and relevant state information. Two variables, lookAheadPerf and lookAheadComp, represent respectively the number of notes in the performance file and the comparison file which will be searched forward from the current position for potential matches. Note that a match that is found ahead in the performance file would indicate potential addition(s), while a match that is found ahead in the comparison file would indicate potential deletion(s). The following code illustrates how the variables are defined and how the general search process works (\(i\) and \(j\) are indexes for the current position in, respectively, the performance file and the comparison file, \(tmp_i\) and \(tmp_j\) are temporary indexes used to search forward from the current position, and \(found\) is a variable which will become a non-zero distance between the current position and the matching note if one is obtained):

\[\text{3. Actually, the parameter is the number of beats minus one, since a beat length isn’t calculated for the last note in the score.}\]
if (minLook > offset) lookAheadPerf = minLook;
else lookAheadPerf = offset;
if (minLook > -(offset)) lookAheadComp = minLook;
else lookAheadComp = -(offset);
{ ... }
while (((!found) && ((tmpi < (i + lookAheadPerf)) || (tmpj < (j + lookAheadComp))))
  
  if (offset > 0)
    
    {first look one note ahead for
     addition...}
    
    {...then look one note ahead for
     deletion}
  else
    
    {first look one note ahead for
     deletion...}
    
    {...then look one note ahead for
     addition}

If no matching note is found within the above search (i.e. if found = 0), then the current note in the performance file is simply marked as a substitution and the variables are incremented to look at the next note. Even if a matching note is found, however, it might not be determined to be the intended note of the performer. The program calculates two metrics: one based on the matching note (i.e. treating the notes in between as addi-
tions or deletions); and one based on the current (incorrect) note (i.e. treating the current note as a substitution, and therefore the match as being spurious).

Each metric has three components: (1) the extent to which the additions or deletions would decrease (or increase) the running count of remaining note surplus or deficit (in other words, how using the match would change the dynamic value of offset; this is a non-zero factor only for the matching note, since using the current note as an incorrect substitution would obviously not change the running count); (2) the number of correctly performed notes following the note in question (after an adjustment in position for the case of additions or deletions); and (3) the difference in the beat locations between the note that should have been played (the correct note from the comparison file) and the note in question (either the current or the matching note). If the overall metric (the combination of all three factors) yields a better score for the matching note, then it is used and the notes in between are marked as additions or deletions. If the metric is better using the current note, then it is treated as an incorrect substitution and the matching note is ignored. The code below shows the computation and comparison of the metrics and the affect that using the match has on the value of offset. The variable offsetReduct is the amount of change to offset by using the match, and can be positive or negative; offsetMetric, numCorrectFound, and beatDiffFound are the components of foundMetric, which treats the match as a true indication of additions or deletions; numCorrectCurr and beatDiffCurr are the components of currentMetric, which treats the incorrect note as a substitution; GetNumCorrect() is a function which returns the number of correct notes in the performance file following a given position; and
GetBeat() is a public function of the file objects which returns the beat location for a given note.

\[
\text{offsetReduct} = \text{abs}(\text{offset}) - \text{abs}(\text{offset} - \text{found});
\]

\[
\text{offsetMetric} = (\text{float}) \left( \frac{(\text{abs}(\text{offset}) + 1) \times (\text{offsetReduct})}{(\text{pfile} \to \text{GetNumNotes}() - i)} \right);
\]

\[
\text{numCorrectCurr} = \text{GetNumCorrect}(	ext{pfile}, i, \text{cfile}, j);
\]

\[
\text{beatDiffCurr} = \text{pfile} \to \text{GetBeat}(i) - \text{cfile} \to \text{GetBeat}(j);
\]

if (found > 0)

\[
\{\text{numCorrectFound} = \text{GetNumCorrect}(	ext{pfile}, \text{tmpi}, \text{cfile}, j);
\]

\[
\text{beatDiffFound} = \text{pfile} \to \text{GetBeat}(\text{tmpi}) - \text{cfile} \to \text{GetBeat}(j);\}
\]

else

\[
\{\text{numCorrectFound} = \text{GetNumCorrect}(	ext{pfile}, i, \text{cfile}, \text{tmpj});
\]

\[
\text{beatDiffFound} = \text{pfile} \to \text{GetBeat}(i) - \text{cfile} \to \text{GetBeat}(\text{tmpj});\}
\]

\[
\text{currentMetric} = \text{numCorrectCurr} - ((\text{beatDiffCurr}) \times (\text{beatDiffCurr})
\]

\[
\text{foundMetric} = \text{numCorrectFound} + \text{offsetMetric} - ((\text{beatDiffFound}) \times (\text{beatDiffFound})
\]
if (foundMetric > currentMetric)
{
    offset = offset - found;
    if (found > 0)
    {
        {Mark all the notes in between as
         additions}
    }
else
    {Mark all the notes in between as
     deletions}}
else
    {Mark the current note as a substitution}

As the above section of code shows, the offset component of the metric is defined as a function of the current position in the piece and both the direction and distance of the matching note. The variable offsetReduct will be positive if the matching note would reduce the discrepancy between remaining notes in the performance file and the comparison file, and will be negative if it would increase the discrepancy (the absolute value calls are necessary in order for this to be true regardless of the direction of the bias). The numerator of offsetMetric is the product of abs(offset)+1 and offsetReduct in order to weight more heavily reductions in large discrepancies. The +1 term ensures that increases in the discrepancy (i.e. negative values of offsetReduct) will be weighted for cases in which the current value of offset is zero. The denominator of offsetMetric is a count of the remaining number of notes in the performance file. This serves to weight a change in the discrepancy more heavily (either positively or nega-
tively) if it is towards the end of the piece, since the ends of the two files should be reached at the same time. The final value of offsetMetric can be either positive or negative (the sign will be the same as that of offsetReduct); positive values support an interpretation of the matching note to give additions or deletions, while negative values support an interpretation of the error as a substitution.

The components to weight the number of correctly performed notes are fairly straightforward, and simply count the number of correct notes which would follow the current error if it is coded as a substitution (numCorrectCurr) or as additions or deletions (numCorrectFound). These components cannot be negative, and larger values tend to indicate that the given interpretation of the error is valid. The components to weight the difference in beat locations are similarly simple, and take the difference between the beat location of the note that should have been played (from the comparison file) and the beat location of the note that was played (from the performance file). This is calculated for coding as a substitution (beatDiffCurr) and as additions or deletions (beatDiffFound), and larger differences would tend to indicate that the performed note should not be interpreted at the given location in the score.

Once all the components have been calculated, the overall metric is defined as the sum of the proper numCorrect variable and offsetMetric (offsetMetric is always zero for currentMetric, since a coding of substitution wouldn’t change offset), minus the square of the proper beatDiff variable. Overall, a large positive value is best. Therefore, the error is coded as additions or deletions only if foundMetric > currentMetric; otherwise, the error is coded as a substitution. After the error has
been properly coded, the indexes are updated and the program moves on to consider the next note.

Note that the relative weighting of the three components of an overall metric is fairly simple. It may be that a different weighting would produce better performance in general, but the usage in this study was utilitarian, in that the main interest was to obtain good results for the two pieces which were performed. Additional experimentation could be conducted in order to establish empirically-obtained weightings of the components which would perform well for a range of performances and musical styles, but that was certainly not the focus of the current study. In particular, it is speculated that a higher weighing of the squared beatDiff terms might produce good results. Finally, note that values of minLook can influence the identification of errors; small values tend to limit the range over which note matches are found and therefore bias the result towards more substitution codings.

For the performances in this study, minLook was generally set to a value of 5, which is a fairly large range in which to detect additions and deletions. Figure 3.1 shows a sample output file from PerfComp, from a piano performance of Ständchen. Figure 3.2 shows an log file of verbose reporting from PerfComp, from the same performance. Notice that the first error was categorized as a substitution (Ds5 played instead of E) because a match was not found. The next error found a match and determined that the A5 was an added note, reducing offset from 2 to 1. The third error found a match but did not use it on the basis of the metrics; therefore the D6 was determined to be a substitution for E. The final error found a match and categorized the Ds6 as an added note, reducing offset to
<table>
<thead>
<tr>
<th>Chn</th>
<th>NoteOn</th>
<th>NoteOff</th>
<th>Dur</th>
<th>Pitch</th>
<th>Name</th>
<th>Oct</th>
<th>Vel</th>
<th>IOI</th>
<th>Beat</th>
<th>CmpBeat</th>
<th>BeatDif</th>
<th>Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>6492</td>
<td>6601</td>
<td>109</td>
<td>69</td>
<td>A</td>
<td>5</td>
<td>26</td>
<td>245</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>ok</td>
</tr>
<tr>
<td>1</td>
<td>6737</td>
<td>6877</td>
<td>140</td>
<td>70</td>
<td>As</td>
<td>5</td>
<td>28</td>
<td>219</td>
<td>1.3741</td>
<td>1.332</td>
<td>-0.0421002</td>
<td>ok</td>
</tr>
<tr>
<td>1</td>
<td>6956</td>
<td>7112</td>
<td>156</td>
<td>69</td>
<td>A</td>
<td>5</td>
<td>26</td>
<td>238</td>
<td>1.7085</td>
<td>1.666</td>
<td>-0.0424999</td>
<td>ok</td>
</tr>
<tr>
<td>1</td>
<td>7194</td>
<td>8103</td>
<td>909</td>
<td>74</td>
<td>D</td>
<td>6</td>
<td>23</td>
<td>947</td>
<td>2.01791</td>
<td>2</td>
<td>-0.0719116</td>
<td>ok</td>
</tr>
<tr>
<td>1</td>
<td>8141</td>
<td>8339</td>
<td>198</td>
<td>69</td>
<td>A</td>
<td>5</td>
<td>23</td>
<td>318</td>
<td>3.51792</td>
<td>3.5</td>
<td>-0.0179234</td>
<td>ok</td>
</tr>
<tr>
<td>1</td>
<td>8459</td>
<td>8602</td>
<td>143</td>
<td>67</td>
<td>G</td>
<td>5</td>
<td>28</td>
<td>219</td>
<td>1.3741</td>
<td>1.332</td>
<td>-0.0422497</td>
<td>ok</td>
</tr>
<tr>
<td>1</td>
<td>8666</td>
<td>8802</td>
<td>136</td>
<td>69</td>
<td>A</td>
<td>5</td>
<td>41</td>
<td>219</td>
<td>4.31957</td>
<td>4.332</td>
<td>0.0124331</td>
<td>ok</td>
</tr>
<tr>
<td>1</td>
<td>8885</td>
<td>9048</td>
<td>163</td>
<td>67</td>
<td>G</td>
<td>5</td>
<td>39</td>
<td>238</td>
<td>4.65397</td>
<td>4.666</td>
<td>0.0120339</td>
<td>ok</td>
</tr>
<tr>
<td>1</td>
<td>9123</td>
<td>9716</td>
<td>593</td>
<td>74</td>
<td>D</td>
<td>6</td>
<td>42</td>
<td>667</td>
<td>5.01738</td>
<td>5</td>
<td>-0.0173779</td>
<td>ok</td>
</tr>
<tr>
<td>1</td>
<td>9790</td>
<td>10198</td>
<td>408</td>
<td>67</td>
<td>G</td>
<td>5</td>
<td>35</td>
<td>646</td>
<td>6.03585</td>
<td>6</td>
<td>-0.0358467</td>
<td>ok</td>
</tr>
<tr>
<td>1</td>
<td>10436</td>
<td>11226</td>
<td>790</td>
<td>69</td>
<td>A</td>
<td>5</td>
<td>32</td>
<td>947</td>
<td>7.02225</td>
<td>7</td>
<td>-0.0222497</td>
<td>ok</td>
</tr>
<tr>
<td>1</td>
<td>11383</td>
<td>11615</td>
<td>232</td>
<td>67</td>
<td>G</td>
<td>5</td>
<td>47</td>
<td>312</td>
<td>8.46826</td>
<td>8.5</td>
<td>0.0317383</td>
<td>ok</td>
</tr>
<tr>
<td>1</td>
<td>11695</td>
<td>11843</td>
<td>148</td>
<td>67</td>
<td>G</td>
<td>5</td>
<td>47</td>
<td>233</td>
<td>8.94467</td>
<td>9</td>
<td>0.0553341</td>
<td>ok</td>
</tr>
<tr>
<td>1</td>
<td>11928</td>
<td>12136</td>
<td>208</td>
<td>65</td>
<td>F</td>
<td>5</td>
<td>59</td>
<td>259</td>
<td>9.30044</td>
<td>9.332</td>
<td>0.0315561</td>
<td>ok</td>
</tr>
<tr>
<td>1</td>
<td>12187</td>
<td>12258</td>
<td>71</td>
<td>63</td>
<td>Ds</td>
<td>5</td>
<td>30</td>
<td>142</td>
<td>9.69592</td>
<td>9.666</td>
<td>-0.0299206</td>
<td>E</td>
</tr>
<tr>
<td>1</td>
<td>12329</td>
<td>13403</td>
<td>1074</td>
<td>69</td>
<td>G</td>
<td>5</td>
<td>51</td>
<td>1857</td>
<td>9.91275</td>
<td>10</td>
<td>0.0872536</td>
<td>ok</td>
</tr>
<tr>
<td>1</td>
<td>14186</td>
<td>14302</td>
<td>116</td>
<td>69</td>
<td>A</td>
<td>5</td>
<td>40</td>
<td>229</td>
<td>12.7483</td>
<td>13</td>
<td>0.251727</td>
<td>ok</td>
</tr>
<tr>
<td>1</td>
<td>14415</td>
<td>14543</td>
<td>128</td>
<td>70</td>
<td>As</td>
<td>5</td>
<td>33</td>
<td>194</td>
<td>13.0979</td>
<td>13.332</td>
<td>0.234058</td>
<td>ok</td>
</tr>
<tr>
<td>1</td>
<td>14609</td>
<td>14769</td>
<td>160</td>
<td>69</td>
<td>A</td>
<td>5</td>
<td>35</td>
<td>236</td>
<td>13.3942</td>
<td>13.666</td>
<td>0.271832</td>
<td>ok</td>
</tr>
<tr>
<td>1</td>
<td>14845</td>
<td>15656</td>
<td>811</td>
<td>77</td>
<td>F</td>
<td>6</td>
<td>32</td>
<td>975</td>
<td>13.7545</td>
<td>14</td>
<td>0.245474</td>
<td>ok</td>
</tr>
<tr>
<td>1</td>
<td>15820</td>
<td>16011</td>
<td>191</td>
<td>69</td>
<td>A</td>
<td>5</td>
<td>32</td>
<td>361</td>
<td>15.2433</td>
<td>15.5</td>
<td>0.256708</td>
<td>ok</td>
</tr>
<tr>
<td>1</td>
<td>16181</td>
<td>16219</td>
<td>38</td>
<td>69</td>
<td>A</td>
<td>5</td>
<td>39</td>
<td>153</td>
<td>15.7945</td>
<td>999</td>
<td>999</td>
<td>ADD</td>
</tr>
<tr>
<td>1</td>
<td>16334</td>
<td>16551</td>
<td>146</td>
<td>67</td>
<td>G</td>
<td>5</td>
<td>38</td>
<td>221</td>
<td>16.0281</td>
<td>16</td>
<td>-0.0281391</td>
<td>ok</td>
</tr>
<tr>
<td>1</td>
<td>16551</td>
<td>16668</td>
<td>113</td>
<td>69</td>
<td>A</td>
<td>5</td>
<td>34</td>
<td>187</td>
<td>16.3656</td>
<td>16.332</td>
<td>-0.0335922</td>
<td>ok</td>
</tr>
<tr>
<td>1</td>
<td>16742</td>
<td>16854</td>
<td>112</td>
<td>67</td>
<td>G</td>
<td>5</td>
<td>24</td>
<td>181</td>
<td>16.6511</td>
<td>16.666</td>
<td>0.0148697</td>
<td>ok</td>
</tr>
<tr>
<td>1</td>
<td>16923</td>
<td>17241</td>
<td>318</td>
<td>74</td>
<td>D</td>
<td>6</td>
<td>23</td>
<td>436</td>
<td>16.9275</td>
<td>17</td>
<td>0.0724945</td>
<td>E</td>
</tr>
<tr>
<td>1</td>
<td>17359</td>
<td>17875</td>
<td>516</td>
<td>75</td>
<td>Ds</td>
<td>6</td>
<td>20</td>
<td>603</td>
<td>17.5933</td>
<td>999</td>
<td>999</td>
<td>ADD</td>
</tr>
<tr>
<td>1</td>
<td>17962</td>
<td>18317</td>
<td>209</td>
<td>74</td>
<td>D</td>
<td>6</td>
<td>12</td>
<td>294</td>
<td>18.514</td>
<td>18.5</td>
<td>-0.0139961</td>
<td>ok</td>
</tr>
<tr>
<td>1</td>
<td>18256</td>
<td>19036</td>
<td>780</td>
<td>72</td>
<td>C</td>
<td>6</td>
<td>9</td>
<td>975</td>
<td>18.9629</td>
<td>19</td>
<td>0.0370827</td>
<td>ok</td>
</tr>
<tr>
<td>1</td>
<td>19231</td>
<td>19451</td>
<td>220</td>
<td>70</td>
<td>As</td>
<td>5</td>
<td>33</td>
<td>325</td>
<td>20.4517</td>
<td>20.5</td>
<td>0.048317</td>
<td>ok</td>
</tr>
<tr>
<td>1</td>
<td>19556</td>
<td>19688</td>
<td>132</td>
<td>70</td>
<td>As</td>
<td>5</td>
<td>39</td>
<td>219</td>
<td>20.9479</td>
<td>21</td>
<td>0.052063</td>
<td>ok</td>
</tr>
<tr>
<td>1</td>
<td>19775</td>
<td>19906</td>
<td>131</td>
<td>69</td>
<td>A</td>
<td>5</td>
<td>27</td>
<td>225</td>
<td>21.2823</td>
<td>21.332</td>
<td>0.0496635</td>
<td>ok</td>
</tr>
<tr>
<td>2</td>
<td>20000</td>
<td>20176</td>
<td>176</td>
<td>67</td>
<td>G</td>
<td>5</td>
<td>38</td>
<td>245</td>
<td>21.6259</td>
<td>21.666</td>
<td>0.040102</td>
<td>ok</td>
</tr>
<tr>
<td>1</td>
<td>20245</td>
<td>21500</td>
<td>1255</td>
<td>69</td>
<td>A</td>
<td>5</td>
<td>29</td>
<td>0</td>
<td>22</td>
<td>22</td>
<td>0</td>
<td>ok</td>
</tr>
</tbody>
</table>

Figure 3.1: Sample output from PerfComp
Analyzing performance in p6Standchenflat.asc, compared to StandchenQuant.asc

Writing output to p6Standchenflat.out

Octave Diff = 0

32 notes in comparison file
34 notes in performance file

Offset = 2

- i = 14  j = 14
  lookAheadPerf = 5  lookAheadComp = 5
  Not Found.  Offset = 2
  i = 21  j = 21
  lookAheadPerf = 5  lookAheadComp = 5
  FOUND!  Found = 1  Offset reduction = 1  Offset Metric = 0.230769
  # correct notes after current = 0  # correct notes after found = 2
  Beat Diff Current = -0.205483  Beat Diff Found = 0.0281391
  Current Metric = -0.0422234  Found Metric = 2.22998

  Used

  Offset = 1

  i = 25  j = 24
  lookAheadPerf = 5  lookAheadComp = 5
  FOUND!  Found = -1  Offset reduction = -1  Offset Metric = -0.222222
  # correct notes after current = 0  # correct notes after found = 0
  Beat Diff Current = -0.0724945  Beat Diff Found = -1.57249
  Current Metric = -0.00525545  Found Metric = -2.69496

  ***Not Used***

  Offset = 1

  i = 26  j = 25
  lookAheadPerf = 5  lookAheadComp = 5
  FOUND!  Found = 1
  Offset reduction = 1  Offset Metric = 0.25
  # correct notes after current = 0  # correct notes after found = 6
  Beat Diff Current = -0.822192  Beat Diff Found = 0.0139961
  Current Metric = -0.82192  Found Metric = 6.2498

  Used

  Offset = 0

2 substitutions
2 additions
0 deletions
4 total errors

Figure 3.2: Sample verbose log from PerfComp
zero. Also, notice that the output includes BeatDif, defined as the comparison beat minus the performed beat (therefore positive values reflect a speeding up while negative values reflect a slowing down). BeatDif is the main dependent variable for performed timing, although duration and IOIs are analyzed as well. BeatDif is not defined for additions and deletions since there are not comparison notes; flags of 999 and -999 are used in the output, and these lines were removed before analyses. The other dependent variable in the output is Vel (MIDI velocity), which defines the dynamics of the performance.

Summary

Much of the existing literature of empirical music performance studies consists of MIDI performance data drawn from keyboard instruments. An earlier study (Walker, 1996) addressed the effects that different instruments have on music performance, and obtained some interesting results. However, that study was limited in its ability to include multiple instruments and to collect data in a non-intrusive manner. Therefore, the current study was developed in order to establish a methodology for studying performances from an arbitrary set of instruments and with no intrusion of the recording process on the performers. Expressive and mechanical performances were recorded in this study directly to a digital audio format from four different instruments: clarinet, guitar, piano, and trumpet. The recordings were later transcribed into MIDI data, using a combination of automatic software translation and manual correction, in order to derive measures of error type, timing, articulation, and dynamics. A heuristic computer algorithm compared each performance file to a quantized score file in order to identify and categorize note errors.
Quantitative analyses and discussion of the performances are given in the following chapter.
CHAPTER 4

PERFORMANCE RESULTS AND DISCUSSION

This chapter describes the analyses of the performance data and discusses some interpretations of the findings. Some general expectations are outlined, then the analyses are presented separately for each major type of dependent variable (errors, timing, articulation, and dynamics). For each, the analyses for the two musical compositions (Melody 1 and Ständchen) are described separately. Finally, some implications of the pattern of results are discussed.

Measures of errors, timing, articulation, and intensity were obtained for each performance, as described in Chapter 3, and were analyzed as functions of instrument, expressive level, and event location. Separate analyses were conducted for each musical piece. Expectations were that patterns of event timing and intensity would differ across instruments, and that the differences would be more apparent in expressive than in mechanical performances. Although mechanical performances should display reduced systematic variability in event timing and intensity, the timing and intensity profiles for mechanical performances should not be completely flat; performers should not be able to totally strip
expressive elements away from their performance (e.g. Drake & Palmer, 1993; Kendall & Carterette, 1990; Palmer, 1989a, 1996b; Seashore, 1938).

The nature of all of the differences between instruments is difficult to precisely predict, but it was expected that the wind instruments (i.e. clarinet and trumpet) would display different patterns than guitar and piano due to the physical demands of performance. Constraints of breathing affect the timing patterns across long phrases and necessitate strategic planning for particularly difficult notes and passages, such as relatively high notes, much more so than on non-wind instruments. For example, results from Kendall and Carterette (1990) indicated that trumpet performances displayed a different pattern of event timing than other instruments in that study.

Furthermore, observed effects are likely to differ across musical pieces, since each score presents its own set of challenges and opportunities in its interpretation by a performer. Melody 1, with its quick scalar runs and large pitch intervals, was expected to be more technically challenging than Ständchen. This would suggest that Melody 1 should produce higher error rates and more disruptive shifts in timing, perhaps more so for the wind instruments. Ständchen, on the other hand, should show greater expressive variability and a more pronounced difference between the expressive and mechanical performances.

When errors do occur, they were not expected to be random, either in type or in location. Errors are often drawn from the diatonic scale defined by the musical key of piece (Palmer & van de Sande, 1993, 1995) and are harmonically related to the correct note. However, errors which arise as a result of physical performance demands were expected to
be more likely non-diatonic, or “chromatic”, errors. For example, large pitch jumps in a score, such as those in Melody 1, could result in “undershooting”, resulting in an incorrect note below the intended note. These types of errors should be more common in wind instruments than in the other instruments. Non-diatonic errors may also be more likely in guitar performances than those on piano due to the relative difficulty of sight-reading on guitar; standard musical notation for guitar allows for a variety of fingerings and fretting, and could result in chromatic errors of note location.

One guitar performer struggled with Melody 1; he played the piece without note errors, but was not able to adequately perform with the notated timing. Therefore, these performances were treated as outliers and were not included in the following analyses.

Error Analyses

Table 4.1 shows the error rates for each instrument separately for each musical piece; errors are categorized separately as substitutions, additions, and deletions. Totals, averages, and standard errors are shown for the 16 performances by each instrument (8 performers x 2 expression types), as well as the maximum number of errors made by a performance by each instrument. Overall, more errors were made in performances of Melody 1 (61) than of Ständchen (28), $\chi^2(1) = 12.24, p = .0005$. Performers tended to make different types of errors (predominantly substitutions) on both Melody 1, $\chi^2(2) = 56.49, p = .0000$, and Ständchen, $\chi^2(2) = 19.14, p = .0001$. 

58
<table>
<thead>
<tr>
<th>Instrument</th>
<th>Melody 1 Subs</th>
<th>Melody 1 Adds</th>
<th>Melody 1 Dels</th>
<th>Ständchen Subs</th>
<th>Ständchen Adds</th>
<th>Ständchen Dels</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clarinet</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Guitar</td>
<td>20</td>
<td>4</td>
<td>24</td>
<td>12</td>
<td>2</td>
<td>1</td>
<td>15</td>
</tr>
<tr>
<td>Piano</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Trumpet</td>
<td>27</td>
<td>7</td>
<td>2</td>
<td>36</td>
<td>4</td>
<td>3</td>
<td>7</td>
</tr>
</tbody>
</table>

Table 4.1: Error rates for Melody 1 and Ständchen by instrument
Error rates also differ as a function of instrument, both for Melody 1, $\chi^2(3) = 61.82$, $p = .0000$, and for Ständchen, $\chi^2(3) = 16.29$, $p = .0010$. Performances on clarinet and piano contained very few errors, while guitar and trumpet contained more. In particular, the highest error rates appear in trumpet performances of Melody 1, indicative of the physical demands which that piece places on trumpet players. As an additional note, every guitar performer played both pieces one octave lower than the other instruments, even though the notation was in the same register for every instrument. No performers on any other instrument played any octave transpositions.

The data were also analyzed by instrument and error type to determine whether there was an interaction between the two factors. Because of the low counts overall and the presence of zero-valued cells, a value of 0.1 was added to each cell\(^1\) and the analyses were performed by using a bootstrap method for approximating probability (e.g. Efron & Tibshirani, 1993; Patefield, 1981). The bootstrap simulation was implemented within the open-source statistical environment R (Ihaka & Gentleman, 1996) by the chisq.test() function (a member of the “ctest” package), using $B = 50,000$ replicates. There was not a significant interaction between instrument and error type for Melody 1, $\chi^2 = 6.729$, $p = .2493$, or for Ständchen, $\chi^2 = 5.322$, $p = .4258$.

Although their relatively small numbers prohibit a true statistical analysis of the nature of the errors, an attempt was made to categorize the location and type of each error.

---

1. Different small values were tried in order to eliminate a denominator of zero; the results were fairly stable across values, so 0.1 was arbitrarily chosen as the final value to report.
Melody 1

For Melody 1, only one clarinet performer made any errors, which consisted of a single chromatic substitution in a scalar run. Five guitar performers made errors, across seven performances. One performer consistently misplayed the two eighth-notes in measures 1, 3, and 5, resulting in diatonic substitutions. Three performers had difficulty with the sequence of low notes in measure 4, resulting in substitutions or deletions. Also, three performers had difficulty with the sixteenth-note pattern in the last measure, resulting in substitutions or deletions. There were no additions in any of the guitar performances of Melody 1. In all, 2 of the 20 (10%) substitutions in guitar performances were chromatic (i.e. non-diatonic); the rest were diatonic. No piano performers made any errors on Melody 1. Five trumpet performers made errors, across 10 performances (i.e. all five performers made errors on both expressive and mechanical performances). Three performers made substitutions in the scalar runs across measures 2 and 3 and measures 6 and 7. Four performers had difficulty with the high note at the beginning of measure 6; most of these performances contained an addition prior to the note, one full step lower. In other words, the performers “undershot” the pitch jump by one scale degree, and then played the proper note. In all, 7 of the 34 (20.6%) substitutions and additions in trumpet performances were chromatic; the rest were diatonic.

Ständchen

For Ständchen, no clarinet performers made any errors. Three guitar performers made errors, across 4 performances. All of these errors occurred in one of the triplets, and
all of the substitutions and additions were diatonic. One piano performer made errors in Ständchen, on both the expressive and the mechanical performance. These errors consisted of 2 additions, one of which was chromatic, and 2 (chromatic) instances in which E-flat was played instead of E. Three trumpet performers made errors in Ständchen, across three performances. There were no deletions, and 2 of the 7 (28.6%) substitutions and additions of trumpet performances were chromatic; the rest were diatonic.

Timing Analyses

Three different measures of timing were available: IOI, duration, and beat difference. Beat difference is defined as the difference between the performed and notated (i.e. quantized) onsets for each note, and was analyzed in a 3-way analysis of variance (separately for each piece) with factors of performance instrument (clarinet, guitar, piano, or trumpet), expression type (expressive or mechanical), and beat location (the location of each note in the score). Because they have obvious confounds with the notated length of each note, IOI and duration were each analyzed in a 2-way analysis of variance with factors of performance instrument and expression type.

Melody 1

For Melody 1, there was a significant 3-way interaction of Instrument, Expression Type, and Location on Beat Difference, $F(126, 1130) = 2.065$, $p = .0000$. Figure 4.1 shows a graph of the interaction; it is fairly complicated, but it appears that there is a large timing shift at beat 11, which is a pitch jump of over an octave. The trumpet and clarinet
Figure 4.1: Melody 1 timing by instrument, expression type, and location
performances, in particular, slow down to prepare for the leap, then speed up on the next note to compensate, creating a spike in the timing curve. The curve is more pronounced in the expressive performance than in the mechanical. There are also significant 2-way interactions of Instrument and Location, $F(126, 1131) = 1.381, p = .0051$, shown in Figure 4.2, and of Expression Type and Location, $F(42, 1130) = 3.005, p = .0000$, shown in Figure 4.3. Figure 4.2 shows the large spike at beat 11 discussed earlier and also the relative speed with which the trumpet performances played the first half of the piece. Figure 4.3 shows the increase in timing variability for expressive performances, as expected. Finally, there are significant main effects of Location, $F(42, 1131) = 10.86, p = .0000$, and of

![Figure 4.2: Melody 1 timing by instrument and location](image-url)
Expression Type, $F(1, 23) = 4.517, p = .0445$. The main effect of Location simply reflects the fact that the beat difference changes systematically as a function of the location in the piece, and the main effect of Expression Type shows a higher average beat difference for expressive performances (0.0368) than for mechanical performances (0.0018).

There was a significant main effect of Instrument on IOI, $F(3, 25) = 4.823, p = .0088$. Table 4.2 shows the average IOIs by instrument, which indicate a slower tempo for

<table>
<thead>
<tr>
<th>Instrument</th>
<th>Clarinet</th>
<th>Guitar</th>
<th>Piano</th>
<th>Trumpet</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>268.9</td>
<td>365.8</td>
<td>313.9</td>
<td>274.3</td>
</tr>
</tbody>
</table>

Table 4.2: Melody 1 IOIs by instrument
guitar and piano than for clarinet and trumpet. There was also a significant main effect of Expression Type on IOI, $F(1, 27) = 5.513, p = .0265$. Expressive performances had an average IOI of 298.6 ms, while mechanical performances had an average IOI of 308.9 ms, indicating that performers played slower overall in mechanical performances. There was not a significant interaction.

There was a significant main effect of Instrument on Duration, $F(3, 25) = 10.22, p = .0001$. Table 4.3 shows the average durations by instrument, which indicate a similar pattern as the IOIs in Table 3.2. There were no other significant effect on Duration.

<table>
<thead>
<tr>
<th>Clarinet</th>
<th>Guitar</th>
<th>Piano</th>
<th>Trumpet</th>
</tr>
</thead>
<tbody>
<tr>
<td>177.0</td>
<td>287.5</td>
<td>241.9</td>
<td>153.4</td>
</tr>
</tbody>
</table>

Table 4.3: Melody 1 durations by instrument

Ständchen

For Ständchen, there was a significant 2-way interaction of Instrument and Location on Beat Difference, $F(93, 866) = 1.876, p = .0000$, shown in Figure 4.4. The figure shows a similar timing pattern for clarinet and piano, but a difference over the second half of the piece by trumpet performances, and a large difference between the first and second halves.
of guitar performances. Ständchen has a repeating structure, and it appears that the differences are due to the way in which performers approached the second (repeated) half in relation to the first. There were no other significant interactions. There were significant main effects of Location, $F(31, 866) = 8.952$, $p = .0000$, Instrument, $F(3, 27) = 4.045$, $p = .0170$, and Expression Type, $F(1, 27) = 4.233$, $p = .0494$. The main effect of Location, as in Melody 1, reflects the change in beat difference across the piece. The main effect of Instrument shows that average beat differences of clarinet and piano are similar to each other but different from those of guitar and trumpet; averages are given in Table 4.4. The main effect of Expression Type shows that expressive performances tended to be played slightly ahead of the beat (0.0231), while mechanical performances were played slightly behind the beat (-0.0325).

Figure 4.4: Ständchen timing by instrument and location
There was a significant main effect of Expression Type on IOI, $F(1, 28) = 11.41, p = .0022$. Expressive performances had an average IOI of 543.1 ms, while mechanical performances had an average IOI of 519.2 ms, indicating that performers played faster overall in mechanical performances. There were no other significant effects on IOI.

There was a significant main effect of Expression Type on Duration, $F(1, 28) = 15.90, p = .0004$. Expressive performances had an average duration of 457.8 ms, while mechanical performances had an average duration of 427.7 ms, indicating that performers played shorter notes overall in mechanical performances. There were no other significant effect on Duration.

### Articulation Analyses

Duration scores, by themselves, are obviously related to the notated length of each note in the score. However, the duration of a note as a percentage of its IOI provides a measure of the degree to which a performer is playing relatively staccato or legato. Therefore, the data was transformed to create an additional dependent variable for articulation, defined as the ratio of Duration to IOI. The resulting Articulation scores are thus propor-

<table>
<thead>
<tr>
<th>Instrument</th>
<th>Articulation Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clarinet</td>
<td>0.0407</td>
</tr>
<tr>
<td>Guitar</td>
<td>-0.0719</td>
</tr>
<tr>
<td>Piano</td>
<td>0.0337</td>
</tr>
<tr>
<td>Trumpet</td>
<td>-0.0217</td>
</tr>
</tbody>
</table>

Table 4.4: Ständchen timing by instrument
tions with values falling between 0 and 1; values closer to 0 are more staccato, and values closer to 1 are more legato\(^2\). Articulation was analyzed for each piece in a 3-way analysis of variance with factors of performance instrument (clarinet, guitar, piano, or trumpet), expression type (expressive or mechanical), and beat location (the location of each note in the score.

**Melody 1**

For Melody 1, there was a significant 2-way interaction of Instrument and Location on Articulation, \(F(123, 1104) = 1.271, p = .0304\), shown in Figure 4.5. The performers

---

2. Theoretically, there could be articulation values greater than 1: if the offset of a note occurs later than the onset of the next note, then \(\text{Art} = \frac{\text{Dur}}{\text{IOI}} = \frac{(\text{Offset1}-\text{Onset1})}{(\text{Onset2}-\text{Onset1})} > 1\). However, this situation did not occur in any of the performances in this study.
appear to emphasize notes by playing them longer, particularly at the large pitch jumps at the beginnings of measures 2 and 6, and the spikes appear to be more pronounced for trumpet performances. Figure 4.5 also displays significant main effects of Location, $F(41, 1104) = 4.323$, $p = .0000$, and of Instrument, $F(3, 23) = 11.08$, $p = .0001$. Performances on clarinet and trumpet were overall more staccato than on guitar or piano; averages for each instrument are shown in Table 4.5. There were no other significant effect on Articulation.

<table>
<thead>
<tr>
<th>Clarinet</th>
<th>Guitar</th>
<th>Piano</th>
<th>Trumpet</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5775300</td>
<td>0.7003390</td>
<td>0.6720709</td>
<td>0.4962353</td>
</tr>
</tbody>
</table>

Table 4.5: Melody 1 articulation by instrument

**Ständchen**

For Ständchen, there was a significant 2-way interaction of Instrument and Location on Articulation, $F(90, 838) = 1.537$, $p = .0016$, shown in Figure 4.6. The figure shows a consistent lengthening of the quarter and dotted-quarter notes throughout the piece, and is more pronounced for the clarinet and trumpet performances. Figure 4.6 also shows the significant main effect of Location, $F(30, 838) = 22.58$, $p = .0000$. In addition, there was a significant main effect of Expression Type, $F(1, 27) = 4.797$, $p = .0373$. Mechanical per-
Performances were overall more staccato (.6882) than were expressive performances (.7047). There were no other significant effect on Articulation.

Dynamics Analyses

MIDI velocity was used as a measurement of the intensity each note, and therefore as a pattern of dynamics through a piece of music. Velocity was analyzed in a 3-way analysis of variance (separately for each piece) with factors of performance instrument (clarinet, guitar, piano, or trumpet), expression type (expressive or mechanical), and beat location (the location of each note in the score). Values of velocity were obtained by a process of MIDI transcription from the acoustic recordings (as described in Chapter 2); the experimenter attempted to keep recording levels as consistent as possible across

Figure 4.6: Ständchen articulation by instrument and location
recording sessions, but deviations in a variety of factors could affect the resulting velocities. It is reasonable to assume that recording level discrepancies across performances are random, so that aggregation of performers by instrument provides credible results. However, differences in the attack and sustain acoustic properties of different instruments could produce systematic differences in calculated velocities as a function of instrument. Therefore, the actual values for velocity may not always be directly comparable across performers or instruments. Regardless, recording levels were fixed once the recording began, so variability of velocity (i.e. through the piece or across expression type) should be comparable.

**Melody 1**

For Melody 1, there was a significant 3-way interaction of Instrument, Expression Type, and Location on Velocity, $F(126, 1130) = 1.308, p = .0168$. Figure 4.7 shows a graph of the interaction. Although very difficult to interpret, the most obvious difference seems to be between expressive and mechanical guitar performances; expressive guitar performances appear to show both greater variability across beat location and overall higher levels, more so than differences across other instruments. There are also significant 2-way interactions of Instrument and Location, $F(126, 1131) = 2.329, p = .0000$, shown in Figure 4.8, and of Expression Type and Location, $F(42, 1130) = 1.900, p = .0005$, shown in Figure 4.9. Figure 4.8 shows a complicated pattern that perhaps defies a straightforward interpretation. It is apparent, however, that velocity peaks and dips appear in all instruments at some locations (e.g. peaks at beats 8, 11, and 13.25; dips at beats 4.5, 7.5, 9,
Figure 4.7: Melody 1 velocities by instrument, expression type, and location
Figure 4.8: Melody 1 velocities by instrument and location

Figure 4.9: Melody 1 velocities by expression type and location
10.5 and 12), with greater variability across instruments at other locations. Two aspects stand out in Figure 4.9: the higher peak for expressive performances at beat 11 and the lower values for expressive performances at the end. These two factors combine to create, in expressive performances, an overall decreasing pattern from the dramatic pitch leap at beat 11 through the end of the piece. By comparison, the dynamics of the mechanical piece remain relatively flat. Finally, there are significant main effects of Location, $F(42, 1131) = 2.844$, $p = .0000$, and of Instrument, $F(3, 23) = 3.603$, $p = .0287$. The main effect of Location, similar to the timing analyses, reflects the change in dynamics throughout the piece. Averages for the main effect of Instrument are given in Table 4.6. They show a

<table>
<thead>
<tr>
<th>Instrument</th>
<th>Clarinet</th>
<th>Guitar</th>
<th>Piano</th>
<th>Trumpet</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>40.76890</td>
<td>25.21667</td>
<td>47.33140</td>
<td>54.68513</td>
</tr>
</tbody>
</table>

Table 4.6: Melody 1 velocities by instrument

lower average velocity for guitar performances, although this should be interpreted with appropriate caution, as mentioned above.

Ständchen

For Ständchen, there was a significant 2-way interaction of Instrument and Location on Velocity, $F(93, 866) = 1.791$, $p = .0000$, shown in Figure 4.10. The clarinet and trun-
pet performances seem to show a similar pattern, with both increasing noticeably over the second half of the piece. The guitar and piano performances are each more variable, and have more peaks and dips than do clarinet and trumpet. There is also a significant 2-way interaction of Expression Type and Location, $F(31, 866) = 2.055, p = .0007$, shown in Figure 4.11. This figure mostly shows an amplified dynamics for expressive performance, plus a decreasing pattern over the second half similar to that in Melody 1. There were no other significant interactions. There was also a significant main effect of Location, $F(31, 866) = 3.410, p = .0000$, reflecting the change in dynamics across the piece.
Discussion

The pattern of results shows that performances were affected both by the instrument and by the performer’s expressive intent. Furthermore, the two factors tended to interact, such that expressive content was delivered differently by different instruments. In all cases, the profiles of mechanical performances retained the same structural elements as the expressive performances; expressive deviations were merely attenuated, not eliminated.

It is interesting that every guitar performer played both pieces one octave lower than performers on the other instruments. At least part of this may be attributable to notational convention, but the same register was chosen even by the guitar performers who essen-

Figure 4.11: Ständchen velocities by expression type and location
tially never perform from standard music notation, and who therefore wouldn’t necessarily have acquired such conventions. The register chosen by every guitar performance results in maximizing the number of open strings that can be used to play the melody.

The two musical pieces differ considerably from each other, which was reflected in the performances. The challenges of its pitch intervals and scalar runs resulted in a higher error rate for Melody 1, particularly for guitar and trumpet. The four instruments displayed different rates of errors overall, and in fact the differences may be greater than what is recorded. Only final versions of performances were analyzed, and several of the performers who made errors in a final recorded version had already performed some unused “scratch tracks” of the same piece. Typically, these earlier performances were discarded due to their error rates, so the final versions contained errors which had been reduced to an acceptable level.

When errors do occur, they may have different causes and locations, depending upon the performance instrument. The prevalence of errors in guitar performances may in general be attributable to the relative difficulty of sight-reading, and the mistakes of the performer who consistently misplayed the pair of eighth-notes in Melody 1 may simply be due to an error in reading. The mistakes in the low notes of measure 4 in Melody 1 could be due to a change in hand position as the performer moved from the higher notes of measure 3, which were played on a different set of strings. In comparison, the errors by trumpet performers tended to occur in scalar runs and in leaps to relatively high notes. A large upward pitch leap places demands on wind instruments that simply don’t exist for stringed instruments; a change in hand position is challenging for a guitarist, but the direction of
the change (up or down) is largely irrelevant. By the same logic, one would expect to find
errors in clarinet performances similar to those in trumpet performances, but the error
rates for clarinet were very low. This may be a function of the range of the piece fitting the
clarinet more easily than the trumpet.

The characteristics of the two pieces also influence the patterns of timing, articula-
tion, and dynamics. Effects for performances of Melody 1 tend to be dominated in partic-
ular by the large pitch interval at the beginning of measure 6. The physical demands on
trumpet performers to reach the high note in measure 6 are reflected in the profiles of
errors, timing, articulation, and dynamics. There appears to be an metrical pattern of
articulation within the piece, but the effect is fairly different for each instrument. By con-
trast, the effects for Ständchen tend to be more global in nature. For example, many can
be interpreted as variations across the two halves of the piece. The pattern of articulation
in Ständchen is relatively stable across instruments, and varies only as a matter of degree.
Also, Melody 1 had significant main effects of Instrument on IOI, Duration, Articulation,
and Velocity, showing the differences in interpretation, and had significant main effects of
Expression Type only on Beat Difference and IOI. Ständchen, on the other hand, had
main effects of Expression Type on Beat Difference, IOI, Duration, and Articulation,
showing a difference in degree of expressive interpretation, and had a significant main
effect of Instrument only on Beat Difference. This supports the idea that Ständchen has a
more predictable expressive interpretation, whereas Melody 1 is more neutral and there-
fore more subject to instrumental and individual differences.
Summary

Performances for each composition were analyzed with respect to measures of errors, timing, articulation, and dynamics. Overall, the pattern of results was consistent with the interpretation that Melody 1 is a more technically challenging piece, while Ständchen is more naturally expressive; performers made more note errors on Melody 1 and had more complex interactions and main effects of instrument, while Ständchen showed more main effects of expression type and had a more stable expressive profile. For both pieces, however, the use of instrument appeared to be a significant factor for performance variability. Every dependent variable, for both compositions, resulted in a significant main effect of and/or a significant interaction involving the performance instrument. This supports the general prediction that musical performance is a function of the instrument as well as more abstract expressive intentions of the performer.
CHAPTER 5

PERCEPTUAL DATA

This chapter describes a perceptual experiment using the performance data as stimuli. Participants were asked to identify the original performance instrument from timbrally-neutral MIDI versions of the original recordings. Some relevant background is presented as a context for the experiment, the methodology is detailed, and then the analyses and results are presented. Finally, some general implications of the results are discussed.

There is a wealth of information to study in musical phenomena. The focus of this project is largely on the quantitative analysis of expressive performance characteristics, but music is broadly defined and should be broadly analyzed. Fundamentally, music is performed for a listening audience, so aspects of a piece of music which are important to a performer should be conveyed to the listener. Therefore, the performances were presented as perceptual data to an additional group of participants.

The biggest question that this experiment addresses is the extent to which instrument-specific features of a performance are perceptually obvious. If the kinds of differences that were identified in Chapter 4 are important attributes of musical performance,
then they should be recognized as such by listeners. Furthermore, if there are defining properties of particular instruments (e.g. as profiles of timing, articulation, or dynamics), then listeners should be able to differentiate between instruments even in the absence of timbral cues. The perception experiment presents participants with synthesized versions of the MIDI files, all with the same neutral timbre, and asks them to identify what physical instrument created each performance.

An obvious factor which affects the design of music perception or performance studies is the difference between musicians and non-musicians. Participants in conventional performance studies, by definition, need to be musicians, but participants in perception studies may or may not have formal musical experience. Musicians and non-musicians, naturally, approach musical events differently and will perform differently in many perceptual tasks (e.g. Kendall & Carterette, 1990, Palmer, 1996b). The difference may be more than a differential in sensitivity along a continuum; musicians and non-musicians may behave qualitatively differently in perceptual tasks because they conceive of the domain in a different way and with different terminology, and are therefore actually engaging in different tasks (Babbit, 1958; Cook, 1994). This is a particularly salient issue for studies which employ only musicians, for they need to be careful about generalizations to the listening public at large. With this in mind, the participants for the perceptual experiment were not constrained to have any particular background in music. They were asked whether they played an instrument and how much, if any, formal training they had. The goal was to examine general reactions to the experimental performances, and additionally to consider how participants’ musical backgrounds may have influenced their responses.
Previous work (Walker, 1996) has shown that listeners are sensitive to particular cues of the physical performance instrument; participants were able to identify the original instrument correctly at a better-than-chance level from MIDI performance files played with a consistent or inconsistent timbre. Additionally, average goodness ratings of performances were higher when performance files were played in a timbre consistent with the original instrument, indicating that listeners were sensitive to pairings of timbral and non-timbral cues.

These results suggest that structural, as well as acoustic, factors can be important in communication of expression (Kendall & Carterette, 1990; Repp, 1992). Therefore, listeners may be able to recognize characteristics of the physical instrument which produced a performance, despite hearing the performance with an inconsistent timbre. However, perceptual identification of instrument-specific performance differences may be limited to a categorical discrimination of types, as opposed to a veridical ability to explicitly identify the instrument. Furthermore, the categories employed may be unique to each individual, so that responses grouped across listeners might obfuscate systematic variability in the data.

Perception Experiment

In this experiment, participants will be asked to identify the physical instrument on which the performance was originally produced, regardless of the sounded timbre (similar to Walker, 1996). The MIDI transcriptions from the performance experiment were presented to listeners in this experiment, all with the same neutral timbre. All performances,
both expressive and mechanical, were included. All notes retained their original pitches, timing, and velocities, but otherwise sounded identical to each other. Participants were asked to identify, for each MIDI performance, the original performance instrument. Additionally, they were asked to subjectively rate how good each performance was in relation to the rest. Both musically trained and untrained listeners were included, similar to Kendall & Carterette’s (1990) study. However, this study has the advantage of including different performances of the same piece on each instrument.

Method

Participants

Participants were 48 undergraduate students at Ohio Dominican University in Columbus, OH who received extra credit in a class for participating in the study. They were randomly assigned to one of two groups, to determine which melody they would hear. There was no requirement of musical experience, but participants were asked about their music background in a questionnaire. Of the 48 participants, 28 reported having played an instrument for at least one year, either currently or in the past; 15 of those 28 had played for at least 5 years. The participants ranged in age between 18 and 44 (mean=22.1). The length of experience of playing an instrument (for the 28 participants who had any experience) ranged from 1.5 to 22 years (mean=7.1).
Materials

All 64 MIDI performance files (8 performers x 4 instruments x 2 levels of expression) of each piece of music were played through a software synthesizer\(^1\) and saved as digital audio files. The timbre for each synthesized file was the same flute sound, which was chosen to be as neutral as possible with respect to the original four instruments (clarinet, guitar, piano, and trumpet) while still sounding reasonably like a real instrument. A quantized computer-generated version of each musical piece was also created in order to serve as a practice track. Within each musical piece, a random order was created of the 64 tracks, and then three permutations were made from the original random order, resulting in four different orders for each piece. Several audio compact discs were then burned for each order, with 65 tracks each. The first track was always a quantized performance (either of Melody 1 or Ständchen), and the following 64 tracks were the appropriate ordering of the synthesized performances of the same piece. There were two minutes of silence following quantized performance in the first track, and 15 seconds of silence following each performance in the following tracks.

A master copy of a spreadsheet document was created for participant responses. It contained two response columns, each with pop-up response boxes with data validation so that participants would not have to type anything for their responses. The first response column contained, for each row, possible values of “Clarinet”, “Guitar”, “Piano”, and “Trumpet”. The second response column contained values 1 through 7, and also produced

\(^1\) The timbre was the “Pan Flute” from Emagic’s exsp-24 sound set.
a reminder box which read “Respond on a 1-7 scale, where 1=’very poor’ and 7=’very good’.” The spreadsheet was copied for each participant to serve as a response form.

Procedure

Each participant filled out a brief questionnaire which asked about their demographic information and musical background, read a form describing the nature of the experiment, and signed to indicate voluntary consent. The participant was then given a sheet of instructions to read, and the experimenter answered any questions about the task.

Each participant was seated in front of a computer and was given a pair of headphones to wear. A CD containing one of the orderings of either Melody 1 or Ständchen was placed in the CD-ROM drive of the computer and the CD player application was opened. Participants were instructed that they would hear the same melody many times, once per track on the CD, always with the same synthesized sound. They were told that they should listen to each melody and try to determine whether it had been originally been performed by clarinet, guitar, piano, or trumpet, and also to rate how good of a performance they thought it was. Participants were instructed how to use the spreadsheet to make their responses, and were told to listen to the first (quantized) track and to make their responses. They were given time to ask any questions they might have had and to ensure that both the task and the response sheets were clear. When they were ready, participants were told to listen to the rest of the tracks on the CD and to make responses for each track. Participants made two responses for each trial: 1) what physical instrument they thought originally produced the performance (clarinet, guitar, piano, or trumpet); and 2) overall,
how good they thought each performance was (on a 1-7 scale, where 1 = very poor and 7 = very good).

After all the trials were completed, participants were given a debriefing sheet which explained the goals of the study, asked what characteristics of the performances they thought most influenced their responses, and provided a space for general comments. The responses were exported to a delimited text file for analysis.

Results

Melody 1

Overall, participants were not successful in identifying the performance instrument for Melody 1 performances. They were correct in 24.9% of the trials, essentially the same as chance (25%). However, accurate response is not necessarily the most appropriate metric for the task; participants might be able to distinguish between performances but not be able to place them in the correct instrument category. Furthermore, the basis of categorization may not be the same for all participants.

A 2-way chi-squared analysis was performed for values of the correct instrument and the responded instrument, but did not yield significant results, $X^2(9) = 16.03$, $p = .0663$. One pattern which did emerge was an overall bias towards responding “Clarinet”; comprising 486 of the 1536 responses (31.6%). This is probably due to the fact that the synthesized flute sound was closer to a clarinet’s timbre than to any of the other instruments. Separate chi-squared analyses were performed for responses to the two intended
levels of expression, but did not yield significant results for either the mechanical ($\chi^2(9) = 12.96, p = .1643$) or the expressive ($\chi^2(9) = 7.464, p = .5889$) performances. Since differences by performer may be important for identification of instrument, an additional chi-squared analysis was performed for performer and responded instrument, showing a significant effect, $\chi^2(93) = 159.9, p = .0000$. The 2-way table of response counts is shown in Table 5.1. There is no obvious simple pattern, but it is apparent that for some performers (e.g. t5) responses are relatively uniform, while for others (e.g. p3) there is some consistency in response.

At least in aggregation, participants did not seem to consistently respond to the stimuli on the basis of the original performance instrument. Therefore, the responses were analyzed with respect to global metrics of each performance, in order to determine the factors contributing to categorization. The average IOI of each performance (which serves as a measure of global tempo) was treated as a response variable in a one-way analysis of variance, with participants’ responses treated as factor levels. A significant effect was obtained, $F(3, 65) = 4.574, p = .0057$, showing a differential response to tempo. Mean IOIs per response category are given in Table 5.2, and show a tendency for faster performances to be labelled as “Piano,” while slower performances are labelled as “Trumpet.” A similar analysis was performed on total errors, also with a significant effect, $F(3, 65) = 3.135, p = .0314$. The responses, given in Table 5.3, show that performances with the lowest number of errors tended to be labelled as “Piano,” and those with the highest number of errors tended to be labelled as “Guitar.” There were not significant effects for measures
<table>
<thead>
<tr>
<th>Performer</th>
<th>Clarinet</th>
<th>Guitar</th>
<th>Piano</th>
<th>Trumpet</th>
</tr>
</thead>
<tbody>
<tr>
<td>c1</td>
<td>13</td>
<td>12</td>
<td>7</td>
<td>16</td>
</tr>
<tr>
<td>c2</td>
<td>13</td>
<td>11</td>
<td>17</td>
<td>7</td>
</tr>
<tr>
<td>c3</td>
<td>14</td>
<td>5</td>
<td>19</td>
<td>10</td>
</tr>
<tr>
<td>c4</td>
<td>23</td>
<td>8</td>
<td>10</td>
<td>7</td>
</tr>
<tr>
<td>c5</td>
<td>18</td>
<td>12</td>
<td>11</td>
<td>7</td>
</tr>
<tr>
<td>c6</td>
<td>16</td>
<td>8</td>
<td>14</td>
<td>10</td>
</tr>
<tr>
<td>c7</td>
<td>14</td>
<td>9</td>
<td>11</td>
<td>14</td>
</tr>
<tr>
<td>c8</td>
<td>17</td>
<td>11</td>
<td>9</td>
<td>11</td>
</tr>
<tr>
<td>g1</td>
<td>15</td>
<td>9</td>
<td>10</td>
<td>14</td>
</tr>
<tr>
<td>g2</td>
<td>19</td>
<td>8</td>
<td>6</td>
<td>15</td>
</tr>
<tr>
<td>g3</td>
<td>21</td>
<td>8</td>
<td>10</td>
<td>9</td>
</tr>
<tr>
<td>g4</td>
<td>11</td>
<td>10</td>
<td>4</td>
<td>23</td>
</tr>
<tr>
<td>g5</td>
<td>16</td>
<td>4</td>
<td>15</td>
<td>13</td>
</tr>
<tr>
<td>g6</td>
<td>22</td>
<td>6</td>
<td>9</td>
<td>11</td>
</tr>
<tr>
<td>g7</td>
<td>18</td>
<td>7</td>
<td>10</td>
<td>13</td>
</tr>
<tr>
<td>g8</td>
<td>10</td>
<td>12</td>
<td>11</td>
<td>15</td>
</tr>
<tr>
<td>p1</td>
<td>10</td>
<td>9</td>
<td>20</td>
<td>9</td>
</tr>
<tr>
<td>p2</td>
<td>16</td>
<td>11</td>
<td>13</td>
<td>8</td>
</tr>
<tr>
<td>p3</td>
<td>19</td>
<td>1</td>
<td>11</td>
<td>17</td>
</tr>
<tr>
<td>p4</td>
<td>10</td>
<td>10</td>
<td>7</td>
<td>21</td>
</tr>
<tr>
<td>p5</td>
<td>15</td>
<td>1</td>
<td>15</td>
<td>17</td>
</tr>
<tr>
<td>p6</td>
<td>13</td>
<td>9</td>
<td>7</td>
<td>19</td>
</tr>
<tr>
<td>p7</td>
<td>10</td>
<td>15</td>
<td>9</td>
<td>14</td>
</tr>
<tr>
<td>p8</td>
<td>16</td>
<td>10</td>
<td>10</td>
<td>12</td>
</tr>
<tr>
<td>t1</td>
<td>20</td>
<td>9</td>
<td>9</td>
<td>10</td>
</tr>
<tr>
<td>t2</td>
<td>13</td>
<td>9</td>
<td>3</td>
<td>23</td>
</tr>
<tr>
<td>t3</td>
<td>12</td>
<td>15</td>
<td>8</td>
<td>13</td>
</tr>
<tr>
<td>t4</td>
<td>15</td>
<td>11</td>
<td>16</td>
<td>6</td>
</tr>
<tr>
<td>t5</td>
<td>15</td>
<td>10</td>
<td>11</td>
<td>12</td>
</tr>
<tr>
<td>t6</td>
<td>16</td>
<td>11</td>
<td>7</td>
<td>14</td>
</tr>
<tr>
<td>t7</td>
<td>11</td>
<td>11</td>
<td>16</td>
<td>10</td>
</tr>
<tr>
<td>t8</td>
<td>15</td>
<td>5</td>
<td>17</td>
<td>11</td>
</tr>
</tbody>
</table>

Table 5.1: Melody 1 number of responses of each instrument type by performer
Because categorization by participants could be different from one another, aggregation across all participants may obscure some meaningful effects. Therefore, chi-squared analyses for values of the correct instrument and the responded instrument were performed separately for each participant. These analyses resulted in significant results (at the $\alpha=.05$ level) for 5 of the 24 participants who listened to Melody 1. This may be noteworthy, because all 5 of these participants played instruments, 4 of them for more than five years. Of all of the participants who listened to Melody 1, 12 played instruments, 7 for

more than five years. This means that significant effects were found in 41.7% of the participants with performance experience, 57.1% of those with more than five years of experience, and for the three participants with the most experience. By contrast, no significant effects were seen for any of the 12 participants without performance experience.

Goodness ratings for Melody 1 were analyzed in a 2-way analysis of variance with factors of Instrument and Expression Type. There was a significant main effect of Instrument, $F(3, 69) = 14.01, p = .0000$. The average ratings, shown in Table 5.4, show the highest scores for piano performances and the lowest for trumpet. There were no other significant effects. An additional 2-way analysis of variance was performed with factors of Performer and Expression Type. There was a significant 2-way interaction between Performer and Expression Type, $F(31, 713) = 2.500, p = .0000$. Table 5.5 shows the averages by both factors, and indicates that participants preferred expressive performances better for some performers, but mechanical performances better for others. There was

<table>
<thead>
<tr>
<th></th>
<th>Clarinet</th>
<th>Guitar</th>
<th>Piano</th>
<th>Trumpet</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>4.385</td>
<td>4.102</td>
<td>4.646</td>
<td>3.784</td>
</tr>
</tbody>
</table>

Table 5.4: Melody 1 goodness ratings by instrument
<table>
<thead>
<tr>
<th>Performer</th>
<th>Mechanical</th>
<th>Expressive</th>
</tr>
</thead>
<tbody>
<tr>
<td>c1</td>
<td>3.667</td>
<td>3.917</td>
</tr>
<tr>
<td>c2</td>
<td>4.500</td>
<td>4.458</td>
</tr>
<tr>
<td>c3</td>
<td>5.000</td>
<td>5.083</td>
</tr>
<tr>
<td>c4</td>
<td>4.667</td>
<td>3.958</td>
</tr>
<tr>
<td>c5</td>
<td>4.542</td>
<td>4.125</td>
</tr>
<tr>
<td>c6</td>
<td>4.917</td>
<td>4.750</td>
</tr>
<tr>
<td>c7</td>
<td>4.292</td>
<td>4.208</td>
</tr>
<tr>
<td>c8</td>
<td>3.958</td>
<td>4.125</td>
</tr>
<tr>
<td>g1</td>
<td>3.917</td>
<td>4.333</td>
</tr>
<tr>
<td>g2</td>
<td>4.375</td>
<td>3.292</td>
</tr>
<tr>
<td>g3</td>
<td>4.083</td>
<td>3.875</td>
</tr>
<tr>
<td>g4</td>
<td>3.083</td>
<td>3.833</td>
</tr>
<tr>
<td>g5</td>
<td>4.625</td>
<td>4.375</td>
</tr>
<tr>
<td>g6</td>
<td>4.833</td>
<td>4.917</td>
</tr>
<tr>
<td>g7</td>
<td>4.958</td>
<td>4.042</td>
</tr>
<tr>
<td>g8</td>
<td>3.292</td>
<td>3.792</td>
</tr>
<tr>
<td>p1</td>
<td>4.875</td>
<td>4.708</td>
</tr>
<tr>
<td>p2</td>
<td>4.458</td>
<td>4.458</td>
</tr>
<tr>
<td>p3</td>
<td>4.792</td>
<td>4.708</td>
</tr>
<tr>
<td>p4</td>
<td>4.125</td>
<td>4.458</td>
</tr>
<tr>
<td>p5</td>
<td>4.583</td>
<td>4.750</td>
</tr>
<tr>
<td>p6</td>
<td>5.083</td>
<td>4.833</td>
</tr>
<tr>
<td>p7</td>
<td>5.208</td>
<td>4.458</td>
</tr>
<tr>
<td>p8</td>
<td>4.417</td>
<td>4.417</td>
</tr>
<tr>
<td>t1</td>
<td>4.292</td>
<td>4.500</td>
</tr>
<tr>
<td>t2</td>
<td>3.208</td>
<td>3.958</td>
</tr>
<tr>
<td>t3</td>
<td>2.792</td>
<td>2.917</td>
</tr>
<tr>
<td>t4</td>
<td>2.917</td>
<td>3.667</td>
</tr>
<tr>
<td>t5</td>
<td>3.292</td>
<td>4.292</td>
</tr>
<tr>
<td>t6</td>
<td>4.417</td>
<td>3.042</td>
</tr>
<tr>
<td>t7</td>
<td>3.750</td>
<td>4.167</td>
</tr>
<tr>
<td>t8</td>
<td>4.750</td>
<td>4.583</td>
</tr>
</tbody>
</table>

Table 5.5: Melody 1 goodness ratings by performer and expression type
also a significant main effect of Performer, $F(31, 713) = 7.479, p = .0000$, showing the differential preference for certain performers. There were no other significant effects.

Ständchen

As with Melody 1, participants were not successful in identifying the performance instrument for Ständchen performances. They were correct in 25.3% of the trials, compared to a chance rate of 25%. A 2-way chi-squared analysis was performed for values of the correct instrument and the responded instrument, aggregated across all participants, and showed a significant relationship, $X^2(9) = 21.92, p = .0092$. Table 5.6 shows the 2-way table of counts. The bias towards responding “Clarinet” is present in Ständchen as well, comprising 449 of the 1536 responses (29.2%). However, the bias is less prevalent, particularly for trumpet performances.

<table>
<thead>
<tr>
<th>Instrument</th>
<th>Clarinet</th>
<th>Guitar</th>
<th>Piano</th>
<th>Trumpet</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clarinet</td>
<td>127</td>
<td>76</td>
<td>89</td>
<td>92</td>
</tr>
<tr>
<td>Guitar</td>
<td>120</td>
<td>66</td>
<td>89</td>
<td>109</td>
</tr>
<tr>
<td>Piano</td>
<td>116</td>
<td>65</td>
<td>100</td>
<td>103</td>
</tr>
<tr>
<td>Trumpet</td>
<td>86</td>
<td>99</td>
<td>103</td>
<td>96</td>
</tr>
</tbody>
</table>

Table 5.6: Ständchen responses by performance instrument
A chi-squared analysis for responses to only mechanical performances showed a significant relationship, $X^2(9) = 32.17, p = .0002$. The data are shown in Table 5.7. Participants seem to be distinguishing between wind and non-wind instruments, but making incorrect categorizations; responses of “Guitar” and “Piano” are lower for the actual performances on guitar and piano, but higher for those on clarinet and trumpet. A chi-squared analysis for responses to only expressive performances did not reveal a significant relationship, $X^2(9) = 12.88, p = .1680$. The chi-squared analysis for performer and responded instrument showed a significant effect, $X^2(93) = 153.8, p = .0001$. The 2-way table of response counts is shown in Table 5.8, and again displays a fair degree of variability across performers.

<table>
<thead>
<tr>
<th>Instrument</th>
<th>Clarinet</th>
<th>Guitar</th>
<th>Piano</th>
<th>Trumpet</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clarinet</td>
<td>66</td>
<td>35</td>
<td>51</td>
<td>40</td>
</tr>
<tr>
<td>Guitar</td>
<td>61</td>
<td>25</td>
<td>40</td>
<td>66</td>
</tr>
<tr>
<td>Piano</td>
<td>53</td>
<td>37</td>
<td>50</td>
<td>52</td>
</tr>
<tr>
<td>Trumpet</td>
<td>44</td>
<td>59</td>
<td>50</td>
<td>39</td>
</tr>
</tbody>
</table>

Table 5.7: Ständchen responses by instrument - mechanical performances only
<table>
<thead>
<tr>
<th>Performer</th>
<th>Clarinet</th>
<th>Guitar</th>
<th>Piano</th>
<th>Trumpet</th>
</tr>
</thead>
<tbody>
<tr>
<td>c1</td>
<td>16</td>
<td>10</td>
<td>7</td>
<td>15</td>
</tr>
<tr>
<td>c2</td>
<td>14</td>
<td>10</td>
<td>19</td>
<td>5</td>
</tr>
<tr>
<td>c3</td>
<td>14</td>
<td>9</td>
<td>13</td>
<td>12</td>
</tr>
<tr>
<td>c4</td>
<td>14</td>
<td>6</td>
<td>17</td>
<td>11</td>
</tr>
<tr>
<td>c5</td>
<td>22</td>
<td>10</td>
<td>7</td>
<td>9</td>
</tr>
<tr>
<td>c6</td>
<td>19</td>
<td>10</td>
<td>9</td>
<td>10</td>
</tr>
<tr>
<td>c7</td>
<td>16</td>
<td>7</td>
<td>8</td>
<td>17</td>
</tr>
<tr>
<td>c8</td>
<td>12</td>
<td>14</td>
<td>9</td>
<td>13</td>
</tr>
<tr>
<td>g1</td>
<td>20</td>
<td>9</td>
<td>9</td>
<td>10</td>
</tr>
<tr>
<td>g2</td>
<td>14</td>
<td>4</td>
<td>17</td>
<td>13</td>
</tr>
<tr>
<td>g3</td>
<td>14</td>
<td>12</td>
<td>9</td>
<td>13</td>
</tr>
<tr>
<td>g4</td>
<td>9</td>
<td>11</td>
<td>14</td>
<td>14</td>
</tr>
<tr>
<td>g5</td>
<td>19</td>
<td>11</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>g6</td>
<td>13</td>
<td>6</td>
<td>15</td>
<td>14</td>
</tr>
<tr>
<td>g7</td>
<td>17</td>
<td>9</td>
<td>9</td>
<td>13</td>
</tr>
<tr>
<td>g8</td>
<td>14</td>
<td>4</td>
<td>7</td>
<td>23</td>
</tr>
<tr>
<td>p1</td>
<td>9</td>
<td>11</td>
<td>18</td>
<td>10</td>
</tr>
<tr>
<td>p2</td>
<td>12</td>
<td>7</td>
<td>13</td>
<td>16</td>
</tr>
<tr>
<td>p3</td>
<td>13</td>
<td>10</td>
<td>12</td>
<td>13</td>
</tr>
<tr>
<td>p4</td>
<td>15</td>
<td>8</td>
<td>11</td>
<td>14</td>
</tr>
<tr>
<td>p5</td>
<td>15</td>
<td>7</td>
<td>8</td>
<td>18</td>
</tr>
<tr>
<td>p6</td>
<td>20</td>
<td>6</td>
<td>14</td>
<td>8</td>
</tr>
<tr>
<td>p7</td>
<td>12</td>
<td>8</td>
<td>16</td>
<td>12</td>
</tr>
<tr>
<td>p8</td>
<td>20</td>
<td>8</td>
<td>8</td>
<td>12</td>
</tr>
<tr>
<td>t1</td>
<td>12</td>
<td>5</td>
<td>9</td>
<td>22</td>
</tr>
<tr>
<td>t2</td>
<td>5</td>
<td>14</td>
<td>21</td>
<td>8</td>
</tr>
<tr>
<td>t3</td>
<td>17</td>
<td>14</td>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td>t4</td>
<td>11</td>
<td>11</td>
<td>14</td>
<td>12</td>
</tr>
<tr>
<td>t5</td>
<td>6</td>
<td>20</td>
<td>15</td>
<td>7</td>
</tr>
<tr>
<td>t6</td>
<td>9</td>
<td>15</td>
<td>14</td>
<td>10</td>
</tr>
<tr>
<td>t7</td>
<td>15</td>
<td>8</td>
<td>13</td>
<td>12</td>
</tr>
<tr>
<td>t8</td>
<td>11</td>
<td>12</td>
<td>9</td>
<td>16</td>
</tr>
</tbody>
</table>

Table 5.8: Ständchen number of responses of each instrument type by performer
Global metrics of the performances were analyzed for Ständchen in the same fashion as for Melody 1. A significant effect was obtained for average IOI, $F(3, 69) = 4.061, p = .0102$, and averages for each response are shown in Table 5.9. Responses of “Guitar” and

<table>
<thead>
<tr>
<th>Clarinet</th>
<th>Guitar</th>
<th>Piano</th>
<th>Trumpet</th>
</tr>
</thead>
<tbody>
<tr>
<td>550.4</td>
<td>527.7</td>
<td>532.5</td>
<td>574.0</td>
</tr>
</tbody>
</table>

Table 5.9: Ständchen average IOIs by response category

“Piano” were generally given for faster tempos, while “Clarinet” and “Trumpet” were given for slower tempos. A significant effect was also obtained for average MIDI velocity, $F(3, 69) = 4.587, p = .0055$. The averages for each response, shown in Table 5.10, show

<table>
<thead>
<tr>
<th>Clarinet</th>
<th>Guitar</th>
<th>Piano</th>
<th>Trumpet</th>
</tr>
</thead>
<tbody>
<tr>
<td>37.07</td>
<td>43.27</td>
<td>43.02</td>
<td>40.00</td>
</tr>
</tbody>
</table>

Table 5.10: Ständchen average velocities by response category
that louder performances tended to be labelled as either “Guitar” or “Piano,” while softer performances tended to be labelled as either “Clarinet” or “Trumpet.” There were not significant effects for measures of average articulation, $F(3, 69) = 2.134$, $p = .1039$, or of total errors, $F(3, 69) = 1.233$, $p = .3045$.

Chi-squared analyses were again performed separately for each participant. These analyses resulted in significant results (at the $\alpha=.05$ level) for 2 of the 24 participants who listened to Ständchen (additionally, the analysis for one participant resulted in $p=.0572$). Of the two significant analyses, one of the participants had performance experience, but the other did not.

Goodness ratings for Ständchen were analyzed in a 2-way analysis of variance with factors of Instrument and Expression Type. There was a significant main effect of Instrument, $F(3, 69) = 3.607$, $p = .0175$. The average ratings, shown in Table 5.11, show lower scores for trumpet than for other instruments. There was also a significant main effect of Expression Type, $F(1, 23) = 4.286$, $p = .0498$, showing higher average ratings for mechan-

<table>
<thead>
<tr>
<th>Instrument</th>
<th>Clarinet</th>
<th>Guitar</th>
<th>Piano</th>
<th>Trumpet</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>4.276</td>
<td>4.203</td>
<td>4.255</td>
<td>3.984</td>
</tr>
</tbody>
</table>

Table 5.11: Ständchen goodness ratings by instrument
ical (4.24) than for expressive performances (4.12). There was not a significant interaction. A 2-way analysis of variance was performed with factors of Performer and Expression Type. There was a significant 2-way interaction between Performer and Expression Type, $F(31, 713) = 2.045$, $p = .0008$. Table 5.12 shows the averages by both factors. There were also significant main effects of Performer, $F(31, 713) = 4.043$, $p = .0000$, and of Expression Type, $F(1, 23) = 4.286$, $p = .0498$.

Discussion

The task of identifying the performance instrument from a synthesized MIDI output is very difficult, and it is not surprising that participants were not able to perform accurately. Despite this, it does appear that there are some real effects in the data. The sets of results are somewhat different for Melody 1 and Ständchen, but both display evidence that participants were responding to instrument-specific patterns of event timing and intensity. For example, Ständchen had a significant chi-squared test aggregated over all listeners, indicating that they were responding to cues in order to categorize the performances, even if the categorizations were not veridical. For Melody 1 the aggregate chi-squared was not significant, but the set of individual chi-squared tests produced intriguing results. Musician listeners would be expected to be more sensitive to musical variability than non-musicians, particularly if they have specific experience on one of the performance instruments (e.g. Kendall & Carterette, 1990; Palmer, 1996b), and that is exactly the pattern that was discovered.
Table 5.12: Ständchen goodness ratings by performer and expression type

<table>
<thead>
<tr>
<th>Performer</th>
<th>Mechanical</th>
<th>Expressive</th>
</tr>
</thead>
<tbody>
<tr>
<td>c1</td>
<td>4.333</td>
<td>4.000</td>
</tr>
<tr>
<td>c2</td>
<td>4.750</td>
<td>4.417</td>
</tr>
<tr>
<td>c3</td>
<td>4.417</td>
<td>3.958</td>
</tr>
<tr>
<td>c4</td>
<td>4.292</td>
<td>4.000</td>
</tr>
<tr>
<td>c5</td>
<td>4.458</td>
<td>4.042</td>
</tr>
<tr>
<td>c6</td>
<td>4.708</td>
<td>4.458</td>
</tr>
<tr>
<td>c7</td>
<td>3.833</td>
<td>4.083</td>
</tr>
<tr>
<td>c8</td>
<td>4.250</td>
<td>4.417</td>
</tr>
<tr>
<td>g1</td>
<td>5.292</td>
<td>4.042</td>
</tr>
<tr>
<td>g2</td>
<td>3.958</td>
<td>4.375</td>
</tr>
<tr>
<td>g3</td>
<td>3.375</td>
<td>3.792</td>
</tr>
<tr>
<td>g4</td>
<td>3.750</td>
<td>3.667</td>
</tr>
<tr>
<td>g5</td>
<td>4.417</td>
<td>4.500</td>
</tr>
<tr>
<td>g6</td>
<td>4.875</td>
<td>4.125</td>
</tr>
<tr>
<td>g7</td>
<td>4.375</td>
<td>4.292</td>
</tr>
<tr>
<td>g8</td>
<td>4.667</td>
<td>3.750</td>
</tr>
<tr>
<td>p1</td>
<td>4.750</td>
<td>4.792</td>
</tr>
<tr>
<td>p2</td>
<td>5.000</td>
<td>4.333</td>
</tr>
<tr>
<td>p3</td>
<td>4.458</td>
<td>4.667</td>
</tr>
<tr>
<td>p4</td>
<td>3.917</td>
<td>4.208</td>
</tr>
<tr>
<td>p5</td>
<td>4.417</td>
<td>4.125</td>
</tr>
<tr>
<td>p6</td>
<td>3.333</td>
<td>2.625</td>
</tr>
<tr>
<td>p7</td>
<td>4.000</td>
<td>4.083</td>
</tr>
<tr>
<td>p8</td>
<td>4.625</td>
<td>4.750</td>
</tr>
<tr>
<td>t1</td>
<td>3.875</td>
<td>4.000</td>
</tr>
<tr>
<td>t2</td>
<td>3.542</td>
<td>4.292</td>
</tr>
<tr>
<td>t3</td>
<td>3.292</td>
<td>4.250</td>
</tr>
<tr>
<td>t4</td>
<td>4.000</td>
<td>3.583</td>
</tr>
<tr>
<td>t5</td>
<td>4.542</td>
<td>4.042</td>
</tr>
<tr>
<td>t6</td>
<td>4.583</td>
<td>4.167</td>
</tr>
<tr>
<td>t7</td>
<td>3.875</td>
<td>4.208</td>
</tr>
<tr>
<td>t8</td>
<td>3.750</td>
<td>3.750</td>
</tr>
</tbody>
</table>
However, there are a few cautions in interpreting the data. First, the amount of data from a single participant (64 trials) is a bit low for a 4x4 chi-squared table, so the reported p-values may not be completely accurate. Second, the interpretation of statistical error becomes somewhat difficult when multiple tests are run. With that said, the number of significant results seems to be more than what would be expected from a 5% chance of Type I error across 24 tests. Also, the correspondence with musical experience is fairly persuasive. Less satisfying, though, is the relative paucity of significant individual tests for Ständchen; if the interpretation of the individual tests for Melody 1 is valid, then the performances of Ständchen would be expected to follow the same pattern. One possible explanation is that categorization was made largely on the basis of performance errors, and there were many more errors in Melody 1 than in Ständchen. However, responses to Ständchen did in fact produce a significant relationship in aggregations, so another possibility is simply that differences in responses are due to the stylistic differences between the two pieces.

It is apparent from the analyses of global metrics that participants were responding to features of the performances. Perhaps due to their own experiential biases, however, they often were responding incorrectly to the features. For example, participants’ responses of “Trumpet” tended to correspond with performances which were, on average, softer and slower than others. In fact, the opposite pattern tended to be the case for the actual trumpet performances. Furthermore, there were not significant effects for articulation, even though the performance results showed strong effects of articulation by instrument. In all, the pattern of perceptual results indicate that the participants were sensitive
to non-acoustic musical features, but that they lacked the ability to place performances veridically into categories of instruments. It may be that these skills require a greater degree of musical training than many of the participants possessed.

Analyses of goodness ratings show some effects of performer and instrument; these tend to follow the rates of errors, though, and may simply be explainable on that basis alone. It is somewhat surprising that the goodness ratings do not show a preference for expressive performances over mechanical performances. In fact, listeners of Ständchen displayed a significant effect in the opposite direction, for mechanical over expressive performance. This may be another indication of the difficulty of the task; the impoverished nature of the synthesized stimuli may obscure some of the intended meaning of expressive deviations.

Summary

Expressive features of performances should be perceptible to listeners, both in terms of overall expressive content and in terms of distinguishing characteristics of performances on particular instruments. There is some evidence in the literature that listeners are sensitive to expressive cues in a performance event when the timbre is neutral or inconsistent. Therefore, a perceptual experiment was described which tested listeners’ ability to identify the original instrument from a performance in which all timbral information had been removed. The MIDI files from Chapter 3 were played to participants all in the same neutral timbre, and they were asked to identify the original performance instrument and to rate the quality of the performance. The results were not conclusive, but
showed some evidence that listeners were responding to instrument-specific cues, particularly for those listeners who had some musical experience.
CHAPTER 6

GENERAL DISCUSSION

The central empirical issue of this dissertation is the extent to which a performance instrument influences the expressive interpretation of a piece of music. The performance analyses seem to demonstrate fairly convincingly that there are differences across the instruments in this study; effects of instrument were obtained for every dependent measure, for both musical pieces. However, the nature of the effects was different for the two pieces. Ständchen showed some fairly stable patterns which were modulated by instrumental or expressive nuance, while the patterns of Melody 1 were sometimes radically different across instruments.

An obvious question, then, concerns the generalizability of the empirical findings. If the performance instrument is truly an important component of expressive interpretation, then the effects ought to be apparent across other instruments and across other musical compositions. Undoubtedly, some of the effects from this study are functions of the particular individuals, instruments, and musical stimuli. However, there are also indications of some general phenomena. On Melody 1, for example, the timing perturbations around large pitch leaps are more pronounced for the wind instruments than for the guitar and
piano, and the two wind instruments played overall both faster and more staccato than guitar or piano. For Ständchen, guitar and piano show a greater variability of dynamics than the wind instruments, but less variability of articulation.

One interpretation of the instrumental differences on Melody 1 is simply that the technical demands of the pieces differentially challenged certain instruments, and that the observed differences are akin to performance errors rather than intentional attributes of expression. There may be some merit to this argument. However, the deviations tend be larger for expressive than for mechanical performances, which would not be expected if they simply reflected limitations of performances. Rather than a limitation, then, the characteristic patterning of performance variables may be more like an “accent” in the musical language of an instrument; not necessarily intentional, but nonetheless legitimately expressive.

The effects on articulation seem also to distinguish the wind instruments from guitar and piano. Since wind instruments sustain their tone, they have greater control over note endings, so articulation is a natural aspect to utilize in performance. This may result in a greater use of articulation as an expressive variable for wind instruments. In particular, this interpretation might help to explain the patterns of dynamics and articulation for Ständchen. It may be that expressive content is manifested differentially along several variables, such that the greater variability in dynamics for guitar and piano serves to compensate for a lesser variability in articulation. There are many possible dimensions of variability in a performance, some of which will be more fully exploited than others. Indeed, some dimensions do not exist at all on particular instruments, such as microtonal
pitch bends or timbral variability on acoustic pianos. Therefore, the types of variation in musical performance would be expected to differ across instruments as well as across musical style (e.g. Temperley, 2004).

To be sure, the instrumental differences do not neatly divide themselves between wind instruments and non-wind instruments. For example, the timing patterns for Ständchen show similarities between clarinet and piano performances. Although there may be dimensions of performance which lend themselves preferentially to categories of instruments, such as the discussion of articulation above, it is almost certainly true that each instrument is a complicated set of many dimensions, any subset of which might be relevant to a particular piece of music. The two compositions in this study were deliberately chosen to contrast a technical versus an expressive musical content, and the performances were commensurately different for each. Simple global musical properties, such as tempo, were treated differently in the two pieces; performers played Melody 1 faster in their expressive performances, but played Ständchen slower in their expressive performances. Also, there tended to be more main effects of expression type for Ständchen, and of instrument for Melody 1. Additional compositions drawn from the spectrum of music would certainly reveal many more dimensions along which instrumental performances would vary.

Sources of Variability

Assuming that instrument-specific characteristics, or “accents,” do in fact exist, there are a number of possible contributing factors. First, there are the physical characteristics
of the instrument itself, and the corresponding demands and constraints placed upon the performer. Some obvious examples are the demands of breath and intonation on performers of wind instruments. Performers of wind instruments must be able to take breaths to prepare for relatively long sequences of uninterrupted notes and for particularly demanding notes, such as high pitches or quick intervals. These are the conditions which tended to produce errors in the trumpet performances in this study.

Non-wind instruments also have physical demands. The piano and guitar, for instance, have very different constraints governing the types of chords that can easily be performed. The constituent pitches on a guitar chord must be within a fairly close range on the fretboard, as they must all be fingered by the same hand, and there can be a maximum of six simultaneous pitches, since there are six strings on the guitar. A piano can sound more than six simultaneous pitches, and the constituent notes of a chord must be drawn from within two different local ranges of pitch, corresponding to the locations of the two hands on the keyboard. In addition to constraints on simultaneous notes, the same configurations govern the difficulty of transitions from note to note across hand or string position (Baily, 1977, 1985; Nettheim, 1979; Parncutt, Sloboda, Clarke, Raekallio, & Desain, 1997; Sayegh, 1991).

The trumpet and clarinet, on the other hand, do not normally sound simultaneous notes at all, so chordal events on these instruments (other than arpeggiations) are only realized by multiple performers in an ensemble. Therefore, different instruments are necessarily used to fulfill different musical roles. Non-chordal instruments are rarely used in solo performances or as vocal accompaniments. The implications are that a performer of
an instrument will have a repertoire of stylistically-constrained expressive cues, depending upon his or her training. In particular, instrumentalists who perform notated music within an ensemble have much less freedom to vary attributes such as expressive timing than do solo performers (Friberg & Sundstrom, 2002; Rasch, 1979; Shaffer, 1984). These backgrounds could be internalized to shape performers’ conceptions of expressive content even when the external constraints are removed.

Thus, there is a second set of influences on instrumental characteristics, having to do with the musical background and familiarity with particular musical style. Obviously, these factors arise from the experiences of the performer, but a distinction can be drawn between the influences of stylistic correlates of instruments and purely individual differences among performers. The selective usage of instruments in specific musical applications is not arbitrary; they are chosen at least in part due to implications of their physical properties, including voicing, tonal range, and timbre. A performer’s background, then, can be interpreted to some extent as mediated by the instrument. For instance, performers occasionally remark that they tend to approach music “vertically” or “horizontally,” depending upon whether their instrument is predominantly used harmonically or melodically.

Additionally, there are of course idiosyncrasies unique to the performers themselves. This is the third set of influences on instrumental characteristics. Each performer has his or her own interests and experiences, which will be reflected in their performances. It is possible, however, that even this level of variability is not entirely random. An individual may be drawn to a particular instrument due to its typical stylistic use, its timbral proper-
ties, or even an anatomical “fit” with the instrument, such that the act of playing it feels more comfortable than other instruments. Regardless, any empirical study of performance, including this one, will necessarily contain a component of individual interpretation.

Perceptual Data

The theoretical emphasis of this project has been on the importance of the particular instrument in shaping a musical performance. This implies that the instrument-specific properties should be salient to listeners. The perception data in this study, however, is somewhat equivocal. There are indications that listeners were responding to instrument-specific qualities of performances, but the data was certainly not straightforward.

To some extent, the complexity of the data is not surprising. An implication of the semiological model presented in Chapter 2 is that performers and listeners each operate separately on the acoustic trace which constitutes the performance itself. In other words, information is not necessarily transferred in a direct veridical fashion, so the interpretations of a listener will not necessarily match that of the performer. However, that is certainly not to say that the message is random; listeners might respond differently from each other to performance variables, but there should be some pattern of response nonetheless.

If instrument-specific cues to musical expression are real phenomena, then they should be demonstrable in perceptual data.

One way of interpreting the data is in terms of the consistency in the performance analyses. Melody 1 varied by instrument in a variety of ways and along a number of
dimensions. Listeners may have been responding to different dimensions in their responses, which would explain why analyses of individual responses suggested meaningful patterns of response which may have been partly obscured in the aggregated data. This would also be consistent with the fact that respondents with musical experience performed differently than those without musical experience. Ständchen, on the other hand, displayed a more predictable pattern across instruments, which might help to explain why listeners produced a significant relationship in aggregated responses.

Undeniably, though, task performance in the perception study was rather low. In comparison, participants were able to identify the performance instrument at a better-than-chance rate in Walker (1996). However, two major differences combine to make the current task more difficult than in the 1996 study. First, participants in the current study had four instruments to choose from, compared to only two (piano and guitar) in the 1996 study. Second, the stimuli in this study were always presented in the same neutral timbre. By contrast, stimuli in the 1996 study were always played with either a guitar or a piano timbre, so the task could in fact be reduced to a decision of whether the timbre was consistent with the original instrument or not. Although the timbre pairings were statistically balanced so as to be uncorrelated with the original instrument, it is not true to say that timbre was eliminated as a cue to the extent that it is in the current study.

Timbre is obviously a very important quality of music perception, and the stimuli in this study are greatly impoverished in comparison to the original digital audio recordings. There are a great number of inflectional cues in the acoustic waveform which are not represented at all by MIDI data. Without these cues, the task for a listener not only becomes
very difficult, but may also become qualitatively different; the still-existing systematic
variability in timing, articulation, and dynamics may not be as interpretable without poten-
tial correlates of harmonic overtones, vibrato, or other timbral mechanisms. For example,
the participants who heard Ständchen rated the mechanical performances as significantly
better than the expressive performances, which is opposite of the expected direction. One
possible explanation is that the larger degree of variability in expressive performances was
difficult to contextualize without associated timbral information.

**Methodology**

From an empirical perspective, the goals of this study were to provide a more precise
quantitative understanding of how performers’ musical interpretations and expressive per-
formances are mediated by the characteristics of their respective instruments. Hopefully,
that has been at least partially accomplished; several interesting patterns have emerged
which seem to distinguish among the four instruments.

However, the greater practical importance may be in the establishment of a method-
ology for investigating performances by multiple instruments. Although this is not the
first study to look at performances from a variety of instruments, it has attempted to be as
comprehensive as possible and to build upon the corpus of methodologies in the perform-
ance literature. In particular, the use of MIDI has perhaps been both a blessing and a
curse to performance studies; it certainly simplifies the process of data collection for situ-
ations (i.e. keyboard performances) in which it is appropriate, but it correspondingly tends
to define the nature of experiments. Empirical studies of performance should focus on what is interesting and informative, not just on what is easily accessible.

In its use of MIDI, this study is something of a compromise. It does not rely on MIDI for data collection. However, it does primarily use MIDI information as a method of data representation. By deliberately separating data collection and representation formats, this study is able to exploit the flexibility and ubiquity of MIDI as a computational tool without constraining the nature of the instrumental performances. The process of translation from digital audio to MIDI is time-consuming and intensive, but the results yield a richer set of information (and possibly are less intensive) than separately coding dependent variables manually (e.g. Ashley, 2002; Collier & Collier, 2002; Friberg & Sundstrom, 2002; Gabrielsson & Juslin, 1996).

However, even as a representational format, MIDI is not without its drawbacks. First, the process of translation is not seamless or error-free, and gets much more complicated for polyphonic performances (e.g. Lepain, 1999). The most glaring weakness of MIDI, though, is its treatment of timbral correlates such as vibrato and overtones; a typical MIDI note event merely marks the beginning, end, and key velocity for each note. Some additional features, such as volume envelopes and microtonal pitch bends, are available as MIDI information, but these measure fail to adequately describe the richness present in the acoustic array. In general, properties of timbre and acoustics are very difficult to quantify and therefore have not been sufficiently addressed in empirical studies of performance. Also, note that the lack of attention to acoustic data in performance is compounded by the piano-centric bias, as timbre is not particularly controllable in piano
performances. Future research, particularly as suitable technology develops, should endeavor to address these important topics.

As has been discussed several times in this dissertation, effective use of technology and computational tools is extremely important for an appropriate analysis of performance data. Audio to MIDI transcription would not have been possible without a software package like Logic Audio, but an additional suite of programs were also written to transform the data into all of the necessary intermediate formats. Most of these programs could have been written in essentially any language on any reasonably modern platform. However, an enormous amount of efficiency was gained in this project by using Unix and automating series of formatting commands through shell scripting; in this way, changes to the format could be immediately applied to thousands of files in a matter of seconds. Again, most modern operating systems provide some sort of scripting capability, but the built-in power of Unix file processing commands (e.g. awk and sed) were a great convenience to this project.

The only program that was discussed in any detail in this document was PerfComp, which addresses the non-trivial problem of categorizing performance errors. The interesting aspect of PerfComp is that it works not by following fixed rules for a list of possible note errors but rather by heuristically evaluating sets of possibilities, much the way a human evaluator might. The program was used in a relatively narrowly-defined context in this study, but it could be more fully developed in order to mimic human results more completely and flexibly. For example, a set of data from human evaluators might be
acquired from performances of a variety of compositions, and the metric weightings of PerfComp could be empirically adjusted to obtain the best fit.

The primary focus of this study is on performance characteristics. However, there are also some implications for perceptual methodologies. The perceptual study used stimuli derived from real performance data, and thus attempted to elucidate the connection between performers and listeners. Whether any specific communicative model is posited or not, studies of music perception will of course continue to be concerned with the expressive attributes of music performance (e.g. Balkwill & Thompson, 1999; Kendall & Carterette, 1990; Nakamura, 1987; Repp, 1994; Schellenberg, Krysciak, & Campbell, 2000; Senju & Ohgushi, 1987; Sloboda & Lehmann, 2001; Sundberg & Verillo, 1980; Timmers, 2003). Specifically, the perceptual experiment in this study could be extended in a variety of ways. The musical backgrounds on participants could be explicitly targeted for specific instrumental or stylistic training; the task and number of possible instruments could be varied in order to decrease the overall difficulty; and particular performance attributes, including timbral features, could be deliberately modified in order to assess their importance to perception.

Summary

Overall, the data show that expressive variables of musical performance are patterned differently for different instruments. The shaping of a performance as a function of the instrument is characterized as a musical “accent” of that instrument, and may be a source of aesthetic quality. The instrumental accents may arise from a multitude of
sources, including physical demands of the instrument, stylistic correlates of arrangements, and the background of the performer. Furthermore, the accents are presumed to be multi-dimensional and thus to manifest themselves differently for different types of music. There is some evidence that listeners are sensitive to instrument-specific variation in performance, but task difficulty and differences in response strategies make this data more difficult to interpret. As a practical consideration, this dissertation’s most important contribution is probably its methodology. Quantitative analyses of music performance have relied far too much on data from keyboard instruments, to the near exclusion of other instruments. More empirical studies are needed of performances from a variety of instruments in order to obtain data that is truly generalizable; hopefully this study will help to facilitate that process.
APPENDIX A

NOTATION FOR MELODY 1

\[\text{notation for melody 1}\]
APPENDIX B

NOTATION FOR STÄNDCHEN

\[ \text{Music notation} \]

\[ \text{Music notation} \]
### APPENDIX C

#### AUDIO-TO-SCORE SETTINGS IN LOGIC AUDIO

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Clarinet</th>
<th>Guitar</th>
<th>Piano</th>
<th>Trumpet</th>
</tr>
</thead>
<tbody>
<tr>
<td>Granulation</td>
<td>40 ms</td>
<td>80 ms</td>
<td>70 ms</td>
<td>75 ms</td>
</tr>
<tr>
<td>Attack Range</td>
<td>50 ms</td>
<td>75 ms</td>
<td>40 ms</td>
<td>60 ms</td>
</tr>
<tr>
<td>Smooth Release</td>
<td>6%</td>
<td>18%</td>
<td>8%</td>
<td>5%</td>
</tr>
<tr>
<td>Velocity Threshold</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Minimum Quality</td>
<td>Normal</td>
<td>Normal</td>
<td>Normal</td>
<td>Normal</td>
</tr>
<tr>
<td>Time Correction</td>
<td>0 ms</td>
<td>0 ms</td>
<td>0 ms</td>
<td>0 ms</td>
</tr>
</tbody>
</table>
LIST OF REFERENCES


performance: Analysis of legato, staccato, and repeated notes in performances of the andante movement of Mozart’s Sonata in G Major (K 545). *Journal of New Music Research*, 29, 211-224.


saxophonists, *Perceptual and Motor Skills, 73* (3), 707-713.


