An empirical investigation of a sarcastic tone of voice in instrumental music

DISSERTATION

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

By

Joseph Stephen Plazak

Graduate Program in Music

The Ohio State University

2011

Dissertation Committee:

David Huron, Advisor

Marc Ainger

Anna Gawboy

Copyright by

Joseph Stephen Plazak

2011

Abstract

Music is often regarded as being similar to language. Both music and language are capable of communicating affective information through changes in "tone of voice." This study examines the communication of a sarcastic tone of voice in a musical context. It was hypothesized that the cues of sarcastic music would be similar to the cues of sarcastic speech. The first study investigated the acoustical cues used by instrumentalists to encode a sarcastic tone of voice. These cues were found to include: a noisy timbre, staccato-like articulations, relatively shorter durations, and relatively greater intensity. The second study investigated the acoustical cues used by listeners to decode a sarcastic tone of voice. These cues were found to include: a noisy timbre, staccato-like articulations, relatively shorter durations, and relatively lower pitches. It was further hypothesized that listeners would be able to identify sarcasm in a musical context. In a matching task, listeners were able to distinguish a sarcastic tone of voice better than any of the other seven affective tones of voice (i.e. sadness, joy, sincerity, anger, mysterious, tenderness, neutral) used in the study. In order to test for demand characteristics, a free-response study was devised and found that contextual clues are an important factor in hearing a sarcastic tone of voice in instrumental music.

Acknowledgments

I wish to express my most sincere thanks to my advisor, Dr. David Huron, for his help in developing the ideas and methodologies contained within this document, and also for his sage advice and counsel throughout my graduate studies.

I also wish to thank the members of my dissertation committee, Dr. Marc Ainger, Dr. Anna Gawboy, and Dr. Douglas Nelson, for sharing their time and expertise as I prepared this work, and also for challenging my ideas on this topic.

During my tenure at Ohio State, I benefitted greatly from working with the following faculty members: Dr. Lawrence Feth, Dr. Patricia Flowers, Dr. Lora Dobos, Dr. David Clampitt, Dr. Graeme Boone, Dr. Thomas Bishop, and Dr. Michael Murray.

I am also indebted to the feedback and mentorship of the following teaching coordinators: Dr. Ann Stimson, Pete Tender, and Dr. Ellen Archambault.

In addition, many thanks go to my colleagues in the Cognitive and Systematic Musicology Laboratory, with special thanks to Niall Klynn, Joshua Albrecht, and Brandon Paul, for offering their suggestions on this document.

Finally, I wish to thank the performers and listeners who generously gave their time in order to make this study possible, and the OSU statistical consultants for their invaluable advice.

Vita

May, 2002 .....St. Edwards Central Catholic High School

May, 2007 .....B.A. Music/Psychology, Elmhurst College

August, 2009 .....M.A. Music Theory, Ohio State University

August, 2007 to May, 2011.....Graduate Teaching Associate, Department of Music, Ohio State University

Publications

Plazak, J., & Huron, D. (2011). The first three seconds: listener knowledge gained from brief musical excerpts. *Musicae Scientiae, 15*(1), 29-44.

Plazak, J., Huron, D. & Williams, B. (2010). Fixed average spectra of orchestral instrument tones. *Empirical Musicology Review, 5*(1), 10-17.

Plazak, J. (2008). The effects of attention on frisson-related responses from unexpected musical events. *The Ohio State School of Music Online Journal, 1*(2). http://www.osomjournal.org/.

Fields of Study

Major Field: Music

Table of Contents

Abstract ii

Acknowledgments iii

Vita iv

List of Tables viii

List of Figures x

Chapter 1: Introduction 1

Chapter 2: Background 5

Theoretical Foundations 5

Musical Affect 11

Models of Musical Communication 15

Irony and Sarcasm 21

Hypotheses 27

Chapter 3: Performer Model of Musical Sarcasm 29

Background 29

Methodology 30

Results 33

Chapter 4: Listener Model of Musical Sarcasm 55

Methodology 55

Results 57

Discussion 69

Chapter 5: Concordance Between the Performer and Listener Models 75

Discussion 85

Chapter 6: Free-Response Study 89

Choosing a Musical Corpus 90

Free-Response Study: Methodology 99

Results 102

Discussion 104

Chapter 7: Conclusions and Theoretical Implications 109

References 121

Appendix A: Directions supplied to performers 133

Appendix B: Sample PRAAT Voice Report 135

Appendix C: Performer Multinomial Logit Model 136

Appendix D: Performer Model ANOVAs 137

Appendix E: Listener Study Posting 139

Appendix F: Listener Multinomial Logit Model 140

Appendix G: Listener Model ANOVAs 142

List of Tables

Table 1: Empirical and theoretical findings on the acoustical cues of speech-based sarcasm. See the text above for an explanation of the abbreviated citations. Note that some of the findings are contradictory (e.g. pitch cues). 26

Table 2: Summary of the recording session. Each performer recorded 5 different types of stimuli 4 separate times in order to convey the following tones of voice: joy, sadness, sarcasm, and sincerity. Each session resulted in 20 recordings. 31

Table 3: Results from the multinomial logit model and the associated ANOVA tests. The mean values for each variable are provided as a function of the target tone of voice. The asterisk (*) indicates a mean which was found to be significantly different from the mean for sarcasm at the p=.05 level. Significant differences between the other tones of voice are not reported here, but can be found in appendix D. 35

Table 4: A comparison of the cues used by performers to convey a sarcastic tone of voice in music and the cues of sarcastic speech. In general, there seems to be more similarities between these two contexts of sarcasm than differences. 53

Table 5: Results from the multinomial logit model and the associated ANOVA tests. The mean values for each variable are provided as a function of the target tone of voice. The asterisk (*) indicates a mean which was found to be significantly different from the mean

for sarcasm at the p=.05 level. Significant differences between the other tones of voice are not reported here, but can be found in appendix F..... 61

Table 6: A comparison of auditory cues in sarcastic speech and cues used by listeners to detect a sarcastic tone of voice in instrumental music..... 72

Table 7: Matching success (percentage) as a function of stimulus type and performer's intended tone of voice. Chance level for the 8-alternative-forced-choice task was 12.5%. 78

Table 8: Confusion matrix: listener responses as a function of performer intention. The performers' intentions are organized as rows; listener responses are organized as columns..... 82

Table 9: Comparison of successfully and unsuccessfully identified sarcastic recordings based on the 13 factors from the Listener (L) model..... 84

Table 10: Prevalence of comments that contained words indicating sarcasm or other types of negatively valenced affects as a function of presentation type. 102

List of Figures

Figure 1: Brunswik's (1952) lens model (Adapted from Cooksey, 1996, p.56). Ecological validity is represented by the left side of the figure; cue utilization is represented on the right side. The degree to which these two sides match can be used to estimate the likelihood that the distal variable will be correctly identified. 16

Figure 2: Juslin's modified lens model (adapted from Juslin & Sloboda, 2010). The weighting of the expressive cues for the performer are on the left side; the listener's expressive cue weightings are on the right. Assuming that performers and listeners are consistent, the degree to which these cues match determines the likelihood of successful communication..... 18

Figure 3: Kendall and Carterette's (1990) Three-Part Model of Musical Communication. C= Composer, P=Performer, L=Listener. Compared with Juslin's "modified lens model," this model adds a composer component. Contextual factors are also included for each of the three components. 19

Figure 4: Juslin and Lindström's (2003) Extended Model Lens Model (adapted from Juslin & Laukka, 2004). The model accounts for cues used by composers, performers, and listeners. In this model, the interaction between composer and performer cues is also considered. 20

Figure 5: Global Tempo (in beats per minute) as a function of tone of voice. For each tone of voice, the bold black line indicates the median value, the box indicates the upper and lower quartiles (the middle 50% of the data), the lines extending from the box indicate the highest and lowest values when excluding outliers, and the circles indicate outliers which were more than 3/2 times greater or lesser than the upper and (cont.) (cont.) lower quartile. The global tempo of sarcastic recordings was found to be significantly faster than sad recordings. 42

Figure 6: "Mean Articulation Scores" (d_{io}/d_{ii}) as a function of target "tone of voice." For each "tone of voice," the bold black line indicates the median value, the box indicates the upper and lower quartiles (the middle 50% of the data), the lines extending from the box indicate the highest and lowest values when excluding outliers, and the circles (cont.) (cont.) indicate outliers which were more than 3/2 times greater or lesser than the upper and lower quartile. Articulation scores for sarcastic notes were significantly smaller than sad notes and sincere notes. 44

Figure 7: Change in articulation score from note to note in the "9 repeated note" samples as a function of target "tone of voice." Changes in articulation appear to be patterned. A rising line indicates that a more legato note followed a more staccato note, whereas a descending line indicates the opposite. 45

Figure 8: "Average Intensity" (in dB) as a function of the target tone of voice. For each tone of voice, the bold black line indicates the median value, the box indicates the upper and lower quartiles (the middle 50% of the data), the lines extending from the box indicate the highest and lowest values when excluding outliers, and the circles indicate

outliers which were more than 3/2 times greater or lesser than the upper and lower quartile. Joy, sincerity, and sarcasm were found to have louder average intensities than sadness. 48

Figure 9: Inter-onset intervals (in seconds) between successive notes as a function of the target tone of voice. For each tone of voice, the bold black line indicates the median value, the box indicates the upper and lower (cont.) 49

Figure 10: Change in IOI (in ms) from note to note in the "9 repeated recordings" as a function of the target "tone of voice." An upward sloping line indicates that performers increased IOI from one note to the next, whereas a downward sloping line indicates that performers decreased IOI..... 51

Figure 11: Performer's model for expressing a sarcastic tone in instrumental music. 52

Figure 12: Screenshot of the experiment interface used by participants. Responses were entered by using a computer mouse to click on the tone of voice intended by the performer. The tone of voice responses were in a fixed position on the screen, and were not randomized by the experimental software. 57

Figure 13: The Listener model for detecting a sarcastic tone of voice in instrumental music. 70

Figure 14. A transcription of the trumpet player's improvisation expressing sarcasm. In addition to tone of voice cues, musical and non-musical associations may contribute to hearing sarcasm in a musical context..... 71

Figure 15: Performer and Listener models of a sarcastic tone of voice in instrumental music. In general, evidence was found that performers and listeners have a similar "acoustic code" for a sarcastic tone of voice in music..... 75

Figure 16: A hypothetical composition which contains the musical features which were found to *describe*, but not *define*, a sarcastic tone of voice in music. 88

Figure 17: Prokofiev, op.17, *Sarcasms*, no.1, m.1-11. Potential sarcastic tone of voice features include low toneness, contrastive articulations, contrasting dynamics, and the "*ironico"* performance marking in m.5. 93

Figure 18: Prokofiev, op. 17, *Sarcasms*, no.3, m. 45-51. Potential sarcastic tone of voice cues include rapidly changing dynamics, and the prevalence of dissonance resulting from using two key signatures simultaneously..... 96

Figure 19: Prokofiev, op.17, *Sarcasms,* no. 4, m.12-16. Potential sarcastic tone of voice cues in this passage include changes in register, changes in texture, and the use of grace notes and glissandi. 98

Chapter 1: Introduction

Humans communicate in a variety of ways, some of which are highly sophisticated. Besides communication through language, we are also able capable of communicating through facial expression, posture, tones of voice, etc. Consider the following examples: 1) Imagine that you are waiting at a bus stop and overhear a conversation taking place behind you at the nearby outdoor cafe. A young man says to his friend, "Spring break is too long. I'm ready to go back to school." Even though you cannot see the man, you are able to deduce that he did not mean what he said. 2) Your mother asks your older brother if he has cleaned his room. He replies, “Yeah yeah.” Your mother senses his insincerity and immediately reprimands him. 3) While visiting a foreign country, you encounter a political rally which is being contained by the ruling party's militia. The boisterous protest group is loudly singing the country's national anthem, which for some reason, appears to be upsetting the militia. Even though you don’t understand the native language, you can sense that the protestors are mocking the national anthem.

Not only are we capable of expressing ourselves through words, numbers, music, gestures, etc., we are also capable of expressing to others that our intentions are not to be taken literally. Non-literal communication has become a recent topic of interest in many fields, including linguistics, psychology, literary studies, philosophy, etc. The motivation

for using non-literal communication, as well as the ways in which we signal non-literal intentions, are just beginning to be understood.

How might you have determined that the man at the outdoor cafe was not being earnest about his desire to go back to school? One possibility is that you recognized the contextual incongruity that "breaks" are rarely expressed as being "too long." Another possibility is that you identified something unique about his tone of voice. In the second example, how did your mother know that your older brother was being insincere? One clue may have been that he stated his answer twice. Another possibility is that she heard something unique about his tone of voice. In the third example, how might you have determined that the protesters were mocking the national anthem? One possibility is that you recognized the incongruity between their angry facial expressions and their solemn song. Another possibility is that you noticed the effect that the group's actions had on the militia. Yet another possibility is that you recognized something unique about the protestors' tone of voice.

Research on non-literal communication has primarily focused on non-literal speech (Haiman, 1998) and non-literal writing (Booth, 1974). In recent years, musicology has also begun to examine non-literal communication in music, with a specific interest in irony (Sheinberg, 2000). The topic of non-literal communication is important in music because: 1) it provides insight into the sophistication of our musical abilities, and 2) it shines light on the social functions of music.

In all of the stories above, non-literal intentionality could have been communicated through a variety of means and for a variety of reasons. One way in

which non-literal meaning can be expressed is through “tone of voice” (Gibbs, 1994; Haiman, 1998; Rockwell, 2006). Tone of voice (often referred to throughout this study simply as "tone") can be defined as the collection of non-semantic and non-associative properties of a sound, such as: timbre, pitch, rhythm, intensity, etc. Both speech and music are capable of using different tones of voice. Research has found that tone of voice cues can “successfully” signal the presence of non-literal speech (Rockwell, 2006). However, it is not known if tone of voice cues are capable of signaling non-literal intentions in music.

There are different types of non-literal communication, including irony, figures of speech, jocularity, hyperbole, rhetorical questions, understatements, etc. (Gibbs, 2000). This study investigates a specific type of non-literal music communication: sarcasm. Sarcasm is often considered to be a type of irony with a specific purpose: to criticize someone or something. One of the main features of sarcastic speech is that it can often be distinguished via a distinct tone of voice (Cheang & Pell, 2008). The sarcastic tone of voice has been the focus of recent scholarship in several fields, such as psycholinguistics and psychology. However, with very few exceptions (Fónagy & Magdics, 1967; Haiman, 1998), the ability of music to convey a sarcastic tone of voice has been largely ignored.

This study focuses on a sarcastic tone of voice in instrumental music. The first goal was to determine which musical features can be manipulated by instrumentalists in order to signal a sarcastic tone of voice in instrumental music (chapter 3). The second goal was to determine which musical features listeners hear as signaling a sarcastic tone

of voice in instrumental music (chapter 4). The third goal was to determine if a sarcastic tone of voice could be communicated between instrumentalists and listeners (chapter 5). The fourth goal was to determine the ecological validity of communicating a sarcastic tone of voice in instrumental music (chapter 6).

To foreshadow the results, this study found evidence in support of a high level of concordance between instrumentalists and listeners with regards to a sarcastic tone of voice in instrumental music. Further, this study also found evidence in support of a high level of concordance between a sarcastic tone of voice in instrumental music and sarcastic speech. As with sarcastic speech, the evidence suggests that hearing a sarcastic tone of voice in music depends on the listening context.

Chapter 2: Background

This chapter presents an overview of: 1) the theoretical foundations for this work; 2) musical affect; 3) relevant models of musical communication; and 4) empirical and theoretical findings on the communication of irony and sarcasm.

Theoretical Foundations

There are many similarities between speech and music. They both use sound or text as a medium for communication; they both may systematically manipulate pitch, timbre, and rhythm for different communicative purposes; and they both typically follow a specific syntax. Music has been referred to and debated as the "language of the emotions" (Pratt, 1931). The extent to which music can successfully communicate affect has been extensively studied, and has become a major topic of inquiry within the larger field of music psychology. Several books have been devoted specifically to the topic of musical affect and how it is communicated (Juslin & Sloboda, 2001/2010; Nussbaum, 2007).

There are many theories pertaining to music's ability to convey affect. These theories have been based on a wide variety of ideas, such as the "copy paradigm" (Seashore, 1938), expectation theory (Meyer, 1956), and information theory (Moles, 1958). These communicative models were popular in the past, and each made

presumptions about "successful" communication and how it might be achieved. The limitations of these communicative models were identified in the 1960's by "Reader-Response Criticism" (Tompkins, 1980; Davis & Womack, 2002). This movement faulted the theories above for their inability to explain multiple interpretations of "seemingly" objective stimuli. Barthe's (1968) essay on the "Death of the Author" argued that communicative messages cannot be purely objective, and further that the interpretation of such messages is subject to change over time. In recent years, there has been a renaissance of empirical work on musical communication. It has become possible to side-step some of the arguments of Reader-Response criticism by systematically manipulating the composer, performer, or listener's intentions. A recent trend involves using comparative research methods, such as studying music throughout the lens of disciplines such as linguistics, psychology, or neuroscience. The comparative study of music and language is particularly relevant to this study. Brief overviews of these theories, as well as how they relate to this particular work, are presented below.

Copy Paradigm

Carl Seashore, one of the early pioneers of music psychology, believed that "anything heard, experienced, understood, or felt is directly attributable to a real physical object, even one that is passing through time" (Fiske, 1996, pp.1-2). Seashore (1938) wrote, "everything that is rendered as music or heard as music may be expressed in terms of the concepts of the sound wave" (p.2). Fiske referred to this core belief as the "copy paradigm." This term implies that everything capable of being communicated through

music must be present (i.e. copied) within the sound source. Seashore believed that "esthetic deviation from the regular–pure tone, true pitch, even dynamics, metronomic time, rigid rhythms, etc." was the impetus for communicating affect through music (Seashore, 1938, p.9). Research has shown that many types of musical affect are expressed through the systematic manipulation of Seashore's four basic sound variables: frequency, intensity, duration, and waveform (Juslin & Laukka, 2003). The copy paradigm can be used to describe acoustical features that are associated with musical affects, such as happiness, sadness, tender/longing, seriousness, anger, etc. However, there are several musical events that copy paradigm does not explain, such as feelings of musical tension or resolution. For example, the four basic musical elements listed above cannot account for feelings that accompany a deceptive cadence. Further, the copy paradigm cannot account for musical meaning caused by extra-musical associations. For example, a country's national anthem is likely to evoke nationalistic feelings when heard by its citizens, but it may not evoke any such feelings for non-citizens (Clarke, 2005). Despite the limitations of the copy paradigm, it is a potentially useful starting point for describing some (but not all) of the elements that allow music to communicate affective information.

Expectation Theory

Leonard Meyer's (1956) theory of musical expectation accounts for some of the types of musical affect that the copy paradigm fails to explain. The central thesis of this theory is that "affect is aroused when a tendency to respond is arrested or inhibited"

(Meyer, 1956, p.14). This theory suggests that affective responses, such as the one caused by a deceptive cadence, are the result of inhibited musical tendencies. One advantage of expectation theory is that it can account for responses to dynamic musical elements, such as tonal/melodic tendencies, rhythmic syncopation, the likelihood of chord progressions, or the return of a theme within a sonata-form recapitulation. One problem with expectation theory is that it doesn't provide a means for distinguishing between different affects. Meyer's theory merely states that musical inhibition results in an affective response. It does not account for specific emotions or feelings. Meyer makes a distinction between "affect" and "affective experience." The former is undifferentiated, whereas the latter can be differentiated by "an awareness and cognition of a stimulus situation which always involves particular responding individuals and specific stimuli" (p.19). "Affective experiences" are therefore specific to the context, listener, and the musical stimuli. Claiming that music sounds "happy" or "sad" would be a reference to an "affective experience" rather than an "affect." In order to understand why someone might hear a piece as happy or sad, it would be necessary to examine both the stimuli and the listener's context. In order to research how a listener can perceive affective information in a musical context, one must investigate beyond the inhibition/realization of expectations, and also investigate the experience and context of the listener.

Information Theory

Information theory—a way of determining/calculating the probabilities of a communication system—provides a framework for investigating relatively complex types

of musical communication. "Information theory is a theory of probability…It assumes that understanding is the result of a refined sensitivity to the relative probability of particular (musical) events and their future occurrence" (Fiske, 1996, p.20). An important component of information theory involves calculating the complexity, or the amount of information, present in a given system. Complex systems contain the potential to convey large amounts of information, but they also contain a larger potential for miscommunication. Several scholars were instrumental in applying information theory to music, including Claude Shannon (1948) and Abraham Moles (1958). Music information exists on a variety of levels, including the: physical/acoustical level; psycho-acoustical level (pitch, timbre, etc); cognitive level (form, tonality, etc); and the mental/social/cultural level (value, meaning, emotion, etc) (Clarke, 2005). One method of investigating complex levels of information in musical communication is to study the more basic levels of information that underlie the level of interest. For example, in order to gain insight into "socio-cultural" communication in music, one might study the "psycho-acoustical" and "cognitive" levels that underlie "socio-cultural" meaning. This is referred to as a "bottom-up" approach (Clarke, 2005). A major problem with applying information theory to music is that musical contexts tend to be very specific and weighted according to a variety of factors. Information theory assumes that the meaning of a musical message is related to its complexity, and therefore "copied" into the sound. [1] Information theory fails to account for different interpretations of the same musical

[1] Lerdahl (1988) further states: "Complexity refers not to musical surfaces but to the richness of their (unconscious) derivation by the listener."

stimulus. Despite this limitation, the notion of music communication being understood on a variety of levels is important to the present study.

Comparative Research: Music and Language

Many theories of music perception have been borrowed from or based on theories from linguistics (Heller & Campbell, 1982; Lerdahl & Jackendorf, 1983; Serafine, 1988). Of the many similarities between speech and music, Fiske (1993, p. 50) points out six of the most important:

> "1) Music is a part of all known cultures.
> 2) Music is a uniquely human activity.
> 3) Music occurs in a diverse assortment of styles, types, and genres.
> 4) Particular musical styles, types, and genres change (develop or evolve) with time.
> 5) Musical activity involves the generation and/or reception of tonal-rhythmic patterns of sound.
> 6) Music perception and comprehension requires time and effort."

Again, all six of these points are true for both language and music. Although comparative theories between music and language are not new, the abundance of empirical work that has been done to support these theories is relatively recent (Patel, 2008). For example, Fairbanks and Pronovost (1939) observed that sad speech was typically lower in pitch and more monotone. Recent comparative work has found evidence for a similar phenomenon in music; sad music is typically lower in pitch and exhibits smaller pitch movements (Huron, 2008; Huron & Davis, 2010; Huron, Yim, & Chordia, 2010). One possible explanation for the surge of empirical work comparing music and language may be the result of better technology for measuring sound and

computing large amounts of information. Regardless, music and language both "share a number of basic processing mechanisms…(and) the comparative study of music and language provides a powerful way to explore these mechanisms" (Patel, 2008, p.4).

All the theories described above—Seashore's copy paradigm, Meyer's expectation theory, Moles' information theory, and the recent comparative work on music and language—contribute to the theoretical background of this study. Specifically, Seashore's copy paradigm is invoked in chapters 3 and 4, in which the acoustic features of recorded performances are analyzed for the presence of systematic manipulations of frequency, intensity, duration, and waveform. However, no assumption is made that these acoustic features are interpreted universally, or even outside the context of this study. Elements from Mole's theoretical ideas are used to examine the degree to which performers and listeners successfully communicated information pertaining to musical affect. Again, no assumption is made that all listeners will encode/decode musical stimuli in the same way. Meyer's expectation theory is applied to the results of a third study which investigated musical context and "affect experiences." Finally, throughout this document, comparisons are made between music and language. Comparative studies are not without their limitations, but there are many interesting musical questions which can be addressed through interdisciplinary scholarship.

Musical Affect

The following definition of "affect" is taken from the Oxford Handbook of Music and Emotion: *Affect* is used as "an umbrella term that covers all evaluative—or

‘valenced’ (positive/negative)—states (e.g. emotion, mood, preference). The term denotes such phenomena in general” (Juslin & Sloboda, 2010, p.10). In this sense, affect can be used to talk about qualia, feelings, emotions, moods, etc. In addition to generalities, a functionalist approach to affect defines each of these “evaluative states” by its function or purpose (Juslin, 1997). Panksepp (1998) claimed that all affective systems can be linked to “behavior patterning,” meaning that affects are often paired with specific behaviors. In this regard, affects might be thought of as “motivational amplifiers.” When defining any type of affect, it is important to consider the underlying motivation that is served by the “evaluative state.” For example, understanding the affect of “hunger” should go beyond labeling it as negatively-valenced evaluative state; one must also address that the function of hunger is to amplify the motivation to seek food.

One of the most common distinctions made in music cognition literature is the difference between *perceived* and *felt* affect. It is quite possible for a listener to perceive a musical stimulus as being similar to, exemplary of, or communicating a certain affect without the listener actually feeling the affect itself. For example, it is possible for a listener to hear Samuel Barber’s *Adagio for Strings* as “being” sad, even though it may not make them “feel” sad. The present study is one of *perceived* musical communication rather than *felt* musical affect. It does not investigate induced musical affects.

The language used to discuss musical affect can be confusing. Consider the following: “That music is sad.” The problem with this statement is that music cannot be sad. Affects are not sounds, and further, sounds are not affects. In a similar way, speech cannot be sad. A person who is speaking can be sad, but their speech, which is merely

sound, cannot be sad. This leads to the question: what auditory characteristics contribute to the perception of affect? It is likely that many auditory characteristics play a role in perceiving affect; for the purposes of this study, these features will be collectively referred to as "tone of voice." In general, when music researchers discuss how musical affects are *perceived,* it may be more accurate to refer to affects as "tones of voice" rather than "feelings" or "emotions." A listener cannot hear an emotion, but they can hear a "tone of voice."

Another problem with researching affect pertains to classification. Affects can be organized based on many different factors: ontology, function, prevalence, valence and arousal, etc. Types of affect have been referred to by many names, including "primary," "secondary," "social," "self-conscious," "innate," "learned," "cognitive," "vicarious," and others. Classifying affect can be helpful in discussing different levels or layers of affective experiences, but whatever criterion is used will inevitably have limitations. Rather than organizing this discussion of affective tones of voice on any of the classifications mentioned above, this study elects to use a functional "literary perspective."

Nehaniv (1998) posited the idea of organizing affects based on the extent to which they are "self-oriented" vs. "other-oriented." The most self-oriented affects include motivations that are important for survival, such as "homeostasis, intake, outflow, hunger, pain, and irritation" (ibid, 1998, p.1). These affects might be considered "first-person affects" in that they are commonly viewed from one's own perspective. Nehaniv posits a middle category of affect which is also self-oriented, but includes "a generalized

or specific recognition of other agents or objects" (ibid, 1998, p.1). Affects that fit into this category include: fear, hatred, curiosity, envy, greed, etc. (ibid, 1998). These affects might be considered "second-person affects" in that they can typically be detected from someone else's perspective. This also means that second-person affects tend to be communicative. Detection of second person affects might take place through characteristic facial expressions, tone of voice, posture, or a variety of other cues. The last affect category is almost exclusively "other-oriented." An example of an affect that fits into this category would be empathy.[2] This feeling state typically involves some degree of "self-alienation,"[3] in which the agent removes themselves from their feelings and focuses on, or imitates, the feelings of another. Therefore empathy might be considered a "third-person affect." Nehaniv (1998) notes "It is unclear whether third person considerations are genuinely emotional or even whether they could enter with any phenomenological relevance for a situated agent" (p.3). However, empathy can be considered both a distinctive evaluative state (as required by the definition of "affect") and a motivational amplifier (as suggested by Panksepp). Although this "perspective-based" classification is not without its limitations, it is particularly appropriate for studies pertaining to affective musical communication.

Not all affects are equally capable of being communicated through tone of voice. Some affects have more communicative values than others, and further, only some of these affects are dominantly conveyed through sound (as opposed to facial expression or

---

[2] Nehaniv and I differ on the classification of empathy. Nehaniv considers empathy to be a second-person affect.

[3] This term is borrowed from John Haiman (1998).

posture). Hunger, for example, does not typically result in a specialized "tone of voice." It simply motivates the organism to find a source of energy. However, an affect such as irritation might be considered to have communicative value, especially if the irritation is external. In a survey of studies on musical affect, Juslin and Laukka (2003) found that five of the most popular affects studied in music research included: happiness, sadness, anger, fear, and love/tenderness. The finding that these affects are commonly studied in music research suggests that they all have characteristic acoustical features.

In summary, affect is used in this study as a blanket term indicating all evaluative states. Affects are motivational amplifiers and exist to meet a wide variety of motivations, including biological and social needs. Affects can be *perceived* or *felt.* Certain affects have more communicative value than others. When these evaluative states are communicated through sound, they can be referred to as affective tones of voice. There are many different ways to classify affects, but for studies on musical communication, a "literary-perspective" approach seems particularly appropriate.

Models of Musical Communication

Now that the general concepts of "affect" and "tone of voice" have been examined, it is possible to discuss how they may be communicated within a musical context. Communicating affective information through music is a complex process. Ways of accomplishing this type of communication include: learned associations (e.g. bagpipes), topics (e.g. hunting calls), tones of voice (e.g. prosody, paralanguage), etc.

This study focuses on only one of the several ways in which music is known to communicate affective information, namely, tone of voice cues.

In order for music to "successfully" communicate affective information through tones of voice, the cues used by musical producers (performers, composers) must to some extent match the cues used by musical receivers (listeners). A useful model for examining this so-called "cue concordance" was proposed by Brunswik (1952). This model, which he called the "lens model," was originally intended as a model of visual perception. The lens model estimates the likelihood that an observer will successfully identify a distal variable within the confines of their environment. An estimate of success can be determined based on the degree to which the actual properties of the stimulus match the cues used by the observer. The lens model is illustrated in Figure 1.

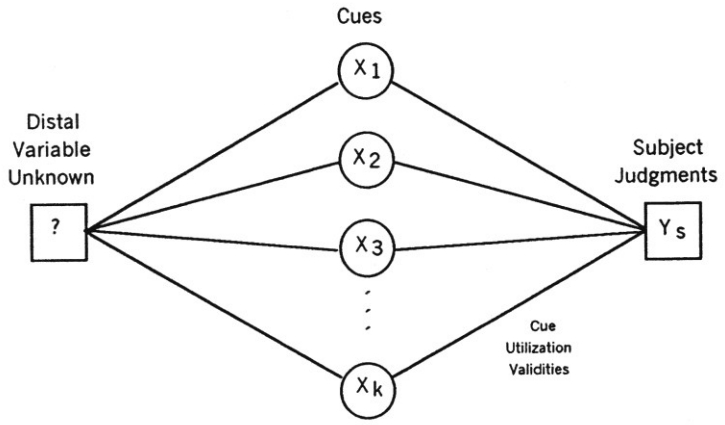


Figure 1: Brunswik's (1952) lens model (Adapted from Cooksey, 1996, p.56). Ecological validity is represented on the left side of the figure; cue utilization is represented on the right side. The degree to which these two sides match can be used to estimate the likelihood that the distal variable will be correctly identified.

The lens model was first applied to musical communication by Patrick Juslin (1995). The main advantage of applying the lens model to affective musical communication is that it can be used to quantify the systematic variations of pitch, rhythm, timbre, harmony, texture, dynamics, etc. used by performers and listeners for distinguishing various affective tones of voice. However, in musical communication, cues used by both performers and listeners may not be the same. Further, cues used by either the performer or the listener may not be consistent (Scherer, 1995; Johnson, 2009). In order to compensate for performer inconsistencies, listeners would presumably need to employ more cues, accept a wider range of cue weights, or ignore cues that are unreliable. In order to quantify the consistency of cues used by performers and listeners, Juslin (2000) added performer and listener consistency components to both sides of the lens model. This "modified" lens model is presented in figure 2.

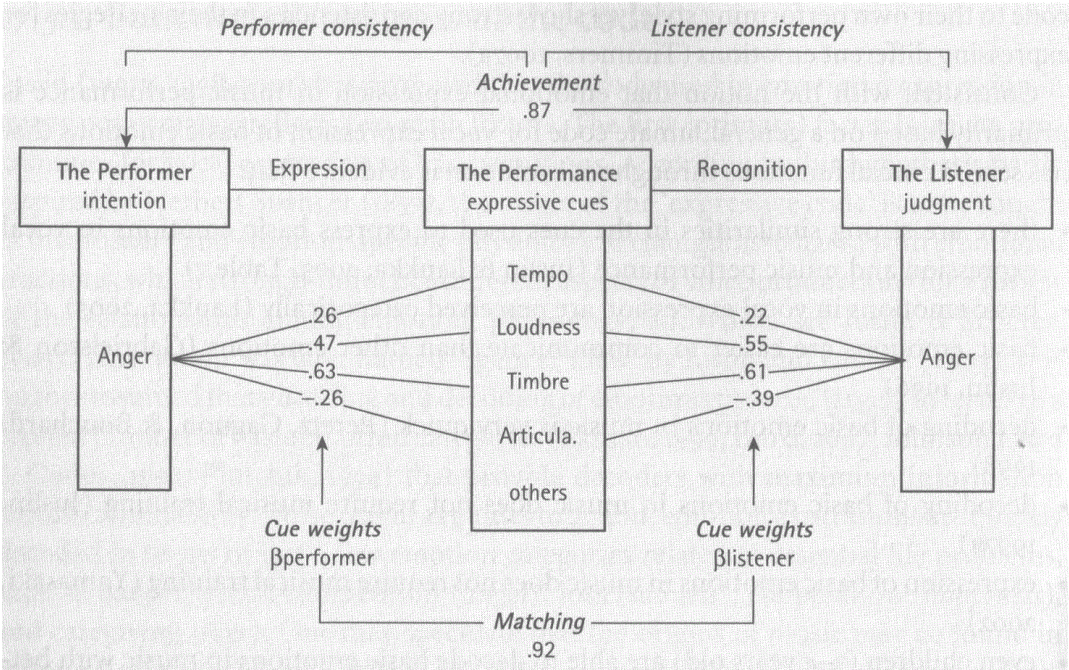


Figure 2: Juslin's modified lens model (adapted from Juslin & Sloboda, 2010). The weighting of the expressive cues for the performer are on the left side; the listener's expressive cue weightings are on the right. Assuming that performers and listeners are consistent, the degree to which these cues match determines the likelihood of successful communication.

Figure 2 shows the concordance of expressive cues used by performers and listeners when communicating an angry tone of voice. The left side of the figure shows the cue weights used by the performers. Weights are actually correlations between the performers intended affect and the utilized cues (Juslin, 2000). In this example, performers primarily rely on timbre cues to express anger (measured as high spectrum energy). The other cues used to encode an angry tone of voice included loud dynamics, somewhat faster tempos, and small articulation variability. The right side of figure 2 indicates that listeners utilize these same cues, but with slightly different weights. The

relatively high correlation of cues on both sides of the modified lens model suggests that performers were able to reliably communicate anger to listeners.

Unlike speech-based communication, musical communication in the Western art-music tradition commonly takes place twice: once between a composer and a performer, and then again between a performer and a listener. This complicates the communication model, especially considering that the context of the composer can be very far removed from the listener. Kendall and Carterette (1990) proposed a three part musical communication model which acknowledges the role of the composer, as well as the importance of composer, performer, and listener contexts. This model is shown in figure 3.

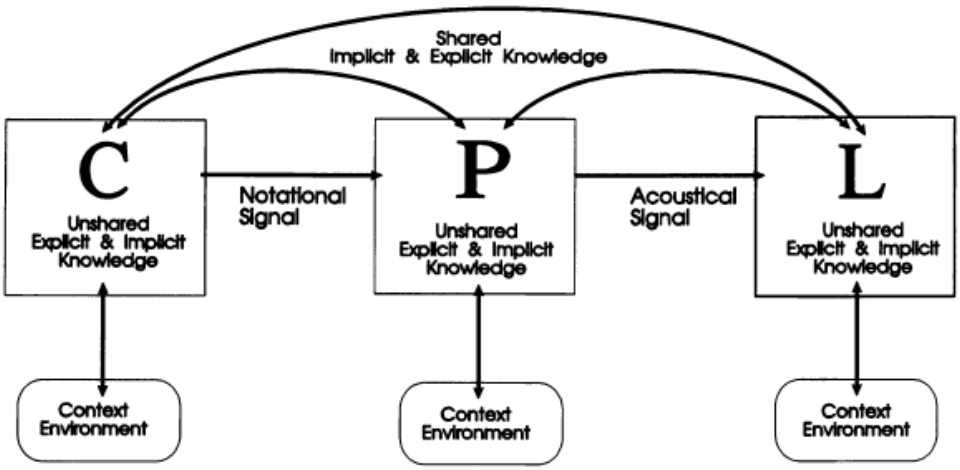


Figure 3: Kendall and Carterette's (1990) Three-Part Model of Musical Communication. C= Composer, P=Performer, L=Listener. Compared with Juslin's "modified lens model," this model adds a composer component. Contextual factors are also included for each of the three components.

Juslin and Lindström (2003) later proposed a three-part model: the "extended lens model" that also included the affective cues used by the composer. Unfortunately, the "extended lens model" is quite complex. In addition to accounting for the cues and consistency of composers, performers, and listeners, it also attempts to account for interactions between the composer and performer cues.

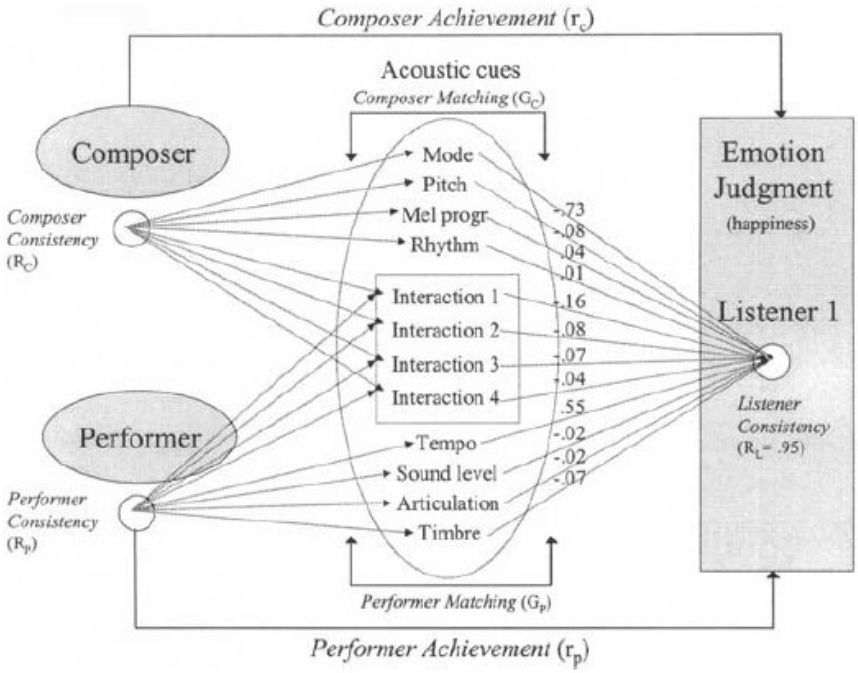


Figure 4: Juslin and Lindström's (2003) Extended Model Lens Model (adapted from Juslin & Laukka, 2004). The model accounts for cues used by composers, performers, and listeners. In this model, the interaction between composer and performer cues is also considered.

By way of review, these models are based on the idea that successful communication stems from the degree to which the sender and receiver possess the same "acoustical code" for a given message. Context and consistency are also key variables. Research has found evidence that acoustical cues used to express affective "tones of

voice" in a musical context are similar to those used in speech contexts (Scherer & Oshinsky, 1977). This finding continues to inspire new research questions, such as: What affects are capable of being expressed in speech that have not been explored by music researchers? What affects have communicative value that might be preserved in a musical context? What affective tones of voice do we hear on a regular basis?

Irony and Sarcasm

Many of the questions above have led to the formation of this study—an investigation of a sarcastic tone of voice in instrumental music. Recent empirical evidence suggests that irony and sarcasm are prevalent in everyday life. Two separate studies found that 7-8% of conversational utterances involve some form of irony (Tannen, 1984; Gibbs, 2000). Gibbs (2000) was able to further break down these findings by type of irony. He found that 50% of ironic statements could be classified as jocularity, 28% as sarcasm, 12% as hyperbole, 8% as rhetorical questions, and 2% as understatements (Gibbs, 2000, p. 14). The frequency of irony and sarcasm in speech has inspired research pertaining to an ironic tone of voice. With regards to music, musicology research has shown some interest in irony (Dill, 1989; Kramer, 1990; Bonds, 1991; Abbate, 1991; Sheinberg, 2000; Brackett, 2002). However, with only a few exceptions (e.g. Fónagy & Magdics, 1963), the topic has been largely ignored by music theory and music cognition researchers.

Irony can be expressed in both textual and spoken forms. There is evidence that irony has been used in writing for hundreds of years (Booth, 1974). Irony is defined as

"the use of words to express something other than and especially opposite the literal meaning" (Merriam-Webster Online, 2011). In text, irony can exist in the form of metaphors, allegories, fables, puns, satire, etc. (Booth, 1974). Many scholars consider sarcasm to be a type of irony (Gibbs, 1994; Haiman, 1998). Specifically, sarcasm can be considered a negatively valenced form of irony typically directed towards an individual (Gibbs, 1986). The word sarcasm can be traced to the Greek word *sarazein*, "meaning to speak bitterly, to tear flesh like dogs" (Lee & Katz, 1998). In order to understand the distinction between irony and sarcasm, consider the following sentence: "Wow! Did you finish the race yet?" If this sentence was said to the winner of the Boston marathon, the sentence would be ironic. If the sentence were said to the last place finisher of the Boston marathon, the sentence could be considered sarcastic. One distinction between irony and sarcasm is whether or not it is used as a form of criticism (Gibbs, 1986), suggesting that one of sarcasm's functions is to express an insult or an insincere form of praise.

One of the main obstacles a study of musical sarcasm (or irony) must overcome is that, unlike language, Western art-music is not typically denotative and so expressing an "opposite literal meaning" is impossible. For example, one cannot say that the pitch F#3 unambiguously indicates the color black, and that playing F#3 with some sort of sarcastic tone therefore unambiguously indicates the color white. However, it has been suggested that a concrete semantic meaning is not needed to detect the presence of a sarcastic tone of voice. The Standard Pragmatic Model view of sarcasm claims that an utterance must first be literally understood before its non-literal meaning can be extracted (Grice,

1975/1978; Searle, 1979). Gibbs (1986) argued that "if this were true, then sarcastic statements would take longer to process than non-sarcastic statements" (p. 3). Gibbs (ibid.) tested this hypothesis empirically and found that sarcastic statements do not take longer to process. As long as listeners were "given adequate information about the speaker's intentions," (typically in the form of context clues) they were able to comprehend the meaning of sarcastic utterances more or less directly. With regards to a sarcastic tone of voice in music, Gibb's (1986) study suggests that a concrete semantic message is not necessary to detect the presence of sarcasm.

> There are instances of sarcasm in which the literal meaning has nothing whatsoever to do with the speaker's intentions. The literal meaning of the sarcastic indirect request "Why don't you take your time getting the ball?" (meaning "Hurry up and get the ball") is just a simple literal question and is not related to the speaker's intention to use this utterance sarcastically as a request" (Gibbs, 1983, as cited in Gibbs, 1986).

Gibbs (1986) found evidence that emphasized the importance of context in detecting sarcasm. Other studies have found that sarcasm is more prevalent between speakers who share a common context (Gibbs, 2000; Caucci, Kreuz, & Buder, 2007). These results suggest sarcasm is more likely to be used among family and friends. A similar phenomenon occurs in textual irony. Texts are read by a variety of readers with varying degrees of familiarity with the author. It has been argued that a reader needs to

reach some degree of familiarity with a writer's style before they can begin to detect ironic or sarcastic statements (Booth, 1974). A parallel phenomenon might exist for a sarcastic tone in music—listeners may need to understand the context of the composition or performance in order to detect sarcasm.

In his theory of sarcasm, Gibbs (1986) highlighted the importance of socially normative behavior. Consider that the utterance, "You're a fine friend" could be spoken literally or sarcastically. However, the phrase "You're a bad friend" typically would not be spoken with a sarcastic tone in order to literally mean "You're a good friend." Support for the social normative component of sarcasm comes from the finding that sarcasm which follows socials norms is recognized faster than non-normative sarcasm (Gibbs, 1986). A musical parallel to Gibb's Social Norm Model might be found in Huron's (2004) study of musical humor. This study found that the loudest laughter in response to humorous music was associated with "mixed genres." Specifically, interjecting so-called low-style music (e.g. bagpipe music) into high-style music (e.g. string quartet music) was significantly more effective in evoking laughter than the inverse (Huron, 2004). It seems that responses to both sarcasm and musical humor are asymmetrical and depend on social context and norms.

Prosodic cues of sarcasm

Many affects have distinct acoustical cues. For example, empirical work has found that the acoustical cues of a sad "tone of voice" in a musical context include: lower pitch, softer amplitude, slower speaking rate, darker timbre, less articulation, and less

pitch variability (Huron, 2008; Schutz, Huron, Keeton, & Loewer, 2008; Turner & Huron, 2008; Post & Huron, 2009; Huron & Davis, 2010; Huron, Yim, & Chordia, 2010). Further, a meta-analysis performed by Juslin & Laukka (2003) on the findings from 41 different studies found that the following musical affects were expressed through distinctive pitch, timing, timbre, and dynamic cues: happiness, anger, fear, sadness, and tenderness.

There is some debate about whether or not sarcasm requires, or even possesses, a unique tone of voice (Bryant & Fox Tree, 2005). Several findings on the prosodic cues of sarcasm are contradictory. Gibbs (2000) noted that:

> "Although speakers need not convey their ironic messages in a special tone of voice, the speaker's intonation is often seen as an important clue to ironic meaning. One difficulty with this conclusion is that there appears not to be a single pattern of prosodic cues when people speak ironically." (p.25)

Table 1 organizes and presents the findings on the acoustical cues of spoken sarcasm. The source for each finding has been coded according to the following scheme: ((First two letters of the first author's last name)(last two digits of the publication year)).

Type of Acoustical Cue	Sarcasm-specific prosodic cue
Fundamental Frequency	Relatively Lower F_0 (Ro00; Ro07; Ch08); Relatively Higher F_0 (Mi97; Br05); Reduced F_0 range (Fo71; Ro07;Ch08); Exaggerated F_0 range (Ad, 96; Ha98; Ro04)
Intensity	Greater Intensity (Ha98; Ro00; Ro07); Softer Intensity (Ca07); Reduced Amplitude Variability (Br05)
Timbre	Unique Resonance Profile (Ch08); Nasalization (Cu74; Sc82; Kr95; Ha98)
Articulation	Heavily/Erratically Stressed (Cu74; Kr95; Ro04); More Pauses (Ca08)
Timing	Slower (Cu74; Kr 95; Ro00; Ca07; Ro07; Ch08; Br10)

Table 1: Empirical and theoretical findings on the acoustical cues of speech-based sarcasm. See the text above for an explanation of the abbreviated citations. Note that some of the findings are contradictory (e.g. pitch cues).

Despite several mixed findings, a sarcastic tone of voice might be loosely considered to exhibit the following acoustical cues: relatively lower F_0, greater intensity, a nasalized timbre, an irregularly stressed rhythm, and a slower speaking rate. Many of these cues, such as nasal timbre, lower F_0, and irregular stress might be associated with negatively valenced affects in general. Despite the idea that "little music is created with the overt intention of evoking negative valence emotions" (Huron, unpublished manuscript) it is hypothesized that a sarcastic tone of voice can be conveyed in music, and further, that musical sarcasm utilizes similar cues to sarcastic speech.

By way of review, sarcasm can be considered a form of irony. Sarcasm tends to be a bitter or caustic form of irony and is often directed towards someone or something. Contextual cues are an important part of perceiving sarcasm, and have been found to be more important than understanding the sender's literal meaning. Sarcasm tends to be

applied to socially normative contexts and it is more prevalent in conversations among close acquaintances. A sarcastic "tone of voice" has been empirically studied in speech research. Although the findings are somewhat mixed, sarcasm loosely exhibits the following auditory characteristics: a relatively lower F_0, greater intensity, a nasal timbre, an irregularly stressed rhythm, and a slower speaking rate.

Hypotheses

A systematic empirical investigation of musical sarcasm has not been previously attempted. The results might be of interest to both musicians and speech researchers. In many ways, the non-denotative nature of instrumental music makes it an ideal candidate for studying what linguists refer to as "tone of voice." Leonard Meyer (1967) said,

> "The mere collection and counting of phenomena do not lead to significant concepts. Behind any statistical investigation must be hypotheses that determine which facts shall be collected and counted." (Meyer, 1967).

Even though this study was exploratory in nature, it was guided by the following hypotheses:

Hypothesis 1: The cues used to convey sarcasm in a musical setting will be similar to the cues used to convey sarcasm in speech.

Hypothesis 2: A sarcastic tone of voice can be reliably communicated within a musical context.

The first hypothesis is examined in chapters 3 and 4 through the use of Fairbanks' (1940) standard paradigm. Two models (Performer and Listener) are devised from this paradigm and subsequently compared to the literature on sarcastic speech. The second hypothesis is examined in chapter 5 by investigating the concordance between the Performer and Listener model. An additional study is presented in chapter 6 that examined the role of context in detecting a sarcastic tone of voice in instrumental music.

Chapter 3: Performer Model of Musical Sarcasm

Chapter 3 presents the first half of a "standard paradigm" study that investigated the cues of a sarcastic tone of voice within an instrumental music context.

Background

In order to investigate the relationship between prosodic features of speech and perceived affect, Fairbanks (1940) devised a methodology which would come to be known as the "standard paradigm" (Juslin & Timmers, 2010). Fairbanks recorded subjects speaking a single text five different times, each time expressing a different affect (contempt, anger, fear, grief, and indifference). These recordings were validated by independent judges and then presented to a larger set of listeners, who were tasked with identifying the speaker's intended affect from a list of 12 possible choices. The results of the study were reported as a confusion matrix. Fairbanks also analyzed the pitch levels for the most and least successfully identified recordings. The results suggested that pitch profiles can be used to discriminate between certain affects in speech.

Carl Seashore (1947) suggested that this same methodology could be applied in order to investigate musical affect. The real advantage of such an approach is that it can be "successful in spite of the comparative looseness in definitions of the emotions studied" (Seashore, 1947, p.176). Seashore further argued, "by this approach, so well

illustrated in speech, we can begin to lay foundations for a science of the expression of specific emotions in song" (Seashore, 1947, p.176). Frick (1985) added the important caveat that this "paradigm cannot be used to show that an emotion, such as anger, can be recognized. Instead, it shows only that the emotion can be discriminated from other emotions" (p.413). Therefore, results for "standard paradigm" studies must be interpreted within the context of the affects that were employed.

Methodology

Fourteen musicians were recruited to record various types of musical excerpts while expressing different tones of voice. The musicians played a variety of musical instruments: alto saxophone, Bb clarinet, double bass, bassoon, flute, oboe, piano, soprano saxophone, trombone, trumpet, tenor saxophone, viola, cello, and violin. Prior to each recording session, participants were sent a page of instructions which contained details on the types of musical material they would be asked to record (a copy of these instructions can be found in Appendix A). All performers were given at least one week to rehearse the material and to ask any questions about the study. Performers were asked to play five different types of musical excerpts: 9 repeated notes, an ascending major 3rd, a descending major 3rd, the song "Happy Birthday," and a 10 second improvisation. With the exception of the 10 second improvisation, all the musical excerpts were notated. Performers recorded each type of musical excerpt with the intention of conveying the following four tones of voice: joy, sadness, sarcasm, and sincerity. None of the performers asked for clarification on any of the target tones of voice. Table 2 summarizes

the 20 recorded excerpts. Performers were allowed to record the 20 excerpts in whatever order they preferred.

Happy Birthday	**Ascending M3**	**Descending M3**	**9 Repeat Notes**	**Improvisation**
Joy	Joy	Joy	Joy	Joy
Sadness	Sadness	Sadness	Sadness	Sadness
Sincerity	Sincerity	Sincerity	Sincerity	Sincerity
Sarcasm	Sarcasm	Sarcasm	Sarcasm	Sarcasm

Table 2: Summary of the recording session. Each performer recorded 5 different types of stimuli 4 separate times in order to convey the following tones of voice: joy, sadness, sarcasm, and sincerity. Each session resulted in 20 recordings.

The five types of musical excerpts were chosen to obtain a range of both highly controlled and open responses. The choice of a single repeated tone was motivated by the goal of identifying acoustic correlates apart from changing pitch. Participants would therefore be forced to use non-pitch manipulations—such as timbre or timing—in order to convey the target affect. Previous literature has found that expressive timing is an important factor in conveying affective information (Bengtsson & Gabrielsson, 1980; Repp, 1992; Gabrielson & Juslin, 1996; Juslin & Madison, 1999; Bhatara, et al. 2011). Ascending and descending intervals were chosen to isolate tone of voice cues on a single pitch transition. The tune "Happy Birthday" was chosen because it has been hypothesized that sarcasm sometimes results from a form of parody or something familiar to the listener (Haiman, 1998). Finally, by allowing performers to improvise a passage for each tone of voice, they were free to create and manipulate any musical characteristic

they wished. Also, by giving performers license to do whatever they wanted it was possible to increase the opportunity of identifying acoustic features that are distinctive of different tones of voice, and also to permit possible interactions between different acoustic features.

The stimuli were recorded in an Industrial Acoustics Corporation sound isolation room using an AKG C1000S stereo microphone. The sound was recorded directly to a compact disc via a Sony CDR-W33 A/D recording device.

Participants were allowed to record the stimuli in any order. No attempt was made to explain the various tones of voice to the performers. Performers were allowed to re-record each stimulus as many times as necessary until they felt they had effectively communicated the intended tone of voice. The instrumentalists themselves chose the recorded passage they deemed to be the best exemplar for each of the four affective tones of voice. In several cases, musicians asked to hear their recordings before deciding whether or not to re-record the stimulus. Throughout the recording, the communicative goal was emphasized: the idea was not for the performer to think their performances were emblematic of different affects, but rather that listeners should clearly recognize the intended tone of voice. On average, the recording session lasted approximately 30 minutes. Afterwards, each musician was given an informal interview to probe their knowledge of sarcasm in musical contexts.

In total, 280 stimuli were recorded (14 musicians x 5 stimuli x 4 affective attitudes). Each recording was analyzed with PRAAT speech analysis software (Boersmam & Weenink, 2010). A PRAAT "voice report" was generated for each

musical sample[4]. An example of a "voice report" is provided in Appendix B. These reports included measurements of excerpt duration, median pitch, mean pitch, pitch standard deviation, minimum pitch, maximum pitch, number of pulses, number of periods, mean period, standard deviation of period, fraction of locally unvoiced frames, number of voice breaks, degree of voice breaks, jitter (local), jitter (local, absolute), jitter (rap), jitter (ppq5), jitter (ddp), shimmer (local), shimmer (local, dB), shimmer (apq3), shimmer (apq5), shimmer (apq11), shimmer (dda), mean autocorrelation, mean noise-to-harmonics ratio, and mean harmonics-to-noise ratio. Explanations of these auditory characteristics are provided below.

Results

General Model

The collection of voice reports generated from the performer recordings were organized by the performer's intended tone of voice and fit to a statistical model. Because this study was exploratory in nature and interested in determining the most important factors in distinguishing between the four affective tones of voice, the dataset was fit using a multinomial logit model (Menard, 2002). The optimized model found four significant musical characteristics used by performers in order to distinguish between the four tones of voice. These features were <u>duration</u>, <u>maximum pitch</u>, <u>fraction of locally unvoiced frames</u>[5], and the <u>number of voice breaks</u>. Combined, these four

[4] Voice reports were generated using PRAAT's default settings.
[5] The fraction of locally unvoiced frames can also be an indicator of timbral quality. A high level of unvoiced frames would be indicative of a "noisy" timbre.

factors accounted for approximately 16.8% of the variance between the four "tones of voice" recorded by performers, and further, this model was found to be statistically significant ($\chi^2 = 130.57$; $p<.001$).[6]

Analysis of Variance

In order to investigate how each of the model's four significant factors were used to convey the four target tones of voice, additional statistical tests were performed. Each of the four factors (i.e. musical cues) was subjected to a one-way analysis of variance (ANOVA) test. Three of the four factors—duration, fraction of locally unvoiced frames, and number of voice breaks—were found to significantly distinguish sarcasm from at least one of the other tones of voice. These results are shown in table 3 and discussed below.

[6] The coefficients (cue weights) for the Performer model can be found in Appendix C.

Multinomial Logit Results	ANOVA Results		
Factor	Tone of Voice	Mean	
Duration	Sad	11.09s	*
	Sincerity	8.36s	
	Sarcasm	6.53s	
	Joy	6.15s	
Maximum Pitch	Joy	396 Hz	
	Sadness	370 Hz	
	Sarcasm	369 Hz	
	Sincerity	369 Hz	
Unvoiced Frames	Sarcasm	28.6%	
	Joy	12.7%	*
	Sadness	11.0%	*
	Sincerity	9.8%	*
Voice Breaks	Sarcasm	7.31 breaks	
	Joy	5.42 breaks	
	Sincerity	4.95 breaks	
	Sadness	3.61 breaks	*

Table 3: Results from the multinomial logit model and the associated ANOVA tests. The mean values for each variable are provided as a function of the target tone of voice. The asterisk (*) indicates a mean which was found to be significantly different from the mean for sarcasm at the p=.05 level. Significant differences between the other tones of voice are not reported here, but can be found in appendix D.

Duration

The first significant factor in the Performer model was total duration. PRAAT measures duration as the time from the onset of the first note until the offset of the last note. While this measurement might be meaningful for speech studies, PRAAT's definition of duration is perhaps less meaningful in a musical context. If all the musical

excerpts used in this study were identical, then PRAAT's measurement of duration would be a reasonable indicator of musical tempo. However, bear in mind that the dataset contained improvisations in which performers were asked to improvise for no longer than 10 seconds. Some of the improvisations contained many notes, whereas others contained only a few. There was quite a bit of variability in the length of the improvisation recordings (M=9.4s, SD=3.7). The means and SD's of the durations for the four target tones of voice were: *joy* (M=6.2s, SD=5.1); *sarcasm* (M=6.5s, SD=6.1); *sincerity* (M=8.4, SD=6.4); and *sadness* (M=11.1s, SD=9.0).

An analysis of variance revealed a significant difference between the durations of the four tones of voice, $F(3, 276) = 7.68$, $p < .001$. The Tukey's honestly significant difference (HSD) post hoc test revealed a significant difference between the durations used to express *sarcasm* and *sadness*.

Maximum Pitch

The second factor in the multinomial logit model was "maximum pitch." Maximum pitch should not be confused with mean pitch (which was not a significant factor in the Performer model). All the pitches used in the study, with the exception of the improvisations, were notated as fixed pitches (see appendix A). Not all instruments were capable of playing the passages at the same pitch level. For example, the "9 repeated-note" excerpt expressing joy was recorded by the bass using the pitch A2, whereas the flautist recorded this passage using the pitch A5. Therefore, analyzing the differences in maximum pitch as a function of the 14 instruments would not yield

meaningful results. However, it was possible to analyze the coded differences in maximum pitch as a function of intended affect—comparing the means of all the recordings which expressed a given tone of voice—since each instrument (and tone of voice) was equally represented in the data. The main disadvantage to this type of analysis was that different pitch registers of each instrument resulted in a high level of variability. An alternative analysis might have examined differences in pitch for different affects within the same instrument, but this was not done.

The average maximum pitch for each of the four tones of voice was: *joy* (396 Hz, SD=97); *sadness* (370 Hz, SD=123); *sarcasm* (369 Hz, SD=120); *sincerity* (369 Hz, SD=119). An analysis of variance did not reveal a significant difference between the maximum pitch of the four tones of voice, $F(3, 276) = 0.95$, $p=.41$.

Fraction of locally unvoiced frames

The third factor revealed by the multinomial logit model was the "fraction of locally unvoiced frames." According to the PRAAT user manual (Boersma & Weenink, 2010), the fraction of locally unvoiced frames is:

> "the fraction of pitch frames that are [sic] analysed as unvoiced. If the pitch floor is 75 Hz … there will be 100 pitch frames. If 86 of these are locally voiced, the fraction will be 14 percent."

Measurements of locally unvoiced frames are used to detect pathologies in the voice; if more than 20% of the frames are unvoiced, the speaker might sound like they have a cold (Boersma & Weenink, 2010). The fraction of locally unvoiced frames can also be an indicator of timbral quality. A high level of unvoiced frames would be indicative of a "noisy" timbre. The "fraction of locally unvoiced frames" means and SD's for each of the four affective tones of voice were: *sarcasm* (M=28.6%, SD=19.6); *joy* (M=12.7%, SD=10.5); *sadness* (M=11.0%, SD=12.2); and *sadness* (M=9.8%, SD=9.0).

An analysis of variance revealed a significant difference between the "fraction of locally unvoiced frames" for the four tones of voice, $F(3, 276) = 30.10$, $p < .001$. The Tukey's HSD post hoc test revealed a significant difference between the "fraction of locally unvoiced frames" used to express *sarcasm* and *joy, sarcasm* and *sadness,* as well as *sarcasm* and *sincerity.*

Number of voice breaks

The fourth factor found by the multinomial logit model was the "number of voice breaks." In a PRAAT voice analysis, "All inter-pulse intervals longer than 16.7 milliseconds are regarded as voice breaks" (Boersma & Weenink, 2010). Within a musical context, this factor could be thought of as a measurement of the number of staccato articulations or dramatic pauses used by performers. Much like the caveat that applied to the measurements of duration, the improvisation excerpts may have caused errors regarding the true number of voice breaks used to convey each of the four target tones of voice. This caveat notwithstanding, the mean number of voice breaks for each

of the four target "tones of voice" was as follows: *sarcasm* (M=7.3 breaks, SD=7.9); *joy* (M=5.4 breaks, SD=6.7); *sincerity* (4.95 breaks, SD=7.0); and *sadness* (M=3.6 breaks, SD=5.8).

An analysis of variance revealed a significant difference between the "mean number of voice breaks" for the four tones of voice, $F(3, 276) = 3.45$, $p<.05$. Tukey's HSD post hoc test revealed a significant difference between the "mean number of voice breaks" used to express *sarcasm* and *sadness*.

In summary, the Performer model found four cues which were used to distinguish between the four target tones of voice. These cues included duration, maximum pitch, fraction of locally unvoiced frames, and mean number of voice breaks. Further, performers used the following cues to distinguish a sarcastic tone from the other target tones of voice: relatively shorter durations, a high level of unvoiced frames, and a relatively greater number of voice breaks. The analysis for the "maximum pitch" factor of the model was found to be inconclusive.

Additional Analyses

Although the PRAAT voice reports were capable of quantifying many aspects of the recordings, they were not comprehensive. Several studies have found that expressive performance is largely due to timing variations (Bengtsson & Gabrielsson, 1980; Gabrielson & Juslin, 1996; Juslin & Madison, 1999). Kendall and Carterette (1990) had participants identify musical affect from synthesized musical samples, and found that "timings alone, with reasonably natural signals, seem to be sufficient for categorization

accuracy that is nearly equal to that obtained with real instruments" (pp.155-156). Four potentially important measurements which were not measured by the PRAAT voice reports included: global tempo, articulation, note intensity, and inter-onset interval variability. Additional analyses for these four measurements were performed on a small subset of the data, namely the "9 repeated note recordings." This subset consisted of 56 recordings (14 performers x 4 target tones of voice). These results for these four additional analyses are discussed below.

Global Tempo

An important cue for encoding affect is global tempo. For example, one of the cues for sad music is a relatively slower global tempo (Post & Huron, 2009). The term global in this context implies a sense of long-term regularity[7]. In order to quantify the global tempo for each of the "9 repeated note" recordings, precise measurements of the length of time between note onsets were made using PRAAT. This measurement is often referred to as the inter-onset interval (IOI).

The notation supplied to the performers (see appendix A) consisted of eight 8th notes followed by a half note in common time. For a musical passage of 9 notes, 8 measurements of IOI can be made. In order to estimate the global tempo of each performance, the 8 IOI measurements were averaged together to calculate the recordings mean IOI. Global tempo was then computed with the following formula:

[7] On the order of a musical phrase length or longer

$$\text{Tempo (bpm)} = 60/(\text{Mean IOI} * 2). \qquad (1)$$

Note that in the above formula, the mean IOI had to be doubled because the 8 repeated notes were subdivisions of the beat (8th notes in a measure of common time). After global tempo had been calculated for all 56 recordings, the data were subjected to a one way analysis of variance (ANOVA) test. This test found a significant effect of the target tone of voice on global tempo, $F(3, 52) = 8.20$, $p < .001$. Post hoc comparisons using the Tukey HSD test indicated that the global tempo for *sarcasm* (M = 90.6 bpm, SD = 40.4) was significantly different from *sadness* (M = 47.2 bpm, SD = 17.5). The difference between *sarcasm* and *sincerity* approached, but did not reach statistical significance (M = 62.1 bpm, SD = 20.4, p=.10). There was not a significant difference between the global tempo for *sarcasm* and *joy* (M = 100.1 bpm, SD = 42.0). The results are depicted in figure 5, and suggest that performers use a relatively faster tempo to distinguish sarcasm from sadness, and to a lesser extent, sarcasm from sincerity.

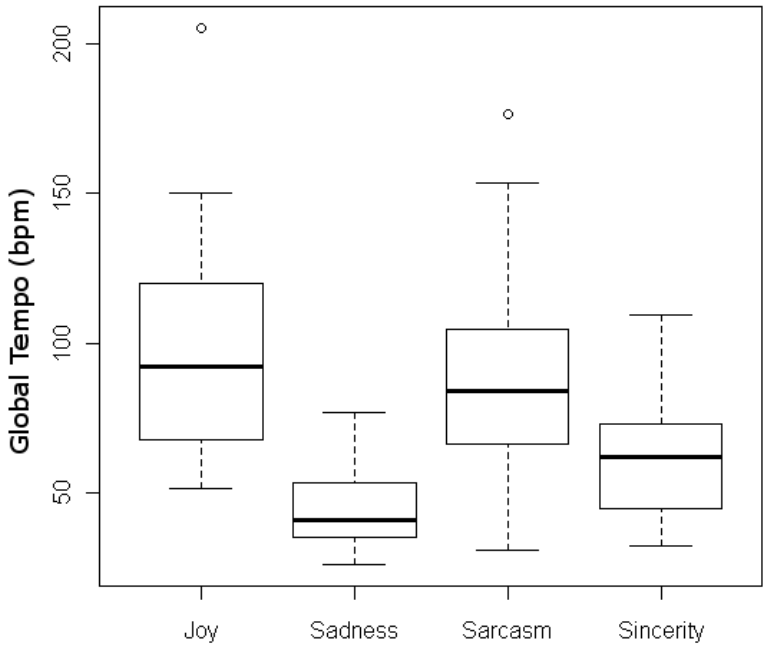


Figure 5: Global Tempo (in beats per minute) as a function of tone of voice. For each tone of voice, the bold black line indicates the median value, the box indicates the upper and lower quartiles (the middle 50% of the data), the lines extending from the box indicate the highest and lowest values when excluding outliers, and the circles indicate outliers which were more than 3/2 times greater or lesser than the upper and (cont.) (cont.) lower quartile. The global tempo of sarcastic recordings was found to be significantly faster than sad recordings.

Articulation

The articulation of a musical note can be described and measured in many ways. One particularly simple method of measuring musical articulation is described in Bengtsson & Gabrielsson (1980). Their methodology involves making two measurements: the length of time between two consecutive note onsets (d_{ii}), and also the length of time between the onset and offset of a note (d_{io}). Articulation is calculated by

dividing d_{io} by d_{ii}. The resulting score will always be a value between 0 to 1, with higher scores being more legato-type articulations and lower scores being more staccato-type articulations. For example, if a note had a d_{io} of 750ms and a d_{ii} of 1000ms, the articulation score would be .750, indicating a more legato-type articulation.

In order to investigate the articulation of sarcastic tone of voice in music, articulation scores were calculated for every note from the collection of "9 repeated note" recordings using the same formula as Bengtsson & Gabrielsson (1980). PRAAT's graphical waveform editor was used to make the most accurate measurements possible. Measurements were recorded to the nearest millisecond. It was not possible to calculate the articulation score for the final note of each recording (i.e. a posterior articulation score). The articulation scores for each recording were averaged together to calculate the "mean articulation score" for each recording. These mean articulation scores were subjected to a one way analysis of variance (ANOVA) test. This test found a significant effect of the target tone of voice on articulation, $F(3, 52) = 8.9328$, $p<.001$. Post hoc comparisons using the Tukey HSD test indicated that the mean articulation scores used to express *sarcasm* (M = 0.70, SD = 0.21) were significantly different from those used to express *sadness* (M =0.93 , SD = 0.10) and *sincerity* (M = 0.93, SD = 0.05). However, mean articulation scores used to express *sarcasm* were not significantly different from those used to express *joy* (M = .80, SD = 0.16). These results are shown in figure 6, and suggest that performers use more staccato-like articulations for expressing sarcasm than for expressing sadness or sincerity.

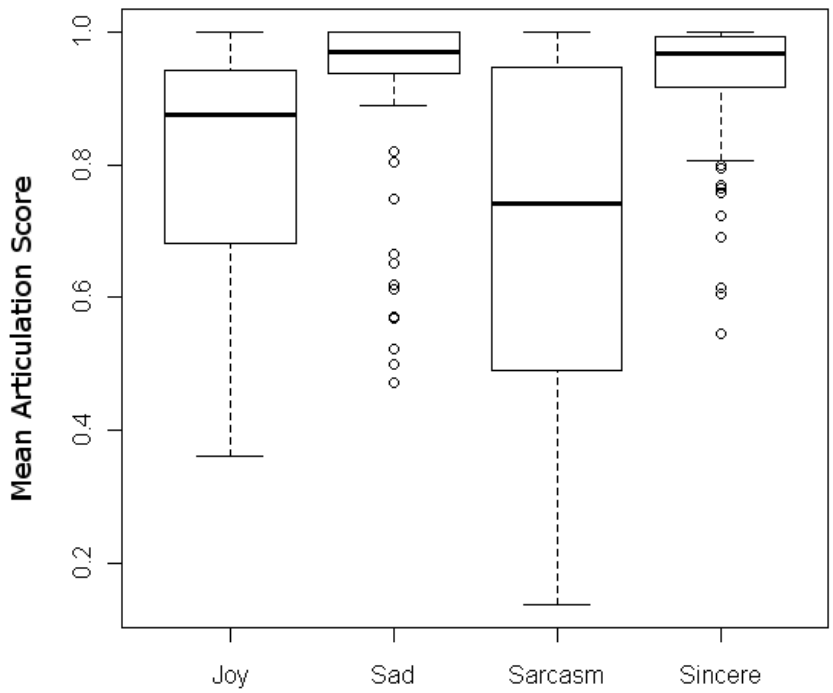


Figure 6: "Mean Articulation Scores" (d_{io}/d_{ii}) as a function of target "tone of voice." For each "tone of voice," the bold black line indicates the median value, the box indicates the upper and lower quartiles (the middle 50% of the data), the lines extending from the box indicate the highest and lowest values when excluding outliers, and the circles (cont.) (cont.) indicate outliers which were more than 3/2 times greater or lesser than the upper and lower quartile. Articulation scores for sarcastic notes were significantly smaller than sad notes and sincere notes.

Figure 6 shows that there was a very large amount of variability in the "mean articulation scores" for expressing a sarcastic tone of voice. In order to probe for systematic variations of articulation, an additional analysis was performed to examine the change in articulation score from one note to the next (i.e. second order changes in articulation). By comparing the changes in articulation from note to note, it was possible

to see if performers were keeping their articulations consistent, or patterning articulations in some particular way. Each recording contained 8 articulation scores, and therefore, 7 second order changes in articulation. These values were grouped according to affect and then averaged. The results are presented in figure 7.

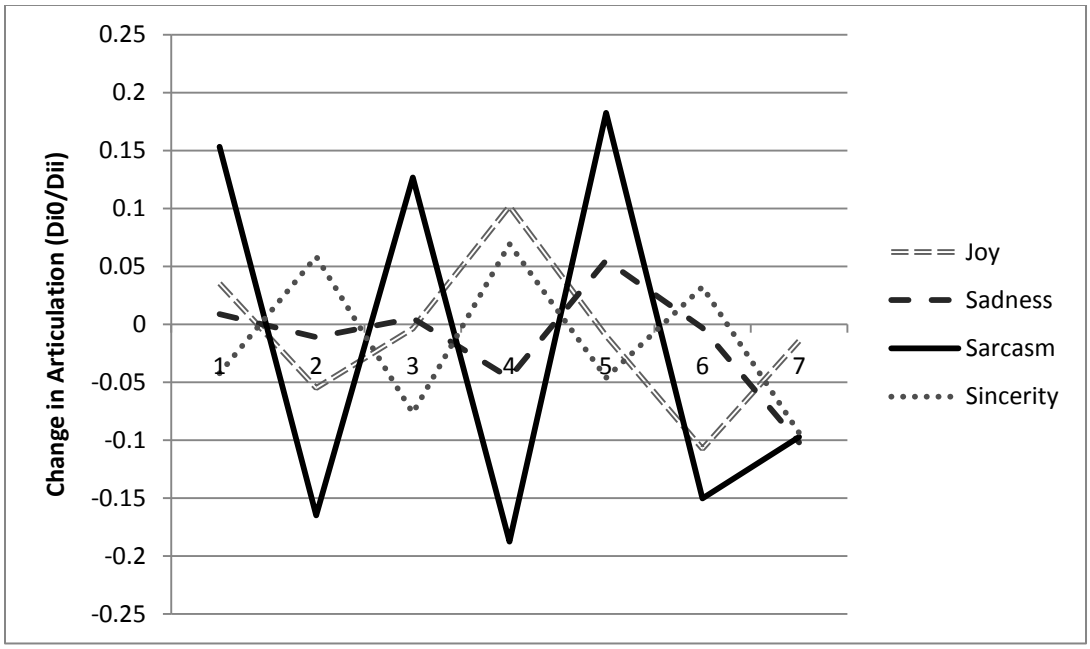


Figure 7: Change in articulation score from note to note in the "9 repeated note" samples as a function of target "tone of voice." Changes in articulation appear to be patterned. A rising line indicates that a more legato note followed a more staccato note, whereas a descending line indicates the opposite.

As can be seen in figure 7, recordings expressing a sarcastic tone of voice tended to exhibit a greater change in articulation between notes than did any other of the recorded affects. Specifically, recordings of sarcastic tones tended to systematically alternate between staccato and legato articulations. This pattern can also be seen in other tones of voice, which suggests that performers in general might systematically alternate

between longer and shorter articulations. However, this pattern is particularly pronounced when performers are expressing a sarcastic tone.

Intensity

The systematic manipulation of intensity to convey the target tones of voice was also analyzed. PRAAT was used to measure the maximum intensity (in dB) of each of the 9 notes from every recording. Measurements (in decibels) were expressed relative to the minimum and maximum representable digitized signals. Keeping the PRAAT default settings, the mean pressure was subtracted from each measurement. Even though the measurements were not calibrated to a standard, the differences between measurements can still be compared[8]. During the recording process, the gain value was set for each individual instrument so as to obtain a strong signal. Once this level was set, it remained constant for the recording session. Because each recording was made with a different gain setting, it was not possible to analyze the differences in maximum intensity as a function of the 14 instruments. However, because each instrument (and tone of voice) was equally represented in the data, it was possible to analyze the differences in average maximum intensity as a function of intended affect. The main disadvantage to this type of analysis was that many factors determine the intensity of a signal, including the size of each instrument, the distance between the sound source and the microphone, etc. These

[8] According to Paul Boersma, "a measurement of "70 dB" in Praat is still 20 dB stronger than a measurement of "50 dB" in the same recording (or even across recordings with the exact same set-up). It does not matter whether these sounds are in reality e.g. 60 and 40 dB, or 85 and 65 dB, or whatever: the *difference* in dB will be the same, whether the sound is calibrated or not." Taken from a forum the PRAAT user forum. *http://uk.dir.groups.yahoo.com/group/praat-users/message/2767*, accessed 5/28/11.

factors resulted in a high level of variability. Once again, an alternative analysis might have examined differences in intensity for different affects within the same instrument, but this was not done.

In total, 504 measurements were made (9 notes x 4 affects x 14 performers). The 9 intensity values were averaged for each recording, creating what may be considered an "average maximum intensity." These measurements were grouped by target "tones of voice" and submitted to a one-way ANOVA test The test found a significant effect of the target affect on intensity at the $p<.05$ level. $F(3, 52) = 7.7617$, $p<0.001$. Post hoc comparisons using the Tukey HSD test indicated that the mean score for *sarcasm* (M = 78.3dB, SD = 3.8) was significantly different from *sadness* M = 71.7dB, SD = 5.0). However, there was not a significant difference between *sarcasm* and *jo*y (M = 78.7dB, SD = 3.1) or *sincerity* (M = 76.7dB, SD =5.1). Results are depicted figure 8 and suggest that performers use a relatively greater intensity to distinguish sarcasm, joy, and sincerity from sadness.

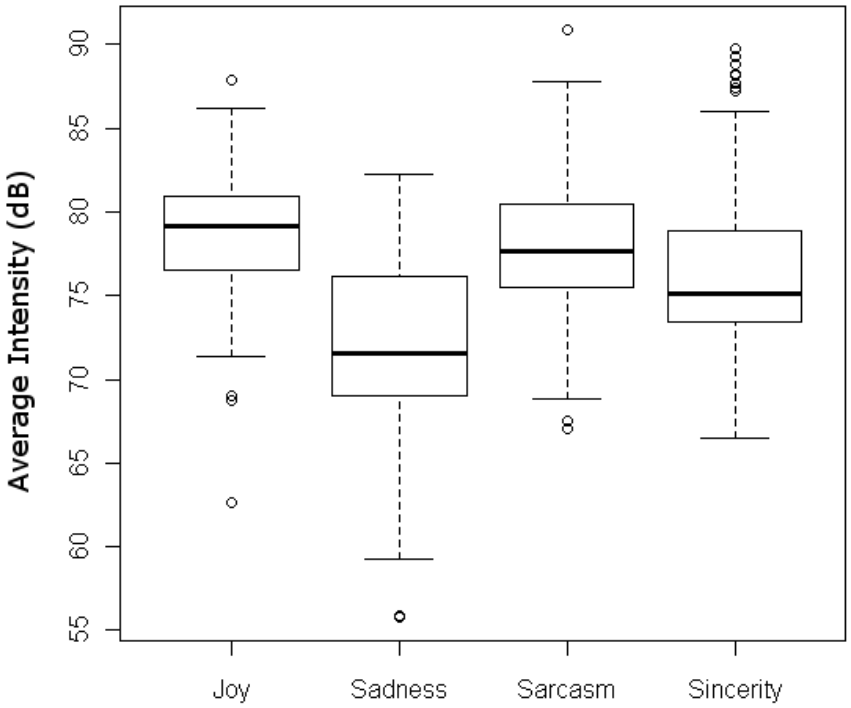


Figure 8: "Average Intensity" (in dB) as a function of the target tone of voice. For each tone of voice, the bold black line indicates the median value, the box indicates the upper and lower quartiles (the middle 50% of the data), the lines extending from the box indicate the highest and lowest values when excluding outliers, and the circles indicate outliers which were more than 3/2 times greater or lesser than the upper and lower quartile. Joy, sincerity, and sarcasm were found to have louder average intensities than sadness.

Inter-onset interval

PRAAT was also used to measure the amount of time between consecutive note onsets (IOI) for each note in the "9 repeated note" samples. Values were measured to the nearest millisecond using PRAAT's graphical waveform editor. The average of the 8 IOI measurements for each recording was calculated and may be considered a "mean IOI."

Mean IOI's were grouped by the target tone of voice and subjected to a one-way ANOVA. There was a significant effect of the target tone of voice on "mean IOI," $F(3, 52) = 9.8209$, $p<.001$. Post hoc comparisons using the Tukey HSD test indicated that *sarcastic* "mean IOI's" (M = 406ms, SD = 212) were significantly different from the sad "mean IOI's" (M = 704ms, SD = 237). However, *sarcastic* "mean IOI's" were not different from the "mean IOI's" of recordings expressing either *sincerity* (M = 533ms, SD = 177) or *joy* (M = 347ms, SD = 131). Results are depicted in figure 9 and suggest that performers use shorter IOI's for expressing sarcasm than for expressing sadness.

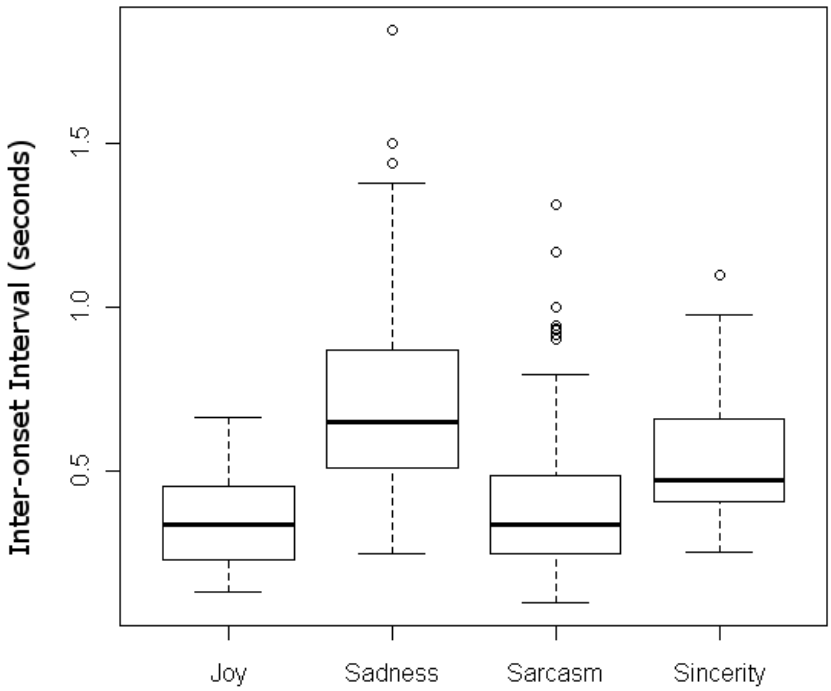


Figure 9: Inter-onset intervals (in seconds) between successive notes as a function of the target tone of voice. For each tone of voice, the bold black line indicates the median value, the box indicates the upper and lower (cont.)

(cont.) quartiles (the middle 50% of the data), the lines extending from the box indicate the highest and lowest values when excluding outliers, and the circles indicate outliers which were more than 3/2 times greater or lesser than the upper and lower quartile. Sarcastic recordings were found to have a shorter mean IOI than sad recordings.

The number of outliers in figure 9 suggested that performers may have been systematically varying their IOIs. Similar to the analysis that was performed on articulation score, the second order data on IOI was also analyzed. This information is depicted in Figure 10. Note that all of the IOIs increase towards the end of the excerpt. This is consistent with the findings of Repp (1992) who found that performers tend to slow down towards the end of musical phrases. It is also noteworthy that the sarcastic tone of voice had a lengthening of IOI between the 4th and 5th beats. This happens to occur on one of the relatively strong beats of the measure, and signals that expressing a sarcastic "tone of voice" might involve violating (and in this case delaying) metrical expectations.[9] Future research may wish to specifically test this hypothesis.

[9] See Huron (2006) for a discussion on violations of metrical expectation.

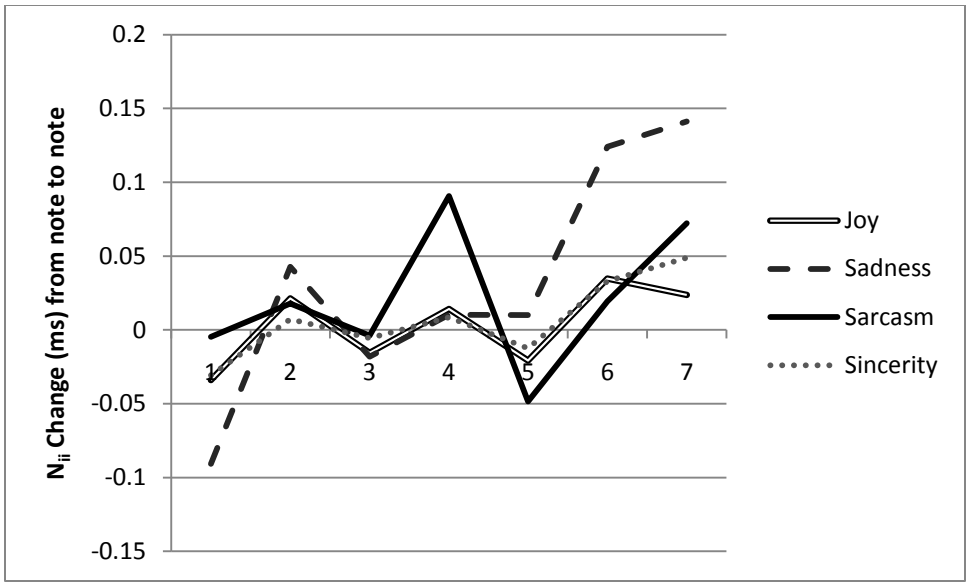


Figure 10: Change in IOI (in ms) from note to note in the "9 repeated recordings" as a function of the target "tone of voice." An upward sloping line indicates that performers increased IOI from one note to the next, whereas a downward sloping line indicates that performers decreased IOI.

Discussion

Through a variety of analytical tools, some of the acoustic cues used by performers to distinguish sarcasm from the other target tones of voice included: relatively shorter durations[10], a high level of unvoiced frames, more staccato-like articulations[11], and relatively greater intensity. Combining the cues from the multinomial logit model and the additional analysis yields what might be considered the "Performer Model" for expressing a sarcastic tone within a musical context. This model is depicted in figure 11.

[10] This was measured in two ways: PRAAT total duration, and mean IOI.
[11] This was also measured in two ways: "mean number of voice breaks" and the d_{io}/d_{ii} score.

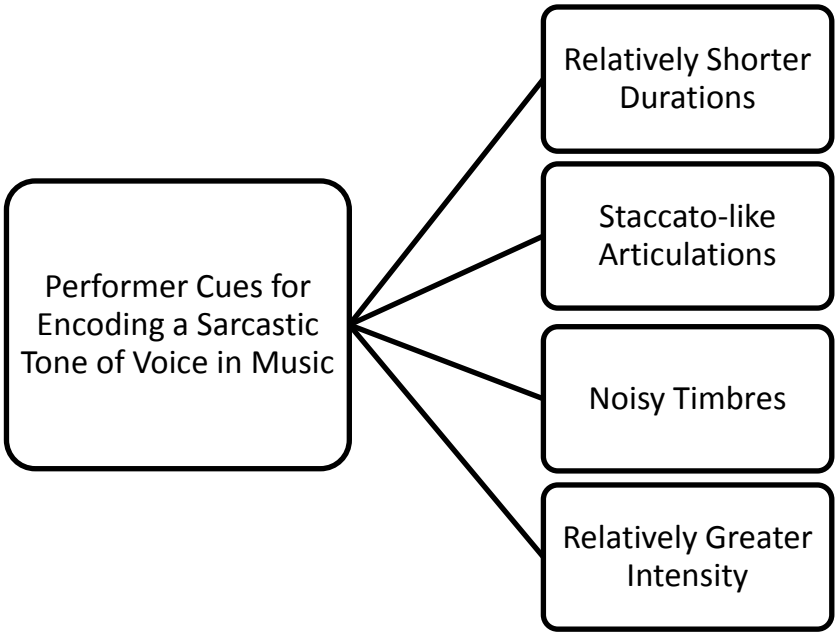


Figure 11: Performer's model for expressing a sarcastic tone in instrumental music.

Recall that the model shown in figure 11 does not imply that these cues are understood by listeners (concordance will be examined in chapter 5). Also note that this model represents a relatively small amount of performers (n=14). Further, it is likely that amateur and professional musicians would use different cues, or differing amounts of various cues, in order to distinguish between the target tones of voice. For example, one of the more experienced performers in this study used very subtle cues to express each of the target tones of voice, whereas many performers had a tendency to use obvious or extreme musical cues.

Now that the cues used by instrumentalists to encode a sarcastic tone of voice have been examined, it is possible to make a comparison to the acoustic cues of sarcastic speech. This comparison is provided in table 4.

Sarcastic Speech	Cue	Sarcastic Music
Inconclusive	Pitch/F_0	Inconclusive
Relatively Greater	Intensity	Relatively Greater
Nasalized	Timbre	High Level of Unvoiced Frames (Noisy)
Stressed, Erratic	Articulation	Staccato-like
Slower	Timing	Relatively Faster

Table 4: A comparison of the cues used by performers to convey a sarcastic tone of voice in music and the cues of sarcastic speech. In general, there seems to be more similarities between these two contexts of sarcasm than differences.

Table 4 shows that there are many similarities between sarcastic speech and a sarcastic tone of voice in instrumental music. These similarities include: a noisy or nasalized timbre, stressed articulations, and relatively greater intensity[12]. Some researchers have argued that sarcastic speech involves a distinct timbre (Cheang & Pell, 2008) whereas others point out that sarcastic utterances can be communicated with a dry tone of voice (Bryant & Fox Tree, 2005). In a musical context, not all instruments are capable of changing timbre, and further, not all instruments are capable of creating a nasal or noisy sound. Like spoken sarcasm (Cutler, 1974), most instruments are capable of employing stressed or erratic articulations. Not only did performers in this study use more staccato-articulations when expressing a sarcastic tone, they also seemed to pattern their articulations. Finally, the intensity of sarcastic speech tends to be greater than non-

[12] Although the intensity factor was found to be statistically significant, it is more a reflection of sadness having a low intensity than sarcasm having a greater intensity.

sarcastic speech (Rockwell, 2000); a parallel cue might be present in music. Intensity, like timbre, is strongly linked to the capacities of a musical instrument. However, further testing of intensity with additional affective tones of voice is needed.

A large amount of variability was expected in this study. The cues which a performer can manipulate in order to express varying tones of voice are highly dependent on the capabilities of their musical instrument. Despite the large amount of variability, the statistical model and follow-up tests were able to reveal several similarities between the cues used to encode sarcastic speech and a sarcastic tone of voice in instrumental music.

Chapter 4: Listener Model of Musical Sarcasm

Chapter 4 presents the second half of the "standard paradigm" study that investigated the cues of a sarcastic tone within instrumental music.

Methodology

Twenty-four participants were recruited from the Ohio State University School of Music subject pool; none were familiar with the purpose of the study[13]. These participants were asked to listen to the recordings from the previous study, and then identify the performer's intended tone of voice. Each listener heard one half of the total stimuli (140 recordings).

Participants were seated in an Industrial Acoustics sound isolation room. They were told that they would hear a series of passages played using various instruments, and that each passage conveyed one of the following tones of voice: sarcasm, sadness, sincerity, joy, tenderness, mysteriousness, anger, or neutrality. Their task was to determine which tone of voice the performer was attempting to convey. Note that the last four tones of voice were added as lures in order to increase the task difficulty. Other

[13] A copy of the posting for recruiting participants can be found in Appendix E.

studies have also used “lures” to make the experimental task more difficult (Fisher, 1940; Madison, 2000).

Only one participant was tested at a time. Recordings were presented randomly by the experimental software: Affect 4.0 (Hermans, Clarysse, Baeyens, & Spruyt, 2010), which sent the audio signals to two loudspeakers. Before beginning the study, participants were allowed to adjust the volume to a comfortable level. They were also given the opportunity to ask any procedural questions.

A screen shot of the experimental software is provided in Figure 12. Trials were initiated when the listener used a computer mouse to click a portion of the screen labeled “Click here to listen.” They were allowed to listen to each excerpt multiple times in order to determine the performer’s intended tone of voice. Responses to the “8-alternative-forced-choice” task were entered by clicking the appropriate portion of the computer screen. Once a response was selected, it could not be changed. Participants were allowed to practice several trials with the experimenter present before the data collection began. After the participant was comfortable with the interface, the experimenter left the room and data collection began. Each participant heard and identified the performer’s intended tone of voice for 140 recordings. They were encouraged to take a break if they were tired, but none elected to do so. The experiment took approximately 30 minutes to complete. After the experiment, the participants were debriefed, interviewed, and offered a chance to ask questions about the study.

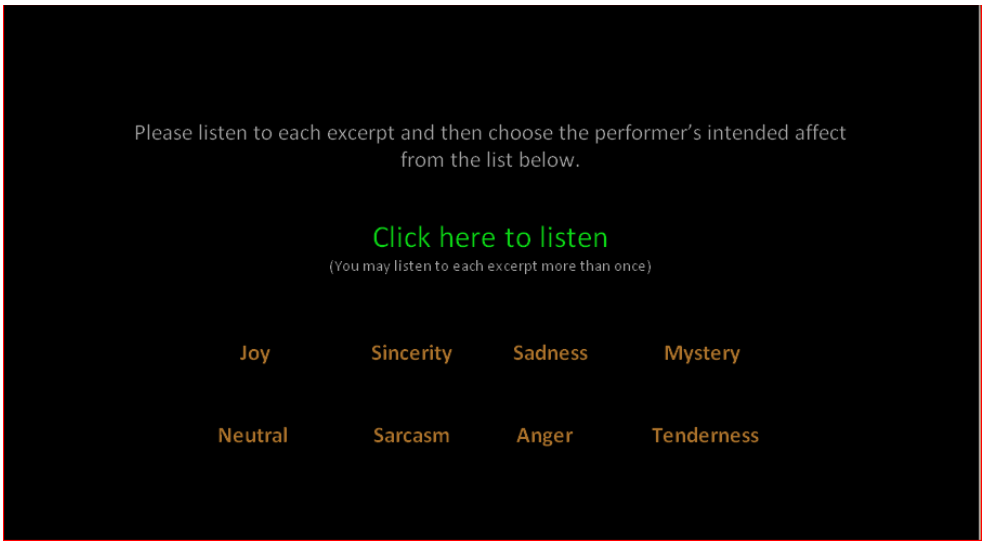


Figure 12: Screenshot of the experiment interface used by participants. Responses were entered by using a computer mouse to click on the tone of voice intended by the performer. The tone of voice responses were in a fixed position on the screen, and were not randomized by the experimental software.

Results

In total, each of the 280 recordings was judged 12 times, resulting in 3,360 judgments of the performers' intended tone of voice. The results presented below only pertain to the acoustic characteristics that were present when a listener selected a specific tone of voice. These results do not investigate the extent to which listeners were able to successfully identify the performers' intended tone of voice; this latter type of data will be examined in chapter 5.

Listener responses were first examined for frequency. The most frequently selected tone of voice was *neutral* (16.5%), followed by *joy* (14.4%), *sadness* (13.1%), *sarcasm* (13.0%), *tenderness* (11.9%), *sincerity* (11.8%), *mystery* (11.4%), and *anger*

(7.9%). The high prevalence of the *neutral* response might have been due to listeners selecting this tone of voice when they were unsure of the performer's intention.

General Model

A second multinomal logit model was created to investigate which cues were used by listeners in discriminating between the target tones of voice. The independent variables were the 27 measurements contained in the PRAAT voice reports. In order to allow a comparison with the first study, the model used 5 levels of the dependent variable: joy, sadness, sarcasm, sincerity, and "other." All the "lure" responses were amalgamated into the "other" variable and were discarded from the rest of analysis. As an example, every time a listener identified a passage as *sarcastic*, the PRAAT voice report for that recording was added to the Listener model as a set of sarcastic acoustic characteristics. Therefore, each of the 3,360 "tone of voice" judgments resulted in a PRAAT voice report being assigned to one of the five independent variables. Further, since each recording was judged 12 times, it was possible for a voice report to be included with multiple independent variables. Also, recall that the performer's intention was irrelevant for this model; voice reports were assigned to the five independent variable regardless of whether or not the listener correctly identified the performer's intended tone of voice.

According to the Listener Multinomial Logit model, 13 factors were found to be significant in distinguishing between the target affect: duration of excerpt, pitch standard deviation, minimum pitch, maximum pitch, mean period, standard deviation of period,

fraction of locally unvoiced frames, number of voice breaks, jitter (ppq5), shimmer (local), shimmer (apq3), shimmer (apq11), and mean harmonics-to-noise ratio. These factors are explained below. Combined, the 13 factors above accounted for approximately 13.7% of the variance between the five levels of the independent variable, and further, this finding was found to be statistically significant ($\chi^2 = 1308.6$ ($p<.001$).[14]

Analysis of Variance

In order to investigate how each of the model's 13 significant factors were used to decode the four target tones of voice, additional statistical tests were performed. Each of the 13 factors (i.e. musical cues) was subjected to a one-way analysis of variance (ANOVA) test. Eleven of the thirteen factors were found to significantly distinguish sarcasm from at least one of the other tones of voice. These results are shown in table 5 and discussed below.

[14] The coefficients (cue weights) for this model can be found in Appendix F.

Multinomial Logit Results	ANOVA Results		
Factor	Tone of Voice	Mean	
Duration	Sadness	10.86s	*
	Sincerity	9.29s	
	Sarcasm	8.43s	
	Joy	7.43s	
Pitch Standard Deviation	Joy	57.62Hz	*
	Sarcasm	49.07Hz	
	Sincerity	47.21Hz	
	Sadness	39.12Hz	*
Minimum Pitch	Sadness	197.21Hz	
	Sarcasm	205.76Hz	
	Joy	209.41Hz	
	Sincerity	215.00Hz	
Maximum Pitch	Joy	423.17Hz	*
	Sincerity	394.20Hz	
	Sarcasm	383.45Hz	
	Sadness	348.72Hz	*
Mean Period	Sadness	4.4ms	
	Sarcasm	4.3ms	
	Sincerity	3.7ms	*
	Joy	3.6ms	*
Period Standard Deviation	Sarcasm	0.83ms	
	Joy	0.83ms	
	Sadness	0.77ms	
	Sincerity	0.69ms	
Fraction of Locally Unvoiced Frames	Sarcasm	27.80%	
	Joy	16.13%	*
	Sadness	10.82%	*
	Sincerity	10.07%	*
Voice Breaks	Sarcasm	8.6 breaks	
	Joy	8.5 breaks	
	Sincerity	5.2 breaks	*
	Sadness	3.7 breaks	*
Jitter (ppq5)	Sarcasm	36.20%	
	Joy	35.20%	
	Sincerity	20.70%	*
	Sadness	19.50%	*

(Continued)

(Continued)

Shimmer (Local)	Sarcasm	5.81%	
	Joy	4.75%	*
	Sadness	3.61%	*
	Sincerity	3.55%	*
Shimmer (apq3)	Sarcasm	2.23%	
	Joy	1.73%	*
	Sadness	1.50%	*
	Sincerity	1.35%	*
Shimmer (apq11)	Sarcasm	3.82%	
	Joy	3.34%	
	Sadness	2.50%	*
	Sincerity	2.50%	*
Harmonic to Noise Ratio	Sincerity	27.32dB	*
	Sadness	27.21dB	*
	Joy	23.90dB	*
	Sarcasm	22.33dB	

Table 5: Results from the multinomial logit model and the associated ANOVA tests. The mean values for each variable are provided as a function of the target tone of voice. The asterisk (*) indicates a mean which was found to be significantly different from the mean for sarcasm at the p=.05 level. Significant differences between the other tones of voice are not reported here, but can be found in appendix F.

Several of the factors in Table 5 might be considered redundant measurements (e.g. the 3 different measurements of shimmer). For discussion purposes, the 13 factors were informally amalgamated into 4 larger categories: Timing (duration), Pitch (pitch standard deviation, minimum pitch, maximum pitch, mean period, and period standard deviation), Articulation (number of voice breaks), and Timbre (locally unvoiced frames, jitter[15], shimmer[16], and harmonic to noise ratio). These four factors are discussed below.

[15] Jitter is a measurement of period variability. Large amounts of jitter are associated with vocal tremors or hoarseness (Shaker, unpublished article).
[16] Shimmer is a measurement of amplitude variation. High percentages of shimmer are associated with vocal "roughness" (Shaker, unpublished article).

Further, when discussing statistically significant differences between the target tones of voice, only the significant differences related to sarcasm will be discussed. The results of the each ANOVA (as well as the post hoc tests) can be found in appendix G.

Timing

Duration

Duration was found to be a significant factor used by listeners to distinguish between the target affects. In the context of this study, duration referred to the total length of each recording. The beginning and end points for each recording were measured using PRAAT's graphical waveform editor. The mean duration for each of the four target "tones of voice," as distinguished by the listeners, was as follows: *sadness* (M=10.9s, SD=9.2), *sincerity* (M=9.3s, SD=7.3), *sarcasm* (M=8.4s, SD=7.4), and *joy* (M=7.4s, SD=5.4).

An analysis of variance revealed a significant difference between the "durations" of the four identified tones of voice, $F(3, 1753) = 17.58$, $p<.001$. Tukey's HSD post hoc test revealed a significant difference between the "durations" of *sarcasm* and *sadness*. Listeners distinguished the performer's intention as sarcastic when duration was relatively shorter.

Pitch

Mean Period

Mean period is the reciprocal of mean frequency; the two factors are inversely related (Rossing, Moore, & Wheeler, 2002). Frequency is measured in cycles per second (Hz), whereas period is measured as a unit of time. Low frequencies have larger periods whereas high frequencies have smaller periods. Because of this relationship, mean period could be considered an indicator of mean f_0. The mean period for each of the four target tones of voice, as distinguished by the listeners, was as follows: *sadness* (M=4.5ms, SD=2.0), *sarcasm* (M=4.3ms, SD=2.3), *sincerity* (M=3.7ms, SD=1.6) and *joy* (M=3.6ms, SD=1.6).

An analysis of variance revealed a significant difference between the mean period of the four identified tones of voice, $F(3, 1753) = 19.87$, $p<.001$. Tukey's HSD post hoc test revealed a significant difference between the mean periods of *sarcasm* and *sincerity*, as well as *sarcasm* and *joy.* Listeners tended to distinguish the performers' intentions as sarcastic when the mean period was relatively longer.

Minimum and Maximum Pitch

Minimum and maximum pitch were also significant factors in the Listener model. The average minimum pitch for each of the four tones of voice, as distinguished by the listeners, was as follows: *sad* (M=197Hz, SD=107), *sarcasm* (M=205Hz, SD=94), *joy* (M=209Hz, SD=94) and *sincerity* (M=215Hz, SD=95). An analysis of variance revealed

no significant differences between the minimum pitches of the four identified tones of voice, $F(3, 1753) = 2.49$, $p=.058$.

The average maximum pitch for each of the four tones of voice, as distinguished by the listeners, was as follows: *sadness* (M=349Hz, SD=118), *sarcasm* (M=384Hz, SD=124), *sincerity* (M=394Hz, SD=103), and *joy* (M=423Hz, SD=90). An analysis of variance revealed a significant difference between the maximum pitch of the four identified tones of voice, $F(3, 1753) = 36.34$, $p<.001$. Tukey's HSD post hoc test revealed a significant difference between the maximum pitch of *sarcasm* and *joy*, as well as *sarcasm* and *sadness.*

Pitch/Period Standard Deviation

Fundamental frequency (f_0) standard deviation has been found to be one of the most important factors in listeners' judgments of intended affect (Scherer & Oshinisky, 1977). This measurement was calculated by PRAAT using the mean f_0 for each passage as the baseline. In all of the passages except for the improvisations, fundamental frequencies were essentially fixed. The average pitch standard deviation for each of the four tones of voice, as distinguished by the listeners, was as follows: *joy* (M=58Hz), *sarcasm* (M=49Hz), *sincerity* (M=47Hz), and *sadness* (M=39Hz). An analysis of variance revealed a significant difference between pitch standard deviations of the four identified tones of voice, $F(3, 1753) = 28.35$, $p<.001$. Tukey's HSD post hoc test revealed a significant difference between the pitch standard deviations of *sarcasm* and *joy*, as well as *sarcasm* and *sadness.*

The findings on period standard deviation were similar. The average period standard deviation for each of the four tones of voice, as distinguished by the listeners, was as follows: *sarcasm* (M=0.83ms), *joy* (M=0.83ms), *sadness* (M=0.77ms), and *sincerity* (M=0.69ms). An analysis of variance revealed a significant difference between period standard deviations of the four identified tones of voice, $F(3, 1753) = 2.71$, $p<.05$. However, Tukey's HSD post hoc test did not reveal any significant differences pertaining to sarcasm.

Articulation

Voice breaks

In the Listener model, the mean "number of voice breaks" was found to be a significant factor. Voice breaks are identified by PRAAT when the interpulse interval is longer than 16.7ms (Boersma & Weenink, 2010). A high number of voice breaks in a musical context could be the result of staccato-like articulations or dramatic pauses. The mean number of voice breaks for each of the four tones of voice, as distinguished by the listeners, was as follows: *joy* (M=8.5 breaks, SD=8.4), *sarcasm* (M=8.6 breaks, SD=8.3), *sincerity* (M=5.2 breaks, SD=7.0), and *sadness* (M=3.7 breaks, SD=5.9).

An analysis of variance revealed a significant difference between the mean number of voice breaks for the four identified tones of voice, $F(3, 1753) = 46.38$, $p<.001$. Tukey's HSD post hoc test revealed significant differences between *sarcasm* and *sincerity*, as well as *sarcasm* and *sadness*. This finding suggests that listeners tended to

distinguish the performers' intentions as sarcastic when the mean number of voice breaks was relatively high.

Timbre

Fraction of locally unvoiced frames

The fraction of locally unvoiced frames was found to be a significant factor in the listener model. This factor is expressed as the percentage of the frames in each pitch that are unvoiced. In speech studies, any percentage over 20% typically signifies some type of pathology and results in a voice that sounds like the speaker has a cold (Boersma & Weenink, 2010). The fraction of locally unvoiced frames for each of the four tones of voice, as distinguished by the listeners, was as follows: sa*rcastic* (M=28%, SD=21), *joy* (M=16%,SD=12), *sad* (M=11%, SD=11), and *sincerity* (M=10%, SD=8).

An analysis of variance revealed a significant difference between the fraction of locally unvoiced frames for the four identified tones of voice, $F(3, 1753) = 141.92$, $p<.001$. Tukey's HSD post hoc test revealed significant differences between *sarcasm* and all 3 of the other target tones of voice. Listeners tended to distinguish the performers' intention as sarcastic when the fraction of locally unvoiced frames was high. This suggests that listeners may have relied on a noisy or nasalized timbre as a cue to distinguish a sarcastic tone of voice from the other target affects.

Jitter

Jitter, measured by the ppq5 method, was also found to be a significant factor in the listener model. Jitter is a measurement of small-scale period variability and can be measured in various ways. The ppq5 method (i.e. five-point Period Perturbation Quotient) is "the average absolute difference between a period and the average of it and its four closest [sic] neighbours, divided by the average period" (Boersma & Weenink, 2010). A high level of jitter results in a voice that sounds "shaky." Jitter is also somewhat related to what musicians refer to as "vibrato," but on a much smaller level. The percentage of jitter for each of the four tones of voice, as distinguished by the listeners, was as follows: as *sarcastic* (M=36%, SD=45), *joyous* (M=35%, SD=47), *sincere* (M=21%, SD=28), and *sadness* (M=20%, SD=25).

An analysis of variance revealed a significant difference between percentage of jitter for the four identified tones of voice, $F(3, 1753) = 24.91$, $p<.001$. Tukey's HSD post hoc test revealed significant differences between *sarcasm* and *sincerity*, as well as *sarcasm* and *sadness.* Listeners tended to distinguish the performers' intention as sarcastic when the percentage of jitter was high.

Shimmer

Shimmer, measured in three different ways, was also found to be a significant factor in the listener model. Shimmer is a measurement of amplitude variation. Three measurements of shimmer were found to be significant in the performer model. These measurements included: 1) the "local" method: "the average absolute difference between

the amplitudes of consecutive periods;" 2) the apq3 method: "the average absolute difference between the amplitude of a period and the average of the amplitudes of its [sic] neighbours, divided by the average amplitude;" and 3) the apq11 method: "the average absolute difference between the amplitude of a period and the average of the amplitudes of it and its ten closest [sic] neighbours, divided by the average amplitude" (Boersma & Weenink, 2010). High percentages of shimmer can be heard as vocal "roughness" (Shaker, unpublished article). All three measurements of shimmer had similar results, so only the "local" shimmer data will be discussed here. The percentage of shimmer for each of the four tones of voice, as distinguished by the listeners, was as follows: *sarcasm* (M=5.8%, SD=5.3), *joy* (M=4.74%, SD=4.3), *sadness* (M=3.6%,SD=4.3), and *sincerity* (M=3.5%, SD=3.7).

An analysis of variance revealed a significant difference between percentage of shimmer for the four identified tones of voice, $F(3, 1753) = 25.13$, $p<.001$.[17] Tukey's HSD post hoc test revealed significant differences between *sarcasm* and all 3 of the other target tones of voice. Listeners tended to distinguish the performers' intention as sarcastic when the percentage of shimmer was high. This suggests that listeners may have relied on a "rough" sounding timbre as a cue to distinguish a sarcastic "tone of voice" from the other target affects.

[17] The ANOVA reported here pertains to local shimmer. The ANOVAs for the other measurements of shimmer followed a similar pattern and can be found in appendix G.

Mean harmonic to noise ratio

The last significant factor in the Listener model was the mean harmonic to noise ratio (HNR). Mean HNR measures the degree of acoustic periodicity.

> "Harmonicity is expressed in dB: if 99% of the energy of the signal is in the periodic part, and 1% is noise, the HNR is $10*\log_{10}(99/1) = 20$ dB. A HNR of 0 dB means that there is equal energy in the harmonics and in the noise" (Boersma & Weenink, 2010).

The mean HNR for each of the four tones of voice, as distinguished by the listeners, was as follows: *sincere* (M=27dB, SD=7.1), *sadness* (M=27dB, SD=7.3), *joy* (M=24dB, SD=6.5), and *sarcastic* (M=22dB, SD=7.0).

An analysis of variance revealed a significant difference between the mean HNR for the four identified tones of voice, $F(3, 1753) = 48.91$, $p<.001$. Tukey's HSD post hoc test revealed significant differences between *sarcasm* and all 3 of the other target tones of voice. Listeners tended to distinguish the performers' intention as sarcastic when the mean HNR was low. This suggests that listeners may have relied on a noisy sounding timbre as a cue to distinguish a sarcastic tone of voice from the other target affects.

Discussion

In summary, the model revealed 13 factors used by listeners in order to distinguish between the target tones of voice. Specifically, listeners were most likely to

label a performer's intended tone of voice as being sarcastic when the recording had the following auditory characteristics: relatively shorter durations, relatively lower mean period, a higher level of unvoiced frames, a higher level of voice breaks, a higher level of jitter, a higher level of shimmer, and a lower mean harmonic to noise ratio. These cues were combined into what might be considered "the Listener Model" for expressing a sarcastic "tone within a musical context. This model is depicted in figure 13.

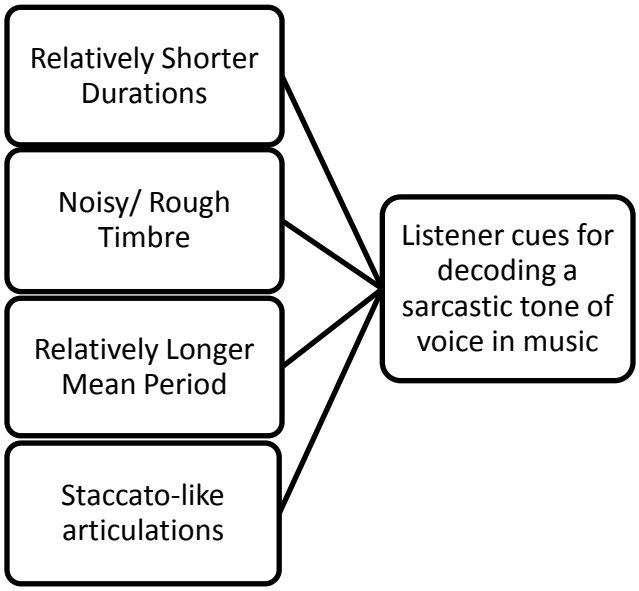


Figure 13: The Listener model for detecting a sarcastic tone of voice in instrumental music.

The Listener model represents a large amount of independent tone of voice judgments (n=3360). It is important to note that the results presented above only reflect the cues that were utilized by participants when distinguishing the performer's intended

tone of voice from the other possible responses. These results do not reflect the extent to which performers and listeners agreed upon (or matched) expressive cues. For example, the sarcastic tone of voice cues listed above reflects the acoustic characteristics of the recordings that were selected as intending "sarcasm," regardless of whether or not the performer had intended sarcasm (concordance between the Performer and Listener models is examined in chapter 5). Also, although the Listener model contained 13 significant factors, it should not be assumed that these are the only factors involved in decoding a sarcastic tone in music.

Also, it cannot be determined that listeners were actually listening to tone of voice cues when distinguishing between the target affects. This limitation is particularly important when considering the matching success of the improvisation data. In addition to tone of voice cues, listeners may have also used various associations in order to distinguish the performer's intent. For example, at least one of the improvisations expressing sarcasm consisted of what might be considered the "schoolyard taunt." This tune is shown in figure 14.

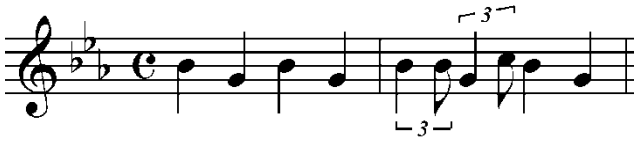


Figure 14. A transcription of the trumpet player's improvisation expressing sarcasm. In addition to tone of voice cues, musical and non-musical associations may contribute to hearing sarcasm in a musical context.

Associations with certain melodies, implied harmonies, meter, and mode all may have played a role in distinguishing between the target tones of voice. The experimental protocol allowed performers to play anything they wished in the improvisation recordings in order to convey the intended target affect.

Table 6 presents an indirect comparison between the Listener model and the known acoustical cues of sarcastic speech. The reason this comparison is indirect is that the speech literature rarely examines decoding cues in isolation. Typically only encoding cues, as well as *successful* decoding cues, are analyzed in the speech literature.

Sarcastic Speech	**Cue**	**Sarcastic Music (Listener Model)**
Inconclusive	Pitch/F_0	Relatively Lower (i.e. longer mean period)
Relatively Greater	Intensity	N/A
Nasalized	Timbre	Low HNR; High Unvoiced Frames; High values of Jitter(hoarseness); high values of shimmer (roughness)
Stressed, Erratic	Articulation	High Number of Voice Breaks
Slower	Timing	Relatively Shorter Durations (i.e. Faster)

Table 6: A comparison of auditory cues in sarcastic speech and cues used by listeners to detect a sarcastic tone of voice in instrumental music.

Something notable about Table 6 is the large number of timbre cues that were found to be statistically significant in the Listener model. Gabrielsson and Lindström (2010) noted that there has been "no systematic research on how the timbre of different musical instruments may influence perceived emotional expression" (p.389). This study provides further support that such systematic research may yield fruitful results. With regards to sarcasm, not all instruments are capable of creating a noisy timbre. Therefore, it could be hypothesized that some musical instruments are better suited for expressing sarcasm than others. In general, the timbre cues used by listeners to decode a sarcastic tone in music seem to match the timbre cues for sarcastic speech.

The results regarding pitch cues used by listeners to decode a sarcastic tone of voice were found to be inconclusive. Although a sarcastic tone of voice was found to be distinguished by a relatively longer mean period, further evidence is warranted. Findings regarding pitch cues in sarcastic speech are also inconclusive. Some research has found that a sarcastic tone of voice tends to be lower in pitch (Cheang & Pell, 2008; Rockwell, 2000), whereas other research has found sarcasm to be conveyed through higher pitches (Bryant & Fox Tree, 2005).

Listeners used relatively shorter durations as another cue for distinguishing a sarcastic tone from the other target affects. This is inconsistent with the research on sarcastic speech, which has been found to have relatively slower speaking rate (Cutler, 1974; Rockwell, 2000/2007; Cheang & Pell, 2008, Bryant, 2010; Caucci, et al. 2007). One possible explanation for this inconsistency (which was also discussed for the Performer model) is that spoken sarcasm may slow down in order to allow the listener to

capture the true meaning of the message, whereas this may not be necessary in a musical context. Another possibility is that this finding is an artifact of the target tones of voice used in this study. For example, some evidence was found that a sarcastic tone might be slower than other affective tones of voice which were not included in the Listener model (e.g. neutral or anger).

Regarding articulation, listeners distinguished a sarcastic tone in music through a relatively high level of voice breaks. In the context of this study, this measurement may be associated with staccato-like articulations and dramatic pauses.

The PRAAT voice reports did not contain any information pertaining to intensity. In the speech literature, a sarcastic tone of voice has been found to use greater intensities (Rockwell, 2000/2007) and reduced amplitude variability (Bryant & Fox Tree, 2005). Additional analyses and future research could investigate whether or not these acoustical characteristics are also present for sarcasm in a musical context.

In general, there seem to be many similarities between the cues used by listeners for distinguishing a sarcastic tone of voice in music and the known cues of sarcastic speech.

Chapter 5: Concordance Between the Performer and Listener Models

This chapter examines how well performers are able to communicate a sarcastic tone of voice to listeners.

Performer Listener Concordance

Figure 15 presents the performer and listener models derived in chapters 3 and 4 side-by side. Several similarities can be seen between the models. The similar cues used by performers and listeners included shorter durations, noisy timbre cues, and staccato-like articulations.

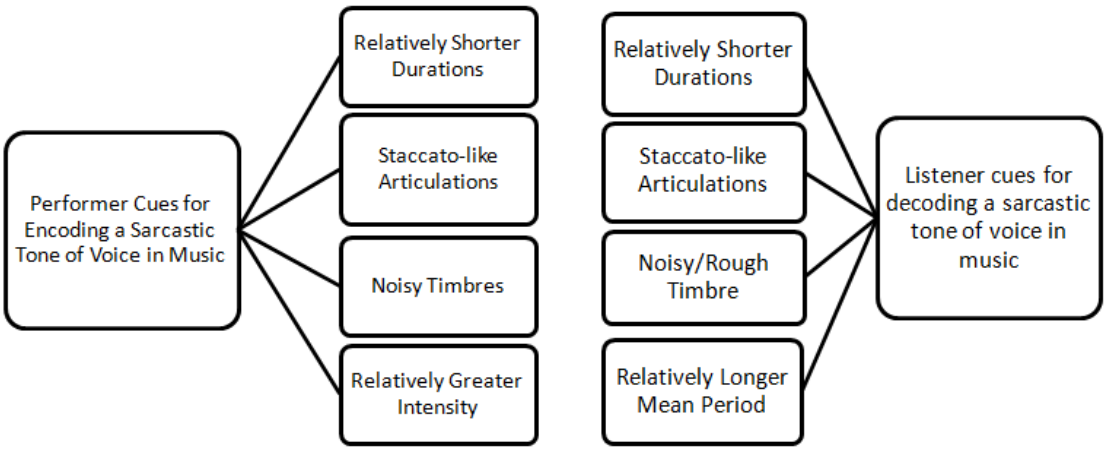


Figure 15: Performer and Listener models of a sarcastic tone of voice in instrumental music. In general, evidence was found that performers and listeners have a similar "acoustic code" for a sarcastic tone of voice in music.

Comparison with sarcastic speech

Both the performer and the listener models contained relatively shorter durations as a cue for a sarcastic tone of voice in music. This is contrary to the longer durations that are found in sarcastic speech (Rockwell, 2007; Cheang & Pell, 2008; Bryant, 2010). Future research on both sarcastic speech and sarcastic music could investigate this discrepancy by digitally manipulating the duration of sarcastic recordings (both speech and music) presented to listeners.

The performer and listener models also both contained more staccato-like articulations as a cue for a sarcastic tone of voice in music. This was primarily measured by the number of voice breaks in each recording. A similar articulation cue can be found in the sarcastic speech literature: sarcastic speech tends to be stressed and erratic (Cutler, 1974; Utsumi, 2000).

Both the performer and the listener models contained cues that are suggestive of a noisy timbre. In both models combined, timbre was characterized using four measures: the fraction of locally unvoiced frames, jitter, shimmer, and the mean harmonic-to-noise ratio. These measurements suggest that sarcastic tones typically contain some degree of noisiness. Literature on the timbre of sarcastic speech has found that it includes a unique resonance profile (Cheang & Pell, 2008), nasalization (Cutler, 1974; Utsumi, 2000), and a “chest tone” (Fonagy & Magdics, 1963). Although several attempts have been made, music researchers have struggled with objectively describing timbre (Grey, 1978; McAdams, 1999). If nothing else, a sarcastic tone in music seems to involve some

differentiation or markedness from a normal voice. Based on the findings of the statistical models, adding noise to the signal is likely to be one of the distinguishing characteristics of a sarcastic tone of voice.

The results suggest that performers and listeners associate different pitch cues with a sarcastic tone of voice in music. The data on the performers' use of pitch cues to encode a sarcastic tone in music was inconclusive, whereas listeners tended to decode musical sarcasm with relatively longer mean periods (i.e. lower pitch). This dichotomy is especially interesting because it echoes the mixed results that are found in the speech literature (see table 1). One possible explanation for these mixed results might be that absolute pitch height is not actually a cue for a sarcastic tone of voice. Rather, performers and listeners might utilize/hear a reserved portion of the sender's potential pitch range (high or low) as a cue of intended sarcasm. Evidence in support of this account comes from Cheang and Pell's (2008) finding that sarcastic speech tends to have a unique resonance profile. Further, resonance profiles can be used by listeners to normalize a speaker's voice (Patterson et al., 2008). Therefore, a more sophisticated measurement might be needed to explain the mixed results on pitch cues with regards to a sarcastic tone of voice.

Matching Data

The performer and the listener models are in some ways incomparable. For example, the performers were asked to differentiate four target tones of voice. It was found that they manipulated at least four cues to accomplish the task. The listeners were

asked to distinguish between 8 affective tones of voice. The Listener model found that at least 13 cues (some of which might be considered redundant) were used to accomplish the discrimination task. Distinguishing between more affective tones of voice presumably requires using more auditory cues.

Despite these methodological differences, both tasks involved demonstrating knowledge of different affective tones of voice in a musical context. Because of this, it is possible to examine the concordance between performers and listeners by calculating the success of their communication. This was accomplished by examining those recordings that were successfully distinguished by listeners. The results are presented in table 7.

Table 7: Matching success (percentage) as a function of stimulus type and performer's intended tone of voice. Chance level for the 8-alternative-forced-choice task was 12.5%.

	9 notes	Ascending M3	Descending M3	Happy Birthday	Improvisations	**Total**
Joy	13.7%	22.0%	12.5%	42.3%	61.9%	**30.5%**
Sadness	31.0%	22.6%	29.8%	36.3%	35.1%	**31.0%**
Sarcasm	19.6%	26.8%	32.7%	51.2%	29.2%	**31.9%**
Sincerity	9.5%	14.9%	14.9%	23.2%	28.0%	**18.1%**
Total	**18.5%**	**21.6%**	**22.5%**	**38.2%**	**38.5%**	***27.9%***

Table 7 shows that the best distinguished affective tone of voice was sarcasm (32%), followed sadness (31% percent), joy (31%), and sincerity (18%). All four target attitudes were identified above chance levels (12.5%). However, it is important that these results be considered within the context of the study. For example, it would be erroneous to think that a sarcastic tone in music could, on average, be correctly *identified* in any

type of music about 32% of the time. Recall that this study only examined the extent to which sarcasm could be *distinguished* from the 7 other target tones of voice (Frick, 1985). Moreover, the list of affects to be identified might be expected to have primed listeners to expect that some examples would sound sarcastic. Also, it was found that the results varied considerably depending on the type of musical stimulus (i.e. type of recording).

Table 7 also presents the percentage of correct matches, in which the listener was able to correctly distinguish the performers' intended tone of voice, as a function of recording type. In general, the more information that listeners received, the better they performed on the discrimination task. This finding is in accordance with previous literature on acquired musical knowledge (Plazak & Huron, 2011). Information in this context refers to the type of stimulus that was heard (9 recorded notes, intervals, "Happy Birthday," improvisations). Specifically, participants were able to correctly distinguish 19% of the "9 repeated note" recordings, 22% of the ascending major 3rd recordings, 23% of the descending major 3rd recordings, 38% of the "Happy Birthday" recordings, and 39% of the improvisation recordings. When averaged across the target tones of voice, all 5 types of recordings were distinguished with better than chance level accuracy.

Listeners successfully distinguished the performer's intention at better than chance levels when presented with the "9 repeated note" recordings. This supports a similar finding by Juslin and Madison (1999). In these "9 repeated note" recordings, performers were limited to manipulations of timing and timbre in order to convey the target tone of voice. The only affective type which was not distinguished above chance

level for the "9 repeated notes" stimuli was *sincerity* (M=10%). The most recognized tone of voice for this stimulus type was *sadness* (M=31%), which is also congruent with the findings of Juslin and Madison (1999).[18]

The percentage of successful communication was slightly higher for the ascending and descending intervals than it was for the "9 repeated note" recordings. With only one exception, all of the tones of voice were distinguished above chance levels through hearing this single isolated interval. The best identified tone of voice for this stimulus type was *sarcasm* (ascending M3: 27%; descending M3: 33%), whereas the poorest identified tone of voice was *sincerity* (ascending M3: 15%; descending M3: 15%). There was a notable asymmetry between correctly identifying *joy* through ascending (22%) and descending (12.5%) major 3rds. The former type of interval was recognized above chance level, whereas the latter was recognized at chance level. Future studies of affective musical communication may wish to investigate the impact of pitch contour direction on the ability to express specific musical tones of voice. These data are also relevant to the literature on the emotional qualia of musical intervals (Maher & Berlyne, 1982; Kaminska & Woolf, 2000), as well as literature on how performers signal various tones of voice for fixed pitch intervals.

Communication of the target tone of voice was more successful for the "Happy Birthday" recordings than it was for the "9 repeated note" and interval recordings. All four of the target affects were communicated above chance levels, with *sincerity* (23%) being the least and *sarcasm* (51%) being the best recognized. This finding provides

[18] Juslin and Madison examined the difference between angry, sad, fearful, and happy expressions.

support for the hypothesis that sarcasm is easier to distinguish when it parodies something known to the listener (Haiman, 1998).[19]

The highest overall communication success rate resulted from the improvisation recordings. 39% of all the improvisation recordings were successfully distinguished. The best distinguished tone of voice for this stimuli-type was *joy* (62%); the least distinguished tone of voice was *sincerity* (28%). All four of the target tones of voice were recognized above chance level. It is worth noting that improvisations expressing *joy* were the best distinguished recordings of the entire study. Future research may wish to investigate why this tone of voice, when combined with this stimulus-type, had such a high rate of communication success.

In addition to determining which stimulus types and affective tones of voice were most successfully communicated, the most common errors were also investigated (Fairbanks, 1940). Scherer (1995) states that "errors are not randomly distributed and the patterns of misidentification provide important information on the judgment process" (p.237). The most common confusion for the tone of voice *joy* was *neutral* (21% confusion rate); the most common confusion for *sadness* was *tenderness* (22% confusion rate); the most common confusion for *sarcasm* was *joy* (15% confusion rate); and the most common confusion for *sincerity* was *neutral* (21% confusion rate). Table 8 presents a confusion matrix for the listeners' responses.

[19] Assuming that the listeners in this experiment were familiar with "Happy Birthday."

Table 8: Confusion matrix: listener responses as a function of performer intention. The performers' intentions are organized as rows; listener responses are organized as columns.

		Listener Response							
		Target "tones of voice"				Lure "tones of voice"			
		Joy	Sadness	Sarcasm	Sincerity	Anger	Tenderness	Mysterious	Neutral
Intention	Joy	30.5%	4.4%	9.8%	13.3%	8.8%	6.3%	5.8%	21.1%
	Sadness	1.0%	31.0%	5.1%	10.4%	1.1%	21.8%	20.7%	9.1%
	Sarcasm	15.5%	5.1%	31.9%	5.4%	14.5%	3.0%	9.6%	15.0%
	Sincerity	10.7%	11.9%	5.2%	18.1%	7.1%	16.7%	9.5%	20.7%

As can be seen in table 8, four specific confusions rates were above 20%. These confusions occurred between: *sadness* & *tenderness* (22%); *joy* & *neutral* (21%), *sincerity* & *neutral* (21%), and *sad* & *mysterious* (21%). In the affect literature, sadness is considered to be a negatively valenced affect whereas tenderness is a positively valenced. However, both tones of voice are characterized as low arousal affects and have similar structural cues (Juslin & Timmers, 2010). The high confusion rate between *joy* and *neutral*, as well as between *sincerity* and *neutral*, could be attributed to many factors. One possible explanation is that listeners may have selected "*neutral*" if the cues were deemed to be too subtle to discriminate. Said another way, it may have been possible that the participants were using the "*neutral*" response as an "unsure" or "undecided" response. It is difficult to speculate why there was a high confusion rate between *sadness* & *mysteriousness*. Hevner (1935) found that both sadness and mysteriousness are associated with the minor mode, but additional research is needed on musical mysteriousness.

With specific regards to *sarcasm*, this tone of voice was most confused with *joy* (15.5%), *neutral* (15.0%), and *anger* (14.5%). It was least confused with *tenderness* (3.0%), *sadness* (5.1%), and *sincerity* (5.4%). It is puzzling how sarcasm, which was defined earlier as a negatively valenced affect, could be most commonly confused with a positively valenced affect such as joy. One might expect that sarcasm would have been more commonly confused with another negatively valenced affect such as anger or sadness. A few speculative explanations for this finding are provided later.

Yet another way to investigate the ways in which sarcasm is successfully communicated between musicians and listeners is to compare the differences between recordings that were successfully identified and those that were not. This comparison could be considered one way of checking the accuracy of the Listener model presented in chapter 4.

PRAAT voice reports for the sarcastic recordings were separated into two groups based on whether or not they were successfully distinguished. Listeners successfully distinguished 31.9% of the sarcastic recordings; the PRAAT voice reports from these successful matches constituted one data group. The other 68.1% of the responses, and their associated PRAAT voice reports constituted the other data group. Note that each recording intending to convey sarcasm was heard 12 times. Therefore each voice report is represented in this analysis 12 separate times. Also, it was possible for a voice report to be in both categories (successfully distinguished and unsuccessfully distinguished). Descriptive statistics for the successfully and unsuccessfully distinguished recordings are

presented in Table 9. Of the 27 variables in each PRAAT voice report, only the 13 significant factors determined by the Listener model are reported.

Table 9: Comparison of successfully and unsuccessfully identified sarcastic recordings based on the 13 factors from the Listener (L) model.

Factor	Sarcasm relative to other target affects	Successfully Distinguished	Unsuccessfully Distinguished
Duration	Shorter	7.847s	9.359s
Pitch SD	Higher	50.038Hz	47.559Hz
Minimum Pitch	Lower	203.499Hz	209.352Hz
Maximum Pitch	Higher	383.809Hz	382.9Hz
Mean Period	Higher	4.31ms	4.34ms
Period SD	Higher	0.86ms	0.78ms
Unvoiced Frames	Higher	35.2%	16.1%
Voice Breaks	Higher	9.0 breaks	7.9 breaks
Jitter (ppq5)	Higher	42%	26.8%
Shimmer (Local)	Higher	6.6%	4.5%
Shimmer (apq3)	Higher	2.46%	1.87%
Shimmer (apq11)	Higher	4.57%	2.68%
HTN Ratio	Lower	21.1dB	25.2dB

Table 9 shows that successfully distinguished sarcastic recordings had more pronounced auditory cues for all but one of the 13 significant factors in the Listener model. Correctly distinguished sarcastic recordings were characterized by shorter durations, greater pitch variability, lower minimum pitches, higher maximum pitches, greater period variability, a higher percentage of unvoiced frames, a larger number of voice breaks, a higher percentage of jitter, a higher percentage of shimmer, and a lower harmonic to noise ratio. This suggests that the Listener model was fairly accurate in determining the most important cues for successfully decoding a sarcastic tone of voice.

The only Listener model factor which was less pronounced for the successfully distinguished recordings was mean period. However, the difference between successful and unsuccessful matches for this factor was very small.

Discussion

How could a sarcastic tone of voice in music be so well recognized, despite how relatively rare it is in common musical practice? Figure 15 showed a great deal of concordance between the known cues used by both performers and listeners for communicating a sarcastic tone of voice in music. Despite this, it still seems improbable that sarcasm could be distinguished as well or better than more prevalent musical affects such as joy or sadness. One possible explanation for this finding will be provided using the framework of Robert Hatten's (1994) concept of musical markedness.

Hatten (1994) defines markedness in music "quite simply as the valuation given to difference" (p.34). He argues there is a relationship between markedness and meaning, namely that: 1) differentiations yield oppositions/variations, 2) oppositions/variations often have an unequal probability, and 3) unequal probabilities result in various musical meanings. As an example of the relationship between markedness and meaning, consider the statement, "What a nice day." The word "day" is an unmarked term. It could mean either the time from sunrise to sunset, or it could refer to a 24 hour period. An opposing term for day is "night." The statement, "What a nice night." takes on a much narrower meaning. Night is a marked term, and therefore the latter sentence takes on a more specific meaning. Of the multiple types of oppositions used by Hatten, privative

oppositions are the most relevant to this specific study. Privative oppositions occur when the relationship between the marked and unmarked terms is a matter of being present or absent; they are not antithetical. As a musical example, consider the relationship between the major and minor mode. The major mode is more common than the minor mode, and therefore, it is the unmarked term. The relationship between these two modes is not antithetical (i.e. the major mode is not the opposite of the minor mode). Rather, the minor mode only implies the absence of the major mode (i.e. a privative opposition).

Earlier it was stated that negatively-valenced music is rare. In this sense, negatively valenced music might be considered to be marked. Because of its opposition from the more probable positively-valenced music, it would be expected to carry a more narrow and specific meaning. Recall that in this study, listeners were asked to distinguish between 8 different tones of voice. Of these 8 terms, only 3 were what might be considered negatively valenced (sadness, sarcasm, and anger), and further, only two might be considered to involve a relatively high level of arousal (sarcasm and anger). One reason that sarcasm may have been so well recognized from the other target tones of voice might be attributed to its markedness from being negatively valenced. If this were true, then one of the most common confusions for sarcasm should have been anger. Table 8 showed that 14.5% of all sarcastic recordings were mislabeled as anger, which provides some support for this conjecture. However, sarcasm was more commonly confused with joy and neutral, neither of which is negatively valenced. With regards to the *joy/sarcasm* confusion, recall that many of the same auditory cues were used to express these two tones of voice. The major distinction between joy and sarcastic tones

of voice was a unique timbre. It seems plausible that some of the instruments used in this study may not have been capable of producing the timbre cues associated with a sarcastic tone, but were capable of producing other sarcastic tone cues. Another possible explanation for this confusion may have been that listeners "missed" the more subtle sarcastic recordings and misinterpreted them as joy, similar to the way listeners might "miss" a sarcastic message in speech. With regards to the *neutral/sarcasm* confusion, two possible explanations will be provided. First, it is possible that listeners used the response "neutral" whenever the performers' cues were too subtle to distinguish. In this case, the confusion would simply indicate that the listeners just didn't know the answer, rather than a true confusion between sarcastic and neutral tones of voice. Secondly, on a spectrum of valence strength, one might argue that sarcasm tends be less negatively valenced (i.e. more neutral) than other affects like anger or disgust. Future research might explore what other types of negatively valenced "tones of voice" can be expressed in music, and also the relative strength or weakness of these musical affects.

In addition to the hypothesis that sarcasm was highly distinguishable from the other tones of voice due to its markedness, there are other possible explanations. One criticism of the Performer and Listener models presented in figure 15 is that they only *describe* the sound of musical sarcasm, they do not *define* it. For example, consider the following hypothetical piece of music in figure 16.

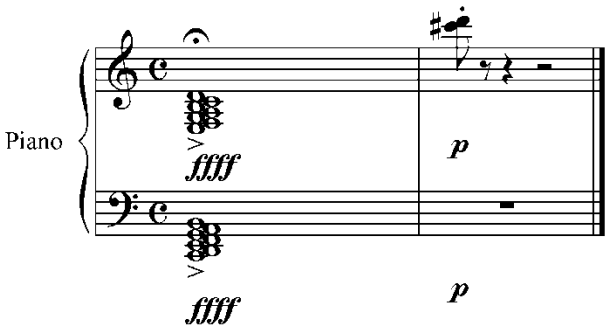


Figure 16: A hypothetical composition which contains the musical features which were found to *describe*, but not *define*, a sarcastic tone of voice in music.

The nonsense piece of music in figure 16 contains most of the cues which were found to describe sarcasm, including loud dynamics, contrasting articulations, a noisy sonority (due to the cluster chord and the minor second dyad), the unpredictability of the fermata, etc. However, does the presence of these musical structures reliably signal a sarcastic tone of voice? In order to investigate this question, as well as the ecological validity of the findings on a sarcastic tone of voice in music, an additional empirical study was devised and is presented in chapter 6.

Chapter 6: Free-Response Study

This chapter investigates listeners' free responses to a set of nominally sarcastic pieces of music, namely the five pieces in Prokofiev's op.17, for piano entitled *Sarcasms*. Based on the high degree of concordance between the Performer and Listener models, it was hypothesized that listeners would be able to identify a sarcastic tone of voice in a free-response listening task. To foreshadow the results, no evidence was found to support this hypothesis. Without contextual clues, listeners did not make spontaneous comments pertaining to a sarcastic "tone of voice" after hearing the piece. However, listeners did make comments about other types of negatively valenced affects (fear, evil, creepiness, etc.). By contrast, after being prompted with the title of the piece (i.e. *Sarcasms*), many of the comments made by listeners pertained to sarcasm. This finding suggests that contextual clues are an important component to hearing a sarcastic tone of voice in music.

This chapter is organized into four sections: rationale for using Prokofiev's op.17, *Sarcasms*; methodology of the free-response study; results; and a discussion about the importance of context for hearing a sarcastic tone of voice in instrumental music.

Choosing a Musical Corpus

There were several challenges in choosing a musical corpus to investigate listeners' ability to identify a sarcastic tone in music. These challenges included: 1) finding pieces of instrumental music that were intended to be sarcastic, or could be reasonably identified as being sarcastic; 2) finding instrumental pieces in which sarcasm was used for a sustained period of time, as opposed to just a brief sarcastic moment, and 3) finding pieces of instrumental music that contained the acoustic characteristics of a sarcastic tone.

An informal method was used to select a piece of music for the free-response study. The performers and listeners who took part in the "standard paradigm" study were asked during the post-study interviews for suggestions of composers or pieces that they thought were sarcastic. Participants made a number of suggestions. Of the collected suggestions, very few musical works could be considered to express sarcasm for a sustained period of time. This echoes the finding that sarcastic speech tends to occur in one conversational turn, rather than being used consistently throughout a conversation (Nelms, 2002, as cited in Rockwell, 2006). Many of the suggestions involved some type of musical quote or parody, not necessarily a sarcastic tone of voice. Other suggestions provided by the participants pertained to notably sarcastic composers, sarcastic musical instruments, and sarcastic special effects rather than specific pieces of music.

Of the pieces suggested by participants, one stood out as being particularly appropriate: Prokofiev's op.17, *Sarcasms* for solo piano. Not only does Prokofiev's work

offer a pertinent title, it also contains (in notated form) many of the acoustical cues found in the Performer and Listeners models for a sarcastic tone of voice.

Prokofiev's op.17, *Sarcasms*, for piano

The *Sarcasms* were written from 1912-1914 and were first performed by Prokofiev in December of 1916 (Boylan, 1979). Prokofiev continued to perform these pieces for many years after they were composed, and even made an attempt to orchestrate them (Phillips, 2006). Despite this, the *Sarcasms* are relatively obscure compared to Prokofiev's other piano works. Features of Prokofiev's compositional style are apparent throughout op.17, including: "unexpected harmonic juxtapositions and the alternation of acerbic dryness, lyricism, and motoric rhythms," as well as "dissonance and satire" (Burkholder, Grout, & Palisca, 2006, p. 877). Musical features of the *Sarcasms* include: distortion of conventional harmonic structures, out-of-phase voice leading, and a balance of consonance and dissonance that often results in tonal ambiguity (Boylan, 1979). In light of the title and of the various musical features, it is not unreasonable to assume that Prokofiev intended these pieces to be heard as expressing sarcasm. In 1941, Prokofiev wrote in his autobiography: "I would prefer my music to be described as 'scherzo-ish' in quality, or else by three words describing various degrees of the scherzo—whimsicality, laughter, mockery" (Prokofiev & Palmer, 1992, p. 249). Scherzos are often associated with musical jokes. In parts of his diary, he refers to op.17 as "my sarcastic pieces" rather than by its formal title (Phillips, 2006, p.580). In addition to op.17's colorful title,

a program was written for the last of the five *Sarcasms*, which Prokofiev copied into his autobiography. Prokofiev wrote:

> "We often indulge in malicious laughter at someone or something, but when we pause to look we see how pitiful and sad is the object of our ridicule: and then we grow ashamed: the mocking laughter rings in our ears, but it is we who are its object now" (Prokofiev & Palmer, 1992, p.255).

Prokofiev's program suggests both irony and sarcasm. It is ironic in that it depicts a double meaning: speaking poorly of others often speaks poorly of us. The program is sarcastic in that the irony Prokofiev refers to has been used maliciously against someone else.

Certain assumptions are made in discussing the musical characteristics of op.17 which could be considered sarcastic. In general, tone of voice cues are subjective whereas musical scores are objective. Musicians might disagree about whether a sound has a noisy timbre, but they are likely to agree that a notated melody involves drastic changes in register or dynamics. Up until this point in the study, the primary focus of investigation has been on subjective musical characteristics (i.e. tone of voice cues). However, in defending the use of Prokofiev's op. 17 for this study, several objective musical characteristics are assumed to result in the cues of a sarcastic tone of voice. These cues included: noisy timbres (both low "toneness" and low sensory consonance),

irregularly stressed rhythms, relatively greater dynamics, and contrastive articulations. Three exemplary musical excerpts are provided and discussed below. The degree to which these tone of voice cues can be heard is highly dependent on the performer's interpretation of the score. [20]

Musical excerpts from op.17

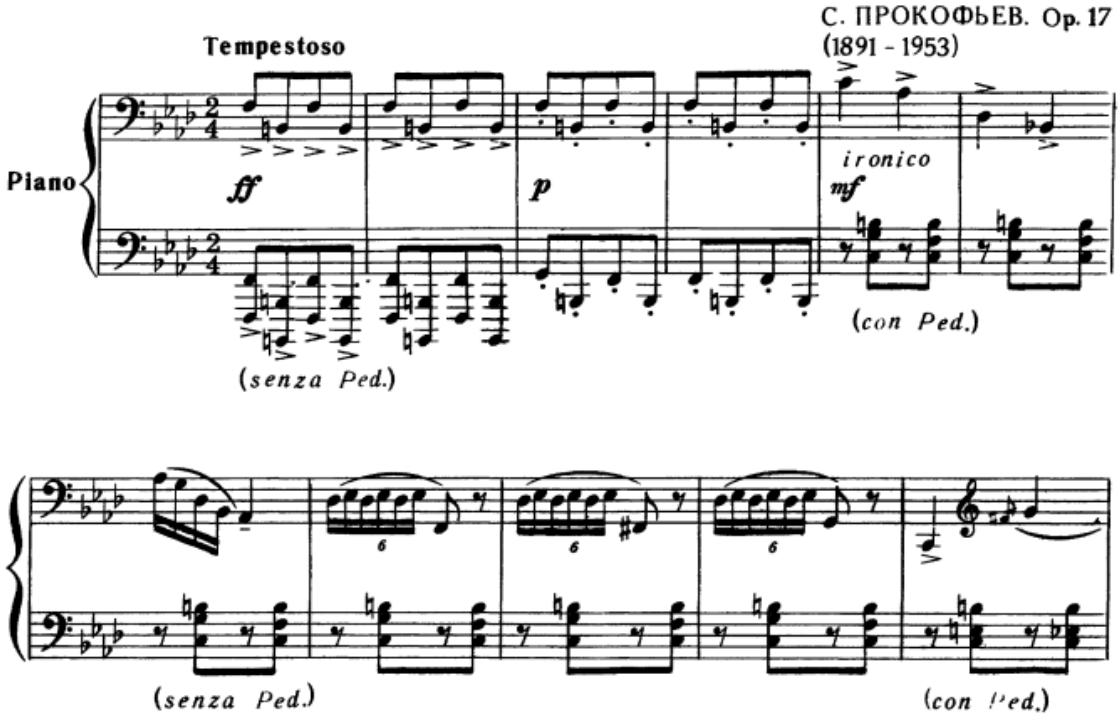


Figure 17: Prokofiev, op.17, *Sarcasms*, no.1, m.1-11. Potential sarcastic tone of voice features include low toneness, contrastive articulations, contrasting dynamics, and the "*ironico"* performance marking in m.5.

[20] Excerpts were taken from the following edition of the score: Masters Music Publications (1998). Sarcasms : for solo piano, op. 17 by Sergei Prokofiev. Boca Raton, FL: 1-20.

Figure 17 shows the first 11 measures of Prokofiev's op.17. Of particular interest is the "*ironico"* performance marking in measure 5. The presence of this marking may imply that Prokofiev believed performers would be able to accomplish some type of ironic playing style. There is minimal musical context that precedes the *"ironico"* performance indication. On the surface level, the *"ironico"* could merely indicate a change from the "*tempestoso"* marking in measure 1. On a deeper level, one might suspect that *"ironico"* indicates something about the way Prokofiev intended the transition between measures 4 and 5 to be performed. However, it is unclear if *"ironico"* suggests that the transition between these contrasting sections (m.1-4 and m.5-11) should be accentuated or subdued. Another possible interpretation is that the marking has nothing to do with the preceding measures, and instead only refers to the section beginning in m.5. If this were the case, then *"ironico"* might refer to the ambiguity of the downbeat resulting from the juxtaposed parts (i.e. left and right hand), or it might refer to the articulation of the piece's main theme.

In the 1920's, Prokofiev made several piano roll recordings while in Paris (Arnest, 1996). Several realizations of these piano roll recordings are commercially available. Fortunately, among these recordings are the first two *Sarcasms*. By listening to Prokofiev's own interpretation of the "*ironico"* marking, some insight can be gained into the composer's intention. In these recordings, the transition that occurs between measures 4 and 5 is rather subdued, suggesting that *"ironico"* refers instead to the section beginning in m.5. The volume of the right hand melody clearly dominants the left hand accompaniment in m. 5-11; therefore, the juxtaposed rhythm is not accentuated. Based

on Prokofiev's piano roll recording, what seems to be most implied by the "*ironico*" marking in measure 5 is a mechanical performance of the main theme (which is played in the right hand, m. 5-6). Prokofiev performs this theme as plainly as possible. Deviations in tempo, articulation, and intensity are minimal. This is in stark contrast to the rubato-style that is so prevalent throughout the rest of Prokofiev's op.17 recording. The *"ironico"* therefore seems to be an instruction to perform the main theme in an unexpected, and perhaps unmusical, style.

One of the most distinctive cues of a sarcastic tone is a noisy timbre. In general, there is little that can be done to change the timbre of the piano. However, the piano does have an exceptionally large pitch range. In addition to obvious differences in pitch, each octave on the piano might be considered to have a unique timbre. As a way of overcoming the relatively fixed timbre of the piano, an argument could be made that Prokofiev used different pitch registers in order to convey a sarcastic tone of voice in op.17. Transposing pitches up or down can signal certain affective qualities such as aggressiveness or politeness (Huron, Kinney, & Precoda, 2006). A distinct characteristic of the lower pitches, such as those on the piano, is that they often sounds "muddy" or "unclear" (Kameoka and Kuriyagawa, 1969). This auditory characteristic has been referred to as "toneness" (Huron, 2001). Prokofiev uses the low register of the piano extensively throughout op.17, often scoring both the left and right hand parts in the bass clef, resulting in many instances of low toneness.

Another sarcastic tone of voice cue that can be seen in figure 17 is irregularly stressed rhythms and contrastive articulations. Op.17 begins with a syncopated figure

which is achieved through register stress (m. 1-4). In m. 5-11, this syncopation continues in the left hand, but the right hand melody sounds on the beat. The juxtaposition of these two parts continues until m. 14, resulting in irregular metrical accents. With regard to contrastive articulations, Prokofiev specifies the articulation of almost every note in the opening four measures, and further, systematically varies these articulations in a contrastive manner. In addition to the marcato and staccato makings, articulation contrasts are also specified in the pedal markings. It bears reminding that these particular musical features may or may not result in a sarcastic tone of voice depending on the performer's interpretation.

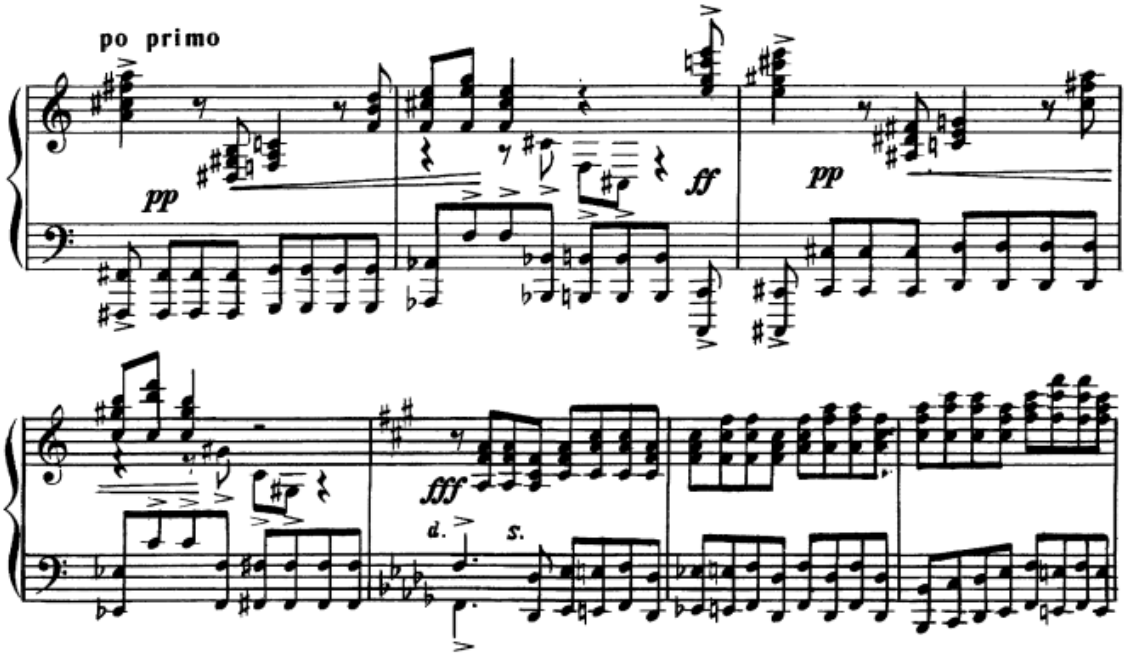


Figure 18: Prokofiev, op. 17, *Sarcasms*, no.3, m. 45-51. Potential sarcastic tone of voice cues include rapidly changing dynamics, and the prevalence of dissonance resulting from using two key signatures simultaneously.

Sarcastic tone of voice cues which can be seen in figure 18 include the prevalence of dissonant harmonic intervals and contrastive dynamics. Prokofiev's simultaneous use of two key signatures (see m.49) results in many dissonant intervals. The prevalence of harmonic 7ths and 9ths is especially clear in m.49-51. These intervals have been found to have a low degree of sensory consonance (Kaestner, 1909, as cited in Huron, 1991; Plomp and Levelt, 1965; Vos, 1986) and have been found to be less frequently used in many types of music (Huron, 1991). As stated earlier, the timbre of the piano is relatively fixed; it is hard to make the piano sound "noisy." One possible explanation for the prevalence of low sensory consonance intervals in this passage may be to compensate for the piano's relatively fixed timbre. If the concept of "noisy" could be considered to include the presence of "harsh or unpleasant" sounds, then using many intervals with low sensory consonance is one possible way to obtain a noisy piano sound.

Another sarcastic tone of voice cue that can be seen in figure 18 is rapid changes in dynamics. The piece rapidly crescendos from pianissimo to fortissimo twice in a five measure span (m.45-49). According to Hatten (1994), these contrasts signal musical markedness. Prokofiev's drastic dynamic contrasts, especially within a relatively short period of time, suggest a high degree of markedness. Marked terms take on a narrower meaning (Hatten, 1994); extreme markedness may take on an extremely narrow meaning. Prokofiev's use of dynamics is also consistent with the intensity cues of sarcastic speech, which has been found to be more intense (Cutler, 1974; Rockwell, 2000) and distinguished through contrast (Attardo, et al., 2003).

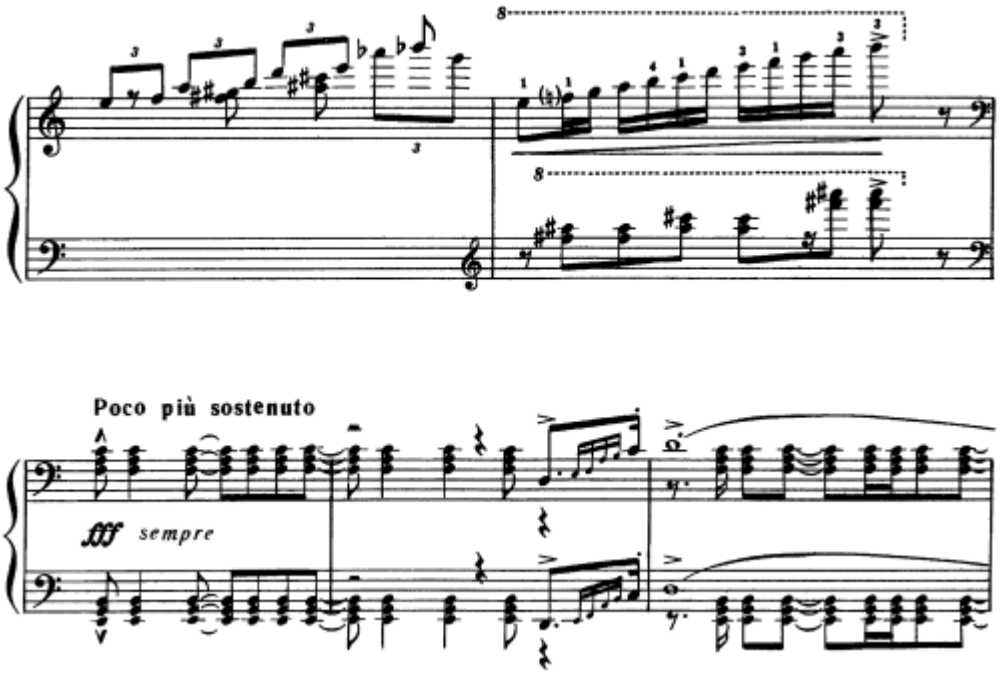


Figure 19: Prokofiev, op.17, *Sarcasms,* no. 4, m.12-16. Potential sarcastic tone of voice cues in this passage include changes in register, changes in texture, and the use of grace notes and glissandi.

Two final cues used by Prokofiev to signal a sarcastic tone include the use of grace notes and extreme shifts in pitch register, both of which can be seen in figure 19. Throughout op.17, Prokofiev uses a variety of embellishments, including grace notes, rolled chords, and glissandi. Even though these embellishments are sometimes considered to have no rhythmic value, they play an important role in performance expressivity (Desain & Honing, 1995). These embellishments, which can be performed ahead of the beat or on the beat, might be considered as one possible way to mimic the irregularly stressed rhythms of sarcastic speech[21]. Figure 19 also shows one of

[21] This is highly dependent on the way in which the performer chooses to performer the embellishments.

Prokofiev's more extreme changes of register from op.17. In measure 13, both the left and right hand parts are scored in treble clefs with octave signs. After a dramatic climb to B7 (almost the highest note on the piano), the music briefly rests before plummeting over 3 octaves (m.14). The radical register shift is further accentuated by the fortissimo dynamic marking, syncopated rhythm, and a change in texture. This use of contrastive register is consistent with the findings of Attardo et al. (2003), who found that sarcasm is often conveyed through contrasts.

By way of summary, op. 17 was one of several pieces mentioned in the post-experiment interviews as being notably sarcastic. Based on Prokofiev's writings, it is reasonable to assume that he intended this piece to be sarcastic. Other reasons for using this piece in a free-response study of a sarcastic tone in music included Prokofiev's use of noisy timbres (both low toneness and low sensory consonance), irregularly stressed rhythms, contrastive dynamics, and contrastive articulations, all of which are assumed to contribute to a sarcastic tone of voice. Further, the title of the piece, the program that accompanied the last of the five pieces, and the unusual *"ironico"* performance marking, suggest that Prokofiev intended this piece to be heard as expressing sarcasm.

Free-Response Study: Methodology

In order to investigate the ability of listeners to *identify* sarcasm in a free-response task, an empirical study was devised. Nine listeners, a mixture of music majors and non-music majors registered in a music theory fundamentals course, agreed to participate in the study for course extra credit. They were asked to listen to three of the five pieces

from Prokofiev's op.17, *Sarcasms,* and provide verbal feedback about any aspect of the music. Pieces were randomly chosen for each participant.

Recordings were presented to listeners in 3 different conditions: 1) An audio-only condition using Eteri Andjaparidze's 1997 recording of Prokofiev's op.17. In this condition, music was played to listeners through loudspeakers; no contextual information about the work was provided. 2) An audio-visual presentation (Elena Simonyants, 2008), in which participants watched a video performance of Prokofiev's op.17. The video was presented via a computer monitor and loudspeakers; again, no contextual information was provided to the participants. 3) A title-prompt audio-only presentation in which listeners were told prior to hearing the music: "This piece was composed by Sergei Prokofiev; it is entitled *Sarcasms*." This third condition again used the Andjaparidze (1997) recording.

Each of the three conditions was devised to investigate a specific type of communication. The audio-only presentation was considered to be the most ecologically valid of the three listening conditions. If a sarcastic tone of voice could be heard in this condition, it would provide support for Seashore's idea that musical meaning, such as sarcasm, can be encoded directly into the sound wave. The audio-visual condition was used to test Rockwell's (2006) idea that "nonverbal cues are the primary cue in revealing the speaker's sarcastic intent" (p.69). Non-verbal cues of sarcasm are normally communicated through exaggerated facial expressions, such as eye movement, eyebrow movement, sneers, fake smiles, etc. (Rockwell, 2006). However, some research has found that a blank or expressionless face can also be used to signal sarcasm (Hay et al., 2002, as cited in Rockwell, 2006). The video performance that was used in this study

(Elena Simonyants, 2008) contained minimal facial expression. If a sarcastic tone could only be heard through an audio-visual presentation, it would provide support for Rockwell's theory that nonverbal cues are primary in revealing a sarcastic intention. The title-prompt condition was used to test the importance of extra-musical context in hearing a sarcastic tone of voice in music. By prompting listeners with the name of the piece, it was expected that they would contextualize their comments according to the suggestive title.

Participants provided verbal feedback after listening to each piece. These comments were transcribed by the researcher. Participants were told that there were no right or wrong answers, and that they were free to speak about any aspect of the music that came to mind. In order to avoiding asking leading questions, the experimenter used neutral prompts: "Can you say more?" "What do you mean?" "What else can you tell me about the piece" and "Would you like to move on to the next listening?"

Pieces from op.17 were chosen at random for each participant and presented in one of two sequences: 1) audio-only, then audio-visual, then the title prompted presentation, or 2) audio-visual, audio-only, then title prompted presentation. The reason for using two different sequences was to control for any possible order effect. Each recording was played once, after which participants were given as much time as they needed to discuss the recording. Each session lasted approximately 20 minutes. At the end of each session, listeners were thanked and debriefed on the purpose of the study.

Results

In total, the 9 participants provided 265 comments. None of the participants reported that they were familiar with the piece. Participant comments were organized by the type of presentation (audio; audio-visual, title prompt) and then coded for two distinct types of comments: 1) comments indicating sarcasm, and 2) comments indicating other negatively-valenced affects (excluding sarcasm). The prevalence of these elements is presented in table 10.

Table 10: Prevalence of comments that contained words indicating sarcasm or other types of negatively valenced affects as a function of presentation type.

	Presentation Type		
	Audio	Audio-Visual	Title Prompt
Sarcasm-specific comments	0%	0%	24%
Negatively-valenced comments	31%	19%	19%

Table 10 shows that no comments were made about sarcasm in either of the unprompted presentations. In carrying out an ANOVA, certain assumptions are made that all the data are independent. When running an ANOVA on the 265 collected listener comments, this assumption was violated. The reported *p* values are therefore likely to be inflated. This caveat notwithstanding, a one-way ANOVA was performed and found that the number of sarcastic comments in the title prompt presentations (M=24%) was significantly different from the audio (M=0%) and audio-visual (M=0%) presentations $F(1, 263) = 38.827$, $p < .001$]. It was not until participants were prompted with the title of

the piece that they began to make comments pertaining to the nominally sarcastic nature of the piece. Sample comments included:

- "Very sarcastically and non-chalantly changed between major and minor."
- "I could definitely see how the music was setup: heavy at the beginning and then it dies off; opposites, like saying something sarcastic."
- "After you gave me the title, I can understand why it's called *Sarcasms*."
- "You can't call a piece *Sarcasms* and not do something unexpected."

Negatively valenced comments occurred for each type of presentation. A one-way ANOVA was performed and found no significant differences for the number of negatively-valenced comments as a function of presentation type, $F(1, 263) = 3.64$, $p > .05$. However, the fact that listeners made comments pertaining to negatively valenced affects in each condition is noteworthy. Several of these comments included:

- "It reminded me of a nightmare."
- "Creepy happy feel."
- "Full of emotion and very eerie."
- "Sounded evil, like I should have been running from it."

Two additional statistical tests (independent-measures *t*-tests) were performed on the dataset. No significant difference was found between the number of sarcasm-related

comments made by music majors and non-music majors; $t(21.876) = -0.3511$, $p=.729$. Further, there was no significant difference between the two orders of presentation; $t(22.419) = 0.6518$, $p=.521$.

Discussion

Despite the prevalence of sarcastic features in Prokofiev's *Sarcasms,* this simple study demonstrated that listeners did not make spontaneous comments suggesting that they heard a sarcastic tone of voice in music. Only when listeners were told the title of the piece did they begin to comment on the apparent sarcastic nature of the piece. This finding echoes the literature on the perception of irony in text: direct cues from the author, such as carefully chosen titles, are one of the most reliable ways to signal a hidden meaning (Booth, 1974).

If we assume that listeners did not spontaneously perceive sarcasm in Prokofiev's op.17 until they were aware of the title, then how did listeners in the "standard paradigm" study distinguish sarcasm with greater accuracy than any other target affect? One explanation is that there was a strong demand characteristic. It is possible that merely having the word "sarcasm" as one of the target affects in the standard paradigm study was enough to encourage listeners to interpret aural cues differently than they would had the experimental task involved an open-ended question. Another explanation is that the listeners in the study were not familiar enough with Prokofiev's style in order to discern his sarcastic intentions; the title of the work served as the context they needed in order to interpret the sounds that they heard. Support for this latter interpretation comes from

research that has found sarcasm to be more frequent between people who are familiar with one another or share a common context (Caucci, et al. 2007). Also, Booth (1974) theorized that listeners need to be familiar with a writer's style before they are able to identify sarcastic writing; the same theory might also be applicable to sarcastic music.

One of the most commonly discussed characteristics of sarcasm in psycholinguistic literature is the role of context (Rockwell, 2006). "Linguists argue that receivers can only detect sarcasm by comparing the sarcastic utterance to its immediate context and then noting any obvious discrepancies" (Rockwell, 2006, p.64). Not only is context important, and perhaps necessary, in hearing a sarcastic tone in speech, it also appears to be equally important in music. Sarcastic speech, although prevalent in everyday life, seems to occur in very specific contexts. Sarcastic speech is more likely to occur between family and friends than strangers (Rockwell, 2003; Caucci, et al., 2008), more likely to occur in private rather than in public conversations (Toplak & Katz, 2000), and is more likely to occur when entertaining others than when working collectively on a task (Schwarz, 1998). The context in which sarcasm is likely to be expressed in music has not yet been explored. However, it could be hypothesized that, like sarcastic speech, a sarcastic tone of voice in music only occurs in limited musical contexts. Further, these contexts may be difficult to recreate in a research laboratory.

Even if a musical context suggests the presence of a sarcastic tone, there is no guarantee that listeners will be able to hear it. Listeners have a choice whether to incorporate extra-musical narrative, suggestive titles, or program notes into the context of

their listening. James Hepokowski (1992) related "program music" to a game played between composers and listeners. This game involves:

> "the calling back and forth of some sort of 'meaning' from a text (which, when lacking an established principle of interpretation, is normally capable of multiple readings), the reader is encouraged to interpret the text on the basis of the generic or descriptive implications of an overriding title" (Hepokowski, 1992, p.136).

As was seen in this study, some listeners choose to "play the game." That is, some listeners framed their listening and subsequent free-responses within the context implied by Prokofiev's title, *Sarcasms*. The findings from this study suggest that it does not take much in the way of contextual clues for composers or performers to initiate contextual "game-playing;" merely giving a piece a descriptive title is enough information to frame the listener's context. As an example, not many listeners are familiar with Krzystof Penderecki's work entitled *8'37"*. However, a large number of listeners are familiar with the piece entitled *Threnody: To the Victims of Hiroshima* (1960). Both of these pieces are in fact the same. Penderecki's work was originally composed as an abstract composition. By adding a provocative title, *Threnody* won a "much larger audience than it would otherwise have had" (Burkholder, Grout, & Palisca, 2006, p.929). This example is profound because it implies that the title of a piece may represent something greater than a title; the title may actually be able to "reframe" a listener's context.

Hepokowski's idea that listeners can "opt in" or "opt out" of an extra-musical context suggests that extra-musical contexts can be both beneficial and detrimental. Most of the early studies on extra-musical context have found that it results in an increase of listener enjoyment and greater overall satisfaction with the work (Damon, 1933; Keston, 1954; Hartshorn, 1958). A study by Zalanowski (1986) examined how different types of extra-musical content effect listeners' enjoyment; these types included: free-form mental images, abstract verbal programs, concrete analytical programs, and story programs. It was found that imagery and story programs resulted in the biggest gains of listener enjoyment (Zalanowski, 1986). However, a more recent study by Margulis (2010) found that extra-musical context had the opposite effect—a decrease in enjoyment of the music. Margulis suggests that "listeners may seek to be swept away by the music, without explicit information of its constituent elements" (Margulis, 2010, p.295), and therefore an extra-musical context may actually inhibit the listener's enjoyment. When these findings are considered together, it seems that there are reasons why a listener may or may not choose to incorporate extra-musical context into their listening.

By way of summary, context is an important issue for identifying sarcasm in both speech and music. Research has shown that extra-musical contexts, such as program notes or descriptive titles, can both enhance and inhibit musical enjoyment. Hepokowski's concept of program music as a game between composers and listeners highlights that listeners choose whether or not they wish to incorporate extra-musical features into their listening experience. Future research may wish to examine what extra-

musical contexts encourage or enable the perception of a sarcastic tone of voice within instrumental music.

Chapter 7: Conclusions and Theoretical Implications

The findings of this study, as they pertain to the two initial hypotheses, are reiterated in this chapter. Limitations and theoretical implications for these findings are also addressed.

This study began with two hypotheses, each of which is discussed below.

Hypothesis 1: The cues used to convey sarcasm in a musical setting will be similar to the cues used to convey sarcasm in speech.

Hypothesis 2: A sarcastic tone of voice can be reliably communicated within a musical context.

Hypothesis 1

The "standard paradigm" study found evidence in support of the hypothesis that the cues used to convey a sarcastic tone of voice in music are similar to the cues used to convey sarcasm in speech. Acoustical cues of a sarcastic tone of voice in both contexts included: a noisy or nasalized timbre, staccato-like articulations, and relatively greater intensity. One cue that was not shared between speech and musical contexts pertained to duration. Sarcastic speech has been found to exhibit slower durations, whereas sarcastic music was found to have relatively faster durations. Future research might investigate if

this difference is significant, or merely an artifact of the target "tones of voice" that were used in the "standard paradigm" study. Further, there may be an interaction between the duration and articulation of a sarcastic tone of voice. For example, staccato-like speech can often cause the total length of an utterance to increase. However, it is not clear if this also holds true for instrumental music. Another sarcastic tone cue that was found to be inconclusive was absolute pitch height. Mixed results have been found for the absolute pitch height of a sarcastic tone of voice in both speech and musical contexts. There is some evidence that normalized pitch level, rather than absolute pitch level, might be a more appropriate pitch cue for a sarcastic tone (Cheang & Pell, 2008). These caveats notwithstanding, many similarities were found between the cues of sarcastic speech and the cues of a sarcastic tone of voice in music.

Hypothesis 2

Evidence was found in support of the hypothesis that a sarcastic tone of voice can be communicated through music. There was a high degree of concordance between the Listener and Performer models for hearing a sarcastic tone of voice in music. Sarcasm was distinguished by listeners better than any of the other target affects (i.e. sadness, joy, and sincerity) used in this study. However, caution must be used in interpreting this result. There is an important difference between identifying and distinguishing musical affects (Frick, 1985). Although sarcasm could be successfully *distinguished*, an additional study found that sarcasm was not *identified* by listeners <u>until</u> they were given a contextual clue that implied the presence of sarcasm. This finding parallels a major

theme in the sarcastic speech literature that suggests context is an important component for detecting a sarcastic tone of voice (Rockwell, 2006).

The communicative value of sarcasm

Although this study found similarities between a sarcastic tone of voice in both speech and music contexts, it has yet to be discussed why these similarities might exist. The prevalence of sarcasm within a culture might suggest the degree to which it is valued. Not all cultures equally value sarcasm (Rockwell & Theriot, 2001). In order to discuss the communicative value of sarcasm, and also why there might be similarities between sarcastic music and speech, the various ways in which sarcasm might be used are examined below. The four types of sarcastic communication that are considered include: direct communication, indirect communication, vicarious (third-person) communication, and self-directed communication.

Direct Sarcasm

When sarcasm is used in direct communication, it might be considered a mild form of anger. Direct communication in this sense implies that the receiver of the sarcastic message is also the person who is being criticized. Consider Gibbs' (1994) sarcastic example, "You're a fine friend." When used to "directly" criticize another person, this message might be considered to communicate a mild type of anger. Although dependent on the context of the conversation, sarcastically saying "You're a fine friend" is probably less severe than stating the literal meaning, "You're a bad

friend." In such cases, sarcasm functions to lessen the severity of the intended insult or anger (Claridge, 2002, as cited in Rockwell, 2006). The utterance above might be considered to be "half-angry" in that it is semantically positive but intended to be negative.

One might consider if there is a musical equivalent for this type of direct sarcasm—one in which performers use sarcasm to criticize their listeners. This possibility seems unlikely. However, there are some performers who have expressed that that the sole motivation behind their music is to make money. In popular musical studies, this "money motivation" can often cause an artist to be perceived as insincere (Parzer, 2009). Regardless, there is no rational reason for listeners to support a musician who deliberately insults them. Therefore, it seems unlikely that the "direct communication" of sarcasm has much value in a musical context.

Indirect Sarcasm

When sarcasm is used in indirect communication, it might be considered a way of signaling social disgust. Indirect communication refers to a situation in which the person who is being criticized through sarcasm is not being directly addressed, but rather the criticism is being communicated to a third party. Consider Rockwell's (2006) sarcastic example: "I bet that resume will get him the job." When this sentence is spoken sarcastically to a third party, its purpose may be to express a form of social disgust. Stating the literal meaning of this sentence ("I bet that resume will not get him the job") might be considered less poignant than the non-literal utterance. In this case, using

sarcasm seems to enhance the sender's message (Toplak & Katz, 2000). Rather than merely stating that "the resume" is not good enough, the sarcastic version of this message may also imply that "the person who wrote the resume" is not good enough. Indirect sarcasm in this case appears to covertly criticize the author under the cover of criticizing the work. However, recall that in this "indirect communication" scenario, the author is not even being addressed. If the speaker is using sarcasm to express social disgust, to whom are they expressing it? One theory regarding the communicative value of sarcasm is that it can be used to signal social alliances (Ducharme, 1994). As social beings, humans rely upon one another to meet basic survival needs. Sarcasm can function to covertly signal a speaker's negative feelings about someone to a third party, often in a semantically "safe" manner (Giora, 2001). Those who share the same beliefs as the speaker, or listeners who know the speaker well, are likely to decode the speaker's true intention. Gibbs (2000) found that sarcasm is normally directed towards a person—such as the speaker, listener, or a person who is not present—rather than an object or event. Since there is no reason to form an alliance with an object or an event, this finding supports the hypothesis that one of sarcasm's functions is to signal social alliances.

The question arises if "indirect" sarcasm is relevant to musical communication. Is there any benefit for musical producers (i.e. composers and performers) to utilize sarcasm to signal social alliances? Through sarcasm, it may be possible for composers and performers to express negativity or potentially insulting viewpoints in a socially normative way. Composers might be able to signal alliances to listeners who share similar beliefs, while remaining somewhat semantically neutral to those with different

beliefs. An example in which this type of sarcasm would be valuable would be revolutionary music (political, cultural, social, etc). Sheinberg (2000) noted this type of sarcasm in the music of Shostakovich. Sarcasm has been shown to not only promote group solidarity (Ducharme, 1994), but also to be a form of protest in powerless groups (Giora, 2001). Of the many functions of sarcasm, one that may be common to both sarcastic speech and sarcastic music is the signaling of social alliances.

Vicarious Sarcasm

Another context for considering the communicative value for sarcasm is vicarious (i.e. third person) communication. In this case, vicarious communication refers to self-alienation, where a speaker can communicate from another's point of view (Haiman, 1998). This type of communication is related to Nehaniv's (1998) category of third-person affects. For example, empathy refers to the ability to try on the emotions of another. The value of an affect like empathy is that it allows us to tune our context so as to understand the context of others. Haiman (1998) claimed that sarcasm involves some degree of self-alienation. As opposed to an affect like sympathy, which involves perceiving the pain of another, sarcasm may involve speaking and "expelling" the words of those that disgust us. Rozin and Fallon (1987) note that the limitations of self are important in determining what disgusts us. For example, bodily fluids, such as blood, salivia, etc, are not disgusting when they are inside our sense of self space. However, as soon as these fluids leave our sense of bodily self, they become disgusting. In this regard, "vicarious" sarcasm may function to expel or vent feelings of social disgust.

Could there be a musical equivalent for vicarious sarcasm? One example would be musical parody. Sarcasm and parody are related, but they are not the same thing. Clift (1999) describes the difference between these two as "sarcasm is a form of parody restricted by tone. Parody is… a form of irony restricted by…form" (Clift, 1999, p.549). A musical parody may very well involve the type of self-alienation described by Haiman (1998). It may also be functioning as a form of social disgust. Beyond parody, musical sarcasm may also take the form of making music from another perspective. An example of such musical sarcasm might be found in Bernhard Lang's opera entitled, "I Hate Mozart."

Self-directed Sarcasm

One last context for considering the communicative value for sarcasm is self-communication. In this case, self-communication refers to the use of sarcasm as a form of self-criticism. Self-directed sarcasm has been theorized to be more frequently used by females (Rockwell, 2006). For example, imagine a student who has just made an incredibly basic mistake in front of their class and then says: "Wow, aren't I smart?" In this example, the self-directed sarcasm seems to be used for self-preservation. The literal meaning of this sentence is incongruent with the context (the student just made an unacceptable mistake). However, the non-literal sarcastic meaning of the utterance ("Wow, I am not smart") is also somewhat incongruent because sarcasm tends to be an indicator of verbal ability and education (Ball, 1965).

Could there be a musical equivalent for self-directed sarcasm? One case in which musicians might need to preserve their sense of self is after making a mistake. In order for a musician to both acknowledge the mistake and to preserve their status, they might elect to draw attention to the mistake, or possibly even repeat it. Consider the following quote by the famous jazz violinist Joe Venuti: "If you're gonna make a mistake, make it so loud everybody else sounds wrong.[22]" This quote suggests that self-directed sarcasm may also serve a self-preservation function in music.

By way of summary, there are multiple contexts in which the value of sarcasm might be considered, including direct, indirect, vicarious, and self-directed communication. With regards to musical sarcasm, likely functions include signaling social alliances, expressing forms of anger or disgust, and self-preservation. Even though these functions have been presented as singular and concrete, the function of sarcasm likely involves a mixture of functions. Further, the functions of sarcasm are likely to change depending on the context of the situation.

Limitations & Future Directions

This study had several limitations: 1) the standard paradigm suffered from a low level of ecological validity and a high level of variability, 2) assumptions had to be made regarding the interpretation of notational symbols when discussing Prokofiev's op.17, and 3) demand characteristics were likely to be present in the free response study.

[22] Obtained from: http://athensjazztet.com/jazz-quotations/

The results obtained from the standard paradigm study make no claim to being ecologically valid. Performers were specifically instructed to record each passage until they were satisfied that they had conveyed sarcasm, as well as joy, sadness, and sincerity. The acoustical cues in the Performer model are not assumed to be applicable to other musical contexts as they are on the performer's recordings. In a similar manner, the acoustical cues in the Listener model may not apply to other musical contexts. Chapter 5 examined the concordance between the Performer and the Listener models; in some ways, these two models were incomparable. The Performer model was constructed from a task which required musicians to distinguish between *four* different tones of voice. The Listener model was constructed from a task in which participants were required to distinguish between *eight* different tones of voice. Had the performers and listeners been given the same experimental task (i.e. both distinguished between the same four tones of voice), it is likely that the degree of cue concordance and the amount of "successful" matches would have increased. The lure target affects which were used in the listener study were chosen informally. Future research may wish to adopt a more formal method for choosing lure targets.

The data from the standard paradigm study contained a large amount of variability. The 27 variables from the PRAAT voice reports were used to account not only for the four target tones of voice, but also the 5 different types of passages which were played on 14 different instruments. Despite this large variability, the multinomial logit model found four significant factors used by performers to convey a sarcastic tone of voice in instrumental music. This does not imply that there are only four factors which

distinguished the four affective tones of voice. The amount of relevant factors is unknown. Now that some exploratory work has been completed, future studies may wish to limit their experimental paradigms to one or two factors of a sarcastic tone of voice in music.

The study in chapter 6 examined listeners' free responses to Prokofiev's op. 17, *Sarcasms*. This piece was chosen for several reasons, but the main reason was that *when performed*, it contained several acoustical cues of a sarcastic tone of voice. Notational symbols and acoustical symbols should not be confused (Kendall and Carterette, 1990). Musical notes written on a staff do not have a tone of voice. It is only when these notes are realized (i.e. performed) that they can have a tone of voice. Therefore, when discussing listeners' free-responses to Prokofiev's op.17, it is important to note that the tone of voice cues do not stem directly from the notation, rather these cues come from a combination of both Prokofiev's notational symbols and the way in which the performer chooses to interpret them. Different performances of Prokofiev's op.17 should be capable of resulting in different degrees of perceived musical sarcasm. Future research may wish to pursue this hypothesis. Also, the techniques for conveying irony and sarcasm in speech are different from the techniques for conveying these same non-literal meanings in a written text. Future studies might examine the concordance between non-literal writing and non-literal music notation. Some of the known techniques used in non-literal writing include: repetition (Gibbs, 1994), unexpected pauses (Caucci, et al., 2008), and unexpected foreign terms (Kruez & Caucci, 2007).

One last limitation pertains to a likely demand characteristic in the free-response study presented in chapter 6. This study encouraged listeners to speak at length about the musical excerpts that they heard. When asking listeners (some of who were not trained musicians) to speak about any aspect of the music, it seems somewhat obvious that they would discuss the excerpt with regards to its title. However, the constant probing of the experimenter may have caused participants to make comments about the title of the piece that they normally would have omitted. Musical context is difficult to control in an experimental setting. One way in which future research might glean more information from free-response studies (such as the one in ch.6) is to account for the order of the participants' comments. It would be interesting to know if participants' free-responses were patterned. For example, did participants begin by discussing the most obvious features of the music? Did they build their free-responses around a common theme? It seems reasonable that the participant comments following experimenter prompts for more information might be considered less ecologically valid.

Conclusions

Researchers from many fields, including English, psychology, linguistics, psycholinguistics, communication, philosophy, and neuroscience, have addressed the topic of sarcasm. Along with these disciplines, music research can also contribute to the study of sarcasm. This study found evidence in support of the hypothesis that a sarcastic tone of voice in music is similar to a sarcastic tone of voice in speech. From this finding, it might be further hypothesized that sarcastic music and sarcastic speech serve similar

functions, such as signaling social alliances, expressing anger and disgust, and acting as a form of self-preservation. Communicating non-literal intentions represents a high degree of communication sophistication. This type of communication appears to be possible in both music and speech. Future research will hopefully continue to explore the limits and boundaries of musical communication and our musical abilities.

References

Abbate, C. (1991). *Unsung voices: opera and musical narrative in the nineteenth century*. Princeton, NJ: Princeton University Press.

Adachi, T. (1996). Sarcasm in Japanese. *Studies in Language, 20*(1), 1-36.

Andjaparidze, E. (1997). *Prokofiev piano music* (vol.1) [sound recording]. Naxos.

Arnest, M. (1996). Prokofiev as Pianist. Retrieved April 29, 2011, from http://home.earthlink.net/~marnest/prokofiev.html.

Attardo, S., Eisterhold, J., Hay, J., & Poggi, I. (2003). Multimodal markers of irony and sarcasm. *International Journal of Humor Research, 16*(2), 243-260.

Ball, D.W. (1965). Sarcasm as sociation: The rhetoric of interaction. *Canadian Review of Sociology and Anthropology, 2*(4), 190-198.

Barthes, R. (1968). *Image – Music – Text* (Stephen Heath, trans.). New York, NY: Harper Collins.

Bengtsson, I., & Gabrielsson, A. (1980). Methods for analyzing performances of musical rhythm. *Scandinavian Journal of Psychology, 21,* 257-268.

Bhatara, A., Tirovolas, A.K., Duan, L.M., Levy, B., and Levitin, D.J. (2011). Perception of emotional expression in musical performance. *Journal of Experimental Pyschology: Human Perception and Performance*. Advance online publication.

Boersma, P.,& Weenink, D.(2010). Praat: doing phonetics by computer [Computer program]. Version 5.1.32, retrieved 30 April 2010 from http://www.praat.org/.

Bonds, M.E. (1991). Haydn, Laurence Sterne, and the origins of musical irony. *Journal of the American Musicological Society, 4*(1), 57-91.

Booth, W.C. (1974). *A rhetoric of irony*. Chicago, IL: University of Chicago Press.

Boylan, S. (1979). *An analysis of tonal structure in Sergei Prokofiev's Sarcasms, op. 17, for piano.* (Master's thesis). Bloomington, IN: University of Indiana.

Brackett, D. (2002). "Where's It At": Postmodern theory and the contemporary musical field. In J.I. Lochhead, J.H. Auner (Eds.), *Postmodern music/postmodern thought* (pp.207-234). New York, NY: Routledge.

Brunswik, E. (1952). *The conceptual framework of psychology*. (International Encyclopedia of Unified Science, Volume 1, Number 10.) Chicago, IL: The University of Chicago Press.

Bryant, G.A., & Fox Tree, J.E. (2002). Recognizing verbal irony in spontaneous speech. *Metaphor and Symbol*, *17*(2), 99-117.

Bryant, G.A. (2010). Prosodic contrasts in ironic speech. *Discourse Processes, 47*, 545-566.

Burkholder, J.P., Grout, D.J., & Palisca, C.V. (2006). *A history of western music* (7th ed.). New York, NY: W.W. Norton & Company.

Caucci, G.M., Kreuz, R.J., Buder, E.H. (2008, November). "I'm always sincere": What is salient in sarcasm? Poster presented at the 49th annual meeting of the Psychonomic Society, Chicago, IL.

Caucci, G. M., Kreuz, R. J., & Buder, E. H. (2007, November). Acoustic analyses of the sarcastic tone of voice. Paper presented at the 48th annual meeting of the Psychonomic Society, Long Beach, CA.

Cheang, H.S., & Pell, M.D. (2008). The sound of sarcasm. *Speech Communication, 50,* 366-281.

Claridge, C. (2002). Hyperbolic expressions in speech. Paper presented at the 23rd, International Conference on English Language Research in Computerized Corpora of Modern and Medieval English, Göteborg, Sweden, May 22-26.

Clarke, E.F. (2005). *Ways of listening: An ecological approach to the perception of musical meaning.* New York, NY: Oxford University Press.

Clift, R. (1999). Irony in conversation. *Language and Society, 28*, 523-553.

Cooksey, R. W. (1996). *Judgment Analysis: Theory, methods, and applications*. San Diego, CA: Academic Press.

Cutler, A.H. (1974). On saying what you mean without meaning what you say. *Papers from the annual meeting of the Chicago Linguistic Society* (pp. 117-127). Chicago: University of Chicago Press.

Damon, K. F. (1933). *Program notes for the listener to music.* New York: Freybourg.

Davis, T.F., & Womack, K. (2002). *Formalist criticism and reader reponse theory.* New York, NY: Palgrave.

Desain, P. & Honing, H. (1995). *Music, mind, machine: computational modeling of temporal structure in musical knowledge and music cognition*. [unpublished manuscript].

Dill, H.J. (1989). Romantic irony in the works of Robert Schumann. *The musical quarterly, 73*(2). 172-195.

Ducharme, L. (1994). Sarcasm and interactional politics. *Symbolic Interactions, 17*, 51-62.

Fairbanks, G., & Pronovost, W. (1939). An experimental study of the pitch characteristics of the voice during the expression of emotions. *Speech Monographs, 6*, 87-104.

Fairbanks, G. (1940). Recent experimental investigations of vocal pitch in speech. *Journal of the Acoustical Society of America, 11,* 457-466.

Fiske, H.E. (1993). *Music cognition and aesthetic attitudes*. Edwin Mellen Press: Lewiston, NY.

Fiske, H.E. (1996). *Selected theories of music perception*. Edwin Mellen Press: Lewiston, NY.

Fónagy, I., & Magdics, K. (1963). Emotional patterns in intonation and music. *Zeitschrift für Phonetik Sprachwissenschaft und Kommunikationsforschung, 16,* 293-326.

Fónagy, I. (1971). Synthèse de l'ironie. *Phonetica, 23,* 42-51.

Frick, R.W. (1985). Communicating emotion: the role of prosodic features. *Psychological Bulletin, 97*(3), 412-429.

Gabrielson, A., & Juslin, P.N. (1996). Emotional expression in music performance: between the performer's intention and the listener's experience. *Psychology of Music, 24,* 68-91.

Gabrielsson, A., & Lindström, E. (2010). The role of structure in the musical expression of emotions. In P.N, Juslin & J.A. Sloboda (Eds.), *Handbook of music and emotion: theory, research, and applications* (pp.367-400). New York, NY: Oxford University Press.

Gibbs, R.W. (1983). Understanding sarcasm and irony. In R. Hoffman (Chair)*, Recent research on metaphor.* Symposium conducted at the annual meeting of the American Psychological Association, Anaheim, CA.

Gibbs, R.W. (1986). On the psycholinguistics of sarcasm. *Journal of Experimental Psychology: General, 115*(1), 3-15.

Gibbs, R.W. (1994*). The Poetics of the Mind.* Cambridge University Press: New York, NY.

Gibbs, R.W. (2000). Irony in talk among friends. *Metaphor and Symbol, 15*(1&2), 5-27.

Giora, R. (2001). Masking one's themes: Irony and the politics of indirectness. In M. M. Louwerse's and W. van Peer (Eds.). *Thematics: Interdisciplinary Studies* (pp. 283-300). Philadelphia, PA: John Benjamins Publishing Co.

Grey*,* J.M. (1978)*.* Timbre discrimination in musical patterns. *Journal of the Acoustical Society of America, 64,* 461-412.

Grice, H.P. (1975). Logic and conversation. In P. Cole & J.L. Morgan (Eds.), *Syntax and Semantics: Vol. 3, Speech Acts* (pp.41-58), New York, NY: Academic Press.

Grice, H.P. (1978). *Pragmatics.* New York, NY: Academic Press.

Haiman, J. (1998). *Talk is cheap: sarcasm, alienation, and the evolution of language*. New York, NY: Oxford University Press.

Hartshorn, W. C. (1958). The role of listening. In W. C. Hartshorn (Ed.), *Basic concepts in music education* (pp. 261-291). Chicago, IL: Chicago University Press.

Hatten, R. (1994). *Musical meaning in Beethoven: markedness, correlation, and interpretation*. Bloomington, IN: Indiana University Press.

Hay, J., Attardo, S., Nelms, J., Poggi, I. (2002). Multimodal markers of irony and sarcasm. Paper presented at the *Georgetown University Roundtable on Linguistics*. Washington, D.C.

Hevner, K. (1935). The affective character of the major and minor modes in music. *The American Journal of Psychology, 47*(1), 103-118.

Heller, J., & Campbell, W. (1982). Music communication and cognition. *Bulletin of the Council for Research in Music Education, 72,* 1-15.

Hepokoski, J. (1992). Fiery-Pulsed Libertine or Domestic Hero? : Strauss's Don Juan Reinvestigated. In Bryan Gilliam (ed.), *Richard Strauss : New Perspectives on the Composer and His Work* (pp. 135-175). Durham, NC: Duke University Press.

Huron, D. (1991). Tonal consonance versus tonal fusion in polyphonic sonorities. *Music Perception, 9*(2), 135-154.

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

Huron, D. (2004). Music-engendered laughter: an analysis of humor devices in PDQ Bach. In S.D. Lipscomb, R. Ashley, R.O. Gjerdingen, and P. Webster (Eds.), *Proceedings of the 7th International Conference on Music Perception and Cogition.* pp.673-676. Adelaide: Casual Productions.

Huron, D. (2006). *Sweet anticipation: music and the psychology of expectation*. Cambridge, MA: The MIT Press.

Huron, D. (2008). A comparison of average pitch height and interval size in major- and minor-key themes: Evidence consistent with affect-related pitch prosody. *Empirical Musicology Review, 3*(2), 59-63.

Huron, D. (2010). An ear for music [unpublished manuscript].

Huron, D., & Davis, M. (2010). The effect of scale degree modifications on average interval size. In S.M. Demorest, S.J. Morrison, P.S. Campbell (Eds.), *Proceedings*

of the 11th International Conference on Music Perception and Cognition. Seattle, Washington: Causal Productions, pp. 439-444.

Huron, D., Kinney, D., & Precoda, K. (2006). Influence of pitch height on the perception of submissiveness and threat in musical passages. *Empirical Musicology Review, 1*(3), 170-177.

Huron, D., Yim, G., & Chordia, P. (2010). The effect of pitch exposure on sadness judgments: An association between sadness and lower than normal pitch. In S.M. Demorest, S.J. Morrison, P.S. Campbell (Eds.), *Proceedings of the 11th International Conference on Music Perception and Cognition.* Seattle, Washington: Causal Productions, pp. 63-66.

Johnson, R.B. (2009). *Selected Topics in the Perception and Interpretation of Musical Tempo*. (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses. (UMI Number: 3425288).

Irony. 2011. In *Merriam-Webster Online Dictionary*. Retrieved April 02, 2010, from http://www.merriam-webster.com/dictionary/irony.

Juslin, P.N. (1995). Emotional communication in music viewed through a Brunswikian lens. In G. Kleinen (Ed.), *Music and expression: Proceedings of the Conference of DGM and ESCOM*, Bremen, 1995 (pp.21-25). Bremen, Germany: University of Bremen.

Juslin, P.N. (1997). Emotional communication in music performance: A functionalist perspective and some data. *Music Perception, 14*(4), 383-418.

Juslin, P.N. (2000). Cue utilization in communication of emotion in music performance: relating performance to perception. *Journal of Experimental Psychology: Human Perception and Performance, 26*(6), 1797-1813.

Juslin, P.N., & Laukka, P. (2003). Communication of emotions in music performance through cognitive feedback. *Musicae Scientiae, 4,* 151-183.

Juslin, P.N., & Laukka, P. (2004). Expression, perception, and induction of musical emotions: a review and a questionnaire study of everyday listening. *Journal of New Music Research, 33,* 217-238.

Juslin, P.N., & Lindström, E. (2003). Musical expression of emotions modeling composed and performed features. Paper presented at *the Fifth Conference of the European Society for the Cognitive Sciences of Music (ESCOM)*, Hanover, Germany, September 2003.

Juslin, P.N., & Madison, G. (1999). The role of timing patterns in recognition of emotional expression from music performance. *Music Perception, 17*(2), 197-221.

Juslin, P.N., & Sloboda, J.A. (2001). *Music and emotion: theory and research*. New York, NY: Oxford University Press.

Juslin, P.N., & Sloboda, J.A. (2010). Introduction: aims, organization, and terminology. In P.N, Juslin & J.A. Sloboda (Eds.), *Handbook of music and emotion: theory, research, and applications* (pp.3-12). New York, NY: Oxford University Press.

Juslin, P.N., & Timmers, R. (2010). Expression and communication of emotion in music performance. In P.N, Juslin & J.A. Sloboda (Eds.), *Handbook of music and emotion: theory, research, and applications* (pp.453-489). New York, NY: Oxford University Press.

Kaestner, G. (1909). Untersuchungen über den Gefuehlseindruck unanalysierter Zweiklaenge. *Psychologische Studien, 4,* 473-504.

Kameoka, A., & Kuriyagawa, M. (1969). Consonance theory, part I: Consonance of dyads. *Journal of the Acoustical Society of America, 45*(6), 1451-1459.

Kaminska, Z. & Woolf, J. (2000). Melodic line and emotion: Cooke's theory revisited. *Psychology of Music, 28,* 133-153.

Kendall, R.A., & Carterette, E.C. (1990). The communication of musical expression. *Music Perception, 8*(2), p. 129-163.

Keston, M. J. (1954). An experimental evaluation of the efficacy of two methods of teaching music appreciation. *Journal of Experimental Education, 22*, 215-226.

Kramer, L. (1991). *Music as cultural practice: 1800-1900*. University of California Press: Berkeley, CA.

Kreuz, R.J., & Roberts, R.M. (1995). Two cues for verbal irony: hyperbole and the ironic tone of voice. *Metaphor and Symbolic Activity, 10*, 21-30.

Lee, C.J., & Katz, A.N. (1998). The differential role of ridicule in sarcasm and irony. *Metaphor and Symbol, 13*(1), 1-15.

Lerdahl, F. & Jackendoff, R. (1983). *A Generative Theory of Tonal Music*. Cambridge, MA: The MIT Press.

Lerdahl, F. (1988). Cognitive constraints on compositional systems. In J.A. Sloboda (Ed.), *Generative processes in music: The psychology of performance, improvisation, and composition* (pp. 231-259). New York, NY: Clarendon Press/Oxford University Press.

Madison, G. (2000). Properties of expressive variability patterns in music performances. *Journal of New Music Research, 29*(4), 335-356.

Maher, T. F., & Berlyne, D.W. (1982). Verbal and exploratory responses to melodic musical intervals. *Psychology of Music, 10,* 11-27.

Margulis, E. (2010). When program notes don't help: music descriptions and enjoyment. *Psychology of Music, 38*(3), 285-302.

Masters Music Publications (1998). Sarcasms : for solo piano, op. 17 by Sergei Prokofiev. Boca Raton, FL: 1-20.

McAdams, S. (1999). Perspectives on the contribution of timbre to musical structure. *Computer Music Journal, 23*(3), 85-102.

Menard, S. (2002). *Logistic Regression Analysis*. Thousand Oaks, CA: Sage.

Meyer, L.B. (1956). *Emotion and Meaning in Music*. Chicago, IL: University of Chicago Press.

Meyer, L.B. (1967). *Music, the arts, and ideas.* Chicago, IL: University of Chicago Press.

Milosky, L. M., & Ford, J. A., (1997). The role of prosody in children's inferences of ironic intent. *Discourse Processes, 23,* 47–61.

Moles, A. (1958/1966). *Théorie de l'information et perception esthétique*. Paris. Trans. (1966) as *Information theory and aesthetic perception.* Urbana, IL: University of Illinois Press.

Nehaniv, C. (1998). The first, second, and third person emotions: grounding adaptation in a biological and social world. In: D. Cañamero, C. Numaoka, P. Petta, (Eds.), *From Animals To Animats: Fifth International Conference of the Society for Adaptive Behavior* (SAB 98) Workshop on Grounding Emotions in Adaptive Systems (Zürich, Switzerland - 21 August 1998), pp. 43-47.

Nelms, J. (2002). Responding to sarcasm and irony. Paper presented at the *Annual Convention of the International Society of Humor Studies* (July 7). Forli, Italy.

Nussbaum, C. O. (2007). *The musical representation: meaning, ontology, and emotion.* Cambridge, MA: The MIT press.

Panksepp, J. (1998). *Affective neuroscience: the foundations of human and animal emotions.* New York, NY: Oxford University Press.

Parzer, M. (2009). Ich hör' alles querbeet! Zum Bedeutungsverlust von Authentizität als ethisches Bewertungskriterium populärer Musik. *SAMPLES: Online-Publikationen des Arbeitskreis Studium Populärer Musik, 8.*

Patel, A.D. (2008). *Music, language, and the brain*. New York, NY: Oxford University Press.

Patterson,R.D., Smith, D., Dinter, R., Walters, T. (2008). "Size information in the production and perception of communication sounds." In W.A. Yost, A.N. Popper, and R.R. Fay (Eds.), *Auditory Perception of Sound Sources* (pp.43-75). New York, NY: Springer.

Phillips, A. (tran). (2006). *Sergey Prokofiev diaries, 1907-1914: prodigious youth*. Ithica, NY: Cornell University Press.

Plazak, J., & Huron, D. (2011). The first three seconds: listener knowledge gained from brief musical excerpts. *Musicae Scientiae, 15*(1), 29-44.

Plomp, R. & Levelt, W.J.M. (1965). Tonal consonance and critical bandwidth. *Journal of the Acoustical Society of America, 38*, 548-560.

Post, O., & Huron, D. (2009). Music in minor modes is slower (except in the Romantic Period). *Empirical Musicology Review, 4*(1), 1-9.

Pratt, C.C. (1931). *The meaning of music: a study in psychological aesthetics.* New York, NY: McGraw-Hill.

Prokofiev, O. (tran.) & Palmer, C. (ed.). (1991). *Sergei Prokofiev: Soviet diary 1927 and other writings.* Boston, MA: Northeastern University Press.

Repp, B.H. (1992). Diversity and commonality in music performance: an analysis of timing microstructure in Schumann's "Träumerei.*" Journal of the Acoustical Society of America, 92*(5), 2546-2568.

Rockwell, P. (2000). Lower, slower, louder: vocal cues of sarcasm. *Journal of Psycholinguistic Research, 29*(5), 483-495.

Rockwell, P. (2004). The sarcastive: a linguistic and paralinguistic analysis. *Language and Social Interaction Division, International Communication Association Annual Convention*, May, 2004, New Orleans, LA.

Rockwell, P. (2006). *Sarcasm and other mixed message: the ambiguous ways people use language.* Lewiston, NY: Edwin Mellen Press.

Rockwell, P. (2007). Vocal features of conversational sarcasm: a comparison of methods. *Journal of Psycholinguistic Research, 36*, 361-369.

Rockwell, P., & Theriot, E. (2001). Culture, gender, and gender mix in encoders of sarcasm. *Communication Research Reports, 18*, 44-52.

Rossing, T., Moore, F., & Wheeler, P. (2002). *The Science of Sound* (3rd ed). San Francisco: Addison Wesley.

Rozin, P. & Fallon, A.E. (1987). A perspective on disgust. *Psychological Review, 94*(1), 23-41.

Serafine, M.L. (1988). *Music as cognition: the development of thought in sound.* New York, NY: Columbia University Press.

Schaffer, R. (1982). Are there consistent vocal clues for irony? In C. S. Masek, R. A., Hendrick, & M. F. Miller (Eds.). *Papers from the Parasession on Language and Behavior* (pp. 204–210). Chicago, IL: Chicago Linguistic Society.

Scherer, K.R. (1995). Expression of emotion in voice and music. *Journal of Voice, 9*(3), 235-248.

Scherer, K. R., & Oshinsky, J.S. (1977). Cue utilization in emotion attribution from auditory stimuli. *Motivation and Emotion, 1*(4), 331-346.

Schutz, M., Huron, B., Keeton, K., & Loewer, G. (2008). The happy xylophone: Acoustic affordances restrict an emotional palate. *Empirical Musicology Review, 3*(3), 126-135.

Schwarz, N. (1998). Communication in standardized research situations: A Gricean perspective. In S. Fussell & R.J. Kreuz (Eds.), *Social and cognitive approaches to onterpersonal communication* (pp. 39-66). Mahwah, NJ: Erlbaum.

Shaker, H.A. Physiology of the Voice [Online reference article]. Retrieved March 27, 2011 from http://www.hani-shaker.com/Voice/V_Physiology.pdf.

Searle, J. (1979). *Expression and meaning*. Cambridge: Cambridge Univ. Press.

Seashore, C.E. (1938). *The psychology of music.* New York, NY: McGraw-Hill.

Seashore, C.E. (1947). *In search of beauty in music*. New York, NY: The Ronald Press Company.

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

Sheinberg, E. (2000). *Irony, satire, parody, and the grotesque in the music of Shostakovich*. Burlington, VT: Ashgate.

Simonyants, E. (2008). Sergei Prokofiev: Sarcasms – five pieces for piano, op.17 [music video]. Retrieved November 1, 2010 from http://www.youtube.com/watch?v=e-Swxz9QctU.s

Sloboda, J. (1985). *The musical mind: the cognitive psychology of music*. New York, NY: Oxford University Press.

Spruyt, A., Clarysse, J., Vansteenwegen, D., Baeyens, F., & Hermans, D. (2010). Affect 4.0: A free software package for implementing psychological and psychophysiological experiments. *Experimental Psychology, 57*(1), 36-45.

Tannen, D. (1984). *Conversational Style: Analyzing talk among friends*. Norwood, NJ: Ablex.

Tompkins, J.P. (1980). An introduction to reader-response criticism. In J.P. Tompkins (Ed.), *Reader-response criticism: from formalism to post-structuralism* (pp. xi-xxvi). Baltimore, MD: John Hopkins University Press.

Toplak, M. & Katz, A.N. (2000). On the use of sarcastic irony. *Journal of Pragmatics, 32*, 1467-1488.

Turner, B. & Huron, B. (2008). A comparison of dynamics in major- and minor-key works. *Empirical Musicology Review, 3*(2), 64-68.

Utsumi, A. (2000). Verbal irony as implicit display of ironic environment: distinguishing ironic utterances from nonirony. *Journal of Pragmatics, 32*, 1777-1806.

Vos, J. (1986). Purity ratings of tempered fifths and major thirds. *Music Perception, 3*(3), 221-258.

Zalanowski, A.H. (1986). The effects of listening instructions and cognitive style on music appreciation. *Journal of Research in Music Education, 34*(1), 43-53.

Appendix A: Directions supplied to performers

In this research we are studying how musicians are able to communicate different feelings. Specifically, for this study we are looking at joy, sadness, sincerity, and sarcasm. We want you to practice five simple passages in each of the four target feelings. (Refer to the accompanying notation.) After one week, you will be asked to record the 20 passages (5 passages x 4 target feelings) that you have practiced.

The first passage is "Happy Birthday." Attached are four notations in different transpositions. Choose one of the transpositions that seems right for your instrument or vocal range. Then practice the piece developing four different interpretations until you are convinced that a listener will recognize that you are conveying either joy, sadness, sincerity or sarcasm. During the recording session we will ask you to play all four versions.

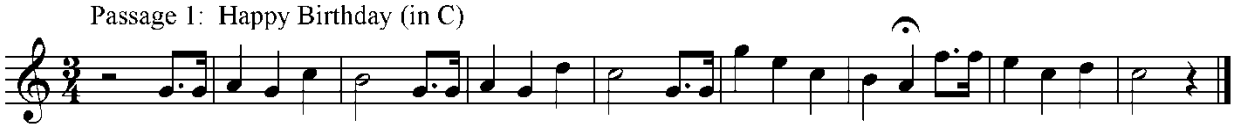
For the second passage we want you to practice playing an ascending major third (see notation) in the four target feelings: joy, sadness, sincerity or sarcasm.

For the third passage, we want you to practice playing a descending major third — again using the four target feelings.

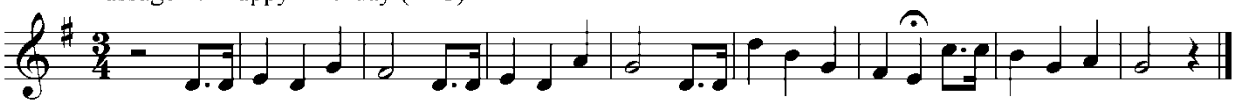
For the fourth passage, we want you to practice playing nine repeated notes (see notation) — once again conveying the four target feelings to your listeners.

Finally, we want you to make up your own passages. You can play whatever you like for each of the four target feelings, but each passage should not be longer than 10 seconds in duration. You can play different things for each of the four feelings.

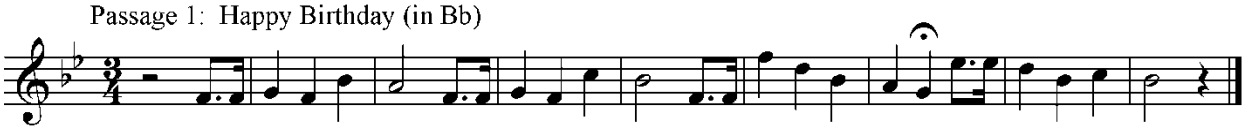
Passage 1: Happy Birthday (in C)



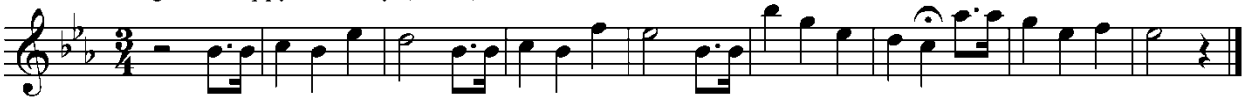
Passage 1: Happy Birthday (in G)



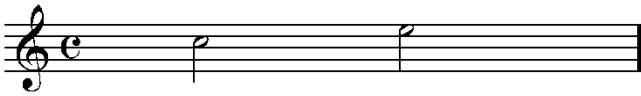
Passage 1: Happy Birthday (in Bb)



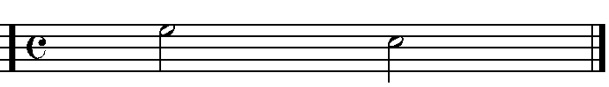
Passage 1: Happy Birthday (in Eb)



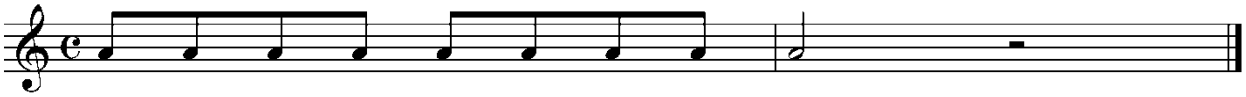
Passage 2: Ascending Major 3rd
(Can be played in any octave)



Passage 3: Descending Major 3rd
(Can be played in any octave)



Passage 4: Nine Repeated Notes (Can be played in any octave)



Passage 5: Improvisation (Must be less than 10 seconds)

Appendix B: Sample PRAAT Voice Report

-- Voice report for 1. LongSound asax_hb_sar --

Time range of SELECTION
  From 0 to 16.092358 seconds (duration: 16.092358 seconds)
Pitch:
  Median pitch: 289.382 Hz
  Mean pitch: 290.618 Hz
  Standard deviation: 58.469 Hz
  Minimum pitch: 219.511 Hz
  Maximum pitch: 463.550 Hz
Pulses:
  Number of pulses: 2350
  Number of periods: 2326
  Mean period: 3.441654E-3 seconds
  Standard deviation of period: 0.657451E-3 seconds
Voicing:
  Fraction of locally unvoiced frames: 50.125% (805 / 1606)
  Number of voice breaks: 22
  Degree of voice breaks: 48.426% (7.792825 seconds / 16.092358 seconds)
Jitter:
  Jitter (local): 0.249%
  Jitter (local, absolute): 8.556E-6 seconds
  Jitter (rap): 0.103%
  Jitter (ppq5): 0.118%
  Jitter (ddp): 0.310%
Shimmer:
  Shimmer (local): 2.862%
  Shimmer (local, dB): 0.395 dB
  Shimmer (apq3): 0.731%
  Shimmer (apq5): 1.092%
  Shimmer (apq11): 2.312%
  Shimmer (dda): 2.193%
Harmonicity of the voiced parts only:
  Mean autocorrelation: 0.984186
  Mean noise-to-harmonics ratio: 0.017503
  Mean harmonics-to-noise ratio: 24.930 dB

Appendix C: Performer Multinomial Logit Model

Baseline=Joy	Estimate	Standard Error	t-value	Pr(>|t|)	
Sadness	0.21477	0.7233839	0.2969	0.766546	
Sarcasm	-0.5122517	0.7234343	-0.7081	0.478894	
Sincerity	0.973187	0.6802032	1.4307	0.152508	
Sadness:Duration	0.2467821	0.0453488	5.4419	5.27E-08	***
Sarcasm:Duration	0.0064992	0.0490756	0.1324	0.894642	
Sincerity:Duration	0.1319026	0.0426736	3.091	0.001995	**
Sadness:Maximum Pitch	-0.003505	0.0017546	-1.9976	0.045756	*
Sarcasm:Maximum Pitch	-0.0019539	0.0017076	-1.1443	0.252516	
Sincerity:Maximum Pitch	-0.0033962	0.0016395	-2.0715	0.038312	*
Sadness:Unvoiced Frames	0.0151767	0.0193084	0.786	0.431858	
Sarcasm:Unvoiced Frames	0.0657239	0.0147237	4.4638	8.05E-06	***
Sincerity:Unvoiced Frames	-0.0345031	0.0215173	-1.6035	0.108822	
Sadness:Voice Breaks	-0.2406302	0.056851	-4.2326	2.31E-05	***
Sarcasm:Voice Breaks	-0.0068261	0.0353176	-0.1933	0.846741	
Sincerity:Voice Breaks	-0.0451249	0.0375744	-1.2009	0.229772	

Appendix D: Performer Model ANOVAs

Duration~Affect

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Affect	3	1067.6	355.87	7.6826	5.997e-05 ***
Residuals	276	12784.7	46.32		

	diff	lwr	upr	p adj
Sadness-Joy	4.9377833	1.964234	7.911333	0.00014
Sarcasm-Joy	0.3772287	-2.596321	3.350778	0.98781
Sincerity-Joy	2.2097333	-0.763816	5.183283	0.22168
Sarcasm-Sadness	-4.560555	-7.534104	-1.58701	0.00054
Sincerity-Sadness	-2.72805	-5.701599	0.2455	0.08507
Sincerity-Sarcasm	1.8325047	-1.141045	4.806054	0.38434

Maximum Pitch~Affect

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Affect	3	37657	12552	0.947	.4183
Residuals	276	3658454	13255		

	diff	lwr	upr	p adj
Sadness-Joy	-26.326386	-76.62765	23.97488	0.5301467
Sarcasm-Joy	-26.685657	-76.98693	23.61561	0.5184776
Sincerity-Joy	-27.297643	-77.59891	23.00363	0.4987145
Sarcasm-Sadness	-0.3592714	-50.66054	49.942	0.9999977
Sincerity-Sadness	-0.9712571	-51.27253	49.33001	0.9999553
Sincerity-Sarcasm	-0.6119857	-50.91325	49.68928	0.9999888

UnvoicedFrames~Affect

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Affect	3	16385	5461.8	30.088	2.20E-16***
Residuals	276	50101	181.5		

	diff	lwr	upr	p adj
Sadness-Joy	-1.719686	-7.606147	4.166775	0.8744225
Sarcasm-Joy	15.942871	10.05641	21.82933	0
Sincerity-Joy	-2.951043	-8.837504	2.935418	0.5663729
Sarcasm-Sadness	17.662557	11.776096	23.54902	0
Sincerity-Sadness	-1.231357	-7.117818	4.655104	0.9489737
Sincerity-Sarcasm	-18.893914	-24.78038	-13.00745	0

VoiceBreaks~Affect

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Affect	3	492.1	164.029	3.4538	1.70E-02*
Residuals	276	13107.7	47.492		

	diff	lwr	upr	p adj
Sadness-Joy	-1.8142857	-4.825164	1.196592	0.4047624
Sarcasm-Joy	1.8857143	-1.125164	4.896592	0.3696856
Sincerity-Joy	-0.4714286	-3.482307	2.539449	0.9775545
Sarcasm-Sadness	3.7	0.6891221	6.710878	0.0089635
Sincerity-Sadness	1.3428571	-1.668021	4.353735	0.6570898

Appendix E: Listener Study Posting

Posting:

Perceived Performance Affect: What types of affects (i.e. joy, sadness, etc.) can musicians convey through performance? Are some instruments more capable of expressing certain affects than others? Does a banjo ever sound sad?!? In this study, participants will listen to recordings of various instruments and attempt to identify the performer's intended affect. The study requires approximately 35 minutes. Contact the experimenter via email for more information and to schedule an appointment.

Contact: Joe Plazak plazak.1@buckeyemail.osu.edu

Appendix F: Listener Multinomial Logit Model

Baseline= Joy	Estimate	Std. Error	t-value	Pr(>|t|)	
Sadness	1.56E+00	1.46E+00	1.0651	0.286841	
Sarcasm	-8.61E-01	1.35E+00	-0.64	0.522157	
Sincerity	-1.35E+00	1.48E+00	-0.9107	0.362447	
Other	1.98E+00	1.14E+00	1.7396	0.081933	.
Sadness:Duration	2.33E-01	1.91E-02	12.1958	<2.2e-16	***
Sarcasm:Duration	9.16E-02	1.88E-02	4.8751	1.09E-06	***
Sincerity:Duration	1.40E-01	1.87E-02	7.4847	7.17E-14	***
Other:Duration	1.43E-01	1.63E-02	8.75	<2.2e-16	***
Sadness:PitchSD	-2.82E-02	6.41E-03	-4.3981	1.09E-05	***
Sarcasm:PitchSD	-3.59E-03	5.17E-03	-0.6935	0.487999	
Sincerity:PitchSD	-9.22E-03	5.79E-03	-1.5912	0.111575	
Other:PitchSD	-1.83E-02	4.30E-03	-4.2623	2.02E-05	***
Sadness:MinPitch	5.21E-03	1.63E-03	3.1968	0.00139	**
Sarcasm:MinPitch	3.28E-03	1.46E-03	2.2396	0.025116	*
Sincerity:MinPitch	2.68E-03	1.51E-03	1.7683	0.077016	.
Other:MinPitch	3.82E-03	1.20E-03	3.1839	0.001453	**
Sadness:MaxPitch	-4.05E-03	1.90E-03	-2.127	0.033423	*
Sarcasm:MaxPitch	8.45E-04	1.58E-03	0.5365	0.591634	
Sincerity:MaxPitch	-4.95E-03	1.82E-03	-2.7205	0.006519	**
Other:MaxPitch	-2.08E-03	1.38E-03	-1.5111	0.130764	
Sadness:MeanPeriod	1.16E+02	1.17E+02	0.9967	0.31892	
Sarcasm:MeanPeriod	3.89E+02	1.08E+02	3.6021	0.000316	***
Sincerity:MeanPeriod	-1.99E+02	1.24E+02	-1.5977	0.110116	
Other:MeanPeriod	1.22E+02	9.68E+01	1.2604	0.207537	
Sadness:PeriodSD	1.41E+03	2.48E+02	5.6848	1.31E-08	***
Sarcasm:PeriodSD	-3.21E+02	2.20E+02	-1.4565	0.145248	
Sincerity:PeriodSD	7.42E+02	2.41E+02	3.0808	0.002064	**
Other:PeriodSD	4.38E+02	1.89E+02	2.3122	0.020767	*
Sadness:UnvoicedFrames	-3.56E-03	8.40E-03	-0.4236	0.671838	
Sarcasm:UnvoicedFrames	6.12E-02	6.26E-03	9.7758	<2.2e-16	***
Sincerity:UnvoicedFrames	-2.96E-02	8.93E-03	-3.3122	0.000926	***
Other:UnvoicedFrames	1.98E-02	5.91E-03	3.3461	0.00082	***
Sadness:VoiceBreaks	-2.19E-01	2.02E-02	-10.8291	<2.2e-16	***
Sarcasm:VoiceBreaks	-7.34E-02	1.35E-02	-5.4459	5.16E-08	***
Sincerity:VoiceBreaks	-8.24E-02	1.47E-02	-5.5954	2.20E-08	***
Other:VoiceBreaks	-1.73E-01	1.26E-02	-13.7335	<2.2e-16	***

(Continued below)

(Continued from above)					
Sadness:Jitter(ppq5)	-2.09E+00	5.49E-01	-3.8003	0.000145	***
Sarcasm:Jitter(ppq5)	-2.42E-01	3.20E-01	-0.756	0.449678	
Sincerity:Jitter(ppq5)	-9.96E-01	4.13E-01	-2.4127	0.015833	*
Other:Jitter(ppq5)	-1.21E+00	2.77E-01	-4.3481	1.37E-05	***
Sadness:Shimmer(local)	-3.46E-01	1.15E-01	-2.9997	0.002703	**
Sarcasm:Shimmer(local)	-8.37E-02	6.79E-02	-1.2335	0.217384	
Sincerity:Shimmer(local)	9.10E-03	1.09E-01	0.0838	0.933177	
Other:Shimmer(local)	8.64E-02	6.32E-02	1.3662	0.171879	
Sadness:Shimmer(apq3)	8.82E-01	1.94E-01	4.559	5.14E-06	***
Sarcasm:Shimmer(apq3)	1.87E-01	1.18E-01	1.5902	0.111792	
Sincerity:Shimmer(apq3)	3.69E-01	1.85E-01	1.9966	0.045873	*
Other:Shimmer(apq3)	2.09E-01	1.10E-01	1.9019	0.057182	.
Sadness:Shimmer(apq11)	-5.77E-03	2.28E-03	-2.5327	0.01132	*
Sarcasm:Shimmer(apq11)	-7.47E-03	1.84E-03	-4.0568	4.98E-05	***
Sincerity:Shimmer(apq11)	2.28E-03	2.29E-03	0.997	0.318776	
Other:Shimmer(apq11)	-7.19E-03	1.65E-03	-4.367	1.26E-05	***
Sadness:HTNratio	-3.42E-02	3.38E-02	-1.0118	0.311637	
Sarcasm:HTNratio	-5.91E-02	2.98E-02	-1.984	0.047255	*
Sincerity:HTNratio	9.29E-02	3.32E-02	2.7947	0.005195	**
Other:HTNratio	-1.67E-02	2.53E-02	-0.6609	0.508689	

Appendix G: Listener Model ANOVAs

Duration~Response

	Df	Sum_Sq	Mean_Sq	F_value	Pr(>F)	
Response	3	2892	964.03	17.581	3.02E-11	***
Residuals	1753	96121	54.83			

	diff	lwr	upr	p.adj
Sadness-Joy	3.432815	2.178545	4.687086	0
Sarcasm-Joy	0.996927	-0.2596	2.253451	0.173678
Sincerity-Joy	1.855027	0.564772	3.145281	0.001285
Sarcasm-Sadness	-2.43589	-3.72188	-1.1499	7.2E-06
Sincerity-Sadness	-1.57779	-2.89675	-0.25882	0.011434
Sincerity-Sarcasm	0.8581	-0.46301	2.179208	0.339805

Pitch Standard Deviation~Response

	Df	Sum_sq	Mean_sq	F_Value	Pr(>F)
Response	3	79823	26607.7	28.349	< 2.2e-16 ***
Residuals	1753	1645321	938.6		

	diff	lwr	upr	p.adj
Sadness-Joy	-18.506	-23.6953	-13.3167	0
Sarcasm-Joy	-8.55022	-13.7488	-3.35161	0.000145
Sincerity-Joy	-10.4132	-15.7514	-5.07504	3.5E-06
Sarcasm-Sadness	9.9558	4.63529	15.27631	9.7E-06
Sincerity-Sadness	8.09282	2.635874	13.54977	0.000816
Sincerity-Sarcasm	-1.86298	-7.32879	3.602832	0.817067

Minimum Pitch~Response

	Df	Sum_sq	Mean_sq	F_Value	Pr(>F)
Response	3	70856	23618.5	2.4926	0.05849
Residuals	1753	16610376	9475.4		

	diff	lwr	upr	p.adj
Sadness-Joy	-12.1987	-28.6869	4.289414	0.22729
Sarcasm-Joy	-3.64852	-20.1663	12.86924	0.941575
Sincerity-Joy	5.593023	-11.3682	22.5542	0.831446
Sarcasm-Sadness	8.550211	-8.35488	25.45531	0.562636
Sincerity-Sadness	17.79176	0.453153	35.13036	0.041747
Sincerity-Sarcasm	9.241543	-8.12523	26.60831	0.519431

Maximum Pitch~Response

	Df	Sum_sq	Mean_sq	F_Value	Pr(>F)	
Response	3	1301810	433937	36.347	<2.2e-16	***
Residuals	1753	20928428	11939			

	diff	lwr	upr	p.adj
Sadness-Joy	-74.4499	-92.9575	-55.9423	0
Sarcasm-Joy	-39.7191	-58.26	-21.1783	2E-07
Sincerity-Joy	-28.967	-48.0056	-9.92841	0.000549
Sarcasm-Sadness	34.73077	15.75514	53.7064	1.61E-05
Sincerity-Sadness	45.48291	26.02068	64.94514	0
Sincerity-Sarcasm	10.75214	-8.74171	30.24599	0.487963

Mean Period~Response

	Df	Sum_sq	Mean_sq	F_Value	Pr(>F)	
Response	3	0.000223	7.42E-05	19.865	1.16E-12	***
Residuals	1753	0.006548	3.74E-06			

	diff	lwr	upr	p.adj
Sadness-Joy	8.07E-04	0.000479	0.001134	0
Sarcasm-Joy	6.72E-04	0.000344	0.001	9E-07
Sincerity-Joy	8.11E-05	-0.00026	0.000418	0.925964
Sarcasm-Sadness	-1.34E-04	-0.00047	0.000201	0.73158
Sincerity-Sadness	-7.26E-04	-0.00107	-0.00038	4E-07
Sincerity-Sarcasm	-5.91E-04	-0.00094	-0.00025	6.52E-05

Period Standard Deviation~Response

	Df	Sum_sq	Mean_sq	F_Value	Pr(>F)	
Response	3	5.35E-06	1.78E-06	2.7085	4.38E-02	*
Residuals	1753	0.001155	6.59E-07			

	diff	lwr	upr	p.adj
Sadness-Joy	-5.37E-05	-1.91E-04	8.38E-05	0.747055
Sarcasm-Joy	1.39E-06	-1.36E-04	1.39E-04	0.999994
Sincerity-Joy	-1.38E-04	-2.79E-04	3.76E-06	0.05975
Sarcasm-Sadness	5.51E-05	-8.59E-05	1.96E-04	0.746727
Sincerity-Sadness	-8.40E-05	-2.29E-04	6.06E-05	0.441373
Sincerity-Sarcasm	-1.39E-04	-2.84E-04	5.75E-06	0.065131

Unvoiced Frames~Response

	Df	Sum_sq	Mean_sq	F_Value	Pr(>F)	
Response	3	86332	2.88E+04	141.92	<2.2e-16	***
Residuals	1753	355462	202.8			

	diff	lwr	upr	p.adj
Sadness-Joy	-5.31E+00	-7.72E+00	-2.90E+00	1E-07
Sarcasm-Joy	1.17E+01	9.25E+00	1.41E+01	0
Sincerity-Joy	-6.06E+00	-8.54E+00	-3.58E+00	0
Sarcasm-Sadness	1.70E+01	1.45E+01	1.94E+01	0
Sincerity-Sadness	-7.50E-01	-3.29E+00	1.79E+00	0.872244

Voice Breaks~Response

	Df	Sum_sq	Mean_sq	F_Value	Pr(>F)	
Response	3	7863	2.62E+03	46.377	<2.2e-16	***
Residuals	1753	99065	56.51			

	diff	lwr	upr	p.adj
Sadness-Joy	-4.76E+00	-6.03E+00	-3.49E+00	0
Sarcasm-Joy	9.77E-02	-1.18E+00	1.37E+00	0.997295
Sincerity-Joy	-3.29E+00	-4.60E+00	-1.98E+00	0
Sarcasm-Sadness	4.86E+00	3.55E+00	6.16E+00	0
Sincerity-Sadness	1.47E+00	1.27E-01	2.81E+00	0.025313
Sincerity-Sarcasm	-3.3922347	-4.733424	-2.051045	0

Jitter~Response

	Df	Sum_sq	Mean_sq	F_Value	Pr(>F)	
Response	3	10.687	3.56E+00	24.905	8.86E-16	***
Residuals	1753	250.742	0.143			

	diff	lwr	upr	p.adj
Sadness-Joy	-1.57E-01	-2.21E-01	-9.27E-02	0
Sarcasm-Joy	9.68E-03	-5.45E-02	7.39E-02	0.980183
Sincerity-Joy	-1.45E-01	-2.11E-01	-7.91E-02	1E-07
Sarcasm-Sadness	1.66E-01	1.01E-01	2.32E-01	0
Sincerity-Sadness	1.17E-02	-5.56E-02	7.91E-02	0.970141
Sincerity-Sarcasm	-0.1547161	-0.222191	-0.087241	0

Shimmer (local)~Response

	Df	Sum_sq	Mean_sq	F_Value	Pr(>F)	
Response	3	1485	4.95E+02	25.13	6.44E-16	***
Residuals	1753	34540	19.7			

	diff	lwr	upr	p.adj
Sadness-Joy	-1.14E+00	-1.89E+00	-3.89E-01	0.000576
Sarcasm-Joy	1.06E+00	3.08E-01	1.81E+00	0.001697
Sincerity-Joy	-1.20E+00	-1.97E+00	-4.27E-01	0.000398
Sarcasm-Sadness	2.20E+00	1.43E+00	2.97E+00	0
Sincerity-Sadness	-5.99E-02	-8.51E-01	7.31E-01	0.997374
Sincerity-Sarcasm	-2.2619413	-3.053879	-1.470004	0

Shimmer (apq3)~Response

	Df	Sum_sq	Mean_sq	F_Value	Pr(>F)	
Response	3	188.8	62.935	13.016	2.08E-08	***
Residuals	1753	8475.9	4.835			

	diff	lwr	upr	p.adj
Sadness-Joy	-2.21E-01	-5.94E-01	1.51E-01	0.421737
Sarcasm-Joy	5.08E-01	1.35E-01	8.82E-01	0.002645
Sincerity-Joy	-3.71E-01	-7.54E-01	1.18E-02	0.061435
Sarcasm-Sadness	7.30E-01	3.48E-01	1.11E+00	5.9E-06
Sincerity-Sadness	-0.150247	-0.541915	0.2414209	0.75721
Sincerity-Sarcasm	-0.8797842	-1.272088	-0.48748	1E-07

Shimmer (apq11)~Response

	Df	Sum_sq	Mean_sq	F_Value	Pr(>F)	
Response	3	547.1	182.365	18.857	4.90E-12	***
Residuals	1741	16837.4	9.671			

	diff	lwr	upr	p.adj
Sadness-Joy	-8.39E-01	-1.37E+00	-3.12E-01	0.000262
Sarcasm-Joy	4.79E-01	-5.22E-02	1.01E+00	0.094103
Sincerity-Joy	-8.43E-01	-1.39E+00	-3.01E-01	0.000386
Sarcasm-Sadness	1.32E+00	7.75E-01	1.86E+00	0
Sincerity-Sadness	-0.00445076	-0.558383	0.5494818	0.999997
Sincerity-Sarcasm	-1.32270738	-1.88062	-0.764795	0

Harmonic to Noise Ratio~Reponse

	Df	Sum_sq	Mean_sq	F_Value	Pr(>F)	
Response	3	7159	2386.34	48.911	<2.2e-16	***
Residuals	1753	85528	48.79			

	diff	lwr	upr	p.adj
Sadness-Joy	3.30E+00	2.12E+00	4.49E+00	0
Sarcasm-Joy	-1.23E+00	-2.41E+00	-4.14E-02	0.039249
Sincerity-Joy	3.42E+00	2.21E+00	4.64E+00	0
Sarcasm-Sadness	-4.53E+00	-5.74E+00	-3.32E+00	0
Sincerity-Sadness	0.1194061	-1.124761	1.3635732	0.994721
Sincerity-Sarcasm	4.6496995	3.403511	5.8958879	0