Animacy, Anthropomimesis, and Musical Line

Dissertation

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

By

George John (Yuri) Broze III

Graduate Program in Music Theory

The Ohio State University

2013

Dissertation Committee:

David Huron, Advisor

David Clampitt

Gregory Proctor
© Copyright by
George John (Yuri) Broze III
2013
Abstract

Music is argued to exhibit animacy cues—perceptual indicators that it is alive. In particular, musical lines mimic three characteristics of humans: voice, gait, and language. Empirical approaches are used to investigate the influence of polyphonic voice multiplicity on musical emotion perception, and to test for correspondences between melodic speed and pitch height in western compositions.
To Nora.
Acknowledgments

This work would not be possible but for the efforts of my colleagues in the CSML, whom I value as friends as well as collaborators. Brandon Paul, Kate Guarna, and Erin Allen directly contributed to the research contained herein. Much is due as well to David Clampitt and Gregory Proctor for their continued guidance and musicianly support. I will forever be grateful to Mom, Dad, Jeff, Sean, and Nora, who guided me through uncountable struggles and joys. And finally, special thanks to David Huron, whose passion and clarity of ideas got me involved with music scholarship in the first place.
Vita

2012 .......................... M.S. Statistics, Ohio State University

2010–2012 ........................ Graduate Teaching Associate, Ohio State University School of Music

2006 ............................. B.S. Chemistry, University of North Carolina at Chapel Hill

Publications


Fields of Study

Major Field: Music Theory
# Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract</td>
<td>ii</td>
</tr>
<tr>
<td>Acknowledgments</td>
<td>iv</td>
</tr>
<tr>
<td>Vita</td>
<td>v</td>
</tr>
<tr>
<td>List of Tables</td>
<td>ix</td>
</tr>
<tr>
<td>List of Figures</td>
<td>x</td>
</tr>
<tr>
<td><strong>1. Introduction</strong></td>
<td>1</td>
</tr>
<tr>
<td>1.1 Conspectus</td>
<td>3</td>
</tr>
<tr>
<td><strong>2. Animacy Cues in Sight and Sound</strong></td>
<td>4</td>
</tr>
<tr>
<td>2.1 A Biological Tale</td>
<td>4</td>
</tr>
<tr>
<td>2.1.1 Linguistic and Musical Animacy</td>
<td>7</td>
</tr>
<tr>
<td>2.2 What is Animacy?</td>
<td>9</td>
</tr>
<tr>
<td>2.3 Animacy in the Visual Sense</td>
<td>10</td>
</tr>
<tr>
<td>2.3.1 Perceptual Causality</td>
<td>10</td>
</tr>
<tr>
<td>2.3.2 Apparent Social Emotion in Geometric Shapes</td>
<td>13</td>
</tr>
<tr>
<td>2.3.3 Perceptual Animacy and the Energy Violation Hypothesis</td>
<td>15</td>
</tr>
<tr>
<td>2.4 Animacy in the Auditory Sense</td>
<td>17</td>
</tr>
<tr>
<td>2.5 Animate Musical Lines</td>
<td>20</td>
</tr>
<tr>
<td>2.6 Conclusion</td>
<td>23</td>
</tr>
</tbody>
</table>
# Table of Contents

3. Notes, Beats and Patterns: Three Anthropomimetic Features of Musical Lines 26
   3.1 Percepts of a Social Animal ........................................ 26
      3.1.1 Facial Perceiving-as-Human ................................. 29
   3.2 Three Forms of Musical Anthropomimesis ......................... 31
      3.2.1 Notes and Voice ............................................. 33
      3.2.2 Beats and Gait .............................................. 38
      3.2.3 Patterns and Language ..................................... 42
   3.3 Conclusion ......................................................... 45

4. Voice Multiplicity and Emotion Perception ........................ 47
   4.1 Introduction ...................................................... 47
      4.1.1 Approaches to Musical Texture ............................. 49
   4.2 Experiment 1: Perception of Musical Loneliness ............... 52
      4.2.1 Stimuli ..................................................... 53
      4.2.2 Procedure ................................................ 55
      4.2.3 Results ................................................... 55
      4.2.4 Discussion ............................................... 60
   4.3 Experiment 2: Generalization to Emotion Valence and Sociality 61
      4.3.1 Stimuli and Procedure .................................... 63
      4.3.2 Results and Discussion .................................. 63
   4.4 Experiment 3: Emotion Rating Strength and Voice Denumerability 66
      4.4.1 Procedure and Stimuli ..................................... 68
      4.4.2 Results and Discussion .................................. 69
   4.5 General Discussion ............................................... 75
      4.5.1 Numerosity Cues .......................................... 76
      4.5.2 Future Directions ......................................... 79

5. Musical Line Mediates Musical Pitch-Speed Relationships .... 81
   5.1 Introduction ...................................................... 81
      5.1.1 Measuring Musical Speed .................................. 83
   5.2 Study 1: Notewise Relationship Between Pitch Height and Duration 86
      5.2.1 Methods ................................................... 86
      5.2.2 Results .................................................. 87
      5.2.3 Discussion ............................................... 91
   5.3 Study 2: Partwise Distribution of Musical Activity .......... 92
      5.3.1 Methods ................................................... 93
      5.3.2 Results .................................................. 94
      5.3.3 Discussion ............................................... 97
List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1 Emotions Used in Experiment Two</td>
<td>62</td>
</tr>
<tr>
<td>4.2 Voice Denumeration Confusion Matrices</td>
<td>70</td>
</tr>
<tr>
<td>5.1 Notewise Correlations between Pitch Height and Note Duration</td>
<td>90</td>
</tr>
<tr>
<td>5.2 Note Distributions Between Musical Parts</td>
<td>95</td>
</tr>
<tr>
<td>5.3 Instruments, Ranges, and Observed Speeds</td>
<td>101</td>
</tr>
<tr>
<td>5.4 Ornament Distribution Between Musical Parts</td>
<td>105</td>
</tr>
</tbody>
</table>
### List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Perceptual Causality Animation, after Albert Michotte.</td>
<td>11</td>
</tr>
<tr>
<td>2.2</td>
<td>Still from Heider and Simmel’s (1944) animated display.</td>
<td>14</td>
</tr>
<tr>
<td>2.3</td>
<td>Sample displays from Stewart’s (1982) studies of perceptual animacy.</td>
<td>16</td>
</tr>
<tr>
<td>3.1</td>
<td>Two examples of pareidolic faces</td>
<td>29</td>
</tr>
<tr>
<td>3.2</td>
<td>Face Perception Exploited for Artistic Purposes</td>
<td>32</td>
</tr>
<tr>
<td>3.3</td>
<td>Biological motion point-light display, after Johansson (1973).</td>
<td>39</td>
</tr>
<tr>
<td>4.1</td>
<td>Perceived Loneliness in Beethoven’s Pathétique Sonata, Second Movement, mm. 37–38 Albrecht (2012)</td>
<td>51</td>
</tr>
<tr>
<td>4.2</td>
<td>Subjects of Fugues Sampled</td>
<td>54</td>
</tr>
<tr>
<td>4.3</td>
<td>Loneliness Ratings for Minor Fugue Excerpts</td>
<td>56</td>
</tr>
<tr>
<td>4.4</td>
<td>Median Emotion Ratings</td>
<td>57</td>
</tr>
<tr>
<td>4.5</td>
<td>Strength of Emotion Ratings</td>
<td>59</td>
</tr>
<tr>
<td>4.6</td>
<td>Emotion ratings from Experiment Two: Valence and Sociality</td>
<td>64</td>
</tr>
<tr>
<td>4.7</td>
<td>Voice Denumerations and Error Rates</td>
<td>71</td>
</tr>
<tr>
<td>4.8</td>
<td>Voice Denumeration and Normalized Emotion Ratings</td>
<td>73</td>
</tr>
<tr>
<td>4.9</td>
<td>Note and Onset Density in Stimuli</td>
<td>78</td>
</tr>
<tr>
<td>Section</td>
<td>Page</td>
<td></td>
</tr>
<tr>
<td>------------------------------------------------------------------------</td>
<td>------</td>
<td></td>
</tr>
<tr>
<td>5.1 Stars and Stripes Forever (excerpt)</td>
<td>84</td>
<td></td>
</tr>
<tr>
<td>5.2 Note Distributions for Four Pieces</td>
<td>88</td>
<td></td>
</tr>
<tr>
<td>5.3 Notewise Correlations between Pitch Height and Note Duration</td>
<td>89</td>
<td></td>
</tr>
<tr>
<td>5.4 Relative Note Distributions by Average Part Tessitura</td>
<td>96</td>
<td></td>
</tr>
<tr>
<td>5.5 Instrumental Tessitura and Melodic Speed for 16 Solo Instruments</td>
<td>102</td>
<td></td>
</tr>
</tbody>
</table>
Chapter 1: Introduction

We begin with a sestet penned by Robert Herrick, a seventeenth-century poet whom playwright A. C. Swinburne once called “the greatest song-writer ever born of English race.”

**To Musick. A Song.**

Musick, thou Queen of Heaven, Care-charming spel,
That strik’st a stillness into hell:
Thou that tam’st Tygers, and fierce storms, that rife,
With thy soule-melting Lullabies:
Fall down, down, down, from those thy chiming spheres,
To charme our soulues, as thou enchant’st our eares.

—Robert Herrick, *Hesperides* (1648)

Just what is it that makes music so compelling? Herrick’s Musick is a deity that lives and breathes with a power ferocious in its docility. And it is not merely the poetically-inclined who have used animate descriptions to put a face upon music. Consider the following, written in 1813 by German music critic E. T. A. Hoffman on the subject of Beethoven’s instrumental compositions.

...in this artful structure there alternate in restless flight the most marvelous pictures in which joy and grief, melancholy and ecstasy, come side by side or intermingled to the fore. Strange figures begin a merry dance, now floating off into a point of light, now splitting apart, flashing and sparkling, evading

---

1(Herrick, 1846). Swinburne’s quotation can be found in the preface to Lawrence & Bullen’s edition, 1898.
and pursuing one another in various combinations, and at the center of the
spirit realm thus disclosed the intoxicated soul gives ear to the unfamiliar lan-
guage and understands the most mysterious premonitions that have stirred
it. (Hoffmann, 1998, p. 1196)

Music is commonly said to push, to pull, to strive, to yearn, to reach, to move, to
desire, to ease, to relent, to rise, to fall, to abate, to explode. Music can be said to roar
to life, or to die out with a final whimper. Perhaps more amazingly, music can be happy,
sad, distraught, ebullient, proud, or morose—all emotions that can be felt only by living,
 sentient beings. Moreover, it seems that music’s ineffable ability to evoke all of these
ideas and feelings is intertwined with its ability to command attention and compel us to
move and feel along with it. But when we do, just what is the ‘it’ that we respond to,
and why should this be so?

Many perspectives on these questions have been offered over the course of western
history, many of which have appealed to deities, numerological relations, or a divine
resonance with the soul. The present naturalistic approach will hopefully be at least as
profitable (as well as more currently fashionable). Specifically, it will be argued that those
elements of music we are most likely to hear and respond to are musical lines—melodies
or other horizontally-connected segments of music with perceptual relevance. As real
neurally-encoded ‘objects’ that are abstracted from surface textures, musical lines have a
concrete reality that makes them reasonable objects of study.

By removing musical lines from their artistic context and pinning them to the dissec-
tion table we do not wish to trivialize their role in musical artworks or expression. Quite
to the contrary, we propose that musical lines gain their power from their quantifiable
resemblance to exactly the sort of sound that humans have always found particularly
enchanting: those created by other humans.

Like other forms of culture and communication, music can be seen as a ritualization
of preexistent biological tendencies and behaviors. But although the musical art in this
sense could be seen as derivative of more commonplace human behaviors, this does not
render music in any way deficient. Rather, music might instead be the artful distillation of our perceptual proclivities, appearing at times more human than even humans typically do. Perhaps this is precisely what one would expect from a Queen of Heaven.

1.1 Conspectus

The present study will argue that music is compelling precisely because it is likely to be understood as being alive, and having originated from a human source. This is not simply because we learn to associate music with its human creators from experience, or because we apply artistic license when describing it. Rather, it is because music exhibits anthropomimetic animacy cues that compel our sensory apparatus to interpret it as coming from a living human source. In essence, music is a human creation that resembles human voice, movement, and intellect, even when performed instrumentally or mechanically.

A biological account of perceptual animacy will be provided in Chapter 2, along with an argument that the categorization of sensations into ‘animate’ and ‘inanimate’ types is a perceptually primitive, bottom-up process. In Chapter 3, it will be suggested that humans possess perceptual systems that are specifically attuned to human sounds. Three distinct musical features will be seen to imitate recognizable human characteristics: First, musical lines resemble human voices in their pitch and timbre. Second, the recurring beats characteristic of meter resemble human locomotion. Finally, because music exhibits syntactic patterns of rhythm and intonation, this suggests the presence of an intelligent human producing a structured linguistic signal.

If musical lines are processed as though they come from individual humans (even when they do not), then one might expect emotional responses to vary according to the implied social situation. Chapter 4 is a test of perceptual voice numerosity’s effect on perceived musical emotion. Finally, Chapter 5 provides an empirical test of the notion that higher music is faster, with results suggesting that the answer depends on musical organization into lines.
Chapter 2: Animacy Cues in Sight and Sound

Abstract

A biological motivation is given for the existence of *animacy cues*—those features of a percept that indicate the presence of a living organism. Experiments and theories regarding visual animacy cues are reviewed. Possible animacy cues in the auditory modality are suggested, and musical implications are discussed.

2.1 A Biological Tale

Any biologically-grounded account of music depends on an understanding of perception in general. Following the suggestion of psychologist James J. Gibson (1966), the senses are perceptual systems that inform an organism of its surroundings—but rather than creating an unbiased representation, these systems ought to provide all and only the information useful for the organism’s continued functioning and ultimate survival. Percepts are colored according to an animal’s needs, filtered and classified at many stages of processing to ensure efficiency and maximize utility.

All perceptual systems are at their core either energy detectors or chemical detectors. The eyes detect electromagnetic waves through photoreception, and the ears detect sound waves through mechanoreception. Touch is sensitive to kinetic energy of both large objects and of vibrating particles (*i.e.*, temperature), while olfaction and taste both detect particular chemical structures. In all of these cases, sensations can be said to reflect energy—organized or disorganized, great or small.
Although both vision and audition detect energy, they differ as to the type of energy sources they are tailored to respond to. In the case of vision, most of what our evolutionary ancestors would have seen were visual echoes, reflected light energy originating from something other than the perceived object. By contrast, most sound energy detected by the ear is emitted directly from its source. Even in the case of echolocation and reverberation, in which sounds convey information about the size, shape, and surface texture of a space, the initial, unreflected noise is almost invariably louder than its echo.

The auditory system is therefore prone to perceive sound events coming from sound sources far more often than, say, the visual system encounters ‘sight events’ from ‘sight sources.’ It has been said that the ears point the eyes: audition detects noises in all directions, and if a sound is threatening or interesting, the organism will likely turn its eyes toward it. Because the auditory system can perceive objects and events that are not directly before us, audition acts as an excellent alarm system, alerting an animal to unseen threats. Modern alarms from clocks to sirens make use of this characteristic of the auditory modality as well, and as Huron (1992) has suggested, composers also can exploit the human ‘orienting response’ to sudden, loud, or unexpected sound.

As an animal’s auditory system monitors its surroundings, a process of source identification and categorization affixes meaning to the incoming percepts. Sound signals are classified in order to be efficiently processed, and the derived information about the auditory objects becomes consciously accessible. Of all possible noises, one might imagine that the sound categories most relevant to survival ought to be the most readily detected. In other words, particularly important sound types should be more perceptually salient, presenting themselves with distinctive phenomenological qualia to ensure they are noticed and acted upon.

Sound categorization based on a source’s perceived animacy would be a particularly ability. An animal would surely stand to gain by recognizing those sounds caused by other living things, because they tend to offer more affordances for immediate action than sounds from inanimate sources. A prey animal that can successfully distinguish a
stalking predator from wind-rustled leaves is more likely to survive than one that cannot, and the inverse would be true for a predator. Because other animals can present considerable danger and opportunity, one might even expect animacy detectors to be slightly over-responsive, as would be desirable whenever false negatives are more costly than false positives.¹

Imagine walking into an unfamiliar room in an office building. A seemingly reasonable first task might be to orient oneself by surveying the room’s layout and contents. By noting the room’s floorplan, furniture, and decor, it is possible to learn what the room is for and formulate a plan for action. Although it may be attractive to imagine that we dispassionately scan for information, in reality we are unlikely to methodically construct an accurate representation of the space. Instead, we would probably seek out the most immediately useful information, such as whether or not we are alone in the room. If there is a person already in the room, then the first order of business quickly becomes whether and how to greet them. Because the script we follow depends so much on whether another person is present, detecting animate agents must be performed at least as early as any other information-gathering task.

Based on the immense importance of social interactions in human life, one might even expect that the identity and quantity of animate agents present at any given time would be a piece of information always kept readily in mind. The feeling of being alone has a completely different quale from knowing somebody else is in the room, which is likewise distinct from the feeling of being completely surrounded by people at Grand Central Terminal. That we constantly monitor the number of people present is made plain when one considers what the consequences would be if a prankster were to sneak up undetected as you make your morning coffee, hide for several minutes, and suddenly leap from behind the kitchen counter and shout ‘boo!’. It is possible that manipulating

¹That is, a prey animal should be more likely to mistake wind-rustled leaves for a predator than a predator for wind-rustled leaves.
a listener’s ‘animate agent’ count might have musical implications as well, as will be investigated in Chapter 4.

Characteristics of an object that suggest it is alive could be called animacy cues. It is clear enough that animacy cues must exist; otherwise it would be impossible to distinguish the living from the nonliving altogether. A more interesting question is the degree to which animacy is a perceptual phenomenon as opposed to a higher level distinction mediated by cognition. If animacy can be rapidly inferred based on simple perceptual features, this could explain the ease with which we can monitor the number of living things nearby, how readily we notice when living things appear, and why animals are so interesting to observe.

While a complete account of perceptual animacy cues remains to be constructed, there is already evidence that perceptual animacy can arise from very primitive perceptual cues (Stewart, 1982). The capacity to distinguish animate from inanimate things also appears very early in life (Premack, 1990; Gelman et al., 1995), with even six-month old infants able to form animacy-dependent expectations (Leslie & Keeble, 1987). Adults can abstract animacy cues from visual stimuli as minimal as a single moving dot (Tremoulet & Feldman, 2000). From a musical standpoint, it is worthwhile to consider whether animacy might be inferrable from a single moving tone as well.

2.1.1 Linguistic and Musical Animacy

In the field of linguistics, ‘animacy’ refers to a grammatical category, important because some human languages distinguish nouns referring to animate things from those referring to inanimate ones. In fact, English is one of these languages; the third-person pronouns ‘she’ and ‘he’ refer specifically to animate things, whereas ‘it’ is reserved for inanimate things. The animacy of a noun can even determine the intended meaning of some words: “it is depressed” means that the valley is lower than the surrounding area; “he is depressed” means that your friend is sad.
While animacy could be seen as ‘merely’ a grammatical category, this category obviously reflects very real differences between the inanimate and the animate. Unlike doors, books, and buildings, animate things are sentient: they can experience emotions and possess intentionality. Doors probably cannot have intentional states, but people, cats, and musk oxen probably can, and are therefore referred to using personal pronouns. Possibly even crickets and ants have intentional states as well, but even if they do not, it doesn’t preclude Anglophones from occasionally referring to a cricket as ‘he.’ One could infer that this must be because a cricket exhibits animacy cues, resembling as it does a living creature in its morphology, movement, and the sounds it emits.

A similar situation arises in the case of music. If music can be said to push, strive, or yearn, this would seem to imply that it is a living thing with intentional states. If music can be morose or ebullient, this would imply that music can also experience emotional states. Of course, because music is not a sentient being, it is incapable of either intention or emotion. These statements surely do not actually mean that notes on a page or vibrations in the air are alive, but merely that they can be compared to living things. In a similar way, a pitcher who has retired nine batters in a row is said to be ‘on fire,’ but there is clearly no combustion involved. Or likewise, a well-delivered speech might ‘get through’ to a particularly ‘dense’ person, even if no drilling was necessary. Are these lifelike descriptions of music are simply metaphorical, or might something deeper be at work?

Linguist George Lakoff and philosopher Mark Johnson (1980) have argued that many linguistic metaphors are more than mere artifice, and in fact reflect deeper conceptual metaphors used by the cognitive faculty in general. By their account, the mind itself employs conceptual metaphors when it understands one idea or experience in terms of an idea or bodily state from some other domain. These ideas are by no means uncontroversial. See, for instance, Fauconnier (1999).

---

2Philosophical treatments of this problem can be found in Kivy (1980) and Davies (1994).
3These ideas are by no means uncontroversial. See, for instance, Fauconnier (1999).
cross-domain mapping from theme to figure reflects not only a flair for literary description, but also a deeply-rooted way of conceptualizing music.

While literary metaphors are only limited by a writer’s imagination, one would expect cognitive metaphors to be shaped by very real physiological, cognitive, or perceptual constraints (Lakoff, 1987). If this is true, then perhaps animate descriptions of music are so prevalent precisely because music specifically invites this interpretation. In short, music is constructed in such a way as to exhibit perceptible animacy cues. Although in the end, music itself could not possibly be alive, perhaps we yet have a tendency to perceive it at some intermediate stage as though it were.

2.2 What is Animacy?

At the heart of the notion of animacy lies an ontological and metaphysical question of enormous breadth and depth. To avoid delving too deeply into philosophical discussions of mind, it will suffice to list several characteristics of some, but not all, animate things. In roughly ascending order of complexity, these include:

- Animate agents are *self-propelling* or locomotive, and can move under their own power.
- Animate agents possess *intentionality*, and exhibit goal-directed behavior.
- Animate agents are *communicative*, and employ signalling systems.
- Animate agents are *sentient*, and can experience subjective feelings, percepts, and emotions.
- Animate agents are *intelligent*, capable of rational thought.
- Animate agents can be *self-conscious*, having metacognitive states about their own consciousness and that of others.
Not all animate agents exhibit every one of these characteristics, nor must these characteristics necessarily indicate the immediate presence of an animate agent. However, all of these characteristics appear together in the most important animate organism in the human environment, namely other humans. Therefore, it seems reasonable that observing one or two of these characteristics might lead one to assume the others. For example, cat owners will often speak to their pets as though they are intelligent or even self-conscious, even though cats are probably not able to engage in metacognition about their minds or others.

If any of these features of animate agents are immediately perceptible in some modality, they could potentially provide a primitive perceptual basis for animacy classifications. Animacy cues would therefore be those visual, auditory, tactile, or haptic sensations which tend to correlate with actual animacy. As will be seen, there is psychological evidence that several of these animacy characteristics leave visual traces that can be detected by the perceptual system, including notions of locomotion (Michotte’s caterpillar, 1963), sentience (Heider & Simmel, 1944), and intentionality (Bassili, 1976; Stewart, 1982; Dasser et al., 1989).

2.3 Animacy in the Visual Sense

In lieu of an exhaustive review of studies relevant to perceptual animacy, three important threads of research will be highlighted: Albert Michotte’s experiments on perceptual causality early in the twentieth century, Fritz Heider and Marianne Simmel’s (1944) work concerning the perception of social behavior and intentionality, and Judith Ann Stewart’s (1982) dissertation research specifically targeting perceptual animacy.

2.3.1 Perceptual Causality

The psychological study of animacy has its roots in pioneering work on causality perception carried out by Belgian psychologist and philosopher Albert Michotte. His
Belgian psychologist Albert Michotte used animations such as this sequence in his pioneering experiments on perceptual causality. Michotte argued that causality is directly inferred from primitive features of a visual scene, such as the distance between the two displayed objects and their temporal coincidence. For a summary of these experiments, see Michotte (1963).

Michotte’s famous experiments employed projected displays of moving discs and rectangles to instantiate certain rudimentary types of motion and interaction. A representative example of a ‘launching effect,’ in which one moving disc is observed to transmit its energy to a second, is depicted in Figure 2.1. By systematically varying the speeds of the objects, the distances between them, and the directions of their motion, the subjective
perception of causality was linked to simple quantifiable features of the visual displays. Based on his results, Michotte theorized that the visual perception of causality depends on three factors: the spatial proximity of the participating objects, the temporal coincidence of cause and effect, and a correspondence of ‘form and matter’ between the effect and the cause.

Although his experimental techniques were oriented toward abstract iconic shapes and simple causal percepts, Michotte was well aware that certain simple types of motion could induce the impression of animacy. Says Michotte,

Among the examples of causality mentioned, . . . several, it will have been noticed, were concerned with human activity. This is no accident. Situations in which mechanical causality can actually be perceived in the world of inanimate nature are in fact very rare (except in the case of machines made by man) . . .

Besides their relations with external objects, the movements made by a man or an animal possess a special character which, in the normal way, differentiates them clearly from the movements of inanimate objects, and makes it easy to recognise the presence of animal life, an important fact from the biological point of view. These movements are not only spontaneous, like those of inert objects which begin to move without external cause, but they also have the appearance of being activities of which the object itself seems to be the source. Thus when observers describe what they see in simple language, they say without exception that they see an object which is ‘moving of its own accord’, or ‘is going under its own steam’ (Michotte, 1963, p. 184).

Thus inspired, Michotte conducted several foundational studies of visual animacy cues, using geometric shapes that specifically suggested locomotion. One display in particular featured a rectangle that mimicked a caterpillar inching its way across the screen. Subjects who are shown motion of this type voluntarily would exclaim ‘Caterpillar!’ or ‘Worm!’ in instant recognition, even though the rectangle itself had no other outward resemblance to a living creature.

One final insight of Michotte’s is worth noting. After having observed innumerable reactions to these visual displays, it seemed that notions of causal motion could also be linked to emotions. Many colloquial phrases, such as ‘It shook me to pieces’, ‘It got me
down’, or ‘It knocked me flat’ provide a crucial connection. “All these terms not only have a kinematic significance; in addition they clearly imply mechanical action” (Michotte, 1963, p. 283). To Michotte, these linguistic idioms implied cognitive processes about emotional states that make direct and consistent use of kinetic and spatial imagery. In making this observation, Michotte preceded Lakoff and Johnson’s (1980) contribution to cognitive linguistics by some thirty-five years.

### 2.3.2 Apparent Social Emotion in Geometric Shapes

Michotte was primarily concerned with the immediate phenomenological experience of causality, but research in this tradition continued to investigate more complex perceptual impressions. Around the same time as the publication of *La perception de la causalité*, Austrian psychologist Fritz Heider and his student Marianne Simmel were investigating the perception of social emotion at Smith College in Massachusetts. In a landmark study, Heider & Simmel (1944) demonstrated that simple geometric shapes can spontaneously be perceived as animate, social agents with goals and feeling states. A film depicting moving geometric shapes was shown to several participants. In this film, depicted in Figure 2.2, a triangle and two circles move within and around a crude ‘enclosure,’ pantomiming a dramatic narrative. Viewers spontaneously described the scene using not only basic emotion words like ‘happy’ or ‘afraid,’ but also reported observing more complex socially-tinged affective states such as frustration, rage, hesitation, love, and vengeance.

What is more, participants in the study volunteered this sort of description whether they had been told that the shapes represented people or not—the study participants had spontaneously inferred from some properties of the display that they represented living things, and proceeded to describe the events in emotional terms. Several years later, it was found that participants would use the same emotional descriptions even when they were specifically admonished to describe the scene in geometric terms only (Hashimoto,
Observers of this film, which displayed moving and interacting shapes, described the scene using intentional and emotional language usually reserved for sentient agents. The motion of the shapes, it seems, was simply too compellingly lifelike for sterile geometric descriptions to suffice.

Because geometric shapes such as Heider and Simmel’s (Figure 2.2) lack outward morphological resemblance to living creatures, it seems that their motion and interactions must have cued viewers to perceive them as being animate. At least two characteristics of animate beings might have been directly perceived. First, the shapes might have appeared to locomote under their own power. Second, the shapes might have appeared to intentionally pursue certain goals, based on their trajectories. Although the study participants might have inferred sentience, intelligence, or even meta-cognition, it seems unlikely that these could have been directly perceived from shapes moving on a screen. Presumably, these higher-level characteristics of animate creatures might have been inferred from the shapes’ apparent locomotion or intention.

The excitement surrounding the idea that simple moving shapes can suggest emotional states might seem somewhat perplexing to musicians and music lovers, whose art
has long used ‘simple’ vibrations in the air to express emotions or to construct dramatic scenes. Nonetheless, Heider and Simmel’s results raise the question of why these sorts of descriptions are so naturally attached to musical sounds in the first place. One wonders whether a similar experiment using minimal musical stimuli would produce analogous results, and perhaps suggest what low-level feature might be triggering animate descriptions of music.

2.3.3 Perceptual Animacy and the Energy Violation Hypothesis

A comprehensive theoretically-driven study of perceptual animacy was provided by Judith Ann Stewart in her 1982 dissertation. Stewart observed that one characteristic feature of animals is the ability to initiate changes in movement using stored chemical energy. Because animate objects have internal sources of energy, Stewart conjectured that objects which appear to respond to unseen forces should appear to be animate. This hypothesis could be called an ‘energy violation’ hypothesis of animacy cues (Scholl & Tremoulet, 2000).

Using several displays depicting the motion of a single dot (as in Figure 2.3), Stewart investigated whether apparently self-initiated changes of direction and velocity would increase the likelihood the dot would appear alive. For example, if a dot appears to bounce off a flat barrier, its change in acceleration could be easily attributed to the barrier’s presence. However, an identical display with the barrier removed ought to leave the viewer wondering what might have been the source of the dot’s change in direction. For the most part, Stewart’s predictions turned out to be correct, with direction-changing acceleration strongly predicting the perception of animacy. Subsequent experiments using minimal visual simuli have replicated and extended these findings (Gelman et al., 1995; Tremoulet & Feldman, 2000), making the energy violation hypothesis of animacy cues quite compelling.
Figure 2.3: Sample displays from Stewart’s (1982) studies of perceptual animacy.

Stewart (1982) tested many properties of moving shapes for their effect on perceived animacy, including initiation of motion, impact, velocity and acceleration, angularity of deflection, goals objects, and directness of path.

Because the visual displays used by Michotte, Heider and Simmel, and Stewart all depend on the *motion* of onscreen shapes, they implicitly provide evidence only for *locomotive* animacy cues. However, one could easily argue that an energy violation hypothesis would apply equally well to other characteristics of animacy.\(^4\) Communication by its nature relies on channeling energy into visual, acoustic, chemical, or electric signals. Moreover, biological minds that have sentience, intelligence, self-consciousness or metacognitive capacities all exhibit a form of anti-entropic organization that requires the

\(^4\)Realizing this, Stewart (1982) also conducted explicit tests for perceptual intentionality by manipulating whether ‘goal objects’ were present in some displays. However, this inference would also depend on observing directed locomotion.
continual investment of energy—brains require a constant supply of chemical fuel to operate, and in humans the brain consumes more energy than any other organ.

Because animacy connotes so many different possible capacities, perceptual animacy might not be as simple as a binary distinction. Instead, it might represent an entire class of different perceptual cues operating simultaneously. It has even been proposed that movements such as those appearing in Heider and Simmel’s display might be directly perceived as chasing, evading, courting, being courted, fighting, or playing (Blythe et al., 1999). Each of these scenarios in turn could represent many of the aforementioned characteristics of animacy, including locomotion, communication, intelligence, or even presence of self-conscious emotions like love. As research in this area continues, an important question to resolve will be the degree to which these percepts rely on higher-level cognitive processes, and how much they are attributable to lower-level perceptual features.

2.4 Animacy in the Auditory Sense

If the perception of an apparent energy source is crucial for the visual perception of animacy, then one might think that the auditory system, itself a non-trivial energy detector, would be particularly attuned to the animacy of sound sources. However, few have yet investigated the existence of auditory animacy cues (Scholl & Tremoulet, 2000). As an initial step, several possible cues will here be suggested.

Auditory and visual perception are completely different modalities with different psychophysics. While visual experiments have dealt mainly with locomotive animacy cues, auditory perception might be better attuned to other features of animate things altogether. The characteristics of animacy already enumerated—locomotion, intentionality, communication, sentience, intelligence, and self-consciousness—will be used to generate several hypotheses. Future work might later reveal whether these cues produce mandatory, immediate animacy percepts.
**Locomotion.** Animals produce a wide variety of sounds when they move, including footsteps, tapping, scratching, breaking, knocking, brushing, and rubbing. Several organisms produce distinctly recognizable periodic sounds by their movements, such as a horse that trots or gallops. Other animals produce irregular noises such as rustles and crackles as they move through foliage. In both of these cases, sustained motion would produce sounds that do not become softer over time, which could provide perceptual evidence that there is a driving source of energy. In the case of music, there are many obvious correspondences with physical motion, especially with regard to speed, investigated in Chapter 5.

**Intentionality.** Intentionality could be described as the possession of goal directed mental states. In order for intentionality to be perceived, however, it must first be translated into some observable action by a creature exercising *volition*. One form of volitional action could be the goal-directed motion employed by (Stewart, 1982). However, a simpler volitional exercise might be in the form of making a simple binary choice, akin to flipping a power switch on or off. That is, sounds that suddenly start, stop, or resume might indicate that the sound source chose to move in such a way, and is therefore animate.

Imagine the stalking behavior of a house cat hunting a field mouse. In order not to alarm its prey, the cat will not simply dash toward its target as soon as the mouse can be seen, because the probability of a successful strike from a distance is low. Instead, it would benefit the cat to strike from nearby. In order to reach such a position, the cat will divide its approach into several stealthy movements, interspersed with periods of stillness. This discontinuous behavior, which could also result in discontinuous sounds, would reflect volitional modulation of an internal energy source.

**Communication.** Animate things often produce communicative signals, both to coordinate behavior with allies and to issue warnings to enemies. One common signalling modality is vocal communication, which constitutes a suite of readily identifiable sounds produced by passing air over a vibrating membrane. Because only animate things can
communicate vocally, sounds resembling vocal noises in pitch and timbre might well serve as animacy cues, as will be explored in the next chapter.

Another type of acoustic communication makes use of percussive sounds, such as a rattlesnake’s warning rattle or a gorilla’s chest thumping. If these sounds start and stop suddenly, they might suggest volition, and if they maintain their volume (or even get louder), they would suggest the continual investment of energy. If these percussive sounds are organized into distinctive patterns that could not be easily generated by naturally-occurring pendular motion, this could provide additional evidence that a willful agent might have produced them. Hence, many types of percussive sound—not only those representing locomotion—could potentially serve as animacy cues.

Note that both vocal and percussive communication sounds exhibit periodicity resulting from the investment of energy, corresponding to a sort of non-randomness or decreased entropy. Thus, the simple presence of periodicity in general could suggest a concentrated source of energy available for work, and therefore an animate presence.

**Sentience.** A sentient being can experience sensations, emotions, and affective states. In the auditory modality, this could be perceived in terms of the pitch, loudness, and contour of vocalizations. It has often been suggested that music can appear to express emotions by merit of resemblance to vocal prosody (Kivy, 1980; Juslin & Laukka, 2003; Ilie & Thompson, 2006). In order to account for systematic regularities in both musical and vocal expressivity, an ‘ethological model’ of musical emotion has recently been proposed (Huron, 2012).

One might expect that in order to interpret prosodic cues, it would first be necessary to identify a sound as vocal in origin. On the other hand, perhaps the vocal origin and affective content of a sound could be simultaneously recognized. As German musicologist Viktor Zuckerkandl (1956, p. 60) has noted, “tone sensations. . . are always colored by feeling. We do not hear flute tones or trombone tones; we hear charming flute tones, solemn or threatening trombone tones. Low tones sound serious, high tones gay, and so on.”
**Intelligence.** Another promising way in which listeners might infer animacy is by detecting evidence of intelligence. One way this could occur is if meaningful syntactic patterns could be identified in an auditory percept. The presence of a coded message would strongly imply the existence of an animate agent using a language. Indeed, ‘intelligent’ is an adjective often applied in musical description, and several pieces of music could be described as ‘cerebral’ by merit of the complex relational patterns they exhibit.

While human language is an obvious place where syntactic sounds can be found in nature, it should be noted that syntactic sounds can be found in other places as well. For example, whales and songbirds have been observed to make grammatical sounds that are learned through experience. Whether or not these sounds reflect ‘intelligence’ as usually understood is unclear, but it is indisputable that whales and songbirds are at least alive.

**Self-consciousness.** Directly perceiving self-consciousness through sound would seem rather unlikely. However, it remains possible that sound sources could be imputed to have metacognitive states if they were to exhibit sufficient animacy cues of other sorts. Indeed, the shapes in Heider and Simmel’s display might have seemed to some participants to exhibit complex social emotions such as love, shame, or jealousy, which would depend on metacognitive capacities. Relationships between musical elements and social emotions are directly investigated in Chapter 4.

While auditory animacy cues remain little studied from a psychological perspective, there would seem to be ample anecdotal evidence that they might exist. A more thorough investigation of auditory animacy might reveal much about how we interact with all aspects of our auditory environment, musical and nonmusical alike.

### 2.5 Animate Musical Lines

When we listen to music, there are innumerable things to which we can pay attention, including melodies, basslines, rhythms, harmony, texture, and timbre. In part,
the multidimensional complexity of music is what makes it possible for one composition to provide several different unique listening experiences. One could focus on a piece’s bassline, attempt to pick out motivic transformations, follow harmonic functions, or simply let timbres guide the ear. There are as many possible ways to listen to a Webern string quartet as there are segmentational analyses of the surface, a very large number indeed. If, as philosopher Charles Nussbaum has suggested, “the acousmatic realm of the composer is an animistic realm” (Nussbaum, 2007, emphasis in original), we might then inquire which aspect of a musical surface is animated. Is it that the entire surface sounds lifelike, or is it that individual entities or agents within the music sound alive?

Many music theorists are familiar with Heinrich Schenker’s penchant for organicist analogy and animism (Solie, 1980; Pastille, 1984; Hubbs, 1991). At least two musical elements were described in animate terms by Schenker over the course of his career: individual tones, and entire motifs. In *Harmony*, Schenker compares motives directly to biological creatures, complete with reproductive urges (Schenker, 1909/1954).

Man repeats himself in man; tree in tree. In other words, any creature repeats itself in its own kind, and only in its own kind; and by this repetition the concept “man” or the concept “tree” is formed. Thus a series of tones becomes an individual in the world of music only by repeating itself in its own kind; and, as in nature in general, so music manifests a procreative urge, which initiates this process of repetition... We should get accustomed to seeing tones as creatures. We should learn to assume in them biological urges as they characterize living beings (p. 6).

Later Schenkerian thought centered around the *Tonraum*, a world of pitches in which tones themselves were animate agents. To Schenker, these tones seemed to have intentionality—*Tonwille*—and the most sensitive composers could help the tones realize their urges, both to generate octaves, fifths, and thirds and to find their ultimate consummation in the descent of the *Urlinie*.

---

3In its later stages, Schenker’s theory of musical organicism eschewed the notion of motive in favor of a synchronic, multi-layered outgrowth from a fundamental structure (see Schenker, 1935/1979, §50, §254, §308).
In order to understand why Schenker might have chosen to animate tones and lines instead of rhythm, harmony, meter, or loudness, it is useful to consider how a musical surface is perceived. A fundamental process of music perception is akin to a score segmentation problem in the analysis of atonal music, in which vertical and horizontal units are discretized based on various structural, functional, or acoustic dimensions. The perceptual segmentation process is an important component of a more general capacity of the auditory system, known as auditory scene analysis (Bregman, 1990). Through this process, raw acoustic vibrations are transformed into a coherent ‘auditory image’ of one’s surroundings, with sounds apparently emanating from the same source grouped into perceptual units, or auditory objects.

The process of scene analysis allows us to perceptually link together discrete musical sounds into larger streams of sounds: melodies (Dowling, 1968). Using acoustic aspects of sound such as pitch proximity and temporal spacing, the auditory system automatically identifies which sequential sounds were likely to come from the same source, in a process known as stream segregation. We have some volitional control over our percepts as well: we can choose to follow the melody line, the bass part, or inner voices of a chorale in a way similar to how one can tune into one speaker at a time at a cocktail party.

When perceiving animate agents in the auditory environment, it seems clear enough that the sound source itself is the thing that is animate, not the entire scene. As the preeminent instances of auditory streams in music, melodies provide the best targets for the possibility of animacy. This is consistent with common experience as well; it would seem to be nearly impossible to listen to music without hearing any lines at all. Lines are of utmost importance to Schenker as well: “Every linear progression shows the eternal shape of life—birth to death. The progression begins, lives its own existence in the passing tones, ceases when it has reached its goal—all as organic as life itself” (Solie, 1980).
Melodies have often been described as being alive, beautiful, and expressive. The opening quotation by critic E. T. A. Hoffman directly compares Beethoven’s motives—their linear melodies—to living creatures. Philosopher Georg W. F. Hegel claimed that melody, in contrast to meter, rhythm, and harmony, is the “free-sounding soul in music” (Schnädelbach, 2010, p. 84). Hegel is of course not alone; for example, the German aesthete Eduard Hanslick (1986) claimed in his famous monograph on musical aesthetics, “Unconsumed and inexhaustible, melody holds sway over all, as the basic form of musical beauty” (p. 28).

As music theorist Fred Maus (1992) has pointed out, Hanslick was prone to describe music as though one or more interacting people are in the room, despite his avowed formalist stance.6 This would seem to closely align with the drama that unfolded in Heider and Simmel’s (1944) display: Participants in their experiment did not describe the entire scene, but rather the shapes within the scene. In the case of music, it seems possible that listeners hear each musical line as an individual animate agent.

### 2.6 Conclusion

We have proposed that animate descriptions of music are not merely literary anthropomorphisms, but instead reflect a fundamental mode of perception that distinguishes the animate from the inanimate. Underlying this process are perceptual animacy cues that reflect characteristics such as intentionality, sentience, or intelligence. For something to be animate, it must have some internal energy source that fuels its mind and movement, whether this leads to locomotion, periodic sounds, or the production of organized patterns. Because musical lines are more readily perceivable as sound sources than other aspects of a musical surface, melodies and motives are more likely to be perceived as animate agents than, for instance, harmonies, textures, or formal structure.

---

6Perceptual systems conceived in the Gibsonian sense are unlikely to operate in a formalistic way, despite the protests of their possessors. For his part, Hanslick clarified that such animate descriptions were to be understood metaphorically only.
Although the present account adopts a modern scientific perspective, we have in fact resurrected a very old idea. Long ago, the Aristotelian Problemata discussed music’s moral or ethical character, ethos (῏ ηθος) (Problems XIX, Problems Connected with Harmonia). The twenty-seventh problem reads as follows:

Why does what is heard, alone of perceptible objects, possess ethical character [῏ ηθος]? Indeed, even if a melody is without words, it nonetheless possesses ethical character; but neither color nor smell nor flavor possess it. Is it because (what is heard) alone possesses movement, though not that which the sound moves in us? For such movement exists in the other (perceptible objects) too—indeed, even color moves our sight; but we perceive a movement that follows this sort of sound. This movement has a likeness (to ethical character) both in the rhythms and in the arrangement of high and low notes, not in their mixture. But consonance [συμφωνία] has no ethical character. This character, however, does not exist in the other perceptible objects. But these movements are connected with action, and actions are signs of ethical character. (Aristotle, 2011, Problems XIX.27)

In the words of Classics scholar Andrew Barker (1984), this problem’s author observed that “Moral character is represented by movement from note to note, not by notes played simultaneously” (p. 197). Perhaps the auditory perception of animacy cues attached to musical lines could explain why the Greeks heard ethical melodies, why neither the Harmonicists nor Rameau were never completely able to unseat melody as the heart of the musically beautiful, and why Schenker elevated melody to the pinnacle of his theory of musical organicism.

The presence of animacy cues in music would suggest that it ought to demand a particular sort of attention that inanimate, static percepts would not. However, this does not explain why music should be as uniquely captivating as it often seems to be; mere animacy must not be enough. After all, we tend not to become completely entranced by each and every one of the many animate agents we encounter each day. It is also easy to find decidedly nonmusical things that seem to exhibit animacy cues: even if a fire alarm is perceived animately because of its voicelike timbre, it would still seem unlikely

\[ ^7 \text{Parenthetical expressions are the translator’s; square brackets enclose the original Greek.} \]
to enchant many ears. Evidently, simple animacy cues are not quite enough to make music.

One possible missing component is whether or not a listener is deliberately paying aesthetic attention. As John Cage has vividly demonstrated, music—or at least aesthetic auditory experience—can be found in a broad range of sounds. It is possible that all that is missing to make an animate sound musical is the appropriate aesthetic mindset, as avant-garde composers have often reasoned. Consider as an example Arseny Avraamov’s Symphony of Sirens, which debuted in 1922 in Baku, Azerbaijan. The piece combined a bewildering array of modern industrial and military sounds, including heavy artillery and factory sirens (Wendela, 2012). Undoubtedly this performance captured the attention of those who were nearby to witness it. However, it is unclear whether it would invite willful attention, repeat performances, or humming along in the same way a traditional orchestral symphony or hit single can.

While Avraamov’s Symphony of Sirens is a human product that exploits certain biological tendencies for artistic purposes, it seems not to be a ritualization of them. One must conclude that not all apparently animate sounds closely resemble the sort of thing that usually results from human culture. Not only must there be a sense of animacy, it would seem that there must be a certain sense of humaneness: although music is a Queen of Heaven, a siren is a shrieking demon, alive but inhuman.
Chapter 3: Notes, Beats and Patterns: Three Anthropomimetic Features of Musical Lines

Abstract

Of the innumerable ways in which music can be constructed and appreciated, three features stand out as defining characteristics. First, music tends to utilize pitched tones. Second, music often is organized around regular beats. Third, music is constructed using recognizable patterns. It is argued from psychological and neurological evidence that these three musical features are explicitly anthropomimetic, causing musical sounds to be especially compelling and capable of expressing human emotion. Importantly, all three of these features are present in musical lines.

If music exhibits auditory animacy cues, it joins a broad category of percepts including barking dogs, revving engines, and trotting horses. But among a diverse array of apparently animate sounds encountered daily, music seems to be particularly expressive and emotionally compelling. It will be seen that music’s peculiar ability to captivate us could be attributed to its anthropomimesis. In other words, not only is music created by humans, but it also specifically resembles basic human features.

3.1 Percepts of a Social Animal

There are few creatures on the planet as social as human beings, we who rely on relationships with other people to fulfill innumerable short-term and long-term needs. Other people are our friends and loved ones, those with whom we build communities and on whom we rely for support. Other humans are also authority figures, teachers,
teammates, or rivals. Each of us relies on a network of social alliances to help us navigate the densely populated world we inhabit. The human impulse toward interpersonal interaction is so strong that depriving somebody of human contact is among the most torturous of punishments, and chronic depression is often nearly indistinguishable from chronic loneliness (Cacioppo & Patrick, 2008).

This overwhelming human fascination with other humans is abundantly evident in the art and media we produce. Flipping through television or radio stations, one finds human likeness in sound or image practically inescapable. Even when nonhuman landscapes or structures appear onscreen, they are almost invariably accompanied by the sound of human narration. One would no more expect to find a motionless piece of furniture televised than grass growing or paint drying. Literature proves to be no different; fictional works concern humans and their affairs, and even if actual human animals never appear, anthropomorphized animals or objects are ready surrogates. In many ways, the world we inhabit is constructed by and of human beings and their relationships.

Because other people are so important to our lives, it would make sense for our perceptual and cognitive systems to be specifically attuned to other humans, their appearance, and the sounds they make. It appears that many social species have anatomical and neural structures devoted to the task of identifying conspecifics (that is, members of the same species) and gathering relevant information about them (e.g., Romanski & Averbeck, 2009). If music is able to activate human-detecting neural pathways in our brains, this could provide a partial explanation for why music can be so bewitching.

While mimetic theories of music have existed from antiquity, only a subset have explicitly proposed that music mimics human beings. Instead, several other subjects have commonly been suggested. The Pythagoreans felt that music instantiates musica universalis, and compels the soul to resonate sympathetically with divine harmonies. For Charles Batteaux, eighteenth-century French philosopher and aesthete, music serves as a ‘portrait of the passions,’ imitating Nature’s beauty as all beaux arts do (Neubauer, 1986, p. 61). Batteaux’s fellow French académicien Jean-Baptiste Dubos believed that
music imitates ‘all those sounds with which nature herself expresses her sentiments’ (ibid. p. 63). A century later, German philosopher Arthur Schopenhauer claimed that music represents ‘the will itself with nothing intervening.’ (Dahlhaus, 1982, p. 42). The field of possible referents seems limitless.

Nonetheless, the view that instrumental music mimics the human voice has also been prevalent since antiquity. Several, including Plato and Galileo, have argued the inferiority of instrumental music to vocal music from precisely this premise. Still others have pointed out resemblances to human features other than the voice. For example, Enlightenment lawyer and music theorist Roger North wrote:

> musick is a true pantomime, or resemblance of humanity in all its states, motions, passions, and affections. And in every musicall attempt reasonably designed, humane nature is the subject, and so penetrant that thoughts, such as mankind occasionally have, and even speech it self, share in the resemblance so that an hearer shall put himself into the like condition, as if the state represented were his owne. (Musicall Recollections III: f.48v, cited in North, 1728/1990, p. 46)

North’s stance on musical anthropomimesis can be viewed as a direct precursor of the present perspective.

An argument that a particular type of sensation is at some level perceived as being human could be supported by several types of evidence. First, a distinct mode of perception specialized for other humans ought to be automatic and mandatory, possibly producing a strong and distinct phenomenology. Second, presuming it is more costly to mistake a human for something non-human than the reverse, one would expect to find several instances where non-human percepts are perceived as human. Third, these faculties ought to appear very early in life, because perceiving-as-human would enable infants to know which sensations to learn from. Fourth, there ought to be identifiable neural pathways that support such percepts, whether they are completely innate or acquired from experience. Finally, if the ability to perceive something as human is a special, modularized capacity, then one might expect to find specific deficits in this ability.
3.1.1 Facial Perceiving-as-Human

An excellent example of an everyday ‘perceiving-as-human’ experience can be found when one gazes at another person’s face. Faces are of primary importance in our social world; we use them not only to recognize the identities of others, but also their intentional and affective states. Moreover, it appears that facial perception might be automatic and mandatory, producing a particular phenomenological quale (Palermo & Rhodes, 2007). The strength of this committed face perception apparatus is demonstrated by the human penchant to see faces in many places other than on other people, as the two images in Figure 3.1 make vivid.

The image on the left depicts a rock formation identified by the Viking 1 spacecraft when photographing Martian Cydonia in 1976. This mountain in particular bears a striking likeness to a human visage. On the right is a more commonly-encountered ‘face’: an electrical outlet. Although either image has only a coarse, minimal resemblance to an actual human face, both images are nonetheless strongly suggestive, creating an irresistible compulsion to see eyes and a mouth. The mind’s hair-trigger for face identification also leads us to commonly imagine faces in clouds, the moon, and the front of
some automobiles. This tendency to incorrectly find meaningful content in inappropriate places has been called *pareidolia*, literally, ‘wrong image.’

It is clear that despite the spuriousness of such pareidolic percepts, things ‘perceived-as-human’ can nonetheless be quite compelling. A plausible explanation for this phenomenon would be that certain low-level features of face-like images—such as eye-like and mouth-like shapes in an upright orientation—automatically activate our face-processing neural machinery, a view supported by recent neurological evidence (Hadjikhani et al., 2009). Face processing has been linked to specific cortical tissue in the fusiform gyrus (Sergent et al., 1992; Kanwisher et al., 1997) as well as a subcortical pathway involving the superior colliculus, pulvinar and amygdala (Johnson, 2005). Rapid, spontaneous activation of these pathways in response to face-like stimuli could explain why facial perception seems to be so automatic, rapid, and irresistible.\(^1\) Even neonates with an average age of 36 hours have the capacity to recognize faces and mimic the facial expressions they see (Field et al., 1982).

The existence of a special facial perception faculty is also suggested by pathological deficits in face discrimination, recognition, and perception, together referred to as *prosopagnosia*, or ‘face-blindness.’ Although otherwise unaffected, people with acute or congenital prosopagnosia are often unable to recognize the faces of friends, family, or even themselves. Because this deficiency is specific to faces, it indicates there might be some modular capacity underlying facial perception that could be individually disabled.

Notably for the present study of musical anthropomimesis, biological tendencies of the human perceptual system are also ripe for artistic exploitation, as is shown by several works of *trompe l’œil* art. C. Allan Gilbert’s famous illustration *All is Vanity*, shown in Figure 3.2, deliberately manipulates the human propensity to see faces in minimal stimuli. The anthropomimetic characteristics of pareidolic faces can also trigger perception

---

\(^1\)Some parts of the ‘fusiform face area’ have been observed to activate in response to other objects as well (Grill-Spector et al., 2006), suggesting the region might represent more generalized ‘expertise.’ This would not conflict with the idea that facial perception recruits specialized pathways.
of emotional or affective states. As Davies (2010) has pointed out, when a visual percept has features that resemble a human emotional expression, we might perceive it as expressing an emotional state. It seems likely that musical art could produce its compelling expressivity in similar ways.

3.2 Three Forms of Musical Anthropomimesis

Music resembles at least three types of human-produced sounds, each of which might activate specialized human-perceiving neural pathways. First, music makes use of pitched tones that resemble (or are in fact produced by) the human voice, corresponding to a communicative animacy cue. Second, musical beats evoke imagery of human ambulation, corresponding to a locomotive animacy cue. Finally, pattern-based aspects of musical construction indicate an intelligent agent might have created the signal. In all, music’s trifold similarity to voice, gait, and language could potentially activate three separate perceiving-as-human systems. If so, this would explain its ability to capture our attention, as well as its capacity to express affective, intentional, and emotional states so strongly.

The view that music mimics humans ought to be clarified in light of the many other things music can imitate. Musicians and composers have often availed themselves of music’s ability to imitate any number of natural sounds, settings, or topics, from Josquin’s “El Grillo” to Edvard Grieg’s “Morgenstemning” or Leroy Anderson’s “Sleigh Ride.” This fact of artistic freedom is completely compatible with the idea that low-level processing of perceived sounds might yet follow preordained paths. Certain perceptual machinery may always treat music at some level as though it came from a human source, whether the composer used musical word-painting or not.

A final caveat should be mentioned regarding the scientific research that will be cited. The field of neuroscience is young and moves quickly; current neuroscientific studies are often limited to identifying possible biological underpinnings of long-held assumptions, ancient theories, or more recent psychological results. While the precise neurological
Figure 3.2: Face Perception Exploited for Artistic Purposes

American illustrator C. Allan Gilbert’s *All is Vanity* (1892). The visual pun in this image takes advantage of the human tendency to identify faces based on minimal perceptual cues, representing a ritualized artistic manipulation of biological tendencies.
evidence cited here might later be superceded, the salient point is the plausibility of such mechanisms, if not their detailed structure.

3.2.1 Notes and Voice

The idea that music shares close kinship with the human voice is anything but obscure: the voice has retained an important place in music throughout western history. Plainchant and Medieval polyphony are thought to have been performed completely without instrumental accompaniment, a practice that gives rise to the musical meaning of *a cappella* (Leech-Wilkinson, 2002). Writers from the *Enchiridion* author (c. 850) to Leonard Bernstein (1976) have invoked linguistic comparisons with music, with countless authors suggesting that language and music share a common origin (e.g., Rousseau, 1782). To better understand relationships between music and the voice, it will be necessary to consider how musical and vocal sounds stand out from other ecological noises.

From a biological perspective, the sound of a human voice can be understood as just one instance of vocal communication sounds in general. A great many species signal one another using pitched tones against a backdrop of unpredictably random acoustic noise. Signalling apparatus such as vocal folds or syrinxes are able to produce sounds with energy concentrated at just a few frequencies, making pitched tones audible at longer distances than the same amount of sound energy would be if spread out over the sonic spectrum. This basic principle of vocal communication could be seen as what Dennett (1995) might call a Good Trick: an excellent evolutionary solution to a problem that would be expected to arise again and again.\(^2\)

To humans, voices are indeed quite special sounds. A human voice, whether speaking, laughing, wailing, or humming, potentially carries as much information as does a human face (Belin et al., 2004). We recognize our family members and best friends merely by the sound of their voices, and can also tell if they are excited, bored, sarcastic,

\(^2\)Sure enough, in the context of radio broadcast, humans have reinvented the same conceptual solution.
or frightened. We can quickly identify the sex and approximate age of people simply by hearing them on the phone, and we can even make a guess as to what part of the world they come from. Because of its distinctive character and informational richness, a human voice, much like a face, might be expected to evoke a particular mode of perceiving-as-human.

As previously suggested, certain sounds—such as factory whistles or alarms—might be perceived as being voicelike if they carry pitch or undulate like a voice could. This similarity is precisely why sirens are said to wail, engines to roar, and wheels to squeal. If metaphors such as these are in fact Lakovian cognitive metaphors rather than mere linguistic ones, they might reflect underlying perceptual predispositions. All sounds that are perceptually similar enough to pitched vocal sounds could be processed to some extent as though they were vocal sounds. The subset of these sounds that specifically resembles the timbre and pitch range of the human voice could move beyond mere perceptual animacy and instead extend to perceptually-relevant anthropomimesis.

Just as our hypersensitivity to face-like percepts can lead us to see faces where they do not exist, hypersensitive voice identification processes can result in hearing pareidolic voices. The auditory equivalent of a face on Mars or in an electrical socket might be something like hearing pained groans in the radiator system, or perhaps hearing an alto saxophone as a weeping widow. It is easy to imagine how blowing wind could be heard as a voice as well, whether at a seance or a campfire. Auditory pareidolias have even fueled occult movements: some twentieth century spiritualists have claimed to hear voices in radio static, tape noise, or other mechanical sources, believing them to be paranormal voices from the deceased or divine (Banks, 2001).

There appear to be particular regions of the primate and human brain specifically attuned to conspecific vocalizations (Belin et al., 2000; Fecteau et al., 2004; Petkov et al., 2008; Romanski & Averbeck, 2009), as well as areas devoted to recognizing vocalizations in general (Lewis et al., 2005). Similar to prosopagnosia, cases of phonagnosia—problems with voice recognition and discrimination—occur in both acute and congenital forms.
Adding to the evidence that voice perception is a specialized faculty, it appears that infants selectively attend to voices at a very young age (Vouloumanos & Werker, 2007).

**Hearing Voices in Musical Lines**

For facial perception, it seems that two rudimentary eyes and a mouth, oriented properly, are sufficient to trigger a specific face perception mode. What might the analogous perceptual features be that indicate something should be perceived as a voice?

Several psychophysical studies have investigated the sound features that listeners use to classify sounds, such as harmonicity, spectral shape, frequency, and amplitude envelope (see, e.g., Kidd & Watson, 1987; Christensen & Humes, 1996). A recent study by Gygi et al. (2007) asked four participants to make 10,000 pairwise similarity ratings for a collection of short sounds. A multi-dimensional scaling technique was then used to determine which three composite factors could best explain the similarity of the sounds. The first, most important dimension corresponded to the sound's harmonicity, with animal sounds, speech, laughter, and musical instruments clustered together on one end of the dimension, and inharmonic noises such as a typewriter, drums, footsteps, or hammering on the other.

Because vocalizations are distinguished from other environmental sounds by the presence of harmonic overtones—concentrated energy at particular frequencies—the harmonicity of sounds might be a good indicator that they are vocalizations. Assuming that conspecific vocalizations are particularly special sounds, it would make sense that the human hair-trigger for voice perception might be wired to harmonicity detectors.

Neuroimaging studies have provided additional evidence that musical instruments and the voice are cortically processed in similar ways by merit of their high degree of harmonicity. It has been proposed that vocal sounds are processed along a cortical pathway in which acoustic features such as harmonicity are used to route the signal toward increasingly specialized processing locations (Lewis et al., 2009; Talkington et al., 2012).
At lower levels of processing, musical instruments and voices would presumably be processed identically. Consistent with this view, at least one phonagnost has exhibited impaired ability to recognize musical instruments as well as voices (Hailstone et al., 2010).

In a representative neuroimaging study, Leaver & Rauschecker (2010) employed functional magnetic resonance imaging techniques to measure cortical activation when hearing different sounds. Participants heard recordings of human speech sounds (isolated phonemes), musical instruments, songbirds, and other animals. Of these four, both human voices and musical instruments evoked responses strongly in the same region of the anteroventral auditory cortex, consistent with the view that their processing overlaps. Interestingly, Leaver & Rauschecker’s musical instrument recordings exhibited a slightly higher harmonics-to-noise ratio than speech and a much higher pitch strength. If the hypothesis that harmonicity is an important identifying perceptual marker for speech is correct, then perhaps musical instruments could be viewed as a distillation of this salient feature of the voice.

In addition to harmonicity, there are other similarities between human voices and instrumental music that distinguish them from animal sounds. Both utilize the same approximate pitch range and clef-based notation, the standard form of which possibly arose from vocal ranges. Moreover, both speech and musical melodies exhibit similar interval sizes, with smaller intervals generally favored and intervals a fifth or larger being relatively rare (Patel, 2008, p. 221; Huron, 2006, p. 74; Vos & Troost, 1989). Even though several instruments can transcend the pitch and speed-based constraints of the voice, it seems anecdotally that the most memorable melodies are those that can be sung. It is to this idea that we now turn.

**Motor Theories of Speech and Music**

When considering the evolution of vocal communication, it is clear that any sound meant to be heard would be completely unhelpful if it could not be perceived. This
could be viewed as a basic problem of any communication system, which must rely on the signaller and receiver using the same ‘code.’ The faculties of vocal production and vocal perception ought therefore to evolve concurrently and congruently, including the musculoskeletal and neural machinery for both.

While the neural substrates for voice production and perception could theoretically coevolve as two separate systems, it might be more efficient and parsimonious for the two to overlap. In other words, the perceptual process might be expected to rely in part on the premotor and motor neurons responsible for the production of vocal sounds, engaging them in an ‘off-line’ mode during perception. This ‘simulationist’ account is the essence of the ‘motor theory’ of speech perception proposed by Haskins Laboratory researchers in the 1950s (see Liberman et al., 1967; Liberman & Mattingly, 1985). Because speech sounds are so complex and vary so much from speaker to speaker, Liberman and his colleagues proposed that it is not phonemes per se that are perceived, but rather anatomical configuration of the vocal apparatus. This would account for how listeners are able to perfectly understand speech from a broad range of speakers, even though the constituent sounds themselves might vary dramatically.

The motor theory of speech perception has been bolstered by experimental evidence collected using direct neuronal measurements. Studies with nonhuman animals have revealed that the perception and production of vocal sounds can activate identical neurons. For example, several species of birds exhibit activation of motor nerves when perceiving birdsong (e.g., Williams & Nottebohm, 1985). Evidence for a broader simulationist theory of behavior perception has also been observed in nonhuman primates. Certain neurons in the monkey premotor cortex have been identified which respond both to actions performed and perceived, across both visual and auditory modalities (Di Pellegrino et al., 1992; Gallese et al., 1996; Kohler et al., 2002). The existence of these so-called ‘mirror neurons’ suggests that that simulationist hypotheses of perception might extend quite broadly (Galantucci et al., 2006).
It has been suggested that a motor theory of perception might have considerable explanatory power for music as well (Nussbaum, 2007; Zatorre et al., 2007). For those who would suggest that music is able to ‘resonate’ with us, or that it can induce affective states through an empathic sense, mirror neurons might provide a reasonable anatomical substrate. If this is the case, then when we hear musical lines, we hear voice-like stimuli that not only draw our attention, but activate the same neural mechanisms that we use to produce musical sounds ourselves.

3.2.2 Beats and Gait

In addition to vocal mimicry, music exhibits a strong resemblance to the human rhythms of walking. A great deal of music worldwide is arranged according to an underlying pulse, and moving to music appears to be a cross-cultural universal (Brown, 2003). Not only do compositions tend to be written at tempi close to a typical walking tempo of 120 beats per minute (Murray et al., 1964; Fraisse, 1982; Moelants, 2002), this tempo also appears to be the easiest to synchronize one’s paces with (Styns et al., 2007), and is approximately the optimal range for musical beat perception (Fraisse, 1982). The many parallels between biological motion and musical meter have led van Noorden & Moelants (1999) to propose a ‘resonance theory of tempo perception’ that incorporates both driving energy and dampening factors. Experimental evidence has even linked the characteristic tempo changes of final ritardandi with the velocity of runners coming to a stop (Friberg & Sundberg, 1999). It is worth considering whether rhythmically metered music is naturally perceived not only in terms of auditory constraints at the extremes, but in terms of physiological optima in the center.

Much like the face and voice, a person’s gait provides a great deal of information about them, including their identity, affective state, and intentions. The sound of walking is easily distinguished from sprinting or limping, and the sound of speeding up is quite different from the sound of slowing down. A toddler has a considerably different gait from his grandmother, and the two would almost certainly be aurally distinguishable.
Swedish psychologist Gunnar Johansson used animations such as this sequence to study the perception of biological motion. This figure adapted from Kim et al. (2005).

Even a person’s sex can be determined solely by the sound of their footsteps (Li et al., 1991). Because of the information these sounds potentially carry, one might expect gait-like percepts to follow specialized processing pathways that extract relevant information about the walker’s behavior, mood, or identity.

In the visual domain, there is abundant evidence that human gait is readily perceived with only minimal cues. Moving displays of a few strategically-placed dots are sufficient to suggest motion characteristic of human gait (Johansson, 1973). Figure 3.3 depicts a point-light display of the type used by Swedish psychologist Gunnar Johansson, who demonstrated that biological motion could be perceived with no information about the form of the walker’s body or limbs themselves. Not only is this impression uncanny and irresistible when observed, the capacity to perceive biological motion from moving dots appears even in two-day old newborns, who attend more to upright biological motion displays than their inverted counterparts (Simion et al., 2007). Analogously to prosopagnosia and phonagnosia, pathological deficits in biological motion perception also appears in some people, such as schizophrenics (Kim et al., 2005) and autists (Blake et al., 2003; Parron et al., 2008). In all, it would seem that the visual perception of locomoting animals and humans could rely on specialized systems, whether fully innate or acquired through experience.

3Interestingly, both autism and schizophrenia are also associated with deficits in the perception of vocal prosody, or ‘speech melody.’
The auditory perception of biological motion is somewhat less studied than visual perception, but scholarly literature regarding musical beat perception does exhibit several parallels to that of faces, voices, and biological motion. The perception of regular beats seems to quickly recruit perceptual pathways that predict their continued occurrence; disruption of regular rhythms can elicit characteristic preconsious brain wave responses (see Näätänen & Alho, 1995, for a review). Additionally, infant studies have suggested that the ability to perceive beat and meter appears at a very early age (Phillips-Silver & Trainor, 2005, 2007). Specific deficiencies in the auditory perception of beat and meter have also been identified. At least one person with congenital amusia has been diagnosed with musical ‘beat deafness,’ without exhibiting other musical deficits (Phillips-Silver et al., 2011), suggesting a discrete contribution of beat-processing pathways to musical enjoyment.

Recalling the motor theory of vocal perception, visual perception of biological motion and auditory perception of beat both appear to make use of premotor cortical regions. Biological motion in point-light displays reliably activates parts of the premotor cortex, consistent with a simulationist theory of ambulation perception (Saygin et al., 2004; Saygin, 2007). Theorizing along these lines, psychologist Neil P. M. Todd (1999) proposed an explicit role for motor pathways in the perception of musical rhythm, and more recent studies have revealed that the perception of musical beat does in fact activate the premotor cortex as well (Grahn & Brett, 2007). If musical beat is processed by the same neural architecture that underlies ambulation, then this could provide some explanation for our tendency to spontaneously entrain to music, synchronizing our movements with the beat by tapping, nodding, rocking, or dancing. In both the visual and auditory senses, it seems that locomotion can result in a simulationist mode of perceiving-as-human.

Evidence from medical science has also suggested surprising links between beat perception and ambulation. A striato-thalamo-cortical network in the brain has been proposed to mediate beat perception, linking the premotor cortex with the basal ganglia,
which are thought to mediate purposeful movement and timing tasks (Zatorre et al., 2007; Grahn & Rowe, 2009; McAuley et al., 2012). When the basal ganglia are compromised in patients with Parkinson’s disease, deficits arise in controlling and pacing muscle movements, resulting in these patients’ characteristic Parkinsonian gait. In addition to problems with movement and walking, patients with Parkinson’s disease have problems perceiving musical beats and performing timing tasks (Artieda et al., 1992; Elsinger et al., 2003; Grahn & Brett, 2009). In fact, it seems that deficiencies in timekeeping might lie at the root of walking difficulties in Parkinson’s patients: providing somebody with Parkinson’s with an external beat to synchronize with can cause dramatic improvements in his gait (McIntosh et al., 1997).

Musical relationships with motion need not stop at the beat produced by walking, and many theorists have related musical rhythm to kinematics more broadly (see Repp, 1993; Todd, 1992, 1995). When MacDougall & Moore (2005) used accelerometers to record the daily motion of several study participants, several oscillation frequencies in whole-number ratios were detectable in various body parts, such as the hip, head, or ankle. There is also some evidence that individual differences in body morphology might contribute to preferred musical tempi (Todd et al., 2007; Utley, 2009; Grahn & McAuley, 2009). Although relationships between beat processing and walking have not been entirely worked out, it would seem that close ties between beat and body motion could provide a low-level biological basis for the infectious quality of musical beat.

Finally, it is worth considering that without the ability to move one’s body, there would be very little reason for an advanced timing faculty to evolve in the first place. Organisms without articulated musculature (such as plants) seem not to have any advanced sense of time, instead relying on various environmental cues to regulate their growth patterns. It was previously suggested that a motor theory of vocalization would be evolutionarily parsimonious, because both the productive and perceptual capacity could share the same neural substrate. The same argument could be made for locomotion and metric timekeeping.
3.2.3 Patterns and Language

In addition to having voice-like and gait-like characteristics, musical lines are language-like in that they exhibit syntactically arranged patterns of notes and rhythms. Any sight or sound that appears to have repeated subunits might suggest that some energy source put them in order. Whether or not these units actually convey information, complex patterns and embedded regularities could indicate that an intelligent creature lies behind the percept. Therefore, sounds that seem to contain embedded information, such as spoken language, morse code, or radar blips, might be expected to encourage more careful listening than comparatively less organized sounds, such as chuckles, grunts, or undulating sirens. Music, with its tendency to exhibit hierarchical organization and repeating subunits, certainly gives the impression of being syntactically arranged, and it very often is.

Would it be reasonable to postulate a specific ‘mode of perception’ for pattern recognition? It would surely seem that humans are adept at recognizing patterns, from points of light in the night sky to tesselations of brickwork. Chess grandmasters can identify complex board positions they have encountered before, and even make intuitive judgments regarding ones that they have not. There is even evidence that people can implicitly learn organizational patterns behind temporal sequences of visual and auditory stimuli, even without being consciously aware of what they are learning (Reber, 1967). For instance, both infants and adults appear able to implicitly learn novel grammars of pitched tone sequences (Saffran et al., 1999). But by far the most striking instance of human pattern perception is language. Extracting meaningful information from patterns of printed and spoken words is a monumental feat, surpassed only by a child’s ability to acquire this capacity without explicit instruction.

On the surface, the processing of linguistic patterns resembles other types of perceiving-as-human, in that specialized neural machinery seems to underlie the process. Without delving too deeply into debates over the domain-specificity of the linguistic capacity,
there can be no doubt that at least two language processing centers, Broca’s and Wernicke’s areas, exist on the human cortex. The ability to recognize linguistic patterns also appears very early in life, with infants even preferring the rhythmic cadence of their mother’s language over that of other tongues. Well-described linguistic deficits, or aphasias, have also been linked to lesions in Broca’s or Wernicke’s areas, suggesting some modularity of the ability to understand and produce language.

Similarly to the modes of perception already discussed, it appears that there are several examples where humans are prone to hyperactive pattern recognition (see Foster & Kokko, 2009). This mistaken discovery of patterns in random or meaningless data has been called *apophenia,* sometimes manifesting as conspiracy theories or superstitions, and other times as the simple perceptual recognition of spurious visual or auditory patterns. For example, a fast-spreading rumor that British rock musician Paul McCartney had died grew rapidly in 1969 after ‘hidden reversed messages’ on a Beatles recording were heard (Reeve, 1994). In the coming decade, fundamentalist Christians soon began to complain about Satanic messages that were audible in reversed popular records, such as Led Zeppelin’s 1971 rock anthem “Stairway to Heaven” or Styx’s 1981 release “Snowblind” (Smith, 2011). In each of these examples, a voice-like sound was certainly present, but patterns of speech that were absent (i.e., English words) were mistakenly identified. These incorrectly heard backward messages could be described as instances of auditory apophenia, and ones that involved anthropomimetic hallucinations at that.

Because musical sounds tend to exhibit deliberate syntactic organization, it seems possible that the built-in pattern detectors normally applied to language or recognizing other structured events might also be engaged while listening to music. In fact, there is considerable evidence that music and speech processing share overlapping cognitive resources, which lends credence to a resource-sharing hypothesis of language and music (Patel, 2003). Both musical and linguistic signals can evoke similar event-related brain potentials (Besson & Faïta, 1995; Patel et al., 1998), and the processing of musical syntax appears to make use of linguistic areas in the brain (Patel, 2008, p. 276). To the extent
that languages are perceived as coming from human sources, it would seem that syntactic music would be too.

Could the processing of spoken syntax and patterns of musical tones really be mediated by the same neural pathways? After all, most western languages utilize words composed of vowels and consonants as their syntactic elements, not musical notes. However, this apparent separation between language and tone is not a cultural universal—in fact, the majority of the world’s languages are tone languages, with some exhibiting up to five discrete level tones (Patel, 2008; Fromkin, V. (Ed.), 1978). In tone languages, linguistic syntax explicitly depends on the sequences of tones being used, highlighting the connection between musical and lexical syntax. Instruments such as drums, whistles, or stringed instruments can even serve as ‘speech surrogates’ by speakers of tonal languages. Games can be played based on guessing the linguistic message communicated on a guitar or drum, and important messages can even be whistled across long distances (Patel, 2008, p. 48). The existence of such speech surrogates provides direct evidence that some musical instrumental sounds can be readily ‘perceived as human.’

In addition to syntactic orderings of pitch, it is possible that some rhythmic regularities in music are processed by the same brain regions responsible for linguistic rhythm. Iverson et al. (2008) reports that Japanese and American listeners tend to group identical rhythmic stimuli differently, in ways reflective of their native languages’ characteristics. There is also neurological evidence that similar regions of the brain are involved in processing both musical and linguistic phrase boundaries (Steinhauer et al., 1999; Knösche et al., 2005). Interestingly, tapping musical polyrhythms seems to activate linguistic areas of the brain (Vuust et al., 2011), suggesting that some forms of musical rhythmic complexity is closely related to language processing.

Musical lines therefore could be said not only to mimic human noises by having the harmonicity and contour of vocal sounds (as laughs, cries, grunts, wails, or screams do), but also because they use tonal and rhythmic units syntactically, resulting in a distinct mode of musical perception and enjoyment (see, e.g., Meyer, 1956; Cooke, 1959; Lerdahl
& Jackendoff, 1983; Narmour, 1990; Huron, 2006). Because musical lines make use of patterns—even if they do not map directly onto any known language, they still might suggest an intelligent human being produced them.

3.3 Conclusion

This chapter has argued that music is compelling in part because of the anthropomimetic character of notes, beats, and patterns, which in turn reflect three aspects of human-created sounds: voice, gait, and language. These modes of perceiving-as-human could underlie music’s capacity not only to be perceived as emotional or intentional, but also its ability to capture our attention so strongly. A melodic line resembles a voice by merit of being pitched, can resemble gait by being metric, and can contain information by having motivic patterns of rhythm and pitch.

Ethnomusicologists and anthropologists have long observed that music plays an integral role in human cultures, with singing, dancing, and group participation being crucial aspects. To western musicians and nonmusicians alike, it’s enjoyable to sing along or move along with music—consistent with the idea that music perceived-as-human tends to be processed using the same premotor pathways that would be used to create the sounds.

That not all western art music is meant for singing or dancing appears to be somewhat unusual among musical cultures. It would be interesting to know the degree to which our perceptual capacity constrains western art music to remain within the realm of sounds that could be produced by the body itself. If motor theories of musical processing are correct, then even a person with no prior biases might still be unlikely to enjoy music incompatible with their physiological and psychological dispositions, even if the music falls within the extremes of their perceptual limits.

If musical lines are perceived in terms of their anthropomimetic character, this leads naturally to some predictions about how music might express certain types of emotion. For example, polyphonic instrumental music with many musical parts might imply a
social situation with many participants, whereas instrumental music with only a single musical line would imply only a limited number of participants. The next chapter will describe three psychological experiments regarding the perception of multiple polyphonic voices and perceived musical emotion.
Chapter 4: Voice Multiplicity and Emotion Perception

Abstract

Three experimental studies investigated how the number of musical voices in a passage—the music’s voice multiplicity—influences the perception of musical emotions. Brief (5s) extracts from polyphonic keyboard works by Johann Sebastian Bach were chosen to represent conditions of one, two, three, or four concurrent musical voices, and listeners rated each for their perceived emotional content. In general, denser polyphonic textures were associated with more positive emotional valence. While this effect generalized to a broad range of positive and negative emotions, it appeared not to depend on whether an emotion connoted social situations. In a polyphonic voice denumeration task using the same stimuli, response patterns corresponded closely with emotion ratings, suggesting a single musical feature or percept might play a role in both. Aspects of voice numerosity—the subjective impression of the number of voices present in music—are discussed. Biological origins for the observed effect are proposed.

4.1 Introduction

It has long been speculated that music is expressive of emotion in ways similar to the speaking voice (Rousseau, 1782; Spencer, 1857; von Helmholtz, 1863; Kivy, 1980). More recent empirical evidence has indicated that emotion cues in music and speech overlap considerably (Juslin & Laukka, 2003; Ilie & Thompson, 2006; Thompson et al., 2012), and many musical features are reliably associated with specific emotions (see Gabrielsson

& Lindström, 2010; Gabrielsson & Juslin, 2003). Nonetheless, explanations by simple analogy or homology to speech cannot easily account for all musical situations. A great deal of music does not come from a single voice, and instead reflects the coordinated effort of many musicians performing simultaneously (see, e.g., Merriam, 1964; Blacking, 1973; Nettl, 2005). Indeed, unaccompanied melodies seem to represent only a fraction of the music typically encountered, while most music heard today features multiple simultaneous musical parts, lines, or accompaniments. If musical emotions are perceived in part because musical and linguistic emotion cues coincide, then how do listeners perceive emotion in music that resembles more than one voice?

Relatively few psychological studies of musical texture and emotion appear to have been conducted, and those that have been published have produced conflicting results. Kastner & Crowder (1990) asked children aged three to twelve to respond to both monophonic and block-chord accompanied melodies by pointing to drawings of faces depicting positive and negative emotions. Overall, unaccompanied stimuli appeared to be perceived more positively than accompanied stimuli. A similar correspondence was identified by Webster & Weir (2005) using four-part harmonizations. In their study, undergraduate participants consistently rated nonharmonized melodies as happier than harmonized melodies for both major and minor stimuli using a continuous scale. Nonetheless, other research has produced precisely the opposite response pattern. In an attempt to replicate Kastner & Crowder (1990), Gregory et al. (1996) asked both children and adults to identify accompanied and unaccompanied melodies as sounding either happy or sad. Unaccompanied melodies in fact sounded less happy than accompanied melodies, contradicting the earlier result. Similar results were obtained by McCulloch (1999), whose research also suggested that accompanied melodies are perceived more positively than unaccompanied melodies.

1In fact, the authors analyzed their data using a ‘hit or miss’ score depending on whether the emotion response was congruent with major-positive/minor-negative associations. That unaccompanied melodies were perceived more positively can be seen by reflecting the minor-mode results in their Figure 3 about the x-axis (Kastner & Crowder, 1990, p. 199).
The lack of consensus regarding the emotional effects of textural density potentially reflects the different versions of dense textures used in these studies. Kastner & Crowder (1990) employed block chords played on the piano, and Webster & Weir (2005) appear to have used basic chorale-style harmonization. McCulloch (1999) used accompaniments taken from published songbooks, and Gregory et al. (1996) appears to have done the same. One could imagine several musical features which might have differed between these harmonized or accompanied dense textures. The present study will diverge from these previous approaches by using contrapuntal polyphonic stimuli of four different levels of textural density. Hopefully, our results will provide a helpful third perspective.

4.1.1 Approaches to Musical Texture

When describing different textures, musicians often refer to relationships between musical parts, or voices. In this context, a ‘musical voice’ does not necessarily need to be sung, but instead can refer generically to any horizontal musical line. As a result, the relationship between the number of musical voices and the number of performers is not always one-to-one: many complex arrangements are possible (Cambouropoulos, 2008). A monophonic texture could arise both from a single person chanting or from several instrumentalists performing in unison. Similarly, polyphonic music with multiple musical voices could be performed by a single musician using a polyphonic instrument. While polyphony in the broadest sense can refer to any music with more than one simultaneous note, it also can be used in a more specific sense to designate contrapuntal textures with a high degree of independent linear motion (such as occurs in canons, fugues, and Renaissance motets).

In addition to monophony and polyphony, several other types of texture have been identified by music theorists. Homophonic music employs a texture based on chords: many different notes appear at once, but their rhythms are aligned, as in traditional

---

2A musical voice might even refer to imagined melodies that are not fully articulated, as occurs in pseudopolyphony (Davis, 2006).
chorale settings (Clendinning & Marvin, 2005; Laitz, 2012). Some have identified ‘heterophony’ as a distinct texture in which a single melody is simultaneously presented in multiple variations (Cooper, 1981). Berry (1976) adopted an open-ended interpretation of musical texture, accommodating many different intervallic, rhythmic, and melodic components. Dimensional approaches have also appeared: Huron (1989a) describes monophony, polyphony, homophony, and heterophony as four regions in a two-dimensional texture space, representing different degrees of semblant motion and onset synchrony.

The present study investigates the perception of musical emotion in terms of musical voice multiplicity: the number of musical parts or voices simultaneously present in a texture. Because our primary goal was to measure a purely musical effect (and not one which would actually reflect the number of musicians performing), we used stimuli performed by a single person on a polyphonic instrument. A straightforward definition of musical voice was adopted: the number of notated musical lines specified by the composer of contrapuntal works.

Because the voice multiplicity of a piece of music might implicitly or explicitly designate certain social settings, we conjectured that listeners would be more likely to perceive musical emotions corresponding to the musical texture’s social connotation. Specifically, we hypothesized that music with fewer polyphonic voices would be expected to sound lonelier than music with many voices. This hypothesis was inspired by recent exploratory research on musical affect: Albrecht’s (2012) complete surface-level affective analysis of the slow movement from Beethoven’s Sonata No. 8, Op. 13 (“Pathétique”). In Albrecht’s ‘progressive exposure’ paradigm, over two hundred listeners were asked to judge the emotional content of five-second excerpts representing the entire duration of the movement. One of the rated emotions, loneliness, appeared to show sensitivity to the number of musical voices present (see Figure 4.1). Note that the present hypothesis that thin textures should sound more lonely runs somewhat contrary to Webster &
Loneliness ratings for the minor-mode episode’s first two measures are given as z-scores normalized per person. Whiskers indicate the 75th and 25th percentiles of all rating z-scores. Loneliness ratings begin rather high when the melody is initially heard, but drop sharply after an additional voice enters the texture. The present investigation. Figure adapted from Albrecht (2012, Appendix A).

Weir (2005) and Kastner & Crowder (1990), who reported that monophony tends to be associated with increased positive affect over harmonized (multivoice) textures.

In our first experiment, participants rated several brief musical excerpts for four emotions, including perceived loneliness. We predicted that music with fewer musical voices would tend to sound lonelier than music with many musical voices. A second experiment tested whether voice multiplicity effects generalize to emotional valence and sociality. Finally, a third experiment allowed subjects to generate their own emotion labels and directly measured their ability to denumerate the voices in the specific stimuli used. To anticipate our results, perceptual loneliness will exhibit a voice multiplicity effect in which higher voice multiplicity corresponds to less perceived loneliness in polyphonic music. This effect appears to apply more generally to several positive and negative emotions. In general, response patterns for perceived emotion will bear a strong resemblance
to the perceived number of musical voices present, suggesting that they might derive from the same underlying cues.

4.2 Experiment 1: Perception of Musical Loneliness

Our conjecture that textures with fewer voices should sound more lonely implies that listeners form some perceptual or cognitive representation of the number of voices present. We shall refer to this proposed subjective representation as the perceived voice numerosity of a piece of music.\(^3\) The distinction between voice multiplicity and voice numerosity (presuming it exists) reflects the division between external acoustic sound sources and internal auditory percepts, similar to that between amplitude and loudness, or frequency and pitch. Note that music’s voice numerosity need not match its nominal voice multiplicity: perceptual errors could be made. In this case, perceived musical loneliness would be expected to reflect subjective voice numerosity, including any perceptual errors that might have arisen.

As a subjective psychological construct, voice numerosity cannot be directly measured. However, an indirect measure could be used: performance on a voice denumeration task, in which a listener reports the number of voices present in a polyphonic texture. Voice denumeration appears to be subject to certain limitations. While listeners are quite accurate in identifying one-voice and two-voice textures, they increasingly underreport the voices present in denser textures (Huron, 1989b; Parncutt, 1993). In fact, Huron’s listeners reported four-voice textures as having only three voices more often than they answered correctly (see Table 4.2). It would therefore appear that as polyphonic textures grow denser, voice multiplicity levels become less perceptually distinct.\(^4\)

\(^3\)This use of ‘numerosity’ roughly aligns with its use in the study of numerical cognition (Cheatham & White, 1954; White, 1963; ten Hoopen & Vos, 1979; Dehaene, 1997).

\(^4\)As Huron pointed out, this is reminiscent of the ‘\textit{un, deux, trois, beaucoup}’ pattern of responses identified by Descoeudres (1921) in her studies of numerical cognition in children. However, as listeners often misidentified four voices as only three, perhaps ‘one, two, many’ would be more apt to describe polyphonic voice numerosity.
Presuming these voice denumeration results reflect perceptual voice numerosity, we additionally made the prediction that perceived loneliness should change more between one- and two-voice textures than between three- and four-voice textures.

4.2.1 Stimuli

Fugal compositions were chosen because they feature polyphonic textures with clearly defined voice multiplicities. Fugues typically begin with a single monophonic melody: the fugue’s subject. Additional voices then accumulate, with each new entry repeating the subject melody. By excerpting from fugue expositions, sets of stimuli matched for melodic content, tempo, and other composition-specific features can easily be generated. Accordingly, we chose fugues from J. S. Bach’s *Well-Tempered Clavier*, which includes one fugue in each of the 24 major and minor keys (Bach, 1722). Recordings by a professional harpsichordist playing a 1624 clavecin were used (Verlet, 1993). Because the harpsichord’s plucking mechanism results in relatively uniform intensity, articulation, and timbre across its range, performers have limited ability to highlight certain musical voices over others.

Because loneliness is a negatively-valenced emotion, we expected it would be most perceived in minor-mode pieces of music. Hence, all minor-mode fugues of at least four voices were identified. For each, the beginning and ending of each statement of the fugal subject was marked, and 5s excerpts from the center of each multiplicity region were generated, including 200ms logarithmic fade-ins and fade-outs. The G minor fugue was excluded because its exposition does not include stepwise accumulation of voices, and fugues without 5 contiguous seconds for each multiplicity were also excluded. This left seven minor-mode fugues with four isotextural multiplicity conditions each. Two major fugues from the same recording were additionally sampled in order to increase the modal diversity of the stimuli. Ratings for major-mode stimuli were not analyzed in the first experiment.
From each of nine fugues, four 5s stimuli were generated; one for each voice multiplicity condition. Fugues are from J.S. Bach’s *Well-Tempered Clavier* Book I, except for the E major fugue which was taken from Book II.
4.2.2 Procedure

Twenty-eight Ohio State University School of Music undergraduates enrolled in sophomore-level aural skills classes participated for course credit (17 female, 11 male, aged 19-21). Participants were asked to provide emotion ratings for randomly-ordered musical excerpts played through free field speakers at a comfortable level. All trials took place in an Industrial Acoustics Corporation sound-attenuation chamber.

In addition to loneliness, three other emotions were rated to disguise the target emotion, as well as to provide data for exploratory analysis. Subjects rated the prototypical negative and positive emotions of ‘sadness’ and ‘happiness’, and ‘pride’ as a socially-relevant positive emotion. Participants were instructed to rate the emotions they perceived in the music (not necessarily the emotions they might themselves experience). Each randomly-ordered trial presented the prompt: “How [emotion] does this music sound?” Responses were collected using a continuous slider interface labeled only on the left and right as “not at all [emotion]” (scored as 1.0) and “very [emotion]” (scored as 7.0). After each rating, the slider was reset to the center position (4.0). After 5 practice trials, participants rated all 144 stimulus-emotion combinations in random order. A second block of 72 randomly-selected stimulus-emotion pairs immediately followed, used to gauge the reliability of emotion ratings. Post-experimental interviews were conducted to check for demand characteristics and explore strategies used.

4.2.3 Results

Across all repeated subject-stimulus ratings for loneliness, the test-retest correlation was $r_s = .65$, and the average difference in rating was unbiased. A 95% confidence interval around the difference in rating was calculated to be ($-0.15, 0.14$). Hence, duplicate measurements were averaged prior to further analysis in order to increase analytic precision.

Figure 4.3 shows median loneliness ratings for each minor fugue in the study, as affected by voice multiplicity condition. In general, loneliness ratings decrease as voice
Median loneliness ratings for stimuli from each fugue used in experiment one are plotted. Stimuli were 5 second excerpts of fugal expositions from Bach’s *Well-Tempered Clavier*. Keys are given to identify the fugue, and do not indicate a generalization of effects to key. Perceived loneliness drops as voice multiplicity increases, with listeners rating one-voice stimuli as 1.89 points lonelier than four-voice stimuli on average, $t(27) = 8.71$, $p < .001$. Moreover, this effect leveled off as voice multiplicity increased: loneliness ratings dropped more between one- and two-voice conditions than between three- and four-voice conditions, $t(27) = 5.21$, $p < .001$.

multiplicity rises, regardless of fugue. On average, listeners rated a given fugue’s one-voice condition as 1.89 points lonelier on a 7-point scale than its four-voice presentation, $t(27) = 8.71$, $p < .001$, consistent with the first hypothesis. A second hypothesis predicted that loneliness ratings would decline the least between three- and four-voice conditions, reflecting these textures’ supposed phenomenological similarity (as suggested by
Median emotion ratings for negatively- and positively-valenced emotion pairs from experiment one. Filled points (●, ▲) represent the nonsocial emotions sadness and happiness, while open points (○, △) represent the social emotions loneliness and pride.

Voice denumerability limitations. On average, listener ratings dropped 0.99 points more between one- and two-voice conditions than between three- and four-voice conditions, $t(27) = 5.21, p < .001$. This result is consistent with the idea that loneliness ratings and voice denumeration are both mediated by perceptual voice numerosity.

Other Emotions

The effects of voice multiplicity on the perception of happiness, sadness, and pride were also explored (Figure 4.4). The positively-valenced emotions happiness and pride both exhibited increased ratings for higher voice multiplicity, while ratings for the negatively-valenced emotions of sadness and loneliness showed the opposite effect. It would appear that the positive/negative valence of an emotion might dictate the direction of
the relationship. Additionally, the effect of voice multiplicity on ratings for the tested social emotions (loneliness and pride) appears to be stronger than for the nonsocial emotions (sadness and happiness). Comparing the effect on sadness to that on loneliness, a post hoc statistical test revealed that a user’s difference in emotion ratings between one- and four-voice textures is 0.65 points greater for loneliness than for sadness, $t(27) = -3.09$, $p = .003$. The same result is found for positively-valenced emotions: pride ratings changed 0.46 points more than happiness ratings ($t(27) = 2.55$, $p = .009$). That social emotion perception would be more strongly affected by voice multiplicity than nonsocial emotions suggests that a social setting might somehow be inferred by listeners. Experiment two will test this hypothesis explicitly (but fail to find corroborating evidence).

**Strong Ratings of Musical Emotion**

So far, ratings have been treated as being measureable from the scale’s ‘zero’ point (‘not at all [emotion]’). However, an alternate interpretation would consider ratings as directed deviations from the slider’s center point. Viewed this way, emotion ratings seem to be more neutral when more voices are present, and stronger for lower voice multiplicities (see Figure 4.4). To explore this idea, the strength of each emotion rating was calculated by measuring its absolute deviation from the scale’s midpoint (4.0). Histograms of absolute rating strength were prepared for each voice multiplicity (Figure 4.5). Results of six post hoc Mood median tests suggested that the monophonic texture’s median rating strength was higher than each of the other three ($p < .0001$ for every comparison), while other comparisons were statistically nonsignificant. In other words, emotion ratings in experiment one appear to have been particularly strong for monophonic stimuli.
Histograms summarizing a post hoc analysis of emotion rating strength in experiment one. The absolute deviation of each slider-based emotion rating from the center position for different voice multiplicities was calculated for all four emotions. Medians are indicated with arrows (↓). Pairwise Mood median tests indicate that the median emotion rating for the 1-voice stimuli is different from all three others ($p < .0001$), indicated by a star (*).
4.2.4 Discussion

The first experiment provided evidence consistent with the experimental hypotheses, suggested possible generalizations of the effect to emotion valence and sociality, and implied that the strength of emotion ratings might depend on voice multiplicity.

First, the experimental results are consistent with the idea that lower voice multiplicities increase the tendency to perceive musical loneliness. Post-experimental interviews corroborated this finding, as many participants reported that textural density informed some emotion judgments. Moreover, the effect appears to correspond with perceptual limitations on voice denumerability: voice multiplicity effects arose predominantly between one-, two-, and three-voice conditions. We interpret this to implicate perceptual voice numerosity as mediating emotion perception differences. While it is possible that one or more musical variables apart from voice multiplicity could also explain the results (such as event density or complexity), such concomitant changes in the musical surface might themselves serve as voice multiplicity cues. This idea will be addressed in the general discussion to follow.

Second, the results raise the possibility that voice multiplicity effects might generalize to many emotions based on their valence (positive or negative), their sociality (social or nonsocial), or both. However, another plausible interpretation would be that these apparent effects are simply due to a demand characteristic. Each participant rated all four emotions, and several expressed awareness of the $2 \times 2$ design in post-experimental interviews. To address this issue, our second study attempted to replicate these results using a between-subjects design with many exemplars of each emotion category.

Finally, post hoc analysis indicated that emotion ratings might have been stronger when only a single voice was present in the music. This suggests that monophonic and polyphonic textures might differ qualitatively, at least in regard to perceived emotion. To anticipate the results of studies two and three, the finding that monophonic stimuli reliably evoke stronger emotion ratings will be replicated for many experimenter-supplied emotions, but not for emotions that participants themselves volunteer.
4.3 Experiment 2: Generalization to Emotion Valence and Sociality

While results of experiment one suggest that voice multiplicity might influence perceived emotion ratings in interaction with emotional valence and sociality, the initial study design introduced a possible demand characteristic. This second experiment provides an explicit test for these effects. A between-groups design was used with regard to valence, and within each valence group, three social and three nonsocial emotions were chosen. In all, three hypotheses were tested: First, positive emotion ratings should increase, and negative emotion ratings should decrease as voice multiplicity rises. Second, social emotions should exhibit larger-magnitude voice multiplicity effects than nonsocial emotions. Third, we hoped to reproduce the result that monophonic stimuli evoke stronger responses than any other multiplicity level.

Categorizing emotions into ‘social’ and ‘nonsocial’ types draws on experimental evidence from developmental studies (Stipek et al., 1990; Kagan, 1981) and more recently, functional neuroimaging (e.g., Burnett et al., 2009). Experiencing certain emotions—such as pride, shame, guilt, embarrassment, and envy—depends on the ability to model the mental states of oneself and others. These emotions have been variously referred to as social emotions (Teroni & Deonna, 2008; Minzenberg et al., 2006), self-conscious emotions (Lewis et al., 1989; Tracy et al., 2007), or moral emotions (Eisenberg, 2000; Tangney et al., 2007). Social emotions often overlap with so-called ‘secondary emotions,’ distinguishable from the ‘primary’ or ‘basic emotions’ which appear early in life (Lewis & Michalson, 1983), and have cross-culturally recognizable facial expressions (Ekman, 1973). Both loneliness (Weiss, 1975; Cacioppo & Patrick, 2008) and pride (Tracy & Robins, 2007; Williams & DeSteno, 2008) have been described as social emotions.

Although we interpreted the results of Experiment 1 as suggesting an effect due to sociality, emotions like loneliness and pride could also be distinguished from sadness and

---

5It was expected that the distinction between social and nonsocial emotion types would not be immediately evident to subjects.
Table 4.1: Emotions Used in Experiment Two

<table>
<thead>
<tr>
<th>Valence Group</th>
<th>Sociality</th>
<th>Emotion</th>
<th>Prototypicality</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive</td>
<td>Social</td>
<td>Pride</td>
<td>3.14</td>
<td>13,805</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Love</td>
<td>3.94</td>
<td>153,466</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Compassion</td>
<td>3.62</td>
<td>5,461</td>
</tr>
<tr>
<td></td>
<td>Nonsocial</td>
<td>Happiness</td>
<td>3.77</td>
<td>8,177</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Contentment</td>
<td>2.92</td>
<td>796</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Excitement</td>
<td>3.51</td>
<td>9,231</td>
</tr>
<tr>
<td>Negative</td>
<td>Social</td>
<td>Shame</td>
<td>3.44</td>
<td>9,319</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Loneliness</td>
<td>3.41</td>
<td>3,121</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Envy</td>
<td>3.58</td>
<td>3,355</td>
</tr>
<tr>
<td></td>
<td>Nonsocial</td>
<td>Sadness</td>
<td>3.68</td>
<td>4,734</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Fear</td>
<td>3.83</td>
<td>49,410</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Disgust</td>
<td>3.42</td>
<td>2,879</td>
</tr>
</tbody>
</table>

Note. Emotions rated in the second experiment on emotion valence and sociality. Prototypicality scores are average ratings on a 4-point scale provided by 112 university students (Shaver et al., 1987). Word frequency reflects the number of occurrences in a corpus of 450 million words used in contemporary American English (Davies, 2008).

happiness by other factors, such as categorical prototypicality (Rosch, 1973, 1975; Lakoff, 1987), or familiarity (Zajonc, 1968; Bornstein, 1989). In order to choose a well-stratified collection of emotions, we made use of the hierarchic analysis of Shaver et al. (1987), who clustered 135 emotions into categories and subcategories using results of a sorting task. Six positive and six negative emotions were chosen such that no two emotions fell in the same subcategory (see Table 4.1). Shaver et al. (1987) also provided subjective ratings of the emotions' prototypicality: the degree to which listeners consider it to be an emotion. To estimate each emotion word’s familiarity, its frequency ranking in the Corpus of Contemporary American English was found (Davies, 2008). These two covariates would be used for post hoc exploratory analysis.
4.3.1 Stimuli and Procedure

Sophomore-level music students aged 19–21 participated for Aural Skills course credit. Participants were alternately assigned to the positive-emotion ($N = 18$) or the negative-emotion group ($N = 19$) in the order in which they arrived. Using identical slider interfaces as in the first study, the positive-emotion and negative-emotion groups rated identical stimuli for the emotions given in Table 4.1, in randomized order. Because the major-mode and minor-mode excerpts did not appear to elicit different response patterns, results for all 36 stimuli were analyzed together (7 minor fugues and 2 major fugues with four voice multiplicity conditions).

4.3.2 Results and Discussion

Results for the positive and negative emotion groups are compared in Figure 4.6. It is clear that positive and negative emotion ratings responded in opposite directions to increasing polyphonic voice multiplicity. The mean difference in emotion rating moving from monophonic to four-voice music was $-0.83$ for the negative emotion group ($t(18) = 6.10, p < .001$) and $+1.44$ for the positive emotion group ($t(17) = 12.19, p < .001$). In fact, all twelve emotions responded in the expected direction, with positive emotion ratings increasing with voice multiplicity and negative emotion ratings decreasing. In short, it seems that thinner polyphonic textures are associated with negative emotional valence and thicker textures with positive emotional valence. Once again, the effect appeared to taper off as multiplicity rose.

To test for an effect of emotion sociality, two-way mixed models were fit to the responses of each the positive and negative emotion groups, using multiplicity and sociality to predict emotion ratings, blocking by both subject and fugue.$^6$ The hypothesis that social emotions would respond more strongly than nonsocial emotions would predict that

$^6$Although the slider interface introduced ceiling and floor effects, these would be expected to reduce statistical power, increasing false negatives, not false positives. Model diagnostics indicated that residuals were well-behaved.
Median emotion ratings per emotion are given for the negative (N=19) and positive (N=18) emotion groups in experiment two. As predicted, positive emotion ratings increased with increasing voice multiplicity while negative emotion ratings decreased. Using ANOVA analyses, main effects were identified for voice multiplicity in both groups. However, emotion sociality did not appear to change the magnitude of this effect, contrary to the experimental hypothesis.
an interaction between voice multiplicity and sociality should be present. To control for multiple tests, p values below .01 were treated as statistically significant, for a familywise error rate under .06.

Main effects were identified for voice multiplicity (considered linearly) in both the negative and positive groups, with higher multiplicity associated with lower negative emotion ratings, $\chi^2(1) = 183$, $p < .001$, and higher positive emotion ratings, $\chi^2(1) = 556$, $p < .001$. Additionally, a main effect was found for emotion sociality in the positive emotion group ($\chi^2(1) = 129$, $p < .001$); social positive emotions got higher ratings than nonsocial positive emotions overall. However, no significant interaction effect between voice multiplicity and emotion sociality was found for either emotion valence group, failing to support the hypothesis that social emotion perception ought to depend on multiplicity more strongly than nonsocial emotion perception. ($\chi^2(1) = 0.18$ for negative emotions; $\chi^2(1) = 0.91$ for positive emotions)

A post hoc reanalysis using the assembled emotion prototypicality and frequency covariates instead of sociality indicated that emotion ratings did not depend on prototypicality, but might have depended on an emotion term’s frequency of occurrence. Specifically, positive emotions were rated higher when the word was more frequent ($\chi^2(1) = 10.1$, $p = .002$) and negative emotions were rated lower ($\chi^2(1) = 21.4$, $p < .001$). This potentially reflects a version of the ‘mere exposure’ effect, wherein more frequently-encountered stimuli are perceived more positively (Zajonc, 1968). This would indicate that emotion ratings at least partially reflect listener preferences and attitudes toward the music or the task demands, not only the music itself. Because of the high familiarity of the positive social emotions (particularly ‘love’), the observed main effect for sociality might be spurious. Although the absolute ratings appeared to be affected by emotion word frequency, neither covariate seems to have impacted the absolute strength of emotion ratings.
Finally, monophonic stimuli again resulted in extreme emotion ratings more frequently than any other multiplicity level, replicating the finding from experiment one (logistic $z = 7.76$, $p < .001$).

### 4.4 Experiment 3: Emotion Rating Strength and Voice Denumerability

Results of the first two experiments indicate that as polyphonic textures thicken, negative emotion perception decreases while positive emotion perception increases. Additionally, the effect appears to taper off in a manner similar to limitations in voice denumerability, with 3-voice and 4-voice textures exhibiting smaller differences in emotion perception than 1- and 2-voice textures. This is consistent with the idea that perceptual voice numerosity mediates both responses.

However, this interpretation relies on denumeration data from separate studies using different stimuli and procedures: Huron (1989b) and Parncutt (1993). To provide internal validity, experiment three endeavors to replicate Huron’s polyphonic voice denumeration results using the current protocol. Additionally, we will attempt to bolster the external validity of emotion perception findings by using only the most expressive subset of stimuli and by allowing listeners to freely specify which emotions they perceive.

In order to make more refined predictions regarding voice denumerability, it is worthwhile to theorize possible perceptual or cognitive processes which could produce a subjective sense of voice numerosity or otherwise determine denumeration responses. First, listeners could estimate the number of voices, making approximate, nonsymbolic judgments of magnitude based on some feature of the sound, such as event density. Second, listeners might be able to quickly identify certain voice multiplicity levels without counting, in a process which has been called ‘subitizing’ in the context of vision and numeric
Subitizing would seem most likely in a musical context for small voice multiplicities such as one or two. Third, listeners could deliberately count the number of voices by shifting their attention between perceived auditory streams (Bregman & Campbell, 1971; Bregman, 1990). A counting strategy could be facilitated by recognizing voice entries and exits, but need not rely on them exclusively.

Note that these processes differ as to the response profile they would be expected to produce with regard to central tendency (accuracy) and statistical dispersion (precision). For example, subitizing ought to have both high accuracy and high precision, as it represents a direct apprehension of the voice multiplicity. By contrast, estimation would be expected to be somewhat imprecise, and potentially systematically biased as well (e.g., Tversky & Kahneman, 1974). Counting processes could result in several response patterns, possibly depending on a listener’s streaming capacity or their working memory. In general, one might expect counting responses to be reasonably precise, but potentially biased in one direction or another, depending on whether voices are actively entering or exiting the music.

Because they listened to music played from the beginning, listeners in Huron’s denumeration study would presumably have had all three processes available; however, the current stimuli might not offer such flexibility. Huron’s listeners would have been able to identify entries and exits, but the stimuli used here are isotextural, containing no voice entries or exits by design. The excerpts’ short length could create further problems, as stream segregation appears to be cumulative (see Bregman, 1978): listeners tend to hear one stream at first before subsequently resolving more. In all, because our stimuli are both short and isotextural, counting would not be expected to be a successful strategy in the present denumeration task.

There is some debate as to whether subitizing in fact reflects a faculty distinct from counting. For present purposes, we treat it separately based on the rapidity with which it appears to operate.
In addition to replicating Huron’s results, experiment three tests the hypothesis that voice denumeration and emotion ratings will follow similar response patterns with regard to voice multiplicity. Additionally, we again tested whether monophonic stimuli would be more likely to evoke strong emotion ratings.

4.4.1 Procedure and Stimuli

Procedures were similar to those of the first two experiments, but attempted to minimize a possible design artifact due to nonexpressive musical excerpts: If listeners do not hear the music as emotionally expressive of the presented emotion (or any emotion for that matter), then their emotion ratings might have limited meaning. In order to ensure the fugues employed were sufficiently expressive, all $36 \times 12 = 432$ stimulus-emotion combinations from Experiment 2 were ranked in terms of their median rating’s absolute deviation from the middle point of the scale. The rank sum of these deviations across each fugue’s four multiplicity levels for all emotions was used as a measure of the fugue’s expressivity. Of the nine fugues, the B minor, E major, C major, and B♭ minor fugues were selected as the most expressive.

Participants were undergraduates enrolled in sophomore-level aural skills classes and participated for course credit (N=24, 10 females, 14 males, age 19–21). The study was divided into three parts: emotion generation, emotion rating, and denumeration. In the first part, subjects listened to four excerpts as many times as they wished. For each, subjects generated three emotion labels completing the sentence, “I hear ____ in the music.” They subsequently chose the one emotion label which fit the excerpt the best. The four stimuli used to generate emotions represented each of the four fugues and each of the four voice multiplicity levels; each of the 24 subjects listened to one of the possible subsets of stimuli satisfying this constraint in random order.

The second part of the experiment employed the slider-rating paradigm as before, using the four listener-selected emotion labels in combination with the 16 different musical
excerpts for 64 total randomized emotion-excerpt trials. In the third part, subjects denu-
merated the voices in excerpts by typing a single digit into an empty text field. All sixteen
excerpts were presented twice, in two randomized blocks. Participants then completed
the Ollen Musical Sophistication Index battery (Ollen, 2006).

Volunteered emotions were classified as ‘positive’, ‘negative’, or ‘neutral / don’t
know’ by an independent group of 6 raters. Emotions for which five of six raters agreed
were grouped into valence categories for analysis.

4.4.2 Results and Discussion

Denumeration

Voice denumeration responses (Table 4.2) ranged from one to six voices, and the
rest-retest correlation between the two denumeration blocks was $r_s = .69$. For each
multiplicity level tested, listeners’ most common voice denumerations matched those
of Huron (1989b), with the mode response in both studies being ‘1, 2, 3, 3’ for the
first four multiplicities. Average denumeration responses are quite concordant as well,
with underreporting being much more common than overreporting (see Figure 4.7, top
panel). Listeners with higher OMSI scores made fewer errors overall (logistic $z = 2.10$,
$p = .036$).

We additionally tested whether the form of denumeration error types was concordant
with Huron’s results. To do so, the confusion matrices from each study were normalized
by number of trials per multiplicity level, and denumerations greater than five were
truncated. The probabilities of each incorrect response type in the confusion matrices
(i.e., non-diagonal entries) were tested for statistical dependence using Kendall’s tau. The
patterns of erroneous responses were highly correlated, $\tau(16) = .713$, $p < .001$, reflecting
rank-order correspondence of both bias and dispersion effects between the studies.

Despite this rank-order correlation, error rates in the present study were substan-
tially higher for 2-voice and 3-voice stimuli than in Huron’s previous results (Figure 4.7,
bottom panel). This could reasonably be attributed to the aforementioned differences
Table 4.2: Voice Denumeration Confusion Matrices

(a) Huron (1989b)

<table>
<thead>
<tr>
<th>Multiplicity</th>
<th>One</th>
<th>Two</th>
<th>Three</th>
<th>Four</th>
<th>Five</th>
</tr>
</thead>
<tbody>
<tr>
<td>Denumeration Response</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>One</td>
<td>309</td>
<td>6</td>
<td>15</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>Two</td>
<td>19</td>
<td>1368</td>
<td>63</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Three</td>
<td>0</td>
<td>127</td>
<td>1121</td>
<td>12</td>
<td>0</td>
</tr>
<tr>
<td>Four</td>
<td>0</td>
<td>77</td>
<td>961</td>
<td>608</td>
<td>14</td>
</tr>
<tr>
<td>Five</td>
<td>0</td>
<td>22</td>
<td>165</td>
<td>466</td>
<td>399</td>
</tr>
</tbody>
</table>

(b) Experiment 3

<table>
<thead>
<tr>
<th>Multiplicity</th>
<th>One</th>
<th>Two</th>
<th>Three</th>
<th>Four</th>
<th>Five</th>
</tr>
</thead>
<tbody>
<tr>
<td>Denumeration Response</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>One</td>
<td>185</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Two</td>
<td>2</td>
<td>151</td>
<td>37</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Three</td>
<td>0</td>
<td>87</td>
<td>95</td>
<td>9</td>
<td>0</td>
</tr>
<tr>
<td>Four</td>
<td>0</td>
<td>43</td>
<td>82</td>
<td>64</td>
<td>1</td>
</tr>
</tbody>
</table>

*Note.* Results from Huron (1989b) are continuous denumeration responses by five musicians listening to J. S. Bach’s ‘Saint Anne’ organ fugue. Results from Experiment 3 are from 24 undergraduates enrolled in a music course, and correspond to isolated 5s excerpts of harpsichord music. Mode responses per multiplicity level are in bold text. In both cases, underreporting predominated, with the most common responses following a ‘1, 2, 3, 3’ pattern for the first four multiplicities.
Comparison of average voice denumeration response and error rates obtained in Huron (1989b) and experiment three. Huron’s results are from five expert musicians who listened to J. S. Bach’s ‘Saint Anne’ organ fugue and continuously reported the voices present. In the present study, 24 undergraduate music students responded to five-second harpsichord stimuli. In both cases, underreporting was more common than overreporting, and the central tendency of responses was similar. Error rates were much higher in the present study for 2-voice and three-voice textures. Error rate whiskers are raw 95% binomial confidence intervals.
between the two tasks: while Huron’s listeners would have been able to count voice entry and exits over a more typical listening experience, our subjects had only five seconds of music with no entries or exits.\(^8\) The uniformly high precision and accuracy for monophonic stimuli suggest that instant recognition (subitizing) might have occurred in both studies for one-voice textures. While Huron’s listeners would have potentially been able to apply counting techniques to detect the entry of a second or third voice, our listeners were presumably much more likely to use estimation because the stimuli were isotextural. Finally, in both cases did listeners appear to resort to estimation processes when denumerating voices in four-part textures, resulting in very high error rates.

**Emotion Perception**

As predicted, positive emotion ratings increased with voice multiplicity while negative emotion ratings decreased, replicating the emotion valence effect identified in Experiment 2. Median positive and negative emotion ratings are displayed in Figure 4.8, with the negative emotion rating scale inverted to facilitate direct comparison. Qualitatively, it would appear that median emotion ratings do indeed follow a similar pattern to average voice denumeration outcomes. Both positive emotion ratings \(r = .584\) and negative emotions \(r = -.558\) were strongly correlated with average voice denumerations. As an inferential test, a linear model was used to predict normalized emotion ratings using a subject’s first set of denumeration responses, blocking by subject and fugue. A main effect for voice denumeration was identified \(\chi^2(5) = 120.1, p < .0001\), consistent with the idea that the two tasks depend on the same information.

One could describe the similarity of denumeration and emotion responses as appearing biphasic: both seem to respond approximately linearly for voice multiplicities of two.

---

\(^8\)Additionally, Huron’s participants were five expert musicians, while our participants were undergraduate music students. Given that higher OMSI scores correlate with denumeration performance, this disparity could also be attributed to differences in skill.
Voice denumeration compared with median ratings for negative and positive emotions in experiment three. Emotions were classified by an independent panel of judges; negative emotion ratings are plotted on an inverted scale to facilitate comparison. Both voice denumeration and emotion ratings tend toward similar response patterns, consistent with the idea that perceptual voice numerosity contributes to both. Subjectwise average voice denumerations and average emotion ratings per multiplicity level were strongly correlated, for both positive emotions ($r = .584$) and negative emotions ($r = -.558$). Denumeration whiskers are raw 95% $t(23)$ confidence intervals; emotion rating whiskers are 95% binomial confidence intervals around the median.
and higher, but monophonic stimuli are qualitatively different. This could correspond to estimation processes and subitizing, respectively.

**Emotion Strength and Monophonicity**

Our final hypothesis predicted that the proportion of strong emotion ratings would be highest for monophonic stimuli. Surprisingly, the probability that a listener would make a strong rating had no significant relationship with voice multiplicity in experiment three, failing to support the hypothesis (logistic \( z = -1.28 \), \( p = .2 \)). It would appear that the previously-observed results in the first two experiments should be interpreted in light of the particular experimental task employed. Specifically, monophonic stimuli were associated with stronger emotion ratings when emotions were experimenter-provided, but not when they were listener-generated.

One might suppose that when participants were provided with emotion words in the first two experiments, these terms were understood as *personal emotions* describing affective experience or emotional expression of an individual. Perhaps monophonic stimuli were rated more strongly simply because they resemble a single person’s voice. In this view, textures with more than one voice might have resembled multiple agents to which differing emotions could be attributed: hence, the emotional expression would be much less clearly defined. An alternative view would be that monophonic stimuli simply have fewer relationships present, making emotion perception more straightforward. There might also be attentional influences: unattended musical voices appear to influence the perception of attended musical voices (Davison & Banks, 2003; Fujioka et al., 2005).

Regardless of the reason, it would seem that experimenter-provided emotion words led participants to seek melodies by which to judge emotional content.

By contrast, the third experiment requested that listeners provide their own emotion labels after listening to musical stimuli of all four multiplicity levels. These emotions might have been understood as *musical emotions* with specific musical meanings and referents. Thus, participants in experiment three might have been making ratings based
on an excerpt’s musical similarity to the one that initially evoked the emotion instead of to a more typical understanding of individually experienced or expressed emotions. Because emotion terms were generated for all four multiplicities, it is understandable why monophonic stimuli would not always evoke the strongest responses.

4.5 General Discussion

Overall, our results are consistent with the idea that textural density exerts a general influence on the perception of musical emotion, with thicker polyphonic textures associated with increased emotional positivity. The specific pattern of emotion responses appears to coincide with voice denumeration results, suggesting that both could depend in part on the same underlying percept: voice numerosity. While the relationship between voice denumeration and emotion ratings is correlational, their relationships to voice multiplicity appears to be causal.

Why might increased voice multiplicity be associated with positive emotions? A ‘musical depiction’ explanation would suggest that the music is interpreted as representing happy or otherwise positive social situations. Parties, dances, celebrations, ceremonies all occur in group settings marked by positive affect. On the other hand, one could also imagine instances where high voice numerosity would not necessarily be associated with positive affect, depending on the social or situational context. A traveling merchant would surely prefer to hear the sound of one marauding warrior instead of one hundred. Perhaps an opposite voice multiplicity effect would be expected for loud, aggressive, or threatening timbres.

An explanation based on musical depiction might incorporate some degree of cognitive appraisal of artistic communication, but this would not necessarily be required. One could instead conjecture that voice numerosity or related information is continuously used by humans as we monitor our auditory environment. Considered this way, it would not be unexpected if voice numerosity percepts directly lead to emotional responses much like other auditory information would. In such a way, implicit processes
related to the interpretation of vocal signals might be expected to lead to evaluations of musical emotion.\(^9\)

There are many reasons to believe that auditory voice numerosity percepts would be comforting to social organisms, provided that the vocalizations are nonthreatening. Many gregarious animals who live in groups for protection, foraging, or social interaction appear to use auditory signals known as ‘contact calls’ (or ‘cohesion calls’) to communicate their presence to conspecifics, including horses, cows, pigs, chickens, and primates. One could speculate that music is sometimes processed as though it were composed of human contact calls—sound made simply to communicate presence. In this view, any sound resembling several nonthreatening human voices would be perceived positively. Indeed, following the advent of recording technology, music seems to have become widespread in public spaces, with background music filling in socially dangerous silences, placating callers on hold, or calming early elevator passengers (Lanza, 2004).\(^{10}\)

4.5.1 Numerosity Cues

If voice numerosity is indeed associated with more positive emotion perception, then one would expect that thicker musical textures ought to produce the effect only to the extent that they increase perceived voice numerosity. If so, then voice numerosity might be a third variable which could account for the conflicting results of Kastner & Crowder (1990) and Webster & Weir (2005), in which thicker textures were associated with decreased positivity. We will here suggest a broader conjecture concerning musical texture’s emotional effects based on two possible musical numerosity cues: note density and onset density.

When hearing music, listeners might perceive textures in terms of the note density of the music, defined as the average number of individual notes occurring per unit time. As

\(^9\)In experiment two, emotion ratings were possibly affected simply by the familiarity of the emotion terms, suggesting that responses might at least be affected by implicit processes.

\(^{10}\)One could also argue that this state of affairs is not new at all: whistling, singing, humming, and instrumental performance could make music equally ubiquitous in nonindustrial cultures.
voice multiplicity rises, note density would be expected to increase.\footnote{Although perhaps not strictly linearly: different voices might contain different numbers of notes (see Chapter 5).} Perception of note density would depend on processing both synchronous and asynchronous note onsets (representing intervals and block chords on one hand and individual notes on the other). Another related musical feature would be temporal onset density: the number of event onsets per unit time, treating notes, intervals, and block chords as single temporal events.

Although note density would always be expected to increase as more voices are added to a texture, onset density might not. If an added melody is constrained to the same rhythm as the first melody, a homophonic musical texture results. In this case, onset density would be unchanged even as note density rises. On the other hand, if the added melody’s notes do not coincide with any previously occupied metric positions, a polyphonic texture is formed, and onset density and note density rise identically. Hence, the degree of onset synchrony found in a piece could be used to measure its homophonicity (Huron, 1989a). The relationship between note density, onset density, and onset synchrony is shown in Figure 4.9 for the 36 excerpts used in experiment two.

Knowledge of any two of note density, onset density, and onset synchrony would imply the third, and any could potentially be used to perceive or estimate how many voices are present. Specifically, polyphonic textures imply both high onset densities and low onset synchrony, which appear to increase listener ability to resolve individual voices (Rasch, 1981; Huron, 1993, 2001). Polyphonic textures would thus be expected to produce higher voice numerosity percepts due to their increased onset density. Notably, emotion ratings from experiment two appear to be more closely related to average onset density than average note density, because their greatest changes occurred between smaller voice multiplicities.\footnote{In general, onset density cannot rise indefinitely as voices are added to metered music, because all metric positions will eventually be filled.} One could therefore conjecture that subjective voice numerosity percepts might depend on musical onset density, so increased onset density ought to be associated with increased positivity of perceived emotion. Note that this
Average note and onset density in the minor-mode fugal stimuli used in the first study, expressed in events per second. Note density increases with voice multiplicity in a fairly linear relationship. Onset density, on the other hand, changes most between one-voice and two-voice textures, before leveling off. The shaded region represents the music’s onset synchrony. Increasing onset synchrony would lower onset density, resulting in a more homophonic texture; low onset synchrony would result in high onset density, characteristic of polyphonic textures. The degree of onset synchrony is a measure of the ‘harmonic depth’ of typical events.
conjecture concords with independent observations that greater rhythmic complexity and faster tempi are associated with positive valence.

If this is true, then the present stimuli might be most saliently distinguished from those of Kastner & Crowder (1990) and Webster & Weir (2005), because the present stimuli have high onset densities characteristic of polyphonic writing. This would lead to a comparatively higher perceptual voice numerosity. This feature seems to be hared by the accompaniments used by McCulloch (1999) and Gregory et al. (1996) which also led denser textures to be associated with more positive emotional valence.

While this would explain why the low onset density music used by Kastner & Crowder (1990) and Webster & Weir (2005) would not increase the music’s valence, it does not immediately explain why these researchers would in fact have observed the opposite response: increased negative valence for thicker textures. A tentative reason could be proposed, however, based on the result from experiments one and two that monophonic stimuli evoke stronger ratings than thicker textures. Instead of using separate scales for ‘happiness’ and ‘sadness’, Webster & Weir (2005) asked listeners to use a single continuous valence scale from ‘sad’ to ‘happy’ to make their ratings. On the whole, their excerpts tended to be perceived on the positive side of the spectrum, with unharmonized melodies being perceived as more positive. While this could be interpreted as indicating that the harmonized melodies were perceived ‘more negatively,’ an alternative interpretation would be that the emotion ratings were simply more neutral (‘less positive’), because the thicker homophonic texture sounded less like an individual voice, or otherwise obscured the melodic/prosodic cues listeners sought.

4.5.2 Future Directions

The observed effect of musical texture on voice denumeration and emotion perception would benefit from generalization to other musical styles, modes, composers, and instruments. Especially, one would want to more rigorously test the idea that voice
numerosity per se mediates the emotion effect, rather than actual voice multiplicity. Direct means of manipulating and measuring the number of musical streams listeners are able to hear could be developed, potentially exploiting pseudopolyphony or interactions with the visual modality. The idea that homophonic music provides fewer numerosity cues than polyphony could be tested using specially constructed stimuli that independently manipulate note density, onset density, or onset synchrony. Specifically, the idea the onset density serves as a voice numerosity cue would be straightforward to verify experimentally.

Fundamental to the ‘social organism’ explanation is the idea that human speech sounds ought to produce a voice numerosity effect on perceived emotion similar to that of music. While intuitively appealing, this conjecture assumption must be subjected to empirical verification before it can be accepted. One could also determine whether spoken voice denumerability exhibits similar limitations to those for musical voices, or shares the proposed numerosity cue of onset density. One would hope to find these effects cross-culturally.

Finally, it appears that task demands and implicit effects play important roles in musical emotion rating paradigms. Much might be gained by carrying out several replications for any musical emotion study, each implementing a slightly modified protocol.
Chapter 5: Musical Line Mediates Musical Pitch-Speed Relationships

Abstract

We conducted four tests of the conjecture that higher musical pitch coincides with faster musical speeds in composition and performance. First, a ‘notewise’ examination of western musical scores tested whether longer (i.e., slower) notes tend to have lower pitches. Results were genre-dependent, with three of six sampled styles exhibiting the predicted effect. A second study considered an independent sample of Western music part-by-part and found that lower musical voices tend to have significantly fewer notes than higher voices. The third study used instrumental recordings to directly measure event onset densities in notes per second. A strong correlation ($r_s = 0.736$, $p < .002$) between performed note speed and an instrument’s pitch range (tessitura) was found. Finally, a fourth study indicated that Baroque ornaments are more likely to appear in higher musical parts. Considered together, these four studies suggest a pitch-speed relationship that is most evident when the methodology preserves the notion of musical ‘line.’ We outline several possible origins for the observed effect.

5.1 Introduction

To a musician or lover of music, the idea that lower voices in western music move at slower rates than higher voices might seem intuitively obvious. Some of the most memorable musical passages are virtuosic coloratura arias or dazzling violin cadenzas, both of which typically feature rapid passagework in a high pitch range. Similarly, some

Note. This chapter is adapted from Broze, Y. & Huron, D. (accepted). Is higher music faster? Pitch-speed relationships in western compositions. Music Perception.
of the most dramatic music written in lower pitch ranges is relatively slow, as is the case for funeral marches or laments. Nonetheless, it is possible that this supposed pitch-speed relationship does not accurately reflect compositional and performance practice, and instead reflects cognitive or perceptual biases.

Musical dimensions such as pitch and speed exhibit perceptual interaction effects, in which a change in one dimension can influence the perception of another (see, e.g., Prince et al., 2009). Specifically, Collier & Hubbard (2001) found that listener perception of the speed of isochronous stimuli is affected by pitch height and contour in addition to tempo: higher-pitched tone sequences were perceived as being faster than lower-pitched tone sequences. Boltz (2011) recently replicated and extended these findings, again using isochronous pitch sequences. Melodies of higher pitch and brighter timbre were judged to be faster than comparison melodies of the same nominal tempo. Presuming that perceptual pitch-speed interactions apply to typical listening situations, one would predict that higher-register music will generally sound faster than it truly is. Consequently, the intuition that higher music tends to be faster might be partially or even entirely illusory. The primary purpose of the present investigation is to empirically test whether higher music truly does tend to be faster.

To explain the observed perceptual effects, Boltz (2011) proposed that perceptual interactions might derive from regularities in one’s auditory environment. The reasoning is thus: if high pitches and fast speeds tend to coincide, then information about one dimension could provide useful information about the likely state of the other. Therefore, our test for an objective pitch-speed relationship in music carries implications regarding the origin of the perceptual interaction itself.

We conducted four separate studies testing the hypothesis that fast musical speeds tend to coincide with high musical pitch in western composition and performance practice. Because the notion of musical speed does not precisely correspond to a single obvious measurement technique, we applied four complementary methods, hoping to provide a more complete account of pitch-speed effects in musical organization.
5.1.1 Measuring Musical Speed

Measuring musical speed in practice poses challenges not faced in controlled experimental conditions. Stimuli from experiments are often isochronous sequences of notes, whose speed is straightforward to measure as onset densities (i.e., notes per second) or interonset intervals (IOIs) (Collier & Hubbard, 2001; Boltz, 2011). In short, the speed of isochronous sequences can simply be treated as identical to the music’s tempo, yielding a measure of speed which is both intuitive and practical.

However, most western musical compositions do not consist solely of monophonic sequences of isochronous notes. Instead, music is typically polyphonic: pieces include multiple sounding notes at once, organized into melodies, harmonies, basslines, and accompaniment patterns. Additionally, most western rhythmic schemes are considerably more complex than simple isochrony: note onsets and durations are usually specified in terms of a note’s position in a hierarchic metric framework. An underlying pulse (the tactus) is subdivided into several metric levels, and musical rhythms are specified relative to the prevailing meter. In western music, ‘tempo’ is understood not as a characteristic of notes and melodies, but as a description of the meter’s tactus.

The distinction between meter and rhythm helps explain why tempo alone cannot account for all types of musical speed. For example, statements such as ‘the violin is playing much faster than the cello in this piece’ must depend on factors apart from tempo, since both parts would be playing within the same meter. This suggests that many possible forms of perceptual musical speed could be recognized, potentially with contributions from metric structure, tempo, rhythms (Kuhn, 1987), articulation (Geringer & Madsen, 2006), or even distances in ‘pitch space’ (c.f. Boltz, 1998). Hence, one might expect that a definitive study of musical speed would depend on a nuanced model of what exactly is meant by fast and slow musical speeds. Because no such model has yet been widely accepted, the present research employs multiple different operational definitions of ‘musical speed’ in order to test the hypothesis that western compositions exhibit pitch-speed relationships.
Excerpt of piccolo *obbligato* and tuba part from John Philip Sousa’s March, *Stars and Stripes Forever* (Philadelphia: Theodore Presser Co., public domain). Both parts are given in concert pitch. The piccolo part is considerably higher in pitch and noticeably faster. In particular, the piccolo plays seventeen notes to the tuba’s eight, includes melodic trills, and features notes with shorter duration.

In order to motivate possible indices of musical speed, we will briefly discuss a musical example: John Philip Sousa’s march, *Stars and Stripes Forever*. The *Grandioso* section of this turn-of-the-century march contains a fast-paced obbligato part which is assigned to a piccolo player (Figure 5.1). The tuba part, by contrast, plays the ‘oom’ of a typical ‘oom-pah’ march pattern. Here, the piccolo and tuba parts appear to be moving at different speeds, despite being played at the same tempo.

The apparent cooccurrence of fast and slow speeds in this piece suggests several musical features apart from tempo that might contribute to musical speed. At least three easily-measurable factors are apparent. First, the average *notated duration* in the faster piccolo part is somewhat shorter than in the slower tuba part. Second, the piccolo part plays *more notes* than the tuba part: seventeen versus the tuba’s eight. Third, the piccolo contains two trills: *melodic ornaments* that specify rapid alternation between two
These three temporal and rhythmic features—note length, note count, and ornaments—might all be useful when measuring musical speed. Specifically, these features appear to relate to either *melodic speed* (the amount of ‘musical motion’ present in certain melodic line) or *musical activity* (how active the music seems to be). In the following, we employ all three features as indices of musical speed.

One additional methodological challenge remains: the question of how to interpret notated durations, which specify metric time internal to the piece instead of clock-time. While some compositions include tempo indications in beats per minute, others give only imprecise written instruction such as ‘Adagio’, or omit tempo markings altogether. There are other complications as well: compositions might change tempi several times, and performers often are free to employ expressive timing deviations such as rubato, accelerando, and ritardando, whether or not they are expressly indicated. Interpreting notated durations therefore presents a potential methodological pitfall: even in the optimal case that two pieces have identical tempo markings, the length of a notated quarter note in one composition might still differ from a notated quarter note in the other. Fortunately, one can minimize the problem of tempo mismatches by making duration comparisons only *within* an individual musical work, and not between multiple pieces. Furthermore, one could seek converging evidence from sources other than musical scores, such as performance recordings. In a following four studies, both strategies are applied.

The first study measures melodic speed using note duration, where short durations would be associated with fast melodic speeds, and *vice versa*. Because each note is considered individually in terms of its pitch height and its duration, we could call this a ‘notewise’ approach. Pitch-duration correlations were calculated to measure pitch-speed

---

1 One could also point out that the piccolo part covers a wider span of pitch space (an eleventh) than does the tuba (a fourth). Spatial metaphors for pitch height imply that a one-second melody which traverses a large pitch interval would be perceived as ‘faster’ than one which traverses a smaller interval in the same amount of time (c.f. Boltz, 1998). In the present study we restrict our discussion of musical speed to the temporal dimension.
relationships independent of musical part or instrument. Inspired by the fact that music tends to be organized into individual ‘parts’ or ‘voices,’ a second study employs note counts to measure musical ‘activity,’ predicting that higher musical parts would contain more notes. We consider this to be a ‘partwise’ approach to measuring musical speed. Study 3 measured musical speeds in notes per second from recordings, testing for instrument effects on likely performance speeds. Musical speeds were correlated with an index of instrumental tessitura: the instrument’s typical range of sounding pitches. Finally, a fourth study tested whether melodic ornaments in Baroque keyboard music tend to occur in higher musical voices.

To anticipate our results, we will report empirical evidence consistent with a pitch-speed relationship in composition and performance, in which higher music does indeed tend to be faster. Furthermore, it seems that this relationship is most evident when the experimental method preserves the notion of musical ‘line.’

5.2 Study 1: Notewise Relationship Between Pitch Height and Duration

5.2.1 Methods

If there is a relationship between pitch height and musical speed, then one might predict that high-pitched notes should typically exhibit shorter durations than low-pitched notes, regardless of instrument or musical part. To test this hypothesis, we queried six existing databases of notated western art music, representing contrasting styles and periods. Specifically, our convenience sample included 34 assorted vocal motets by English composers Leonel Power (c. 1370–1445), John Dunstable (c. 1390–1453), and Thomas Morley (1557–1602), 21 movements from the Brandenburg concertos by Johann Sebastian Bach (1685–1750), 24 movements from orchestral symphonies 99–104 by Franz Joseph Haydn (1732–1809), 15 movements from the first, third, fifth, and seventh symphonies by Ludwig van Beethoven (1770–1827), 24 piano preludes by Frédéric Chopin (1810–1849), and 44 piano rags by Scott Joplin (1867–1917). These works span
nearly six hundred years of musical history, but were not chosen to be a stratified sample representative of all western music. We therefore analyzed each subsample individually.

For a given score, each note was characterized according to its pitch distance in semitones from middle C (C4) and its notated duration (measured in quarter durations). For example, a half note G4 would be coded as having the pitch +7 and duration 2.0. Any transposed orchestra parts were restored to their actual sounding pitch before processing, and piano sustain pedal markings were appropriately expanded. Examples of this pitch-duration representation are shown in Figure 5.2.

To determine whether long notes tend to be low-pitched, we calculated the sample correlation $r$ between pitch height and note duration for each individual score, providing a measure of their linear relationship. Positive (+) pitch-duration correlations indicate that long durations are associated with higher pitches. By contrast, the hypothesized pitch-speed relationship predicts that long durations (i.e., slower music) would be associated with lower pitches. Therefore, negative (−) correlations would be consistent with the pitch-speed hypothesis. To address the aforementioned tempo problem of score-based studies, we made only intra-opus comparisons of pitch and duration, computing a separate correlation for each individual piece.

Statistical inference was carried out on the correlations via binomial sign tests. Although composers might indicate explicit tempo and meter changes within a given work, statistical inference based on this test should not be affected, provided the marginal distribution of pitches does not change and the direction and location of tempo changes is not systematically biased.

5.2.2 Results

Figure 5.3 depicts the distribution of notewise pitch-duration correlations for each subsample. While the English motets, Bach concertos, and Joplin rags tend to show negative correlations between pitch height and note duration, the Haydn symphonies,
Figure 5.2: Note Distributions for Four Pieces

Distribution of notated pitch and durations for notes in four sample works used in Study 1. For pitches, the various octaves of C are marked, with C4 corresponding to “middle C.” Note durations are given in notated quarter durations. Circular glyphs are scaled to reflect the frequency of occurrence at each point, and sample correlations are provided. Note the marked difference in scale for note duration between the choral/symphonic works and the piano works.
Figure 5.3: Notewise Correlations between Pitch Height and Note Duration

Boxed regions correspond to the 25th through 75th percentiles for each corpus, and the dark bars represent the median correlations per sample. Outliers are plotted as points if they fall outside $\frac{3}{2}$ the interquartile range. Whiskers represent maximum and minimum nonoutlying values, and are not confidence intervals. Composer names refer only to the sample studied and are not representative of their entire output. Starred (*) samples exhibit statistical significance in binomial sign tests (Bonferroni-adjusted $\alpha = .008$).
Table 5.1: Notewise Correlations between Pitch Height and Note Duration.

<table>
<thead>
<tr>
<th>Sample</th>
<th>Pieces (Total Notes)</th>
<th>Mean</th>
<th>(+)</th>
<th>(−)</th>
<th>Binomial p</th>
</tr>
</thead>
<tbody>
<tr>
<td>English Motets</td>
<td>34 motets (28,013 notes)</td>
<td>−0.172</td>
<td>2</td>
<td>32</td>
<td>p &lt; .0001 **</td>
</tr>
<tr>
<td>Bach Concertos</td>
<td>21 mvts. (103,780 notes)</td>
<td>−0.093</td>
<td>4</td>
<td>17</td>
<td>p = .004 *</td>
</tr>
<tr>
<td>Haydn Symph.</td>
<td>24 mvts. (157,808 notes)</td>
<td>−0.003</td>
<td>12</td>
<td>12</td>
<td>p = .58</td>
</tr>
<tr>
<td>Beethoven Symph.</td>
<td>14 mvts. (169,576 notes)</td>
<td>0.021</td>
<td>10</td>
<td>4</td>
<td>p = .97</td>
</tr>
<tr>
<td>Chopin Preludes</td>
<td>24 preludes (18,871 notes)</td>
<td>−0.059</td>
<td>11</td>
<td>13</td>
<td>p = .72</td>
</tr>
<tr>
<td>Joplin Rags</td>
<td>44 rags (64,915 notes)</td>
<td>−0.105</td>
<td>8</td>
<td>36</td>
<td>p &lt; .0001 **</td>
</tr>
</tbody>
</table>

Counts of positive and negative correlations between pitch height and note duration, calculated note by note for individual pieces of music. The average correlation across all works in a given corpus is provided as a summary statistic, using the Fisher $z$-transform (see Silver & Dunlap, 1987). If higher notes tend to be faster (i.e., shorter), one would expect more negative correlations than positive correlations to appear. The number of pieces exhibiting positive and negative correlations in each corpus are given, along with one-tailed $p$-values based on the binomial sign test. Starred (*) $p$-values indicate statistical significance (adjusted $\alpha < .008$); double-stars (**) indicate stronger evidence ($p < .0001$).

Beethoven symphonies, and Chopin piano preludes do not. There are some distributional differences between the subsamples as well: the Chopin preludes in particular exhibit a much broader range of possible correlations.

Table 5.1 summarizes the correlations between a notated pitch height and duration for compositions in all six musical subsamples. For each genre, the number of works exhibiting positive and negative correlations is tabulated, where negative correlations are consistent with the hypothesis that higher notes tend to be shorter. A rough indication of the strength of the association within each sample is provided by an average correlation.² Every subsample except the Beethoven symphonies skews in the predicted direction.

²Correlations fall between -1.0 and 1.0; average correlations were calculated under Fisher’s $z$ transformation (Fisher, 1921; Silver & Dunlap, 1987).
One-sided Fisher sign tests were accordingly applied to each corpus’ pitch-duration correlations. Using the Bonferroni correction for multiple tests, $p$-values below $\alpha = \frac{.05}{6} = .008$ were considered statistically significant. By this criterion, we conclude that English motets ($B(34) = 32, p < .0001$), Bach concertos ($B(21) = 17, p = .004$), and Joplin rags ($B(44) = 36, p < .0001$) tend to have longer notes at low pitch and higher notes at high pitch, consistent with the pitch-speed hypothesis. However, the results for the Haydn symphonies, Beethoven symphonies, and Chopin preludes are statistically nonsignificant.

5.2.3 Discussion

Based on the observed correlations between pitch height and note duration, it appears that certain types of music do indeed exhibit faster ‘notewise’ speeds in higher pitch registers in the sense that higher notes tend to be shorter. Nonetheless, this notewise pitch-speed relationship does not seem to generalize across all musical styles and composers. In the case of the symphonies and piano preludes, other (perhaps culturally-derived) compositional determinants might exert more powerful organizational effects.

On the other hand, extant pitch-speed relationships might have been obscured by one or more confounds, such as expressive timing deviations, tempo changes, or instrumentation effects. Notably, the strongest pitch-speed relationships were found in vocal motets and piano rags, two styles characterized by having short length, steady tempos, little expressive modification by performers, and uniform instrumentation. By contrast, symphonies and concertos tend to be longer pieces punctuated by many different textures involving several instrument types. Moreover, the works of Chopin are commonly associated with the use of tempo rubato, making accurate interpretation of notated durations especially difficult. Although these score-based results are compelling, additional

---

3An earlier version of this study gave slightly different results due to three differences in the analytical methodology. In the present work, (1) notes in instrumental transpositions have been restored to concert pitch, (2) double-stops, split spines, and chords have been incorporated, and (3) sustain pedal markings in piano scores have been expanded (cf. Broze & Huron, 2012).
studies of actual performances could provide a more complete understanding of notewise relationships between pitch and duration.

It could also be argued that the correlations used here cannot adequately capture all types of perceptible musical pitch-speed relationships. Perhaps note duration alone cannot always measure musical or melodic speed without additional contributions from inter-note pitch intervals or time intervals. In other words, musical speed might emerge primarily when successive tones are understood as comprising musical ‘lines.’ If this is true, then a sufficiently sensitive measurement of speed would require integration across multiple notes. To explore this possibility, our second study tests for the presence of the hypothesized pitch-speed relationship insofar as it is mediated through musical organization into parts.

5.3 Study 2: Partwise Distribution of Musical Activity

In western music, it is common to organize music according to musical ‘parts’ or ‘voices.’ These are typically distinguished by the person, instrument, or group of instruments performing the part. For example, a choral work might have four separate musical parts—soprano, alto, tenor, and bass—and each would be performed by a different group of musicians. Musical parts can also be understood in a more general sense as being distinct melodies occupying a certain region in pitch space. In this way, music written for a polyphonic instrument such as the organ might also be organized into parts, despite there being only a single performer. It is possible that pitch-speed relationships in musical practice arise and are perceived according to such partwise organization. In order to test for a partwise pitch-speed correspondence in compositions, we formulated a second testable prediction: if compositions tend to be organized such that higher music is faster, then one would predict that higher musical parts should contain more notes than lower musical parts—that is, higher parts should be more ‘active’ than lower parts.

Measurements of the effect of musical part per se might be easily confounded by instrumental effects. In a given ensemble, it is common for particular instruments to carry
certain musical parts. For instance, in a woodwind quintet, the bass is generally assigned to the bassoon, with the treble voices assigned to the flute and/or oboe. Different instruments tend to exhibit different acoustic and mechanical properties, possibly affecting the general speed with which an instrument can be played. For example, one might expect that a valve trombone can be played faster than a slide trombone. It is therefore possible that any observed differences in speed across parts might actually arise from properties of an instrument rather than of the musical part’s position in the texture. In order to minimize instrumental confounds, we sampled only those musical works which employ similar instruments on all voices or parts.

Using a previously-encoded database, we selected three groups of suitable scores based on the dual criteria of instrumental similarity within the compositions and clear part-based organization. Specifically, our convenience sample consisted of 37 four-part keyboard fugues by Johann Sebastian Bach, 100 six-part mass movements by Giovanni Palestrina (c. 1525–1594), and 100 string quartet movements primarily by Mozart and Beethoven, with additional quartets by Schubert, Mendelssohn, and Brahms. In the case of the Bach fugues, additional voices occasionally entered the texture toward the end of the piece; this material was excluded from analysis.

5.3.1 Methods

We first verified that for the Bach fugues, the Palestrina masses, and the string quartets, the ordinal positions of the musical parts in the score does indeed reflect the actual pitch heights of the parts. In order to describe the relative pitch heights of musical parts, it is useful to employ the musical term tessitura, which refers to the typical pitch range of a melody, passage, or musical part. For each piece, the tessitura of each musical part was measured by averaging the pitch heights of every note, measured in semitones from middle C (C4). We found that these average note heights do correspond to the nominal ordering of the parts; averages across each subsample are reported in Table 5.2.
In each subsample, we measured musical activity by counting the notes in each musical part across all works in the sample. These totals were then expressed as normalized proportions to aid interpretation. For each part, the note tally was divided by the total number of events in all parts, and subsequently multiplied by the number of parts in the sample. Using this scheme, the expected ratio of note counts for each part would be 1.0 in the absence of any pitch-speed effects. Values greater than 1.0 indicate parts that are especially active relative to the other parts. Conversely, values below 1.0 indicate parts that are more sparse in their activity.

We also tested whether the average pitch distance separating musical parts predicts their relative activity levels better than their ordinal positions alone. In other words, if two parts have a larger pitch separation in one piece than in another, will they show a greater difference in their note tallies? To this end, we computed the difference in mean pitch height between outer voices for each piece as an index of tessitural spread, and measured the difference in outer part note numbers, normalized to the total number of notes in the piece. A positive correlation between these two values would mean parts spaced more widely are more likely to differ in activity than narrowly-spaced outer voices.

5.3.2 Results

Table 5.2 reports tessituras (i.e., average pitch heights), note tallies, ratios of note distribution, and statistical results for the Bach fugues, Palestrina mass movements, and string quartet movements. For all three samples, higher voices tend to have more notes. This pattern is consistent with the hypothesis that higher voices show more activity than do lower voices. Chi-squared tests for significance were performed between each adjacent pair of voices.\(^4\) With eleven comparisons, we took marginal tests to be statistically significant when their \(p\)-values were below \(0.05/11 = 0.004\). All pairwise comparisons exhibited

---

\(^4\)The chi-squared test is here used as a large-sample approximation to the binomial test.
Table 5.2: Note Distributions Between Musical Parts

### Bach Fugues (4 parts, 37 fugues)

<table>
<thead>
<tr>
<th>Part</th>
<th>Tessitura</th>
<th>Notes</th>
<th>Ratio</th>
<th>$\chi^2(1)$</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Soprano</td>
<td>12.91</td>
<td>14,850</td>
<td>1.08</td>
<td>2.55</td>
<td>$p = .11$</td>
</tr>
<tr>
<td>Alto</td>
<td>6.40</td>
<td>14,576</td>
<td>1.06</td>
<td>9.63</td>
<td>$p &lt; .002^*$</td>
</tr>
<tr>
<td>Tenor</td>
<td>-1.70</td>
<td>14,051</td>
<td>1.02</td>
<td>206</td>
<td>$p &lt; .001^*$</td>
</tr>
<tr>
<td>Bass</td>
<td>-9.65</td>
<td>11,745</td>
<td>0.85</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Palestrina Masses (6 parts, 101 movements)

<table>
<thead>
<tr>
<th>Part</th>
<th>Tessitura</th>
<th>Notes</th>
<th>Ratio</th>
<th>$\chi^2(1)$</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cantus I</td>
<td>11.64</td>
<td>16,440</td>
<td>1.08</td>
<td>.004</td>
<td>$p = .95$</td>
</tr>
<tr>
<td>Cantus II</td>
<td>7.21</td>
<td>16,429</td>
<td>1.08</td>
<td>3.79</td>
<td>$p = .05$</td>
</tr>
<tr>
<td>Altus</td>
<td>4.32</td>
<td>16,784</td>
<td>1.10</td>
<td>29.8</td>
<td>$p &lt; .001^*$</td>
</tr>
<tr>
<td>Tenor</td>
<td>0.53</td>
<td>15,799</td>
<td>1.04</td>
<td>168</td>
<td>$p &lt; .001^*$</td>
</tr>
<tr>
<td>Baritone</td>
<td>-1.40</td>
<td>13,578</td>
<td>0.89</td>
<td>81.6</td>
<td>$p &lt; .001^*$</td>
</tr>
<tr>
<td>Bassus</td>
<td>-7.20</td>
<td>12,130</td>
<td>0.79</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### String Quartets (4 parts, 100 quartets)

<table>
<thead>
<tr>
<th>Part</th>
<th>Tessitura</th>
<th>Notes</th>
<th>Ratio</th>
<th>$\chi^2(1)$</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Violin 1</td>
<td>14.80</td>
<td>60,989</td>
<td>1.232</td>
<td>955</td>
<td>$p &lt; .001^*$</td>
</tr>
<tr>
<td>Violin 2</td>
<td>7.61</td>
<td>50,661</td>
<td>1.023</td>
<td>197</td>
<td>$p &lt; .001^*$</td>
</tr>
<tr>
<td>Viola</td>
<td>1.38</td>
<td>46,296</td>
<td>0.935</td>
<td>441</td>
<td>$p &lt; .001^*$</td>
</tr>
<tr>
<td>Cello</td>
<td>-8.65</td>
<td>40,124</td>
<td>0.810</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Tessituras correspond to the average pitch height of each musical part, calculated for each piece, and averaged across all pieces in the corpus with each piece weighted equally. Reported note counts are summed across all pieces in the corresponding corpus; a normalized ratio of note counts is supplied to facilitate comparison. If notes were evenly distributed across parts, a ratio of 1.00 would be expected. Chi-squared tests for statistical significance are shown for each adjacent pair of voices. A starred (*) p-value indicates a statistically significant difference in note density compared with the musical voice immediately below. With the exception of the uppermost parts in the Palestrina masses and Bach fugues, all other neighboring pairs of voices display activity differences consistent with the hypothesis.
Figure 5.4: Relative Note Distributions by Average Part Tessitura

Relationship between the relative number of notes in a part compared with the part’s average tessitura, indexed by mean pitch height within a part with each musical work per corpus weighted equally. Relative notes per part was calculated by tallying the notes found in each musical part across all pieces in the sample. Counts were then divided the total number of notes in a sample, and multiplied by the number of parts. Using this normalization scheme, values of 1.00 would be produced by an equal distribution of notes across parts. While the string quartets show a gradual increase in note frequency from lowest to highest part, the Palestrina masses and Bach fugues show more even distributions among the top voices.
statistical significance except for three: the Soprano/Alto pairing in the Bach Fugues, and the Cantus 1 / Cantus 2 and Cantus 2 / Altus pairings in the Palestrina Masses. Based on these results, one would conclude that higher musical parts do indeed show more note activity.

A visual representation of these data is given in Figure 5.4, where each corpus’ distribution of notes across parts is plotted against each part’s tessitura. All three genres of music display some form of the hypothesized pitch-speed relationship. For the Bach fugues and Palestrina masses, the greatest differences in musical activity appear to involve the lowest voices, whereas the top voices display reasonable parity. By contrast, the string quartets have significantly more musical activity in the top voice than any other. That is, the pattern of partwise pitch-speed association appears to differ qualitatively between the string quartets and the other two samples.

If pitch-speed effects are mediated by actual pitch height as well as by a part’s ordinal position in a musical texture, then one would expect that musical voices spread across a large range of pitches should exhibit greater differences in speed than those with more similar tessituras. To test the effect of tessitural ‘spread’ on the relative activity of the outer voices, we employed Pearson’s correlation $r$. The Bach fugues ($r(35) = .110$, $p = 0.52$) and string quartets ($r(98) = .087$, $p = 0.39$) showed no detectable effect, while the Palestrina masses showed somewhat higher linear dependence ($r(99) = .196$, $p = .05$). While all three correlations were positive, none reached statistical significance. Hence, it would appear that speed depends more on ordinal position in a texture than on absolute pitch height.

5.3.3 Discussion

From these results, one could conclude that organization of music into parts does predict regularities in note distribution. In particular, lower parts seem to be considerably less active than higher parts, with the lowest part typically being most active. This organizational pattern appears to relate directly to the ordinal position of the voices
in a texture instead of the actual pitch height of the musical parts. Part-based organization of music might thus be a more important determinant of compositional choices than pitch height \textit{per se}. It seems reasonable that listeners might learn to comprehend music in the same way.

While relative partwise activity might be perceptually salient, one could still object that the note-tallying measurement method does not necessarily reflect high musical speed: a string quartet movement in which the cello has bursts of rapid playing intermixed with long and frequent multimeasure rests could also result in a low note count. To address this possible confound, we repeated the partwise study using only those measures in which all musical voices participate, and compared them to the original results. This \textit{post hoc} rest-eliminating method produced similar results overall, but with somewhat reduced effect sizes. The complete sample results (representing net musical activity) and the rest-eliminating results (representing typical contribution to four-voice textures) had an average samplewise correlation of .821. We interpret this to indicate that while multimeasure rests do account for a measurable portion of partwise activity differences, there remains a substantial effect attributable to musical speed differences.

In this approach, musical ‘part’ was defined and tested in such a way as to allow contributions from several sources of variance apart from simple ordinal position in a musical texture. For example, the observed difference in activity between string quartet cello and viola parts might reflect their different instrumentation or their different musical purposes: cellos often perform basslines and violas often perform accompaniment figures. Similarly, the difference between the first and second violin parts could reflect the tendency for melodies to appear in the top musical voice. In all, the data are consistent with the presence of partwise pitch-speed effects, even if the precise causes remain to be fully explored.
5.4 Study 3: Effect of Instrumentation

The first ‘notewise’ study indicated that western music of some styles exhibits a systematic relationship in which higher notes are shorter; the second ‘partwise’ study similarly suggests that higher musical parts display more musical activity. Study three focuses specifically on the question of whether instrumentation might play a role in mediating these effects, testing the hypothesis that lower-pitched instruments generally exhibit slower musical speeds. In pursuit of evidence converging with that of score-based methods, actual performed musical speeds were measured for several common western instruments using a sample of recorded performances.

The pitch-speed hypothesis suggests that low-pitched instruments typically perform slower music than high-pitched instruments. Based on the results of the second study, one might reasonably expect that instrumental effects in ensemble music would be difficult to distinguish from effects due either the ordinal position of a part, or its typical melodic, harmonic, or supportive role. In an effort to treat each instrument as equitably as possible, we studied instrumental performance in the context of instrumental solo repertoire. In such music, a single featured instrument typically performs the primary melodic content of the piece, while other musicians provide supportive accompaniment. Because music played by featured soloists would presumably be subject to different compositional constraints than more typical ensemble music, this repertoire could provide a perspective on pitch-speed relationships complimentary to those of the notewise and partwise studies. Specifically, we reasoned that solo repertoire should be less likely to reflect the compositional determinants of typical part-based ensemble writing, and more likely to reflect what is idiomatic to the instrument itself.

5.4.1 Methods

Sixteen common monophonic orchestral instruments representing string, brass, and woodwind families were identified (Table 5.3). For each, the nominal highest and lowest
sounding pitches are listed as reported in Samuel Adler’s *The Study of Orchestration, 3rd Edition* (2002). The midrange pitch was used to characterize the instrument’s tessitura.\(^5\)

In order to assemble a representative sample of solo literature, we used the Naxos Music Library’s website for access to recordings (Naxos, 2011). Naxos is a major record label known for its extensive catalogue of recorded western art music. At the time of this study, the catalogue was reported to contain 880,711 individual tracks. A sample of solo recordings for each target instrument was assembled by using the search string “[instrument name] recital.” We selected the first four albums returned by the search for each of the 16 target instruments, which were typically collections of works for solo instrument, usually with piano accompaniment, sometimes with guitar or harp, and occasionally with a larger instrumental ensemble.

From each album, we randomly selected three tracks, subject to certain constraints. If any track appeared to have two or more soloists, or was a recording of a piece already sampled, it was discarded. For the euphonium and piccolo, four albums could not be identified, so the twelve tracks were drawn from only three unique albums. Similarly, the bass trombone proved to be an uncommon instrument in recital recordings; six tracks each were selected from the two available albums.

For each track, the soloist’s initial speed was measured in notes per second by counting the number of notes played in the first 10 seconds following the soloist’s entry. Trills were counted as multiple notes. For each instrument, speeds were calculated for each of the twelve tracks, and the median speed was taken to represent an instrument’s probable melodic speed.
Table 5.3: Instruments, Ranges, and Observed Speeds

<table>
<thead>
<tr>
<th>Instrument</th>
<th>Range</th>
<th>Midrange (from C4)</th>
<th>Median Speed (notes/sec)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Strings</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Violin</td>
<td>G4 - B7</td>
<td>25.5</td>
<td>2.60</td>
</tr>
<tr>
<td>Viola</td>
<td>C3 - A6</td>
<td>10.5</td>
<td>1.55</td>
</tr>
<tr>
<td>Cello</td>
<td>C2 - E6</td>
<td>2.0</td>
<td>2.55</td>
</tr>
<tr>
<td>Double Bass</td>
<td>C1 - G4</td>
<td>−14.5</td>
<td>1.70</td>
</tr>
<tr>
<td><strong>Woodwinds</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Piccolo</td>
<td>D5 - C8</td>
<td>31.0</td>
<td>3.35</td>
</tr>
<tr>
<td>Flute</td>
<td>C4 - D7</td>
<td>19.0</td>
<td>2.00</td>
</tr>
<tr>
<td>Oboe</td>
<td>B♭3 - A6</td>
<td>15.5</td>
<td>3.10</td>
</tr>
<tr>
<td>Clarinet (A)</td>
<td>C♯3 - F♯6</td>
<td>9.5</td>
<td>3.05</td>
</tr>
<tr>
<td>Alto Saxophone</td>
<td>D♭3 - B♭5</td>
<td>5.5</td>
<td>2.40</td>
</tr>
<tr>
<td>Bassoon</td>
<td>B♭1 - E♭5</td>
<td>−5.5</td>
<td>1.95</td>
</tr>
<tr>
<td><strong>Brass</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trumpet (C)</td>
<td>F♯3 - C6</td>
<td>9.0</td>
<td>2.55</td>
</tr>
<tr>
<td>Tenor Trombone</td>
<td>E2 - F5</td>
<td>−1.5</td>
<td>2.20</td>
</tr>
<tr>
<td>Horn</td>
<td>B1 - F5</td>
<td>−4.0</td>
<td>1.60</td>
</tr>
<tr>
<td>Bass Trombone</td>
<td>B♭1 - B♭4</td>
<td>−8.0</td>
<td>1.65</td>
</tr>
<tr>
<td>Euphonium</td>
<td>G♭1 - B♭4</td>
<td>−10</td>
<td>1.40</td>
</tr>
<tr>
<td>Tuba</td>
<td>B0 - G4</td>
<td>−15</td>
<td>1.20</td>
</tr>
</tbody>
</table>

Instruments and ranges adapted from Adler (2002, Appendix A). Range is as reported for professional-caliber orchestras. Midrange indicates the note whose pitch lies in the middle of the range, measured in semitones from middle C (C4). Speeds are measured in notes per second for the first 10 seconds of recital recordings.
Figure 5.5: Instrumental Tessitura and Melodic Speed for 16 Solo Instruments

Tessitura is indexed by professional ranges adapted from Adler (2002, Appendix A). Median melodic speed was calculated from the first 10 seconds of twelve recorded performances per instrument (Naxos, 2011). The Spearman rank correlation between the two is $r_s = .736$, suggesting a strong relationship between instrumental tessitura and performed melodic speed.
5.4.2 Results

Spearman’s rank correlation $r_s$ was used to measure the relationship between median melodic speed and instrumental tessituta. The correlation between instrumental midrange and median melodic speed was quite high ($r_s = .736$, $p < .002$), consistent with the hypothesis that high-range instruments tend to play faster than instruments with low range. Figure 5.5 summarizes this relationship graphically, showing a clear positive association between instrumental tessitura and typical musical speed. For example, the piccolo exhibits both the highest midrange pitch and the fastest melodic speed, whereas the tuba exhibits the lowest midrange pitch and the slowest speed.

It would seem that even when instruments are performing with the same nominal musical role—that of featured soloist—musical pitch-speed relationships appear. Because these results depend on sound recordings instead of computerized scores, the instrumental pitch-speed relationship identified here presumably reflects some combination of compositional decisions and performance practices. Further experimental work could elucidate the relationship between composer and performer choice, perhaps by asking several performers to perform music without tempo indications, or asking several composers to write music appropriate for different instruments.

5.5 Study 4: Trills and Ornaments

While carrying out the third study using instrumental recordings, it became clear that a great deal of rapid melodic activity is associated with ornaments such as trills. Inspired by this observation, we carried out a brief fourth study to examine the distribution of melodic ornaments in relation to different pitch levels.

Melodic ornaments include trills, mordants, turns, and other such musical embellishment figures that typically involve rapid sequences of pitches. Ornaments are often

---

5In the case of string instruments (violin, viola, cello, double bass), this midrange pitch and the observed mean pitch heights from Study 2, have a Pearson correlation of $r = .997$, indicating that the calculated midrange bears a strong relation to an instrument’s practical tessitura.
abbreviated using purpose-specific symbols placed above a given note in a score. The presence of an ornament therefore indicates a sequence of notes more rapid than the surrounding music. If a pitch-speed effect is present, one might expect higher musical parts to exhibit more ornaments than lower musical parts. Our test of this hypothesis focuses on Baroque music, where ornaments are a relatively common feature.

Accordingly, we used a convenience sample of fifteen two-part inventions and fifteen three-part sinfonie by J. S. Bach, predicting that there would be more ornaments in the upper part of the two-part inventions and in the highest of the three parts in the sinfonie. Results are given in Table 5.4. Substantially more ornaments were notated in the uppermost voice, a result consistent with the pitch-speed hypothesis. In all, we identified 150 soprano-voice ornaments and only 61 bass-voice ornaments, a statistically significant difference ($\chi^2(1, N = 211) = 37.5, p < .001$).

### 5.6 General Discussion

In all, our results are consistent with the idea that high musical speeds and high musical pitch tend to coincide in compositional and performance practice. Moreover, this correspondence seems to be instantiated in many different ways. In summary, the first ‘notewise’ study found that low notes tend to be longer in certain styles. The second ‘partwise’ study suggested that the distribution of musical activity favors parts with higher ordinal positions in the texture. Our third study directly measured instrumental performance speeds in notes per second, finding that typical performance speeds are highly correlated with typical instrumental tessituras. Finally, a fourth study indicated that fast-paced melodic ornaments tend to occur in the uppermost voice. Although pitch-speed interaction effects might bias one’s perception of musical speeds, the intuitive notion that higher music is faster appears to be largely substantiated.
Table 5.4: Ornament Distribution Between Musical Parts

<table>
<thead>
<tr>
<th>Part</th>
<th>Ornaments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Soprano</td>
<td>123</td>
</tr>
<tr>
<td>Bass</td>
<td>58</td>
</tr>
</tbody>
</table>

Bach Inventions
(2 parts, 15 Inventions)

<table>
<thead>
<tr>
<th>Part</th>
<th>Ornaments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Soprano</td>
<td>27</td>
</tr>
<tr>
<td>Alto</td>
<td>19</td>
</tr>
<tr>
<td>Bass</td>
<td>3</td>
</tr>
</tbody>
</table>

Bach Sinfoniae
(3 parts, 15 Sinfonia)

Distribution of notated ornaments in J. S. Bach’s Inventions and Sinfoniae. The occurrence of these rapid trills and mordants is more frequent in the highest musical voice ($\chi^2(1, N = 211) = 37.5, p < .001$).

5.6.1 Musical Speed and Musical Line

Our first ‘notewise’ correlational study indicated a musical pitch-speed relationship for some of the genres studied, but not all. On one hand, the hypothesized association might simply be absent in these subsamples. An alternate interpretation would posit that the hypothesized relationship does exist, but the notewise operationalization of musical speed is too crude to detect it in some musical styles. To help determine which account is correct, it would be useful to directly compare the performance of the notewise and partwise measurements when used on the same musical sample.
In a post hoc test, the notewise method of the first study was applied to the pieces sampled for the second study. For each score, pitch-duration correlations were calculated. The string quartets exhibited 7 positive correlations and 93 negative correlations—a striking imbalance consistent with the pitch-speed hypothesis, and in agreement with the second study’s results. The Palestrina masses exhibited 30 positive and 71 negative correlations. While less pronounced, these results are also concordant with the second study’s result. However, applying the notewise method to the Bach fugues produced null results, with 18 positive and 19 negative correlations. These results stand in contrast to the partwise results of the second study, which identified statistically significant results for the fugue’s three lower voices.

Therefore, it seems that the first study’s notewise method might not have captured all types of pitch-speed organization: pitch-speed relationships in musical practice might depend more directly upon the activity of musical parts than upon isolated notes. Another way of saying this is that pitch-speed relationships become most obvious when tones are organized into melodic ‘lines.’ This is intuitively sensible from a perceptual point of view—after all, things which seem to have any speed at all must first somehow be perceived as being ‘objects.’ Indeed, the process through which listeners identify musical lines as independent auditory streams has been well-studied (see, e.g., Miller & Heise, 1950; Dowling, 1973; van Noorden, 1975; Bregman, 1990). In short, it seems that notes don’t have speed, but lines do.

5.6.2 Five Theoretical Accounts

Here, we offer several preliminary accounts for compositional pitch-speed organization, roughly organized into five categories: acoustic, kinesiologic, music theoretic, sensory, and perceptual/psychological. These explanations are not intended to be exhaustive, but to provide a basis for future work.
Acoustic. Perhaps the pitch-speed relationship originates in the physics of sound production on musical instruments, which in turn could influence compositional and performance practice. For example, informal observation suggests that larger, lower-pitched instruments tend to have longer initial attack transients before notes reach full amplitude. It is possible that such limitations would restrict a low-pitched instrument’s ability to play music effectively at rapid speeds. A straightforward investigation of acoustic limitations on performance speeds could test whether otherwise identical music tends to be performed more rapidly on instruments with shorter onset times. However, while acoustic explanations are attractive, it seems unlikely that they could be solely responsible for the particular pitch-speed relationships observed in the present investigation.

Kinesiologic. Pitch-speed relationships might arise from the human movements involved in vocal and instrumental performance. As lower-pitched instruments tend to be larger and heavier, they could place increased demands on a performer’s strength or agility. For instance, valves and fingerboards for large instruments tend to require more strength to operate than those of smaller instruments. It might be physically taxing to rapidly move slides or fingers the distances required in low pitch ranges.

Some musically-relevant motor limitations might be manifestations of Fitts’s Law, a model of human motion (Fitts, 1954). According to Fitts’s Law, muscular motion is subject to a tradeoff between simultaneous speed and accuracy: fast movements are unlikely to be very precise, especially when distances are long. Interestingly, melodic behavior does appear to be consistent with Fitts’s Law (Huron, 2001), although it remains to be seen whether this reflects motor constraints or perceptual ones. In general, it would seem that kinesiologically derived pitch-speed effects should become most evident when performing very challenging music. A direct test for physiological effects on musical

---

6There is evidence that such instrumental attack transients are of a musically meaningful duration. For example, Pickering (1986) found that on some strings, violins can take up to 300 ms or more before attaining a steady tone, corresponding to a common-time eighth note at 100bpm.

7We thank an anonymous reviewer for this suggestion.
speed might endeavor to develop a complete picture of the muscle movements involved in instrumental or vocal performance, and experimentally quantify the effects of each.

**Music Theoretic.** A music scholar might point to the particular history of western music’s development as giving rise to the observed relationship. An association between speed and pitch height would naturally result from the compositional practices of twelfth century Aquitainian vocal polyphony, in which faster voices were composed atop extant chant melodies (Yudkin, 1989). Music theoretic reasons also apply to homophonic compositions: here, bass notes might tend exhibit less embellishment than other voices due to their privileged status in western music, resulting in an overall slower speed. Cross-cultural studies could test whether the observed pitch-speed relationship is unique to the western musical tradition, or if it represents a more general pattern.

**Sensory.** The idiosyncratic design of the auditory system could account for pitch-speed relationships in music in several different ways. One explanation pertains to limitations in pitch perception: as tones become lower in frequency, pitch perception itself becomes increasingly difficult for short tones (Houghty & Garner, 1947; Pollack, 1968; Robinson & Patterson, 1995). Perhaps lower notes tend to be longer in order to accommodate this perceptual limitation.

Another account would note that uppermost voices in polyphonic textures tend to be the most perceptually salient (Huron, 1989b). Melodies, which are often faster than accompanying parts, would sensibly be featured by placing them in the highest voice (and perhaps by ornamenting them as well). While the fourth study on ornaments supports this explanation, the partwise study suggests that inter-part differences more frequently involve the lowest lines in a texture.

**Perceptual/Psychological.** There are two explanations related to the psychology of musical expectation that deserve discussion. The first is based on perceptual pitch-speed interaction effects of the sort described by Collier & Hubbard (2001) and Boltz (2011). The existence of these interactions indicates that listeners use information from multiple
perceptual domains to determine the most likely speed of music, which could be particularly helpful if information in one or another modality is noisy or unreliable. Based on pitch-speed interactions, one would expect the piccolo obbligato in *Stars and Stripes Forever* to sound subjectively faster when played on a piccolo than if it were played several octaves lower on a tuba. However, the existence of interaction effects cannot immediately explain why composers would choose to amplify the perceptual effect by writing music in which lower pitches move slower yet. One possibility would be that composers tend to write music following whatever pitch-speed correspondences they usually perceive. If this is true, then any musical tradition ought to slowly drift toward more pronounced pitch-speed relationships, as long as the perceptual effect persists.

Speed judgments might also be affected in a different way: one could argue that listeners should make speed judgments relative to context-dependent expectations. If listeners expect higher music to be faster in general, then shouldn’t they tend to evaluate higher music against this adjusted baseline? Surprisingly, this ‘expectation calibration’ argument appears to make a prediction directly opposite to that of interaction effects. In this view, a tuba performance of the *Stars and Stripes* obbligato would probably seem impressively fast by tuba standards, even if it were objectively slower than typical piccolo speeds. This argument is particularly attractive in that there are many ways composers and performers could exploit context-dependent musical expectations (see, e.g., Huron, 2006; Meyer, 1956).

Although the predicted effects of ‘perceptual interaction’ and ‘expectation calibration’ might appear contradictory, the two are not in fact mutually exclusive. One could interpret pitch-speed interactions as representing a strategy to maximize the chance of making a correct judgment (or producing an accurate percept) within a given contextual frame. For example, a collection of stimuli used in a perceptual experiment would implicitly define the contextual frame within which speed judgments are to be made. By contrast, the ‘expectation calibration’ argument would apply when a listener is in the process of choosing an appropriate contextual frame, such as when they watch a tubist
walking onto the stage. An experiment testing for the latter effect could not use a set
of predetermined stimuli, but would instead need to measure speed judgments within
listener-defined contexts.

Put another way, pitch-speed interactions might be most useful to determine what
happened, and expectation calibration most useful to predict what is likely to happen
next.

5.6.3 Future Directions

Given the diversity of plausible causes for pitch-speed correspondences, there is rea-
son to suppose that a complex network of interactions might underlie the observed re-
lationhip. For example, small perceptual effects may favor certain musical practices,
which are then amplified by cultural norms. Testing for a perceptual effect may there-
fore lead one to erroneously believe that the proximal cause is perceptual, when in fact
the effect is primarily a consequence of cultural inertia. Ultimately, progress may depend
on the development of causal models rather than testing single-cause conjectures.
Chapter 6: General Summary

We began by asking what makes music so uniquely motive and emotionally compelling. In response, it was argued that music exhibits perceptual animacy cues, indications that it emanates from a living source. Because it would be useful for any animal to distinguish living things from nonliving things, it is reasonable to suppose that humans might have a special perceptual capacity to do so. Both animate and inanimate sound sources tend to evoke the experience of an auditory streams. As musical analogs of auditory streams, musical lines may be necessary, though not necessarily sufficient, for the perception of musical animacy.

In addition to being perceived as animate, musical sounds might in fact be perceived as being distinctly human in origin and character. Three anthropomimetic properties of music might help to distinguish musical lines as animate rather than inanimate streams. By merit of being pitched sounds with a large degree of sonic harmonicity and appropriate pitch height, they strongly resemble vocally produced sounds. Because of their rhythmicity, they could preferentially activate perceptual systems dedicated to processing human movement. Finally, because musical lines exhibit organized acoustic patterns, they could suggest that an intelligent or linguistic creature might have produced them. By combining the essence of three prototypically human residues, music might act as a ‘supernormal stimulus’—one that hijacks our tendency to strongly respond to the sound of others of our own kind.
Empirical approaches were then used to investigate the influence of polyphonic voice multiplicity on musical emotion perception. Nominal voice multiplicity was distinguished from perceptual voice numerosity, which was itself investigated by employing a voice denumeration task. It was found that in polyphonic western compositions, one should expect textures with more polyphonic voices to sound less lonely, and to be associated with increased positive affect more generally. Some evidence was found that ‘personal emotions’ and ‘musical emotions’ could be somewhat distinct, at least in the context of experimental protocol. Further work would be needed to determine whether the multiplicity–positivity connection is indeed due to an implied social situation, or could be attributed to another cause, such as enjoyment of compositional complexity.

Finally, compositional organization according to musical line was found to mediate correspondences between melodic speed and pitch height. This relationship appears both when considering music part by part, as well as instrument by instrument. It would seem that composers treat musical lines as independent musical agents with different speeds and trajectories just as listeners are likely to perceive them as such. A complete account of the causal chain that has led to this practice might, however, be considerably complex and rely on as-yet uncollected experimental evidence from several sources.
Bibliography


Albrecht, J. D. (2012). *Affective Analysis of Music Using the Progressive Exposure Method: The influence of bottom-up features on perceived musical affect.* PhD thesis, Ohio State University, Columbus, OH.


122


