MUSIC ANALYSIS EXPERIENCE

New Perspectives in Musical Semiotics

Edited by
Costantino Maeder and Mark Reybrouck

Leuven University Press
# Table of Contents

**Preface**  
Costantino Maeder and Mark Reybrouck  
9

**Part One. Setting the Stage: Music-In-Action, Semiotics, and Intermediality**  
17

Music, Speech and Meaning in Interaction  
*Ian Cross*  
19

Capturing the Music: The Thin Line Between Mediation and Interference  
*Zelia Chueke and Isaac Chueke*  
31

Performativity Through(out) Media: Analyzing Popular Music Performance in the Age of Intermediality  
*Maurizio Corbella*  
43

**Part Two. Representation, Interpretation, and Meaning**  
59

Reading a Work of Music from the Perspective of Integral Interpretation  
*Mieczysław Tomaszewski*  
61

*La musique au second degré:* on Gérard Genette’s Theory of Transtextuality and its Musical Relevance  
*Paulo F. de Castro*  
83

Semiotic Narrativization Processes  
*Nicolas Marty*  
97

Musical Understanding: Wittgenstein, Ethics, and Aesthetics  
*Paulo C. Chagas*  
115

Where to Draw the Line? Representation in Intermedial Song Analysis  
*Lea Maria Lucas Wierød*  
135
Table of Contents

**Part Three. Experience, Cognition, and Affect** 149

The Ability of Tonality Recognition as One of Human-Specific Adaptations 151

*Piotr Podlipniak*

Musical Semiotics and Analogical Reference 167

*Lawrence M. Zbikowski*

The Other Semiotic Legacy of Charles Sanders Peirce: Ethology and Music-Related Emotion 185

*David Huron*

**Part Four. Intermediality and Transdisciplinarity** 209

The Death of Klinghoffer: From Stage to Screen 211

*Maarten Nellestijn*

The Perception of Art Songs Through Image: A Semiotic Approach 225

*Mônica Pedrosa de Pádua*

“What Kind of Genre Do You Think We Are?” Genre Theories, Genre Names and Classes within Music Intermedial Ecology 239

*Gabriele Marino*

**Part Five. Analysis and Beyond** 255

A Story or Not a Story? Pascal Dusapin’s Opera *Roméo & Juliette* and New Ways of Musical Narratives 257

*Małgorzata Pawłowska*

Authorship, Narrativity and Ideology: The Case of Lennon-McCartney 277

*Dario Martinelli*

The Sublime as a Topic in Beethoven’s Late Piano Sonatas 301

*Jamie Liddle*
Table of Contents

Melodic Forces and Agential Energies: An Integrative Approach to the Analysis and Expressive Interpretation of Tonal Melodies 315
Robert S. Hatten

The Embodiment of Yearning: Towards a Tripartite Theory of Musical Agency 331
Rebecca Thumpston

Editors 349

Contributors 351
Charles Sanders Peirce was a remarkable scholar with broad ranging interests in many areas of philosophy, logic, and metaphysics. The influence of Peirce’s ideas in the modern field of semiotics is well known. Less well known is the influence Peirce had on another discipline, the field of ethology (the study of animal behavior). Largely a European undertaking, the founders of ethology—notably Konrad Lorenz and Nico Tinbergen—were influenced by semiotic concepts in their studies of animal displays and calls. This presentation pays tribute to the achievements of ethology by returning insights from animal behavior back to the humanities and perhaps philosophy. Ethological concepts are used to provide a novel approach to understanding music-related emotion.

1. Signals and cues

Inspired by his reading of Peirce, Konrad Lorenz distinguished between two kinds of animal communications: signals and cues (Lorenz, 1939; 1970). This distinction remains a foundational concept in modern ethology (e.g., Smith and Harper, 2003). As defined by Lorenz, a signal is an evolved purposeful communicative behavior, such as evident in a rattlesnake’s rattle. The rattle makes use of a purpose-evolved anatomical device to communicate a specific message to the observer. A cue is a non-purposeful artefact that is nevertheless informative, such as the buzzing sound produced by a mosquito suggesting imminent attack. Both the rattling of the rattlesnake’s rattle and the buzzing of the mosquito presage the possibility of an attack. However, in the former case the communication is intentional whereas in the latter case it is an unintended consequence of the need for the insect to flap its wings. Signals
involve innate behavioral and physiological mechanisms, whereas cues are commonly learned artefactual behaviors.

2. Size matters

When two animals interact, many behaviors can be classified as either agonistic (e.g., aggression), or affinitive (e.g., submission, greeting, sharing). In general, when animals create threatening displays they behave in ways that tend to make them appear larger. This includes raised hair, ruffled feathers, standing upright, arching of the back, looming, and other behaviors that make the animal seem bigger. Conversely, when animals create submissive or friendly displays, they typically behave in ways that tend to make them appear smaller, such as bowing, squatting, sitting, head-lowering, limb-withdrawing, etc.

These visual behaviors have acoustic parallels. One of the best generalizations one can make about acoustics is that large masses or large volumes tend to produce low frequencies of vibration whereas small masses and small resonant cavities tend to produce higher frequencies. The ethologist Eugene Morton (1977) carried out a seminal comparative study of vocalizations in 28 avian and 28 mammalian species. Morton found that high pitch is associated with submissive and affinitive behaviors whereas low pitch is associated with threatening and aggressive behaviors.

Bolinger (1978) observed the same relationship in a cross-cultural sample of human speech intonation. In general, high vocal pitch is associated with appeasement, deference or politeness. Conversely, low pitch is associated with aggression or seriousness. This same pattern can be observed using musical stimuli. For example, Huron, Kinney and Precoda (2006) played unfamiliar Western folksongs to listeners and had them judge the melodies according to such criterion as politeness, heaviness, and aggressiveness. Unknown to the participants, each melody appeared in three different transpositions spanning two octaves. Consistent with ethological observations, it was found that transposing a melody to a higher pitch causes the melody to be judged more polite and more submissive (see also Morton, 2006).

In another study, Huron and Shanahan (2004) coded the degree of sociability for three hundred characters in randomly selected opera scenarios. Friendly and altruistic characters were rated as exhibiting high sociability whereas self-centred or aggressive characters were rated as exhibiting low sociability. Not surprisingly, there is a significant association between the tessitura of the voice and the character’s sociability: heroes are mainly tenors and sopranos, whereas villains are mainly basses and contraltos.
As with many mammals, the human voice can be broadly characterized as having two acoustical components: a source (vocal folds) and a filter (vocal tract). The frequency of the vocal folds is mainly determined by their mass and tension. The resonant frequency of the vocal tract is determined by the length and volume of the air cavity. In humans, the frequencies of the source and filter are under independent voluntary control. For example, we can produce a low pitch and low resonance when we utter a low vowel (e.g., [a]) with a low pitch. Speaking a high vowel (e.g., [i]) with elevated pitch produces a high pitch and a high resonance. However, we can also produce mixtures, such as speaking [i] with a low voice (low pitch + high resonance) or speaking [a] with a high intonation (high pitch + low resonance).

2.1 The smile

The linguist John Ohala extended Morton’s observations regarding sound-size symbolism to include the vocal filter (resonant frequency) as well as the vocal source (pitch). As with pitch, resonant frequency arising from body cavities is also correlated with body size (Ohala, 1980; 1982; 1983; 1984).

Ohala argued that sound-size symbolism can be used to account for the human smile (Ohala, 1982; 1994). For over a century, scholars have pondered the apparent enigma of the smile: why would showing one’s teeth (commonly associated with aggression) be construed as a sign of friendliness? Ohala drew attention to the fact that one can hear a smile (see also Tartter, 1980). Without seeing a person smiling, the smile is nonetheless evident in the sound of the voice. Flexing the zygomatic muscles characteristic of smiling causes the flesh of the lips to be drawn tight against the teeth. This effectively shortens the length of the vocal tract and so shifts the resonance of the voice upward. In short, the sound of the smile is the sound of a smaller resonant cavity. The upward shift of the spectral centroid is consistent with sound-size symbolism, which, throughout the animal kingdom, is a ubiquitous way of conveying friendly or non-aggressive intent. Accordingly, Ohala suggested that the smile originated as an acoustical display that later became generalized to include the visual component. Ohala proposed that the evolutionary origin of the smile is auditory, not visual.

2.2 The pout

Ohala extended his observations regarding the smile in the opposite direction. Instead of retracting the lips against the teeth, one can thrust the lips forward away from the teeth—lengthening the vocal tract with a characteristic drop
in the first formant frequency. Ohala refers to this as the “o-face”. An example is evident in the human “pout”. According to sound-size symbolism, this lowering of the frequency should be associated with anti-social rather than pro-social behavior. Indeed, the classic “brutish” or “loutish” voice involves extending the lips away from the teeth. The cliché sound of the aggressive hooligan offers a polar contrast with the sound of smiling—consistent with sound-size symbolism.

3. Multimodal signals

Are displays such as smiling or frowning ethological signals or ethological cues? Ethologists have identified a number of ways in which signals can be distinguished from cues. One property of signals is that they tend to exhibit **redundancy** where the signal is repeated or sustained over time and over multiple sensory channels (Wiley, 1983; Johnstone, 1997; Partan & Marler, 1999). Since signals are intended to be communicated, a “subtle” signal is unlikely to have the intended effect. Employing more than one sensory modality makes the signal more conspicuous. For example, in the case of the rattlesnake’s rattle, there is both a distinctive acoustical component (the sound of the rattle) as well as a distinctive visual component (the raised shaking tail). Ostensibly, even if an observer is only able to hear, or just see the snake, the signal could nevertheless be successful communicated. By contrast, many (though not all) cues do not exhibit multimodal features. This simply reflects the fact that cues are behavioral artefacts (like the buzzing of a mosquito’s wings), and not explicitly intended to be communicative. We have already seen an example of the tendency for multimodal displays in the case of the smile: Ohala’s main claim is that it would be wrong to regard the smile as solely a visual display.

In the past, emotion researchers have tended to focus on the visual aspects of facial expressions without considering other sensory modes. From an ethological perspective, we would expect many facial expressions to qualify as **signals**—and therefore tend to be accompanied by distinctive acoustical features, not just visual features. Facial expressions that are not accompanied by distinctive acoustical features are more likely to be artefactual cues rather than signals.

Notice that the smile and the pout displays involve only one frequency-related component of the voice—namely the filter or vocal-cavity component. Recall that we can independently manipulate the source or pitch of the voice. Once again, high pitch is associated with pro-social intent whereas low pitch is generally associated with anti-social or aggressive intent. In light
of the multimodal tendency of signals, if the pitch of the voice represents an authentic signal, then we ought to see a distinctive visual element that accompanies the higher/lower vocal pitch. Two studies are pertinent. The first study was carried out by Huron, Dahl and Johnson (2009; see also commentary by Ohala, 2009). We asked forty-four non-musician participants to sing neutral, high and low pitches while their faces were photographed. The high and low photographs were paired together and independent judges were asked to identify which face is friendlier. Photographs of high-pitch faces were easily perceived as friendlier than the low-pitch faces. A careful examination of the photographs revealed that, when singing a low pitch, participants tend to drop the chin, frown, and lower their eyebrows. Conversely, when singing a high pitch, participants tend to raise the chin, smile, and raise the eyebrows. In a follow-up experiment, we cropped the photographs so that only the region above the nose-tip was shown. Once again, independent judges found the high-pitch faces friendlier than the low-pitch faces. The eyebrows alone appear to provide a sufficient feature for judging the friendliness.

Producing a low pitch appears to have a causal relationship with eyebrow movement. That is, when asked to sing a low pitch, people naturally tend to lower their eyebrows. What about the reverse relationship? Does moving your eyebrows cause your voice to move up or down in pitch? A second study was carried out by Huron and Shanahan (2013). Thirty-one participants were asked to read aloud short sentences placing their eyebrows in a high, low, or neutral position. Eyebrow placement was found to have a significant (though small) effect on pitch height. That is, compared with neutral and low eyebrow placement, speaking with raised eyebrows causes the pitch to rise.

This relationship is consistent with existing research concerning eyebrow placement. Cross-culturally, low eyebrow placement tends to be symptomatic of aggression whereas high eyebrow placement is indicative of friendliness. Ethologist Irenäus Eibl-Eibesfeldt (1989) has noted that the eyebrow flash (quick up-and-down movement) is a common greeting for humans and that analogous displays are evidence in other primates.

In summary, there is a strong relationship between pitch height and eyebrow placement that appears to be bi-causal: moving the pitch tends to cause the eyebrows to move in tandem, and moving the eyebrows tends (to a lesser extent) to cause the pitch to move in tandem. This bi-directional causality suggests a shared or common source in the motor cortex, consistent with a single unified display. Moreover, the multi-modal connection is consistent with the existence of an ethological signal.

Once again, this apparent signalling system can be observed in music. Bonfiglioli, Caterina, Incasa and Baroni (2006) carried out a qualitative study
of facial expressions from video recordings of performing musicians. They found that when the musical texture involves predominantly low pitches there is a tendency for the instrumentalist to lower her/his eyebrows. Conversely, when the music involves predominantly high pitches, there is a tendency for the instrumentalist to raise his/her eyebrows.

3.1 Sarcasm

Another basic facial expression described in the literature is the so-called contempt or sneer facial expression (Ekman, 1972). The sneer is regarded as a variant of the disgust expression. Specifically, the disgust expression involves characteristic flexion of the levator labii superioris muscles that elevate the upper lip and the depressor septi muscle that constrict the nostrils. This is presumed to have originated in efforts to reduce the inhaling of offensive odours. The sneer display is essentially “one-half” of a disgust expression. That is, flexion occurs asymmetrically on one side of the face only. Unlike the disgust display, the sneer is regarded as a social display: As Ekman notes, the disgust response effectively says “I find this disgusting” whereas the sneer says “I find you disgusting.”

Plazak (2011) carried out a seminal study in which instrumentalists were asked to play various passages in a sarcastic fashion. Through sound alone, listeners were readily able to recognize musical sarcasm compared with other affective conditions. Acoustic analyses using speech-based methods showed the sarcastic renditions exhibited elevated nasality measures. That is, the instrumental sounds approached the nya nya timbre associated with vocal taunts characteristic of the sneer or contempt. Once again, from an ethological perspective, the contempt or sneer facial expression is correlated with a distinctive auditory effect (in this case nasalization). The facial expression and sound go hand-in-hand. And once again, the same auditory features can be observed in a musical context.

3.2 Cuteness

Ethologists also make use of the Peircean concept of an index. The index conveys information by virtue of its factual connection to an object. Amphibians, such as frogs, grow continuous throughout their lives. As a consequence, the size of a frog provides a good index of its longevity, and so a reasonable index of its capacity to survive. The pitch of a frog’s croak is directly related to its size, and females preferentially mate with male frogs that produce the lowest croak—that is, those male frogs who have survived the longest.
From infancy to adulthood, the human vocal tract similarly increases in length. At birth, the typical vocal tract length is roughly 8 centimetres. By adulthood, the length has doubled to roughly 16 centimetres in length (Vorperian et al., 2005).

Some years ago, we carried out an experiment where listeners judged the “cuteness” of a wide variety of sound-producing objects. Listeners are in broad agreement when judging the relative cuteness of sounds. We learned that the sound-producing objects judged “most cute” share two physical attributes. They all share a small resonant cavity activated by a small amount of energy. The dimensions of the most cute-sounding objects resemble the size of an infant’s vocal tract.

Of course infants are capable of generating rather loud sounds that are manifestly not cute. However, when cooing or gurgling, the small energies activating the small cavity produce sounds that readily evoke nurturing and protective behaviors—in short, they engender parenting behaviors. Imagine the dire consequences if human parents did not find their infants cute. Without this affective response, humans would cease to survive for long.

What is interesting is that the response generalizes beyond infant sounds. What a soprano recorder, ocarina, and music box share in common is a small resonant cavity activated by a small amount of energy. Moreover, in each case, the sounds are readily described as “cute”. We hear these sounds a vulnerable, innocent and helpless.

For the first year of life, the vocal tract grows rather slowly. After 12 months, the rate of growth increases notably (Vorperian et al., 2005). It is in the first year, that the vocal tract produces the cutest sounds—a period in which the neonate is most vulnerable, and most in need of parental attention. The infantile quality of the voice is echoed in characteristic visual features, including a round (rather than long) face, large eyes (relative to the head size), small nose, and small ears. These are the classic “baby face” features.

The musical use of cuteness is apparent in the recordings of the American popular singer from the 1920s, Helen Kane (whose voice became the model for the Betty Boop cartoon character). Kane had a diminutive stature, including a small head, and correspondingly short vocal tract. Her singing style is quintessentially “cute” sounding. Unlike the intentional lowering of vocal pitch, or the voluntary smiling resonance, the bulk of the vocal tract length is fixed by anatomy—and therefore a true index of infantile vulnerability in both the Peircean and ethological senses.
4. Faces and voices

Let’s pause and summarize. We have seen some evidence consistent with the multimodal tendencies of ethological signals. Specifically, we have observed that smiling exhibits both characteristic visual and characteristic auditory features, and that the auditory component is consistent with sound-size symbolism. We have also observed that pouting exhibits both characteristic visual and characteristic auditory features, and that the auditory component is consistent with sound-size symbolism.

We have also observed a close relationship between vocal pitch height and eyebrow position. Specifically, voluntary efforts to raise or lower the pitch of the voice produce involuntary tendencies to move the eyebrows in a parallel fashion. At the same time, voluntary efforts to raise or lower the eyebrows produce involuntary tendencies to move the pitch of the voice in a parallel fashion. This bi-causal relationship is consistent with multimodal redundancy whose purpose is to increase the conspicuousness of signals. Moreover, the auditory component of this display is consistent with the sound-size symbolism.

We have further observed an association between the sneer/contempt expression, where flexing the nose causes a distinctive visual expression accompanied by audible nasalization of the voice—consistent with the multimodal tendency of ethological signals. Finally, we have seen an example of an index in the form of auditory cuteness—with the acoustical characteristics of a short vocal tract linked with characteristic baby-face visual features. In each case, all of the acoustic-related features (with their corresponding affects) can be observed in musical contexts.

5. Sadness and grief

Charles Darwin (1872) made an important distinction between sadness and sorrow. Here we propose to use the terms sadness and grief instead. Sadness is an affective state characterized by low physiological arousal. When sad, a person typically exhibits slow heart rate, shallow respiration, slumped posture, loss of appetite, sleep, reduced engagement with the world, a tendency to avoid conversation (i.e., mute), and contemplation/reflection (thinking sad thoughts). Grief, by contrast, is an affective state characterized by high physiological arousal. When in a state of grief, a person typically exhibits fast heart rate, erratic respiration, flushed face, tears, nasal congestion, pharyngeal constriction, vocalizing (anything from quiet sobbing to loud wailing), and
ingressive vocalizing (sound production while inhaling). Sadness and grief are often interleaved together; that is, periods of psychic pain commonly involve alternating periods of (quiet) sadness and (louder) grief.

Both sadness and grief are associated with distinctive sounds. People who are sad or depressed typically speak with a (1) quieter voice, (2) slower speaking rate, (3) low pitch, (4) small pitch movement, (5) poor articulation, and (6) dark timbre (Kraepelin, 1899). These same features have been observed in nominally sad music. For example, music in the minor mode is quieter in dynamic level (Turner & Huron, 2008), exhibits a slower tempo (Post & Huron, 2009), is slightly lower in overall pitch (Huron, 2008), employs smaller average melodic intervals (Huron, 2008), involves more mumbled articulation, and makes use of darker timbres (Schutz, Huron, Keeton, & Loewer, 2008).

People who experience grief also exhibit characteristic vocalizations. Grief vocalizations can range from quiet moaning to loud wailing. The vocalizations are commonly high in pitch, exhibit gliding (often descending) pitch contours, sniffing, ingressive phonation (vocalizing while inhaling), punctuated exhaling, and involve pharyngealized voice (due to the constricted pharynx) (e.g. Fox, 2004). The constricted pharynx introduces vocal instability—producing distinctive alternation between modal and falsetto phonation (commonly called “cracking” or “breaking” voice). Breaking voice is perhaps the most telltale sound associated with grief.

Paul and Huron (2010) studied the role of “breaking” voice in music. Country music fans were recruited; they identified 31 instances of cracking or breaking voice in their record collections. Each identified song was paired with a matched (control) song from the same album sung by the same singer—a song that did not contain any instance of breaking voice. Lyrics were assembled for both the target and control songs. Without hearing the music, independent judges rated the lyrics for grief-related content. Breaking voice was found to correlate positively with grief-related lexical content in the lyrics.

Why, we might ask do voices break? In general, grief exhibits a highly distinctive set of physiological characteristics, including watery eyes, nasal congestion, constriction of the throat, and erratic breathing. When crying for an extended period, the face tends to become “puffy” with notable inflammation around the eyes. Oddly, researchers on crying have failed to notice that, in isolation, any medical doctor would diagnosis these symptoms as characteristic of a systemic allergic response. Moreover, inflammation—such as that seen in the face after a long bout of crying—is caused by histamines. These are the same histamines that cause an allergy sufferer to reach for a bottle of antihistamines.
Notice that the allergic response leads to characteristic visual (facial) features, and also leads to distinctive vocalizations through the accompanying pharyngeal constriction. In short, crying appears to borrow the systemic allergic response, leading to characteristic visual and auditory features consistent with an ethological signal. This sort of physiological “borrowing” is known as an *exaptation* (Gould & Vrba, 1982).

The features of crying are not limited to the effects of an allergic response. For example, allergy sufferers are not compelled to vocalize. In the case of crying, however, the tendency to vocalize is so strong that the vocal cords remain engaged even when inhaling. The ingressive phonation characteristic of crying is consistent with an innate compulsion to make a sound.

Notice that crying bears all the hallmarks of an ethological signal. Crying appears to commandeer the allergic response as an exaptation that produces both distinctive visual features as well as distinctive acoustical features. That is, grief entails multimodal elements congruent with the goal of conspicuousness.

If crying is a signal, what does it signal? Limitations of space preclude any detailed exposition here. Jeffrey Kottler has proposed that weeping is the human “surrender” signal (Kottler 1996; Kottler & Montgomery, 2001). Kottler has documented how weeping “turns off” aggression or argument and leads to sympathetic altruistic behaviors directed toward the person crying. For the person crying, assistance is purchase at the cost of a loss of social status. That is, crying parallels the submission/surrender displays found in many other social animals. Support for Kottler’s theory comes from the work of Gelstein et al. (2011) on the olfactory effects of tears. Tears were collected from women volunteers who had been induced to weep by watching a sad scene from a movie. For comparison purposes they also collected saline solution that was trickled down the women’s cheeks. Men were then asked to smell both the real and imitation tears. They couldn’t tell the difference: neither had any noticeable odour. Nevertheless, the real psychic tears produced a marked physiological effect: testosterone levels dropped significantly when the men were exposed to the real tears. In addition, other measures showed that sniffing the tears significantly impeded sexual arousal. The results of this study suggest that psychic tears contain a chemical pheromone—an odourless air-borne hormone that influences the behavior of others.

### 6. Sadness as cue

Recall that sad speech is associated with six acoustic features: quieter, slower, lower in pitch, more monotone, mumbling, and dark timbre. What, we might
ask, do all six features share in common? It turns out that all six features can be plausibly attributed to low physiological arousal. Low energy is associated with low epinephrine levels and low acetylcholine levels. Acetylcholine has a marked impact on muscle tone and reactivity. Specifically, low acetylcholine leads to weakness (flaccid muscle tone) and sluggishness (slow muscular reactivity). Slow muscle movement causes sluggish movement of the lips, tongue and chin. That is, the articulatory muscles move slower producing a slower rate of speaking as well as a more mumbled articulation. When the pulmonary muscles (involved in breathing) are relaxed, the subglottal air pressure drops, producing a quieter sound. Similarly less tense vocal folds result in a lower overall pitch. Slow movement of the cricothyroid muscle results in less responsive pitch movements, leading to smaller pitch intervals or a more monotone pitch inflection. Finally, the relaxed facial musculature includes weak zygomatic activity; there is no active smile, so the lips tend to pull away from the teeth resulting in a longer vocal tract, and consequently a darker timbre. In short, all of the features of “sad voice” can be plausibly regarded as artefacts of low physiological arousal.

An important observation to be made about sad voice is that people tend to be mute when sad: sad people don’t vocalize much. This contrasts with grief. Although crying can be done quietly, there is a strong compulsion to vocalize when crying. As we have seen, the compulsion to vocalize is so strong that weeping tends to engage the vocal folds even when inhaling—a rare phenomenon.

In continuing research we have been looking at the uniqueness of nominally sad facial expression and vocalization. Although the research is not complete, it appears that there is no distinctive or unique “sad” facial expression. A presumed “sad” face appears to be indistinguishable from a “sleepy” or “relaxed” face. The “glum” faces commonly observed on a public bus or train are often deemed sad. However, people thought to appear sad are often simply relaxed. We are currently carrying out an experiment to test whether listeners can distinguish between “sad” voice and “sleepy” voice. We predict that it is difficult or impossible for listeners to distinguish sadness from sleepiness.

Summarizing, we might contrast sadness with grief as follows:
1. Unlike grief, sadness is not associated with a compulsion to vocalize.
2. Unlike grief, sadness does not appear to exhibit a clearly unique facial expression.
3. All of the characteristics of sad speech can be attributed to low physiological arousal—that is, they are artefacts of low energy.
4. Sad voice may not be distinguishable from sleepy voice (Shanahan and Huron, in progress).

In short, sadness looks like an ethological cue whereas grief looks like an ethological signal.

7. Acoustic ethological model

With this background, we might now address the question: How do we reconcile the seemingly contradiction claims that low pitch is associated with aggression and that low pitch is also associated with sadness? We have claimed that sadness is a covert affect. As a cue, sadness entails no overt expression. Nevertheless, observers learn to infer sadness through its association with low physiological arousal. Moreover, the low pitch is linked to other features arising from low physiological arousal, notably quiet voice. That is, the combination of low pitch and low intensity are likely to be interpreted by experienced listeners as indicative of sadness. Notice, however, that the acoustic features linked to sadness are the same as those associated with other states of low physiological arousal—including sleepiness and relaxation. This suggests that sleepiness, relaxation, and sadness share the same acoustic features, and should be easily confused with one another.

Aggression, by contrast, taps into the sound-size symbolism evident in calls throughout the animal kingdom. Accordingly, the association between low pitch and aggression or seriousness is likely to be a true ethological signal. If this is the case, then the link between low pitch and aggression ought to be biologically prepared—in contrast with sadness. Similarly, high pitch is also likely to be interpreted according to sound-size symbolism, and also likely to be an ethological signal.

Table 1 provides a summary of the theory presented here. We might refer to this as the Acoustic Ethological Model (AEM). This model can be regarded as a refinement of the model proposed by Morton (1977). Specifically, the AEM introduces a second dimension: adding intensity to pitch. Accordingly, the model distinguishes four acoustical conditions: (1) high pitch and high intensity is associated with fear or alarm, (2) high pitch and low intensity is associated with appeasement or friendliness, (3) low pitch and high intensity is associated with aggression or seriousness, and (4) low pitch and low intensity is associated with sadness, sleepiness, and relaxation. Three of the four conditions are candidate ethological signals, with one quadrant regarded as a candidate ethological cue.
The Other Semiotic Legacy Of Charles Sanders Peirce

<table>
<thead>
<tr>
<th>Quiet</th>
<th>Loud</th>
</tr>
</thead>
<tbody>
<tr>
<td>High pitch</td>
<td>appeasement, friendliness</td>
</tr>
<tr>
<td>Low pitch</td>
<td>sad, relaxed, sleepy</td>
</tr>
<tr>
<td></td>
<td>fear, alarm</td>
</tr>
<tr>
<td></td>
<td>aggression, seriousness</td>
</tr>
</tbody>
</table>

Table 1. Acoustic Ethological Model

8. To signal or not to signal

In ethology, the purpose of a signal is to change the behavior of the observer (Bradbury & Vehrenkamp, 1998). For example, when being attacked, a wolf can signal its submission to a conspecific aggressor by rolling over on its back, exposing its belly and whimpering. The immediate effect of this behavior is to terminate the aggression of the dominant animal. The surrender display signals to the dominant animal that it has won the altercation. The remarkable part of this interaction is how the signal transforms the behavior of the observing animal: the angry aggressive attack immediately dissolves.

From an evolutionary perspective, one needs to ask the question “Why would any animal make a signal?” Evolutionary logic compels us to the conclusion that a signal will be made only if it is to the benefit of the signalling animal. If a signal reduces the fitness of the signalling animal, then the signalling behavior would be selected against.

Notice that there are plenty of affective states that should remain covert—that is, not communicated to others. Suppose for example, that you have stolen my food. I might be angry with you, but if you are clearly more powerful than me, it would be foolhardy for me to express anger in your presence. A better strategy would be to mask my feelings, and wait for an appropriate opportunity (such as assembling an alliance) that could ultimately prevail over you. Conversely, if I am the more powerful individual, there might be value in my overtly expressing anger—even if I do not actually feel anger. An expression of anger might make you respond in a deferential way—for example, abandoning your food so that I can take it. In short, in some cases an expression of anger can have a beneficial effect, even in the absence of any matching feeling. In other cases, no expression of anger should take place, even if one feels angry.
Here we see compelling reasons for separating affect from expression. For example, there are good reasons to distinguish two forms of anger: hot anger (anger that is displayed) and cold anger (anger that is felt but not displayed). Whether anger is expressed depends on whether the signal is beneficial to the signalling animal.

Notice that this logic is incompatible with the most popular understanding of emotional display — what might be called the Emotion Communication Model (ECM). This model might be described as follows: A person feels an emotion (such as happiness), and this causes them to generate an appropriate display (such as smiling). An observer perceives (recognizes) the display and infers that the individual feels happy. In the ECM model, it is assumed that the purpose of an emotional display is to communicate the affective state of the individual making the display. However, this assumption makes no biological sense. Instead, an expression should be viewed as other-directed: the expression is intended to change the behavior of the observer to the benefit of the signaller (Bradbury & Vehrenkamp 1998). Expressions of anger, sadness, happiness, etc. may certainly be regarded as communicative acts, but it is wrong to assume that they are intended to convey one's affective state.

Of course there are many variations to the basic ECM paradigm. An individual might engage in some deception—such as displaying an emotion that is not truly felt. Or an individual might attempt to mask or hide an otherwise spontaneous expression, such as turning away from an observer so they cannot see one's tears.

Several emotional displays are thought to be cross-culture universals (e.g., Ekman, 1972). At the same time, Ekman has suggested that purported universal expressions are often modified by local “display rules”—shaping various aspects of the display in culturally unique ways. Some evidence suggests that certain emotions may be culture-specific (e.g., Lutz, 1988).

Ethologists offer a very different perspective concerning the nature and function of displays, such as smiling or frowning. As we have seen, ethological research raises serious objections to the ECM model just described.

Tomkins (1980) has characterized emotions as motivational amplifiers—internal feeling states that encourage or compel an individual to behave in particular ways. Understood as motivational states, there are good reasons why some emotions would be experienced without any accompanying expression. Many affective states can exercise a transforming effect on behavior without being communicated: e.g., jealousy, love, hunger, disappointment, suspicion, pride, curiosity, etc. Of course, some affective states may indeed be recognizable even though they are not expressive signals (e.g., sleepiness, pain, etc.). But these states are recognized because of the spill-over of physiological...
concomitants that observers learn to decipher through past experience. The observed features for these states are artefacts rather than intentional communications; that is to say, they are ethological cues rather than signals.

In the past, some psychologists have tended to reify emotions as their expressions. In Ekman’s work, for example, there is a clear tendency to equate emotions with distinctive (facial) expressions. Any feeling-state that has no expression is deemed not to be an emotion. From an evolutionary and ethological perspective, these views are clearly problematic.

With this background, we can return to consider the contrast between sadness and grief.

9. Depressive realism

If sadness and grief are different affective states, we might ask what purpose they serve, and why they tend to co-occur.

Consider the etiology or causes of sadness. “Clueless Carl” is eager to date beautiful women. He approaches several beautiful women, each of whom declines his invitation for a date. After a series of such failures, Carl experiences feelings of sadness. Research indicates that sadness leads to reflection and reconsideration of life strategies (Nesse, 1991). Carl is likely to consider his own assets and liabilities, and recognize that he is not an especially handsome or accomplished man. He might consider other women he knows who are less physically beautiful but have other attractive qualities. In short, a bout of sadness is likely to cause Carl to reevaluate his romantic strategy, and to encourage him to set more realistic goals.

In general, people tend to hold overly optimistic self-appraisals (Ross & Nisbett, 1991). People tend to think they are more attractive, more intelligent, and more interesting than others judge them to be. We tend to look at the world through rose-tinted lenses. One might expect that when we are sad, we become pessimistic, underestimating ourselves. Instead, when sad, we are more realistic in our self-appraisals. This phenomenon is referred to as depressive realism (Alloy & Abrahamson, 1979). Compared with happiness, sadness encourages more detail-oriented thinking, less judgment bias, less reliance on stereotypes (Clore & Huntsinger, 2007) and greater memory accuracy (Storbeck & Clore, 2005). Listening to nominally sad music is known to induce depressive realism (Brown & Mankowski, 1993).
10. Mourning

When bad things happen in people’s lives, they often experience alternating periods of active grieving (crying) and quiescent sadness. We might refer to this oscillating pattern as mourning. Recall that sadness induces depressive realism and is typically accompanied by periods of reflection. Sadness causes us to think of how we might adapt to problematic circumstances (Nesse, 1991).

As we have seen, crying exhibits all of the hallmarks of an ethological signal. Recall that signals are intended to change the behavior of the observer. And indeed, crying does have a profound affect on others. In particular, crying leads to affiliative, supportive, and compassionate behaviors.

When bad things happen in our lives, there are two kinds of resources one may call upon to mount an effective response. One resource is our friends and family: people around us can come to our assistance. The second resource is ourselves. By thinking-through the situation, we can formulate strategies that help us cope with the difficulty.

My claim is that crying and sadness are different emotions that serve different (yet complimentary) purposes. Sadness is intended to change my behavior: reflection causes me to lower my expectations and contemplate different strategies that are better adapted to the environment. My crying is intended to change your behavior: crying encourages observers to become more altruistic. Said another way, sadness is an personal/covert emotion, whereas grief is a social/overt emotion. When we experience difficulties in life, we adapt through a combination of our own resources (sadness) plus help from others (solicited through crying).

Notice that this theory explains why crying would be an ethological signal whereas sadness would be an ethological cue. Sadness is simply not designed to be communicative. This does not necessarily mean that sensitive observers cannot recognize sadness in others, although it does suggest that sadness can be mistakenly attributed to another person, whereas assessments of grief are likely to be accurate.

11. Honest signalling

If crying is a signal—intended to influence the behavior of observers—then why do we often cry in private? Moreover, why do we often try to mask or hide the fact that we are crying?
To the extent that observers respond to signals in a biologically prepared manner, signals can be used deceptively (Smith & Harper, 2003; Zahavi, 1977). Observers need reassurance that a signal is authentic rather than deceptive. Considerable research has been carried out regarding mechanisms intended to ensure the honesty of signals (Pentland, 2008). Several mechanisms have been proposed.

One approach is to make a signal involuntary. Unlike the social (voluntary) smile, the genuine or “Duchenne” smile, for example, is difficult or impossible to “fake” (Freitas-Magalhães, 2006). Similarly, blushing is a response not under voluntary control. People who blush easily usually dislike this, however, observers are usually delighted by people who blush easily. These attitudes can be traced to honest signalling: we appreciate honesty in others, but feel vulnerable when our own signals are beyond our control. If crying is to remain an effective social signal, this can be best assured by making the response involuntary. As an involuntary response, we can expect it to arise even in non-social settings. People cry alone, not because crying is not intended to be observed, but because crying is an honest signal. Finally, the phenomenal experience of someone crying is hardly that of a person engaged in a Machiavellian trick to con bystanders—even if the evolved purpose of crying is to solicit help from others. The unconscious mind knows when to appeal for help, even if the conscious mind is an unwilling participant.

Another approach to honest signalling is the handicap principle, where making a signal must be “costly” for the individual making the signal (Zahavi & Zahavi, 1997). Like all appeasement displays, crying incurs the cost of the loss of social status. If an individual does not want to pay this cost, then they should attempt to hide or suppress their crying.

12. Affect induction

A core question in music-related affect is how music might lead listeners to experience some emotion: How does affect induction take place? Several plausible mechanisms have already been identified, including associative, empathetic, and cognitive emotional generators (Huron, 2002; Tuuri & Eerola, 2012). To these sources of emotion, ethological research suggests adding yet another—what might be called signalling. First, let us review three commonly identified emotional generators.

A first mechanism is simple association. As in the case of a conditioned response, certain sounds or sound patterns may become associated with past emotional experiences. The associations may be entirely arbitrary, as when a
nominally sad passage reminds a listener of a past happy event such as winning the lottery.

A second mechanism is empathetic. In this case, a listener recognizes acoustic features associated with particular emotions. Mirror neurons (for example) might induce an observer to vicariously experience feelings akin to those being displayed. For example, a listener might hear acoustic features associated with sadness, and consequently be induced to feel sad through some sort of mirror process.

A third mechanism is cognitive. Conscious thoughts can lead a listener to a particular experience. For example, when listening to Beethoven’s fifth symphony, a listener might be reminded of Susan McClary’s discussion of Beethoven’s work as a depiction of rape, and by interpreting the sounds in light of McClary’s suggestion, experience discomfort or alarm while listening to the work.

Signalling theory offers a fourth mechanisms for affect induction. Recall that the purpose of a signal is to change the behavior of the observer. For example, witnessing the submission display of a conspecific, the aggressor animal stops behaving aggressively. That is, the signal changes the behavior of the observer in a way that suggests feelings of anger are replaced by affiliative, playful, or altruistic feelings.

If crying is an ethological signal, then the purpose of crying cannot be to make the observer also experience grief. Instead, crying is effective when it transforms the observer’s state to affiliative, sympathetic and compassionate feelings. Notice that ethologists make the strong claim that the evoked behaviors in the observer are biologically innate. Signals are evolved behaviors whose effectiveness depends on stereotypic patterns of response. Accordingly, exposure to acoustic features of “grief” in music would be expected to induce affiliative, sympathetic, peaceful, altruistic or compassionate feelings, not grief.

Notice that since all four of these purported generative mechanisms are presumed to operate concurrently, one could well imagine more than one affect being induced in response to the same stimulus. For example, upon hearing a “lament”, a listener could well experience sadness/grief (through mirror-neuron-mediated empathetic responses) as well as compassionate feelings due to a biologically prepared response to signal features.
13. Musical repercussions

Ethology offers a number of useful insights for research in music-related affect. First, ethologists argue that displays must confer a fitness benefit for the displaying animal, otherwise the display would be selected against. This insight raises grave difficulties for the Emotion Communication Model described earlier. We should be wary of the idea that feelings are indiscriminately echoed in vocalizations or facial expressions. Only some feeling states ought to be overtly expressed; that is, as behavioral motivators, some emotions should be overt while others remain hidden. Whether an emotion is overtly expressed or covertly masked will depend on the benefit to the signaller. Overt expressions (signals) are intended to change the behavior of the observer, not to induce an emotion similar to that of the signaller. One should not view a display as symptomatic of the signaller’s affective state; instead it should be regarded as an effort to induce in the observer an affective state that is beneficial to the signaller.

In order for signals to be communicated, they should be conspicuous. Accordingly, signals tend to be multimodal. The most likely candidates for signals are those that exhibit both distinctive visual and distinctive acoustical features. In other words, any sound property that originates as a signal is likely to be accompanied by distinctive visible behaviors, such as characteristic facial expressions. Examples of candidate signals including smiling, sneering, and crying. As we have seen, each of these expressions involves distinctive multimodal features.

Some covert affective states can occasionally be inferred by an astute observer. Such covert displays (cues) are unintentionally informative. Signals primarily enhance the fitness of the signaller; cues enhance only the fitness of the observer.

Signals and cues might be regarded as “push” and “pull” forms of information. Signals “push” information into the environment—whether they are observed or not. Cues “pull” information from the environment, even though the information was not intentionally placed in the environment. Once again, through experience, the buzzing of a mosquito can be heard as presaging the possibility of attack, even though the sound is an artefact of rapidly moving wings.

The acoustical features associated with grief—wailing, moaning, sniffling, punctuated exhaling, ingressive phonation, pharyngealized voice, and cracking or breaking voice—are linked to distinctive visual features and exhibit the hallmarks of an ethological signal. By contrast, the acoustical features associated with sadness—quiet dynamic, slow tempo, low pitch, monotone
pitch contour, mumbled articulation, and dark timbre—appear to be simple artefacts of low physiological arousal. Listeners may infer that these features indicate sadness, but these same features will be evident in other states, such as sleepiness and relaxation.

We would therefore expect laments, cry songs, funerary wailing, and other “grieving” expressions to be highly communicative, and cross-cultural in their meaning. By comparison, we might predict that musics associated with low physiological arousal—such as meditative music, New Age music, devotional music, relaxing/easy listening, and lullabies/cradle songs—would be easily confused with sad music. Nominally sad music would therefore exhibit greater cultural confusion than “grief” or “lament” music.

In light of their differences, I have proposed that sadness is a personal/covert emotion whereas grief is a social/overt emotion. Nevertheless, sadness and grief tend to co-occur because they represent complementary strategies for dealing with personal difficulty.

Finally, with regard to affect induction, the concept of an ethological signal provides a previously overlooked mechanism for generating affect in observers. If ethologists are right, these behavioral changes are automatic and species-wide. Although the emotional experience of music is strongly shaped by cultural milieu and individual experience and association, research on signals suggests that signal-features should exhibit a high degree of cross-cultural agreement. Like Ekman’s display rules, the experience of signals might be expected to be modified by local cultural interpretation. However, if ethologists are right, signals should exhibit a common affective core deserving of the adjective “universal”.

14. Coda

Scholars have long been interested in the function, evolution, etiology, ontology, behavior, and phenomenology of emotion (Cornelius, 1996). There are different approaches to the study of emotion however. The Emotion Communication Model has dominated emotion research for several generations. Yet modern ethology suggests that the ECM model is biologically implausible. Displays don’t evolve unless they tend to benefit the displaying individual.

A key insight from ethology has been to distinguish how information benefits actors differently. Displaying and observing animals come to the information transaction with different — sometimes converging, often diverging — aims. Inspired by Peirce’s ideas, early ethologists took concepts such as signal, cue, and index, and applied these to the analysis of animal
interactions. After more than half a century, the results are worthy of the attention of semioticians, including those who have traditionally worked outside of the domain of biology.

In particular, ethology offers considerable insights for those interested in the study of music and emotion. For example, ethology provides a straightforward explanation for why grief displays are more likely to evoke compassion in an observer rather than empathetic grief. Such conceptual tools appear to provide useful methods for handling some of the thorniest problems in music-related emotion, including the difficult problem of distinguishing represented emotion from evoked emotion.*

* N.B. This paper is an expanded version of Huron (2012).

References


David Huron

Heinlein, C. P. (1928). The affective characteristics of the major and minor modes in music. Journal of Comparative Psychology, 8, 103–118.


David Huron


