Proceedings of the 12th International Conference on Music Perception and Cognition and the 8th Triennial Conference of the European Society for the Cognitive Sciences of Music, July 23-28, 2012, Thessaloniki, Greece
Cambouropoulos E., Tsougras C., Mavromatis P., Pastiadis K. (Editors)

Understanding Music-Related Emotion: Lessons from Ethology

David Huron

School of Music, Ohio State University, USA
huron.1@osu.edu

ABSTRACT

A number of musically-pertinent lessons are drawn from research on animal behavior (ethology). The ethological distinction between *signals* and *cues* is used to highlight the difference between *felt* and *expressed* emotion. Several ethologically-inspired studies are described – principally studies related to music and sadness. An ethologically-inspired model is proposed (the Acoustic Ethological Model). The question of how music induces emotion in a listener is addressed, and it is proposed that *signaling* represents a previously unidentified mechanism for inducing affect. An integrated theory of sadness/grief is offered, where sadness is characterized as a personal/covert affect, and grief is characterized as a social/overt affect. Sadness and grief tend to co-occur because they provide complementary strategies for addressing difficult circumstances.

I. INTRODUCTION

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, one approach has tended to dominate psychological thinking regarding emotions. We might refer to this as 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, the purpose of an emotional display is to communicate the affective state of the individual making the display. In this scenario, expressions of anger, sadness, happiness, etc., are communicative acts, intended to convey one's affective state.

Applied to music, it is widely presumed that some acoustical features or gestures imitate various elements of the affective displays, and this provides one of the ways by which music is able to communicate emotions that listeners recognize.

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).

Scholars who study animal behavior (ethologists) offer a very different perspective concerning the nature and function of displays, such as smiling or frowning. In this paper, I draw on ethological principles and offer a number of suggestions regarding the study of music-related emotion. In particular, we will see that ethological research raises serious objections to the ECM model just described.

Before continuing, it is appropriate to acknowledge an important limitation in applying ethological research to music. Since ethologists work with non-human animals, they are unable to infer the affective state of the individual. They cannot ask an animal "how are you feeling?" (Tinbergen, 1951). As a consequence, ethology has nothing to offer concerning the *phenomenological* aspects of music-induced emotion. Ultimately, "how we feel" is arguably the most important aspect of music-related emotion. Nevertheless, we will see that ethological principles provide useful conceptual tools for approaching the study of music-related emotion.

II. SIGNALS AND CUES

Ethologists make a useful distinction between two kinds of animal communications: *signals* and *cues* (Lorenz, 1939; 1970; Smith & Harper, 2003). 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 artifact 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 learned artifactual behaviors.

III. 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 & 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 (in process) coded the degree of sociability for hundreds of characters in randomly selected opera scenarios. Friendly and altruistic characters were rated as exhibiting high sociability whereas self-centered 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 have two acoustical components: a *source* (vocal folds) and a *filter* (vocal tract). The frequency of the vocal folds (F0) 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., [u]) with a low F0. Speaking a high vowel (e.g., [i]) with a high F0 produces a high pitch and a high resonance. However, we can also produce mixtures, such as speaking [i] with a low pitch (low pitch + high resonance) or speaking [u] with a high intonation (high pitch + low resonance).

A. 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 or F0). 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 smiling 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.

B. 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 resonant 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.

IV. 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 channels (Wiley 1983; Johnstone 1997; Partan & Marler, 1999). Since signals are intended to be communicated, a "subtle" signal is less likely 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 artifacts (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 or more likely to be artifactual *cues* rather than *signals*.

Notice that the smile and the pout/lout 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 (F0) 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 our multimodal conjecture we ought to see a distinctive visual element that accompanies the higher/lower vocal pitch. The pertinent study was done by Huron, Dahl and Johnson (2009; see also commentary by Ohala, 2009). We asked 44 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. What about the reverse relationship? Does moving your eyebrows cause your voice to move up or down in pitch? Huron and Shanahan (2012) carried out the pertinent experiment. 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 (F0). 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 signal for humans and that analogous displays are evidence in other primates.

In summary, there is a strong relationship between pitch height (F0) 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 signaling 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 musician to lower her/his eyebrows. Conversely, when the music involves predominantly high pitches, there is a tendency for the musician to raise his/her eyebrows.

C. 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 labli superioris* muscles that elevate the upper lip and the *depressor septi* muscle that constricting the nostrils. This is presumed to have originated in efforts to reduce the inhaling of offensive odors. The sneer is essentially "one-half" of a disgust expression. That is, flexion occurs asymmetrically on one side of the face. As Ekman notes, the disgust response says "I find *this* disgusting" whereas the sneer is a social display saying "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.

V. 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 vice produces an involuntary tendency to move the eyebrows in a parallel fashion. At the same time, voluntary efforts to raise or lower the eyebrows produces an involuntary tendency 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.

Finally, we have observed an association between the sneer/contempt expression, where pinching the nose causes a distinctive visual expression accompanied by audible nasalization of the voice – consistent with the multimodal tendency of ethological signals. In each case, these same features can be observed in musical contexts.

VI. 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 rumination (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 mixed 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, 1899/1921). 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, sniffling, ingressive phonation (vocalizing while inhaling), punctuated exhaling, and involve pharyngealized voice (due to 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, erratic breathing, and puffy face. 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.

VII. 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 movements or a more monotone pitch infection. 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 artifacts 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 anticipate 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 artifacts of low energy.
4. Sad voice may not be distinguishable from sleepy voice (Shanahan & Huron, in progress).

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

VIII. 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 the last quadrant regarded as a candidate ethological cue.

Table 1. Acoustic Ethological Model.

	Quiet	**Loud**
High pitch	appeasement, friendliness	fear, alarm
Low pitch	sad, relaxed, sleepy	aggression, seriousness

IX. 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 signaling animal. If a signal reduces the fitness of the signaling animal, then the signaling 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 signaling animal.

Notice that this logic is incompatible with the Emotion Communication Model (ECM) described earlier. The assumption in the ECM is that an expression is intended to convey what an individual feels. 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 signaler (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.

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 *artifacts* 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.

X. 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 rumination 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).

XI. 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 rumination. Sadness causes us to think of how we might adapt to problematic circumstances (Nesse, 1991).

As we've 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: rumination 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.

XII. 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 signaling: 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 isn't 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 signaling 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 doesn't want to pay this cost, then they should attempt to hide or suppress their crying.

XIII. 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 *signaling*. First, let's 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 to 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.

Signaling 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.

XIV. 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 signaler. Overt expressions (signals) are intended to change the behavior of the observer, not to induce an emotion similar to that of the signaler. One should not view a display as symptomatic of the signaler's affective state; instead it should be regarded as symptomatic of what affective state the signaler hopes to induce in the observer.

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 signaler; 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 artifact 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 artifacts 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 unidentified 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."

REFERENCES

Alloy, L.B., & Abramson, L.Y. (1988). Depressive realism: Four theoretical perspectives. In L.B. Alloy (Ed.), *Cognitive processes in depression,* (pp. 223-265). New York: Guilford.

Bauer, H.R. (1987). Frequency code: Orofacial correlates of fundamental frequency. *Phonetica*, 44, 173-191.

Bolinger, D.L. (1964). Intonation across languages. In J.H. Greenberg, C.A. Ferguson & E.A. Moravcsik (Eds.), *Universals of Human Language, Vol. 2: Phonology* (pp. 471-524). Stanford, CA: Stanford University Press.

Bonfiglioli, L., Caterina, R., Incasa, I., & Baroni, M. (2006). Facial expression and piano performance. In M. Baroni, A.R. Addessi, R. Caterina, M. Costa (Eds.), *Proceedings of the 9th International Conference on Music Perception and Cognition,* pp. 1355-1360.

Bradbury, J.W. & Vehrenkamp, S.L. (1998). *Principles of Animal Communication*. Sunderland, MA: Sinauer.

Brown, J.D., & Mankowski, T. (1993). Self-esteem, mood, and self-evaluation: Changes in mood and the way you see you. *Journal of Personality and Social Psychology*, 64, 421-30.

Calhoun, C., & Solomon, R.C. (1984). *What is an Emotion? Classic Readings in Philosophical Psychology*. Oxford: Oxford University Press.

Clore, G.L., & Huntsinger, J.R. (2007). How emotions inform judgment and regulate thought. *Trends in Cognitive Science,* 11, 393–399.

Cornelius, R.R. (1996). *The Science of Emotion: Research and Tradition in the Psychology of Emotion*. Upper Saddle River, New Jersey: Prentice-Hall.

Cruttenden, A. (1981). Falls and rises: meanings and universals. *Journal of Linguistics*, 17, 77-91.

Darwin, C. (1872). *The Expression of the Emotions in Man and Animals*. London: John Murray.

Eggebrecht, H. H. (1955). Das Ausdrucks-Prinzip im musikalischen Sturm und Drang, *Deutsche Vierteljahrsschrift für Literaturwissenschaft und Geistesgeschichte*, 29, 323-349.

Eibl-Eibesfeldt, I. (1989). *The Biology of Peace and War: Men, Animals, and Aggression*. New York: Viking.

Ekman, P. (1972). *Emotion in the Human Face: Guide-Lines for Research and an Integration of Findings*. New York: Pergamon Press.

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

Fox, A.A. (2004). *Real Country: Music and Language in Working-Class Culture*. Durham, NC: Duke University Press.

Freitas-Magalhães, A. (2006). *The Psychology of Human Smile*. Oporto: University Fernando Pessoa Press.

Frey II, W.H. (1985). *Crying: The Mystery of Tears.* Minneapolis: Winston Press.

Frijda, N. (1986). *The Emotions: Studies in Emotion and Social Interaction*. Cambridge: Cambridge University Press.

Gelstein, S., Yeshurun, Y., Rozenkrantz, L., Shushan, S.,Frumin, I., Roth Y., & Sobel, N. (2011). Human tears contain a chemosignal. *Science*, 331, 226-230.

Gould, S.J. & Vrba, E.S. (1982). Exaptation – a missing term in the science of form. *Paleobiology*, 8, 4-15.

Heinlein, C. P. (1928). The affective characteristics of the major and minor modes in music. *Journal of Comparative Psychology*, 8, 101-142.

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

Huron, D. (2002). A six-component theory of auditory-evoked emotion. *Proceedings of the 7th International Conference on Music Perception and Cognition*. Sydney, Australia.

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

Huron, D., Dahl, S., & Johnson, R. (2009). Facial expression and vocal pitch height: Evidence of an intermodal association. *Empirical Musicology Review*, 4, 93-100.

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

Huron, D., & Shanahan (2012). Eyebrow movements and vocal pitch height: Evidence consistent with an ethological signal. Unpublished manuscript.

Huron, D., Yim, G., & Chordia, P. (2010). The effect of pitch exposure on sadness judgments: An association between sadness and lower than normal pitch. In S.M. Demorest, S.J. Morrison,

P.S. Campbell (Eds.), *Proceedings of the 11th International Conference on Music Perception and Cognition* (pp. 63-66). Seattle, Washington: Causal Productions.

Johnstone, R.A. (1997). The evolution of animal signals. In J.R. Krebs & N.B. Davies (Eds.), *Behavioural Ecology* (pp. 155-178). Oxford: Oxford University Press.

Kottler, J.A., & Montgomery, M.J. (2001). Theories of crying. In Ad J.J.M. Vingerhoets & Randolph R. Cornelius (Eds.), *Adult Crying: A Biopsychosocial Approach*. Hove, East Sussex, UK: Brunner-Routledge.

Kraepelin, E. (1899/1921). *Psychiatrie. Ein Lehrbuch für Studierende und Arzte, Klinische Psychiatrie. II*. Leipzig: Johann Ambrosius Barth, 1899. Trans. by R.M. Barclay as *Manic-Depressive Insanity and Paranoia*. Edinburgh: E. & S. Livingstone, 1921.

Ladinig, O., & Huron, D. (2010). Dynamic levels in Classical and Romantic keyboard music: Effect of musical mode. *Empirical Musicology Review*, 5, 51-56.

Lorenz, K. (1939). *Vergleichende Verhaltensforschung. Zoologische Anzeiger*, *Supplement*, 12, 69-102.

Lorenz, K. (1970). *Studies in Animal and Human Behaviour*, Vol. 1. London: Methuen.

Lutz, C.A. (1988). *Unnatural Emotions: Everyday Sentiments on a Micronesian Atoll & Their Challenge to Western Theory*. Chicago: University of Chicago Press.

Morton, E.S. (1977). On the occurrence and significance of motivation-structural rules in some bird and mammal sounds. *American Naturalist*, 111, 855-869.

Morton, E. (1994). Sound symbolism and its role in non-human vertebrate communication. In L. Hinton, J. Nichols & J. Ohala (Eds.), *Sound Symbolism*, (pp. 348-365). Cambridge: Cambridge University Press.

Morton, E. (2006). Commentary on `The influence of pitch height on the perception of submissiveness and threat in musical passages' by David Huron, Daryl Kinney, and Kristin Precoda. *Empirical Musicology Review*, 1, 178-179.

Nesse, R. (1991). What good is feeling bad? The evolutionary benefits of psychic pain. *The Sciences*, 31, 30-37.

Ohala, J. (1980). The acoustic origin of the smile. *Journal of the Acoustical Society of America*, 68, S33.

Ohala, J. (1982). The voice of dominance. *Journal of the Acoustical Society of America*, 72, S66.

Ohala, J. (1983). Cross-language use of pitch: an ethological view. *Phonetica*, 40, 1-18.

Ohala, J. (1984). An ethological perspective on common cross-language utilization of F0 in voice. *Phonetica*, 41, 1-16.

Ohala, J. (1994). The frequency code underlies the sound-symbolic use of voice pitch. In L. Hinton, J. Nichols & J. Ohala (Eds.), *Sound Symbolism*, (pp. 325-347). Cambridge: Cambridge University Press.

Ohala, J. (2009a). Signaling with the eyebrows - Commentary on Huron, Dahl, and Johnson. *Empirical Musicology Review*, 4, 101-102.

Ohala, J. (2009b). The ethological basis of certain signals of affect and emotion. In S. Hancil (Ed.), *The Role of Prosody in Affective Speech*. (pp. 17-30). Bern: Peter Lang.

Partan, S., & Marler, P. (1999). Communication goes multimodal. *Science*, 283, 1272-1273.

Paul, B., & Huron, D. (2010). An association between breaking voice and grief-related lyrics in Country music. *Empirical Musicology Review*, 5, 27-35.

Pentland, A. (2008). *Honest Signals: How They Shape Our World*. Cambridge, MA: MIT Press.

Plazak, J. (2011). *Instrumental irony and the perception of musical sarcasm*. PhD dissertation. School of Music, Ohio State University.

Post, O., & Huron, D. (2009). Music in minor modes is slower (Except in the Romantic Period). *Empirical Musicology Review*, Vol. 4, No. 1, pp. 1-9.

Scherer, K. R. (2004). Which emotions can be induced by music? *Journal of New Music Research*, 33, 239–251.

Scherer, K., London, H., & Wolf, J.J. (1973). The voice of confidence: Paralinguistic cues and audience evaluation. *Journal of Research in Personality*, 7, 31-44.

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

Smith, J.M. & Harper, D. (2003). *Animal Signals*. Oxford: Oxford University Press.

Storbeck, J., & Clore, G.L. (2005). With sadness comes accuracy; With happiness, false memory: Mood and the false memory effect. *Psychological Science*, 16, 785-791.

Tartter, V.C. (1980). Happy talk: Perceptual and acoustic effects of smiling on speech. *Perception & Psychophysics*, 27, 24-27.

Tinbergen, N. (1951). *The Study of Instinct*. Oxford: Clarendon Press.

Tomkins, S.S. (1980). Affect as amplification: some modifications in theory. In R. Plutchik & H. Kellerman (Eds.), *Emotion: Theory, Research and Experience* (pp. 141-164). New York: Academic Press.

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

Tuuri, K., & Eerola, T. (2012). Formulating a revised taxonomy for modes of listening. *Journal of New Music Research*, 41(1): 1-16.

Wiley, R.H. (1983). The evolution of communication: information and manipulation. In T.R. Halliday & P.J.B. Slater (Eds.), *Communication* (pp. 82-113). Oxford: Blackwell.

Williams, C.E., & Stevens, K.N. (1972). Emotions and speech: Some acoustical correlates. *Journal of the Acoustical Society of America,* 52, 1238-1250.

Zahavi, A. (1977). Reliability in communication systems and the evolution of altruism. In B. Stonehouse & C.M. Perrins (Eds.), *Evolutionary Ecology* (pp. 253-259). London: Macmillan.

Zahavi, A., & Zahavi, A. (19970). *The handicap principle: A missing piece of Darwin's puzzle.* New York: Oxford University Press.

Zentner, M., Grandjean, D., & Scherer, K.R. (2008). Emotions evoked by the sound of music: Characterization, classification, and measurement. *Emotion*, 8, 494-521.