



CrossMark
click for updates

Opinion piece

Cite this article: Huron D. 2015 Affect induction through musical sounds: an ethological perspective. *Phil. Trans. R. Soc. B* **370**: 20140098.
<http://dx.doi.org/10.1098/rstb.2014.0098>

One contribution of 12 to a theme issue 'Biology, cognition and origins of musicality'.

Subject Areas:

behaviour, cognition

Keywords:

music, emotion, ethology, signals, cues, mirror neurons

Author for correspondence:

David Huron

e-mail: huron.1@osu.edu

Affect induction through musical sounds: an ethological perspective

David Huron

School of Music and Center for Cognitive and Brain Sciences, Ohio State University, 1866 College Road, Columbus, OH 43210, USA

How does music induce or evoke feeling states in listeners? A number of mechanisms have been proposed for how sounds induce emotions, including innate auditory responses, learned associations and mirror neuron processes. Inspired by ethology, it is suggested that the ethological concepts of *signals*, *cues* and *indices* offer additional analytic tools for better understanding induced affect. It is proposed that ethological concepts help explain why music is able to induce only certain emotions, why some induced emotions are similar to the displayed emotion (whereas other induced emotions differ considerably from the displayed emotion), why listeners often report feeling mixed emotions and why only some musical expressions evoke similar responses across cultures.

1. Introduction

When asked why they listen to music, people commonly allude to the pleasure of how music makes them feel. Music is capable of evoking a wide range of feeling states from the pedestrian to the sublime [1,2]. Not all emotions are equally easy to evoke, however. Music might evoke feelings of exuberance, compassion or tenderness, but it is less common that music evokes or portrays emotions such as jealousy or guilt [3].

Conceptually, a distinction can be made between emotional representations or portrayals in music and the emotions that are evoked or induced by music. Theoretically, a musical passage might be regarded as portraying (say) fear, but evoking (instead) jocularity. Considerable research has focused on the associations between various acoustical features and the representation or portrayal of various affects through music (see [4] for summary). It is generally acknowledged, however, that the more difficult problem is identifying the means by which affective states are evoked in listeners.

Several possible mechanisms have been proposed to account for musically evoked emotions. These include innate auditory responses, learned associations, mirror neuron processes and many other possible sources (e.g. [5,6]). Efforts to identify possible sources have been partly confounded by the variability of responses both between different cultural communities and between individuals within a given culture. Even with regard to the perception or recognition of affective portrayals in music, extant research has documented considerable variability. For example, whether a particular musical passage is perceived as expressing or representing a given emotion is known to be influenced by personality: listeners who score high on Neuroticism are more likely to characterize a passage as 'sad-sounding' [7].

The variability found between individuals and cultures has led some researchers to conclude that music-induced affect must arise principally or exclusively due to learning. This variability further implies that any presumed biological foundations for music are likely to be shallow, and so calls into question several proposals for possible evolutionary bases for the origins of music.

Both conditioned responses and learned associations may have an arbitrary relationship to the musical stimulus: a song that typically evokes happiness might, for some listener, evoke sadness due to an association with a failed romantic relationship. While some learned associations are widespread across a population (such as the use of a tune in a popular television show), many

associations are idiosyncratic to specific listeners. Even if it were the case that learned associations provoke the strongest and most reliable emotions in listeners, unless these associations are broadly shared across listeners, there is little incentive for musicians to use these associations as compositional tools for evoking particular responses. The existence of a musical culture depends on shared experiences.

From a musical perspective, purported mirror neuron processes appear to offer a more attractive method for inducing listener emotions since mirror neurons imply a close echoing of expressed and evoked emotion (at least when there is a mirrored action involved). Gallese *et al.* [8] have suggested that mirror neurons provide a plausible neurological mechanism for empathetic responses. Singer *et al.* [9] carried out brain scans while they inflicted obvious pain on a romantic partner of the person being scanned. They found similar patterns of activation in brain regions associated with pain in the observing romantic partner—suggesting that empathetic responses rely on emulating the emotional experiences of others. Although Singer's work does not specifically link empathetic responses to mirror neuron processes, several studies are indeed consistent with the idea that mirror neuron systems lead to empathetic feeling states [10,11]. Evidence consistent with auditory-evoked mirror responses exists [12,13] and Molnar-Szakacs & Overy [6] have explicitly proposed that music evokes empathetic affective states in listeners through a mirror neuron process.

The promise of mirror neurons notwithstanding, in this paper, other ways by which music might induce emotions are proposed and explored. Drawing on concepts in ethology, it is suggested that ethological signals, cues and indices provide additional useful analytic tools for approaching auditory-induced affect. In particular, it is suggested that these concepts can help to decipher whether a given musical or acoustical feature is likely to induce similar affective responses in listeners across cultures.

2. Size matters

One of the best cross-species generalizations one can make about behaviour is that appearing to be large is associated with threat, whereas appearing to be small is associated with deference or submission. This principle is not restricted to vision. One of the best generalizations one can make regarding acoustics is that large masses or volumes vibrate at a lower frequency. Morton [14] observed this principle in a broad sample of vocalizations from 56 species. In general, high-pitched calls are associated with fear, affinity or submission, whereas low-pitched calls are associated with threat or aggression. This same principle has been observed in human speech. Across cultures, high pitch is associated with friendliness or deference, whereas low pitch is associated with seriousness or aggression [15].

The same relationship has been observed in music. When melodies are transposed lower in pitch, they are perceived as less polite, less submissive and more threatening [16]. This effect is clearly apparent in opera, where heroic roles tend to be assigned to high-pitched tenors and sopranos, whereas villainous roles tend to be assigned to basses and contraltos [17]. In speech, an important acoustical distinction is made between the source and filter components [18]. The vocal folds provide the sound source whose vibrations determine the fundamental

frequency (F_0) or pitch of the voice. The vocal tract (pharynx, oral cavity and nasal passages) provides the filter component. The volume and length of the oral cavity can be manipulated, changing the resonance of the ensuing sound. Low-resonance vowels are associated with a dropped chin and protruding lips that both lengthen the vocal tract and increase the oral cavity volume. Conversely, high-resonance vowels are associated with a raised chin and retracted lips—resulting in higher frequency resonances. The independence of the source and filter in speech means that there are two distinct frequency components: the fundamental frequency and the vocal tract resonance.

Ohala [19] proposed a theory of the human smile that deserves to be better known. Why would a display that commonly reveals a person's teeth be construed as anything other than a threat display? Tarter [20,21] showed that listeners easily recognize the sound of a smiling voice. Pulling the lips taut against the teeth shortens the vocal tract length (VTL)—shifting the main resonance higher in frequency. Ohala suggested that the smile originated as an auditory rather than as a visual display. Specifically, the upward shift of the smiling-voice resonance is consistent with the association of small size with affiliative or deferential behaviours.

Ohala's conjecture suggests that the 'opposite' of the smile may not be the frown, but the pout—where the lips are thrust outward. The association of a pout or pucker display with threat has been observed in many animal species [22–27]. Among humans, the aggressiveness of the pout is evident in film portrayals of the hooligan where the speaker's lips are pushed forward—producing a bellicose or hostile-sounding timbre. Although Ohala has drawn attention to the auditory properties of smiling, as we will see, the combination of visual and auditory features likely represents the more seminal observation.

3. Signals and cues

Consider the distinction made by ethologists between signals and cues [28,29]. A *signal* is an evolved purposeful communicative behaviour, such as evident in a rattlesnake's rattle. The rattle is a special anatomical organ that has evolved explicitly for communicative purposes. When threatened, the rattling behaviour alerts the observer to the possibility of a pre-emptive or retaliatory attack.

By contrast, an ethological *cue* is a non-functional behaviour that is nevertheless informative. An example would be the buzzing sound produced by a mosquito suggesting an imminent attack. When hearing the buzzing of a mosquito, past experience suggests that there is a good likelihood of being bitten. However, in this case, the buzzing sound is simply an unintended consequence of the insect flapping its wings—not a functional communicative act.

Both the rattlesnake's rattle and the mosquito's buzzing presage the possibility of an attack. But one is a functional signal, whereas the other is an artefactual cue. Both sounds are informative, but only one sound is overtly communicative.

Signals are important because their purpose is to change the behaviour of the observer—to the mutual benefit of signaler and observer [30,31]. Among ethologists, the ensuing changes in behaviour are thought to arise because the signal activates co-evolved circuits intended to influence motivation. Among humans, these motivational circuits are more commonly deemed 'emotions' [32]. Unlike cues, signals are

evolved behaviours that induce stereotypic feeling states in an observer—as when crying evokes feelings of compassion in a bystander [33–35]. In music, those acoustical features that imitate true signals (in the ethological sense) are most likely to induce affective states in listeners that are similar cross-culturally. That is, signal-like musical features are likely to be interpreted similarly.

Moreover, with regard to affect induction, the concept of an ethological signal provides a commonly overlooked mechanism for generating affect in observers. If ethologists are correct, these behavioural changes are largely automatic and species-wide.

4. Multimodal signals

Determining whether a particular behaviour represents a cue or signal is non-trivial. Ethologists have proposed a number of characteristics that help in making this distinction. Most importantly, signals tend to be conspicuous [36,37]. If the purpose is to communicate, then the display behaviour is unlikely to be timid or subtle. One of the best ways to increase conspicuousness is by making the signal multimodal. In the case of the rattlesnake, for example, the raised shaking tail is visually distinctive, even if the rattling sound is not heard. Cues may also be multimodal, but as the behaviours are artefacts, they are less likely to exhibit multimodal features. In short, cues tend to be unimodal, whereas signals tend to be multimodal [30].

Whether or not Ohala's theory of the acoustical origin of the smile is correct, the important point is that smiling exhibits both characteristic visual and auditory features—consistent with a multimodal tendency for ethological signals. That is, the smile display is consistent with a co-evolved communicative function.

Note that the smile and pout relate only to the filter component of the source-filter foundations of vocalization. This raises the question of the signalling status of the source (F_0) component of vocalization. Unlike the smile, it is not immediately apparent whether changes in F_0 are accompanied by a distinctive visual element. Does a low voice *signal* aggression (in the ethological sense), or is it merely an artefactual cue? If F_0 is a signal, we might expect to see an accompanying visual correlate consistent with the multimodal tendency of ethological signals.

In the case of chimpanzees, F_0 is positively correlated with teeth and lip opening distances [38]. However, among *Homo sapiens*, the key visual element appears to reside in the eyebrows. Huron *et al.* [39] showed that the height of the eyebrows track vocal pitch. When asked to raise or lower the pitch of the voice, participants' eyebrows tend to rise and fall in tandem with the pitch. Moreover, assessments of the resulting facial expressions are consistent with affective assessments of F_0 : photographs of faces when producing a high-pitch vocalization were deemed much friendlier than matched facial photographs of low-pitch vocalization [39]. When producing a low pitch, the head tends to tilt downward, the chin drops and the eyebrows are lowered; conversely, when producing a high pitch, the head tends to tilt upward, the mouth tends to form a smile and the eyebrows are raised.

In order to control for a possible confounding motor relationship between the larynx and chin position, a second control experiment was carried out in which the photographs were cropped so that subjects could view only the region

above the nose. Nevertheless, faces were clearly deemed more friendly or aggressive when producing high and low pitches, respectively. In a subsequent study [40], the reverse causation was also demonstrated. When instructed to move their eyebrows up or down, F_0 was found to move in tandem. The observations that moving pitch causes the eyebrows to move and that moving the eyebrows causes the pitch to move, implies the existence of a central motor process that controls both the eyebrows and the vocal folds. This multimodal relationship is consistent with an ethological signal—suggesting that eyebrow position and vocal pitch may be components of a single communicative display.

In light of these apparently successful reinterpretations of the smile and pitch/eyebrow displays, we might consider applying a similar ethological analysis to other facial expressions—and to consider their possible repercussions for understanding music-induced affect.

5. Sneer as signal

Ekman & Friesen [41] published a seminal description of the 'contempt' facial expression (cross-culturally replicated in [42]). Ekman proposed that contempt is a variant of the disgust display. When forming a disgust expression, the upper lip is elevated and the nostrils are flexed [41]. Contempt can be viewed as an asymmetrical version of disgust, in which the elevation of the upper lip and nostril flexion occurs on only one side of the face. According to Ekman, disgust is a general expression of revulsion (such as when exposed to repugnant smells or tastes), whereas contempt is a social expression—used to communicate hostility towards another person or social situation (see also [43]). Applying the ethological question, we might once again ask whether contempt is a signal? More specifically, is there a characteristic sound that accompanies a characteristic visual display?

Contemptuousness and sarcasm are readily conveyed through speech prosody. Even if one does not understand the language, contempt is often apparent. Sarcastic or contemptuous speech appears to be linked with high nasalization. The nasal quality of contempt is exemplified by playground taunts like *nya, nya*.

In the vocal tract, the nasal passages normally behave as static acoustical cavities with a fixed formant or resonance. However, the overall acoustical effect is influenced by the amount of airflow through the nose. Nasalization arises when the airflow is high [44]. So-called 'nasals'—phonemes such as 'm' and 'n'—are accompanied by high nasal airflow. The acoustical effect of nasalization is that the first formant becomes broader. As a consequence, more energy is shifted to the upper partials.

The volume of nasal airflow is determined by the size of the velopharyngeal opening. In general, the bigger the opening, the greater the nasality. Sundberg has suggested that nasalization is not directly caused by flexing nose muscles. Instead, there is some motor connection between the nose and the velum/pharynx that causes one to influence the other (J. Sundberg 2014, personal communication).

With regard to music, pertinent work has been carried out by Plazak [45] who asked instrumentalists to perform various passages (including single tones) in happy, sad, angry and sarcastic manners. In a forced-choice paradigm, Plazak found that listeners had little difficulty distinguishing these different

represented or expressed affects. Plazak then analysed the various sounds in an effort to identify acoustical correlates with musical sarcasm. For those instruments most capable of conveying sarcasm, Plazak found that the sarcastic renditions exhibited elevated levels of *nasality* as measured using standard speech analysis methods. Although further research is warranted, these observations suggest an association between a facial expression ('sneer') and a recognizable change in vocal sound ('nasalization')—consistent with multimodal tendencies of ethological signals. Moreover, we see evidence that these features generalize beyond speech—and can be observed in musical timbres as well.

6. Grief as signal: sadness as cue

Darwin [46] distinguished *sadness* (a low-arousal affect) from *grief* (a high-arousal affect). Compared with normal speech, sad speech is slower, quieter, lower in pitch, more monotone, more mumbled and exhibits a dark timbre (lower spectral centroid). Grief-related vocalizations, by comparison, exhibit high pitch, pharyngealization, breaking voice and ingressive phonation [47].

Huron offered a detailed argument that sadness bears the hallmarks of an ethological cue, whereas grief bears the hallmarks of an ethological signal [47]. Space limitations preclude reviewing the argument here. By way of summary, all of the features of sad speech can be attributed to low physiological arousal, and cannot be easily distinguished from similar low-arousal states such as relaxed speech or sleepy speech. Similarly, a nominally sad facial expression is not easily distinguished from a relaxed or sleepy facial expression. Moreover, sad individuals tend to remain mute—which is not consistent with a communicative function. By contrast, grief exhibits highly distinctive visual and acoustical features. Grief is accompanied by a compulsion to vocalize—exemplified by the rare phenomenon of phonating while inhaling (gasp sound). In addition, Gelstein *et al.* [33] have shown that psychic (emotionally evoked) tears contain a pheromone that induces observers to behave in a more compassionate way. In short, grief exhibits multimodal characteristics (visual, acoustical and olfactory) that result in notable changes in the feeling states of observers—consistent with an ethological signal. Sadness, by contrast, is easily confused with relaxation, and these confusions are echoed in musical confusions, such as disagreements among listeners about whether New Age music is sad-sounding or relaxing-sounding. As noted in §1, these judgements are known to be confounded by personality—suggesting that any presumed communicative function is not robust.

An especially distinctive acoustical feature of grief is the sound of 'breaking' voice—caused by an abrupt transition between falsetto and modal phonation. The breaking sounds are evident in many musical styles, including Western opera and country music. Paul & Huron [48] carried out a study of breaking voice in commercially recorded music and found that moments of breaking voice are positively correlated with grief-related content in the lyrics.

Recall that ethological signals are produced in order to change the behaviour of the observer to the benefit of the signaler. In the case of grief, crying behaviour has a tendency to evoke strong feelings of compassion in observers. This suggests that listening to musical passages that emulate grief-related acoustical

features should tend to evoke—not feelings of sadness or grief, but feelings of compassion and tenderness in listeners.

7. Cuteness as index

In an exploratory study of auditory cuteness, we found that listeners exhibit high intersubjective agreement when characterizing the degree of 'cuteness' for various sounds. Listeners heard a wide range of sounds, including those produced by assorted toys and musical instruments. Sounds scoring high in cuteness (such as squeeze toys, ocarina, soprano recorder or music box) were described by listeners using terms such as fragile, delicate, innocent and vulnerable. The sounds appear to induce nurturing and protective attitudes. In short, the responses are consistent with the evoking of parenting behaviours appropriate for interacting with infants.

An analysis of the sound-producing mechanisms established that 'cute' sounds tend to be associated with a small resonant cavity activated by a small amount of energy. Although no formal tests were carried out, the optimum size of a 'cute' cavity appears to resemble the vocal tract of infants. Vorperian *et al.* [49] measured VTLs across various ages, from neonates to adults. A typical newborn infant has a vocal tract near 8 cm in length, whereas adult VTLs are typically 14–17 cm.

Fitch [50] has noted that VTL is consistent with the ethological concept of an *index*. An index is defined as a type of signal whose variability is causally related to the quality being signalled—and which cannot be faked [51]. An example of an index is the marking of trees by tigers. Tigers mark trees by reaching as high as they can along the trunk. Consequently, the position of the smell provides a reliable indicator of the size of the animal.

In the acoustical domain, formant dispersion has been identified as a reliable indicator of size [51]. Specifically, formant dispersion is directly proportional to VTL, and VTL is strongly correlated with body size [51]. In this regard, formant dispersion contrasts notably with vocal pitch. Although the sex, age and size of an individual establish a general tessitura or range for vocal pitch, within this range, pitch variability can be voluntarily controlled. In short, pitch provides a useful signal reflecting affective states such as aggressive or submissive attitudes, whereas formant dispersion provides a useful index of the individual's size—largely independent of affective state or intention. As noted in §2, the VTL is susceptible to some modification (due to retracting or extending the lips). However, in contrast to pitch variability, VTL is much more constrained—much less susceptible to voluntary control, and therefore a reliable or honest indicator of body size.

In music, examples of 'cuteness' are readily found. A good example is the innocent 'little girl' voice of Helen Kane—a diminutive popular American singer from the 1930s. Kane became the model for the subsequent well-known Betty Boop cartoon character. In light of the idea that VTL represents an (honest) ethological index of size, it is understandable how these sounds might provoke descriptions suggestive of parenting behaviours—specifically, that the sounds might be described as innocent and vulnerable, while evoking feelings of care, tolerance or nurturing in listeners.

Signals, cues and indices are classes of adaptations. Apart from the adaptive concepts, Tinbergen would encourage

consideration of some of the causal mechanisms that may be involved. To this end, we might contrast the ethological concepts with two plausible mechanisms for music-induced affect, notably mirror processes and associations.

8. Tempo as mirror

Mirror neuron systems are closely linked to motor behaviours. Those stimuli most apt to evoke mirror responses are those implying a particular motor action rather than an emotion *per se*. When we hear someone speaking with a 'frog' in his/her throat, we may feel an unconscious compulsion to want to clear our own throat. Any evoked emotions are presumed to arise from a Jamesian process of self-perception [52]. Mirror systems may be the basis for emotional contagion [53].

The close relationship to motor action suggests that mirror responses are most likely to be evoked by musical features related to action. A large body of research has examined the relationship between music and movement. The foremost behaviour examined has been locomotion, notably walking.

Recording acceleration data from different body parts over sustained periods of daily activity, MacDougall and Moore [54] showed that the distribution of preferred musical tempos is very closely related, not to the movement of legs or arms, but to the motion of the head. (When walking, head movements are twice the frequency of individual leg or arm movements.) Consistent with other music research, the key to tempo perception appears to be vestibular activation. What makes 40 beats min^{-1} sound sluggish and lethargic (whereas 120 beats min^{-1} sounds spirited and zestful) is that these frequencies match head movements for lethargic and spirited walking, respectively. Moreover, it is not simply that these slower and faster beat rates represent or convey lethargic and spirited activity; as beat rate is echoed in the motor cortex, mirror processes might be expected to evoke feelings of lethargy or energy (respectively) in an engaged listener. In short, a mirror process may explain how fast and slow music succeeds in inducing languid or exuberant feeling states. Although further research is warranted, the induced affects arising from different tempos appear consistent with a mirror process linked to the vestibular system.

9. Vibrato as association

As noted, many associations are idiosyncratic to individual listeners. However, some associations are widespread across a culture—such as the association of *Taps* with funeral or memorial events. Other associations might be deemed 'universal' due to shared experiences. For example, virtually all humans have broadly similar experiences with gravity, breathing, walking, thirst, happiness and other phenomena. One of these stock phenomena is the experience of *fear*.

When in a state of fear, trembling may result. The trembling is attributable to high levels of adrenaline and associated peripheral acetylcholine which increases muscle reactivity. All skeletal muscles are influenced, including the muscles involved in vocal production. Fear or nervousness is often evident in the trembling sound of the voice, with a distinctive modulation of pitch. Although trembling may also be visible in hand motions, trembling is most likely to become evident first, in the voice.

The stylized vocal trembling musicians call *vibrato* can be observed in many musical cultures around the world. Moreover, the same frequency modulation is commonly found in purely instrumental music, where special performance techniques are required in order to emulate the acoustical effects of trembling voice. The effect of vibrato is commonly described as rendering the sound 'more emotional'—consistent with the etiology of trembling voice [55].

It is possible that vocal vibrato represents an ethological signal for which a learned association provides the proximal affect-inducing mechanism. However, the trembling is a straightforward consequence of a general physiological response—and so has the appearance of being artefactual. This suggests that vocal trembling is more likely to be a cue or a learned association.

At this point, it is appropriate to consider how cues might differ from the psychological concept of a learned association. In the first instance, learned associations are mechanisms that may or may not be adaptive, whereas cues (by definition) are necessarily adaptive. Also, cues may or may not be innate. On the one hand, mosquitoes have an innate disposition to fly in the direction of an increasing carbon dioxide gradient. That is, mosquitoes are attracted to the exhaled breath of animals. As the production of CO_2 is an artefact of respiration, it is not likely to represent a functional communicative behaviour. At the same time, mosquitoes do not need to learn to follow a CO_2 trail. In many other cases, simple associative or Pavlovian learning is sufficient to account for adaptive behaviours without appealing to an innate biological mechanism. That is, for ethologists, cues can entail responses that can be either wholly learned or innate. When wholly learned, associations may be regarded as a causal mechanism that enables cues.

At least in Western culture, vibrato is typically slower than fear-induced trembling. Presumably, this connotes a more controlled effect—suggesting high emotionality, without necessarily evoking fear or terror in a listener.

10. Conclusion

In attempting to understand the emotional dimensions of music, how sounds induce emotions in listeners is one of the more difficult questions. This paper has explored several potential emotion-inducing approaches including signals, cues, indices, mirrors and associations. Learned associations may be emotionally compelling, but they are of limited compositional utility unless the associations are widely shared among listeners. As associations can be entirely arbitrary (with no resemblance between the sounds and the evoked memory), they may be useful when accounting for idiosyncratic or paradoxical responses to particular stimuli. Mirrors appear to be important mechanisms that provide a ready explanation for how expressed emotion may be echoed as evoked emotion—with the important caveat that mirrors are restricted to action-oriented behaviours. Signals, by contrast, offer a unique opportunity to induce feeling states that are not simply reflections of the displayed affect. Signals warrant particular interest since, as evolved communicative behaviours, both the displaying behaviour and the observer's response are innate, and so presumably shared across all human listeners. In humans at least, the responses are not the involuntary innate releasing mechanisms envisioned by

early ethologists. Nevertheless, the feeling states evoked by observing a signal do tend to be stereotypic, making them especially attractive tools when composing music intended to induce emotions in listeners. Finally, indices are notably 'honest' signals whose connotations ought to be self-evident to listeners.

The affect-inducing concepts discussed here are not intended to be exhaustive or exclusive. Juslin, for example, has outlined a framework for music-induced emotion that posits eight mechanisms, referred to using the acronym BRECVEMA: brain stem reflexes, rhythmic entrainment, evaluative conditioning, contagion, visual imagery, episodic memory, musical expectancy and aesthetic judgement ([56] revised in [5]). As with signals and cues, the specific physiological/cognitive pathways for each mechanism discussed in this paper await detailed exposition. It may turn out that some of the mechanisms distinguished here share identical neural origins that warrant a single description.

In addition, none of the mechanisms for auditory-evoked affect discussed here should be considered mutually exclusive. For example, learned associations, mirror processes and cognitive appraisals can conceivably coexist, and any given acoustical stimulus might well activate each of these mechanisms concurrently. Parallel systems suggest the possibility of evoking mixed emotions—a state that is commonly reported among listeners [57]. Crying provides an illustrative example. Mirror mechanisms suggest that observing someone crying might lead to an empathetic response in which the observer also feels disposed to cry. However, crying also bears the hallmarks of an ethological signal. Here, the research suggests that crying commonly evokes sympathetic feelings of compassion that motivate the observer to offer assistance (or to terminate aggressive behaviours). As a result, the combination of mirror and signalling processes implies that for many observers, a crying display may evoke simultaneously both empathetic feelings of grief or sadness, mixed with sympathetic feelings of compassion and tenderness. To the extent that a musical passage

emulates crying-related acoustical features (such as breaking voice), we might expect many listeners to experience a similar mixture of grief, sadness, compassion and/or tenderness.

The different mechanisms for auditory-induced emotion distinguished in this paper provide possible starting points for addressing five puzzles in music-related affect: Why is music able to induce only certain emotions (e.g. exuberance, compassion and tenderness), but not others (e.g. jealousy and guilt)? Why are some induced emotions similar to the displayed emotion, whereas in other cases the induced emotion differs from the displayed emotion? Why do listeners often report feeling mixed emotions? Why are some emotional connotations similar across musical cultures, whereas others are not? Why do musicians appear to rely on some emotion-inducing mechanisms more than others?

Particular stress has been placed here on the potential role of ethological signals in music-related emotion. Identifying signals provides a potentially useful strategy for discerning those aspects of musical expression that are likely to be shared across cultures. One useful heuristic in this effort is attending to possible multimodal aspects of purported displays. As signals are intended to be communicative, they tend to be conspicuous, and one way of ensuring conspicuousness is via multimodal displays. For decades, emotion researchers have tended to focus on facial expressions without attending to acoustical concomitants of emotion. Researchers interested in acoustical aspects of emotion would be wise to avoid a similar myopia—by attending to multimodal correlates.

Data accessibility. As a review article, this paper does not present new data that has not already been published. However, the data for three key papers published by the author (and cited in the manuscript) have not been heretofore made available. The data for these three studies will be deposited with the Dryad curated digital repository. Specifically, the studies include: [17,39,40].

Conflict of interests. The author declares that he has no competing interests with regard to employment, sources of funding, personal financial interests or membership of relevant organizations.

References

- Gabrielsson A, Lindström S. 2003 Strong experiences to music: a descriptive system. *Music Sci.* **7**, 157–217.
- Gabrielsson A. 2011 *Strong experiences with music*. Oxford, UK: Oxford University Press.
- Zentner M, Grandjean D, Scherer KR. 2008 Emotions evoked by the sound of music: characterization, classification, and measurement. *Emotion* **8**, 494–521. (doi:10.1037/1528-3542.8.4.494)
- Juslin PN, Laukka P. 2003 Communication of emotions in vocal expression and music performance: different channels, same code? *Psychol. Bull.* **129**, 770–814. (doi:10.1037/0033-2909.129.5.770)
- Juslin PN. 2013 From everyday emotions to aesthetic emotions: towards a unified theory of musical emotions. *Phys. Life Rev.* **10**, 235–266. (doi:10.1016/j.plrev.2013.05.008)
- Molnar-Szakacs I, Overy K. 2006 Music and mirror neurons: from motion to 'e'motion. *Soc. Cogn. Affect. Neurosci.* **13**, 235–241. (doi:10.1093/scan/nsl029)
- Ladinig O, Schellenberg EG. 2012 Liking unfamiliar music: effects of felt emotion and individual differences. *Psychol. Aesthetic Creat. Arts* **6**, 146–154. (doi:10.1037/a0024671)
- Gallese V, Keysers C, Rizzolatti G. 2004 A unifying view of the basis of social cognition. *Trends Cogn. Sci.* **8**, 396–403. (doi:10.1016/j.tics.2004.07.002)
- Singer T, Seymour B, O'Doherty SKE, Kaube H, Dolan RJ. 2004 Empathy for pain involves the affective but not sensory components of pain. *Science* **303**, 1157–1162. (doi:10.1126/science.1093535)
- Carr L, Iacoboni M, Dubeau MC, Mazziotta JC, Lenzi G. 2003 Neural mechanisms of empathy in humans: a relay from neural systems for imitation to limbic areas. *Proc. Natl Acad. Sci. USA* **100**, 5497–5502. (doi:10.1073/pnas.0935845100)
- Gallese V. 2003 The roots of empathy: the shared manifold hypothesis and the neural basis of intersubjectivity. *Psychopathology* **36**, 171–180. (doi:10.1159/000072786)
- Aziz-Zadeh L, Iacoboni M, Zaidel E, Wilson S, Mazziotta J. 2004 Left hemisphere motor facilitation in response to manual action sounds. *Eur. J. Neurosci.* **19**, 2609–2612. (doi:10.1111/j.0953-816X.2004.03348.x)
- Buccino G, Riggio L, Melli G, Binkofski F, Gallese V, Rizzolatti G. 2005 Listening to action-related sentences modulates the activity of the motor system: a combined TMS and behavioral study. *Cogn. Brain Res.* **24**, 355–363. (doi:10.1016/j.cogbrainres.2005.02.020)
- Morton ES. 1977 On the occurrence and significance of motivation-structural rules in some bird and mammal sounds. *Am. Nat.* **111**, 855–869. (doi:10.1086/283219)
- Bolinger DL. 1964 Intonation across languages. In *Universals of human language, vol. 2: phonology* (eds JH Greenberg, CA Ferguson, EA Moravcsik),

- pp. 471–524. Stanford, CA: Stanford University Press.
16. Huron D, Kinney D, Precoda K. 2006 Influence of pitch height on the perception of submissiveness and threat in musical passages. *Empir. Musicol. Rev.* **1**, 170–177.
 17. Shanahan D, Huron D. 2014 Heroes and villains: the relationship between pitch tessitura and sociability of operatic characters. *Empir. Musicol. Rev.* **9**, 46–59.
 18. Fant G. 1960 *Acoustic theory of speech production*. The Hague, The Netherlands: Mouton.
 19. Ohala J. 1980 The acoustic origin of the smile. *J. Acoust. Soc. Am.* **68**, S33. (doi:10.1121/1.2004679)
 20. Tartert VC. 1980 Happy talk: perceptual and acoustic effects of smiling on speech. *Percept. Psychophys.* **27**, 24–27. (doi:10.3758/BF03199901)
 21. Tartert VC, Braun D. 1994 Hearing smiles and frowns in normal and whisper registers. *J. Acoust. Soc. Am.* **96**, 2101–2107. (doi:10.1121/1.410151)
 22. Epplé G. 1967 Vergleichende Untersuchungen über Sexual und Sozialverhalten der Krallenaffen (Hapalidae). *Folia Primatologica* **7**, 37–65. (doi:10.1159/000155095)
 23. Fitch WT. 1994 *Vocal tract length perception and the evolution of language*, PhD dissertation, Brown University, Providence, RI, USA.
 24. Fox MW, Cohen JA. 1977 Canid communication. In *How animals communicate* (ed. TA Sebeok), pp. 728–748. Bloomington, IN: Indiana University Press.
 25. Kaufman JH. 1967 Social relations of adult males in a free-ranging band of rhesus monkeys. In *Social communication among primates* (ed. SA Altmann), pp. 73–98. Chicago, IL: University of Chicago Press.
 26. Miller EH. 1975 A comparative study of facial expressions of two species of pinnipeds. *Behaviour* **53**, 268–284. (doi:10.1163/156853975X00227)
 27. Pruitt CH, Burghardt GM. 1977 Communication in terrestrial carnivores. Mustelidae, Procyonidae and Ursidae. In *How animals communicate* (ed. TA Sebeok), pp. 767–793. Bloomington, IN: Indiana University Press.
 28. Lorenz K. 1939 Vergleichende Verhaltensforschung. *Zool. Anzeiger Suppl.* **12**, 69–102.
 29. Lorenz K. 1970 *Studies in animal and human behaviour*, vol. 1. London, UK: Methuen.
 30. Maynard Smith JM, Harper D. 2003 *Animal signals*. Oxford, UK: Oxford University Press.
 31. Silk JB, Kaldor E, Boyd R. 2000 Cheap talk when interests conflict. *Anim. Behav.* **59**, 423–432. (doi:10.1006/anbe.1999.1312)
 32. Tomkins SS. 1980 Affect as amplification: some modifications in theory. In *Emotion: theory, research and experience* (eds R Plutchik, H Kellerman), pp. 141–164. New York, NY: Academic Press.
 33. Gelstein S, Yeshurun Y, Rozenkrantz L, Shushan S, Frumin I, Roth Y, Sobel N. 2011 Human tears contain a chemosignal. *Science* **331**, 226–230. (doi:10.1126/science.1198331)
 34. Kottler JA. 1996 *The language of tears*. San Francisco, CA: Jossey-Bass.
 35. Lane CJ. 2006 *Evolution of gender differences in adult crying*, PhD dissertation, University of Texas at Arlington, Arlington, TX, USA.
 36. Wiley RH. 1983 The evolution of communication: information and manipulation. In *Communication* (eds TR Halliday, PJB Slater), pp. 82–113. Oxford, UK: Blackwell.
 37. Johnstone RA. 1997 The evolution of animal signals. In *Behavioural ecology* (eds JR Krebs, NB Davies), pp. 155–178. Oxford, UK: Oxford University Press.
 38. Bauer HR. 1987 Frequency code: orofacial correlates of fundamental frequency. *Phonetica* **44**, 173–191. (doi:10.1159/000261793)
 39. Huron D, Dahl S, Johnson R. 2009 Facial expression and vocal pitch height: evidence of an intermodal association. *Empir. Musicol. Rev.* **4**, 93–100.
 40. Huron D, Shanahan D. 2013 Eyebrow movements and vocal pitch height: evidence consistent with an ethological signal. *J. Acoust. Soc. Am.* **133**, 2947–2952. (doi:10.1121/1.4798801)
 41. Ekman P, Friesen WV. 1986 A new pan-cultural facial expression of emotion. *Motiv. Emot.* **10**, 159–168. (doi:10.1007/BF00992253)
 42. Ekman P, Heider KG. 1988 The universality of a contempt expression: a replication. *Motiv. Emot.* **12**, 303–308. (doi:10.1007/BF00993116)
 43. Kelly DR. 2011 *Yuck! The nature and moral significance of disgust*. Cambridge, MA: MIT Press.
 44. Ladefoged PN. 1962 *Elements of acoustic phonetics*. Chicago, IL: University of Chicago Press.
 45. Plazak J. 2011 *Encoding and decoding sarcasm in instrumental music: a comparative study*, PhD dissertation, Ohio State University, Columbus, OH, USA.
 46. Darwin C. 1872 *The expression of the emotions in man and animals*. London, UK: John Murray.
 47. Huron D. 2012 Understanding music-related emotion: lessons from ethology. In *Proc. 12th Int. Conf. on Music Perception and Cognition* (eds E Cambouropoulos, C Tsougras, P Mavromatis, K Pasiadis), pp. 473–481. Thessaloniki, Greece: ESCOM.
 48. Paul B, Huron D. 2010 An association between breaking voice and grief-related lyrics in country music. *Empir. Musicol. Rev.* **5**, 27–35.
 49. Vorperian HK, Kent RD, Lindstrom MJ, Kalina CM, Gentry LR, Yandell BS. 2005 Development of vocal tract length during early childhood: a magnetic resonance imaging study. *J. Acoust. Soc. Am.* **117**, 338–350. (doi:10.1121/1.1835958)
 50. Fitch WT. 1997 Vocal tract length and formant frequency dispersion correlate with body size in rhesus macaques. *J. Acoust. Soc. Am.* **102**, 1213–1222. (doi:10.1121/1.421048)
 51. Maynard Smith J, Harper D. 1995 Animal signals: models and terminology. *J. Theor. Biol.* **177**, 305–311. (doi:10.1006/jtbi.1995.0248)
 52. Laird JD. 2007 *Feelings: the perception of self*. Oxford, UK: Oxford University Press.
 53. Hatfield E, Cacioppo JT, Rapson RL. 1994 *Emotional contagion*. Cambridge, UK: Cambridge University Press.
 54. MacDougall HG, Moore ST. 2005 Marching to the beat of the same drummer: the spontaneous tempo of human locomotion. *J. Appl. Physiol.* **99**, 1164–1173. (doi:10.1152/jappphysiol.00138.2005)
 55. Seashore CE. 1932 *The vibrato*. Iowa City, IA: University of Iowa Press.
 56. Juslin PN, Västfjäll D. 2008 Emotional responses to music: the need to consider underlying mechanisms. *Behav. Brain. Sci.* **31**, 559–621. (doi:10.1017/S0140525X08005293)
 57. Hunter PG, Schellenberg EG, Schimmack U. 2010 Feelings and perceptions of happiness and sadness induced by music: similarities, differences, and mixed emotions. *Psychol. Aesthet. Creat. Arts* **4**, 47–56. (doi:10.1037/a0016873)