Human emotions track changes in the acoustic environment

Weiyi Ma (马维毅)1,2,3 and William Forde Thompson4,5,6

*Australian Research Council Centre of Excellence in Cognition and its Disorders, Macquarie University, Sydney, NSW 2109, Australia; 5School of Linguistics and Literature, University of Electronic Science and Technology of China, Chengdu 610061, China; and 6Department of Psychology, Macquarie University, Sydney, NSW 2109, Australia

Edited by Dale Purves, Duke University, Durham, NC, and approved October 13, 2015 (received for review August 11, 2015)

Emotional responses to biologically significant events are essential for human survival. Do human emotions lawfully track changes in the acoustic environment? Here we report that changes in acoustic attributes that are well known to interact with human emotions in speech and music also trigger systematic emotional responses when they occur in environmental sounds, including sounds of human actions, animal calls, machinery, or natural phenomena, such as wind and rain. Three changes in acoustic attributes known to signal emotional states in speech and music were imposed upon 24 environmental sounds. Evaluations of stimuli indicated that human emotions track such changes in environmental sounds just as they do for speech and music. Such changes not only influenced evaluations of the sounds themselves, they also affected the way accompanying facial expressions were interpreted emotionally. The findings illustrate that human emotions are highly attuned to changes in the acoustic environment, and reiterate a discussion of Charles Darwin’s hypothesis that speech and music originated from a common emotional signal system based on the imitation and modification of environmental sounds.

Emotional responses to environmental events are essential for human survival. In contexts that have implications for survival and reproduction, the amygdala transmits signals to the hypothalamus, which releases hormones that activate the autonomic nervous system and cause physiological changes, such as increased heart rate, respiration, and blood pressure (1). These bodily changes contribute to the experience of emotion (2), and function to prepare an organism to respond effectively to biologically significant events in the environment (3).

Throughout the arts and media, environmental conditions have been used to connote an emotional character. For example, the acoustic soundscape of film and television can powerfully affect a viewer’s perspectives on the narrative (4). Thus, human emotions appear to track changes in the acoustic environment, but it is unclear how they do this. One possibility is that the acoustic attributes that convey emotional states in speech and music also trigger emotional responses in environmental sounds. This possibility is implied within Charles Darwin’s theory that speech and music originated from a common precursor that developed from “the imitation and modification of various natural sounds, the voices of other animals, and man’s own instinctive cries” (5). Darwin also argued that this primitive system would have been especially useful in the expression of emotion. Modern day music, he reasoned, was a behavioral remnant of this early system of communication (5, 6).

This hypothesis has been elaborated and restated by modern researchers as the “musical protolanguage hypothesis”: speech and music share a common ancestral precursor of a songlike communication system (or musical protolanguage) used in courtship and territory and in the expression of emotion, which is based on the imitation and modification of environmental sounds (6–10). Environmental sounds carry biologically significant information reflected in our emotional responses to such sounds. To express an emotional state, early hominins might have selectively imitated and manipulated abstract attributes of environmental sounds that have broad biological significance, vocally modulating pitch, intensity, and rate while disregarding the attributes of sound that are specific to individual sources. Extracting and transposing biologically significant cues in the environment to contexts beyond their original source allowed a new channel of emotional communication to emerge (11–14).

The musical protolanguage hypothesis is supported by recent evidence that speech and music share underlying cognitive and neural resources (15–22), and draw on a common code of acoustic attributes when used to communicate emotional states (23–31). In their review of emotional expression in speech and music, Justin and Laukkka found that higher pitch, increased intensity, and faster rate were associated with more excited and positive emotions in both speech and music (23). More recently, it has been demonstrated that the spectra associated with certain major and minor intervals are similar to the spectra of excited and subdued speech, respectively (26, 27), a finding corroborated in acoustic analyses of South Indian music and speech (28). Furthermore, deficits in music processing are associated with reduced sensitivity to emotional speech prosody (32), whereas enhancements of the capacity to process music are correlated with improved sensitivity to emotional speech prosody (33, 34). For example, a study on individuals with congenital amusia, a neurodevelopmental disorder characterized by deficits in processing acoustic and structural attributes of music, showed that amusic individuals were worse than matched controls at decoding emotional prosody in speech, supporting speculations that music and language share mechanisms that trigger emotional responses to acoustic attributes (32).

Significance

Emotions function to optimize adaptive responses to biologically significant events. In the auditory channel, humans are highly attuned to emotional signals in speech and music that arise from shifts in the frequency spectrum, intensity, and rate of acoustic information. We found that changes in acoustic attributes that evoke emotional responses in speech and music also trigger emotions when perceived in environmental sounds, including sounds arising from human actions, animal calls, machinery, or natural phenomena, such as wind and rain. The findings align with Darwin’s hypothesis that speech and music originated from a common emotional signal system based on the imitation and modification of sounds in the environment.

Author contributions: W.M. and W.F.T. designed research; W.M. performed research; W.M. and W.F.T. analyzed data; and W.M. and W.F.T. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

1W.M. and W.F.T. contributed equally to this work.

2To whom correspondence may be addressed. Email: weiyi.ma@mq.edu.au or Bill.Thompson@mq.edu.au.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1515087112/-/DCSupplemental.

www.pnas.org/cgi/doi/10.1073/pnas.1515087112

PNAS | November 24, 2015 | vol. 112 | no. 47 | 14563–14568
Changes in three acoustic attributes are especially important for communicating emotion in speech and music: frequency spectrum, intensity, and rate (23–25). Darwin’s hypothesis implies that these attributes are tracked by human emotions because they reflect biologically significant information about sound sources, such as their size, proximity, and speed. More specifically, the musical protolanguage hypothesis predicts that acoustic attributes that influence the emotional character of speech and music should also have emotional significance when arising from environmental sounds (5).

The present study tested the hypothesis that changes in the frequency spectrum, intensity, and rate of environmental sounds are associated with changes in the perceived valence and arousal of those sounds (23–25). Because the sources and nature of environmental sounds vary considerably according to geographic location, environmental sounds are defined as any acoustic stimuli that can be heard in daily life that are neither musical nor linguistic. Thus, four types of environmental sounds were considered (human actions, animal sounds, machine noise, sounds in nature), each containing six exemplars. For each of these 24 environmental sounds, we manipulated the frequency spectrum, intensity, and rate. In accordance with the circumplex model of emotion, we obtained ratings of the perceived difference in valence (negative to positive) and arousal (calm to energetic) for stimulus pairs that differed in just one of the three manipulated attributes (35, 36). Although not all environmental sounds have a clearly perceptible fundamental frequency, research on pitch sensations for nonperiodic sounds confirm that individuals are sensitive to salient spectral regions and can detect when such regions are shifted (37, 38).

Results

Two preliminary stimulus verification tests confirmed that all manipulation-types resulted in highly discriminable stimuli, and none of the manipulations had a significant effect on the perceived naturalness of stimuli (Materials and Methods). Next, the three types of manipulation (frequency spectrum, intensity, rate) were examined separately in Exps. 1a–1c, respectively, to minimize fatigue effects and to increase response reliability. Exp. 1a focused on the frequency spectrum of stimuli. For each stimulus, listeners rated the difference in valence and arousal between the high- and low-frequency versions. The same procedure was adopted for manipulations of intensity (Exp. 1b) and rate (Exp. 1c).

Experiment 1a (Frequency Spectrum). Valence and arousal ratings were subjected to separate two-by-four ANOVAs with repeated measures on frequency spectrum height (high, low) and sound-type (human actions, animals, machinery, natural phenomena). We observed a main effect of frequency spectrum height for ratings of valence \([F(1, 49) = 88.37, P < 0.001, \eta^2_p = 0.64]\) and arousal \([F(1, 49) = 129.76, P < 0.001, \eta^2_p = 0.73]\), with a higher frequency spectrum associated with more positive valence and greater arousal (Fig. 1). For valence ratings, we also observed a

\[F(1, 49) = 28.96, P < 0.001, \eta^2_p = 0.37\] for the sound-type factor, with human actions rated as more positive than the other sound types, which did not differ.

For arousal ratings, the sound-type factor was not significant \([F(1, 49) = 1.14, P = 0.29]\), but there was a significant interaction between frequency spectrum height and sound-type \([F(1, 49) = 3.15, P = 0.08, \eta^2_p = 0.06]\), indicating that the effect of frequency spectrum height on arousal varied across sound types.

![Fig. 1. Mean valence (A) and arousal (B) ratings for frequency spectrum, intensity, and rate manipulations. For each manipulation-type, the mean ratings (and SEs) for the increased and decreased stimuli for each sound-type are displayed relative to mean ratings for control stimuli. The ratings of rate manipulations are averaged across long and short versions. Separate paired-samples t tests showed that both valence and arousal ratings significantly decreased from increased stimuli to controls then to decreased stimuli within each type of manipulation.](www.pnas.org/cgi/doi/10.1073/pnas.1515087112)
significant main effect of sound-type \[ F(3, 147) = 5.29, P = 0.002, \eta^2 = 0.10 \] with higher ratings associated with human actions and lower ratings associated with animals, machinery, and natural phenomena. The main effect of sound-type was not significant for arousal ratings \( (P = 0.28) \).

**Experiment 1b (Intensity).** As in Exp. 1a, valence and arousal ratings were subjected to separate two-by-four ANOVAs with repeated measures on intensity (loud, soft) and sound-type. We found a main effect of intensity for ratings of valence \( F(1, 49) = 82.90, P < 0.001, \eta^2 = 0.63 \) and arousal \( F(1, 49) = 125.90, P < 0.001, \eta^2 = 0.72 \), with louder sounds associated with more positive valence and greater energy relative to softer sounds (Fig. 1). There were no main effects of sound-type for either valence \( (P = 0.07) \) or arousal \( (P = 0.37) \) ratings.

**Experiment 1c (Rate).** Because rate manipulations alter the duration of stimuli as well, the pairs of rate-manipulated stimuli always differed in both rate and duration. To address this confound, we created long (7–7.5 s) and short versions (5 s) of each original stimulus before manipulating its rate. This procedure allowed us to separate the effects of duration and rate on the emotional character of sounds. Long and short versions of the original stimuli were subjected to rate manipulations. Separate two-by-two-by-four ANOVAs were conducted with repeated measures on duration (long, short), rate (fast, slow), and sound-type. There were no reliable effects of duration on ratings of valence \( (P = 0.89) \) or arousal \( (P = 0.69) \). However, there was a main effect of rate for both valence \( F(1, 49) = 136.39, P < 0.001, \eta^2 = 0.74 \) and arousal \( F(1, 49) = 88.43, P < 0.001, \eta^2 = 0.64 \) with the fast rate associated with more positive valence and greater arousal, relative to the slow rate (Fig. 1). We also found a main effect of sound-type for ratings of valence \( F(3, 147) = 11.13, P < 0.001, \eta^2 = 0.19 \) and arousal \( F(3, 147) = 6.72, P < 0.001, \eta^2 = 0.12 \), with the highest ratings associated with human actions and lower ratings associated with other sounds.

These findings confirm a close correspondence between attributes that carry emotional information in speech and music and the attributes that carry emotional information in environmental sounds. The acoustic attributes (higher frequency spectrum, increased intensity, and faster rate), which have been associated with valence and arousal in speech and music (23–25), are also tracked by human emotions when they occur in environmental sounds. Frequency, intensity, and rate may interact with human emotions because they carry biologically significant information about the size, proximity, and degree of energy of a sound source. For example, larger species tend to produce lower-pitched calls than smaller species (39); intensity is inversely related to the distance (40) and positively related to the body size and muscle power (41) of the sound source; rate is inherently determined by the speed of motion.

Exp. 2 was designed to validate this result by asking participants to rate the emotional connotation of sounds presented in isolation rather than in pairs. This design addressed the possibility that our acoustic manipulations only resulted in differences in emotional connotation because contrasting sounds were presented successively, whereas ratings of isolated sounds might not yield reliable effects of our manipulations.

The three types of manipulations (frequency spectrum, intensity, rate) were re-examined in Exps. 2a–2c, respectively. Another group of participants completed the emotional rating task. On each trial, listeners heard a single sound and were asked to rate its valence and arousal level. Consistent with Exp. 1, Exp. 2 showed that sounds with higher frequency spectrum, intensity, and rate were rated as more positive and energetic, relative to sounds with lower frequency, intensity, and rate (SI Materials and Methods and Fig. S1).

In the first two experiments, participants directly judged the emotional connotations of changes in environmental sounds. Exp. 3 tested the emotional consequences of changes in environmental sounds without asking participants to evaluate the sounds themselves. If human emotions automatically track the acoustic environment, then the presence of environmental sounds should affect emotional judgments in other domains (e.g., vision).

A subset \( (n = 8) \) of environmental sounds was used. On each trial, participants were presented with a change in an environmental sound (e.g., an increased sound followed by a decreased sound) along with a change in visual stimuli (e.g., a happy face followed by a neutral face) and were instructed to judge as quickly as possible the emotional change implied by the visual stimuli. Thus, the acoustic channel was irrelevant to the task, allowing us to evaluate any interference that it might have on judgments of the visual channel. We predicted that if the acoustic stimuli changed in a manner that was congruent with the visual change, then classification of the visual change should be rapid; if it was incongruent, however, then classification of the visual change should be slower.

Average reaction times were calculated for the congruent and incongruent trials for frequency spectrum, intensity, and rate manipulations within separate valence and arousal blocks. Reaction times were subjected to two-by-three ANOVAs with repeated measures on congruity (congruent, incongruent) and manipulation-type (frequency spectrum, intensity, rate) within each block. We observed main effects of congruity in both valence \( F(1, 39) = 10.40, P = 0.003, \eta^2 = 0.21 \) and arousal \( F(1, 39) = 7.98, P = 0.007, \eta^2 = 0.17 \) blocks with congruent trials associated with faster reaction time (Fig. S2). Thus, the mere presence of changes in (irrelevant) background sounds affected the emotional decoding of facial expressions by participants, suggesting that human emotions continuously and automatically track changes in the acoustic environment. The results also suggest that the acoustic environment can shape our visual perception of emotion: our interpretation of what we see is affected by what we hear.

**Discussion**

This investigation demonstrates that human emotions systematically track changes in the acoustic environment, affecting not only how we experience those sounds but also how we perceive facial expressions in other people. Such effects are consistent with Darwin’s musical protolanguage hypothesis, which posits a transition from emotional communication based on the imitation of environmental sounds to the evolution of speech and music. In contemporary theories that draw from Darwin’s insights, the capacity to imitate and modify sensory input is thought to have been crucial for the evolution of human cognition (8, 42). Contemporary theories of cognitive evolution posit a sequence of critical transitions (8, 42), including a transition in which a concrete, time-bound representation of the environment evolved into an abstract representation by extracting key features from the environment. An abstract representation provided individuals with an understanding that sensory attributes were not tied to specific environments but had significance independently of circumstances. This transition would have made it possible for individuals to communicate the meaning of stimulus attributes in novel contexts and channels of communication, including vocalizations. This evolutionary stage has been referred to as mimic mimetic cognition (42), and is thought to have mediated the transition that led to human cognitive capacities.

According to this account, mimesis represents a generative and intentional form of communication that can occur in facial expressions, gestures, and vocalizations based on the abstract features of communicative meanings (42). By using vocalizations to imitate acoustic changes in the environment that had biological significance, early hominins could convey environmental conditions to other individuals, communicating second-hand knowledge that could be acted upon adaptively. It is possible that emotive
vocalizations gradually came to stand for complex states of conspecifics. Thus, emotional mimesis may have allowed early hominins to share biologically significant information efficiently. Human memory eventually became inadequate for storing and processing our accumulating collective knowledge, creating the need for a more efficient and effective communication system. This led to the next transition: the invention of language, revolutionary for the evolution of human cognition (7, 8, 12, 43). Humans started to construct elaborate symbolic systems, ranging from cuneiforms, hieroglyphics, and ideograms to alphabetic languages and mathematics (44).

At the same time, it has been suggested that protolanguage was insufficient to convey the range of complex emotions associated with large social groups (8). This furthered the separation of speech and music, eventually leading to a fully developed musical system for expressing emotions efficiently and effectively, facilitating group coordination, signaling fitness and creativity, and nurturing social bonds (6, 8). Of course, hypotheses on the precise sequence of evolutionary transitions leading to speech and music are necessarily speculative, given that the early vocalizations that preceded speech and music left no “bones” (45), and there is scant paleontological evidence with which to trace their emergence and subsequent development. What is important for the purposes of this investigation is that the present results are consistent with Darwin’s hypothesis (8) that vestiges of a musical protolanguage (5) should be evident in the emotional parallels between speech, music, and environmental sounds.

Darwin’s musical protolanguage hypothesis can account for the present results, but it should be acknowledged that there are other possible interpretations. For example, it is plausible that emotional responses to a primitive vocalization system are primary and emotional responses to music and environmental sounds derive from this emotional code. Although it is not possible to adjudicate between competing interpretations based on the evidence to date, it is relevant to note that evolution promotes development in the direction toward selective advantage. Thus, it is reasonable to suggest that the capacity to track changes in the acoustic environment evolved before the development of a vocalization system for emotional communication (e.g., protolanguage).

Regardless of the evolutionary implications of the effect, the findings illustrate the emotional power of environmental sounds on both our experience of sounds and our evaluations of acoustic stimuli (separated by a 2-s pause) were the same or different (0 = no difference, 1 = subtle difference, 2 = obvious difference). The task consisted of four blocks, each testing one manipulation-type (frequency spectrum, intensity, rate-long, rate-short). Each block had 30 trials, consisting of 24 tests where an increased and decreased versions of a stimulus were presented (six increased and six decreased trials for each sound-type). The presentation order of the four blocks was counterbalanced across participants and the trial order within a block was randomized. The task lasted ~30 min. All acoustic manipulation resulted in highly discriminable stimuli (Table S2).

Another group of 20 participants rated the naturalness of stimuli presented individually (1 = unnatural, 4 = moderately natural, 7 = completely natural). The task consisted of four blocks, each testing one manipulation-type (frequency spectrum, intensity, rate-long, rate-short). Each block had 48 trials (six increased and six decreased trials for each sound-type). The presentation order of the four blocks was counterbalanced across participants and the trial order within a block was randomized. The task lasted ~40 min. For each participant, average ratings of increased and decreased stimuli were calculated within each type of manipulation to confirm that they sounded equally natural, as any differences in naturalness would confound interpretation of our primary experiment on the emotional consequences of the manipulations. Separate two- (manipulation: increased, decreased) by-four (sound-type) repeated-measures ANOVAs were conducted to assess the effect of each type of manipulation on naturalness. The main effect of manipulation was not significant for frequency spectrum (P = 0.20), intensity (P = 0.32), or rate (long version: P = 0.79; short version: P = 0.23) manipulation. The effect of manipulation decreased when the manipulation was not directed at perceptual naturalness, and that perceived naturalness was unaffected by our manipulations (SI Materials and Methods and Table S3).

**Experiment 1.**
**Participants.** Fifty participants (M = 23.32 y, SD = 2.29; 27 males) completed an emotional rating task.
**Procedure.** The effects of the three acoustic manipulations (frequency spectrum, intensity, rate) were examined in Exp. 1a–1c, respectively. In Exp. 1a (frequency spectrum), there were 24 experimental trials (six for each type of sound: human actions, animals, machinery, natural phenomena) and 6 control trials (two identical sounds, selected randomly from the set of unmanipulated sounds). On each trial, participants heard increased and decreased versions of each exemplar presented consecutively and separated by a 2-s pause. Participants then rated the difference in emotional character between the two versions. In the experimental trials, high- and low-frequency versions of a sound were presented: half with the high-frequency spectrum version presented first (three trials per sound-type) and half with the low-frequency spectrum version presented first. Twenty-five participants rated the first sound in each pair relative to the second; the other 25 participants rated the second sound relative to the first. The order of the three part-experiments was counterbalanced across participants, and the order of trials within each part-experiment was randomized independently for each participant. The same procedure was adopted for manipulations of intensity (Exp. 1b) and rate (Exp. 1c).
Participants were seated in a sound-attenuated booth, and given a demonstration of the task based on an instruction written in Chinese: “Sometimes our environment suggests a mood or emotion. This connection is often seen in films and story-telling, which often describes weather conditions, and details of the environment. In this study, we are examining the possibility that sounds we hear in our environment can sometimes suggest an emotional tone or mood. Your task is to compare the two sounds and rate the emotional connotation of the second sound in relation to the first by using two scales: one ranging from more negative to more positive (1 = more negative, 4 = no change, 7 = more positive); and the other ranging from more calm to energetic (1 = more calm, 4 = no change, 7 = more energetic). To help you do the task, you might imagine that you are a film director, and your sound editor has introduced a series of environmental sounds in order to create an overall mood for the film. Imagine that you are giving your sound editor feedback about her choice of sounds, and the moods that they evoke.” The auditory stimuli were binaurally presented via headphones. The sound pressure level remained constant across participants. The task lasted ~40 min.

Experiment 2. Participants. Fifty UESTC students (M = 23.10 y, SD = 1.58; 39 males) completed an emotional rating task.

Procedure. The procedure was similar to Exp. 1 but participants only heard one stimulus on a trial, and the middle point of the valence and arousal scales were revised to be “neutral”. Exps. 2a and 2b contained 48 trials (six increased and six decreased trials for the four types of sounds); Exp. 2c contained 96 trials (short and long exemplars of the fast and slow versions of each type of sound). On each trial, listeners heard a single sound and were asked to rate each pole of the 2D model of emotion: valence (1 = more negative, 4 = neutral, 7 = more positive), and arousal (1 = more calm, 4 = neutral, 7 = more energetic). The order of the three-part experiments was counterbalanced across participants, and the order of trials within each part-experiment was randomized independently for each participant. The experiment lasted ~45 min. For each participant, we calculated the average ratings of the increased and decreased stimuli for each sound-type and manipulation.

Experiment 3. Participants. Forty UESTC students (M = 23.70 y, SD = 1.37; 27 males) completed a visual emotional judgment task while the environmental sounds were presented at the same time as the irrelevant signal.

Stimuli preparation.

Visual stimuli. Facial expressions of 58 actors/actresses were selected from the Radboud Faces Database (49). For each actor/actress, three facial expressions (happy, neutral, surprised) recorded from the frontal angle were used (Fig. S3).

Auditory stimuli. Eight environmental sounds, two from each sound-type (clapping, typing, cat, horse, car engine, train, thunder, waves), were used. Because Exps. 1 and 2 showed that duration did not confound emotional perception for rate manipulations, only the long version was used for rate manipulations. Thus, each sound had six versions, manipulated in either the frequency spectrum (high, low), or intensity (loud, soft), or rate (fast, slow).

Procedure. An experiment consisted of a valence block and an arousal block, each containing 48 experimental trials and 10 control trials. In the valence block, participants saw an experimental trial two facial expressions (happy, neutral) of an actor/actress presented sequentially on a computer monitor, with the order of the three part-experiments was counterbalanced across participants, and the order of trials within each part-experiment was randomized independently for each participant. The experiment contained 116 trials and lasted ~40 min. The same set of actors/actresses was used in the two blocks.

To ensure reliable reaction time data, responses that took longer than 4 s were excluded from data analysis. A 4-s limit was chosen because it gave ample time to reliably judge the appearance of a visual stimulus (50). The exclusion criterion resulted into 4.8% of responses being excluded across participants.

ACKNOWLEDGMENTS. We thank Diankun Gong for assistance in data collection at the University of Electronic Science and Technology of China. The Australian Research Council supported this research through a research fellowship (W.M.) in the Australian Research Centre Centre of Excellence in Cognition and its Disorders, and through Discovery Grant DP130101084 (to W.T.F.); W.M. is also supported by Chinese Ministry of Education Social Science Funding (11YJC880079) and Chinese National Educational Research Key Project (GPA115005).
20. Musacchia G, Sams M, Skoe E, Kraus N (2007) Musicians have enhanced subcortical auditory and audiovisual processing of speech and music. Proc Natl Acad Sci USA 104(40):15894–15898.

21. Besson M, Chobert J, Marie C (2011) Transfer of training between music and speech: Common processing, attention, and memory. Front Psychol 2:94.

22. Patel AD (2003) Language, music, syntax and the brain. Nat Neurosci 6(7):674–681.

23. Juslin PN, Laukka P (2003) Communication of emotions in vocal expression and music performance: Different channels, same code? Psychol Bull 129(5):770–814.

24. Ilie G, Thompson WF (2006) A comparison of acoustic cues in music and speech for three dimensions of affect. Music Percept 22(4):319–329.

25. Ilie G, Thompson WF (2011) Experiential and cognitive changes following seven minutes exposure to music and speech. Music Percept 28(3):247–264.

26. Bowling DL, Gill K, Choi JD, Prinz J, Purves D (2010) Major and minor music compared to excited and subdued speech. J Acoust Soc Am 127(3):491–503.

27. Curtis ME, Bharucha JJ (2010) The minor third communicates sadness in speech, mirroring its use in music. Emotion 10(3):335–348.

28. Bowling DL, Sundararajan J, Han S, Purves D (2012) Expression of emotion in Eastern and Western music mirrors vocalization. PLoS One 7(3):e31942.

29. Quinto L, Thompson WF, Keating FL (2013) Emotional communication in speech and music: The role of melodic and rhythmic contrasts. Front Psychol 4:184.

30. Huron D (2008) Science & music: Lost in music. Nature 453(7194):456–457.

31. Hauser MD, McDermott J (2003) The evolution of the music faculty: A comparative perspective. Nat Neurosci 6(7):663–668.

32. Thompson WF, Marin MM, Stewart L (2012) Reduced sensitivity to emotional prosody in congenital amusia rekindles the musical protolanguage hypothesis. Proc Natl Acad Sci USA 109(46):19027–19032.

33. Thompson WF, Schellenberg EG, Husain G (2004) Decoding speech prosody: Do music lessons help? Emotion 4(1):46–64.

34. Lima CF, Castro SL (2011) Speaking to the trained ear: musical expertise enhances the recognition of emotions in speech prosody. Emotion 11(5):1021–1031.

35. Russell JA (1980) A circumplex model of affect. J Pers Soc Psychol 39(6):1161–1178.

36. Posner J, Russell JA, Peterson BS (2005) The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology. Dev Psychopathol 17(3):715–734.

37. Howard DM (2014) Pitch, models of. Music in the Social and Behavioral Sciences: An Encyclopedia, ed Thompson WF (SAGE Publication), pp 867–886.

38. Schnupp J, Nelken I, King A (2010) Periodicity and pitch perception: Physics, psychophysics, and neural mechanisms. Auditory Neuroscience: Making Sense of Sound, eds Schnupp J, Nelken I, King A (The MIT Press, Cambridge, MA), pp 93–138.

39. Hauser MD (1993) The evolution of nonhuman primate vocalizations: Effects of phylogeny, body weight, and social context. Am Nat 142(3):528–542.

40. Marten K, Marler P (1977) Sound transmission and its significance for animal vocalization. Behav Ecol Sociobiol 2(3):271–290.

41. Bennet-Clark HC (1998) Size and scale effects as constraints in insect sound communication. Philos Trans R Soc Lond B Biol Sci 353(1367):407–419.

42. Donald M (1991) Origins of the Modern Mind: Three Stages in the Evolution of Culture and Cognition (Harvard Univ Press, Cambridge, MA).

43. Christiansen MH, Kirby S (2003) Language evolution: The hardest problem in science? Language Evolution, eds Christiansen M, Kirby S (Oxford Univ Press, NY), pp 77–93.

44. Daves S (2012) The Artful Species: Aesthetics, Art, and Evolution (Oxford Univ Press, Oxford).

45. Thompson WF (2014) Music Thought & Feeling: Understanding the Psychology of Music (Oxford Univ Press, New York), 2nd Ed.

46. Weingner F, Eyben F, Schuller BW, Mortillaro M, Scherer KR (2013) On the acoustics of emotion in audio: What speech, music and sound have in common. Front Psychol 4:292.

47. Trainer L (2008) Science & music: The neural roots of music. Nature 453(7194):598–599.

48. Zung WWK (1965) A self-rating depression scale. Arch Gen Psychiatry 12(1):63–70.

49. Langner O, et al. (2010) Presentation and validation of the Radboud Faces Database. Cogn Emotion 24(8):1377–1388.

50. Barbalat G, Bazargani N, Blakemore SJ (2013) The influence of prior expectations on emotional face perception in adolescence. Cereb Cortex 23(7):1542–1551.