Cognition 154 (2016) 174–181

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/COGNIT

Original Articles

The source dilemma hypothesis: Perceptual uncertainty contributes to musical emotion

Tanor L. Bonin[a], Laurel J. Trainor[a,b,c], Michel Belyk[a], Paul W. Andrews[a,*]

[a] Department of Psychology, Neuroscience and Behaviour, McMaster University, 1280 Main Street West, Hamilton, Ontario L8S 4K1, Canada
[b] McMaster Institute for Music and the Mind, 1280 Main Street West, Hamilton, Ontario L8S 4K1, Canada
[c] Rotman Research Institute, Baycrest Hospital, Toronto, Ontario, Canada

ARTICLE INFO

Article history:
Received 23 September 2015
Revised 25 May 2016
Accepted 30 May 2016

Keywords:
Auditory scene analysis
Dissonance
Emotion
Evolution
Music
Perceptual uncertainty

ABSTRACT

Music can evoke powerful emotions in listeners. Here we provide the first empirical evidence that the principles of auditory scene analysis and evolutionary theories of emotion are critical to a comprehensive theory of musical emotion. We interpret these data in light of a theoretical framework termed "the source dilemma hypothesis," which predicts that uncertainty in the number, identity or location of sound objects elicits unpleasant emotions by presenting the auditory system with an incoherent percept, thereby motivating listeners to resolve the auditory ambiguity. We describe two experiments in which source location and timbre were manipulated to change uncertainty in the auditory scene. In both experiments, listeners rated tonal and atonal melodies with congruent auditory scene cues as more pleasant than melodies with incongruent auditory scene cues. These data suggest that music's emotive capacity relies in part on the perceptual uncertainty it produces regarding the auditory scene.

1. Introduction

Considerable research has focused on the induction of emotion through music (Juslin & Västfjäll, 2008). This research has engendered disagreement on the extent to which music evokes common emotions across individuals (Koelsch, 2005; Sloboda, 1996), whether musical emotions are similar to emotions evoked in other circumstances (Peretz, 2001; Trainor & Schmidt, 2003), and which specific emotions music is capable of evoking (Panksepp & Bernatzky, 2002). Despite such debates, everyday listeners cite their primary motivation for engaging with music as being the emotional responses that the music evokes (Juslin & Västfjäll, 2008). Western listeners report routinely employing music to induce desired emotions, to complement or change their current emotional state, to provide comfort, and to release stress (Behne, 1997; Gabrielsson, 2001; Juslin & Laukka, 2004; Sloboda & O'Neill, 2001; Zillmann & Gan, 1997).

Musically evoked emotions bear the physiological markers that accompany emotions evoked by other means (Trainor & Schmidt, 2003). Emotional responses to music can elicit physiological responses such as tears, shivering, lump in the throat, and chills (Sloboda, 1991; Sloboda, 2005), and changes in galvanic skin response, breathing rate, blood flow and heart rate (Krumhansl, 1997; Lundqvist, Carlsson, Hilmersson, & Juslin, 2009; Nyklíček, Thayer, & Van Doornen, 1997; Rickard, 2004). In addition, neuroimaging data indicate that music engages dopaminergic pathways and reward centers in the brain (Blood, Zatorre, Bermudez, & Evans, 1999; Koelsch, 2014; Mitterschiffthaler, Fu, Dalton, Andrew, & Williams, 2007; Salimpoor, Benovoy, Larcher, Dagher, & Zatorre, 2011). These neurophysiological responses strongly suggest that music engages the same emotion circuitry as other emotional precipitants. Behavioral studies indicate that there exists considerable agreement across listeners as to which emotion is being expressed in a particular piece of music (Gabrielsson & Lindström, 2001; Hevner, 1936; Terwogt & Van Grinsven, 1991), including to some extent listeners from different cultures (Balkwill & Thompson, 1999; Balkwill, Thompson, & Matsunaga, 2004; Fritz et al., 2009). Such evidence suggests that musical emotion relies at least partly, if not wholly, on the more general mechanisms involved in the production of emotion.

According to evolutionary theory, the biological mechanisms that produce emotion are ancient and evolved to help organisms find adaptive solutions to problems in the environment (Ekman, 1992; Levenson, 1999; Tooby & Cosmides, 1990). Cognitive appraisal processes help map environmental cues onto appropriate emotional responses (Frijda, 1993) by paying attention to those features of the environment that hold adaptive informational content, and producing emotional responses to them along with

* Corresponding author.
E-mail address: paul.andrews@psychology.mcmaster.ca (P.W. Andrews).

http://dx.doi.org/10.1016/j.cognition.2016.05.021
0010-0277/© 2016 Published by Elsevier B.V.

their concomitant adaptive behaviors (Tooby & Cosmides, 1990). A key question, therefore, involves identifying the specific informational content in music that interacts with evolved appraisal mechanisms to produce emotions. Since these appraisal mechanisms are functionally linked to emotions generally, they probably did not evolve for processing music *per se*. Rather, it seems likely that music contains information that triggers appraisal mechanisms that evolved to produce emotion in other contexts.

Ideally, an explanation of how music induces emotional responses would be linked to a functional understanding of why organisms evolved to feel pleasant or unpleasant emotions. Evolutionary theory suggests that pleasant emotions arise when an organism has found an adaptive response to a situation, and the intrinsically rewarding properties of these emotions direct attention, motivation and cognition to maintaining that response. Conversely, unpleasant emotions arise when the organism lacks an adaptive response to a situation, and their function is to direct attention, motivation and cognition towards searching for and implementing adaptive responses (Carver & Scheier, 1990; Levenson, 1999; Thornhill & Thornhill, 1989; Tooby & Cosmides, 1990).

Several lines of research have identified specific aspects of music that could interact with pre-existing appraisal mechanisms to produce emotion. The *motivation-structural hypothesis* states that natural selection has shaped emotional responses to the acoustical structure of vocalizations to elicit particular behavioral responses (Morton, 1977). For example, low pitched and loud sounds are physically associated with large body size. Across species, vocal expressions in aggressive contexts (e.g., anger) include these acoustical features, presumably as a means of signaling large body size (Ohala, 1984). Similarly, it has been proposed that acoustic features of music with emotive properties may mimic those expressed by the voice in emotional contexts (Scherer, 1995). Hence, the emotions evoked by music may to some extent be predicted from acoustical features of the music (Balkwill & Thompson, 1999; Balkwill et al., 2004; Gagnon & Peretz, 2003; Hevner, 1936; Juslin, 2001).

The *expectation hypothesis* suggests that music exploits evolutionarily ancient physiological and cognitive mechanisms for detecting and responding to unexpected events (Huron, 2006; Meyer, 1956; Trainor & Zatorre, 2015). According to this idea, emotional responses to music commonly arise through manipulation of the listener's expectations. A prior musical context sets up probabilistic expectations for subsequent musical events, and the degree of confirmation or violation of these expectations, and the manner in which they are violated, gives rise to emotional responses. Huron's (2006) Imagination-Tension-Prediction-Reaction-Appraisal (ITPRA) hypothesis argues that the mechanisms involved in processing expectancy information evolved to help organisms understand, predict, and react adaptively to their environment. Trainor and Zatorre (2015) provide evidence for the physiological instantiation of expectancy processing in the brain related to musical emotion.

Finally, and of most importance here, a number of theorists have considered the role of *auditory scene analysis* (*ASA*) in musical organization and musical aesthetics (e.g., McAdams & Bregman, 1979; Wright & Bregman, 1987; Huron, 2001; Bonin & Smilek, 2015; Trainor, 2015). The auditory system evolved to accurately determine the identity and location of important objects in their environment, such as predators, conspecifics, mates, food sources, and running water, which directly affect survival and reproduction (Bregman, 1990; Fay & Popper, 2000). ASA describes the organizing principles by which the brain makes inferences about sound sources in its current auditory environment (Bregman, 1990). Each sound source emits a sound wave from some location in physical space with a characteristic spectrotemporal signature. These sound waves are summed and reach the ear as one complex wave. Thus, the auditory system is faced with the challenge of parsing this complex sound wave into a representation of auditory objects on the basis of the temporal and spectral signatures of the sound sources that created them.

The auditory system first performs a spectrotemporal analysis of the incoming complex sound wave. It then uses a number of cues to determine which frequency components ought to be grouped together ("fused") as a single auditory object in perception because they most likely originated from the same sound source, and which should be represented as perceptually distinct ("segregated") because they most likely originated from different sound sources. ASA involves analyzing simultaneous frequency content as well as how it changes over time, as a single sound source such as a melody or spoken sentence can vary across time (Bregman, 1990).

One critical cue for the fusion or segregation of auditory objects is *temporal simultaneity* (Bregman & Pinker, 1978; Dannenbring & Bregman, 1978). Frequency components with common onset and offset times are most likely to have come from the same sound source and thus share a unified representation in auditory perception. A second cue is *parallel motion* (Bregman & Doehring, 1984; McAdams, 1982). Frequency components that move up and down in pitch together likely come from the same sound source and are consequently assigned to a unified auditory object. A third cue is *location*. Interaural time and level differences are computed in the auditory brain stem and can be used to help determine the spatial origin of different sounds (Moore, 2013). Components originating from different physical locations are most likely coming from different sound sources and are thus represented independently. A fourth cue to the number and identity of different sound sources is *timbre*. Timbre is the temporal-spectral quality of sound that allows a listener to differentiate, for example, a flute from a violin, even when those sources are playing a note of the same pitch, loudness, and duration (Caclin, McAdams, Smith, & Winsberg, 2005). Successive sounds that share a common timbre will tend to share a unified representation in auditory perception while those that differ in timbre will be represented independently in auditory perception (Culling & Darwin, 1993; Gregory, 1994).

A fifth cue is *harmonicity* (Dewitt & Crowder, 1987). Natural sounds that give rise to the sensation of pitch typically contain energy at a fundamental frequency, f_0, and at integer multiples of that frequency, called harmonics. For example, a sound with a fundamental of 100 Hz would also have energy at harmonics at 200, 300, 400, ... Hz. Thus, harmonically-related frequency components are most likely to arise from a single sound source and thus fuse in auditory perception, whereas inharmonic frequency content indicates the likely presence of more than one sound source and will tend to result in the perception of more than one sound source segregate. Interestingly, there is a long history of associating harmonic frequency relations to emotions via the concept of sensory consonance and dissonance. According to Plomp and Levelt (1965) interactions between partials close in frequency (which arise, for example, between harmonics of tones whose fundamental frequencies do not stand in simple integer ratios) cause interference patterns and beating on the basilar membrane and lead to unpleasant emotions. More recently, this idea has been modified by experiments showing that *inharmonicity* (the degree to which the frequency components present in a sound deviate from harmonics at integer multiples of a fundamental), rather than roughness and beating, is sufficient to trigger unpleasant emotion (McDermott, Lehr, & Oxenham, 2010). In any case, many studies indicate that the harmonic (or inharmonic) structure of the frequency components present in a stimulus is related to the perceived pleasantness (or unpleasantness) of a sound.

Some theorists have considered how evolved ASA functions may relate to the emotive properties of inharmonicity. Bonin and