Proceedings of the 3rd International Conference on Music & Emotion (ICME3), Jyväskylä, Finland, 11th - 15th June 2013. Geoff Luck & Olivier Brabant (Eds.)

Sonification of Emotion: Strategies for Continuous Display of Arousal and Valence

R. Michael Winters, Marcelo M. Wanderley

Input Devices and Music Interaction Lab (IDMIL)
Center for Interdisciplinary Research in Music Media and Technology (CIRMMT)
Schulich School of Music, McGill University
Raymond.Winters@mail.mcgill.ca

Abstract

Sonification is an interdisciplinary field of research broadly interested in the use of sound to convey information. A fundamental attribute of sound is its ability to evoke emotion, but the display of emotion as a continuous data type has not yet received adequate attention. This paper motivates the use of sonification for display of emotion in affective computing, and as a means of targeting mechanisms of emotion elicitation in music. Environmental sound and music are presented as two possible sources for non-verbal auditory emotion elicitation, each with specific determinants and available features. The review concludes that the auditory-cognitive mechanisms of brain-stem reflex and emotional contagion provide the most advantageous framework for development. A sonification model is presented that implements cues that target these mechanisms. Computationally based strategies for evaluation are presented drawing upon the music emotion recognition literature. Additional aesthetic considerations are discussed that benefit usability and attractiveness of the display.

Keywords: sonification, psychoacoustic cues, affective computing

1. Introduction

Sonification is an interdisciplinary research of research broadly interested in the use of sound (usually "non-speech audio") to convey information (Kramer et al., 1999). A classic example of sonification, the Geiger counter, conveys the amount of radiation in the nearby environment using audible clicks. Although sonification has found many applications, this small sample exemplifies two compelling functions. Namely, sound can i) display a stream of information that is not visually obvious and ii) leave the eyes free to direct attention to other tasks. Like radiation, emotion is not always visually accessible, and displaying emotional information through sound does not require visual attention. Unique to emotion however, sonification can recruit resources from a cognitive apparatus that is well equipped for auditory emotion perception.

In the field of sonification, the subject of continuous emotion display has not yet received adequate attention. Sonification applications have included assistive technologies, bio-acoustic feedback, data exploration, alarms, and process monitoring (Hermann, Hunt, & Neuhoff, 2011), but the subject of emotion is rare. Though it has been recognized for its role in sound quality and interaction (Serafin et al., 2011), and is relevant to preference and pleasantness in sonification aesthetics (Vickers, 2011), only short, discrete sounds have thus far been applied. Such examples include using auditory icons to communicate emotional associations of the weather (Hermann, Drees, & Ritter, 2003) and using earcons for emotional communication in driver-vehicle interfaces (Larsson, 2010) and robotics (Jee, Jeong, Kim, Kwon, & Kobayahi, 2009), but the display of emotion as a continuous realtime data type is absent. The subject as a whole is much more at home in the realms of contemporary research in affective computing (Picard, 1997) and musical emotion (Juslin & Sloboda, 2010), where emotion expression and communication is considered computationally and music's affective capacity is studied in depth.

Furthermore, affective computing and musical emotion stand to benefit from the development of sonification strategies for emotion. Although embodied conversational agents (Hyniewska, Niewiadomski, Mancini, & Pelachaud, 2010) and emotional speech (Schröder, Burkhardt, & Krstulović, 2010) are the predominantly used modalities for affect display and communication, non-speech audio is an *unembodied* medium, requiring neither a face or a voice to be understood, and by extension, leaving visual and verbal attention un-taxed. When used in combination with other display modalities, this auxiliary channel may contribute to a more meaningful data interpretation.

Sonification of emotion can also be useful to the study of musical emotion. A great number of psychological studies have thus far been applied to determining the acoustic, structural (Gabrielsson & Lindström, 2010), and performative (Juslin & Timmers, 2010) elicitors for musical emotion. However, these results have yet to be applied to creating a "systematic and theoretically informed" manipulation of musical stimuli, which according to Juslin and Västfjäll (Juslin & Västfjäll, 2008, p. 574), would be a "significant advance" to stimuli selection. Parallel to psychological studies, music emotion recognition (MER) (Yang & Chen, 2011) has created models for musical emotion using sets of psychoacoustic features, reaching approximately 65% accuracy for categorical emotion recognition in large corpora of music (Kim et al., 2010, p. 261). Sonification offers the possibility of targeting the mechanisms for emotion induction that rely upon the same low-level acoustic cues as these algorithms, increasing (or even *decreasing)* recognition accuracy, leading to interesting conclusions.

This paper motivates the use of sonification for affective computing and presents strategies for continuous auditory monitoring of arousal and valence. After presenting relevant results from environmental sound, a framework is proposed founded upon two mechanisms for emotion induction in music. A sonification model that implements a select number of these acoustic cues is discussed. Goals and methods for evaluation are presented.

2. Background

Affective computing is defined as computing that relates to, arises from or deliberately influences emotion and other affective phenomenon (Picard, 1997). This definition is broad enough to include some uniquely musical pursuits, which would not normally be considered as related to affective computing. The first is music emotion recognition (MER), where automated, computational systems for emotion or "mood" recognition based on audio and/or text-based information have received increasing attention (Kim et al., 2010). The second are systems for affective music generation, where music composition is computationally infused with results from psychological studies of music emotion (e.g. Gabrielsson & Lindström, 2010). Within affective computing, music has been recognized as a "socially accepted form of mood manipulation," (Picard, 1997, p. 234) which for example has been applied to noted performance gains in sports (Eliakim, Bodner, Eliakim, Nemet, & Meckel, 2012), gaming (Cassidy & MacDonald, 2009), and driving mood (Zwaag, Fairclough, Spiridon, & Westerink, 2011).

Among these alternatives, sonification of emotion is most closely related to the development of affective music generation systems. Both share emotional data as input and create an "emotional mapping" to sound parameters. Furthermore, sonification can be listened to musically (Vickers & Hogg, 2006) and even integrated into affective music generation systems (Winters, Hattwick, & Wanderley, 2013). However, they can be distinguished both by the goals of the system designer and the way that they are meant to be listened to. Borrowing from the standard definition of sonification, the goal of a designer is to create a "transformation of data relationships into perceived relations in an acoustic signal for the purposes of facilitating communication or interpretation" (Kramer et al., 1999). In this light, the sound resulting from sonification is most comprehensively a *signal* that for the listener communicates or interprets important data relationships. If the data is emotion, than sonification, even when explicitly borrowing acoustic features from music, is simply a signal that communicates or interprets the data for the user.

The definition of sonification in fact, most closely parallels one of four non-exclusive areas of affective computing: technologies for displaying emotional information or mediating the expression or communication of emotion (Picard, 2009). Although this area most commonly makes use social signals (Vinciarelli et al., 2012) such as facial, gestural and vocal expressions in embodied conversational agents (Hyniewska et al., 2010), and the task of knowing the social display rules that govern *when* to display *which* affect is the "hardest challenge," (Picard, 2009, p. 13) there are contexts in which the relative simplicity of accurate realtime auditory display of emotion would beneficial.

For communication, these contexts arise when social displays of affect are unavailable, misleading, or inappropriate. A social display might be unavailable in cases when an agent is physically removed from or incapable of generating signals recognizable to the receiver. Social displays might be misleading if they are purposely masked, neutralized, or changed in magnitude (Matsumoto, 2009). A social display might be inappropriate if verbal or visual attention needs to be directed elsewhere, like when engaging in other more primary tasks. If paired with a social display, the auditory channel might be likened to the use of music in film, where sound contributes to the emotional expression of a multimodal scene. In visually based analysis tasks, the addition of the auditory channel might draw attention to data relationships not obvious if using visual-only methods.

Sonification of emotion is further motivated by increasingly sophisticated and diverse technologies for realtime emotion measurement and recognition. In these contexts, the subjective experience of emotion is often represented dimensionally (Fontaine, 2009), and the two-dimensional arousal/valence model of Russell is particularly prominent (Russell, 1980). To create a continuous sonification that would be successful in the use-contexts previously described, an objective and systematic mapping of arousal and valence appears most prudent. The content of the next section determines which of the many possible features of non-speech sounds make good candidates for emotion display, and how they might be mapped from realtime arousal and valence coordinates.

3. Determining Best Strategies

Potential sources for auditory display of affect come from two broad categories of sound: environmental sound and music. Though speech is another candidate, the stated goal is to create a display that does not conflict with verbal communication. Although some of the cues used in vocal expression of emotion might be shared by the auditory display (as in music; Juslin & Laukka, 2003), the goal here is not to use speech.

Within environmental sounds and music, there are additional requirements imposed by the conditions of realtime data monitoring as a background task in parallel to other more primary tasks. In sonification, this context is most often associated with process monitoring applications, and the present case is most closely a *peripheral* rather than direct or serendipitous display (Vickers, 2011). As noted by Vickers, common issues raised in process monitoring design are intrusion or distraction, fatigue and annoyance, poor aesthetic or ecological choices, and comprehensibility. These concerns are in turn grounded in the underlying need for appropriate aesthetic and semiotic choices. Through an analysis of acoustic features that communicate emotion in music and environmental sounds, this review shows that ultimately music provides the strongest theoretical framework for development due to the wealth of research and the continuous and malleable nature of its elicitors.

3.1. Emotion in the Acoustic Environment

Research on the acoustic elicitors of emotion in the natural environment has been most commonly presented in the psychoacoustic literature or in the pleasantness or annoyance of product sounds. However, recent research has sought a more ecological approach to sound perception in which psychological determinants take prominence to strictly signal characteristics, and the role of emotion becomes more complex. "Emoacoustics" (emotional acoustics) research (Asutay et al., 2012) embodies this trend towards a focus on listener and context, and contributes intriguing new methods and results.

Perhaps the most thorough review comes from Tajadura-Jiménez who categorizes "auditory-induced emotions" into four determinants (Tajadura-Jiménez, 2008, Ch. 4):

1. Physical Determinants
2. Identification/Psychological Determinants
3. Spatial Determinants
4. Cross-Modal Determinants

Physical determinants are those related to the signal itself and are best studied using "meaningless" sounds (Västfjäll, 2012) like broadband noise, and amplitude or frequency modulated tones, as is done in the psychoacoustics literature (Fastl & Zwicker, 2007). Factors related to identification enter when a sound has meaning due to the recognition and cognitive associations of the listener. Experiments using the *International Affective Digitized Sounds Library* (Bradley & Lang, 2007) have targeted this determinant and found similarities with corresponding affective pictures (Bradley & Lang, 2000). Spatial determinants arise when some aspect of the space contributes to the emotion. Issues of proximity, location, room size (Tajadura-Jiménez, Väljamäe, Asutay, & Västfjäll, 2010), and approaching or receding sound sources (Tajadura-Jiménez, Larsson, Väljamäe, Västfjäll, & Kleiner, 2010) have been studied in combination with different sound types (Hagman, 2010). Cross-modality effects occur when emotionally salient information from one modality impacts another. For sound, visual or tactile information might contribute to the emotional meaning of a sound, though this effect has been studied the least.

Although these categories are valid, only the first three pertain to audio-only display. From these, identifiability requires special consideration. As mentioned in the introduction, identifiable sounds (a.k.a. auditory icons; Brazil & Fernström, 2011) have been applied thus far to conveying emotional associations of the weather (Hermann et al., 2003). Although the affective space occupied by these sounds has been shown to convey a variety of emotions (Bradley & Lang, 2000), sounds notably fall upon two motivations, "appetitive" and "defensive," creating a 'V' shape in the AV space. If this trend were to continue for all identifiable sounds, it would leave gaps that could not be well communicated through sound.

Movement is another problem for the use of identifiable sounds. To convey a transition from high arousal, high valence to low arousal, high valence, would require the interpolation through many sounds. If this transition were to occur rapidly, the identifiability of these sounds might be compromised due to their short length. This problem might be avoided by using *evolutionary objects* (Buxton, Gaver, & Bly, 1994), which, as identified in the auditory icons literature, allow sound properties to be updated while playing (Brazil & Fernström, 2011). If using an evolutionary object, the sound would need to be able to occupy the entire AV space, so it might be best to start with a sound which is more or less emotionally neutral. A promising candidate for this is *self-referential sounds* (Tajadura-Jiménez & Västfjäll, 2008), or sounds related to ones own body and its natural movements (e.g. walking, breathing). The sound of a heartbeat for instance could be changed in tempo or loudness to convey arousal, and perhaps sharpness, roughness, and tonalness to convey valence.

The capacity of using spatial determinants for continuous display is worth mentioning, though is also limited. Increasing room size (reverberation time) creates a systematic decrease in valence and increase in arousal for sounds with neutral emotional connotation (e.g. clarinet, duck quack), but not for negative connotation (e.g. dog growl) (Tajadura-Jiménez, Larsson, et al., 2010). Evidence supporting this effect of neutral sounds is also present in (Västfjäll, Larsson, & Kleiner, 2002), though the effect on arousal was less pronounced. Arousal, in fact, decreased for the condition of highest reverberation, attributed to a decrease in "presence." Presence, though lacking a precise acoustic definition, has been defined as the perceptual illusion of non-mediation (Lombard & Ditton, 1997), and has been strongly connected to the emotion in auditory virtual environments (Västfjäll, 2003), perhaps most analogously correlated with the degree of "realism" (Frija, 1988). Creating the illusion of "approach" is possible by increasing loudness, and creates an increase in emotional intensity, but only for identifiable sounds deemed "unpleasant" (Tajadura-Jiménez, Väljamäe, et al., 2010). Finally, in general, sounds perceived as coming from behind the individual are more emotionally arousing (Tajadura-Jiménez, Larsson, et al., 2010). The use of spatial effects for emotion display or expression is drawn into question as incongruent visual information can diminish the strength of the desired auditory illusion (Larsson, Västfjäll, Olsson, & Kleiner, 2007).

The results are most clear-cut with psychoacoustic literature using broadband-noise, and amplitude or frequency modulated tones (Fastl & Zwicker, 2007). Composite models for sensory pleasantness (p. 245) and psychoacoustic annoyance (p. 328) have been developed using well-defined metrics for roughness, sharpness, loudness, tonality, and fluctuation strength. These have been shown to predictive of ratings of pleasantness and annoyance of product sounds, though they were not designed to be able to predict the position in a full 2-D arousal valence model (Västfjäll, 2012). They make good candidates as features for sonification, though using ecologically valid stimuli should not be abandoned. Results from sonic interaction design (SID) have shown that "naturalness" creates a systematic increase in valence compared to synthesized sounds with similar spectral centroid and tonality (Lemaitre, Houix, Susini, Visell, & Franinović, 2012). However, as in SID, it might be best to consider naturalness as an overall aesthetic property that should be conserved, contributing to the attractiveness of the sound and "usability" of the sonification (Norman, 2004).

This review has accessed different possible features for emotion communication in environmental sounds. If using identifiable sounds, it would be best to use evolutionary sounds, perhaps in some way self-representational. Use of spatial effects can be considered if one is mindful of visual dominance. Psychoacoustic features are the most promising for sonification, but naturalness is a global property that should be conserved. Overall, it would appear that the strongest emotional determinant of environmental sound—identifiability—is not viable for sonification, dramatically diminishing the framework as a whole. The field of emotional acoustics is still developing, and future results might be more favorable. For the time being, a much stronger framework is founded in contemporary research on music and emotion, which will be discussed in the next section.

3.2. Mechanisms of Musical Emotion

On the surface, it would seem that the most useful results for sonification come from the wealth of results linking structural, acoustic, and performative cues in music to defined regions of the arousal/valence space. Instead however, a more rigorous approach first determines which psychological mechanisms are favorable for emotion elicitation given defined properties such as cultural specificity, volitional influence, and induction time. These mechanisms in turn encompass subsets of the available structural/acoustic feature space, making the process of selection easier.

Many psychological studies have been conducted to determine what structural, acoustic, and performative parameters contribute to emotional communication in music (Gabrielsson & Lindström, 2010; Juslin & Timmers, 2010). Additionally, new computational approaches to feature determination have been introduced in the field of music emotion recognition (Yang & Chen, 2011). This literature affirms that there is no dominant single feature, and musical emotion is best predicted using a multiplicity (Kim et al., 2010). The literature on performance cue utilization (Juslin, 2000) has also advanced—recent results have introduced defined ranges for communication of discrete emotions (Bresin & Friberg, 2011).

Collectively, these results offer an abundance of possible features for emotion communication in sonification, but music research offers a more fundamental approach, that of the auditory-cognitive mechanism. In this vein, a collection of six mechanisms for emotion elicitation in music has been proposed (Juslin & Västfjäll, 2008), two of which can be used for continuous auditory display: *brain-stem reflex* and *emotional contagion.* Both have a low-degree of cultural and volitional influence, fast induction speed, and a medium dependence upon musical structure (Juslin & Västfjäll, 2008, Table 4). It is worthy of note that the psychoacoustic features from the environmental sounds literature that are the most viable for sonification are accounted for by these mechanisms, and as noted in (Tajadura-Jiménez, 2008, p. 26), mostly the brain-stem reflex.

Acoustic features drawing upon the brain-stem reflex recruit innate structures of the brain that bear upon the organism's survival. As noted in (Juslin & Västfjäll, 2008, p. 574), these features are most commonly studied in the psychoacoustics literature and include sharpness, loudness, roughness, tonality, and fluctuation strength. In the music literature, a close relative of sharpness is the height of the spectral centroid. Likened to roughness is sensory dissonance, and tonality (a.k.a. "tonalness;" Egmond, 2009, p. 79) refers to how tone-like the timbre is as opposed to broadband. The spatial cues discussed in section 3.1, might be considered in this list in that spatial hearing is also shared and important to an organisms survival, though effects that are dependent upon the sound identification are likely cognitively mediated.

Emotional contagion is a process whereby emotion is induced by perceiving the expression of the stimulus itself and then "mimicking" it internally (Juslin & Västfjäll, 2008). The theory suggests that because music shares many of the acoustic features used in vocal expression of emotion, music becomes like a *super-expressive voice* (Juslin, 2001). Further, musical features are decoded by an "emotion-perception module" (Juslin & Laukka, 2003, p. 803) of the brain that does not distinguish between music and the voice. Evidence supporting this claim comes from an extensive review of literature in musical and vocal expression showing that a number of prominent features governing expression of five discrete emotions were shared in music and speech (Juslin & Laukka, 2003, Table 7). The cross-modal features relevant to this proposed module are tempo, intensity, intensity variability, high-frequency energy, pitch-level, pitch variability, pitch contour, attack and microstructural regularity (taken at the note-to-note level; Bunt & Pavicevic, 2001).

Implementation of these reflex and contagion features requires two levels of acoustic content, timbral and note-based. For the brain-stem reflex and psychoacoustics, spectral content and intensity must be manipulated—the sonification must include a structure that allows malleability of sharpness (amount of high-frequency energy), tonality (amount of noise versus tonal components in the spectra), roughness (including fluctuation strength), and loudness. To use emotional contagion features, a note-based structure must be available for manipulation of tempo, pitch, and attack.

The strength of these features is their low cultural influence, low volitional influence, induction speed and their dependence upon structure. This structure does not have to be "musical" necessarily, for these mechanisms are on the one hand biological, and on the other processed by an emotion-perception module that processes speech as well (Juslin & Laukka, 2003, p. 803). Other acoustic features that rely upon different mechanisms can (and perhaps should) be used in sonification, but they can be expected to be more culturally dependent, with potentially lower induction speed, and subject to volitional influence. Such a feature would be the major-minor mode, which in western classical music can be used to convey positive and negative valence. However, this connotation is not learned until the age of six to eight (Gabrielsson & Lindström, 2010, p. 393), and thus might be accounted for by the mechanism of musical expectancy.

3.3. Summary

Having compared mechanisms for emotional elicitation in both environmental sounds and music, it is clear that sonification of emotion finds more substantive support in the mechanisms described in musical emotion research. From environmental sounds, emotion determined through identification and appraisal of the sound was found to be a strong factor influencing emotion. However, the emotional space occupied by these sounds is incomplete, and the problem of movement suggests the use of emotionally neutral *evolutionary* or *self-representational* sounds for which acoustic properties can be easily manipulated. Though not well researched, "naturalness" should be conserved at a global level to maximize pleasantness.

Ultimately however, the results from this literature are much less developed than those from musical emotion, and factors such as cultural dependency, induction speed, degree of volitional influence have not been adequately accessed. From music research, two viable mechanisms for sonification have been proposed, each with well-defined psychological properties. Further, the *brain-stem reflex* accounts for the psychoacoustic and spatial results in the environmental sounds literature that would otherwise be most promising for sonification. The additional mechanism of *emotional contagion* presents additional musical features for sonification including tempo, attack, pitch information, and regularity.

4. Sonification Model

In designing the sonification model, the goal was to create a simple sound capable of accurately conveying the entire arousal and valence space. Details of the implementation and further discussion can be found in (Winters et al., 2013), but are summarized presently.

A single note forms the basis of the model. This note is created as a bank of resonant modes with independent control of center frequencies, amplitudes and decay times. The resonant object is excited through impulse in alternating left-right stereo channels. The choice of this sound was motivated by its "naturalness"—it is capable of generating sounds that resemble materials like glass, wood, metal, etc. For sonification, tempo, and loudness are mapped to increasing arousal, and the decay time increases with decreasing arousal. Increasing positive or negative valence is conveyed by slowly increasing the loudness of the fifth, third (M3/m3), and octave above the original note. Sensory dissonance is conveyed in the second quadrant by taking an identical copy and pitch-shifting upwards. Loudness of the copy increases with radial proximity to the line $3\pi/4$, and the pitch shift increases with distance from the origin.

4.1. Evaluation and Future Work

The sonification model presented has not yet been formally evaluated, which is the next step for validation. Of utmost importance is determining how well it conveys the underlying arousal/valence space. With this established, it will be necessary to perform user testing to evaluate sonification in the context of realtime peripheral process monitoring.

The decisions for tempo, loudness, and roughness are supported by the present discussion. Tempo is a feature from the *emotion contagion* mechanism, roughness is a feature of the *brain-stem reflex,* and both share loudness. Tempo and loudness increased with increasing arousal as in speech, and increasing roughness and loudness both increased with sensory *un-pleasantness.* The decisions for major-minor and decay are musical features that are not supported by the present discussion but were found to be useful for conveying valence and decreasing arousal respectively. In fact, these two decisions contributed more to the aesthetic appeal of the display than the decisions of loudness, tempo, and roughness. Although mindful that when using features not accounted for by brain-stem reflex and emotion contagion, desired psychological properties (e.g. low cultural specificity) are not guaranteed, the use of cultural associations has been supported in the design of process monitoring sonifications (Vickers, 2011) as well as in aesthetic computing (Fishwick, 2002). Drawing upon a listener's cultural associations can create a convincing display that enhances aesthetic appeal, but the designer should be mindful of its limitations. "Major-happy, minor-sad," for example is culturally learned and may not necessarily be understood by children under six to eight years old.

A yet undeveloped benefit of using strategies from music research, and perhaps most attractive for evaluation, are the growing number of models for music emotion recognition (Yang & Chen, 2011). Using audio-only features, these systems are capable of recognizing emotions categorically or dimensionally, and some systems are designed for time-varying, "second-by-second" emotion detection (Coutinho & Cangelosi, 2011; Schmidt & Kim, 2011). Because these models are sometimes designed for large corpora of music, stretching beyond those of western-classical tradition, the features used for recognition may be less culturally specific. For evaluation, these models can provide a preliminary metric of the accuracy of communication in the arousal valence space.

As of yet, several of the features supported in this analysis have not been implemented. From the brain-stem reflex, these include sharpness, tonalness, and fluctuation strength. From emotional contagion, these include pitch-level and its variation, contour, intensity variability, and attack. The spatial cues of increasing reverberation time and the auditory-illusion of "behind" might also be investigated. The framework of resonant synthesis creates sounds that are relatively more "natural" than other synthesis techniques. This strategy should be continued in further implementations, though using self-representational or evolutionary sounds might be assessed as well.

5. Conclusions

Realtime continuous auditory display of arousal and valence has not yet received adequate attention in the sonification literature, though the pursuit of technologies for realtime emotion recognition makes the data-type eminent. Benefits of sonification include displaying emotional information when visual or verbal cues are unavailable, misleading, or inappropriate, and providing an auxiliary channel for emotional display that can contribute to emotional expression or visual-based data analysis.

Determining the best strategies for display requires careful aesthetic and ecological choices, for which research on the emotional impact of environmental sound and music provides two possible categories for the designer. Currently, the most robust foundation for development is presented by research in musical emotion and specifically cues recruiting the mechanisms of brain-stem reflex and emotional contagion. These mechanisms account for most of the viable acoustic cues from environmental sound and propose additional ones that are shared with speech. These cues can be expected to have a low degree of cultural influence, a high induction speed, and a low degree of volitional influence.

The sonification model discussed explicitly uses some of these features, though others are presented for future work. To evaluate the model, it may be possible to use models for music emotion recognition as a preliminary design metric. With the accuracy of the mapping strategy accessed, user testing needs to evaluate how well the sonification performs when verbal and/or visual attention is already occupied.

References

Asutay, E., Västfjäll, D., Tajadura-Jiménez, A., Genell, A., Bergman, P., & Kleiner, M. (2012). Emoacoustics: A study of the psychoacoustical and psychological dimensions of emotional sound design. *Journal of the Audio Engineering Society, 60(1/2),* 21-8.

Bradley, M. M., & Lang, P. J. (2000). Affective reactions to acoustic stimuli. *Psychophysiology, 37,* 204-15.

Bradley, M. M., & Lang, P. J. (2007). *The international affective digitized sounds (2nd edition; IADS-2): Affective ratings of sounds and instruction manual* (Tech. Rep.). Gainesville, FL: University of Florida.

Brazil, E., & Fernström, M. (2011). Auditory icons. In T. Hermann, A. Hunt, & J. G. Neuhoff (Eds.), *The sonification handbook* (p. 325-38). Berlin, Germany: Logos Verlag.

Bresin, R., & Friberg, A. (2011). Emotion rendering in music: Range and characteristic values of seven musical variables. *Cortex, 47,* 1068-81.

Bunt, L., & Pavicevic, M. (2001). Music and emotion: Perspectives from music therapy. In P. N. Juslin & J. A. Sloboda (Eds.), *Music and emotion: Theory and research* (p. 181-201). New York, NY: Oxford University Press.

Buxton, W., Gaver, W., & Bly, S. (1994). *Auditory interfaces: The use of non-speech audio at the interface.* (Ch. 2: Acoustics and Psychoacoustics)

Cassidy, G., & MacDonald, R. (2009). The effects of music choice on task performance: A study of the impact of self-selected and experimenter-selected music on driving game performance and experience. *Musicae Scientiae, 13(2)*, 357-86.

Coutinho, E., & Cangelosi, A. (2011). Musical emotions: Predicting second-by-second subjective feelings of emotion from low-level psychoacoustic features and physiological measurements. *Emotion, 11(4)*, 921-37.

Egmond, R. V. (2009). The experience of product sounds. In H. N. J. Schifferstein & P. Hekkert (Eds.), *Product experience* (p. 69-89). San Diego, CA: Elsevier.

Eliakim, M., Bodner, E., Eliakim, A., Nemet, D., & Meckel, Y. (2012). Effect of motivational music on lactate levels during recovery from intense exercise. *Journal of Strength and Conditioning Research, 26(1)*, 80-6.

Fastl, H., & Zwicker, E. (2007). *Psychoacoustics: Facts and models* (3rd ed.). Berlin, Germany: Springer-Verlag.

Fishwick, P. A. (2002). Aesthetic programming: Crafting personalized software. *Leonardo, 35(4)*, 383-90.

Fontaine, J. R. (2009). Dimensional emotion models. In D. Sander & K. R. Scherer (Eds.), *Oxford companion to emotion and the affective sciences* (p. 119-20). New York, NY: Oxford University Press.

Frija, N. H. (1988). The laws of emotion. *American Psychologist, 43(5)*, 349-58.

Gabrielsson, A., & Juslin, P. N. (2003). Emotional expression in music. In R. J. Davidson, K. R. Scherer, & H. H. Goldsmith (Eds.), *Handbook of affective sciences* (p. 503-34). New York, NY: Oxford University Press.

Gabrielsson, A., & Lindström, E. (2010). The role of structure in the musical expression of emotions. In P. N. Juslin & J. Sloboda (Eds.), *Handbook of music and emotion: Theory, research, applications* (p. 367-400). New York, NY: Oxford University Press.

Hagman, F. (2010). *Emotional response to sound: Influence of spatial determinants.* Unpublished master's thesis, Chalmers University of Technology, Göteborg, Sweden.

Hermann, T., Drees, J. M., & Ritter, H. (2003, July). Broadcasting auditory weather reports - a pilot project. In *Proceedings of the international conference on auditory display* (p. 208-11). Boston, MA.

Hermann, T., Hunt, A., & Neuhoff, J. G. (Eds.). (2011). *The sonification handbook.* Berlin, Germany: Logos Verlag.

Hyniewska, S., Niewiadomski, R., Mancini, M., & Pelachaud, C. (2010). Expression of affects in embodied conversational agents. In K. R. Scherer, T. Bänziger, & E. B. Roesch (Eds.), *Blueprint for affective computing: A sourcebook* (p. 213-21). New York, NY: Oxford University Press.

Jee, E.-S., Jeong, Y.-J., Kim, C. H., Kwon, D.-S., & Kobayahi, H. (2009). Sound production for the emotional expression of social interactive robots. In V. A. Kulyukin (Ed.), *Advances in human-robot interaction* (p. 257-72). Vukovar, Croatia: InTech.

Juslin, P. N. (2000). Cue utilization in communication of emotion in music performance: Relating performance to perception. *Journal of Experimental Psychology, 26(6)*, 1797-813.

Juslin, P. N. (2001). Communication emotion in music performance: A review and a theoretical framework. In P. N. Juslin & J. A. Sloboda (Eds.), *Music and emotion: Theory and research* (p. 309-37). New York, NY: Oxford University Press.

Juslin, P. N., & Laukka, P. (2003). Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychological Bulletin, 129(5)*, 770-814.

Juslin, P. N., & Sloboda, J. A. (Eds.). (2010). *Handbook of music and emotion: Theory, research, applications.* New York, NY: Oxford University Press.

Juslin, P. N., & Timmers, R. (2010). Expression and communication of emotion in music performance. In P. N. Juslin & J. A. Sloboda (Eds.), *Handbook of music and emotion: Theory, research, applications* (p. 453-89). New York, NY: Oxford University Press.

Juslin, P. N., & Västfjäll, D. (2008). Emotional responses to music: The need to consider underlying mechanisms. *Behavioral and Brain Sciences, 31(5)*, 559-621.

Kim, Y. E., Schmidt, E. M., Migneco, R., Morton, B. G., Richardson, P., Scott, J., et al. (2010, August). Music emotion recognition: A state of the art review. In *Proceedings of the 11th international society for music information retrieval conference* (p. 255-66). Utrecht, Netherlands.

Kramer, G., Walker, B., Bonebright, T., Cook, P., Flowers, J., Miner, N., et al. (1999). *The sonification report: Status of the field and research agenda.* Santa Fe, NM: International Community for Auditory Display (ICAD).

Larsson, P. (2010). Tools for designing emotional auditory driver-vehicle interfaces. In S. Ystad, M. Aramaki, R. Kronland-Martinet, & K. Jensen (Eds.), *Auditory display: 6th international symposium,* *CMMR/ICAD 2009, revised papers* (p. 1-11). Berlin, Germany: Springer.

Larsson, P., Västfjäll, D., Olsson, P., & Kleiner, M. (2007, October). When what you hear is what you see: Presence and auditory-visual integration in virtual environments. In *Proceedings of the 10th annual international workshop on presence* (p. 11-8). Barcelona, Spain.

Lemaitre, G., Houix, O., Susini, P., Visell, Y., & Franinović, K. (2012). Feelings elicited by auditory feedback from a computationally augmented artifact: The flops. *IEEE Transactions on Affective Computing, 3(3),* 335-48.

Lombard, M., & Ditton, T. (1997). At the heart of it all: The concept of presence. *Journal of Computer-Mediated Communication, 3(2).*

Matsumoto, D. (2009). Display rules. In D. Sander & K. R. Scherer (Eds.), *Oxford companion to emotion and the affective sciences* (p. 124). New York, NY: Oxford University Press.

Norman, D. A. (2004). *Emotional design: Why we love (or hate) everyday things.* New York, NY: Basic Books.

Picard, R. (1997). *Affective computing.* Cambridge, MA: The MIT Press.

Picard, R. (2009). Affective computing. In D. Sander & K. R. Scherer (Eds.), *The oxford companion to emotion and the affective sciences* (p. 11-5). New York, NY: Oxford University Press.

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology, 39(6),* 1161-78.

Schmidt, E. M., & Kim, Y. E. (2011, October). Modeling musical emotion dynamics with conditional random fields. In *Proceedings of the 12th international society for music information retrieval conference* (p. 777-82). Miami, FL.

Schröder, M., Burkhardt, F., & Krstulović, S. (2010). Synthesis of emotional speech. In K. R. Scherer, T. Bänziger, & E. B. Roesch (Eds.), *Blueprint for affective computing: A sourcebook* (p. 222-31). New York, NY: Oxford University Press.

Serafin, S., Franinović, K., Hermann, T., Lemaitre, G., Rinott, M., & Rocchesso, D. (2011). Sonic interaction design. In T. Hermann, A. Hunt, & J. G. Neuhoff (Eds.), *The sonification handbook* (p. 87-110). Berlin, Germany: Springer-Verlag.

Tajadura-Jiménez, A. (2008). *Embodied psychoacoustics: Spatial and multisensory determinants of auditory-induced emotion.* Unpublished doctoral dissertation, Chalmers University of Technology, Göteborg, Sweden.

Tajadura-Jiménez, A., Larsson, P., Väljamäe, A., Västfjäll, D., & Kleiner, M. (2010). When room size matters: Acoustic influences on emotional responses to sounds. *Emotion, 10(3),* 416-22.

Tajadura-Jiménez, A., Väljamäe, A., Asutay, E., & Västfjäll, D. (2010). Embodied auditory perception: The emotional impact of approaching and receding sound sources. *Emotion, 10(2),* 216-29.

Tajadura-Jiménez, A., & Västfjäll, D. (2008). Auditory-induced emotion: A neglected channel for communication in human-computer interaction. In C. Peter & R. Beale (Eds.), *Affect and emotion in HCI* (p. 63-74). Berlin, Germany: Springer-Verlag.

Västfjäll, D. (2003). The subjective sense of presence, emotion recognition, and experienced emotions in auditory virtual environments. *CyberPsychology & Behavior, 6(2),* 181-8.

Västfjäll, D. (2012). Emotional reactions to sounds without meaning. *Psychology, 3(8),* 606-9.

Västfjäll, D., Larsson, P., & Kleiner, M. (2002). Emotion and auditory virtual environments: Affect-based judgments of music reproduced with virtual reverberation times. *CyberPsychology & Behavior, 5(1),* 19-32.

Vickers, P. (2011). Sonification for process monitoring. In T. Hermann, A. Hunt, & J. G. Neuhoff (Eds.), *The sonification handbook* (p. 455-91). Berlin, Germany: Logos Verlag.

Vickers, P., & Hogg, B. (2006, June). Sonification abstraite/sonification concrète: An 'aesthetic perspective space' for classifying auditory displays in the ars musica domain. In *Proceedings of the 12th international conference on auditory display* (p. 210-6). London, UK.

Vinciarelli, A., Pantic, M., Heylen, F., Pelachaud, C., Poggi, I., D'Errico, F., et al. (2012). Bridging the gap between social animal and unsocial machine: A survey of social signal processing. *IEEE Transactions on Affective Computing, 3(1),* 69-87.

Winters, R. M., Hattwick, I., & Wanderley, M. M. (2013, June). Integrating emotional data into music performance: Two audio environments for the emotional imaging composer. In *Proceedings of the 3rd International conference on music and emotion* (this volume). Jyväskylä, Finland.

Yang, Y.-H., & Chen, H. H. (2011). *Music emotion recognition.* Boca Raton, FL: CRC Press.

Zwaag, M. D. van der, Fairclough, S., Spiridon, E., & Westerink, J. H. (2011). The impact of music on affect during anger inducing drives. In S. D'Mello (Ed.), *Affective computing and intelligent interaction* (p. 407-16). Berlin, Germany: Springer-Verlag.