# Sonification of Emotion: Strategies and results from the intersection with music

R. MICHAEL WINTERS and MARCELO M. WANDERLEY

Input Devices and Music Interaction Laboratory, CIRMMT, Schulich School of Music, McGill University, 555 Sherbrooke St W, H3A 1E3 Montreal, QC, Canada
E-mail: Raymond.Winters@mail.mcgill.ca; Marcelo.Wanderley@mcgill.ca

**Emotion is a word not often heard in sonification, though advances in affective computing make the data type imminent. At times the relationship between emotion and sonification has been contentious due to an implied overlap with music. This paper clarifies the relationship, demonstrating how it can be mutually beneficial. After identifying contexts favourable to auditory display of emotion, and the utility of its development to research in musical emotion, the current state of the field is addressed, reiterating the necessary conditions for sound to qualify as a sonification of emotion. With this framework, strategies for display are presented that use acoustic and structural cues designed to target select auditory-cognitive mechanisms of musical emotion. Two sonifications are then described using these strategies to convey arousal and valence though differing in design methodology: one designed ecologically, the other computationally. Each model is sampled at 15-second intervals at 49 evenly distributed points on the *AV* space, and evaluated using a publically available tool for computational music emotion recognition. The computational design performed 65 times better in this test, but the ecological design is argued to be more useful for emotional communication. Conscious of these limitations, computational design and evaluation is supported for future development.**

## 1. INTRODUCTION

Sonification is an interdisciplinary field of research broadly interested in the use of sound to convey information (Kramer, Walker, Bonebright, Cook, Flowers, Miner et al. 1999). Though there are many techniques of sonification and many tasks to which it has been applied, a continual problem is that of definition (Supper 2012). Always obfuscating, compromising and testing the mettle of a concise and encompassing delineation are the various artistic and musical practices whereby data is also transformed into sound.

Music has been called a 'language of emotion' (Crooke 1957) and with good cause: a vast and expanding literature describes the ways that music comes to convey or induce an emotion in listeners (Juslin and Sloboda 2010). Sonification, on the other hand, seems to be anything but a language of emotion. Of the approximate 2.4 million standard words in *The Sonification Handbook* (Hermann, Hunt and Neuhoff 2011), the word 'emotion' appears a mere 78 times. Recent discussions of what might be considered a 'sonification of emotion' have even brought contention in the sonification community, due to potential overlaps with music (Preti and Schubert 2011; Schubert, Ferguson, Farrar and McPherson 2011).

In spite of this difficulty, there are several reasons why the field of sonification should consider the representation and communication of emotion more seriously. The first and perhaps most obvious reason is that emotion as a form of data is becoming increasingly common. In the field of affective computing (Picard 1997), algorithms have been designed to detect and measure emotion from all manner of possible sources, including but not limited to physiological process, EEG, and facial, gestural and vocal expression (Picard and Daily 2005). In addition to these indirect measures, technologies for continuous self-report are being used to collect readings of an individual's time-varying emotional experience (Schubert 2010). Just as with other data types, the facilities of audition can be directed to perceiving this information, identifying patterns, and supporting communication when verbal or visual attention is already occupied (Walker and Nees 2011).

Another and perhaps more exciting prospect stems from the utility of the auditory-cognitive system as a non-verbal, non-visual channel for emotional communication. As evidence of the strength of this channel, one need look no further than the importance of music in film, where sound itself brings insurmountable intensity to a scene, even to the point of overriding incongruent visual and verbal emotional cues.

To create a sonification of emotion, however, one does not have to create music. As will be discussed presently, many of the most promising applications benefit from the use of sound as a background display. Music, in all of its cognitive complexity, may obscure communication if it does not systematically convey the data, requires too much attention or uses culturally learned schemas. Instead, by selecting wisely from emotionally salient acoustic cues, many of which are nevertheless used in music, emotion can

*Organised Sound* **19**(1): 60–69 © Cambridge University Press, 2014. doi:10.1017/S1355771813000411

be conveyed as a background information stream with desirable features such as high induction speed and low volitional influence.

After introducing the benefits and contexts favourable to the auditory display of emotion, the current state of research is presented, reiterating the necessary conditions for sound to be considered a ‘sonification of emotion’. Although a number of structural and acoustical cues are used in the expression of musical emotion, a select group is chosen for sonification from desired psychological properties. Two sonification mappings are then presented for conveying arousal and valence but differing in design methodology. The first is designed ecologically using recommendations from the musical emotion literature, while the second is designed computationally using a publically available model for music emotion recognition. Although the latter performs significantly better on a computational test, the former is argued to be more useful for emotion communication. These results help clarify the relationship between music and sonification, identify areas of mutual benefit and facilitate future collaboration in emotion display.

## 2. MOTIVATION AND BACKGROUND

The auditory display of emotion is a timely pursuit supported by research agendas originating in affective computing and musical emotion. Applications arise in both, for either emotional communication or model evaluation. Music research in particular offers a robust framework for development, which is applied to the present research. After these relationships have been presented in detail, the current status of emotion in auditory display is described, highlighting the requirements for a technique to be appropriately termed a sonification of emotion.

### 2.1. Affective computing

Affective computing has been defined as computing that relates to, arises from or deliberately influences emotion and other affective phenomena (Picard 1997). Though this definition is rather broad, technologies for display, expression or communication of emotion constitute the third of four major research foci (Picard 2009). In this context, sonification contrasts and complements existing display modalities, many of which require a face, voice or body for communication. Non-speech sound offers an *unembodied* medium for emotional communication that can be ideal in situations when verbal and/or visual attention is already occupied. By extension, sonification of emotion can be added to an existing emotional display, potentially facilitating communication or expression.

The complexities of the rules governing *when* to display *which* affect has been described as ‘the hardest challenge’ of real-time emotion display (Picard 2009: 13). However, Winters and Wanderley (2013) list three cases in which the relative simplicity of real-time, accurate auditory display of emotion can be beneficial. These contexts arise when social signals (e.g. facial, vocal, gestural; Vinciarelli, Pantic, Heylen, Pelachaud, Poggi, D’Errico et al. 2012) are unavailable, misleading or inappropriate.

A social display might be *unavailable* when an agent is either physically removed from or incapable of generating the social signals that would be otherwise recognisable to a receiver. In the case of autism, for instance, where a person has difficulty utilising social cues that would allow for their emotional reaction to be recognised, sonification might be used to assist the receiver and cue them into an otherwise hidden emotional experience. A social display might be *misleading* when social signals of emotion are consciously or unconsciously masked, neutralised or changed in magnitude (Matsumoto 2009). In this case, verbal and visual attention can remain dedicated to the socially displayed content, but the auditory display once again provides access to a hidden emotional layer, and perhaps a deeper understanding of the agent’s state. Finally, a social display may be *inappropriate* when visual and/or verbal attention need to be directed elsewhere, such as when engaged in complex, more primary tasks.

In any of these contexts, the auditory display needs to be clear but also not so complex as to demand unnecessary attentional resources on the part of the user. This function most closely parallels sonification techniques related to process monitoring (Vickers 2011). Furthermore, because the user’s primary attention is directed elsewhere, but the information content of the display is important to the overall goal, the sonification would be classified as peripheral.

### 2.2. Musical emotion

The auditory display of emotion should not exclusively direct itself towards contexts for real-time emotional communication. To consider this purpose as the exclusive benefit is to miss a potentially advantageous link with a close partner, the study of musical emotion. Musical emotion describes emotions induced or conveyed by music, and, while its discussion is old (Budd 1985), since the late twentieth century its scientific axes have expanded, and a variety of psychophysiological, behavioral and computational methods have been introduced.

Sonification of emotion intersects with musical emotion insofar as the study profits from systematic and theoretically informed mappings of acoustic features. For over three-quarters of a century,

https://doi.org/10.1017/S1355771813000411 Published online by Cambridge University Press

research has been directed to determining the structural and acoustic elicitors responsible for musical emotion (Gabrielsson and Lindström 2010). Although music listening is a multi-faceted process in which cultural learning and cognitive associations are fundamental, this branch has directed itself towards the underlying acoustic details. Though beginning with psychological studies, machine-learning approaches have recently gained momentum, offering signal-level correlates of music perception and composite computational models (Yang and Chen 2011).

Using this background of musical emotion, sonification is afforded a wealth of knowledge on auditory emotion, and can make use of well-developed theories and results. These form the basis for the sonification strategies introduced in Section 3.1. However, sonification can also benefit the study of musical emotion by providing 'systematic and theoretically informed' approaches, which, according to Juslin and Västfjäll (2008: 574) would be a 'significant advance' to stimuli selection. In this way, both fields can profit from the other's research developments.

This benefit is most easily applied to computational models for music emotion recognition, many of which use purely signal/content level attributes for prediction (Kim, Schmidt, Migneco, Morton, Richardson, Scott et al. 2010). These models are complex, using a multiplicity of acoustic features and functions for combination, but can ideally be generalised to large corpora of music, potentially spanning many genres (Ogihara and Kim 2012). Using sonification, these purely computational models can be acoustically instantiated, satisfying a broad range of model requirements, and potentially isolating these low-level acoustic features from the higher-level cultural and cognitive mechanisms involved in music listening. Both of the sonifications presented in Section 3 are measured by such a model, forming the basis for evaluation in Section 4.

### 2.3. The sonification perspective

The subject of emotion is rare in the sonification literature, and at times even contentious for the definition of sonification (Schubert, Ferguson, Farrar and McPherson 2011). To frame the present research, the current state of emotion in the field is addressed, identifying contexts where sonification has thus far been used, its relationship to aesthetics, and the conditions that qualify a technique as a 'sonification of emotion'.

The actual use of sound to communicate or express emotional information has thus far been limited to short, discrete sounds that would qualify either as auditory icons (Brazil and Fernström 2011) or earcons (McGookin and Brewster 2011). Hermann, Drees and Ritter (2003), for instance, have explored the use of auditory icons to communicate emotional associations in auditory weather reports. These emotive markers (e.g. bird, sigh, scream) were played alongside auditory icons indicating more descriptive information such as temperature, windiness and humidity. Later, Larsson (2010) introduced two software tools for designing earcons for communication of urgency in auditory-in-vehicle interfaces. As with the weather reports, the emotive content of these sounds were meant to be paired with descriptive identifiers (e.g. seatbelt reminder, collision warning).

Robotics has been another venue for application. Jee, Jeong, Kim, Kwon and Kobayahi (2009) have studied the use of short musical excerpts to express discrete emotional states such as happiness, sadness or fear. The authors later conducted a review of 275 earcons used for communication of emotion and intention in two popular science-fiction robots (Jee, Jeong, Kim and Kobayahi 2010), applying the results to the design of seven musical sounds for expression in an English-teacher robot.

These uses of sound to convey emotional information can be contrasted with aesthetic and design studies where the discussion of auditory emotion also appears. In sonic interaction design, for example, emotions have been studied in users performing tasks with 'the flops glass', an acoustically and computationally augmented physical object (Lemaitre, Houix, Susini, Visell and Franinovíc 2012). Results suggested that pleasant/positively valenced sounds could make the task seem easier, and provided the user with a stronger sense of control. These results, in combination with similar results from product sound quality, suggest not only that sounds are emotionally differentiable, but that emotions can be predictive of product assessment (Västfjäll, Kleiner and Gärling 2003). In sonification, where sound can take on any number of forms, 'pleasantness' and 'ecological validity' are championed in design, for the reason that their consideration makes the process of listening easier and increases the ability to perceive the desired information content (Vickers and Hogg 2006).

It has recently been posited that music might be considered a sonification of emotion: a potential challenge to traditional definitions of sonification (Schubert et al. 2011). The argument stems from the capacity of music (at times) to successfully communicate emotion – the composer or performer encoding an emotion, and the listener decoding. The conditions introduced by Hermann (2008) can be applied presently to clarify what qualifies as a sonification of emotion.

According to Hermann (2008), a sonification must be objective, systematic, reproducible and able to be used with different data. For sonification of emotion, this fundamentally requires an underlying data space that represents emotion, such that the sound can

https://doi.org/10.1017/S1355771813000411 Published online by Cambridge University Press

reflect properties and relationships in this space. There must furthermore be a precise definition for how each point in this data space becomes a sound, even to the point that sampling the data multiple times at the same coordinate will create structurally identical resulting sounds. As will be clear in the following sections, the sonification strategies introduced presently satisfy all of these criteria, and the features chosen for communication make the association with music secondary.

## 3. TWO MODELS FOR SONIFICATION OF EMOTION

From the previous discussion, the most advantageous avenue for development is the peripheral display of emotion, one that takes advantage of results from musical emotion. Sonification has thus far only made use of auditory-icons and earcons to convey short emotional states, while real-time continuous display has not yet been sufficiently developed. After discussing strategies for auditory display of emotion, two models are introduced for displaying arousal and valence, two theoretical dimensions of emotion. One of the models was designed to be more ecologically valid and pleasant, the other was designed computationally using a tool for music emotion recognition and specially designed software for analysis.

### 3.1. Strategies

Winters and Wanderley (2013) discuss in detail strategies for auditory display of emotion in a process monitoring setting. Although environmental sounds and music are two broad categories of sound, each capable of emotion induction and communication, music is chosen as the framework for development. It proves advantageous because of the flexibility of acoustic elicitors, the encompassing wealth of knowledge, and problems inherent to using environmental sound for emotion display.

Within music, there are many structural and acoustic cues that correlate with musical emotion and that might be used for communication (Gabrielsson and Lindström 2010; Juslin and Timmers 2010). Instead of haphazardly selecting from the available cues, a more psychologically grounded approach first considers psychological properties that would be advantageous to the contexts thus far mentioned. This directs attention to specific auditory-cognitive mechanisms responsible for auditory emotion expression, and the more limited set of acoustic cues to which they respond. The desired psychological properties for this sonification context include high induction speed, low volitional influence, and, importantly, dependence upon structural and acoustic content. Using the framework provided in Juslin and Västfjäll (2008), this narrows the list of potential mechanisms for induction to ‘brain-stem reflex’ and ‘emotional contagion’.

The brain-stem reflex is a biological mechanism, often triggered by sudden or loud changes in sound that bear immediate impact upon an organism’s survival. Structural and acoustic cues that can trigger this mechanism include loudness, sharpness, roughness, tonality and fluctuation strength, all of which are studied in detail in the psychoacoustics literature (Fastl and Zwicker 2007). Emotional contagion is a process whereby a sound triggers an emotion in virtue of having acoustic features that the listener perceives as expressing an emotion, and the listener then ‘mimicks’ this expression internally. Acoustic features that trigger this mechanism are shared with emotional speech (Juslin and Laukka 2003), and include tempo, loudness, loudness variability, high-frequency energy, pitch-level, pitch variability, pitch contour, attack and irregularity at the event-to-event level.

Using these features, it might be possible to create a systematic and reproducible mapping of an ‘emotion’, but to satisfy the objective and different data requirements of Hermann (2008) it is necessary to make a choice of underlying data space. For this purpose, the two-dimensional arousal/valence space is chosen. This so called ‘circumplex’ (Russell 1980) model of affect has been prevalent in both affective computing and musical emotion, and can be contrasted with basic or discrete models of emotion and models using more or different dimensions. In addition to its prevalence, other benefits include the continuous nature of its underlying data space and documented correspondence with discrete emotion models (Eerola and Vuoskoski 2011).

The following two sonifications implement a collection of these cues, differing insofar as they have been designed in two fundamentally different ways. In the first, the desire was to create a mapping strategy that would be pleasant, ecologically valid and perceptually clear for all points on the $AV$ space, such that it might even be usable in a concert setting (Winters, Hattwick and Wanderley 2013). By contrast, the second sonification was designed computationally using software for music emotion recognition that uses a linear combination of nine underlying signal characteristics. After briefly discussing the details of the mapping strategies, they are evaluated in Section 4.

### 3.2. Ecological design

The details of this model are presented in Winters et al. (2013), and are summarised here. The foundation of the sonification is a resonant object created using the DynKlank unit generator in SuperCollider, a programming environment for real-time audio synthesis. By using modal synthesis, DynKlank can

https://doi.org/10.1017/S1355771813000411 Published online by Cambridge University Press

produce realistic sounds resembling physical materials (e.g. wood, ceramic, glass) through independent control of resonant modes, their amplitudes and decay times. As with physical objects, to make sound, the object must be struck (i.e. 'excited'). In this case, excitation always comes through impulse in alternating left–right stereo channels.

To convey emotion, the sonification uses tempo, loudness, decay, roughness and mode. Increasing arousal increases the speed at which the object is excited as well as the overall loudness of the sound. Decreasing arousal increases the length of decay time, the time at which it takes the amplitude of the sound to decay by 60 dB. To convey valence, the original sound is copied and frequency shifted by a major/minor third, perfect fifth and perfect octave. As valence increases in magnitude, either positively or negatively, the amplitudes of the third, fifth and octave increase incrementally such that in a normalised arousal/valence $AV$ space, the third reaches maximum amplitude at $V = \pm 0.5$, the fifth reaches maximum amplitude at $V = \pm 0.75$, and the octave at $V = \pm 1$. The third is major or minor depending on whether valence is positive or negative respectively.[1] Finally, the second quadrant of the $AV$ space (i.e. low valence, high arousal) is conveyed using roughness. While within this region of the space, an identical copy of the original sound is pitch shifted up to 50 Hz with radial distance from the origin, and is increased in amplitude with radial distance from the line $3\pi/4$.

### 3.3. Computational design

The second model was designed with the goal of acoustically instantiating a computational model for music emotion recognition. The model chosen for this purpose was the MIREmotion function (Eerola, Lartillot and Toiviainen 2009) from the MIRToolbox, a MATLAB toolbox with many useful functions for audio-based music information retrieval. The MIREmotion function can generate emotion scores for each of five categorical emotion concepts (happiness, sadness, tenderness, anger and fear), and three emotional dimensions (activity, tension and valence). To determine each score, the model uses a linear combination of four to five audio-based descriptors, determined through a process of multiple linear regression on a database of 110 musical examples and a collection of 29 non-redundant features. Although three dimensions were available, Eerola et al. (2009) demonstrated moderate to high correlation between tension and the other dimensions, while the correlation between activity and valence was marginal. Reasoning that activity and arousal were closely related conceptually, manipulation was directed towards activity and valence.

In the MIREmotion function, activity is determined by the RMS, maximum value of the summarised fluctuation, spectral centroid, spectral entropy and spectral spread. Valence is determined by the standard deviation of the RMS (sdRMS), maximum value of the summarised fluctuation, novelty, mode and key clarity. From these features, the computational sonification manipulates RMS, sdRMS, key clarity and mode. These features were measured using 16-bit, 15-second wave files recorded from the sonification at desired data points in the $AV$ space. To have the greatest degree of control over these features, the fundamental sonification strategy was simplified to a bank of three sinusoidal oscillators, creating a root-position closed major G-chord on G3. To control RMS and sdRMS the sound as a whole was periodically amplitude modulated by a strictly determined square wave at 0.4 Hz. To control key clarity and mode, the amplitudes of the third and fifth were increased or decreased in amplitude. The strategy for conveying valence varied with position in the normalised $AV$ space: from $-1$ to 0 valence, sdRMS was systematically decreased, from 0 to 1 valence, the key clarity and, to a lesser degree, mode were systematically increased. Increasing activity was conveyed by increasing RMS, but at no point was there digital clipping in any of the measured audio files.

## 4. COMPUTATIONAL EVALUATION

Both of the models in Section 3 were designed with the goal of conveying a continuous arousal and valence emotion space. As previously discussed, their mapping strategies vary due to differences in design goals and methods. The first model was designed using acoustic cues suggested by the psychological study of musical emotion, while the second was designed computationally using a publically available tool for music emotion recognition and specially designed software for analysis.

After presenting the software and the computational results, both models are evaluated for their expected utility in both emotional communication and musical emotion research. This comparison brings attention to limitations of computational evaluation, but also its benefits, and the ways in which these difficulties can be addressed.

### 4.1. Software for analysis

For the purpose of evaluation, two GUI frameworks[2] were developed to analyse the output of the

[1] At this point, it is worth mentioning that coincidentally, Schubert et al. (2011) suggested the same mapping of tempo, loudness and mode.

[2] Freely available [Online]: https://github.com/mikewinters/MIREmotion-Visualizer

https://doi.org/10.1017/S1355771813000411 Published online by Cambridge University Press

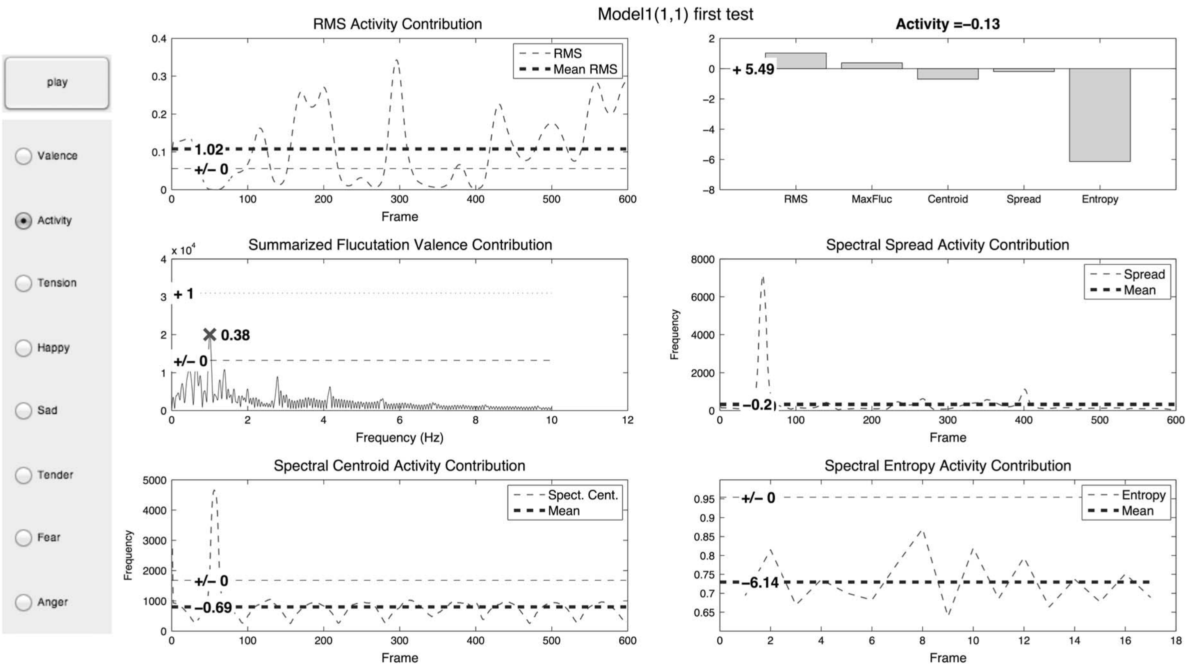

**Figure 1.** A figure displaying an activity visualisation using the myemotion function.

MIREmotion function on both individual and groups of soundfiles. Without these tools, the process of designing sounds is tedious: the default visualisations of the MIREmotion output do not indicate the contribution of the five underlying audio features to the emotion score, and do not represent these constitutive features in ways conducive to their systematic analysis and manipulation.

To analyse individual soundfiles, the 'myemotion' function visualises the audio features determining the emotion score under analysis, including the magnitude of their individual contribution and distance from a reference point, usually $\pm 0$. A 'play' button in the upper left-hand corner allows the user to listen to the analysed file, which is helpful for identifying distortions in the recording or understanding the temporal evolution of measured features. A collection of radio-buttons allows the user to quickly change the emotion dimension or concept under analysis, though only the visualisations for activity and valence have thus far been implemented. To facilitate documentation, if the user creates a title for the graph, it is used to automatically export .eps, .fig and a .wav file copy into a dated directory. A figure displaying the interface for activity is provided in Figure 1 and includes six graphs: one for each of the five constitutive audio features, and a bar-graph summary.

By contrast, the 'avmap' function visualises the distance of multiple individual wave files to desired points in an $AV$ space, and is designed for analysis of a mapping strategy as a whole. Positioned on a two-dimensional plot are the desired point (accumulated from the name of the wave file), the MIREmotion coordinate, and a line connecting the two points. Coloured markers of different shapes help to differentiate the measured points. Adjacent to this plot are two bar graphs displaying in detail the five audio features contributing to each emotion score. Each includes a 'detail' button triggering the myemotion visualisation for that dimension. Clicking on points of the graph makes their line-width and marker size bigger for visual feedback and changes the content of corresponding bar graphs. A unique title is generated for the two-dimensional graph indicating the Euclidean distance of all measured sound-files to their desired point on the graph. An example of an avmap visualisation for 16 soundfiles is provided in Figure 2.

## 4.2. Results

For computational evaluation, the MIREmotion function was applied to a collection of 49 15-second wav files recorded from evenly distributed points on each underlying $AV$ space. Because the function was trained using a seven-point Likert Scale on the interval from [1,7], the collection represents all possible integer combinations of activity and valence. The time scale of 15 seconds was chosen to closely match the average duration of the Soundtrack110 data set used to train the function (Eerola et al. 2009).

Figure 3 shows the comparison of the two sonifications side by side. For the non-computationally designed model, the average distance $d$ from the measured point $(V_m, A_m)$ to the desired point $(V_d, A_d)$

https://doi.org/10.1017/S1355771813000411 Published online by Cambridge University Press

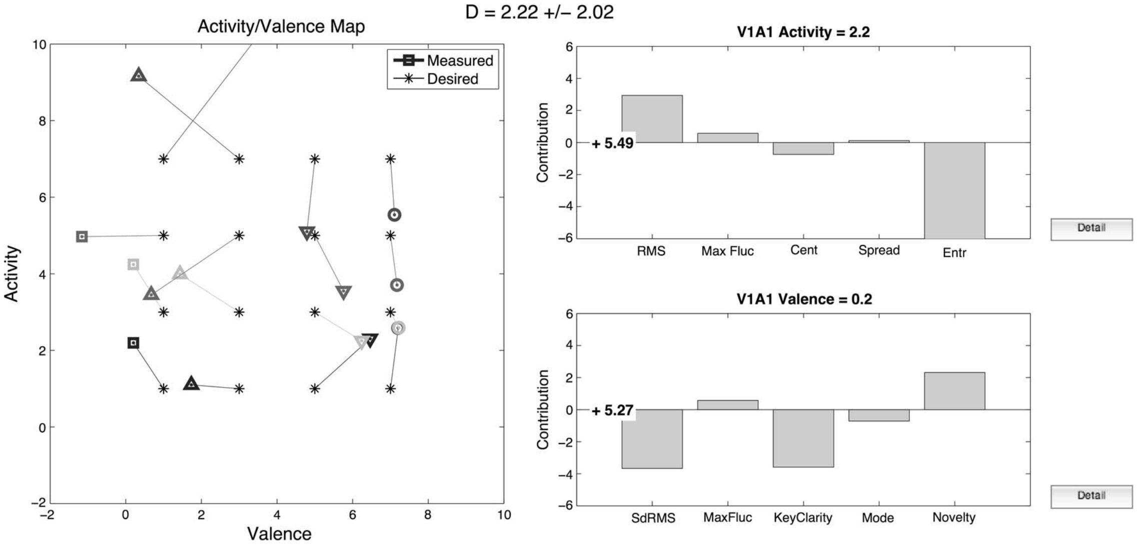

**Figure 2.** A figure displaying the avmap visualisation for a sonification of emotion. The long lines between markers and black stars indicate that the sonification does not conform well to the MIREmotion function.

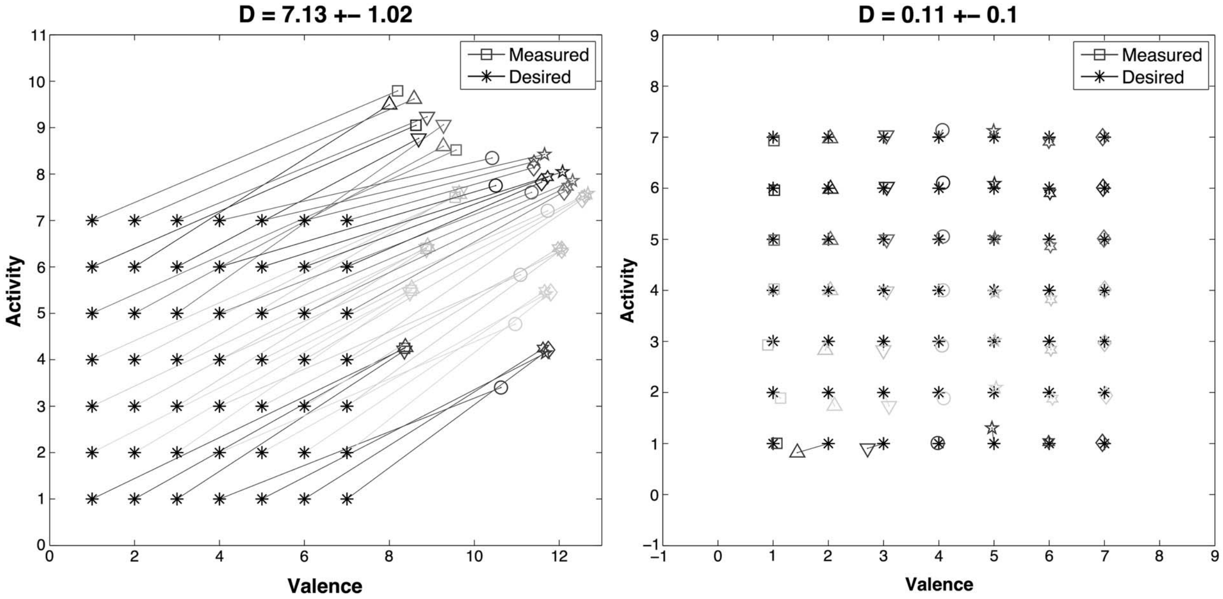

**Figure 3.** A comparison of the two sonifications analysed by the MIREmotion function using the avmap function. To the left, the ecological design, to the right, the computational design.

is $d = 7.13 \pm 1.02$. For the computationally designed model, $d = 0.11 \pm 0.10$, a difference factor of approximately 65.

From visual analysis, it is clear that the computationally designed sonification closely matches most of the desired points in the MIREmotion function. The worst scoring point on the sonification corresponds to $(V_d, A_d) = (2, 1)$ with $d = 0.57$. In general, points $(V_m, A_m)$ of poor performance are found in regions of low activity and valence. This issue stems from the inherent difficulty of creating points in this region for the MIREmotion function. Due to constraints of the model, the solution of a single sinusoid with strict control of both RMS and sdRMS is one of few possibilities. These two audio features however are implicitly connected, making the systematic variation of $V$ and $A$ in this quadrant more challenging.

It is also apparent that the ecological sonification does not conform well to the MIREmotion function. There is a systematic offset of all measured coordinates to a space between $A_m \approx (4, 10)$ and $V_m \approx (8, 13)$, and for all points $(V_m, A_m) > (V_d, A_d)$. Points of equivalent

https://doi.org/10.1017/S1355771813000411 Published online by Cambridge University Press

$A_d$ cluster together for $V_d=[1, 3]$ and $V_d=[5, 7]$, though the latter are systematically higher in valence than the former. Similarly, for every line of equivalent valence, activity incrementally increases from $A_d=[1, 4]$, and to a lesser extent from $A_d=[5, 7]$. In this light, the worst performance is in the region $A_d=[4, 7]$, $V_d=[4, 7]$, which clusters into a very small region between $V_m \approx (10, 12)$ and $A_m \approx (7, 8)$. In spite of these problems, it is interesting to note that the $AV$ structure is more or less preserved. The distribution of points in Figure 2 for instance, is considerably more random.

### 4.3. Analysis

To compare these two models computationally is one method for evaluation. As is evident, benefits include visual graphs (lending itself to visual analysis) and rapid evaluation. Computational models can also be used to direct mapping strategies and to increase the ‘accuracy’ of the sonification with respect to it. However, there are reasons why, in the present case, it would be unwise to base evaluation exclusively upon this method.

In this section it is argued that in spite of its performance in the computational test, the ecological design would still fare better in the contexts of emotional communication thus far mentioned. The reasons for this include the abundance and type of acoustic cues, and the more ‘natural’ sound created by the synthesis. Mindful of these limitations, reasons are provided why the computational approach should continue to be applied in evaluation and design.

#### *4.3.1. Limits of computational design*

As demonstrated here, it is possible to design a sonification of emotion to almost perfectly match a computational model of musical emotion using a small number of acoustic cues. At this limit, changes in the mapping may no longer increase computational accuracy, though may still benefit emotional communication. Further, to attain the highest degree of accuracy, it might even be advantageous to use simple sounds (such as sinusoids or noise) to provide greater systematic control of the constitutive audio features.

Thus, though each model represents an underlying arousal/valence space using structural and acoustic cues shared with musical emotion, it is instructive to highlight reasons why, in the present comparison, the ecological design would probably still be more useful for the communication contexts listed in Section 2.1. The first reason stems from the number and type of cues used in each model. Whereas the ecological design uses three cues to convey arousal (tempo, loudness and decay), the computational design used exclusively RMS (loudness) for this dimension and maintained a constant speed of amplitude modulation (tempo) for the entire $AV$ space. As for valence, similar strategies were used to convey high $V$ (key clarity/majorness), but the two differed in their approaches to low valence. The computational design used sdRMS, and the ecological design used minorness and roughness, a difference not only in number but also in type. Though the use of minor mode was desirable for low $V$, the use of sdRMS of a single sinusoid was dictated by model constraints discussed in Section 4.2. In either case, an abundance of cues is likely to have a greater emotional salience and/or magnitude than a singular cue. Using many cues also provides a degree of redundancy, which might be useful to users who attend to different qualities in the sound.

Besides for the cues, the ecological design also makes use of modal synthesis to create the underlying sound. This type of synthesis lends itself to creating ‘naturalistic’ sounds, which might resemble struck materials such as wood, metal or glass, for instance. On the other hand, the computational design uses a collection of three sinusoids, and for half of the space is limited to just one, centered on G3. The computational model has no mechanism for recognising something like ‘naturalness’, yet from the environmental sounds discussion in Winters and Wanderley (2013), it is a feature that should be preserved, having demonstrated emotional salience and behavioural impact in sonic interaction design (Lemaitre et al. 2012). Similarly, the naturalness in the ecological design might be expected to be preferred to the sinusoids of the computational design, in turn benefiting the utility of the display for communication.

#### *4.3.2. Benefits of computational design*

Although in this case, the ecological design is predicted to perform better in contexts of emotion communication, there are many reasons why the use of computational tools for evaluation and design should continue. Beyond rapid evaluation and graphs, they provide a framework for design, one that is already systematically informed by listeners’ emotional ratings. They are also valuable tools for music emotion research, acoustically instantiating an otherwise abstract mathematical model. The issues encountered in the present case originate in part from restrictions inherent to the model being used (i.e. constraints for low $V$, low $A$) and in part from the desire to clarify and address limitations of the computational approach.

That being said, more cues could be applied in the present computational design – specifically in areas not as restricted as the low $V$, low $A$ quadrant. From the previous discussion, contributing more cues would be beneficial to emotional communication and

https://doi.org/10.1017/S1355771813000411 Published online by Cambridge University Press

computational accuracy may still be maintained. Though the problem of computationally recognising 'naturalness' may persist, other computational models might be expected to be more sensitive to this feature, especially if the model was trained on listening tests including 'natural' and non-'natural' (i.e. sinusoids/noise) test sounds.

Further, neglecting these tools in sonification stymies collaboration with the field of musical emotion, an exchange this paper hopes to demonstrate as mutually valuable. As noted in Section 2.2, sonification offers musical emotion systematic and theoretically informed manipulations of acoustic cues. Although sonification by definition provides a systematic manipulation, and both models are theoretically informed, the computational model goes much further, acoustically instantiating an otherwise exclusively mathematical model of musical emotion and accurately covering a two-dimensional space. Though the ecological design uses suggestions from psychological studies, it follows no theoretical rules for their combination or implementation on an underlying $AV$ space.

By providing this acoustic instantiation, results from listening tests can also be directly applied towards refining the model and extending its predictive power. Although in music emotion recognition the highest scoring classifiers can reach accuracy levels of $\approx$ 65 per cent (Kim et al. 2010), it is possible that future performance would increase if cognitive factors due to recognition or genre preference were minimised. By using sonifications rather than music, these models would also become more predictive of the success of a sonification design than if trained using strictly musical examples. Better tools lead to better sonification designs, and can further contribute to the understanding of musical and auditory-induced emotion more generally.

## 5. CONCLUSION

In this paper, the subject of sonification of emotion was addressed in detail. Contexts favourable to real-time accurate auditory display were identified and the benefit to musical emotion research was highlighted. To frame this research, the current state of emotion in sonification was presented, including a reiteration of the necessary qualifications for a sound to qualify as a sonification of emotion. Strategies for display were presented that draw heavily upon research in musical emotion and target the auditory cognitive mechanisms of brain-stem reflex and emotional contagion. Two sonification mapping strategies were then presented that use these cues to display arousal and valence, two underlying dimensions of emotion. Both were evaluated computationally using the MIREmotion function and custom software for analysis. The significant difference in the performance in this test reflected fundamental differences in their method of design. Though the computational design performed better, the 'naturalness' and the number and type of cues used in the ecological design called to question whether this accuracy would equate to better performance in emotion communication. Mindful of these limitations in the computational approach, its application in sonification of emotion was supported for future research.

In total, this research demonstrates how tools and research in musical emotion can be applied to research in sonification of emotion, and also how sonification might be beneficial to music research. In this reciprocal relationship, computational tools can be applied as a design metric, but listening remains of utmost importance. It is hoped that this research can help to establish the display of emotion as a worthwhile pursuit in sonification, a pursuit that can make use of the wealth of resources from music rather than be confounded by them.

## REFERENCES

Brazil, E. and Fernström, M. 2011. Auditory Icons. In T. Hermann, A. Hunt and J.G. Neuhoff (eds.) *The Sonification Handbook*. Berlin: Logos.

Budd, M. 1985. *Music and the Emotions: The Philosophical Theories*. London: Routledge & Kegan Paul.

Crooke, D. 1957. *The Language of Music*. London: Oxford University Press.

Eerola, T. and Vuoskoski, J.K. 2011. A Comparison of the Discrete and Dimensional Models of Emotion in Music. *Psychology of Music* **39**(1): 18–49.

Eerola, T., Lartillot, O. and Toiviainen, P. 2009. Prediction of Multidimensional Emotional Ratings in Music from Audio Using Multivariate Regression Models. In *Proceedings of the 10th International Society for Music Information Retrieval Conference*. Kobe, Japan, 621–6.

Fastl, H. and Zwicker, E. 2007. *Psychoacoustics: Facts and Models*, 3rd ed. Berlin: Springer.

Gabrielsson, A. and Lindström, E. 2010. The Role of Structure in the Musical Expression of Emotions. In P.N. Juslin and J. Sloboda (eds.), *Handbook of Music and Emotion: Theory, Research, Applications*. New York: Oxford University Press.

Hermann, T. 2008. Taxonomy and Definitions for Sonification and Auditory Display. In *Proceedings of the 14th International Conference on Auditory Display*. Paris, France, 1–8.

Hermann, T., Drees, J.M. and Ritter, H. 2003. Broadcasting Auditory Weather Reports: A Pilot Project. In *Proceedings of the 9th International Conference on Auditory Display*. Boston, MA, USA, 208–11.

Hermann, T., Hunt, A. and Neuhoff, J.G. (eds.) 2011. *The Sonification Handbook*. Berlin: Logos.

Jee, E.-S., Jeong, Y.-J., Kim, C.H., Kwon, D.-S. and Kobayahi, H. 2009. Sound Production for the Emotional Expression of Social Interactive Robots. In V.A. Kulyukin (ed.), *Advances in Human-Robot Interaction*. Vukovar: InTech.

https://doi.org/10.1017/S1355771813000411 Published online by Cambridge University Press

Jee, E.-S., Jeong, Y.-J., Kim, C.H. and Kobayahi, H. 2010. Sound Design for Emotion and Intention Expression in Socially Interactive Robots. *Intelligent Service Robotics* **3**: 199–206.

Juslin, P.N. and Petri Laukka. 2003. Communication of Emotions in Vocal Expression and Music Performance: Different Channels, Same Code? *Psychological Bulletin* **129**(5): 770–814.

Juslin, P.N. and Sloboda, J.A. (eds.) 2010. *Handbook of Music and Emotion: Theory, Research, Applications*. New York: Oxford University Press.

Juslin, P.N. and Timmers, R. 2010. Expression and Communication of Emotion in Music Performance. In P.N. Juslin and J.A. Sloboda (eds.) *Handbook of Music and Emotion: Theory, Research, Applications*. New York: Oxford University Press.

Juslin, P.N. and 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. and Scott, J., et al. 2010. Music Emotion Recognition: A State of the Art Review. In *Proceedings of the 11th International Society for Music Information Retrieval Conference*. Utrecht, Netherlands, 255–66.

Kramer, G., Walker, B., Bonebright, T., Cook, P., Flowers, J. and 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 and K. Jensen (eds.) *Auditory Display: 6th International Symposium, CMMR/ICAD 2009, revised papers*. Berlin: Springer, 1–11.

Lemaitre, G., Houix, O., Susini, P., Visell, Y. and Franinovíc, K. 2012. Feelings Elicited by Auditory Feedback from a Computationally Augmented Artifact: The Flops. *IEEE Transactions on Affective Computing* **3**(3): 335–48.

Matsumoto, D. 2009. Display Rules. In D. Sander and K. R. Scherer (eds.) *Oxford Companion to Emotion and the Affective Sciences*. New York: Oxford University Press.

McGookin, D. and Brewster, S. 2011. Earcons. In T. Hermann, A. Hunt and J.G. Neuhoff (eds.) *The Sonification Handbook*. Berlin: Logos.

Ogihara, M. and Kim, Y. 2012. Mood and Emotion Classification. In T. Li, M. Ogihara and G. Tzanetakis (eds.) *Music Data Mining*. Boca Raton, FL: CRC Press.

Picard, R. 1997. *Affective Computing*. Cambridge, MA: The MIT Press.

Picard, R. 2009. Affective Computing. In D. Sander and K.R. Scherer (eds.) *The Oxford Companion to Emotion and the Affective Sciences*. New York: Oxford University Press.

Picard, R. and Daily, S.B. 2005. Evaluating Affective Interactions: Alternatives to Asking what Users Feel. In *CHI Workshop on Evaluating Affective Interfaces: Innovative Approaches*, Portland, OR.

Preti, C. and Schubert, E. 2011. Sonification of Emotions II: Live Music in a Pediatric Hospital. In *Proceedings of the 17th International Conference on Auditory Display*. Budapest, Hungary.

Russell, J.A. 1980. A Circumplex Model of Affect. *Journal of Personality and Social Psychology* **39**(6): 1161–78.

Schubert, E. 2010. Continuous Self-Report Methods. In P.N. Juslin and J.A. Sloboda (eds.) *Handbook of Music and Emotion: Theory, Research, Applications*. New York: Oxford University Press.

Schubert, E., Ferguson, S., Farrar, N. and McPherson, G.E. 2011. Sonification of Emotion I: Film Music. In *Proceedings of the 17th International Conference on Auditory Display*. Budapest, Hungary.

Supper, A. 2012. The Search for the 'Killer Application': Drawing the Boundaries around the Sonification of Scientific Data. In T. Pinch and K. Bijsterveld (eds.) *The Oxford Handbook of Sound Studies*. New York: Oxford University Press.

Västfjäll, D., Kleiner, M. and Gärling, T. 2003. Affective Reactions to Interior Aircraft Sounds. *Acta Acustica United with Acustica* **89**: 693–701.

Vickers, P. 2011. Sonification for Process Monitoring. In T. Hermann, A. Hunt and J.G. Neuhoff (eds.), *The Sonification Handbook*. Berlin: Logos.

Vickers, P. and Hogg, B. 2006. 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*. London, UK, 210–16.

Vinciarelli, A., Pantic, M., Heylen, F., Pelachaud, C., Poggi, I. and 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.

Walker, B.N. and Nees, M.A. 2011. Theory of Sonification. In T. Hermann, A. Hunt and J.G. Neuhoff (eds.) *The Sonification Handbook*. Berlin: Logos.

Winters, R.M. and Wanderley, M.M. 2013. Strategies for Continuous Auditory Display of Arousal and Valence. In *Proceedings of the 3rd International Conference on Music and Emotion*. Jyväskylä, Finland.

Winters, R.M., Hattwick, I. and Wanderley, M.M. 2013. 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*. Jyväskylä, Finland.

Yang, Y.-H. and Chen, H.H. 2011. *Music Emotion Recognition*. Boca Raton, FL: CRC Press.

https://doi.org/10.1017/S1355771813000411 Published online by Cambridge University Press