

Emotion in Motion: A Study of Music and Affective Response

Javier Jaimovich¹, Niall Coghlan¹ and R. Benjamin Knapp²,

¹ Sonic Arts Research Centre, Queen's University Belfast

² Institute for Creativity, Arts, and Technology, Virginia Tech
{javier, niall, ben}@musicsensorsemotion.com

Abstract. ‘Emotion in Motion’ is an experiment designed to understand the emotional reaction of people to a variety of musical excerpts, via self-report questionnaires and the recording of electrodermal response (EDR) and pulse oximetry (HR) signals. The experiment ran for 3 months as part of a public exhibition, having nearly 4000 participants and over 12000 listening samples. This paper presents the methodology used by the authors to approach this research, as well as preliminary results derived from the self-report data and the physiology.

Keywords: Emotion, Autonomic Nervous System, Physiological Database, Electrodermal Response, Pulse Oximetry, Self-Report Questionnaire.

1 Introduction

‘Emotion in Motion’ is an experiment designed to understand the emotional reactions of people during music listening, through self-report questionnaires and recording of physiological data with on-body sensors. Visitors to the Science Gallery, Dublin, Ireland were asked to listen to different song excerpts while their heart rate (HR) and Electrodermal Response (EDR) were recorded. The songs were chosen randomly from a pool of 53 songs, which were selected to elicit positive emotions (high valence), negative emotions (low valence), high arousal and low arousal. In addition to this, special effort was made in order to include songs from different genres, styles and eras. At the end of each excerpt, subjects were asked to respond to a simple questionnaire regarding their assessment of the song, as well as how it made them feel.

Initial analysis of the dataset has focused on validation of the different measurements, as well as exploring relationships between the physiology and the self-report data, which is presented in this paper.

Following on from this initial work we intend to look for correlations between variables and sonic characteristics of the musical excerpts as well as factors such as the effect of song order on participant responses and the usefulness of the Geneva Emotional Music Scale [1] in assessing emotional responses to music listening.

1.1 Music and Emotion

Specificity of musical emotions vs. ‘basic’ emotions. While the field of emotion research is far from new, from Tomkins theory of ‘discrete’ emotions [2] or Ekman’s [3] studies on the ‘universality’ of human emotions to the fMRI enabled neuroimaging studies of today [4], there is still debate about the appropriateness of the existing ‘standard’ emotion models to adequately describe emotions evoked through musical or performance related experiences. It has been argued that many of the ‘basic’ emotions introduced by Ekman such as anger or disgust are rarely (if ever) evoked by music and that terms more evocative of the subtle and complex emotions engendered by music listening may be more appropriate [5]. It is also argued that the triggering of music-related emotions may be a result of complex interactions between music, cognition, semantics, memory and physiology as opposed to a direct result of audio stimulation [6, 7]. For instance a given piece of music may have a particular significance for a given listener e.g. it was their ‘wedding song’ or is otherwise associated with an emotionally charged memory.

While there is still widespread disagreement and confusion about the nature and causes of musically evoked emotions, recent studies involving real-time observation of brain activity seem to show that areas of the brain linked with emotion (as well as pleasure and reward) are activated by music listening [8]. Studies such as these would seem to indicate that there are undoubtedly changes in physiological state induced by music listening, with many of these correlated to changes in emotional state.

It is also important to differentiate between emotions perceived in musical content, and those felt by the listener [9]. In the study presented on this paper we specifically asked participants how the music made them feel as opposed to any cognitive judgments about the music.

During the last few decades of emotion research, several models attempting to explain the structure and causes of human emotion have been proposed. The ‘discrete’ model is founded on Ekman’s research into ‘basic’ emotions, a set of discrete emotional states that he proposes are common to all humans; anger, fear, enjoyment, disgust, happiness, sadness, relief, etc. [10].

Russell developed this idea with his proposal of an emotional ‘circumplex’, a two or three axis space (valence, arousal and, optionally, power), into which emotional states may be placed depending on the relative strengths of each of the dimensions, i.e. states of positive valence and high arousal would lead to a categorization of ‘joy’. This model allows for more subtle categorization of emotional states such as ‘relaxation’ [11].

The GEMS scale [1] has been developed by Marcel Zentner’s team at the University of Zurich to address the perceived issue of emotions specifically invoked by music, as opposed to the basic emotion categories of much other emotion research. He argues that musical emotions are usually a combination of complex emotions rather than easily characterised basic emotions such as happiness or sadness. The full GEMS scale consists of 45 terms chosen for their consistency in describing emotional states evoked by music, with shorter 25 point and 9 point versions of the scale. These emotional states can be condensed into 9 categories which in turn group into 3 superfactors: vitality, sublimity and unease. Zentner also argues that musically evoked emotions are rare compared to basic/day-to-day emotions and that a random selection of musical excerpts is unlikely to trigger many experiences of strong musically

evoked emotions. He believes that musical emotions are evoked through a combination of factors which may include the state of the listener, the performance of the music, structures within the music and the listening [5].

Lab vs. Real World. Many previous studies into musically evoked emotions have noted the difficulty in inducing emotions in a lab-type setting [12, 13], far removed from any normal music listening environment. This can pose particular problems in studies including measurements of physiology as the lab environment itself may skew physiological readings [14, 15]. While the public experiment/installation format of our experiment may also not be a 'typical' listening environment, we believe that it's informal, open and non-mediated nature which at the very least provides an interesting counterpoint to lab-based studies, and potentially a more natural set of responses to the stimuli.

1.2 Physiology of Emotion

According to Bradley and Lang, emotion has "almost as many definitions as there are investigators", yet "an aspect of emotion upon which most agree, however, is that in emotional situations, the body acts. The heart pounds, flutters stops and drops; palms sweat; muscles tense and relax; blood boils; faces blush, flush, frown, and smile" [16], page 581. A plausible explanation for this lack of agreement among researchers is suggested by Cacioppo et al. in [17], page 174. They claim that "...language sometimes fails to capture affective experiences - so metaphors become more likely vehicles for rendering these conscious states of mind", which is coherent with the etymological meaning of the word emotion; it comes from the Latin *movere*, which means to move, as by an external force.

For more than a century, scientists have been studying the relationship between emotion and its physiological manifestation. Analysis and experimentation has given birth to systems like the polygraph, yet it has not been until the past two decades, and partly due to improvements and reduced costs in physiological sensors, that we have seen an increase in emotion recognition research in scientific publications [18]. An important factor in this growth has been responsibility of the Affective Computing field [19], interested in introducing an emotion channel of communication to human computer interaction.

One of the main problems of emotion recognition experiments using physiology is the amount of influencing factors that act on the Autonomic Nervous System (ANS) [20]. Physical activity, attention and social interaction are some of the external factors that may influence physiological measures. This has led to a multi-modal theory for physiological differentiation of emotions, where the detection of an emotional state will not depend on a single variable change, but in recognizing patterns among several signals. Another issue is the high degree of variation between subjects and low repeatability rates, which means that the same stimulus will create different reactions in different people, and furthermore, this physiological response will change over time. This suggests that any patterns among these signals will only become noticeable when dealing with large sample sizes.

2 Methodology

2.1 Experimental Design

The aim of this study is to determine what (if any) are the relationships between the properties of an excerpt of music (dynamics, rhythm, emotional intent, etc.), the self-reported emotional response, and the ANS response, as measured through features extracted from EDR and HR. In order to build a large database of physiological and self-report data, an experiment was designed and implemented as a computer workstation installation to be presented in public venues. The experiment at the Science Gallery – Dublin lasted for three months (June-August 2010), having nearly 4000 participants and over 12000 listening samples. The music selection included in its 53 excerpts contains a wide variety of genres, styles and structures, which, as previously mentioned, were selected to have a balanced emotional intent between high and low valence and arousal.

To be part of the experiment, a visitor to the Science Gallery would be accommodated by a mediator in front of one of the four computer workstations, and then follow the on-screen instructions to progress through the experiment sections (see Fig. 1 (b)). These, would first give an introduction to the experiment and explain how to wear the EDR and POX sensors. Then, the participant would be asked demographic and background questions (e.g. age, gender, musical expertise, music preferences, etc.). After completing this section, the visitor would be presented with the first song excerpt, which was followed by a brief self-report questionnaire. The audio file is selected randomly from a pool of songs divided in the four affective categories. This was repeated two more times, taking each music piece from a different affective category, so each participant had a balanced selection of music. The visitor was then asked to choose the most engaging and the most liked song from the excerpts heard. Finally, the software presented the participant plots of his or her physiological signals against the audio waveform of the selected song excerpts. This was accompanied with a brief explanation of what these signals represented.

Software. A custom Max/MSP¹ patch was developed which stepped through the different stages of the experiment (e.g. instructions, questionnaires, song selection, etc.) without the need of supervision, although a mediator from the gallery was available in case participants had any questions or problems. The software recorded the participants' questions and physiological data into files on the computer, as well as some extra information about the session (e.g. date and time, selected songs, state of sensors, etc.). All these files were linked with a unique session ID number which was later used to build the database.

Sensors and Data Capture. MediAid POX-OEM M15HP² was used to capture HR using infra-red reflectometry to detect heart rate and blood oxygenation. The sensor was fitted by clipping on to the participant's fingertip as shown in Fig. 1 (a).

¹ <http://cycling74.com/products/max/>

² http://www.mediaidinc.com/Products/M15HP_Engl.htm

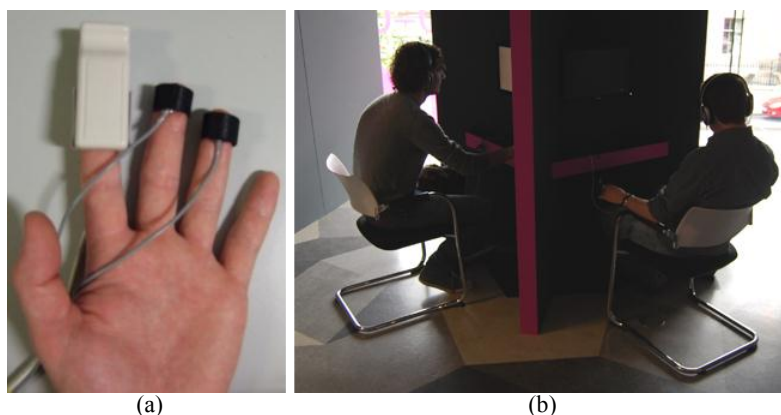


Fig. 1. (a) POX and EDR Sensors and (b) Participants during ‘Emotion in Motion’ experiment.

To record EDR, a sensor developed by BioControl Systems was utilized³. This provided a continuous measurement of changes in skin conductivity. Due to the large number of participants, we had to develop a ‘modular’ electrode system that allowed for easy replacement of failed electrodes.

In order to acquire the data from the sensors, an Arduino⁴ microcontroller was used to sample the analog data at 250 Hz and to send via serial over USB communication to the Max/MSP patch. The code from SARCduino⁵ was used for this purpose. For safety purposes the entire system was powered via an isolation transformer to eliminate any direct connection to ground. Full frequency response closed-cup headphones with a high degree of acoustic isolation were used at each terminal, with the volume set at a fixed level.

Experiment Versions. During the data collection period, variations were made to the experiment in order to correct some technical problems, add or change the songs in the pool, and test different hypothesis. All of this is documented in the database. For example, at the beginning participants were asked to listen to four songs in each session, later this was reduced to three in order to shorten the duration of the experiment. The questionnaire varied in order to test and collect data for different questions sets (detailed below), which were selected to compare this study to other experiments in the literature (e.g. the GEMS scales), analyse the effect of the questions in the physiology by running some cases without any questions, and also collect data for our own set of questions. The results presented in this paper are derived from a portion of the complete database with consistent experimental design.

Scales and Measures.

LEMtool The Layered Emotion Measurement Tool (LEMtool) is a visual measurement instrument designed for use in evaluating emotional responses to/digital media [21]. The full set consists of 8 cartoon caricatures of a figure expressing eight emotional states (Joy/Sadness, Desire/Disgust, Fascination/Boredom,

³ http://infusionsystems.com/catalog/product_info.php/products_id/203

⁴ <http://www.arduino.cc>

⁵ <http://www.musicsensorsemotion.com/2010/03/08/sarcduino/>

Satisfaction/Dissatisfaction) through facial expressions and body language. For the purposes of our experiment we used only the Fascination/Boredom images positioned at either end of a 5 point Likert item in which participants were asked to rate their levels of 'Engagement' with each musical excerpt.

SAM – Self Assessment Mannekin. The SAM is a non-verbal pictorial assessment technique, designed to measure the pleasure, arousal and dominance associated with a person's affective response to a wide range of stimuli [22]. Each point on the scale is represented by an image of a character with no gender or race characteristics, with 3 separate scales measuring the 3 major dimensions of affective state; Pleasure, Arousal, and Dominance. On the Pleasure scale the character ranges from smiling to frowning, on the Arousal scale the figure ranges from excited and wide eyed to a relaxed sleepy figure. The Dominance scale shows a figure changing in size to represent feelings of control over the emotions experienced.

After initial pilot tests we felt that it was too difficult to adequately explain the Dominance dimension to participants without a verbal explanation so we decided to use only the Pleasure and Arousal scales.

Likert Scales. Developed by the psychologist Rensis Likert [23], these are scales in which participants must give a score along a range (usually symmetrical with a mid-point) for a number of items making up a scale investigating a particular phenomenon. Essentially most of the questions we asked during the experiment were Likert items, in which participants were asked to rate the intensity of a particular emotion or experience from 1 (none) to 5 (very strong) or bipolar version i.e. 1 (positive) to 5 (negative).

GEMS – Geneva Emotional Music Scale. The 9 point GEMS scale [1] was used to ask participants to rate any instance of experiencing the following emotions: Wonder, Transcendence, Tenderness, Nostalgia, Peacefulness, Energy, Joyful activation, Tension, and Sadness. Again, they were asked to rate the intensity with which they were felt using a 5 point Likert scale.

Tension Scale. This scale was drafted by Dr. Roddy Cowie of QUB School of Psychology. It is a 5 point Likert scale with pictorial indicators at the Low and High ends of the scale depicting a SAM-type mannekin in a 'Very Relaxed' or 'Very Tense' state.

Chills Scale. This was adaptation from the SAM and featured a 5 point Likert scale with a pictorial representation of a character experiencing Chills / Shivers / Thrills / Goosebumps (CSTG), as appropriate, above the scale. The CSTG questions of the first version of the experiments were subsequently replaced with a single chills measure/question. For the purposes of the statistical analysis, the original results were merged using the mean to give a composite CSTG metric. This process was validated for consistency using both factor analysis and scale reliability test (Cronbach's Alpha of 7.83).

2.2 Song selection and description.

The musical excerpts used in the experiment were chosen by the researchers using several criteria: most were selected on the basis of having been used in previous experiments concerning music and emotion, while some were selected by the researchers for their perceived emotional content. All excerpts were vetted by the researchers for suitability. As far as possible we tried to select excerpts without lyrics, or sections in which the lyrical content was minimal⁶.

Each musical example was edited down to approximately 90 seconds of audio. As much as possible, edits were made at 'musically sensible' points i.e. the end of a verse/chorus/bar. The excerpts then had their volume adjusted to ensure a consistent perceived level across all excerpts. Much of the previous research into music and emotion has used excerpts of music of around 30 seconds which may not be long enough to definitely attribute physiological changes to the music (as opposed to a prior stimulus). We chose 90 seconds duration to maximize the physiological changes that might be attributable to the musical excerpt heard. Each excerpt was also processed to add a short (< 0.5 seconds) fade In/Out to prevent clicks or pops, and 2 seconds of silence added to the start and end of each sound file. We also categorized each song according to the most dominant characteristic of its perceived affective content: Relaxed=Low Arousal, Tense=High Arousal, Sad=Low Valence, Happy=High Valence. Songs were randomly selected from each categories pool every time the experiment was run with participants only hearing one song from any given category.

2.3 Feature extraction from physiology and database built

Database built. Once the signals and answer files were collected from the experiment terminals, the next step was to populate a database with the information of each session and listening case. This consisted in several steps, detailed below.

First, the metadata information was checked against the rest of the files with the same session ID number for consistency, dropping any files that had a wrong filename or that were corrupted. Subsequently, and because the clocks in each acquisition device and the number of samples in each recorded file can have small variations, the sample rates (SR) of each signal file were re-calculated. Moreover, some files had very different number of samples, which were detected and discarded by this process. To calculate the SR of each file, a MATLAB script counted the number of samples of each file, and obtained the SR using the length of the song excerpt used in that recording. Two conditions were tested: a) that the SR was within an acceptable range of the original programmed SR (acquisition device) and b) that the SR did not present more than 0.5% variation over time. After this stage, the calculated SR was recorded as a separate variable in the database.

Finally, the data from each song excerpt was separated from its session and copied into a new case in the database. This means that each case in the database contains variables with background information of the participant, answers to the song

⁶ The full list of songs used in the experiment can be found at http://www.musicsensorsemotion.com/demos/EmotionInMotion_Songs_Dublin.pdf

questionnaire, and features extracted from the physiological signals, as well as metadata about the session (experiment number, SR, order in which the song was heard, terminal number, date, etc.).

EDRtool and HRtool. Two tools developed in MATLAB were used to extract features from the physiological data: EDRtool and HRtool. Extraction of features included detection and removal of artefacts and abnormalities in the data. The output from both tools included the processed feature vectors and an indication of the accuracy of the input signal, which is defined as the percentage of the signal which did not present artefacts. This value can be utilized later to remove signals from the database that fall below a specified accuracy threshold.

EDRtool. EDRtool is a MATLAB function developed to pre-process the EDR signal. Its main function is the removal of electrical noise and the detection and measurement of artefacts and then to separate the signal into phasic and tonic components [24, 25]. These are shifted to start at zero for comparison purposes. Fig. 2 shows an example of the different stages of the EDRtool.

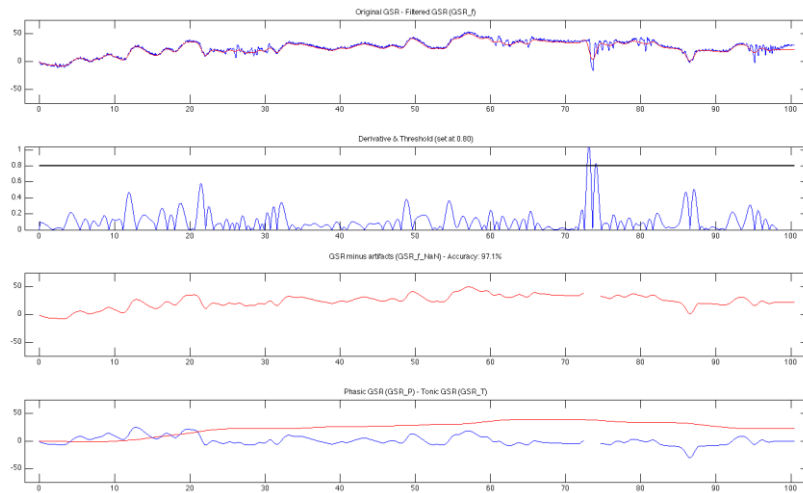


Fig. 2. Stages of the EDRtool on a Skin Conductance signal. The top plot shows the original (blue) signal and the low-passed filtered signal which removes any electrical noise. The next plot shows the artefact detection method, which identifies abrupt changes in the signal by taking the derivative. The 3rd plot from the top shows the filtered signal with the any artefacts removed. The bottom plot shows the phasic (blue) and tonic (red) components of the signal.

HRtool. HRtool is a MATLAB function developed to convert the data from an Electrocardiogram (ECG) or Pulse Oximetry (POX) signal into an HR vector. This involves three main stages, which are the detection of peaks in the signal (which differs for a POX and ECG signal), the measurement of the interval between pulses and calculation of the corresponding HR value. Finally, the algorithm evaluates the HR vector replacing any values that are outside the ranges entered by the user (e.g.

maximum and minimum HR values and maximum change ratio between two consequent pulses).

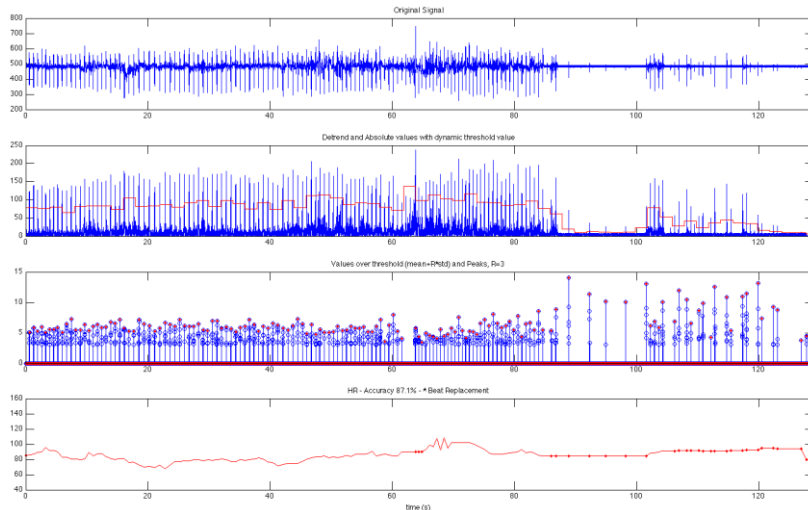


Fig. 3. Stages of the HRtool on an ECG signal. The top plot shows the raw ECG signal. The two middle plots show the peak detection stages, with a dynamic threshold. The bottom plot shows the final HR vector, with the resulting replacements of values that were outside the specified ranges (marked as dots in the plot). In this example, accuracy is at 87.1%, which falls below the acceptance tolerance for this experiment, and would be discarded as a valid case.

3. Research

3.1 Tackling the different research questions

We are not aware of any similar study with a database of this magnitude, which has made it difficult to apply existing methodologies from smaller sized studies [18, 20]. Consequently, a large portion of the research presented in this paper has been dedicated to do exploratory analysis on the results; looking to identify relationships between variables and to evaluate the validity of the questionnaire and physiological measurements.

3.2 Preliminary results from questionnaire

General Demographic Information. After removing all data with artefacts, as described previously, an overall sample size of 3343 participants representing 11041 individual song listens was obtained. The remaining files were checked for consistency and accuracy and no other problems found.

The mean DOB was 1980 (Std. Dev. 13.147) with the oldest participants born in 1930 (22 participants, 0.2%), 47% of the remaining participants were Male, 53%

Female, with 62.2% identifying as 'Irish', 37.8% coming from the 'Rest of the World'.

In the first version of the experiment participants heard 4 songs (1012 participants) with versions 2-5 consisting of 3 songs (2331 participants).

Participants were asked if they considered themselves to have a musical background or specialist musical knowledge, with 60.7% indicating 'No' and 39.3% indicating 'Yes'.

Interestingly, despite the majority of participants stating they had no specialist musical knowledge, when asked to rate their level of musical expertise from '1= No Musical Expertise' to '5= Professional Musician' 41.3% rated their level of musical expertise as '3'.

Participants were also asked to indicate the styles of music to which they regularly listen (by selecting one or more categories from the list below). From a sample of N=3343 cases, preferences broke down as follows: Rock 68.1%, Pop 60.3%, Classical 35%, Jazz 24.9%, Dance 34.2%, Hip Hop 27%, Traditional Irish 17%, World 27.9%, and None 1.2%.

Self-Report Data. An initial analysis was run to determine the song excerpts identified as most enjoyed and engaging. At the end of each experiment session, participants were asked which of the 3 or 4 (depending on experiment version) excerpts they had heard was the most enjoyable and which they had found most engaging. These questions appeared in all 5 versions of the experiment, making them the only ones to appear in all versions (other than the background or demographic questions).

The excerpts rated as 'Most Enjoyed' were James Brown 'Get Up (I Feel Like)' and Juan Luis Guerra 'A Pedir Su Mano' with these excerpts chosen by participants in 55% of the cases where they were one of the excerpts heard. At the other end of the scale, the excerpts rated lowest (fewest percentage of 'Most Enjoyed') were Slayer 'Raining Blood' and Dimitri Shostakovich 'Symphony 11, Op. 103 - 2nd Movement' with these excerpts chosen by participants in 13% of the cases where they were one of the excerpts heard.

Participants were also asked to rate their 'Liking' of each excerpt (in experiment versions 1-3). Having analysed the mean values for 'Liking' on a per-song basis, the songs with the highest means were Jeff Buckley 'Hallelujah' (4.07/5) and The Verve 'Bittersweet Symphony' (4.03/5). The songs with the lowest mean values for 'Liking' were Slayer 'Raining Blood' (2.66/5) and The Venga Boys 'The Venga Bus is Coming' (2.93/5).

The excerpt rated most often as 'Most Engaging' was Clint Mansell's 'Requiem for a Dream Theme' with this excerpt chosen by participants in 53% of the cases where it was one of the excerpts heard. At the other end of the scale, the excerpt rated lowest (fewest percentage of 'Most Engaging') was Ceolteori Chualainn 'Marbhna Luimigh' with this excerpt chosen by participants in 11% of the cases where it was one of the excerpts heard.

Interestingly when the mean values for 'Engagement' for each excerpt were calculated, Clint Mansell's 'Requiem for a Dream Theme' was only rated in 10th place (3.74/5), with Nirvana 'Smells Like Teen Spirit' rated highest (3.99/5), closely followed by The Verve 'Bittersweet Symphony' (3.95/5) and Jeff Buckley 'Hallelujah' (3.94/5). It was observed that while mean values for engagement are all

within the 3-4 point range, there are much more significant differences between songs when participants were asked to rate the excerpt which they found ‘Most Engaging’, with participants clearly indicating a preference for one song over another.

The excerpts with the lowest mean values for ‘Engagement’ were Primal Scream ‘Higher Than The Sun’ (3.05/5) and Ceolteoiri Chualainn ‘Marbhna Luimigh’ (3.09/5).

The excerpts with the highest mean values for Chills / Shivers / Thrills / Goosebumps were Jeff Buckley ‘Hallelujah’ (2.24/5), Mussorgsky ‘A Night on Bare Mountain’ (2.23/5) and G.A. Rossini ‘William Tell Overture’ (2.23/5). The excerpts with the lowest mean values for CSTG were Providence ‘J.O. Forbes of Course’ (1.4/5), Paul Brady ‘Paddys Green Shamrock Shore’ (1.43/5) and Neil Young ‘Only Love Can Break Your Heart’ (1.5/5).

An analysis was also run to attempt to determine the overall frequency of participants experiencing the sensation of CSTG. The number of instances where CSTG were reported as a 4 or 5 after a musical excerpt was tallied, giving 872 reports of a 4 or 5 from 9062 listens (experiment versions 1-3), meaning that significant CSTGs were experienced in around 10% of cases.

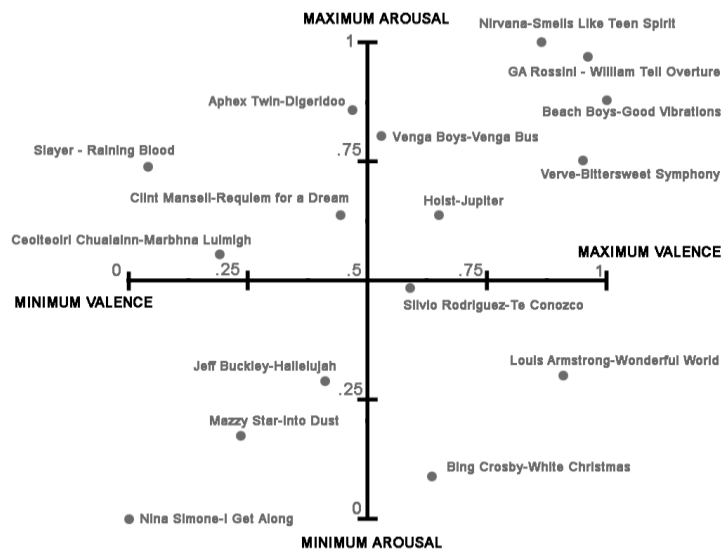


Fig. 4. Circumplex mapping of selected excerpts after a normalisation process to rescale the values 0-1 with the lowest scoring excerpt in each axis as ‘0’ and the highest as ‘1’.

A selection of the musical excerpts used (some of which were outliers in the above analyses) were mapped on to an emotional circumplex (as per Russell 1980), with Arousal and Valence (as measured using the SAM) as the Y and X axes. An overall tendency of participants to report positive experiences during music listening was observed, even for songs which might be categorised as ‘Sad’ e.g. Nina Simone. Arousal responses were a little more evenly distributed but still with a slight positive skew. It seems that while some songs may be perceived as being of negative affect or ‘sad’, these songs do not in the majority of cases induce feelings of sadness. It may

therefore be more appropriate to rescale songs to fit the circumplex from ‘saddest’ to ‘happiest’ (lowest Valence to highest Valence) and ‘most relaxing’ to ‘most exciting’ (lowest Arousal to highest Arousal) rather than using the absolute values reported (as seen on Fig. 4).

This ‘positive’ skew indicating the rewarding nature of music listening corroborates previous findings as documented in Juslin and Sloboda 2001. In future versions of this experiment we hope to identify songs that extend this mapping and are reported as even ‘sadder’ than Nina Simone.

3.3 Preliminary results from physiology

Features Extracted from Physiology. Due to the scope and nature of the experiment, the statistical analysis of the physiological signals has been approached as a continuous iteration, extracting a few basic features from the physiology, running statistical tests and using the results to extract new features. For this reason, the results from the physiology presented in this paper are still in a preliminary stage. The following features have been extracted from the 3 physiological vectors recorded in each case of the database (Phasic EDR, Tonic EDR and HR): Standard deviation of phasic EDR (*STD_EDRP*), mean of Phasic EDR (*mean_EDRP*), Tonic EDR final value divided by duration (*End_EDRT*), Tonic EDR trapezoidal numerical integration divided by duration (*Area_EDRT*), standard deviation of tonic EDR (*STD_EDRT*), difference between tonic EDR vector and linear regression of tonic start and end values (*Lin_EDRT*), mean of the 1st 10 raw EDR values (*Init_EDR*), mean HR (*HR*), mean heart rate variability (*mean_HRV*), HRV end value divided by duration (*End_HRV*), standard deviation of HRV (*STD_HRV*), square root of the mean squared difference of successive pulses (*RMSSD*), HRV low frequency (0.04-0.15Hz) component (*LF_HRV*), HRV high frequency (0.15-0.4Hz) component (*HF_HRV*) and ratio between *HF_HRV* and *LF_HRV* (*HtoL_HRV*).

Exploratory analysis and evaluation of measurements.

Dry skin issue. After removing any EDR signals that presented more than 5% of artefacts (measured by the EDRtool), preliminary analysis on several features extracted from the EDR vectors presented a bimodality in the distribution of the features, which did not correspond to any of the variables measured or changed during the experiment (e.g. gender, age, song, etc.). The mean of the first 10 samples of each signal was calculated and added to the database in order to analyse the initial impedance of each subject. Fig. 5 shows the distribution of this variable.

The distribution shows a clear predominance of a group of participants which presented very high initial impedance (around the 160 mark). Although the origin of this irregularity is not clear, it is equivalent to the measurement of the EDR sensor when it has an open circuit (e.g. no skin connection). Due to the decision to not use conductive gel during the experiment, it is possible that this abnormality corresponds to a large group of participants in which the sensor did not make a good connection with the skin, probably due to them having a drier skin than the rest of the participants. It is also interesting to point out that there were a few hundred cases in which the sensor failed to work correctly (e.g. cases with conductivity near zero). For these reasons, the number of cases used for the analysis was filtered by the *Init_EDR*

variable, looking for values that had normal impedance (above 160 and below 490). This procedure solved the bimodality issue; at the cost of significantly reducing the valid cases in the database by 37%.

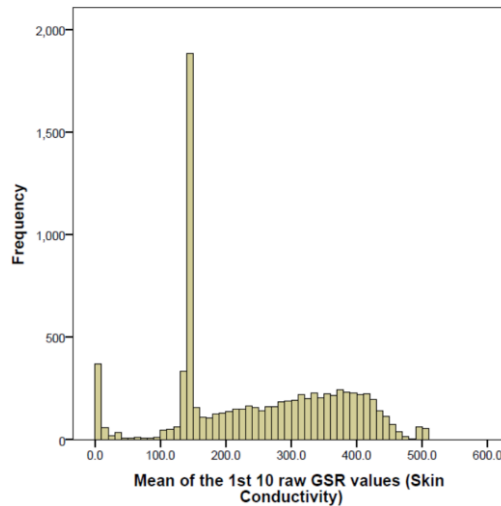


Fig. 5. Histogram of the mean of the 1st 10 samples of the EDR signal; equivalent to the initial conductivity. The histogram shows a large group of participants with an initial conductivity around the 160 mark (high impedance).

Correlation between physiological features and age. As expected, correlation between age and features extracted from physiology showed a negative relationship ($p < 0.01$ level, two-tailed) for several HR features (*STD_HRV*, *RMSSD*, *LF_HRV*, *HF_HRV*, *HtoL_HRV*), being the frequency features the ones with maximum correlation ($r < -0.4$).

Factor analysis of physiological features. Principal Component Analysis (PCA) was performed on a selection of features, excluding features with high degrees of correlation (it is important to state that all physiological features are derived from only two channels, EDR and HR, which can produce problems of multi-collinearity between features. This needs to be addressed prior to running a PCA). Principal Component Analysis shows 3 salient factors after rotation. These indicate a clear distinction between frequency-related features from HRV (Component 1: *STD_HRV*, *HF_HRV*, *LF_HRV*, *Age* and *RMSSD*), features from EDR (Component 2: *Area_GSRT*, *End_EDRT* and *STD_EDRP*) and secondary features from HRV (Component 3: *mean_HRV* and *End_HRV*).

Correlation between factors and questionnaire. The three salient components from PCA were correlated against a selection of the self-report questionnaire: Song Engagement, Song Positivity, Song Activity, Song Tension, Song CSTG, Song Likeness and Song Familiarity. Results show a relationship between components 1 and 2 with the self-report questionnaire (see Table 1).

Table 1. Correlation between components from physiology and questionnaire

Question	Correlation by component ($p < .001$)		
	1	2	3
Song Engagement	-.081	0.75	-
Song Positivity	-	.097	-
Song Activity	-	.110	-
Song Tension	-	.044	-
Song Chills/Shivers/Thrills/Goosebumps	-	-	-
Song Likeness	-.052	.061	-
Song Familiarity	-.060	.083	-

It is important to point out that the correlation coefficients presented above explain only a small portion of the variation in the questionnaire results. Furthermore, it is interesting that there was no significant correlation of CSTG with the 2nd component. Nevertheless, it is exciting to see a relationship between physiological features and self-reports such as song likeness, positivity, activity and tension.

Music Dynamics vs. Physiology. Analysis of temporal changes in correlation with the excerpt's dynamic has been explored. Preliminary results show a relationship between the three physiological vectors; phasic EDR, tonic EDR and HRV, with changes in the audio level of the music content. Fig. 6 shows two examples of pieces that present temporal correlation between physiology and music dynamic.

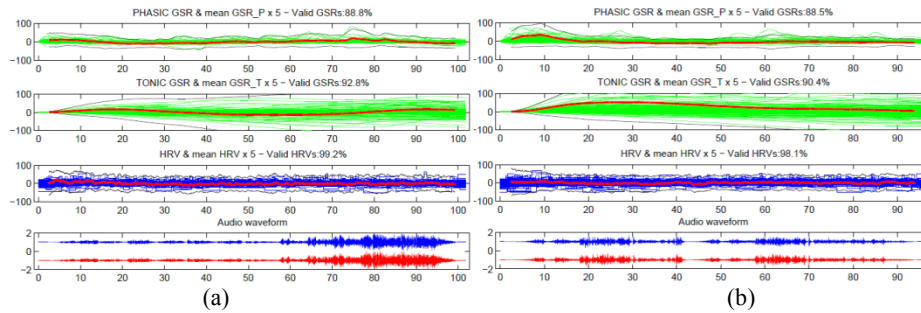


Fig. 6. Plots of Phasic EDR, Tonic EDR, HRV and audio waveform changes (top to bottom), during the duration of the song excerpt. Physiological plots show multiple individual responses overlapped, with the mean and SD overlaid on top in red and black respectively. Fig. 6 (a) plots are for *Mise Eire Muscalit*, of the Irish Traditional composer Sean O’Riada, with 125 cases. The plots in Fig. 6 (b) are for an excerpt of Mussorgsky’s *A Night on the Bare Mountain*, with 209 cases.

4 Discussion

Due to the public nature of this study, work has mainly been focused in improving the acquisition of signals, and the algorithms that correctly identify and remove noise and artefacts. Any unaccounted variation at this stage can impact the validity of the

statistical tests that use physiological measurements. It is important to point out that with the current sensor design, which requires no assistance and can be used by participants briefed with short instructions; we are obtaining almost 75% of valid signals. This has to be taken into account when calculating group sizes for experiments that require physiological sensing of audiences.

The analysis of the physiological measures shows high levels of dispersion between participants for the same feature, which seems to indicate that large sample sizes need to be maintained for future experiments. Furthermore, a significant amount of the participants presented little to no variation in the features extracted from EDR.

Nonetheless, the preliminary results presented in this paper are a significant indication of the possible relationships that explain the way we react to musical stimuli. Correlations between physiology and self-report questionnaire, in groups of this size, are a statement that this relationship undoubtedly exists. We are yet to further define the precise musical cues and variables that trigger changes.

Next steps in the analysis will be focusing on additional physiological descriptors, multimodal analysis of the dataset, looking at temporal changes (versus the current whole song approach) and measures of correlation and entrainment with musical features.

After the implementation in Dublin, 'Emotion in Motion' has been installed in public spaces in the cities of New York, Genoa and Bergen. Each iteration of the experiment has been enhanced and new songs have been added to the pool. We believe augmenting the sample size of this kind of studies is a requirement to start elucidating the complex relationship between music and our affective response to it.

Acknowledgements. The authors would like to thank Dr. Miguel Ortiz-Perez for his invaluable contribution to the software design for this experiment, as well as Dr. Rodderick Cowie and Cian Doherty from QUB for their help with the questionnaire design. Finally, we would like to express our appreciation to the Science Gallery, Dublin for their support and funding of this experiment.

References

1. The Geneva Emotional Music Scales (GEMS) | zentnerlab.com, <http://www.zentnerlab.com/psychological-tests/geneva-emotional-music-scales>.
2. Tomkins, Tomkins, S.S.: *Affect Imagery Consciousness - Volume II the Negative Affects*. Springer Publishing Company (1963).
3. Ekman, P., Friesen, W.V.: The repertoire of nonverbal behavior: Categories, origins, usage, and coding. *Semiotica*. I, 49-98 (1969).
4. Salimpoor, V.N., Benovoy, M., Larcher, K., Dagher, A., Zatorre, R.J.: Anatomically distinct dopamine release during anticipation and experience of peak emotion to music. *Nat Neurosci*. advance online publication, (2011).
5. Zentner, M., Grandjean, D., Scherer, K.R.: Emotions evoked by the sound of music: Characterization, classification, and measurement. *Emotion*. 8, 494-521 (2008).
6. Juslin, P.N., Västfjäll, D.: Emotional responses to music: the need to consider underlying mechanisms. *Behav Brain Sci*. 31, 559-575; discussion 575-621 (2008).
7. Balteş, F.R., Avram, J., Miclea, M., Miu, A.C.: Emotions induced by operatic music: Psychophysiological effects of music, plot, and acting: A scientist's tribute to Maria Callas. *Brain and Cognition*. 76, 146-157 (2011).

8. Trost, W., Ethofer, T., Zentner, M., Vuilleumier, P.: Mapping Aesthetic Musical Emotions in the Brain. *Cerebral Cortex*. (2011).
9. Gabriellsson, A., Juslin, P.N.: Emotional Expression in Music Performance: Between the Performer's Intention and the Listener's Experience. *Psychology of Music*. 24, 68-91 (1996).
10. Ekman, P.: An argument for basic emotions. *Cognition & Emotion*. 6, 169-200 (1992).
11. Russell, J.A.: A circumplex model of affect. *Journal of Personality and Social Psychology*. 39, 1161-1178 (1980).
12. Villon, O., Lisetti, C.: Toward Recognizing Individual's Subjective Emotion from Physiological Signals in Practical Application. Twentieth IEEE International Symposium on Computer-Based Medical Systems, 2007. CBMS '07. pp. 357-362. IEEE (2007).
13. Wilhelm, F.H., Grossman, P.: Emotions beyond the laboratory: Theoretical fundaments, study design, and analytic strategies for advanced ambulatory assessment. *Biological Psychology*. 84, 552-569 (2010).
14. Stranberg, T.E., Salomaa, V.: White coat effect, blood pressure and mortality in men: prospective cohort study. *European heart journal*. 21, 1714-1718.
15. Lantelme, P., Milon, H., Gharib, C., Gayet, C., Fortrat, J.-O.: White Coat Effect and Reactivity to Stress: Cardiovascular and Autonomic Nervous System Responses. *Hypertension*. 31, 1021 -1029 (1998).
16. Bradley, M.M., Lang, P.J.: Emotion and Motivation. *Handbook of Psychophysiology*. pp. 581-607 (2007).
17. Cacioppo, J.T., Bernston, G.G., Larsen, J.T., Poehlmann, K.M., Ito, T.A.: The Psychophysiology of Emotion. *Handbook of Emotions*. pp. 173-91. Guilford Press (2000).
18. Kreibig, S.D., Wilhelm, F.H., Roth, W.T., Gross, J.J.: Cardiovascular, electrodermal, and respiratory response patterns to fear- and sadness-inducing films. *Psychophysiology*. 44, 787-806 (2007).
19. Picard, R.W.: *Affective Computing*. M.I.T Media Laboratory, Cambridge, MA (1997).
20. Kim, J., André, E.: Emotion Recognition Based on Physiological Changes in Music Listening. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*. 30, 2067-2083 (2008).
21. Huisman, G., Van Hout, M.: Using induction and multimodal assessment to understand the role of emotion in musical performance. *Emotion in HCI – Designing for People*. pp. 5-7. , Liverpool (2008).
22. Bradley, M.M., Lang, P.J.: Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry*. 25, 49-59 (1994).
23. Likert, R.: A technique for the measurement of attitudes. *Archives of Psychology;Archives of Psychology*. 22 140, 55 (1932).
24. Van Olst, E.H., Orlebeke, J.F., Fokkema, S.D.: Skin conductance as a measure of tonic and phasic arousal. *Acta Psychologica*. 27, 262 (1967).
25. Christie, M.J.: Electrodermal activity in the 1980s: a review. *J R Soc Med*. 74, 616-622 (1981).