Towards Human-Computer Music Interaction: Evaluation of an Affectively-Driven Music Generator via Galvanic Skin Response Measures

Ian Daly, Asad Malik, James Weaver,
Faustina Hwang, Slawomir J. Nasuto
Brain Embodiment Lab
School of Systems Engineering
University of Reading
Reading, UK
Email: i.daly@reading.ac.uk

Duncan Williams, Alexis Kirke, Eduardo Miranda
Interdisciplinary Centre for Music Research
University of Plymouth
Plymouth, UK

***Abstract*—An affectively driven music generation system is described and evaluated. The system is developed for the intended eventual use in human-computer interaction systems such as brain-computer music interfaces. It is evaluated for its ability to induce changes in a listeners affective state.**

The affectively-driven algorithmic composition system was used to generate a stimulus set covering 9 discrete sectors of a 2-dimensional affective space by means of a 16 channel feed-forward artificial neural network. This system was used to generate 90 short pieces of music with specific affective intentions, 10 stimuli for each of the 9 sectors in the affective space. These pieces were played to 20 healthy participants, and it was observed that the music generation system induced the intended affective states in the participants. This is further verified by inspecting the galvanic skin response recorded from participants.

I. Introduction

Music is a powerful way of inducing changes in a listener's affective state. Consequently, it is extremely widely used for applications such as entertainment or music therapy [1].

We have recently developed an affectively-driven algorithmic music generation (AAC) system [2], [3], [4]. We intend for this system to integrate into a brain-computer music interface (BCMI) for use in affective state modulation. Specifically, we intend to use this as a form of passive BCI [5]. However, prior to this it is first necessary to evaluate the efficacy of the AAC system at inducing changes in a listener's affective state. Thus, this work comprises the first evaluation of the AAC system via the use of physiological signals.

We describe the AAC system and evaluate it via the use of galvanic skin response (GSR) measures of listener skin conductivity (a known correlate of felt emotions [6]).

Brain-computer music interfaces (BCMIs) aim to provide a user with a method for interacting directly with music via their brain activity [7], [8]. Our AAC system is developed for eventual use within a BCMI to generate a sequence of musical pieces to move the listener towards a targeted affective state.

The AAC music generation system is developed with the intention of generating affectively-charged music that might influence the affective state of the listener. We focus here on describing the music generation system. We go on to present our initial findings from the first stage of the development of this system. Specifically, we evaluate the music generator in terms of listener reports of their current affective states and in changes in their physiological measures induced by the music. We then go on to outline the future direction of the work.

II. Methods

A. Music generation system

An affectively-driven algorithmic composition system (AAC) [2], [3], [4] is used as a music generator. AAC is a rapidly developing field which combines affective sciences with computer-aided composition to generate new music with a particular affective intent (a target emotion). The distinction between whether affective compositional intent is induced in, or merely perceived by, the listener is well documented in music psychology [9], [10], [11], [12]. In AAC, the inclusion of bio-physiological readings facilitates a move towards reactive, feedback-driven systems, which might reliably induce affective states in the listener, according to their own preferences and subsequent physiological responses.

Rowe [13] describes three distinct approaches in algorithmic composition systems: generative, sequenced, or transformative. Sequenced systems use pre-composed musical passages, which are ordered by the algorithm. Generative systems create new musical passages according to particular rulesets. Transformative systems, the type evaluated in this paper, take existing musical passages as their source material and derive new sequences according to various functions. For example, a basic transformation might be to reverse the notes of a musical sequence - referred to as a retrograde transformation.

The generative system described here is used to generate the stimulus set for our experiments. The system is a transformative system. Specifically, it generates music by following existing seed musical structures. However, the closeness with which it follows this seed music can be varied. The system is based on a 16-channel feed-forward artificial neural network (ANN), as used in algorithmic music systems in [4], [14], [15], [16]. The ANN is trained on 12 bars of polyphonic piano music in C major at 120 bpm.

In order for the system to be useful, it must be capable of creating a theoretically infinite number of pieces without time constraints. This creates a number of challenges for a music generation system, particularly one which intends to target specific affective responses, as overall musical structure has also been shown to be a significant emotional communicator [17], [18]. We propose incorporating a range of musical features with known affective correlates; tempo, mode, pitch range, timbre, and amplitude envelope.

This material is then manipulated according to an affective mapping, which gives control over the five musical sub-features shown in Table I. Specific variations in each feature are used to imply different affective states in the generated material via a 3x3 Cartesian grid comprising arousal on the vertical axis and valence on the horizontal axis, analogously to the circumplex model of affect [19].

In this system, a Cartesian coordinate of (a, v) is used to refer to a given ratio of the five musical features, for example (1, 1) refers to the lowest corner of the space (low valence and low arousal) and forces the music generator to create stimuli based on its training material incorporating a slow tempo, a primarily minor key, a soft timbre, an amplitude envelope with considerable legato, and a low pitch spread. An example is given in Figure 1.

Fig. 1: Twelve bars of a generated musical figure with a target affective co-ordinate of (v1, a1); low valence and low arousal. Note the low pitch spread and an implied G minor key due to the presence of recurring B♭ and E♭ accidentals.

Co-ordinates with higher arousal generally include a larger pitch spread (range of notes), faster tempo, and harder timbres, whilst co-ordinates with higher valence generally utilize a major key. The particular ratios which inform this musical feature / affect matrix were derived in a large scale series of listening tests that are beyond the scope of this paper, further details may be found in [20].

B. Experiment design

This music generation system was tested to verify its suitability for use in applications such as BCMI systems.

Stimulus	Timbre	Key	Pitch spread	Tempo	Envelope
1-10	Soft	Minor	Low	Slow	Legato
11-20	Soft	Chromatic	Medium	Slow	Legato
21-30	Soft	Major	High	Slow	Legato
31-40	Medium	Minor	Low	Medium	No articulation
41-50	Medium	Chromatic	Medium	Medium	No articulation
51-60	Medium	Chromatic	Medium	Medium	No articulation
61-70	Hard	Minor	Low	Fast	Staccato
71-80	Hard	Minor	Low	Fast	Staccato
81-90	Hard	Major	High	Fast	Staccato

TABLE I: Stimulus number against musical features. Note that low valence is typically targeted by legato amplitude envelope, slow tempo, and softer timbral features, high arousal targeted by faster tempo, staccato envelope, and harder timbral features.

In order to do this, pre-generated music stimuli from the AAC system was played to participants, while recording their galvanic skin response (GSR), a measure known to correlate with felt emotion [6].

A fixation cross was first presented to the participants in each trial for a random duration between 1 - 2 s (drawn with replacement from a uniform distribution). Pre-generated music from the AAC was then played for 20 s and followed by 1 s silence. While the music played the participant was instructed to report their current felt emotions in terms of valence and arousal via the use of a FEELTRACE interface [21]. This was followed by a self-assessment manikin [22], which was used to allow participants to report their current felt affective state in terms of valence and arousal just after the music stopped.

Subsequently, a beep counting task was used as a distractor task. Participants were asked to listen to a sequence of beeps at two different tones, one high pitched and one low pitched. The high pitched tone occurred 20 % of the time and participants were asked to count how many times it occurs in this 15 s period. This was followed by a 2.5 s inter-stimulus interval before the start of the next trial.

The experiment was split into 5 runs with each run containing 18 trials. Between each run the participant was given a break of at least 2 minutes duration.

1) Stimuli: The AAC system was used to generate 90 different musical excerpts, each with duration of 20s. Ten different 20 s excerpts were generated to target each of the 9 discrete regions of the arousal-valence space which can be addressed by the music generation system. The stimuli were presented to each participant in random order.

2) Participants: Twenty individuals participated in the study. Each participant gave informed consent as per the procedures of the University of Reading research ethics committee. The participants had a mean age of 22 (range 19 - 30, standard deviation 1.45). Nine of the participants were female and all were right-handed. Each participant received £10.00 (GBP) for their participation.

3) Recording: Galvanic skin response was recorded, via a BrainAmp GSR sensor (BrainProducts, Germany), from the ventral medial phalanx positions on the index and middle fingers of the left hand. In addition, this study was conducted as a part of a larger study, in which several other physiological signals were also recorded from the participants at the same time. Specifically, the following signals were also recorded.

1) Electrocardiogram (ECG).
2) Galvanic skin response (GSR).
3) Blood oxygenation level (BPS).
4) Respiration rate (RR).
5) Head movement via an accelerometer (Acc.).
6) The electroencephalogram (EEG).

These signals will be described and analysed elsewhere.

C. Analysis

For each participant we attempt to determine whether the AAC music generator is able to place the participant in each of the targeted affective states. This is done in two ways. First, the participant's reports of their music-induced affective states, made while they listen to the music, are compared to the target affective state for each piece of music. Second, physiological measures, in the form of the galvanic skin response (GSR), of the participant's current felt affective states are compared to both the participant's subjective reports of their affective state and the target affective state for each piece of music.

The GSR is chosen because it has been reported to respond to changes in felt/induced arousal [23], [24]. This allows us to determine whether the music generation system is able to induce the target affective state within the participant and whether their subjective reports and physiological signals indicate the elicitation of the target affective state.

The alternative possibility is that participants are perceiving the emotional content of the music but not feeling it. In this case the GSR recorded from these participants is less likely to reflect changes in reported and targeted affective states.

1) Music induced affective states: Participant reports of their affective states are investigated to determine if the music generator induces the targeted affective states. The distinction between induction and perception has been previously documented in [11] and [10]. The mean FEELTRACE reports of valence and arousal recorded during the second half of the music listening period (seconds 10 - 20, relative to the start of the music) are used as measures of participant's reports of music induced valence and arousal. The second half of the music playing period is used because it is reasoned that it will take some time for participants to both become aware of their current affective state and to report it accurately via FFELTRACE. In addition, the participant's reports of their music-induced valence and arousal given via the self-assessment manikins (SAM) are also extracted.

Thus, for each trial the following measures are available.

1) The target affective state the music is aiming to induce. This is measured in the arousal valence space as either low, neutral, or high valence and arousal. Thus, there are a total of 9 different possible target affective states.
2) The participant's reports of their current music induced affective state, supplied by FEELTRACE.
3) Participant reports via SAM of their induced valence and arousal upon the end of the music play period.
4) Recordings of the participant's GSR signals as they listen to the music.

To determine if the music generator is producing the targeted affective states the participant's reports are compared to the affective states targeted by the music generator. Due to the different scales used in labelling targeted affective states and the participant's reporting their affective states, a direct numerical comparison is not possible. Instead Pearson's correlation coefficient is used to determine whether participant's reports of their felt affective states covary with the targeted affective states. This is done for each participant individually.

Additionally, participant's reports of their affective states are coarse grained into high, neutral, and low valence and arousal regions by dividing the available range of possible answers participants can give into three uniform segments. Confusion matrices are then constructed to investigate how accurately the generated music is able to produce the targeted affective states in participants for each region of the valence-arousal space.

2) Physiological signals: Galvanic skin response (GSR) signals are investigated to determine how they relate to the participant's reports of their music-induced affective states. GSR signals recorded from each participant are first low pass filtered (5 Hz, 2nd order Butterworth filter). They are then z-scored on a per participant basis to correct for inter-participant amplitude differences and then baseline corrected on a per trial basis (-4 s to 0 s relative to music start time).

This results in a set of time series of GSR signals, one per music listening trial. This set of signals is divided into sets of trials corresponding to high and low valence and high and low arousal trials by identifying those for which the participants average FEELTRACE rating increases or decrease in the last 10 s of the music playing trial compared to the first 10 s.

To determine if there is a significant difference between these conditions a t-test is used to compare conditions at each point in the time series of relative GSR values.

III. Results

A. Participant reports

Table II lists the correlations between the targeted and reported music-induced affective states in terms of valence and arousal for both FEELTRACE and SAM-based reporting for each participant.

Tables III, IV, V, and VI list the confusion matrices generated over all participants comparing targeted valence and arousal to valence and arousal reported by participants using FEELTRACE and SAM.

It may be observed that the majority of targeted affective states produce the correct response in the participants. However, the neutral affective states target produces the largest number of errors as participants rate these clips as either high or low valence and arousal quite often.

B. Physiological signals

The galvanic skin response (GSR) is investigated to determine if it changes in relation to the participant reports of valence or arousal. Data from all participants are grouped

P.	Valence				Arousal			
	FEELTRACE		SAM		FEELTRACE		SAM	
	r	p	r	p	r	p	r	p
1	0.656	< 0.01	0.649	< 0.01	0.563	< 0.01	0.664	< 0.01
2	0.409	< 0.01	0.435	< 0.01	0.524	< 0.01	0.484	< 0.01
3	0.609	< 0.01	0.634	< 0.01	0.433	< 0.01	0.459	< 0.01
4	0.608	< 0.01	0.609	< 0.01	0.516	< 0.01	0.374	< 0.01
5	0.638	< 0.01	0.614	< 0.01	0.675	< 0.01	0.641	< 0.01
6	0.613	< 0.01	0.537	< 0.01	0.698	< 0.01	0.633	< 0.01
7	0.499	< 0.01	0.521	< 0.01	0.408	< 0.01	0.479	< 0.01
8	0.628	< 0.01	0.573	< 0.01	0.612	< 0.01	0.665	< 0.01
9	0.540	< 0.01	0.441	< 0.01	0.488	< 0.01	0.419	< 0.01
10	0.663	< 0.01	0.644	< 0.01	0.479	< 0.01	0.442	< 0.01
11	0.606	< 0.01	0.495	< 0.01	0.426	< 0.01	0.241	< 0.01
12	0.657	< 0.01	0.592	< 0.01	0.694	< 0.01	0.581	< 0.01
13	0.678	< 0.01	0.662	< 0.01	0.711	< 0.01	0.701	< 0.01
14	0.613	< 0.01	0.643	< 0.01	0.664	< 0.01	0.635	< 0.01
15	0.613	< 0.01	0.576	< 0.01	0.509	< 0.01	0.394	< 0.01
16	0.581	< 0.01	0.629	< 0.01	0.379	< 0.01	0.461	< 0.01
17	0.227	0.031	0.032	0.764	0.125	0.239	0.390	< 0.01
18	0.594	< 0.01	0.595	< 0.01	0.661	< 0.01	0.676	< 0.01
19	0.454	< 0.01	0.558	< 0.01	0.429	< 0.01	0.555	< 0.01
20	0.683	< 0.01	0.602	< 0.01	0.535	< 0.01	0.585	< 0.01

TABLE II: Correlation coefficients (r) and corresponding parametric p-values comparing targeted affective states and reported music-induced affective states via FEELTRACE and SAM for each participant. The higher the correlation the more closely participants reports of their affective states match the affective states targeted by the AAC system.

		Targeted		
		Low	Mid.	High
Reported	Low	0.856	0.385	0.031
	Mid.	0.038	0.352	0.455
	High	0.106	0.263	0.514

TABLE III: Mean confusion matrix for target and reported valence (FEELTRACE).

		Targeted		
		Low	Mid.	High
Reported	Low	0.496	0.400	0.143
	Mid.	0.437	0.420	0.174
	High	0.066	0.179	0.683

TABLE IV: Mean confusion matrix for target and reported arousal (FEELTRACE).

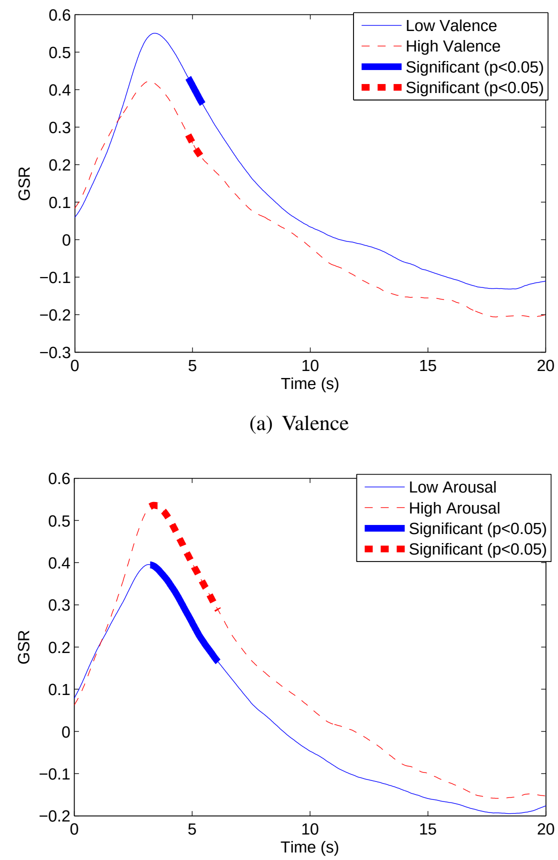
		Targeted		
		Low	Mid.	High
Reported	Low	0.702	0.269	0.047
	Mid.	0.124	0.420	0.492
	High	0.174	0.311	0.460

TABLE V: Mean confusion matrix for target and reported valence (SAM).

		Targeted		
		Low	Mid.	High
Reported	Low	0.451	0.340	0.151
	Mid.	0.432	0.373	0.210
	High	0.117	0.286	0.639

TABLE VI: Mean confusion matrix for target and reported valence (SAM).

together and baseline corrected, GSR time series are then compared between high and low valence and high and low arousal conditions. Valence and arousal are extracted from FEELTRACE and high and low conditions are identified by increases or decreases in the mean rating over the last 10 s compared to the first 10 s. Figure 2(a) illustrates the GSR signals recorded under different valence levels, while figure 2(b) illustrates the GSR signals recorded under different arousal conditions. The GSR signals are compared between conditions on a sample-by-sample basis using t-tests.



(a) Valence

(b) Arousal

Fig. 2: Mean baseline corrected GSR during music listening tasks for high and low valence and arousal conditions. Thick portions of the line indicate periods of significant difference between conditions ($p < 0.05$, t-test).

Significant differences are observed between the two conditions over consecutive samples. However, it is important to verify that these differences are not a result of a multiple comparisons error. To do this surrogate time series are generated by randomly re-shuffling samples. The number of consecutive significant differences between the conditions are then counted over 4,000 bootstrapped time series and compared to the original time series. Using this bootstrapping approach it is found that the consecutive significant differences in GSR signals recorded under high vs. low valence and high vs. low arousal conditions are not the result of a multiple comparisons error ($p < 0.01$).

It may be observed that the amplitude of the peak GSR decreases with increasing valence and increases with decreasing arousal. GSR has been widely reported to relate to changes in levels of stress [6]. Stress relates to conditions of high arousal and low valence. Thus, it may be speculated that GSR may best differentiate subsets of trials containing high valence and low arousal (low stress) reports against subsets of trials containing low valence and high arousal (high stress). This is found to be the case and is illustrated in figure 3.

Bootstrapping is again used to determine if the observed consecutive significant differences could result from a multiple comparisons error. It is observed that these differences could not have occurred by chance ($p < 0.01$).

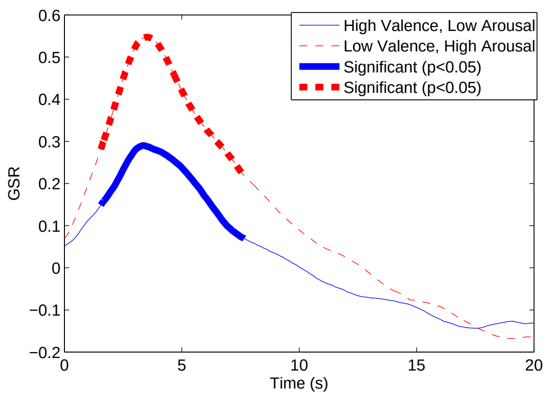


Fig. 3: Mean baseline corrected GSR during music listening tasks for subjectively reported high valence + low arousal and low valence + high arousal conditions. Thick portions of the line indicate periods of significant difference between conditions ($p < 0.05$, t-test).

GSR signals are now compared between musical clips which target high valence, low arousal and low valence, high arousal conditions. Specifically, we explore a different condition in which we compare the emotion the music generation system was targeting, not the emotion that participants reported. This allows us to investigate whether the music generator is actually inducing the targeted emotional states in participants. The GSR signals recorded under these two conditions are illustrated in figure 4. Bootstrapping is used again to determine that the observed consecutive significant differences could not have occurred as a result of a multiple comparisons error ($p < 0.01$).

IV. Discussion

An affectively-drive algorithmic composition (AAC) system is evaluated in terms of its' ability to induce changes in felt emotions. The galvanic skin response is used to provide the first physiological evaluation of the system. This system has a number of potential applications, including entertainment and brain-computer music interfacing (BCMI).

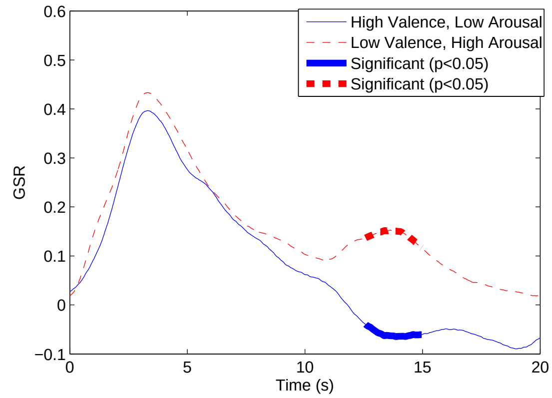


Fig. 4: Mean baseline corrected GSR during music listening for targeted high valence + low arousal and low valence + high arousal conditions. Thick portions of the line indicate periods of significant difference between conditions ($p < 0.05$, t-test).

This AAC system must be able to generate musical clips which are able to induce an affective state in the listener. Thus, it is important to verify that the music generation system does induce an affective state, and moreover that the induced state is the intended one. The enormous potential provided by biophysical sensors to allow control of such systems (for example, to achieve biofeedback derived control of music) was a key motivator for this experiment and has the potential to inform significant further work.

We verify this music generation system with a group of healthy participants. Each of the participants reported induced emotional responses while listening to the generated music that significantly correlated with the affective states the music generation system was targeting (as measured by both FEELTRACE and the self-assessment manikin (SAM)). The results indicate that the AAC system is effective at inducing the targeted affective state changes.

Additionally, we further verify that the music is actually inducing the targeted emotional responses in listeners by inspecting the GSR signals recorded from the participants. GSR peak amplitude, relative to baseline, significantly increases with participant reported arousal and significantly decreases with participant reported valence. GSR peak amplitude, relative to baseline, also significantly covaries with the targeted stress level of the music generation system (as defined by the high arousal, low valence and low arousal, high valence regions of the valence-arousal space).

Significant differences in relative GSR amplitude between conditions are identified by t-tests, which are applied on a sample-by-sample basis. It may be argued that multiple comparisons correction is necessary for this test. However, the traditional multiple comparisons tests (Bonferroni and false discovery rate correction) assume independence between the tests, which is not the case with our time series of GSR values. Indeed the observation that all significant differences in GSR value are consecutive, localised, and structured in time strongly suggests that they are meaningfully significant. We verify this by bootstrapping with shuffled sample orders. This confirms that the significant differences are not the result of a multiple comparisons error.

When comparing high and low valence conditions and high and low arousal conditions significant differences are observed in the first 2-8 s. Music can induce changes in emotional response in as little as 1 s [25]. Additionally, GSR can begin to respond to a stimulus within 1 s [26]. Thus, this response time is not unexpected.

The timing of the difference in the relative GSR amplitude is later in the trial when comparing targeted affective states to reported affective states. This may be due to different subsets of neutral target trials being reported by the participants as high or low valence. It may also be due to participants taking longer to identify and accurately report their current affective state when it is a function of both valence and arousal than when it is a function of just one of these two axes. Additionally, some confusion over neutral trials is also observed in the confusion matrices, in which it may be observed that participants frequently rate a piece of music with a neutral target as either high or low valence or arousal.

An additional important consideration is the inter-participant differences in GSR amplitude, which are reported elsewhere [27]. In order to adjust for these differences we z-scored the GSR signals on a per-participant basis and baseline corrected on a per-trial basis.

We find that the music generator was able to induce a range of targeted emotions. The relationships between these targeted emotions and GSR ratings also suggests the suitability of this music generation system for use in applications such as BCMI.

Our future work will investigate how this music generator can be incorporated into a BCMI. We also plan to investigate inter-participant differences in emotional responses to music and associated changes in a wider range of physiological responses to music. Finally, we will explore other signals, including EEG, as markers of music-induced emotions.

Acknowledgment

This work was supported by the EPSRC grants (EP/J003077/1 and EP/J002135/1). Data relating to this ongoing study will be made publicly available upon completion and publication of the entire study. Further information is available on request from the authors.

References

[1] O. Sacks, "The power of music.," *Brain : a journal of neurology*, vol. 129, pp. 2528–32, Oct. 2006.

[2] A. Mattek, "Emotional Communication in Computer Generated Music: Experimenting with Affective Algorithms," in *Proceedings of the 26th Annual Conference of the Society for Electro-Acoustic Music in the United States*, 2011.

[3] D. Williams, A. Kirke, E. R. Miranda, E. Roesch, I. Daly, and S. Nasuto, "Investigating affect in algorithmic composition systems," *Psychology of Music*, pp. 1–24, Aug. 2014.

[4] D. Williams, A. Kirke, J. Eaton, E. Miranda, I. Daly, J. Hallowell, E. Roesch, F. Hwang, and S. J. Nasuto, "Dynamic Game Soundtrack Generation in Response to a Continuously Varying Emotional Trajectory," in *Audio Engineering Society Conference: 56th International Conference: Audio for Games*, Audio Engineering Society, Feb. 2015.

[5] T. O. Zander and C. Kothe, "Towards passive brain-computer interfaces: applying brain-computer interface technology to human-machine systems in general.," *Journal of neural engineering*, vol. 8, p. 025005, Apr. 2011.

[6] M. V. Villarejo, B. n. G. Zapirain, and A. M. Zorrilla, "A stress sensor based on Galvanic Skin Response (GSR) controlled by ZigBee.," *Sensors (Basel, Switzerland)*, vol. 12, pp. 6075–101, Jan. 2012.

[7] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan, "Brain-computer interfaces for communication and control.," *Clin Neurophysiol*, vol. 113, pp. 767–791, June 2002.

[8] E. Miranda and A. Brouse, "Toward direct brain-computer musical interfaces," in *NIME '05 Proceedings of the 2005 conference on New interfaces for musical*, pp. 216–219, National University of Singapore, May 2005.

[9] K. Kallinen and N. Ravaja, "Emotion perceived and emotion felt: Same and different," 2006.

[10] A. Gabrielsson, "Emotion perceived and emotion felt: Same or different?," *Musicae Scientiae*, vol. Spec Issue, pp. 123–147, 2001.

[11] K. R. Scherer, "Which Emotions Can be Induced by Music? What Are the Underlying Mechanisms? And How Can We Measure Them?," *Journal of New Music Research*, vol. 33, pp. 239–251, Sept. 2004.

[12] J. K. Vuoskoski and T. Eerola, "Measuring music-induced emotion: A comparison of emotion models, personality biases, and intensity of experiences," *Musicae Scientiae*, vol. 15, pp. 159–173, July 2011.

[13] R. Rowe, *Interactive Music Systems: Machine Listening and Composing*. MIT Press, 1993.

[14] P. Purwins, B. Blankertz, and K. Obermayer, "Computing auditory perception," *Organised Sound*, vol. 5, no. 3, 2000.

[15] R. Bresin, "Artificial neural networks based models for automatic performance of musical scores," *Journal of New Music Research*, vol. 27, pp. 239–270, Sept. 1998.

[16] K. Wassermann, P. Verschure, and J. Manzolli, "Live soundscape composition based on synthetic emotions," *IEEE Multimedia*, vol. 10, pp. 82–90, Oct. 2003.

[17] A. Gabrielsson and E. Lindström, "The influence of musical structure on emotional expression.," in *Music and emotion: Theory and research*, pp. 223–248, 2001.

[18] P. Gomez and B. Danuser, "Relationships between musical structure and psychophysiological measures of emotion.," *Emotion*, vol. 7, pp. 377–87, May 2007.

[19] J. A. Russell, "A circumplex model of affect.," *Journal of Personality and Social Psychology*, pp. 1161–1178, 1980.

[20] D. Williams, A. Kirke, E. Miranda, I. Daly, J. Hallowell, F. Hwang, A. Malik, E. Roesch, J. Weaver, and S. Nasuto, "Affective calibration of a computer aided composition system by listener evaluation," in *Proceedings of the Ninth Triennial Conference of the European Society for the Cognitive Sciences of Music (ESCOM2015)*, 2015.

[21] R. Cowie, E. Douglas-Cowie, S. Savvidou, E. McMahon, M. Sawey, and M. Schröder, "'FEELTRACE': An Instrument For Recording Perceived Emotion In Real Time," in *Proceedings of the ISCA Workshop on Speech and Emotion*, pp. 19–24, 2000.

[22] M. M. Bradley and P. J. Lang, "Measuring emotion: the Self-Assessment Manikin and the Semantic Differential.," *Journal of behavior therapy and experimental psychiatry*, vol. 25, pp. 49–59, Mar. 1994.

[23] J. Gutrecht, "Sympathetic skin response.," *Journal of clinical neurophysiology : official publication of the American Electroencephalographic Society*, vol. 11, pp. 519–24, Sept. 1994.

[24] C. A. Torres, A. A. Orozco, and M. A. Álvarez, "Feature selection for multimodal emotion recognition in the arousal-valence space.," *Conference proceedings : ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual Conference*, vol. 2013, pp. 4330–3, Jan. 2013.

[25] E. Bigand, S. Filipic, and P. Lalitte, "The time course of emotional responses to music.," *Annals of the New York Academy of Sciences*, vol. 1060, pp. 429–37, Dec. 2005.

[26] D. R. Bach, G. Flandin, K. J. Friston, and R. J. Dolan, "Time-series analysis for rapid event-related skin conductance responses," *Journal of Neuroscience Methods*, vol. 184, pp. 224–234, Nov. 2009.

[27] P. A. Obrist, "Skin Resistance Levels and Galvanic Skin Response: Unilateral Differences," *Science*, vol. 139, pp. 227–228, Jan. 1963.