IMPLICIT PHYSIOLOGICAL INTERACTION FOR THE GENERATION OF AFFECTIVE MUSICAL SOUNDS

Sylvain Le Groux, Aleksander Valjamae, Jonatas Manzolli, Paul FMJ Verschure
Laboratory for Synthetic Perceptive, Emotive and Cognitive Systems
Audiovisual Institute
Pompeu Fabra University, Barcelona, Spain

ABSTRACT

Music is well known for affecting human emotional states, yet the relationship between specific musical parameters and emotional responses is still not clear. With the advent of new human-computer interaction (HCI) technologies, it is now possible to derive emotion-related information from physiological data and use it as an input to interactive music systems. This raises the question of how musical parameters are mapped to emotional states. We assess this question using both verbal and physiological responses. While most of the work on musical interfaces is based on explicit HCI, e.g. involving gestures, we study the potential of implicit interaction based on human emotional states. Our results show that a significant correlation exists between electrodermal activity, heart rate, heart rate variability and the subjective evaluation of well-defined musical parameters. Hence this demonstrates the feasibility of automated music composition based on physiological feedback. Providing implicit musical HCI will be highly relevant for a number of applications including music therapy, automatic generation of music for interactive virtual story telling and games, music for video games and physiologically-based musical instruments.

1. INTRODUCTION

It is generally acknowledged that music is a powerful carrier of emotions and the effect of music on emotional states has been established using many different self-report, physiological and observational means [8, 13]. Nevertheless, the precise relationship between musical parameters and emotional response is not clear. In the context of a mixed-reality environment called the eXperience Induction Machine (XIM) [2], we developed a real-world interactive composition and performance system which can produce musical structures and sonic textures in real-time, as a result of the interaction between the system and its human and non-human environment. The musical output of an interactive multimedia system is thought as a communication channel that can reflect, express in some way, the sensory, behavioral and internal state of the interactive system itself.

In order to generate original affective music, we investigate the mapping between emotions and the musical output of our real-time composition and performance system. We want to study the relationship between the musical parameters used to generate music and the listener's emotional states. Although several methods are available to assess the emotional state of listeners, time-varying physiological measurements seems to be particularly adequate for real-time interactive applications. Here, we focus on the correlations between the implicit, physiological measure of the emotional state of the listener and the musical parameters used to generate music.

2. BACKGROUND

Recent advances in human computer interaction have provided researchers and musicians with easy access to physiology sensing technologies. Although the idea is not new [9, 17], the past few years have witnessed a growing interest from the computer music community in using physiological data to generate or transform sound and music. In the literature, we distinguish three main trends: the use of physiology to modulate pre-recorded samples, to directly map physiological data to synthesis parameters (sonification), or to control higher level musical structures with parameters extracted from the physiology. A popular example of the first category is the Fraunhofer StepMan sensing and music playback device [3] that adapts the tempo of the music to the speed and rhythm of joggers' step, calculated from biosensoric data. While this approach appears efficient and successful, the creative possibilities are somewhat limited. In other work [1], the emphasis is put on the signal processing chain for analyzing the physiological data, which in turn is sonified, using adhoc experimental mappings. Although raw data sonification can lead to engaging artistic results, these approaches do not use higher-level interpretation of the data to control musical parameters. Finally, musicians and researchers have used physiological data to modulate the activity of groups of predefined musical cells [7]. This approach allows for interesting and original musical results, but the relation between the emotional information contained in the physiological data and the composer's intention is usually not explicited.

In this paper we assess the relationship between physiological response and music generation parameters. If specific musical parameters produce specific physiological responses (thus certain affective states), then those sound parameters can be used as a compositional tool to induce emotional states in the listener.

3. ADAPTIVE MUSIC GENERATOR

3.1. Situated Music, Real-world Composition

Our music generation, interaction and composition tools are based on work on our previous work on the synthetic composition engine RoBoser [12] and the psychoacoustics of emotional sonic expression [11].

The paradigm for musical interaction in XIM, called Real-World composition, is grounded in our work on large-scale interactive multi-media systems [5]. The overall aim is to integrate sensory data from the environment in real time and interface this interpreted sensor data to a composition engine. In this way unique emergent musical structures can be generated. In our previous work on Roboser, we have shown how the dynamics of a real-world system induced novelty in the micro-fluctuations of sound control parameters [12].

Our system has been previously used in various contexts. We designed an interactive music generator where the sensory inputs (motion, color, distance, ...) from a 3D virtual khepera robot living in a game-like environment are modulating musical parameters in real-time[6]. We also developed an automatic soundscape and music generator for a mixed reality space in Barcelona called the eXperience Induction Machine [2]. Finally our system was used to manage audio and music generation in re(PER)curso, an interactive mixed reality performance involving dance, percussion and video presented at the ArtFutura Festival 07 in Barcelona.

Here we propose to use sensory data provided by the listener's physiology to generate musical structures and validate the choice of adequate parameters for inducing specific emotional states. First, we give an overview of our musical generation system (a more detailed description can be found in [11]).

3.2. Parameterizing Music

From the beginning we have chosen a set of standard musical parameters for controlling the generation of music based on the requirement that their modulation should have a clear perceptual effect. We kept a list of parameters that has been extensively studied, and whose effect on emotional expression is widely acknowledged, as described in [8]. Those parameters are tempo, mode, volume register, tonality, consonance and rhythm for musical structure and articulation, brightness and harmonicity for the sound generation. With this set of parameters we are trying to study how the macro and micro levels can be used to express an affective sonification.

3.3. Musical Structure Generation

The generation of music is based on the real-world composition paradigm, where prepared musical material is dynamically modulated as the users interact with the mixed-reality space. All the programming was done in Pure Data [15]. When the interaction between people and the system takes place, these basic musical events are dynamically modified. The initial musical material is amplified, transformed, nuanced, as the interaction between the system and the users evolves. The relation between structural levels of music generation and emotional have been extensively studied elsewhere [8], and we decided to limit the scope of this paper to the quantitative study of a small set of only specific sound features.

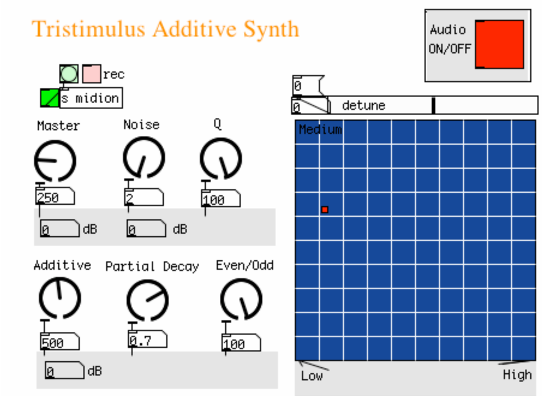

Figure 1. The tristimulus synthesizer allows for intuitive control over even/odd ratio, harmonicity, noisiness, and brightness. All the energy of the spectrum is in the low tristimulus spectral band if the red cursor is in bottom left part of the control grid, respectively the medium tristrimulus at the top left and high tristimulus at the bottom right. The total energy in the three tristimulus spectral bands stay constant.

3.4. Sound Generation

For the generation of the sound itself, we designed several Pure Data modules interfacing MIDI synthesizers and custom software synthesizers providing fine control over modulation of subtle timbral features that are perceptually relevant [8]. For fine timbre control, we implemented a simple tristimulus synthesizer which provides us with a simple and intuitive interface for controlling spectral properties of an additive synthesis model such as spectral content, harmonicity, or odd/even ration [14, 16].

4. MAPPINGS

A comparative review of literature on music and emotion [8] provided us with a set of different musical parameters that have been reported to elicit specific emotional responses . We decided to focus on a subset of parameters such as Loudness, Brightness, Harmonicity, Noisiness and Odd/Even ratio that can be easily produced by our synthesis engine. We also followed the well established bi-polar dimensional theory of emotions: hedonic valence or pleasantness and intensity of activation or arousal [18].

Emotions can then be placed in a two-dimensional emotional space, where the valence scale ranges from pleasantness (happy, pleased, hopeful, positive, etc.) to unpleasantness (unhappy, annoyed, despairing, negative, etc.), and the activation scale extends from calmness (relaxed, sleepy or peaceful) to high arousal (excited, stimulated, energized or alert).

5. METHOD

5.1. Subjects

We performed a pilot study using self-report and physiological measures. A total of 4 students from the university (4 males) ranging from 25 to 30 years of age participated in the experiment.

5.2. Experimental Setup

Each subject was seated in front of a computer, wired to the physiological equipment and listened to our set of sound stimuli via headphones. Sixteen sounds were randomly presented (2 repetitions per sound sample). Each sound snippet was defined by the pair sound feature/feature level (Loudness, Brightness, Harmonicity, Noisiness, Consonance, Tempo / Low, Medium, High). Each sound stimuli was 10 seconds long, and there was a pause of 18s between pause each presentation. During first 10 seconds of the pause, the subjects had to rate each sample in terms of their emotional content on a bidimensional scale (Valence, Arousal) [18] using Self-Assessment manikin pictorial scale (SAM) developed by Lang [10]. Specifically, for the valence dimension, SAM 9-point pictorial scale ranges from a figure showing a wide smile (rated as 9) to a frowning figure (rated as 1). The physiological data was recorded using the g-tech mobilab equipment (www.gtec.at). EDA (ElectroDermal Activity), heartrate and HeartRate Variability (HRV) were collected using 256 Hz sampling rate.

5.3. Stimuli

The stimuli consisted of a set of 8 synthesized sounds from the tristimulus model of timbre where loudness varied from -12 to 0db, a factor of frequency deviation from the harmonic spectrum (or inharmonicity) varied from 0 to 4, the filtered noise component (or noisiness) varied from -12db to 0db, and the factor of attenuation of even partial varied from 0 (all present) to 1 (none present).

6. RESULTS

Figure 3. represents correlations between verbal ratings of valence/arousal and physiological measures of heart rate, heart rate variability and electrodermal activity. As can be seen, subjective level of arousal positively correlates with EDA (p ¡ 0.001) and negatively with HRV data (p ¡ 0.05). Valence ratings positively correlate with HB rate (p ¡ 0.05). These results are in a good agreement with other findings [4] which show that 1) increase in EDA level can be used for monitoring arousal state; 2) heart rate (long term changes) increase for positive states and decrease for negative stimuli and 3) heart rate variability (short term changes) reduces with arousing and attention attracting stimuli.

		valence	arousal	HB	HBvar	EDA
valence	Pearson Correlation	1	.009	.421*	.121	-.185
	Sig. (2-tailed)		.963	.017	.510	.311
	N	32	32	32	32	32
arousal	Pearson Correlation	.009	1	.090	-.411*	.606**
	Sig. (2-tailed)	.963		.626	.019	.000
	N	32	32	32	32	32
HB	Pearson Correlation	.421*	.090	1	-.213	-.295
	Sig. (2-tailed)	.017	.626		.241	.101
	N	32	32	32	32	32
HBvar	Pearson Correlation	.121	-.411*	-.213	1	-.165
	Sig. (2-tailed)	.510	.019	.241		.366
	N	32	32	32	32	32
EDA	Pearson Correlation	-.185	.606**	-.295	-.165	1
	Sig. (2-tailed)	.311	.000	.101	.366	
	N	32	32	32	32	32

Figure 2. Correlation Table

To study the influence of individual sound parameters, we further looked at the correlations between verbal ratings and physiological responses for particular stimulus pairs. For noisiness, noticeable correlations were observed for EDA (p = 0.05, r = 0.8) and HRV (p = 0.09, r = 0.6). For brightness EDA positively correlated with arousal ratings (p ¡ 0.05, r = 0.8) and HRV showed a trend for negative correlation with arousal (p = 0.06, r = -0.7). For even/odd only EDA positively correlated with arousal ratings (p ¡ 0.001, r = 0.9). No significant correlations were found for harmonicity parameter. Loudness showed the negative correlation trend (p = 0.08, r = -0.7) between HRV and arousal ratings. The small number of participants doesn't allow for a more detailed significant analysis, and we are currently collecting more data with a larger set of participants and sensors including facial electromiography (EMG) and respiration. EMG is known to be a reliable measure of stimuli valence [19]and it should complement the heartrate data.

These results suggest that the selected sound stimuli were mainly modulating arousal. Additionaly, looking at the verbal ratings, our musical samples had rather neutral scores on the valence scale. In order to get deeper insights into mapping between musical components and affective states, we are currently working on a new set of samples to represent broader range of emotional responses.

7. CONCLUSIONS AND FUTURE WORK

In this paper we investigated the potential of using physiological data to extract information about emotional states that in turn can be used to control high-level musical attributes. We studied a set of well-defined sound parameters and showed that a variation of those parameters triggered significantly different physiological responses, corresponding to distinct affective states. We propose the use of this high-level emotional information to generate music instead of the low level raw data usually used in many sonification schemes.

We can imagine various applications of this framework in such diverse fields as musicotherapy, automatic generation of music for interactive story telling, music for video games, physiologically-based musical instruments. In particular we are investigating the use of these systems in music therapy for alzheimer patients and autistic children.

Our main objective was to build a system that would generate original interactive music based on the emotional state of its listeners (whether to illustrate or to induce specific emotions).

As a first approach, we have chosen simple synthesis techniques that allow for direct control over timbre, for those parameters that have been shown to have a significant impact on physiology. However a general framework that would allow to map perceptually relevant parameters to synthesis parameters for various, more complex and novel analysis/synthesis paradigms is still to be found. Advanced time series processing techniques would be necessary for learning to generate appropriate low-level synthesis parameters from high-level parameters.

To investigate in more details the potential of musical parameters to induce specific affective states, we also wish to expand our analysis to time-varying parameters, and to co-varying parameters. Our results however demonstrate that a rational approach towards the definition of interactive music systems is feasible.

8. ACKNOWLEDGEMENTS

This work was carried out as part of the PRESENCCIA project, an EU funded Integrated Project under the IST programme (Project Number 27731).

9. REFERENCES

[1] B. Arslan, A. Brouse, J. Castet, R. Lehembre, C. Simon, J. J. Filatriau, and Q. Noirhomme. A real time music synthesis environment driven with biological signals. In *ICASSP Proceedings*, volume 2, pages II–II, 2006.

[2] U Bernardet, S. Bermúdez i Badia, and P.F.M.J. Verschure. The eXperience Induction Machine and its Role in the Research on Presence. In *The 10th Annual International Workshop on Presence, October 25-27*, 2007.

[3] G. Bieber and H. Diener. Stepman - a new kind of music interaction. 2005.

[4] M.E Dawson, A.M Schell, and D.L. Filion. The electrodermal system. In *Handbook of psychopysiology*, 2000.

[5] K. Eng and al. Ada - intelligent space: an artificial creature for the swissexpo.02. *IEEE International Conference on Robotics and Automation*, 3:4154–4159 vol.3, 2003.

[6] Sylvain Le Groux, Jônatas Manzolli, and Paul F. M. J. Verschure. Vr-roboser: real-time adaptive sonification of virtual environments based on avatar behavior. In *NIME '07: Proceedings of the 7th international conference on New interfaces for musical expression*, pages 371–374, New York, NY, USA, 2007. ACM Press.

[7] Robert Hamilton. Bioinformatic feedback: performer bio-data as a driver for real-time composition. In *NIME '06: Proceedings of the 6th international conference on New interfaces for musical expression*, 2006.

[8] Patrik N. Juslin, John A. Sloboda, and Anonymous. *Music and emotion : theory and research*. Oxford University Press, Oxford ; New York, 2001.

[9] Benjamin R. Knapp and Hugh S. Lusted. A bioelectric controller for computer music applications. *Computer Music Journal*, 14(1):42–47, 1990.

[10] P.J. Lang. Behavioral treatment and bio-behavioral assessment: computer applications. In J.B Sidowski, J.H Johnson, and T.A. Williams, editors, *Technology in Mental Health Care Delivery Systems*, pages 119–137, 1980.

[11] Sylvain Le Groux, Jonatas Manzolli, and Paul F.M.J. Verschure. Interactive sonification of the spatial behavior of human and synthetic characters in a mixed-reality environment. In *Proceedings of the 10th Annual International Workshop on Presence*, 2007.

[12] Jonatas Manzolli and Paul F. M. J. Verschure. Roboser: A real-world composition system. *Comput.Music J.*, 29(3):55–74, 2005.

[13] L. B. Meyer. *Emotion and Meaning in Music*. The University of Chicago Press, 1956.

[14] H.F. Pollard and E.V. Jansson. A tristimulus method for the specification of musical timbre. In *Acustica*, volume 51, 1982.

[15] M. Puckette and Anonymous. Pure data: another integrated computer music environment, 1996.

[16] A. Riley and D. Howard. Real-time tristimulus timbre synthesizer. Technical report, University of York, 2004.

[17] David Rosenboom. Biofeedback and the arts: Results of early experiments. In *Computer Music Journal*, volume 13, pages 86–88, 1989.

[18] J. A. Russell. A circumplex model of affect. *Journal of Personality and Social Psychology*, 39:345–356, 1980.

[19] GE Schwartz, SL Brown, and GL Ahern. Facial muscle patterning and subjective experience during affective imagery - sex differences. *Psychophysiology*, 17(1):75–82, 1980.