Musical Engagement that is Predicated on Intentional Activity of the Performer with NOISA Instruments

Koray Tahiroğlu
Department of Media
Aalto University
School of Arts
Design and Architecture
FI-00076 AALTO Finland
koray.tahiroglu@aalto.fi

Thomas Svedström
Department of Media
Aalto University
School of Arts
Design and Architecture
FI-00076 AALTO Finland
thomas.svedstrom@aalto.fi

Valtteri Wikström
Department of Media
Aalto University
School of Arts
Design and Architecture
FI-00076 AALTO Finland
valtteri.wikstrom@aalto.fi

ABSTRACT

This paper presents our current research in which we study the notion of performer engagement within the variance and diversities of the intentional activities of the performer in musical interaction. We introduce a user-test study with the aim to evaluate our system's engagement prediction capability and to understand in detail the system's response behaviour. The quantitative results indicate that our system recognises and monitors performer's engagement successfully, although we found that the system's response to maintain and deepen the performer's engagement is perceived differently among participants. The results reported in this paper can be used to inform the design of interactive systems that enhance the quality of performer's engagement in musical interaction with new interfaces.

Author Keywords

Musical engagement, musical human-computer interaction, NIME, NOISA, novel musical instruments

ACM Classification

H.5.5 [Information Interfaces and Presentation] Sound and Music Computing — Systems, H.5.2 [Information Interfaces and Presentation] User Interfaces — Interaction styles

1. INTRODUCTION

New digital technologies have paradigmatically changed perception, practices and interaction in *New Interfaces for Musical Expression* field. What makes our relationship with the digital and technologically enhanced musical interfaces more interesting is not only the emerging technologies but also the interaction we engage in with them. Researchers, artists and designers have long played critical roles in building interactive systems, aiming to understand the nature of engaging interaction and its design constraints in music [2, 6]. However, the implication is often that it is not quite achieved, indeed most systems are designed to be used for short term entertainment. To allow open and comprehensive exploration of engaging experience, the availability of real-time adaptive features of the interface and an interaction technique that is predicated on a view of intentional activity of the performer should be considered.

In this paper, we present our current project, *Network of Intelligent Sonic Agents* - NOISA, that offers a solution for enhancing the quality of performer's engagement in musical interaction. NOISA project aims to provide an intelligent system that acts to maintain and deepen the performer's engagement by learning from the performer's actions and behavioural patterns while playing NOISA instruments (Figure 1). It monitors the internal states of a performer (e.g. motivation, affective states and reactiveness) and acts to maintain these states; if the performer loses individual motivation or interest, the system reacts by changing its physical control-behaviour of the interface in order to make the interaction interesting again.

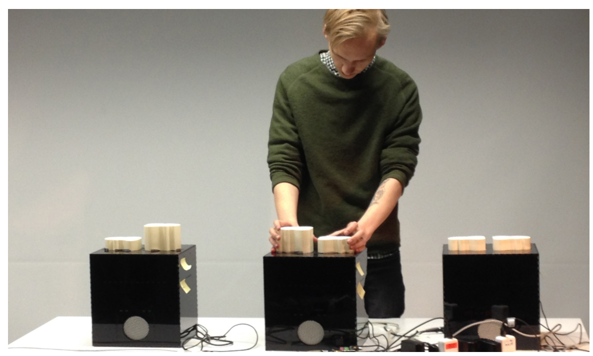

Figure 1: NOISA instruments.

2. RELATED WORK

Engagement has been broadly presented as a state where the performer or listener is emotionally connected and only paying attention to activities related to the music performance [2]. Similarly, Grace [10] suggests a musical engagement model that involves an interdependence between attention, emotion and action. The importance is further highlighted and discussed in performer-instrument relationship context [13]. Using techniques from embodiment theory, engagement is defined as an active state involving repetitive actions and possibly higher-level intentions [12]. It is also referred to "tension" [11] and "activation" [15].

The idea of engagement resembles the Csikszentmihalyi's flow theory [4] on many levels. Flow describes a fully concentrated and rewarding state where the subject is completely absorbed inside an action-cycle of the ongoing task. In the state of flow, the subjects have reported centering of attention, loss of ego, feeling of control, unambiguous goals and immediate feedback. In music context, flow has been shown to be deeper the more skilled the musician is [1].

Due to the complex spatio-temporal nature of a performer's bodily movements there has been an increasing interest in movement notation systems that can efficiently express single movements e.g. the Laban Movement Analysis (LMA) [9]. Initially, LMA was developed and used in dance context. Further, LMA has been used in applications to detect and recognise posture & action [3], as a method to analyse human movement [8] and developing systems for bodily expression recognition [5]. We are interested in the effort category of LMA, which focuses on the subtle characteristics of motion with respect to intentions. LMA states that every human movement is linked with an effort and can therefore reveal the inner aspects and changes behavioural states of the performer. With this approach, we propose a statistically validated multi-modal method that incorporates the use of subjective engagement sampling. The outcome is a real-time prediction of the performer's engagement.

3. NOISA SYSTEM

The NOISA system consists of three semi-autonomous instruments and a model that detects and attempts to maintain a high engagement level for the performer.

3.1 NOISA instrument design - agents

The handle in NOISA instrument was designed to have an ergonomic form for grabbing from several directions. The box itself accommodates the length of the handle, which is attached to a motorised fader (Figure 2). We used the potentiometer in the motorised fader for determining the position of the handle. The DC motor attached to the slider is used to control the slider's position. The handle was 3D printed from PLA plastic and all the other parts of the box were laser cut from MDF and acrylic. The electronics include an Arduino, a Raspberry Pi, two motorised faders with extra large motors and two capacitive touch sensors.

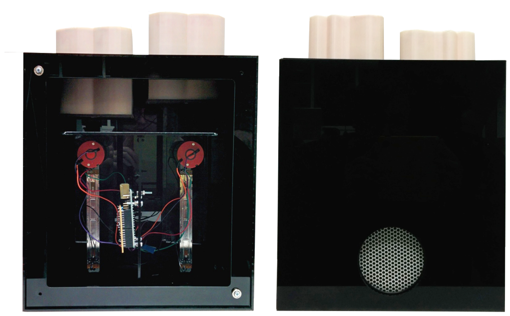

Figure 2: NOISA instrument is an agent in the system, communicating with the central server.

3.2 System description

In addition to three agents, the system consists of an engagement prediction module and a gesture learning module.

3.2.1 *Engagement prediction module*

The engagement prediction module provides a real-time estimation of the performer's engagement by analysing the performer's movements, facial expressions and actions. These features were derived from LMA and our previous study [14]. The features were then compared against a subjectively rated engagement in order to find the characteristics that occur in both states; the performer is engaged and not engaged. The method is presented in the following steps.

1: Attribute selection. First, we selected and formulated a large list of movement features, body postures, facial expressions and playing activity. We adapted the effort qualities of LMA to measure various aspects of the bodily attributes; e.g. *time* was used to measure the duration and acceleration of different body parts; *space* the trunk leaning and orientation of head, *weight* the excitation force and *flow* the smoothness and series of movement. Facial features include the states of the eyes and eyebrows, lips and cheeks, etc. The bodily and facial features were monitored with Kinect SDK 2.0 and the playing activity was derived from the combined amount of change in the raw input data resulting from the performer's interaction with NOISA instruments. In total there were 83 features.

Table 1: Final engagement attributes, attribute ranges and coefficients.

Attribute	Range	Coefficient
Constant		88.145
Slider Activity	[0, 384]	-0.01988
Torso Lean (back-front)	[-1, 1]	10.48
Lip Corner depressor	[-1, 1]	-4.762
Spine Shoulder Acceleration	[0, 51.1]	-0.182
Neck Acceleration	[0, 38.8]	-0.484
Head Acceleration	[0, 61.4]	-0.181
Right Shoulder Acceleration	[0, 75.6]	-0.223
Left Hand Tip Duration	[0, 6.0]	4.485
Right Wrist Duration	[0, 7.9]	1.541

2: Data gathering. We organised a data gathering session in Oct. 2014 and there were 21 volunteer participants to play the NOISA instrument. We videotaped each session and recorded the data of all the attributes. One session lasted approximately for 10 minutes. After each session, the recorded video was projected on the screen and the participant was asked to rate continuously his/her own engagement during the performance by using a slider controller.

3: Attribute and engagement analysis. Next, we compared the subjectively sampled engagement and 83 recorded attributes. By comparing the high-engagement (75^{th} percentile) and low-engagement (25^{th} percentile) segments, we found 22 attributes that showed statistical difference between the two percentiles. Then, we applied multiple linear regression (MLR) analysis to these reduced 22 attributes. Utilising forward feature selection we selected one attribute at a time, which brought the largest improvement in the model prediction. A "leave-one-subject-out" cross-validated paradigm was used. When the model prediction could not be improved any further, the attribute selection was finished and the final model was re-trained using the complete data set. The resulting attributes along with their engagement coefficients are presented in Table 1.

4: Real-time system. In the final stage, we formulated an equation to apply the results of the regression analysis. Engagement is a sum of the products of the attributes and their coefficients that were presented in Table 1. The engagement was calculated every 100ms. The equation is:

$$e = \sum_{i=1}^{n} a_i c_i \tag{1}$$

where e denotes engagement, a an attribute and c the corresponding coefficient.

3.2.2 *Gesture learning module*

There are two modes of operation; active learning mode and the exploratory mode. The system is in the active learning mode when the performer is touching either of the vertical handles, and in the exploratory mode while not touching.

Active learning mode. In the active learning mode the system learns and classifies the musical gestures in real-time based on analysis of local minima, maxima and static time segments. Each gesture is ranked with the engagement level that occurred during its performance. Furthermore, each gesture cluster has an engagement value that is the average of the gestures inside that group. The Density-based Spatial Clustering of Applications With Noise (DBSCAN) algorithm was used for clustering, and dynamic time-warping (DTW) was used for the gesture-wise distance metric.

Exploratory mode. In this mode, the system attempts to maintain and re-gain the engagement by reproducing the players gestures at the right moment. First, it seeks the gesture cluster that has the highest group engagement. Then, inside that cluster, it selects the gesture with the highest engagement and physically moves the instrument thus reproduces the sound. After the gesture is reproduced, it is re-evaluated with the current engagement value. If the gesture deepens the engagement, the corresponding cluster is more likely to be selected also later. If the player is facing towards the active agent, the current engagement will have a higher impact on the updated engagement value.

3.3 Audio synthesis

The first sonic agent creates dynamic rhythmic patterns through pulses generated by the fixed grain sizes in granular synthesis module. The right handler determines the index of the grains in each four grain-player unit and the left handler sets the pitch values to be played for each index of the grains. In the second sonic agent, sound is created by multiple oscillators which are connected to each other within an additive structure. The performer can change the frequency values of the first oscillator and detune the unique tone of the oscillators. The last sonic agent's audio synthesis is based on convolution of two audio signals; the performer can control the frequency values of the sawtooth signals and the playing position of the playback sampler.

4. USER TEST STUDY

We conducted a user-test study to evaluate the engagement prediction accuracy and the automatic response behaviour. The prediction accuracy was evaluated by comparing the subjectively rated engagement and the system's prediction. We tested the system with and without automatic responses. The hypothesis was that the system would monitor performer's engagement level in a high accuracy and respond in such a way that would capture the attention and stimulate action. It is expected that the system would influence the attention-action-engagement loop, as presented by [10], at the right moment and thus produce higher engagement than the non-automatic system. Furthermore, we studied the system's effect on cognitive load by using the NASA Task Load Index (NASA TLX) [7].

We invited 19 volunteer participants (aged 19-61, 7 females), to take part in the study that lasted approximately 45 minutes. Among the participants were eight amateur and two expert musicians. The study was held in a lightly acoustically treated room under the supervision of the researcher. Each session was videotaped. The study consisted of the actual test under two conditions, *automatic* and *manual* mode, in a randomised order. Each participant was first familiarised with the NOISA instruments. Later, the participant was asked to perform an improvised music under both conditions. After each of the two performances, the participant rated his/her continuous engagement during the performance while watching the video (Figure 3). The participants also filled in a NASA-TLX questionnaire.

At the end of the study, the participants filled in a survey question form. The first category was about their personal details (age, gender) and musical expertise on a scale of none-amateur-expert. The second category evaluated the participant's observations of the automatic behaviour with a multiple-choice selection of descriptive words, and with two numerical ratings scaled between 0 (negative)-10 (neutral)-20 (positive). The first rating considered the system predictability. The second rating evaluated how much NOISA encouraged to play and explore new ideas. Third category evaluated the overall experience and the sound design with an adjective selection task and ratings.

Figure 3: User rating his subjective engagement.

4.1 Results

4.1.1 Engagement prediction accuracy

We first compared the predicted engagement and the user reported engagement. The engagement vectors were scaled to match each subject's minimum and maximum ratings, because the behaviour of the system did not depend on the absolute value of the engagement, but the relationship of engagement values at different times during the experiment.

To validate the engagement prediction module, p-values for the correlation between the predicted engagement and the user reported engagement were calculated. The null-hypothesis of no correlation can be rejected, because the p-values were statistically significant for both the automatic (Pearson: $p = 0.0104$, Spearman: $p = 6.27 \cdot -24$) and the manual (Pearson: $p = 3.88 \cdot 10^{-9}$, Spearman: $p = 2.96 \cdot 10^{-16}$) conditions. Figure 4 shows the examples of the prediction results compared to the reported engagement.

4.1.2 NASA Task Load Index

The results of the NASA-TLX were analysed with Wilcoxon signed-rank test. It was found that during automatic condition, there was an increment in the overall task load ($p = 0.024$) and in the mental demand ($p = 0.039$). The mean task load index and standard deviations were $M = 41.4$, $SD = 13.3$ for automatic and $M = 36.6$, $SD = 12.6$ for manual condition. For mental demand the results during the two conditions were $M = 51.8$, $SD = 21.3$ for automatic and $M = 42.2$, $SD = 23.2$ for manual. There were no significant differences on physical demand, temporal demand, frustration, performance or effort.

4.1.3 Automatic behaviour

Two-sample t-test was performed on the mean values of the subjectively rated engagement. Statistically significant differences were not found ($p = 0.769$) between the automatic ($M = 82.3$, $SD = 17.1$) and manual ($M = 80.8$, $SD = 14.2$) conditions. We also analysed how the automatic responses affected the engagement by comparing the five second average engagements before and after the system reaction. This data was recorded from 14 subjects. The effect and its strength (S) varied between participants. The strength was calculated by;

$$S = \frac{\langle e_{after} - e_{before} \rangle}{\langle |e_{after} - e_{before}| \rangle} \quad (2)$$

For five participants the effect was positive and rather large ($S > .1$, $M_+ = .23$, $SD_+ = .09$). For four of them it was negative ($S < -.1$, $M_- = -0.14$, $SD_- = .04$) and for five the strength size was small ($|S| < .1$, $M_n = .05$, $SD_n = .03$).

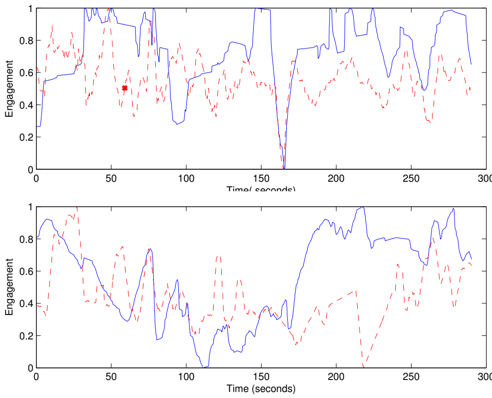

Figure 4: In upper image: reported engagement and prediction result for the automatic condition (# 1). In lower image: reported engagement and prediction result for the manual condition (# 15). The reported engagement is represented with a solid line and the prediction result with a dashed line.

4.1.4 Survey

Overall, NOISA was perceived interactive (79.0%), engaging (63.2%) and active (63.2%). The automatic behaviour of the system was mostly perceived as random (73.7%), mysterious (63.2%) and interesting (52.6%). These qualities were also reflected on the questions that asked to rate the different aspects of NOISA. The overall experience was rated enjoyable ($M = 12.3$, $SD = 4.1$), but the automatic behaviour was unpredictable ($M = 6.3$, $SD = 4.2$). Despite, or rather in conjunction with the unpredictable behaviour, NOISA encouraged the subjects to play and explore ideas ($M = 11.5$, $SD = 4.7$). There was a correlation ($R = 0.49$, $p = 0.03$) on how enjoyable the experience was and how much NOISA encouraged the participants to explore. The sounds were neutrally perceived ($M = 10.1$, $SD = 4.0$).

5. CONCLUSIONS

The results of our study enable us to identify statistically relevant components for the prediction of engagement, as shown by the correlation analysis of the predicted and user reported engagement variables. It is interesting to note that a simple model, such as multiple linear regression, can provide these results. We aim to develop the model further by utilising non-linearity and temporal prediction capabilities.

The analysis shows that the duration of hand movements is longer while they were engaged. In addition, sound producing hand movements become subtle as the participant is more conscious of the consequences of his/her actions and can concentrate on the more delicate qualities of sound. Consequently, majority of the participants who ranked their moments of engagements in our study considered more focused movements, paying more attention to their control actions with the interface. It is useful to consider the ways the subtle movements occur in musical interaction as a subsistent source for articulating musical engagement. Similarly, NASA TLX analysis shows us that the task with the automated responses required more mental focus and effort. The system's responses opened up the possibility to participants to coordinate their actions with stronger demands. Mental demand and effort is key in music performances.

The analysis of the subjectively rated engagement before and after the automatic responses shows that there are subjective preferences on the system's behaviour. Some participants reacted positively while some negatively. It is apt for us to improve the automatic behaviour, which is now perceived quite mysterious. The system should be able to evaluate the impact of its actions more carefully. All in all, the results indicate that NOISA encouraged to play and explore new ideas, providing interactive and engaging experience with enjoyable sound characteristics. Our engagement prediction model will be useful for researchers in the NIME community. Open source release of NOISA project is available at `https://github.com/SopiMlab/NOISA`.

6. ACKNOWLEDGMENTS

This work is supported by Aalto Starting Grant.

7. REFERENCES

[1] A. B. Bakker. Flow among music teachers and their students: The crossover of peak experiences. *Journal of vocational behavior*, 66(1):26–44, 2005.

[2] N. Bryan-Kinns and P. G. Healey. Decay in collaborative music making. In *Proc. of the Int. Conf. on New Interfaces for Musical Expression*, 2006.

[3] A. Camurri, B. Mazzarino, and G. Volpe. Expressive interfaces. *Cognition, Technology & Work*, 6(1):15–22, 2004.

[4] M. Csikszentmihalyi. *Creativity: Flow and the Psychology of Discovery and Invention*. Harper Perennial, New York, 1997.

[5] N. Dael, M. Mortillaro, and K. R. Scherer. The body action and posture coding system (bap): Development and reliability. *Journal of Nonverbal Behavior*, 36(2):97–121, 2012.

[6] M. Gurevich and J. Trevino. Expression and its discontents: Toward an ecology of musical creation. In *Proc. of the Int. Conf. on New Interfaces for Musical Expression*, 2007.

[7] S. G. Hart and L. E. Staveland. Development of nasa-tlx (task load index): Results of empirical and theoretical research. *Advances in Psychology*, 52:139–183, 1988.

[8] A. R. Jensenius. Action-sound:developing methods and tools to study music-related body movement, 2007.

[9] R. Laban and L. Ullmann. *The mastery of movement*. Dance Books, 4^{th} edition edition, 2011.

[10] G. Leslie. *Measuring musical engagement*. PhD thesis, 2013.

[11] L. Meyer. *Emotion and meaning in music*. Chicago: University of Chicago Press, 1956.

[12] J. Nesfield. Strategies for engagement in computer mediated musical performance. In *Proc. of the Int. Conf. on New Interfaces for Musical Expression*, 2012.

[13] D. G. Paine. Gesture and musical interaction: interactive engagement through dynamic morphology. In *Proc. of the Int. Conf. on New interfaces for Musical Expression*, pages 80–86, 2004.

[14] K. Tahiroğlu, T. Svedström, and V. Wikström. Noisa: A novel intelligent system facilitating smart interaction. In *CHI'15 Extended abstracts on Human Factors in Computing Systems*, 2015.

[15] M. Zentner, D. Grandjean, and K. R. Scherer. Emotions evoked by the sound of music: characterization, classification, and measurement. *Emotion*, 8(4):494, 2008.