Non-intrusive Counter-actions: Maintaining Progressively Engaging Interactions for Music Performance

Conference Paper · July 2016

DOI: 10.13140/RG.2.2.10011.46887

3 authors:

Koray Tahiroğlu
Aalto University

Juan Carlos Vasquez
University of Virginia

Johan Kildal
IK4-TEKNIKER

Non-intrusive Counter-actions: Maintaining Progressively Engaging Interactions for Music Performance

Koray Tahiroğlu
Department of Media
Aalto University
School of ARTS
FI-00076 AALTO Finland
koray.tahiroglu@aalto.fi

Juan Carlos Vasquez
Department of Media
Aalto University
School of ARTS
FI-00076 AALTO Finland
juan.vasquezgomez@aalto.fi

Johan Kildal
IK4-TEKNIKER
Iñaki Goenaga 5
Eibar 20600, Spain
johan.kildal@tekniker.es

ABSTRACT

In this paper we present the new development of a semi-autonomous response module for the NOISA system. NOISA is an interactive music system that predicts performer's engagement levels, learns from the performer, decides what to do and does it at the right moment. As an improvement for the above, we implemented real-time adaptive features that respond to a detailed monitoring of the performer's engagement and to overall sonic space, while evaluating the impact of its actions. Through these new features, the response module produces meaningful and non-intrusive counter actions, attempting to deepen and maintain the performer's engagement in musical interaction. In a formative study we compared our designed response module against a random control system of events, in which the former performed consistently better than the latter.

Author Keywords

Engaging interaction; semi-autonomous system; musical interaction; interactive system; NIME; NOISA

ACM Classification

H.5.2 User Interfaces Interaction styles; H.5.5 Sound and Music Computing: Systems

1. INTRODUCTION

In real-world interactions, with the exception of those activities that are largely an accidental accumulation of experience and arbitrary re-actions, people do perform skilled tasks in the way musicians have always done; mastering the control of their actions, applying a systematic approach to their activities, not drifting in lack of interest but recognising self-integration as a result of their actions. In our current line of research, we want to assist these people in the music domain with an interactive system that produces meaningful counter-actions, attempting to deepen and maintain the performer's engagement with a new interface for musical expression (Figure 1).

In the work we present here, we focus on a new development of our interactive music system; the *Network of Intelligent Sonic Agents*, NOISA, that predicts the performer's engagement levels, learns from the performer, decides what to do and does it at the right moment [14]. Addressing the limitations of previous implementation of NOISA [13], we designed and implemented real-time adaptive features that respond to a more detailed monitoring of the performer's engagement at each time, while evaluating the impact of its actions. Our main contribution in this paper is the automatic response-behaviour of the system and the formative user-test study. The major development has been on the sound design, gesture comparison and the audio analysis features of the response module. In a user test study, we focused on these main features to understand how expert musicians would perceive and experience the designed response module in comparison to random responses.

Figure 1: Expert musician and NOISA Instruments

2. RELATED WORK

In regard of developing a supporting response module and generating real-time collaboration between interactive systems and performers, references include Sarkar's [12] efforts to build a synthetic performer that acts simultaneously with live musicians and whose actions cannot be distinguished from human musical expression. Sarkar describes a three-step method to achieve such results in a designed system comprising the following order; an analysis of the overall soundscape, organisation of responses categorised by sonic characteristics, and further improvement of the system's musicianship by the usage of potential adaptive and learnt response mechanisms. Similarly, further advantages of real time spectral analysis for selecting more suitable responses to a determined input have also been explored previously in a live musical context [16].

On the same vein, Ramalho et al. introduce an intelligent jazz player agent [11], in which they integrate three main modules that reflect the division of the agent's tasks; the listener, reasoner and executor. In the chain of execution, the system feeds the performer's resulting response back into the agent's input in order to influence on choices regard-

NIME'16, July 11-15, 2016, Griffith University, Brisbane, Australia.

ing what to play next. One important characteristic of the system is the re-utilisation of musical fragments. The musical fragments are recognised as stylistic alternatives and the system retrieves them back by following predetermined musical rules. These musical rules can be seen in depth in previous research on interaction within musical improvisation context; such as Hudak's and Berger's model [10], which is based on describing hierarchies and interaction between performers and feedback.

On the other hand, considering the multiple definitions and applications of the concept of "agent", the research of multi-agent networks in music shows different branches of development; from attempts towards gaining independence from the composer [6, 8], to collaborative approaches for autonomous, non-deterministic music composition [5], including directed improvisation [9], interactive installations [4] and interactive music systems [15]. To articulate the supportive and non-intrusive actions, it is equally important to consider the notion of replacing adaptive buttons or default values in an interactive system in favour of a better-performing algorithm that can accurately predict the user's next action [7]. We incorporated these ideas into our work to implement a better performing response module that is capable to provide meaningful and non-intrusive responses to the performer's actions in a real time music performance.

3. NOISA RESPONSE MODULE

NOISA, *the Network of Intelligent Sonic Agents* is a distributed musical system designed to provide the performer with autonomous supporting counter-actions. As these responses are correlated to the estimated engagement and previous / current actions of the performer, the main premise of NOISA is, then, attempting to maintain a high engagement level by proposing musically interesting events when the performer is losing interest or motivation.

Figure 2: NOISA Instrument and the system set-up, example of one box containing an agent

Our presented approach functions in two different stages; *engagement prediction* and *response module*. The engagement prediction is based on our previously developed "subjective engagement sampling method" (SESM) [13, 14], which estimates a person's engagement in real-time by monitoring movements, facial expressions and control actions. The response module uses the predicted engagement as an input. It produces aesthetically desired and complementary musical responses when required by the main premise of NOISA, making the responses specially tailored for that particular musical interaction. Overall, NOISA consists of three instruments, central computer, Microsoft Kinect 2 and a Myo armband (see Figure 2). In our current research, the main development has been the re-construction of the response module, being as well the main contribution for this paper.

3.1 Sound Design

In the NOISA project, music space has emerged as a critical factor for encouraging high levels of engagement during sustained amounts of time. Therefore, we have redesigned sonic characteristics of the instruments, using sample-based, frequency-tuned granular synthesis and phase vocoder algorithms. As the physical design of each instrument requires two hands to operate, sound producing actions formed a multilayer sound interaction in capacity of guaranteeing the production of complex sonic textures from a simple gesture input. Regarding the granular synthesis, the left handle modifies the playback speed of every individual grain, with a proportional wet/dry control of an spectral reverberation effect. In the phase vocoder algorithm, the same slider rewinds and triggers the speed parameter, having as a final result a different proportion of time compression/stretching. In both algorithms, the left handler duplicates the modified signal and manipulates the tape head rotation frequency emulator of a pitch shifter device. Satisfactory and encouraging sound results, even in standalone mode, were considered in the design process as prerequisite for engagement.

The physicality of the instrument also played a significant role in the sound design process. We considered the physical properties of the instrument for the potential nature of the sound production events. The result of such discussion shaped the final envelope behaviour; as the handlers approach the box, sharp and shorter attacks appear in contrast to achieve maximum opposite position, which generates an evolving and rich sustained texture for an indefinite period of time. In addition, we implemented a feature for turning off each instrument by setting both handlers in minimum position i.e. silence, so any automatic responses are avoided for the specific instrument in which the "off" gesture is performed.

The control events of the two handlers were inspired in the roles present when performing a string instrument; one hand is in charge of sound production and the other one is responsible for manipulating the produced sounds. In NOISA, the left-handler activates the digital audio processes and the right-handler replicates the previous resulting signal. It features as an extra pitch shifter with tone transportation that is equivalent to the position of the right handler. The intervals that are generated from the relationship of the original signal and the shifted copy, contribute to the richness of the instrument's overall sound production. The inclusion of the Myo armband is also a powerful new resource for musical expression. We implemented a dynamic control with three levels of volume through the EMG data that the system receives from the performer's arm.

Lastly, as the system is fully built upon sample-based audio synthesis processes, it is worth to mention the source material that we used for each of the sound producing algorithms. After many tests, we achieved the most satisfactory aesthetic results by transforming fragments from pieces of the classical music repertoire; Modest Mussourgky's Pictures of an Exhibition in its transcription for Solo Guitar, Johann Sebastian Bach's Partita in A minor BWV 1013 for Solo Flute and Eugene Ysaye's Sonata No. 3 Op. 27 for Solo Violin were the final choices for each of the instruments.

3.2 Non-intrusive Counter-actions

Each one of the NOISA instrument features a *sonic agent* inside it. In NOISA, we refer sonic agent as an entity responsible for learning and reproducing relevant actions in the right timing, with the ability to monitor its surrounding musical space, user actions and performer engagement; altogether forming an intelligent network [13].

The agents are built to act towards supporting a music performance by providing responses without intruding into the overall soundscape. The agents' responses were redesigned to encourage and maintain the communication and engagement of the performer with the system. In previous research we discovered the risk of breaking the flow of a performance while including autonomous responses. "When" and "how" to react to the actions of the performer appeared as crucial questions in our research. In our current contribution, we addressed "how" by aiming towards obtaining an even texture in all frequencies of the audible spectrum through real-time analysis of the overall sonic space and setting a principle of complementarity rules for the responses. The system learns from the performer's actions by recording a series of *gestures* and labels them with the analysed sound properties, which allows NOISA to pick the most appropriate gesture for a particular moment.

"When to react" was addressed by using the predicted engagement level to set up the probability of the automatic responses. In our previous research [13] we also found out that when a person is not engaged, the actions during the performance would be relatively meaningless and unfocused. On the contrary, in the moment when a performer is achieving a high level of engagement, the performer produces highly precised, controlled and focused actions, which result in no need for further encouragement coming from the interactive system. Bearing that in mind, we designed an active deterministic behaviour; when the player is not engaged, the system stimulates further attentive creative activity. In contrast, NOISA produces only occasional and merely supporting responses once the performer is deeply engaged.

3.2.1 Gesture Comparison

As the possibilities for control actions with two handlers are relatively limited, we elaborated a portfolio of aesthetically effective two second-long movements. We translated these actions into rules in order to define what a gesture is within our system (see Figure 3). Once the performer has trespassed the high third of engagement, the performer's gestures are recorded in compliance with these rules. Afterwards, the gestures are labelled with handle data, current predicted engagement and audio properties of the event. As they are stored in a real-time database, the foreseen problem of segmenting the gestures from a stream of continuous data appeared. We tackled this problem by designing an interaction logic where the right handle would perform a dominant function - i.e. triggering a gesture recording event along with a first excitation signal. After this initial event is completed, the system applies a comparison with the rules extracted from the chart. In NOISA, a gesture recording is finalised in two different scenarios; once the sound fades and once a new gesture begins with a new signal excitation, meaning that gestures can have variable length.

3.2.2 Audio Analysis

Each gesture is labelled with three timbral features that are the results of a real time sound analysis. The monitored sound features are *spectral centroid*, *spectral smoothness* and *energy*. Spectral centroid describes where the centre of the spectral mass is in frequency domain. It is an ideal indicator for an appropriate potential response in terms of spectral complementarity i.e. contrasting frequencies generate balance in the overall spectrum.

Spectral smoothness gives an overall idea of the noise content that is present in the spectrum. It provides the system an ability of responding with inverse timbral characteristics and generating an aesthetically desired contrast. Finally, energy is measured in decibels. It is used to keep the gain levels of the system's potential responses always below the decibel measure of the performer's performance, guaranteeing non-intrusiveness. As these descriptors are time-variant, the features are calculated as an average over the total length of a gesture. For obtaining the analysis indicators, we used a custom modular patch in PureData composed of objects from the *flib* library [3].

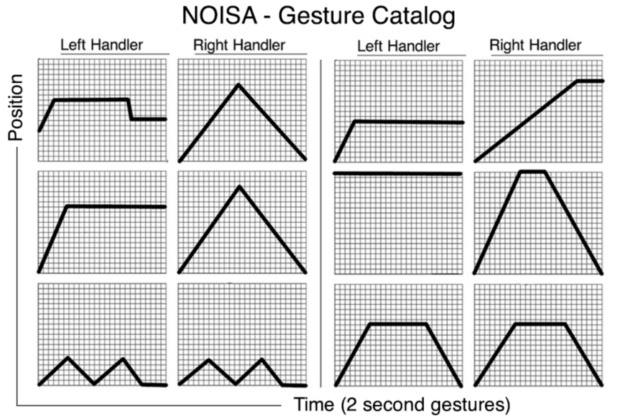

Figure 3: Gesture catalogue showcasing the possibilities of effective musical interaction. Each gesture requires both hands starting simultaneously. Every graph opposes position (Y) and time (X).

4. USER STUDY

Since our main contribution is the response module, we evaluated its effect on a performer's engagement by comparing it against a system producing random responses in a user test study. The main expectation was that our response module, coupled with the engagement prediction, should produce meaningful and interesting events that would improve the engagement and establish a musical dialogue. Furthermore, we were also interested in how NOISA is perceived in overall with AttrakDiff [1], a widely used validated test designed to evaluate interactive systems. AttrakDiff consists of 28 Likert-scale questions, measuring diverse dimensions of a system.

We invited eight expert musicians aged 25-50 to take part in the study. The sessions had an approximate length of 45 minutes, taking place in a room with acoustic treatment and isolation from exterior noises. The NOISA instruments were placed in a row on top of a table. The study consisted of an exploration phase, a task comprising the performance of a musical piece with the designed system (NOISA) and again with the random system (RANDOM), questionnaires and an interview. The musicians were able to freely play with the three NOISA instruments during the familiarisation phase. The questionnaires were completed in after each performance task and the interview was held at the end of the study. The order of the two systems were randomised and the musicians were not informed about the order nor they were informed on how the system operates. The random responses were generated with a random interval of 3-7 seconds and the responses were linear sequences of numbers that resembled the automatically recorded gestures. During the interviews, the conversation was steered in such a way to get more insight in musicians' responses to AttrakDiff questionnaire and in their more detailed descriptions and comparisons of system responses in both conditions.

The musical piece that the musicians were asked to perform had a broad structure with an approximate requested length of three minutes. The musicians had to start quietly with a specific NOISA instrument that was positioned on the edge and were asked to play increasingly loud with the middle instrument until they reached the musical climax on the third instrument. After a climatic section, they would move back to the first instrument, play softer, sporadically and then finish the piece. All the performances were videotaped, and the interviews were audio recorded.

4.1 Results

4.1.1 *AttrakDiff*

The musicians indicated their perception of each system by pairing opposite adjectives in AttrakDiff questionnaires. The results were obtained by evaluating the following indicators: *Pragmatic Quality (PQ)*, it describes rate of success to achieve their goals with the system; *Hedonic Quality - Stimulation (HQ-S)*, it refers to aesthetic pleasure; *Hedonic Quality - Identity (HQ-I)*, it indicates level of identification with the system; and *Attractiveness (ATT)*, a global value of the system based on quality perception.

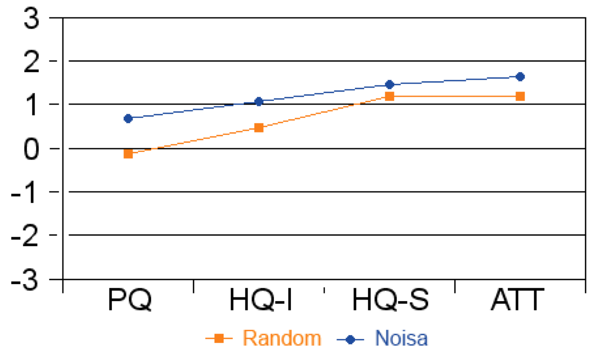

Figure 4: Diagram of average values per category

We analysed both the portfolio of results for NOISA and RANDOM conditions, as well as for the diagram of average values (see Figure 4). Both the pragmatic and hedonic quality results are higher in NOISA than in RANDOM. The confidence interval for pragmatic quality of NOISA is smaller than for RANDOM. In the same way, the confidence interval for hedonic quality of RANDOM is smaller than for NOISA. In both cases, there is a probability of incidental fluctuation greater than 5 percent in the the identity and stimulation aspects of hedonic quality, meaning that are considered not statistically significant (see Figure 4).

In terms of mean values of the word pairs used for evaluation (see Figure 5), there are several extreme values to be noted. In particular, the word pairs “technical - human”, “confusing - clearly structured”, “unruly - manageable”, “isolating - connective”, “alienating - integrating” and “discouraging - motivating” showcase a difference of 1 point or more in a scale from -3 to 3.

4.1.2 *Engagement levels*

According to the monitored engagement values, the normalised mean engagement was improved for the NOISA system in all occasions, except for musician 1 and 7, with negative difference of -0.0024 and -0.0744 respectively. On the other spectrum, musicians 3 and 5 featured a prominent improvement in engagement when using NOISA system.

In percentile values, NOISA improved the engagement 10.79 percent in average for all musicians in comparison to RANDOM. The two more pronounced cases were musician

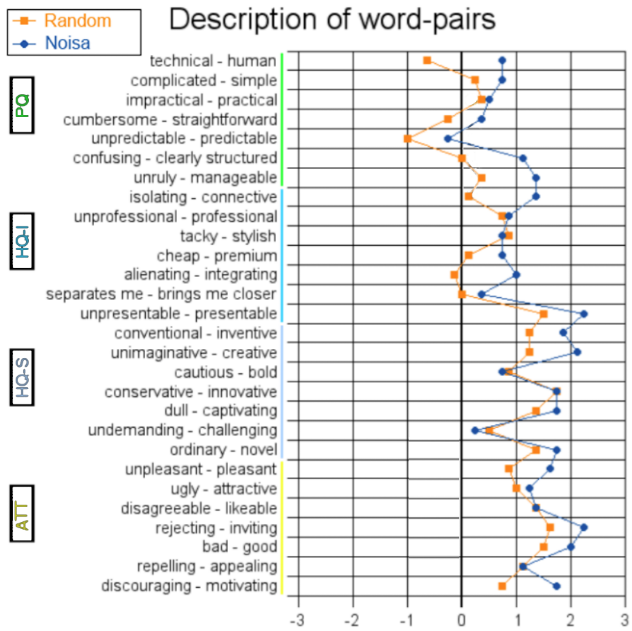

Figure 5: Mean values of the Attrakdiff word pairs for RANDOM and NOISA (designed)

3 (see Figure 7), with an average improvement of 46.64 percent, and musician 5 with an overall improvement of 31.4 percent. In the specific case of musician 3, the engagement average for the designed system was 0.6891, among the values relative to “High Engagement”, while the average for the RANDOM was 0.2227, within the range of the “Low Engagement” category. The engagement over time graph for musician 7 (see Figure 6) demonstrates an example of a case where the designed system response impacted negatively towards the overall measured engagement values. In the latter case, the designed NOISA achieved an engagement average of 0.2738, while RANDOM had an average of 0.3482 for the whole performance. Both values are ranked among the “Low Engagement” category.

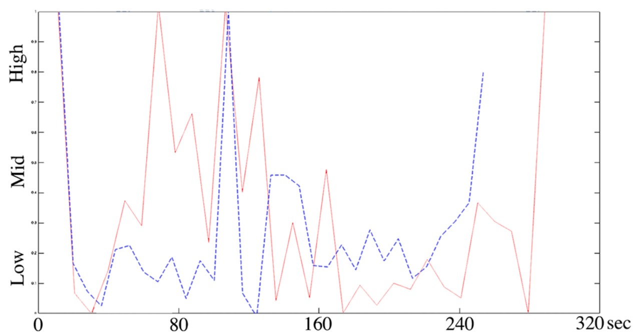

Figure 6: Engagement evolution over time for musician 7 - NOISA (Dotted dark) and RANDOM (Straight bright)

4.1.3 *Analysis of interviews*

The nature of our user study, with AttrakDiff multiple rate comparison and open-ended questions in interviews, guided us to apply thematic analytic method for analysis. The thematic analysis method focuses on identifying patterns and themes within data and has been used as a qualitative analysis method in the field of psychology [2]. With this method,

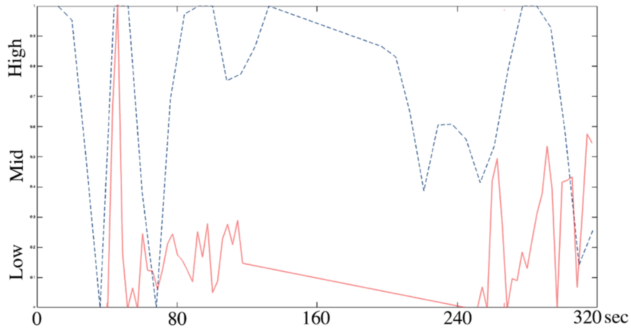

Figure 7: Engagement evolution over time for musician 3 - NOISA (Dotted dark) and RANDOM (Straight bright) comparison

we could be able to identify a certain thematic framework, informed by qualitative responses in the interviews.

After the transcription of the interviews and familiarising ourselves with the qualitative data, we have identified recurrence of certain issues in the interviews. *More intelligent, interact more with what you are doing, more connected, learnt faster and better, unexpected actions, guided by the system, interrupting, playing with "someone", unpredictable, more human*, etc. are the issues emerged from the participants. Looking at how these issues are related to the practice of our system in both conditions of the study, led us to identify higher-level themes: Human-like realness; confused, intelligent and learnt behaviour; and complexity in ambition & simplicity in interaction.

4.1.4 Emerging themes

We can view the individual responses to the interview questions through these higher-level themes;

Human-like Realness reflects the range of responses musicians gave to questions of their overall impression of the system and brief comparison of both tasks. Human-like Realness was often articulated through the responses about intrusive differences based on experience:

"(NOISA) seems more connected, and (RANDOM) more inconsistent. The latter seems that you are not playing with 'someone', but with 'something' instead." (#6)

"Also there was a lot more input from the system in the second one (RANDOM), so the second one was much more strongly guided by the system rather than me guiding it. The shape I had in mind was helped by the first one (NOISA), and contradicted by the second one." (#8)

The second main theme, *Confused, Intelligent and Learnt Behaviour*, was not a specific question in the interview but emerged as a relevant theme in responses musicians gave to a series of questions concerning whether or not the system has provided intruding and unexpected responses.

"It was interrupting what I was doing. (...) Number 2 (NOISA) the instruments learnt faster and better from me than in Number 1 (RANDOM)" (#6)

"It is my first impression, as I would feel better if I could spend more time. I was more confused with the second one (RANDOM)" (#5)

"Yes, although the idea of a task made it a bit more confusing for me. For the way the system looks and operates invites you to freely improvise, so concentrating in changing from one to another in the task was a bit of a distraction for me. However, the first one (NOISA) responded much better." (#7)

There were also a number of responses that link the way the response module reacted under both conditions to the inconsistency or consistency, intelligent and informed behaviour or something that has short-lived interest by virtue only of novelty.

"First one (RANDOM) made me feel like 'trying to', but the second one (NOISA) was 'knowing' what it was doing." (#6)

"I think the second one (NOISA) gave me more ideas. The first one (RANDOM) I don't remember interacting as much, I did what I planned to do, and the second one didn't allow me to proceed, but I worked around this input." (#8)

The last theme, *Complexity in Ambition & Simplicity in Interaction*, is not related to the task description itself but emerged as a theme regarding to the system behaviour under both conditions. Complexity & simplicity is hinted with musicians' responses varying from what they would be eager to do with the system's responses to their own actions:

"It is very easy coming with a sound with this instrument. You can learn how to control the sound morphology (...) Anyway, you can control the colour of the sound, but that seems to be easier in number 2 (NOISA). I don't know why, but they seem to interact more with what you are doing, when you leave them alone they are more vivid." (#7)

"It is kind of a mixture of complexity and simplicity. Complexity in the ambition, definitely sort of signifies that it can be much more complex, or a draft of more complex, but then this particular thing (instruments) is quite simple. You have the prediction of something more complex by the fact that the system tracks your system, but the operation itself is not complex at all." (#6)

"In terms of processing I think is nice and simple, you understand how it works, it gives enough range to be interesting without complicating things too much." (#8)

5. DISCUSSION AND CONCLUSIONS

Results show that the designed system performed consistently better than the random response test module in regard of the pragmatic quality, hedonic quality - stimulation, hedonic quality - identity and attractiveness features. Even though we obtained a probability of incidental fluctuation greater than 5 percent for the portfolio of results, the description of word-pairs show a significant difference in key parameters for our research. Particularly the word pairs that are used for the evaluation of how automatic responses are perceived in both systems, indicate that the designed network of agents pretend to help the performer non-intrusively in a way that resembles another human performer. Identifying NOISA as more "human", "manageable", "integrating" and "connective" in opposition with the random counterpart, states that our research has achieved effective results in terms of the adaptive learning cycle regarding the level of predicted engagement. This is also evident in the analysis of the qualitative data that we collected during our user studies. The main thematic themes demonstrate how the musicians perceived the differences between RANDOM and NOISA responses through their interactions with the system. Similar to our AttrakDiff findings, the perceived interaction leaned towards associating NOISA actions as human-like responses. We attribute this to the fact that NOISA established a communication with musicians more in line with the characteristics of their actions in a situated context, which also prompted them to be more engaged in their own actions.

In regard to the engagement levels, obtaining an average improvement of 10.79 percent, or 0.1079 in our scale from 0 to 1 is considered a success. Even though the expert musicians came from different musical backgrounds, we managed to maintain and improve their levels of engagement. The NOISA system presents a unified and particular musical identity prone to different perceptions depending on the musician's personal taste. The examples of sharp high engagement improvement in opposition to no cases of significant decrease make us to believe that a greater differentiation could be obtained if the number of musicians evaluated were higher.

On the other hand, eventual short sudden peaks of engagement changes were registered in the engagement-over-time comparison graphs. This phenomenon normally could be attributed to momentary lost of user tracking, suggesting that a calibration system may be developed further to cope with possible noise in the data. From the examples where our NOISA system impacted negatively in the overall engagement in comparison with the random response module, it is evident that the irruption of sudden high / low states of engagement complies with a non-realistic user behaviour according to our previous studies. In contrast, the data provided for musician 3 in Figure 7 shows predominantly gradual changes and oscillating ranges in consistent categories of engagement. These results indicate that the prediction module tracked steadily the musician's movements and consequent engagement level. Therefore, further work is necessary to ensure highly precise engagement prediction without interruption.

Further on, after reviewing the musicians' comments, it can be argued that before any attempt to design an interactive system with capabilities to organise given actions upon an increasing engaging basis, it is required an understanding of the basic facts concerning the context in which the interaction is situated. These facts appeared in our higher-level themes; *human-like realness - confused, intelligent & learnt behaviour - complexity in ambition & simplicity in interaction* come together as related main factors linked to interaction, becoming points of discussion within the results of our high level thematic analysis.

While these aspects have been explored in other interactive systems, NOISA provides a basis for an input-output responsiveness model approach within a composite engaging activity. It is certainly not intended here to correlate all the engaging interaction methods into one system. What we hoped to accomplish in the NOISA project was the description of an approach based on engagement prediction as well as a response module for designing an interactive music system. Our approach and the system were found to encourage the exploration of new ideas, while providing an interactive and engaging experience:

"There was a huge difference between both systems. In both occasion it seems to interact with what you're doing, although it seems more intelligent or connected in number 2 (NOISA). I like how sound is treated, really long sustained sounds, but you can change that in the way you operate the boxes." (#7)

Through these new features in response module, NOISA has significantly added to the interaction aspects of its domain. Motivated by the results obtained in the present study regarding engaging interaction for music performance, we acknowledge the potential of further applications of the NOISA system to investigate similar engagement-monitoring techniques in a broader scope of skilled, potentially-mentally-absorbing actions, such as handwork and crafting, or more mundane chores and homework that entail a considerable array of skilled and mentally-demanding actions.

6. ACKNOWLEDGMENTS

This work is supported by the Nokia Research Center. We acknowledge the work of Niklas Pöllönen in the project.

7. REFERENCES

[1] Attrakdiff. `http://attrakdiff.de`. Accessed: 2016-04-13.

[2] V. Braun and V. Clarke. Using thematic analysis in psychology. *Qualitative research in psychology*, 3(2):77–101, 2006.

[3] J. Bullock. *Implementing audio feature extraction in live electronic music*. PhD thesis, PhD thesis, University of Birmingham, 2008.

[4] N. d'Alessandro, R. Calderon, and S. Müller. Room # 81 – agent-based instrument for experiencing architectural and vocal cues. In *Proc. of the Int. Conf. on New interfaces for Musical Expression*, pages 132–135, Oslo, Norway, 2011.

[5] M. Delgado, W. Fajardo, and M. Molina-Solana. Inmamusys: Intelligent multiagent music system. *Expert Systems with Applications*, 36(3):4574–4580, 2009.

[6] A. Eigenfeldt and A. Kapur. An agent-based system for robotic musical performance. In *Proc. of the Int. Conf. on New interfaces for Musical Expression*, pages 144–149, Genoa, Italy, 2008.

[7] K. Z. Gajos, K. Everitt, D. S. Tan, M. Czerwinski, and D. S. Weld. Predictability and accuracy in adaptive user interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1271–1274. ACM, 2008.

[8] M. Goto. An audio-based real-time beat tracking system for music with or without drum-sounds. *Journal of New Music Research*, 30(2):159–171, 2001.

[9] B. Hayes-Roth, L. Brownston, and R. van Gent. Multiagent collaboration in directed improvisation. In *ICMAS*, pages 148–154, 1995.

[10] P. Hudak and J. Berger. A model of performance, interaction, and improvisation. In *Proceedings of Int. Computer Music Conf.*, pages 1–8, 1995.

[11] G. L. Ramalho, P.-Y. Rolland, and J.-G. Ganascia. An artificially intelligent jazz performer. *Journal of New Music Research*, 28(2):105–129, 1999.

[12] M. Sarkar and B. Vercoe. Recognition and prediction in a network music performance system for indian percussion. In *Proc. of the Int. Conf. on New interfaces for Musical Expression*, NYU, USA, 2007.

[13] K. Tahiroğlu, T. Svedström, and V. Wikström. Musical engagement that is predicated on intentional activity of the performer with noisa instruments. In *Proc. of the Int. Conf. on New interfaces for Musical Expression*, pages 132–135, Baton Rouge, USA., 2015.

[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*, Seoul, 2015.

[15] R. D. Wulfhorst, L. Nakayama, and R. M. Vicari. A multiagent approach for musical interactive systems. In *Proc. of the second international joint conference on Autonomous agents and multiagent systems*, pages 584–591. ACM, 2003.

[16] M. Young and S. Lexer. Fft analysis as a creative tool in live performance. In *Proc. of the 6th International Conference on Digital Audio Effects*, 2003.