Shut Up and Play: A Musical Approach to Engagement and Social Presence in Human Robot Interaction

Louis McCallum and Peter W McOwan

Abstract— With a view to studying the development of social relationships between humans and robots, it is our contention that music can help provide extended engagement and open ended interaction. In this paper we explore the effectiveness of music as a mode of engagement using *Mortimer*, a robot able to play a drum kit and employing a composition algorithm to respond to a human pianist. We used this system to conduct a comparative study into the effects of presenting the robot as a social actor or as an instrument. Using automated behavioural metrics, including face tracking, we found that participants in the social actor condition played for longer uninterrupted and stopped the robot mid-performance less. They also spent more time looking at the robot when not playing and less time looking at the piano when playing. We suggest these results indicate greater fluency of playing and engagement and more feelings of social presence towards the robot when presented as a social actor

I. INTRODUCTION

Much work in the field of Social Robotics focusses on either programming social awareness into robots completing tasks alongside humans or designing artificial assistants that leverage social skills to better achieve their goals. Our overall research has the distinct aim of investigating the necessary conditions for sustainable and meaningful human robot relationships. A relationship is something that is necessarily built up over time [1] and a robot's role will determine the frequency, regularity and period of interaction, however, Human Robot Interaction (HRI) trials are rarely extended beyond a single session.

The machines used in longitudinal research that has been conducted include robotic household appliances [2], robotic pets [3] and fully mobile anthropomorphic robots [4], [5]. The interactions afforded by the first two are simply not engaging enough and result in a swiftly declining novelty effect. The latter attempts to use either gesture or language as the basis for interactions and in all cases present, to greater or lesser extent, the frustration of participants at their small range of abilities and often limited scripts. This being said, any time music is played as part of an ensemble, you are guaranteed to have at least two people, in the majority of cases co-located, simultaneously focussing their attention towards the same task and cooperating towards a joint goal.

However, although music does provides these opportunities for affective engagement, shared creativity, and social

activity, it is not our view that a robot best equipped to build sustainable and meaningful relationships with humans will be one that can solely play music. We expect a key factor to be the presentation of a believable system which has some amount of social presence. Believability is prevalent in Sociable Robotics research [6], [7] and is described as the amount to which a person can suspend their knowledge a robot is inanimate and not actually in possession of the human faculties we attempt to make it display. We draw similarities between this notion and the concept of social presence, defined as the feeling of being with another when there is no human physically present [8]. We suspect the addition of simulated social behaviours will provide a sense of believability or social presence, which, along with the engaging musical interaction, will be conducive to forming relationships. As such, we developed *Mortimer*, an interactive robot drummer who can improvise responsively with a human pianist and investigate its presentation as a social actor or instrument through the introduction of a basic artificial personality. We hypothesise that it will have increased engagement and social presence in comparison to the same system exhibiting musical functionality alone.

Section II covers related research, Section III describes our technical development, Section IV details the study we conducted and its findings. These are discussed in Section V. In Section VI we summarise and outline future research directions.

II. RELATED WORK

A study of an assistive eldercare robot by Heerink et al. found that a more socially communicative system elicited more comfortable and expressive interactions [9]. Conversely, Giuliani et al. found little difference between subjective measures of the perception of a robotic bar tender who was either task or socially driven, although they do cite the latter as more efficient [10]. Focussing mainly on assistive technology, Bickmore finds that socially aware systems are more engaging over time [11], demonstrating that the inclusion of simulated social behaviours can be beneficial in such situations. Within this, he finds that dialog which varies over time increases engagement and disproves suggestions that a fictional backstory relayed in the first person may be seen as dishonest [12]. Personality is also part of a 5 category model for human robot companionship defined by Benyon et al. [13].

Research has shown a range of different variables may effect the perceived social presence of virtual or physically embodied agents. For example, in a study with synthetic voices

in a book reviewing scenario, Lee and Nass demonstrate greater social presence when the personality of the voice is matched to the content of the text and to the personality of the user [14]. They also find that an extrovert voice has more social presence than an introverted one. Lee et al. show that participant's feelings of social presence significantly effect whether they are perceive a robotic pet with a complimentary personality more favourably [15]. Further, Nowak and Biocca report higher social presence from an agent represented by less anthropomorphic avatar in comparison to an agent represented by a more anthropomorphic avatar and one with no image [16].

The work of Reeves and Nass shows that humans may treat media as social actors and provides details as to how their personalities can elicit different responses [17]. They find that people will prefer a flattering computer, even if the praise is unwarranted and prefer a computer that praises others. Further, Lee et al. demonstrated verbal behaviour could be used to accurately portray extrovert or introvert personalities and that there was a complimentary personality attraction effect between robot and participants [15]. With regards to memory, Kasap and Magenant-Thalman report that reference to past interactions prevented the usual decline in engagement over time [18].

Bickmore et al. define engagement as the degree of involvement the users choose to have with the system and so use the frequency and length of interactions as a sign of engagement [12]. Sidner et al. define it as 'the process by which two (or more) participants establish, maintain and end their perceived connection' [1], a definition which Castellano et al. also deem suitable for describing social bonding between human and child [20].

Castellano et al. go on to suggest engagement with a robot may entail an affective and attentional component based upon positive feelings and a willingness to maintain interaction [21]. They report a system capable of classifying engagement with 94.79% accuracy when using both contextual information and nonverbal cues such as the user smiling and the user looking at the robot. Similarly, Sanghvi et al. describe a system capable of automatically detecting engagement from laterally captured posture data with an accuracy of 82.2% [22]. Michalowksi et al. also include looking at the robot into their model of spatial engagement [23].

Within Interactive Music Systems (IMS) for duet improvisation, the matching of piano to percussion is uncommon, with most systems taking the homogenous approach of matching percussion with percussion or keyboard to keyboard. For instance, *The Continuator* [24], *Mimi* [25] and *Shimon* [26] construct a model of the users piano input and use this to generate stylistically appropriate tonal response. Motivated to creative ends, the majority of research favours openness of outcome over stability which, as we will elaborate upon in Section III, does not necessarily serve us well. An example of a more stable system is the UK Garage generator for composing subtly varying beats described by

Fig. 1. The Robot

Collins [27].

The use of robots to perform music or as the physical embodiments of IMSs is also well documented and again the focus is often on the creative output or experience. *Haile* [28], *Shimon* [26], the many robots of LEMUR [29], including their *Orchestrion* ensemble built for popular musician Pat Metheny, and Felix Thorne's *Felix's Machines* [30] are amongst the notable.

III. MORTIMER

As stated in Section II, the development of musical robots as physically embodied accompanists, performers, improvisers or art installations is not unique. However, the primary focus of our research is the study of sustainable and meaningful relationships between human and robot based on regular, open ended, musical interaction and we have endeavoured to build a robot to best match these aims. We have identified two key characteristics that have been balanced in development to maintain the long term engagement and elicit the feelings of social presence we feel are necessary.

The system must be stable enough to be run unassisted by a novice user, as any malfunction or erratic behaviour will diverge from the positive experience needed for engagement and break the illusion of believability that creates feelings of social presence. To avoid this confusion, tools from proficient but unsolved research areas such as realtime beat tracking and automatic speech recognition have been avoided.

It is also critical that the robot is responsive to human playing. If the relation between the input and output is explicit enough to be recognised by the player then this will give the impression the robot is listening to them. If the response seems intelligent and adds to the playing experience positively, this will at worst demonstrate the system's competent engineering and provide an enjoyable experience to the listener. This is likely to increase long term engagement. It is our expectation that this will give the impression that the robot has some understanding of the music being played and its emotional content, adding to the feeling of social presence and laying the foundations for a more meaningful relationship.

A. The Robot

Demonstrated in Figure 1, we have built a robot capable of playing a standard drum kit set-up of kick, snare and hi-hat.
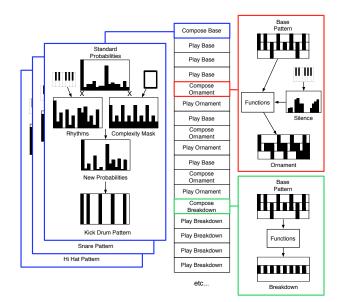
[1][19]

Fig. 2.   Composition of a Single Chorus

Fig. 3.   Social Interaction

It is our intention to present a robot that is clearly mechanoid whilst still triggering a subtle anthropomorphic response for participants. As such, it is built around an aluminium frame and with two identical beater arms and a back projected head that displays simple, cartoonish face. As well as the two striking arms, there is an automated kick drum that sits in front of the robot as it would in a conventional drum set up.

The robot may also respond to and alter the display of a tablet interface. This is used to facilitate simple nonmusical interactions and to control performance parameters.

B. Composition

The algorithm uses the MIDI transcript of previous playing and explicit performance parameters to inform decisions based on inbuilt knowledge. The current performance parameters are complexity, tempo and length. Every session is made up of a number of short tracks, each track consisting of a number of 12 bar choruses, depending on the length parameter. For each chorus, we first compose a form that decides whether each bar will be a base pattern, an ornamentation of this or a breakdown section. The complexity parameter decides the frequency of ornamentation. The music itself is not composed until it is required to be played, thus allowing the algorithm to respond to the pianist using the most recent information.

The base pattern is composed at the beginning of each chorus, taking into account a histogram of previous user rhythmic input and the complexity parameter to taylor its inbuilt knowledge to the current user. Lower complexity favours on-beat notes. If a bar demands an ornament, the base pattern is augmented by a series of ornamentation functions. We aim to fill gaps left by the pianist and so use a histogram of previous silence to inform this decision. Similarly, a breakdown bar is the base pattern augmented by one of a selection of breakdown functions. Figure 2 demonstrates this
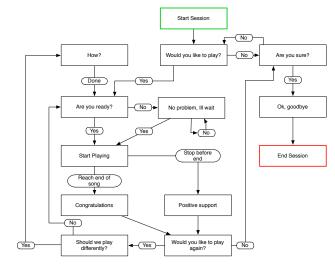
process for a single chorus.

Although the overall tempo is set explicitly by the pianist using the tablet, each of the robot's instruments are run on an individual routine. This allows for each one to implement independent and differing timing deviations from the quantised grid. When transcribing the piano input, we quantise each note to a 16 note grid in relation to the robot's scheduler. However, we also record the actual timing of the notes and the distance from the beat that it was quantised to. We use this to get a feel for the groove of the human input and match this to the groove of the robot.

In addition to composing bars based on a form generated at the beginning, we take note density and mean note velocity as a measure of power for the piano playing. If this dips below a certain threshold then instruments are either removed or thinned by the robot in order to match a perceived sparseness of texture or reduction of dynamics.

C. Artificial Personality

As the generation and processing of natural language are still vast and ongoing research areas in their own right, we decided to take the minimal approach. By restricting human input to multiple choice answers, we have removed the need to process natural language. Figure 3 outlines the script developed to frame the musical session as a social interaction, taking input from a tablet interface. As there is a clear possibility of repeating the script within a session, we used techniques from Natural Language Generation (NLG) to introduce variation into our dialogue. In a simplified version of an approach taken by Skantze and Hjalmarsson [31], we split our utterances into chunks of meaning. For example, 3 analogous phrases each split into 3 analogous chunks allows for 27 unique recombinations, providing considerably more variation than the original set of 3 complete phrases. A speaker is placed behind the head and the appropriate text is synthesised using the inbuilt AppleTalk functionality of Mac OSX.

IV. THE STUDY

A. Participants

Participants were recruited by emailing musician lists and placing adverts on musician recruitment websites. There were 10 participants, 6 male and 4 female between the ages of 26 and 53. 3 classed themselves as beginner pianists, 5 as intermediate and 2 as expert. Although a small sample size, this is a practical constraint of requiring skilled participants.

B. Experimental Setup

We have designed two interfaces for the robot, one similar to a tool or instrument where it is controlled through simple stop and play buttons. The other presents the robot as a social actor, where the sessions are driven by *Mortimer* asking questions and providing positive and supportive feedback. Participants were assigned randomly to one of two experimental conditions, A and B. Both groups were briefed to improvise a piece in the 4/4 time signature with the robotic drummer. They were told that the study would last for 15 minutes but they may leave at any point.

To achieve the control condition, the tablet interface and general interaction experience was different for each of the two groups. For Condition B the tablet had a static interface with two buttons labelled "Play" and "Stop" and three sliders to control "Speed", "Length" and "Complexity". In Condition A, the robot gave a short autobiographical introduction on arrival in a synthesised male voice. It then invited the participant to play, to which they could accept or decline using "Yes" or "No" buttons. The following interactions were all framed in a similar way, with the robot providing supportive, positive feedback and politely verbally requesting any changes in behaviour or action. Figure 3 shows the complete flow of the interaction. The tablet interface changed appropriately depending on the question and possible responses.

C. Measures

We aim to measure engagement with the robot and any feelings of social presence felt by the participants, as we have cited these as crucial to forming sustainable and meaningful human robot relationships. We logged all key presses on the touch interface and transcribed all the piano played and drums composed. The robot has a frontal camera to capture the face and upper body of participants. As the position of both the piano and robot are constant and known, we were able to guarantee that these features would be captured with a stationary camera.

Self report is by far the most common approach to measuring social presence, however, Leite et al. use video analysis in tandem with a questionnaire [32] to evaluate social presence. They recorded events such as "user looking at iCat" and "user talking to iCat" as measures and then cross referenced these with questionnaire results. We view this as critical as, although social presence is concerned with a perceived experience and so could be measured by self report, it is questionable whether users have introspective access to this information. Also, as social presence may be something that

Measure	Mean A	Mean B	$t(8)$	p
Tracks per Session	6.2	9.6	-2.467	0.039
Bars per Track	58.0	38.5	2.668	0.028
Natural Stops (%)	87.0	35.6	2.954	0.018
Button Stops (%)	12.8	61.0	-2.426	0.041
Focus on Piano (%)	11.1	36.4	-3.928	0.004
Focus on Piano - playing (%)	10.2	34.4	-4.300	0.003
Focus on Robot - not playing (%)	47.1	29.3	2.510	0.041

fluctuates over time [8], a post-hoc questionnaire may be too blunt a tool alone to capture this.

Taking the definitions of engagement from Section II into account, we look at the data logs from the robot to find the length of sessions and length of uninterrupted tracks within this, as well as examining the use of the tablet interface and performance parameters. To ascertain the focus of the participants, we also analysed the footage using facial tracking and affect recognition software developed by Soyel and McOwan [33]. In doing so we hope to observe and measure behavioural changes during the study that display differences in engagement and social presence between the two control groups using a t-test. Results are displayed Table I

D. Results

1) Data Logs: All study participants in the B condition stayed for the full 15 minutes, the same is true for Condition A apart from 1 participant who stayed for 13 minutes and 10 seconds. As such we are unable to find any significant difference between the two groups on total length of session. Within the session, we recorded the number of tracks played, defined as a start and stop in play, either triggered manually, or occurring naturally as the robot had reached the end of its precomposed track. In Condition A a stop could also occur through prolonged silence from the human, in which case the robot would stop and give supportive feedback.

We found significantly less individual tracks within a session for Condition A, $t(8)=-2.467$, $p=0.039$, as well as significantly more bars per track in Condition A, $t(8)=2.668$, $p=0.028$. Further, we found that there were significantly more natural stops in Condition A $t(8)=2.954$, $p=0.018$, as well as significantly more manually triggered stops in Condition B $t(8)=-2.426$, $p=0.041$.

We did not find any significant differences in the mean values of any the performance parameters or the frequency with which they were altered, implying that the experimental condition did not have any bearing on how the participants explored them.

2) Automated Video Analysis: As demonstrated in Figure 4, we analysed all footage from the frontal camera using Soyel and McOwan's affect recognition software [33]. Based on Seeing Machine's faceAPI, it provides us with data as to the focus of the participants attention and also infers an affective state. With regards to focus, it distinguishes

Fig. 4. Face Tracking

between the piano, the robot and elsewhere. Initial tests for differences between the groups based on mean times spent in each state across the whole session indicated that Condition B spent significantly more time looking at the piano, $t(8)=-3.928$, $p=0.004$, and spent significantly more time in a thinking state $t(8)=3.402$, $p=0.009$. All other results showed no significant difference between groups. However, social and task-related context is crucial when attempting to infer information about the user's affective state or level engagement [20] so it is arguable how much insight we can gain from a contextless measure that completely ignores what else is happening during the session. For instance, looking at the piano while playing could suggest concentration, whilst looking at the piano while not playing could suggest shyness or avoidance. Or, since the interface was placed on top of the piano, spending time looking at this when not playing could be explained by participants changing parameters.

To investigate further, we measured the mean time spent in each state and took into consideration whether music was being played or not. Any change in behaviour during playing is of particular interest as both groups were interacting with exactly the same system once the music began. Once the conditions had been separated, we found that the difference for attention towards the piano was only significant when the participants were playing, $t(8)=-4.300$, $p=0.003$. This is perhaps explained by user thinking about the robot as nothing more than a sound making device, as one would a loudspeaker, and becoming more immersed in their own playing and the tablet. Further, we found that Condition A spent significantly more time looking at the robot when they were not playing $t(8)=2.510$, $p=0.036$.

With regards to the affective states, the results showed significantly more time in the thinking state when playing for Condition B $t(8)=-3.758$, $p=0.006$ and significantly more aggression when not playing $t(8)=3.011$, $p=0.017$.

3) Manual Video Analysis: When coding for social presence in a multi-session HRI trial, Leite et al. recorded verbalisations both in response to questions and when uninitiated by the robot [32]. We have also noticed verbalisations during the sessions from both groups and manually coded for these, taking it as a signifier of social presence in a system where the user knows the robot cannot hear or understand any verbal communication. 2 participants in

Condition A consistently engaged in back chatter (8 and 19 times) in response to questions. 2 participants from Condition B and 1 from Condition A verbally also greeted the robot before playing. Analysis was completed by three coders and inter-rater reliability was assessed using a two-way, mixed, consistency, average-measures ICC. The resulting ICC was in the excellent range, $ICC = 0.998$, indicating that coders had a high degree of agreement.

V. DISCUSSION

As nearly all the participants completed the full session, this shows that regardless of social framing, the musical interaction was an engaging one. Beyond this, having found that with the inclusion of social dialogue there were less sessions that lasted for more bars, we also suggest that Condition A played with greater fluency and were more engaged. This is enforced by the greater number of natural stops and lower number of manual stops. These latter findings also demonstrate a greater social presence as the participants treated the robot less like a machine or instrument and more like a social actor, waiting for the natural pause rather than stopping mid-performance.

Findings in Section IV-D.2 that those in Condition A looked more at the robot when not playing is perhaps to be expected as the robot was drawing attention to itself by speaking to the user. However, it does suggest a more physically embodied, spatially aware engagement with the robot itself. Indeed, we evidenced in Section II that looking at a robot has shown to be an accurate metric for engagement in previous HRI trials.

The apparent significance of playing and not playing highlights the importance of the distinct roles of mutual gaze in musical and conversational interaction. Gratier says that although gaze is necessary for 'grounding' in conversation, improvising musicians do not need to see each other for this purpose [34]. As such, our finding that focussing on the robot was only a significant difference when not playing is unsurprising and does not detract from our inferences as to the effect of the experimental condition on engagement and social presence. Moreover, the importance of gaze in grounding nonmusical social interaction between humans demonstrates that participants in Condition A were treating the robot as they would a human and allows us to conclude our results display greater feelings of social presence.

We do not put a huge amount of weight in the findings with regards to affective state as the definitions of these come from the context of children playing chess and so are not necessarily transferrable to our domain. This being said, Thompson et al. suggest that facial expressions are often used as *affect displays* in musical performance [35], meaning that automatic facial recognition techniques could be used if developed for a musical context. In future, we plan to extract a more general measure of facial animation from the data as this has been shown to be a signifier of interpersonal relationships [36].

Although their response is initiated by the robot, results in Section IV-D.3 are interpreted as a sign of social presence

as they were specifically told the robot could not understand them. The verbal greeting is perhaps a more interesting finding as those in Condition B were not expecting any interaction at all and so the social behaviour is triggered purely by the anthropomorphic form of the robot

VI. CONCLUSIONS AND FUTURE WORK

We have shown that presenting an interactive music system as a social actor rather than as a instrument changes the way people play and behave. We have found they play with more fluency and are less likely to stop and start the robot mid performance in the former condition. They also look at the robot more when not playing and look at the piano less when playing. We suggest that these results show greater engagement with the robot and the playing and greater sense of social presence when presented as even a rudimentary social actor. Even though the study was a single, short session, the results are promising and embolden us to continue along this path, extending to longitudinal studies with different simulated social behaviours. Recognising the limitations of using improvised music as a domain, further research as to whether our results would transfer to other creative activities or to more commonly occurring social interactions will be instructive.

REFERENCES

[1] C. E. Rusbult and P. A. M. Van Lange, "Interdependence, Interaction, and Relationships," *Annual Review of Psychology*, vol. 54, no. 1, pp. 351–375, Feb. 2003.

[2] J. Sung, H. I. Christensen, and R. E. Grinter, "Robots in the Wild: Understanding Long-Term Use," in *Proc. 2009 Human Robot Interaction Conf.*, San Diego, CA, 2009, pp. 45–52.

[3] Y. Fernaeus, M. Håkansson, M. Jacobsson, and S. Ljungblad, "How do you play with a robotic toy animal?: A long-term study of Pleo," in *Proc. 2010 Interaction Design and Children Conf.*, 2010, pp. 39–48.

[4] M. K. Lee, S. Kiesler, J. Forlizzi, and P. Rybski, "Ripple Effects of an Embedded Social Agent: A Field Study of a Social Robot in the Workplace," in *Proc 2012 Human Factors in Computing Conf*, 2012.

[5] N. Mitsunaga, T. Miyashita, H. Ishiguro, K. Kogure, and N. Hagita, "Robovie-IV: A Communication Robot Interacting with People Daily in an Office," in *Proc 2006 Intelligent Robots and Systems Conf*, Beijing, 2006, pp. 5066–5072.

[6] C. Breazeal, "Designing sociable robots," MIT Press, Cambridge, MA, 2004.

[7] R. S. Aylett, G. Castellano, B. Raducanu, A. Paiva, and M. Hanheide, "Long-term socially perceptive and interactive robot companions: challenges and future perspectives," in *Proc 2011 Multimodal Interaction Conf*, 2011, pp. 323–326.

[8] F. Biocca, C. Harms, and J. K. Burgoon, "Toward a More Robust Theory and Measure of Social Presence: Review and Suggested Criteria," *Presence: Teleoperators and Virtual Environments*, vol. 12, no. 5, pp. 456–480, Oct. 2003.

[9] M. Heerink, K. Ben, B. Wielinga, and V. Evers, "Measuring the influence of social abilities on acceptance of an interface robot and a screen agent by elderly users," in *Proc 2009 British HCI Group Annual Conference on People and Computers*, 2009, pp. 430–439.

[10] M. Giuliani, R. P. A. Petrick, M. E. Foster, A. Gaschler, A. Isard, M. Pateraki, and M. Sigalas, "Comparing task-based and socially intelligent behaviour in a robot bartender," in *Proc 2013 Multimodal Interaction Conf*, 2013, pp. 263–270.

[11] T. W. Bickmore and R. W. Picard, "Establishing and Maintaining Long-Term Human-Computer Relationships," *ACM Transactions on Computer-Human Interaction*, vol. 12, no. 2, pp. 293–327, 2005.

[12] T. Bickmore, D. Schulman, and L. Yin, "Maintaining engagement in long-term interventions with relational agents," *Applied Artificial Intelligence*, vol. 24, no. 6, pp. 648–666, 2010.

[13] D. Benyon and O. Mival, "From human-computer interactions to human-companion relationships," in *Proc 2010 Intelligent Interactive Technologies and Multimedia Conf*, 2010, pp. 1–9.

[14] K. M. Lee and C. Nass, "Designing social presence of social actors in human computer interaction," in *Proc 2003 Human Factors in Computing Systems Conf.*, Fort Lauderdale, FL, 2003, pp. 289–296.

[15] K. M. Lee, W. Peng, S.-A. Jin, and C. Yan, "Can Robots Manifest Personality?: An Empirical Test of Personality Recognition, Social Responses, and Social Presence in Human?Robot Interaction," *Journal of Communication*, vol. 56, no. 4, pp. 754–772, Dec. 2006.

[16] K. L. Nowak and F. Biocca, "The effect of the agency and anthropomorphism on users' sense of telepresence, copresence, and social presence in virtual environments," *Presence: Teleoperators and Virtual Environments*, vol. 12, no. 5, pp. 481–494, 2003.

[17] B. Reeves and C. Nass, *The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places* .

[18] Z. Kasap and N. Magnenat-Thalmann, "Building long-term relationships with virtual and robotic characters: the role of remembering," *The Visual Computer*, vol. 28, no. 1, pp. 87–97, Sep. 2011.

[19] C. L. Sidner, C. Lee, C. D. Kidd, N. Lesh, and C. Rich, "Explorations in engagement for humans and robots," *Artificial Intelligence*, vol. 166, no. 1, pp. 140–164, 2005.

[20] G. Castellano, I. Leite, A. Pereira, C. Martinho, A. Paiva, and P. McOwan, "Detecting Engagement in HRI: An Exploration of Social and Task-Based Context," in *Proc 2012 Privacy, Security, Risk and Trust Conf*, Amsterdam, 2012, pp. 421–428.

[21] G. Castellano, A. Pereira, I. Leite, A. Paiva, and P. W. McOwan, "Detecting user engagement with a robot companion using task and social interaction-based features," in *Proc 2009 Multimodal Interfaces Conf*, 2009, pp. 119–126.

[22] J. Sanghvi, G. Castellano, I. Leite, A. Pereira, P. W. McOwan, and A. Paiva, "Automatic analysis of affective postures and body motion to detect engagement with a game companion," in *Proc 2011 Human Robot Interaction Conf*, 2011, pp. 305–311.

[23] M. Michalowski, S. Sabanovic, and R. Simmons, "A spatial model of engagement for a social robot," in *Proc 2006 Advanced Motion Control Workshop*, 2006, pp. 762–767.

[24] F. Pachet, "The continuator: Musical interaction with style," *Journal of New Music Research*, vol. 32, pp. 333–341, 2003.

[25] A. R. J. François, E. Chew, and D. Thurmond, "Performer-centered visual feedback for human-machine improvisation," *Computers in Entertainment*, vol. 9, no. 3, Nov. 2011.

[26] G. Weinberg, A. Raman, and T. Mallikarjuna, "Interactive jamming with Shimon: a social robotic musician," in *Proc. 2009 Human Robot Interaction Conf.*, 2009, pp. 233–234.

[27] N. Collins, "Algorithmic Composition Methods for Breakbeat Science ," in *Proc 2009 Music Without Walls Conf*, De Montfort University, 2001, pp. 21–23.

[28] G. Weinberg and S. Driscoll, "Robot-human interaction with an anthropomorphic percussionist," in *Proc 2006 Human Factors in Computing Systems Conf*, 2006, pp. 1229–1232.

[29] E. Singer and J. Feddersen, "LEMUR: robotic musical instruments," in *Proc 2004 Multimedia Conf*, 2004, pp. 184–185.

[30] F. Thorne. (2011, Aug.) Felix's Machines. [Online]. Available: http://felixsmachines.com/

[31] G. Skantze and A. Hjalmarsson, "Towards incremental speech generation in dialogue systems," in *Proc. 11th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, 2010, pp. 1–8.

[32] I. Leite, C. Martinho, A. Pereira, and A. Paiva, "As Time goes by: Long-term evaluation of social presence in robotic companions," in *Proc 2009 Robot and Human Interactive Communication Conf.*, Toyama Int Conf Center, 2009, pp. 669–674.

[33] H. Soyel and P. W. McOwan, "Towards an affect sensitive interactive companion," *Computers & Electrical Engineering*, vol. 39, no. 4, pp. 1312–1319, May 2013.

[34] M. Gratier, "Grounding in musical interaction: Evidence from jazz performances," *Musicae Scientiae*, vol. 12, no. 1 suppl, pp. 71–110, 2008.

[35] W. Thompson and P. Graham, "Seeing music performance: Visual influences on perception and experience," *Semiotica*, no. 156, pp. 203–227, 2005.

[36] P. A. Andersen and J. F. Andersen, "Measures of Perceived Nonverbal Immediacy," in *The Sourcebook of Nonverbal Measures: Going Beyond Words*, 2004.