Extending Human-Robot Relationships Based in Music with Virtual Presence

Louis McCallum* and Peter McOwan

***Abstract*—Social relationships between humans and robots require both long term engagement and a feeling of believabilty or social presence towards the robot. It is our contention that music can provide the extended engagement that other open-ended interaction studies have failed to do, also, that in combination with the engaging musical interaction, the addition of simulated social behaviours is necessary to trigger this sense of believability or social presence. Building on previous studies with our robot drummer *Mortimer* that show including social behaviours can increase engagement and social presence, we present the results of a longitudinal study investigating the effect of extending weekly collocated musical improvisation sessions by making *Mortimer* an active member of the participant's virtual social network. Although we found the effects of extending the relationship into the virtual world were less pronounced than results we have previously found by adding social modalties to human-robot musical interaction, interesting questions are raised about the interpretation of our automated behavioural metrics across different contexts. Further, we found repeated results of increasingly uninteruppted playing and notable differences in responses to online posts by *Mortimer* and posts by participant's human friends.**

***Index Terms*—Human-Robot Relationships, Musical Robotics, Virtual Presence.**

I. Introduction

SOCIAL Robotics focusses on developing robots that can interpret human social cues and act appropriately in response. However, trials rarely extend beyond a single session. Whilst this is satisfactory for developing robots that will interact with humans in brief single visit scenarios, such as service robots, it fails to provide insight into situations where a relationship may form between human and robot. A relationship necessarily occurs over multiple interactions and in the longitudinal studies that have been carried out, there is often a notable decline in positive responses over time [1]–[3]. To avoid this, a robot needs to be able to engage the human and maintain and develop this across multiple interactions. Studies have shown that taking part in ensemble music can give participants a sense that they are making an important contribution to the group, feelings of pride in group success and a sense of belonging [4]. Also that music preferences are closely aligned with friendship choices [5] and that humans often use music for mood management [6]. As such, it is our hypothesis that open-ended creative activities, such as musical improvisation, may be able to provide this engagement as they are naturally progressive, involve a high quality of affective interaction and are often fitted into social routines.

In their day-to-day lives, most humans encounter machines and computer programs capable of executing impressively complex tasks to a high standard that may provide them with hours of engagement. However, in order to have anything that could be classed as a social relationship, the human must have the sense that their interactions are taking place with another, a phenomenon known as social presence [7]. This concept addresses similar aspects of a human's perception of a robot as the notion of believability, already prevalent in Sociable Robotics research [8], [9] and is described as the amount to which a person can suspend their knowledge that a robot is inanimate and not actually in possession of the human faculties we attempt to make it display.

As such, we investigate the addition of simulated social behaviours to a robot capable of musical improvisation in order to provide the social presence and long term engagement necessary to form a positive human-robot relationship. To these ends, we developed a robotic drummer *Mortimer*, able to compose music responsively to human pianists in realtime. Our previous work has found that by simply framing the musical improvisation as a social interaction, you can improve both social presence and engagement [10]. Also, in a further longitudinal study, we discovered that the inclusion of musically and socially appropriate nonverbal action resulted in indicators of increased engagement, with longer time spent voluntarily and less interruptions during playing, and increased social presence, with less interruptions during social interaction [11].

Building on these results, in this article we examine the effect of extending the relationship beyond lab based sessions with a physical robot. One of the limitations of our previous research had been that we only have one robot which is only usable in fairly supervised and regulated studio situations. Whilst it is unlikely that two humans would have complete and unadulterated access to each other's time and location, the constraints of the current physical embodiment are more restrictive than those of a human-human relationship. We believe this to be to the detriment of engendering and maintaining long term engagement.

Non-collocated social interactions are common place for much of the world's population, taking part mainly through online social networks such as Facebook. This particular site currently has 1.44 billion active monthly users [12] and since becoming the dominant social network after the decline of

L. McCallum and P. McOwan are with the School of Electrical Engineering and Computer Science, Queen Mary University of London, UK. (e-mail: [l.mccallum, p.mcowan]@qmul.ac.uk).

*This research is supported by the EPSRC as part of the Media and Arts Technology Doctoral Training Centre, EP/G03723X/2

forerunners Friends Reunited and MySpace, it is used by those across generations and across the globe to extend their existing social relationships from the physical world into the virtual world. Although now offering a wide range of services, its original purpose as a place to share photographs online and this remains core to most user's experiences.

As such, we extended *Mortimer's* capabilities to allow him to take pictures during sessions and post them with a supporting comment to Facebook. This paper describes a longitudinal study conducted into any advantages this may have in developing a human-robot relationship when used to supplement sessions focussed around musical improvisation.

II. Related Work

A. Musical Robots

Musical robots have been constructed as art installations, as performers and spectacles and as marquee examples of engineering sophistication. More recently they have also been used as the physical embodiments of interactive music systems and it is these we will focus on.

Georgia Tech's *Haile*, a percussive robot, is equipped with a real time beat tracking module and two beater-arms capable of expressively collaborating with a human player on a Native American Pow Wow drum [13]. Rather than providing a platform for composers to write new music, Weinberg et al. aim to stimulate 'inspiring human machine collaboration' [1]. Similarly, *Shimon*, is a robotic marimba player who presented as a 'Social Robotic Musician' [2]. The researchers validate this with the description of a social module to provide visual cues for human participants via a screen based animated head [14].

Although not physically embodied, Francois Pachet's *The Continuator* [15] holds many interesting and potentially desirable properties with regards to the type of interaction we wish to illicit. *The Continuator* neatly side-steps issues with longer term form in algorithmic compositions systems by allowing the user to be control of the extended form of the improvisation, while the system fills in gaps and provides responses locally. With regards to our interests, the introduction of learning into the system allows for an extended, and widely varied, personalisation process which can avoid problems with lack of personal engagement perceived in many attempts at social systems.

B. Virtual Social Agents

The Facebots project set out with the aim to use data from an online social network to inform social interaction in the physical world [16]. One of the first to situate a robot within a virtual social network, it also used pictures from the sites to inform facial recognition [17].

[1] [13]
[2] [14]

Beyond their use as malware or spammers, programmes that automatically use email and social networks have also been used in both art practice and scientific research. One such example of artificial agents inhabiting social networks alongside humans is *Weavrs*, a web platform allowing the generation and proliferation of artificial, virtual and social agents [18]. So called bots have even been used on mobile dating app *Tinder* [19].

The process of transferring a consistent agent across multiple virtual or physical embodiments is known as migration [20]. Whilst we are not technically migrating our robot across embodiments, similar issues are raised as we are attempting to provide both virtual and physical presences to *Mortimer*. Koay et al. investigate migrating a personality between two different physical embodiments [21] and stress making sure the migration is clear and smooth. Phenomena such as "overlaps" or "gaps" that can happen as a result of technical difficulties can cause uneasiness or anxiety [22]. Robert et al. place similar weight on the importance of consistency between virtual and physical worlds in a mixed reality robot game [23].

III. Mortimer

Our stated aims are investigating how social interactions between social robots and humans can progress or deteriorate over time and the role that open-ended creative activities can take in avoiding a decline in favourable response, with a view to aiding the development and maintainence of positive and sustainable human-robot relationships. This is distinct from a physically embodied interactive music system built to develop creative ideas as a compositional tool or to perform music as a spectacle. As such, whilst the designs of these systems will share common specifications, there are certain nuances to be taken into account that will lead to the development of a system best suited to our particular needs and the robot best equppied to develop positive relationships with humans may not be the most expressive or the one that allows for the greatest variety of musical output.

Mortimer, shown in Fig 2, is a robot developed by the authors over the last 2 years, the musical and social faculties of which are described in previous papers [10], [11]. We will provide a brief description of the composition algorithm used and the social behaviours implemented below. The novel developments to enable virtual communication are covered in Section IV-B.

A. Musical Interaction

Taking the design implications covered in Section III into account, the musical context of a human pianist and robotic drummer was decided upon. *Mortimer* is equipped with two solenoid-driven beater arms which are used to play a closed hi-hat and snare drum. He also has an automated kick drum, completing the minimal requirements for a standard Western drumkit setup.

Each session is made up of tracks, with each track consisting of a number of choruses, each concluded by a breakdown section. As such, each bar within a chorus will either be the replaying of a base groove as is, an ornamented version of a base groove or a reduced version of the base groove. Whilst the form is generated at the beginning, taking into account user-specified length and complexity parameters of the track, the music itself, that is, the score played by the robot, is not composed until these specified points, allowing the composition to take into account the most up to date information about the human's contribution. This includes previous piano input and explicit performance parameters set in between each track, such as complexity.

The base pattern for the snare and kick is generated using 0th order Markov approach. Each semi-quaver position has a manually ascribed probability and is used to stochastically compose a bar of each. Before generation, the base probability tables for snare and kick are augmented by both a histogram of the previous rhythmic input from the human and the explicitly inputted complexity procedure. There are 3 possible hi hat patterns, each with a static probability of being chosen.

In addition to composing bars based on a form composed at the beginning, we take note density and mean note velocity as a measure of power for the piano playing. If this drops below certain thresholds then instruments are either dropped or thinned by the drummer in order to match a perceived sparsening of the texture of reduction of dynamics. *Mortimer* will also try and get a feel for the groove of the human input and try to match this with the timing deviations of the robot. Although human input is taken into account in the composition and timing, overall tempo remains at the constant speed specified by the user in between each track for the track's duration.

B. Social Interaction

For each session, the pianist is greeted by *Mortimer* who communicates via speech sysnthesis software. They are then guided through each session, interacting with *Mortimer* through a tablet placed on top of the piano. Head poses and facial expressions are displayed at socially and musically appropriate times and are described in detail in [11].

IV. Method

A. Participants

Participants were recruited by emailing musical lists and placing adverts on musician recruitment websites. There were 11 participants, 7 male and 4 female between the ages of 18 and 44. They were asked to self-asses their own skill at playing the piano and there was a wide range present in the group (1-5=beginner-expert, min=1, max=5, mean=3.4, SD=1.07). Even though the number of participants is relatively small, which was a practical constraint of needing skilled participants, as each returned multiple times we conducted 66 sessions in total.

B. Experimental Setup

Each participant took part in 6 weekly sessions in a controlled studio environment. They were instructed to stay for a minimum of 20 minutes but could optionally stay for up to 45.

Prior to the start of the study, participants were asked to complete a short background questionnaire composed by the authors with regards to their demographic information and daily social media usage. From this a social media usage score was generated for each participant. Participants were then split into 2 groups, ensuring a diversity of gender and social media usage between experimental group (Condition A) and the control group (Condition B). During the sessions, interactions between groups were socially and musically identical. However, for those in Condition A, a simple interaction based around picture sharing on the social networking site Facebook.com was enacted outside of the sessions.

3 days prior to their first session, a friend request was sent to participant's in Condition A from a Facebook account purporting to be owned by *Mortimer*. Then, during each session, a picture of *Mortimer* and the participant playing was taken automatically by a webcam in the lab and an accompanying comment was generated. Macdorman and Cowley emphasise the need for a robot to have personally specific and developing relationships with humans to build a strong social bond and as such we use this comment to personalise the post to the user and the session [24]. This included either what was happening in the session at the time the photo was taken with regards to performance parameters, for example, how fast they were going, or how the session ranked comparatively to other sessions with regards to performance parameters and session length, for example, if this was the longest session to date. It also included an optional further comment about weather at the time of the session [3].

Once a day a script was manually run which found any pictures that needed to be posted and added them to *Mortimer's* Facebook wall along with the accompanying comment. The user was also tagged in the photo, meaning they received a notification of the photo's posting. Examples of these posts can be seen in Fig 2. *Mortimer* also ran replys to posts through the sentiment analysis algorithm of Narayanan, Arora and Bhatia [25] and if they were deemed positive or neutral, *Mortimer* would post a reply thanking them.

C. Measures

In previous work, we have developed an approach of automated behavioural metrics for investigating the quality of a human-robot relationship. Using this approach we take several domain-specific quantiative measures of the interaction, taking occurences of these to be signifiers of engagement, social presence or a close interpersonal relationship. Measures taken are session length, interupptions, explicit button stops,

[3] http://openweathermap.org/api

proportion of session spent playing and mean length of track. We also determine participant focus using camera-based face-tracking.

Additionally, we recorded the number of "likes" and comments by participants and members of their social network on posts made by *Mortimer* and posts made by themselves that they tagged *Mortimer* in and measured self-reported intention of repeat interaction in an exit survey.

V. Results

A. In Session Results

1) Quantitative Interaction Data: We fitted a random intercept linear mixed effect model for the fixed effects of week, group and the interaction between the two for each measure. Results are displayed in Table I.

We found significant effect of week number (β = 6.21, 95% CI [4.19 8.27], *p*=0.0005) and the interaction between week and group (β = 3.89, 95% CI [0.01 7.79], *p*=0.0005) for the mean number of bars per track. Shown clearly in Figure 1, the mean number of bars per track increases as the weeks continue, and whilst there is no significant difference between the groups overall, the change over time is greater for Condition B.

We further find an effect of the interaction between week and group (β = 84.60, 95% CI [33.13 135.78], *p*=0.0150) for the time spent of the minimum in each session, indicating that Condition B has a positive effect on the trend over time in relation to Condition A. In this case, Figure 1 demonstrates the steady reduction in time, over time, for those in Condition A, whilst those in Condition B see a week on week rise for week 2 to 5 before a drop off in the final session.

2) Automatic Video Analysis: We fitted a random intercept linear mixed effect model for the fixed effects of week, group and the interaction between the two for each category. Due to a technical fault, video was unavailable for analysis for one session from the first week. Therefore, the analysis is done on 65 of the 66 sessions. The results show that there was no significant differences between the groups or over time for any of the categories.

3) Self Report: We found that of all 11 participants, they reported an average of 4.45 out of 5 chance that they would return to do more sessions with *Mortimer* if the opportunity existed. There was, however, no statistical difference between the groups.

B. Facebook Interaction

As by far the most commonly occurring form of interaction with posts, we present the results for likes by others. They show that the total number of likes a post received was considerably higher if the user posted the picture themselves (22 and 63), although this only happened twice, both times in the first week. An exmaple of one of these such posts is

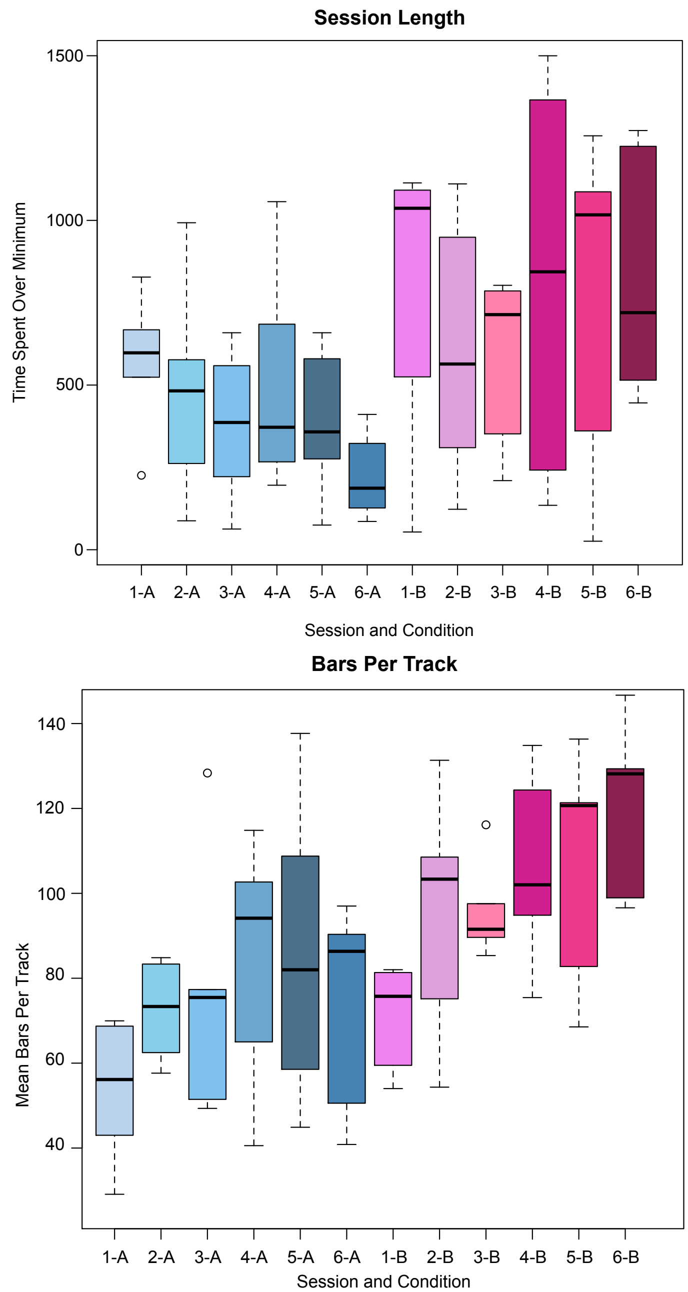

Fig. 1. Results of Random Intercept Linear Mixed Effect Model for time spent over the minimum session length (above) and number of bars per track (below) for each session grouped by Condition

included in Fig 2. It also demonstrates that posts by *Mortimer* did not suffer a novelty effect in the participant's social networks, as the highest proportion of posts liked (5) and the second highest mean number of likes per post (2.17) occurred in the final week.

Likes by participants themselves were rare, with only 3 occurring over the whole 5 weeks. Comments by participants only occurred once and comments by others only 3 times. In contrast, the two posts by participants received 1 and 6 comments, making the comparative means between *Mortimer*-posted and participant-posted 0.1 and 3.5 respectively.

Data	Fixed Effect	Estimate β	CI [5% 95%]	p
Session	Week	-17.11	[-44.57 10.68]	0.3193
Length	Group	303.63	[36.59 591.65]	0.1229
	Week.Group	84.60	[33.13 135.78]	0.0150*
Bars Per	Week	6.21	[4.19 8.27]	0.0005***
Track	Group	24.10	[7.97 40.53]	0.0605
	Week.Group	3.89	[0.01 7.79]	0.0005***
Button	Week	-0.01	[-0.02 0.00]	0.2584
Stops	Group	-0.07	[-0.2 0.08]	0.4968
	Week.Group	-0.02	[-0.05 0.00]	0.3398
Inter	Week	-0.04	[-0.11 0.04]	0.4343
-ruptions	Group	-0.01	[-0.64 0.54]	0.9750
	Week.Group	0.12	[-0.03 0.27]	0.5567
Time	Week	0.71	[0.01 1.40]	0.0960
Playing	Group	4.60	[0.83 8.39]	0.1034
(%)	Week.Group	-0.40	[-1.79 0.93]	0.1109

Random Intercept Linear Mixed Effect Model for quantitative interactional data. P values are estimated from a parametric bootstrap (2000 replicates). Confidence Intervals are estimated from a parametric bootstrap (2000 replicates). $*p<0.05$, $**p<0.01$, $***p<0.001$.

TABLE I
QUANTITATIVE INTERACTION DATA

VI. DISCUSSION

Based on previous studies and the overarching hypothesis running through our work, we would expect the inclusion of additional social modalities to result in increased engagement. We have suggested that the length of time a person voluntarily spends with a robot can key be a indicator of engagement and in our previous work [11] reported that those within the social condition not only spent more time overall but also increased the amount of time they spent as the study continued. A repeated trend is seen in this experiment, with the time spent increasing over the study (see Figure 1, the experimental condition from [11] and Condition B are analagous). Further, we find a significant difference in how the session length changes over the study between the two conditions, however, it is the pro-social condition, Condition A, that actually reduces over time.

With a study design that allowed identical access to the robot for all conditions, as in our previous work, this result would appear to show a decrease in engagement for those in Condition A, however, this was not the case. Participants in Condition A had opportunities for additional contact with *Mortimer* outside of the physical sessions, possibly leading to a reduced need to spend time in the physical sessions. This result highlights the potential issues which may arise from study design when picking experimental measures, in this case, a measure suitable in previous studies is compromised by new modes of interaction. It also raises an interesting question regarding whether humans will spend less time with a physically embodied robot if they know they can interact with it later on a virtual platform.

Fig. 2. A post by a participant (above) and a post by *Mortimer* (below)

Similarly to [11], the length of tracks within sessions increased over the trial period, regardless of experimental condition. We can again draw the conclusion that learning over time is a more important factor than increased social modality in increased fludity in playing.

From the Facebook data, there were considerably more ”likes” from a user’s social network for posts made by a user themselves, as opposed to one posted by *Mortimer* that the user was tagged in. There are several explanations for this outcome, one being the differences in the photos themselves. Both participant posted pictures were ”selfies” in which the participant is facing the camera and, perhaps importantly, the audience. This is in contrast to the photographs posted by *Mortimer*, taken whilst playing and often without the player looking directly at the camera. Alternatively, the effects could have been caused by the quality of the comments generated. It may be that *Mortimer's* comments were viewed as perscriptive or disengenuous and the participant's were more natural and "likeable". It may even have been a poster bias, with just by the fact the picture was posted directly by a friend as opposed to a stranger, in this case *Mortimer*, or that one was posted by a robot, that explains the disproportionate amounts of "likes". One thing to note is that proper comparisons between the two groups, self-posted and robot-posted, are limited due to the data set haveing 30 entries in the latter and only 2 in the former. This makes any generalisations hard, although the size of the difference, even with the limited data, makes it worthwhile considering.

Unlike our previous work [10], [11], where facetracking revealed differences in the focus of the participants, no differences were found between the groups in this experiment. However, unlike the previous studies, the in-session interaction between the groups was identical in this case. From this it may be inferred that any changes caused by the experimental condition are less immediately reflected in the nonverbal behaviour of the participant.

There was a generally high score for self-reported repeat interaction, suggesting that the although the experimenal condition may not have had the expected influence, the in-session activities were engaging for all participants.

VII. Conclusion

Using the methodology of automated behvaioural metrics developed in previous work, we found the effects of extending the relationship into the virtual world were less pronounced than results we have previously found by adding social modalties to human-robot musical interaction. The results also raised a question as to the appropriate use of session length as a measure of engagement in this context. Further, analysis of the Facebook data provided some noteworthy differences in interactions with posts by participants and posts by *Mortimer*, however the former category had too small a dataset to draw any solid conclusions.

Moving forward, more experiments would illuminate whether extending the relationship into the virtual world is simply not a particularly useful tool in this domain or that a higher quality of virtual interaction is required to trigger positive effects on human robot relationships.

References

[1] R. Gockley, A. Bruce, J. Forlizzi, M. Michalowski, A. Mundell, S. Rosenthal, B. Sellner, R. Simmons, K. Snipes, A. C. Schultz, J. Wang, Designing Robots for Long-Term Social Interaction, in: Proc IROS'05, Edmonton, 2005, pp. 1338–1343.

[2] Y. Fernaeus, M. Håkansson, M. Jacobsson, S. Ljungblad, How do you play with a robotic toy animal?: A long-term study of Pleo, in: Proc. IDC'10., Barcelona, 2010, pp. 39–48.

[3] M. K. Lee, S. Kiesler, J. Forlizzi, P. Rybski, Ripple Effects of an Embedded Social Agent: A Field Study of a Social Robot in the Workplace, in: Proc CHI'12, Austin, TX, 2012, pp. 695–704.

[4] D. Kokotsaki and S. Hallam, Higher Education Music Students Perceptions of the Benefits of Participative Music Making, in: Music Education Research, 2007, Vol.9(1), pp.93-109.

[5] M., Selfhout,S. Branje, T. ter Bogt, and W. Meeus, The role of music preferences in early adolescents friendship formation and stability, in: Journal of Adolescence, 2009, Vol.32(1), pp.95-107.

[6] D. J. Hargreaves and A. C. North,, The Functions of Music in Everyday Life: Redefining the Social in Music Psychologym in: Psychology of Music, 1999, Vol.27(1), pp.71-83.

[7] F. Biocca, C. Harms, J. K. Burgoon, Toward a More Robust Theory and Measure of Social Presence: Review and Suggested Criteria, Presence: Teleoperators and Virtual Environments 12 (5) (2003) 456–480.

[8] C. Breazeal, Designing sociable robots, MIT Press, Cambridge, MA (2004).

[9] R. S. Aylett, G. Castellano, B. Raducanu, A. Paiva, M. Hanheide, Long-term socially perceptive and interactive robot companions: challenges and future perspectives, in: Proc ICMI'11, Alicante, 2011, pp. 323–326.

[10] L. McCallum, P. McOwan, Shut up and Play: Engagement and Social Presence Human-Robot Musical Interaction, in: In Proc RO-MAN14, 2014, pp. 138–143.

[11] L. McCallum, P. McOwan, Face the Music and Glance: How Nonverbal Behaviour Improves Human-Robot Relationships Based in Music, in: Proc HRI'15, Porland, OR, 2015, pp. 138–143.

[12] F. Newsroom. http://newsroom.fb.com/company-info/ [online] 2015. Accessed: 15-05-30.

[13] G. Weinberg, S. Driscoll, Robot-human interaction with an anthropomorphic percussionist, in: Proc CHI'06, Montreal, 2006, pp. 1229–1232.

[14] G. Weinberg, A. Raman, T. Mallikarjuna, Interactive jamming with Shimon: a social robotic musician, in: Proc. HRI'09, ACM, San Diego, CA, 2009, pp. 233–234.

[15] F. Pachet, The Continuator: Musical interaction with style, in: Journal of New Music Research, Vol.32, 2003, pp.333–341.

[16] N. Mavridis, On artificial agents within human social networks: Examples, open questions, and potentialities, in: Proc. DEST'10, Dubai, 2010, pp. 685–690.

[17] N. Mavridis, M. Petychakis, A. Tsamakos, P. Toulis, S. Emami, W. Kazmi, C. Datta, C. BenAbdelkader, A. Tanoto, FaceBots: Steps Yowards Enhanced Long-Term Human-Robot Interaction by Utilizing and Publishing Online Social Information, Paladyn 1 (3) (2011) 169–178.

[18] P. Phactory. http://www.weavrs.com [online] 2012. Accessed: 15-05-30.

[19] BBC. http://www.bbc.co.uk/newsbeat/article/31920480/tinder-user-falls-for-robot-woman-at-sxsw-festival [online] 2012. Accessed: 15-05-30.

[20] P. Gomes, E. Segura, H. Cramer, T. Paiva, ViPleo and PhyPleo: Artificial pet with two embodiments, in: Proc ACE'11, New York, NY, 2011, pp. 3:1–3:8.

[21] K. Koay, D. Syrdal, M. Walters, K. Dautenhahn, A User study on visualization of agent migration between two companion robots, in: Proc HCI'09, San Diego, CA, 2009.

[22] E. M. Segura, H. Cramer, P. F. Gomes, S. Nylander, A. Paiva, Revive!: reactions to migration between different embodiments when playing with robotic pets, in: Proc IDC'12, Glasgow, 2012, pp. 88–97.

[23] D. Robert, R. Wistorrt, J. Gray, C. Breazeal, Exploring mixed reality robot gaming, in: Proc TEI'11, Funchal, 2011, pp. 125–128.

[24] K. Macdorman and S. Cowley, Long- term relationships as a benchmark for robot personhood, in: Proc. RO-MAN'06, Hertfordshire, 2006, pp.378-383

[25] V. Narayanan, I. Arora, A. Bhatia, Fast and accurate sentiment classification using an enhanced naive bayes model, in: H. Yin, K. Tang, Y. Gao, F. Klawonn, M. Lee, T. Weise, B. Li, X. Yao (Eds.), Intelligent Data Engineering and Automated Learning, Vol. 8206 of Lecture Notes in Computer Science, 2013, pp. 194–201.