Face the Music and Glance: How Nonverbal Behaviour Aids Human Robot Relationships Based in Music

Louis McCallum
School of Electronic Engineering and Computer Science
327 Mile End Rd
London
l.mccallum@qmul.ac.uk

Peter W McOwan
School of Electronic Engineering and Computer Science
327 Mile End Rd
London
p.mcowan@qmul.ac.uk

ABSTRACT

It is our hypothesis that improvised musical interaction will be able to provide the extended engagement often failing others during long term Human Robot Interaction (HRI) trials. Our previous work found that simply framing sessions with their drumming robot *Mortimer* as social interactions increased both social presence and engagement, two factors we feel are crucial to developing and maintaining a positive and meaningful relationship between human and robot. For this study we investigate the inclusion of the additional social modalities, namely head pose and facial expression, as nonverbal behaviour has been shown to be an important conveyor of information in both social and musical contexts. Following a 6 week experimental study using automatic behavioural metrics, results demonstrate those subjected to nonverbal behaviours not only spent more time voluntarily with the robot, but actually increased the time they spent as the trial progressed. Further, that they interrupted the robot less during social interactions and played for longer uninterrupted. Conversely, they also looked at the robot less in both musical and social contexts. We take these results as support for open ended musical activity providing a solid grounding for human robot relationships and the improvement of this by the inclusion of appropriate nonverbal behaviours.

Categories and Subject Descriptors

H.1.2 [**Information Systems**]: User/Machine Systems; J.4 [**Computer Applications**]: Social and Behavioural Sciences—*psychology*

Keywords

Human-Robot Relationships; Music; Nonverbal Behaviour; Long Term Studies

HRI'15, March 2–5, 2015, Portland, Oregon, USA.

http://dx.doi.org/10.1145/2696454.2696477.

1. INTRODUCTION

There is a paucity of long term studies in the field of Social Robotics, and an even greater draught of those managing to maintain initial positive results over time. Problems with this novelty effect have been seen in research with smart vacuum cleaners [32], robotic dinosaurs [10] and anthropomorphic robots [21]. However, such studies are a necessity for those wishing to investigate human robot relationships, as these are something which will always need time to develop and almost certainly change as they progress. We highlight the maintenance of engagement and a sense of social presence, defined as the feeling of being with someone when they are not physically present [3], as two key factors in developing a positive relationship between human and robot. Further, we suggest improvised music, as a naturally progressive, affective activity around which social bonds often develop as a favourable bedrock to build up such relationships.

Building on our previous findings that framing a session between pianists and their drumming robot, *Mortimer*, as a social interaction resulted in greater feelings of engagement and social presence [23], we present a study in which we investigate the inclusion of further social modalities, namely, head movements and facial expressions.

In their canonical survey of the field, Fong et al. cited realistic facial expression as a key design factor in social robots, especially in the demonstration of affective behaviour [13]. Further, being able to communicate and interpret nonverbal actions can be crucial to the success of social interactions [17]. Noller extends this by claiming nonverbal communication is important for maintaining social bonds, as it allows people to express emotions and to relay how they feel about each other and the relationship [26]. She also reports that the accuracy of decoding of nonverbal cues is often a predictor of relationship closeness and satisfaction. Tickle-Degnen suggests that nonverbal expressivity on the whole tends to have positive social outcomes, including rapport [34]. Within this, Fridlund and Russel claim that faces play a key part in our social interactions [14], indeed, interpreting and imitating facial expressions is one of the first skills an infant learns [28]. Motivated by this, we developed a set of head movements and facial expressions triggered by social and musical cues for *Mortimer*.

We conduct 6 weekly sessions per participant in order to study the effect of a control condition on, and the suitability of musical improvisation in general for, maintaining engagement over time. Following the methodology developed in our

previous research, we rely mainly on automated behavioural metrics [23], analysing data logs to see how participants interacted with the robot and using face tracking to determine where they are focussing their attention during the sessions. In relation to a control group, we expected the use of head movements and facial expressions to increase social presence and engagement within the sessions, seen by increased session time, smoother playing and more displays of behaviour indicative of an interpersonal relationship.

Section 2 covers related research, Section 3 describes our technical development, Section 4 Section 5 detail the study we conducted and its findings. These are discussed in Section 6. In Section 7 we summarise and outline future research directions.

2. RELATED WORK

2.1 Nonverbal Cues in Musical Performance

Nonverbal cues, notably facial expressions, mutual gaze and head movements are used by musicians to convey information about the music either to co-performers or audience members. This serves an especially important role in improvised music.

In almost all acoustic music performance, the body, and in some cases the head and face, are inseparably coupled to the generation of sound [16, 33, 35]. However, they are also used as cues, intentionally or not, to augment the performance and to anticipate or accentuate important events. For example, in an analysis of an improvising jazz guitarist, Gratier demonstrates that musicians may use their bodily movements to convey the structure and meaning of the music [16]. Similarly, Vines et al. discovered that the perceived tension of a performance is most influenced by visual, rather than auditory, cues [35]. They also report that it is a combination of auditory and visual stimulus that effects audience's perception of phrasing in a musical performance, providing the supporting observation that the contours of the performer's body movement tended to align with their phrasing of the music. Further, Thompson et al. find that facial expressions are used to convey timing events, thus increasing musical intelligibility [33]. They also report that facial expressions can be used to make music sound more or less dissonant or to make musical intervals sound further apart or closer together.

Gratier suggests that facial displays of affect may serve the purpose of grounding between improvisers. For example, a musician may smile at a mistake or a particularly satisfying lick [16]. Moreover, whilst drawing comparisons between improvised music and conversation, she reports that mutual gaze is much less constant in the former. This being said, although less frequent, it still serves a crucial role in managing the interaction and tends to occur during moments of structural change or importance in the music.

In a study of a performance by blues guitarist BB King, Thompson et al. find he often used facial expressions to display affect. For example, in moments of tension he takes on an introspective demeanour, looking down and shaking his head. A musicologist interprets this as him signally he feels the emotion but will not submit to it. Alternatively, in moments of release he opens his mouth towards the audience as if in wander. As well as relating to affect, they find King's head movements often react to individual notes and licks and tend to reflect only his performance, rather than that of his band. A study of a Judy Garland performance by the same authors reveal how she uses hand gestures in a more illustrative fashion, literally reflecting the lyrics of the song, displaying the range of purposes bodily movement can play for different performers.

2.2 Facial Expressions in Social Interaction

Since the early 1960s, psychologists have prevalently viewed the face as the key factor in understanding the emotions of humans. However, Chovil makes the argument that facial expressions are not primarily, or even at all, expressions of an internal affective state but serve the purpose of being socially communicative actions [7]. Kraut and Johnston demonstrated that smiles were more likely to occur during social interaction than in situations of happiness in a study of ten-pin bowlers [19]. Further, analysing gold medal ceremonies, Fernandez-Dols and Ruiz-Belda found that a greater proportion of smiles occurred in the interactive stage of the event than elsewhere. This is surprising, considering the whole event is assumed to be one where the athletes will feel intense joy throughout [11]. The rejection of the emotional cause for facial expressions is taken the extreme by Fridlund and Russel, who introduce the Behavioral Ecology View (BEV) [14], providing an alternative socially communicative explanation for the all the expressions which others have claimed are "readouts" of prototypical emotions. For instance, smile moves from "readout of happiness" to a signifier of "readiness to affiliate or play"[1] and "readout of anger" becomes the message "readiness to attack"[2]. Under Fridlund and Russel's treatise, *Mortimer* should use his face to reflect planned intentions and goal states, not emotions.

Regardless of the intention, be it internal affective mirror or socially communicative gesture, it is worth examining what information a face can reliably relay to others within a social interaction. It is reasonable to suggest the face can allow us to distinguish between pleasant and unpleasant expressions and between differing degrees of these expressions [9]. There is also strong agreement between researchers that an eyebrow frown is a sign of negativity or concentration and a smile is a signifier of pleasantness [30]. Beyond this, there is good evidence to show that at least 6 distinct facial expressions can be universally distinguished and recognised [9] and these have been classed as happiness, sadness, surprise, disgust, anger and fear. Smith and Scott outline a further componential model which defines 6 types of behaviours and how they can be expressed [30]. This includes pleasantness, goal obstruction, anticipated effort, attentional activity, certainty, novelty and personal agency and draws from not only their own research but various historical models.

Fernandez-Dols and Carrol demonstrate that although much research treats it as such, it is inherently problematic and reductive to consider facial expressions outside of their context [12]. If we are to clearly and unambiguously use the face of the robot to demonstrate social and musical cues and emotions then we must be aware of the context that they are being produced in, otherwise they may fail to be interpreted as intended. Luckily, in our laboratory experiments, the context is known and controlled to a high degree.

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

Figure 1: Mortimer

2.3 Head Movements in Social Interaction

Head movements are far from arbitrary, they have been shown to reliably occur at certain points in social interactions, serving many functions from emblematic replacements of speech, to turn management and backchannelled affirmation [24]. As a general rule, speaker's heads tend to be in constant motion whereas listeners tend to be relatively static.

In a microanalysis of a corpus of filmed social interactions, McClave found several consistent co-occurrences of head movements and social cues [24]. For example, a lateral sweep is used to demonstrate inclusivity, often concurrently with words such as "everyone" and "whole". Repeated head movements also often coincide with listed items when a speaker is delivering alternatives. Further, head shakes, as well as serving the emblematic purpose of negation, are often used during speech to emphasise a sentiment more intensely or to express uncertainty. This was also seen by Iwano et al., who found that horizontal head movements occurred during denials [18]. Similarly, both found that a head nod, or vertical movement, is often used to demonstrate affirmation, agreement and continuing comprehension. Iwano et al. also found that when speakers are expecting a response, such as preceding a question, they often lift their head up to face their partner directly [18]. In terms of persuasiveness, BriÃśol and Petty find that nodding and head shaking during conversation stands to strengthen or undermine your argument respectively [6]. Head movements can also provide attentional cues that make up our sense of engagement with another [25].

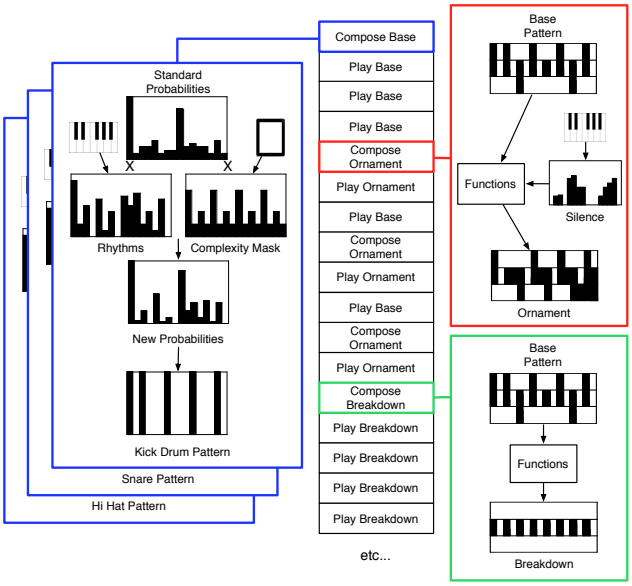

Figure 2: Composition of a Single Chorus

2.4 Musical Robots

Robotic musicians have been used in both research and for performance in many guises, although they are more often than not evaluated on their creative output, as opposed to their skill at building long term relationships. This subtle distinction effects the design of the robot and its composition algorithm. That being said, although distinct in this regard, *Mortimer* still shares much with interactive, anthropomorphic robots such as Georgia Tech's *Haile* [36] and *Shimon* [37].

3. MORTIMER

3.1 The Robot

Pictured in Figure 1, *Mortimer* is a stationary robot with two beater arms and an automated kick drum. His face is a Retina display LCD screen mounted on a servo driven pan tilt system. A tablet device is given to the user to facilitate the social interactions and a speaker fitted in his chest allows for synthesised speech communication in response using the inbuilt AppleTalk functionality of Mac OSX. He takes musical input from a MIDI keyboard.

The sessions are framed a simple social interaction, with *Mortimer* asking questions verbally and receiving responses via the tablet. These include greetings, supportive messages, requests for changes to the style and speed of playing. Phrases are separated into smaller blocks of meaning then recombined for greater variation in dialogue.

3.2 Composition

The composition of drum scores is based upon an underlying statistical model, influenced in realtime by both piano input and explicit performance parameters inputted by the user using the tablet interface. A fuller explanation of the exactitudes of the algorithm and the motivation for the particular approach is available in [23]. A brief description follows.

Each session consists of tracks, which in turn consist of choruses. The structure of each track and the choruses within it is composed at the beginning, with the actual bars not composed until they are to be played as to take into account the most up to date input from the pianist. Each bar in a chorus will be either the base groove, an ornamented version of the base groove, or a breakdown section, with a new base groove generated for each chorus. Figure 2 depicts the composition of a single chorus. Tracks end either upon reaching their precomposed conclusion, after a period of prolonged silence, or may be stopped explicitly by the participant at any time using the tablet device.

When composing the drums, the robot takes into account previous rhythms and a user inputted complexity parameter. The choice and placement of ornaments is aided by the prediction of gaps left by the human and the regularity of ornamentation is also influenced by the complexity parameter. Although the tempo is static throughout each track, the user may change it between tracks using the tablet interface. Further, the individual timing deviations within this overall tempo, or groove, are matched to the user's input in realtime to aid a naturalistic feel.

3.3 Facial Expressions

LaFrance suggests that the causes of facial expressions are far more complicated than the usual "readout" approach that most computer scientists take [20] and the lack of a clear and consistent link between an internal emotional model and facial expressions leads us to approach any such system with caution. However, we have shown in Section 2.2 that facial expressions can be used with satisfactory accuracy and universality to broadly express negative or positive emotions, as well as other more practical social cues such as attention and interestedness.

Following findings in Section 2.1, we used *Mortimer's* face to reflect moments of tension and release in music, as well as moments of concentration. These expressions were also used during musical performance will aid mutual comprehension as the robot enters and exits breakdown sections.

In terms of technical implementation, Fong et al. report that this is often not done well and describe mechanical approaches as often clunky and abrupt [13]. Further, Delaunay et al. suggest the mechanical complexity often comes at a great cost in development and maintenance [8], also, that mechanical android faces are yet to reach levels of humanness necessary to avoid the uncanniness that can lead to anxiety and unease. In fact, this is something to be wary of when attempting any humanoid face, even with smoother animated approaches, such as Brennand and Gordon's Mask Bot [5]. This being said, using the mechanically faced EMYS robot, Ribeiro and Paiva managed to get high classification rates for 5 out of 6 emotions inspired by Ekman's descriptions of distinguishable facial expressions [27].

Given the importance of context and the negative effects of misclassification, we aimed to design facial expressions that are clear and unambiguous in what they attempt to convey and that they occur at appropriate times in concordance with other appropriate actions. As such, we have used a small screen for our robot to allow complex realtime animations that are smooth and easily changeable. We also use a simple, cartoonish face using the basic facets of, but clearly not attempting to replicate, a human face. The most reliable and regularly used facial features are the eyebrows and the mouth, specifically an eyebrow frown and mouth smile so these are the features we have chosen. Further, those who show more positive expressions of affection are more likely to be rated as having good nonverbal skills [17, 26] so we favoured positive facial expressions, such as smiles. Animating the mouth also serves a practical purpose for dialogue.

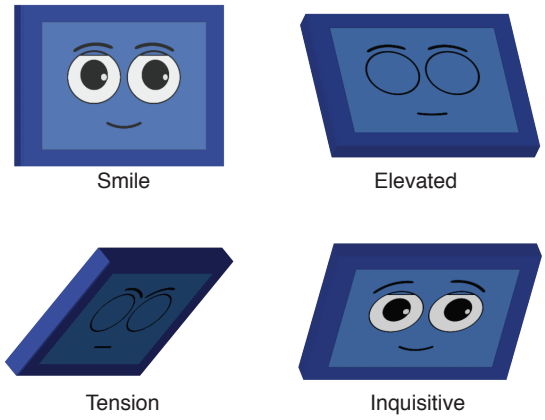

Figure 3: Selected Expressions and Poses

The facial expressions *Mortimer* uses and their triggers are detailed in Table 1 and Figure 3.

3.4 Head Poses

As well as looking to human's use of head poses to influence our design, we are also instructed by previous work in robotics. For instance, Macdorman and Cowley demonstrated that attentive head movements are sufficient to elicit the perception of what they call personhood, a concept that we have shown to have large overlaps with social presence and believability [22]. Head movements have also be used by Weinberg et al. in their musical robot *Shimon* in order to increase its social presence within an ensemble [37]. Breazeal and Fitzpatrick use leaning forwards or recoiling back with the head in order to show willingness to engage or fear, allowing their robot *Kismet* to regulate its personal space [4].

Mounted on a pan/tilt device constructed from two servo motors, *Mortimer's* head has two degrees of freedom. The head poses *Mortimer* uses and their triggers are detailed in Table 1 and Figure 3.

4. METHOD

4.1 Participants

Participants were recruited through by emailing musical lists and placing adverts on musician recruitment websites. There were 10 participants, 5 male and 5 female between the ages of 22 and 54. There was a wide range of self reported skill level (1-5=beginner-expert, min=1, max=5, mean=3.1, SD=1.29). Even though the number of participants is relatively small, a practical constraint of needing skilled participants, as each returned multiple times we conducted 60 sessions in all.

Table 1: Nonverbal Behaviours

What	When	Why
Smile	When you have answered a question with a positive outcome	Smiles used as backchannels
Smile	When positive reassurance is being offered	Agreement
Smile	Following a breakdown	Release
Raised Eyebrows	Before a question	Shows inquisitiveness
Closed Eyes, Eyebrow Frown, Tight Mouth	During breakdown	Shows Tension
Closed Eyes, Eyebrows Raised, Smile	During breakdown	Shows Transportation
Eyebrow Frown	Complicated Ornament	Shows Concentration
Head Nod	When you have answered a question with a positive outcome	Shows Agreement and Affirmation
Head Leans Back	During breakdown	Shows Transportation
Move Head to Side To Side	Complicated Ornament or Breakdown	Shows Intensity
Lean Forward	After question	Demonstrates response expected

4.2 Experimental Setup

Participants were asked to attend 6 identical weekly sessions. After an initial 30 minute session, at each proceeding session they were informed they had to stay for a minimum of 20 minutes, after which they may leave and still fulfil the study requirements. They could also continue to play for anything up to another 25 minutes, leaving at any point. Participants were recompensed £50 upon completion of the study.

During the sessions, participants could freely improvise with *Mortimer*, who facilitated the interactions with a rudimentary artificial personality. The participants were randomly assigned to one of two experimental conditions. For those in Condition A, the robot included all the head poses and facial expressions detailed in Section 3, whilst for those in Condition B, the head and face remained static throughout.

4.3 Measures

As an alternative to the self report measures used by the majority of HRI researchers, we propose that insights into the engagement of a participant and the social presence they experience can be gained from behavioural observation.

We recorded a multitude of quantitative interaction data during the study. Primarily, we measured the time that each participant spent with the robot over the minimum required 20 minutes. This was calculated from when the participant first greets the robot via the tablet interface to when they end the session.

Unlike our previous study, where all but one of the sessions were of equal length [23], the length of sessions ranged from the minimum of 20 minutes right up towards the maximum of 45. As such, the measure of tracks per session used previously is confounded by this variable and not a particularly elucidating one when attempting to investigate the smoothness and immersion of the participant's playing. However, the measure of mean bars per track provides us with a measure of average length of tracks within a session independent of session length.

We examine the number of button stops, as opposed to allowing tracks to finish naturally or due to silence, as our earlier work had indicated that those in the reduced social condition used the tablet to explicitly stop the robot more than the those who experienced *Mortimer* presented as a social actor [23]. They also found that framing the study as a social interaction, and so dividing the session into social and musical interactions, increased both engagement and social presence. Thus, it could be of interest whether the experimental condition would effect the proportion of the session the participant would spend interacting musically or socially.

In order to measure the focus of each participant during the study, we used Soyel and McOwan's face tracking algorithm based upon Seeing Machines faceAPI [31]. Given the robot and participant remain stationary throughout, the algorithm can distinguish whether a participant is looking at the robot, the piano or elsewhere in the room. Given that we had previously found that context, for example, whether the participant is interacting musically or socially, can have a baring on the focus of a participant [23], we took each classification and separated them into playing or not playing.

We also used the NRI-SPV [15], a well validated relationship questionnaire, modified for the use with robots. This was answered by the participants at the midpoint and upon completion of the experiments. The survey provides scores for 9 provisions of the relationship. 7 are positive and 2 are negative and can be amalgamated into overall positive and negative scores.

5. RESULTS

5.1 Quantitative Interaction Data

To measure the effect of experimental condition on the data gleaned from the data logs and its change over time we fitted a random intercept linear mixed effect model for the fixed effects of week, group and the interaction of the two. Results are displayed in Table 2.

We found significant effect of group ($\beta = 49.15$, 95% CI [-378.26 494.81], p=0.047), demonstrating that those in Condition A voluntarily spent more time with the robot. The interaction of group and week was also significant ($\beta = 95.59$, 95% CI [15.80 174.22],p=0.042), demonstrating that the way that those in Condition A changed the amount of time they spent with the robot over the study period differed from

Table 2: Quantitative Interaction Data

Data	Fixed Effect	Estimate β	CI [5% 95%]	p
Session Length	Week	-121.18	[-252.22 3.48]	0.39
	Group	49.15	[-378.27 494.81]	0.047*
	Week.Group	95.59	[15.80 174.22]	0.042*
Bars Per Track	Week	5.261	[3.56 6.83]	0.0005***
	Group	4.91	[-7.13 17.51]	0.5832
	Week.Group	-4.11	[-7.19 -0.85]	0.0005***
Button Stops	Week	-0.01	[-0.02 0.01]	0.4073
	Group	-0.06	[-0.12 0.01]	0.2239
	Week.Group	0.02	[-0.01 0.05]	0.2569
Inter -ruptions	Week	-0.21	[-0.28 -0.14]	0.0005***
	Group	-1.12	[-1.54 -0.71]	0.0045***
	Week.Group	0.23	[0.11 0.36]	0.0005***
Time Playing (%)	Week	-3.552	[0.20 1.18]	0.0305*
	Group	0.700	[-9.52 2.64]	0.4293
	Week.Group	-0.280	[-1.27 0.71]	0.1384

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.

that of those in Condition B. For the group as a whole the mean number of bars per track increased over time, meaning longer tracks and less interruptions during playing (β = 5.26, 95% CI [3.56 6.83],p=0.0045). Interestingly, the rate of increase was greater for those in Condition B (β = −4.11, 95% CI [-7.19 -0.85],p=0.0005). With respect to the proportion of time spent session spent playing piano, we found significant effect of week (β = −3.552, 95% CI [0.20 1.18], p=0.0305), demonstrating that regardless of the experimental condition, all participants spent less time playing with the robot as the study progressed.

For interruptions, we found significant effect of group (β = −1.12, 95% CI [-1.54 -0.71], p=0.004), demonstrating those in Condition A interrupted the robot less over the whole study. There is also a significant decrease in number of interruptions across the trials (β = −0.21, 95% CI [-0.28 -0.14], p=0.005). Further, the rate of reduction of interruptions over the trial was significant higher for those in Condition B (β = 0.23, 95% CI [0.11 0.36], p=0.005). However, the proportion of button stops presented no significant effects for week, the experimental condition or the interaction between the two.

5.2 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. Results are displayed in Table 3.

We found significant effect of group (β = −28.75, 95% CI [-45.10 -12.55], p=0.0350) and the interaction between week and group (β = −4.109, 95% CI [-7.33 -0.80], p=0.0165) for proportion of time spent looking at the robot when playing. There was was also an effect for group for looking at the robot when not playing (β = −26.28, 95% CI [-38.30 -14.16], p=0.0170). This demonstrates that over the course of the whole study, those in Condition A spent less time looking at the robot when playing and when not playing. Also, that way the two groups differed in way this the former category changed as the study progressed.

The only other significant effect was for group for looking at the piano when not playing, (β = 2.933, 95% CI [-1.54 -0.71], p=0.0045), with those in Condition A looking at the piano more when not playing.

Table 3: Automatic Video Analysis

Condition	Fixed Effect	Estimate β	CI [5% 95%]	p
Robot, playing	Week	-1.367	[-2.95 0.34]	0.1984
	Group	-28.75	[-45.10 -12.55]	0.0350*
	Week.Group	-4.109	[-7.33 -0.80]	0.0165*
Robot, not playing	Week	-0.7517	[-3.24 1.51]	0.5787
	Group	-26.28	[-38.30 -14.16]	0.0170*
	Week.Group	1.234	[-3.34 5.74]	0.0520
Piano, playing	Week	0.1283	[-0.87 1.14]	0.8436
	Group	-0.683	[-7.54 6.65]	0.8906
	Week.Group	1.028	[-1.01 3.03]	0.8736
Piano, not playing	Week	1.259	[-1.72 7.58]	0.3473
	Group	2.933	[-1.54 -0.71]	0.0045***
	Week.Group	-0.744	[-2.44 1.06]	0.1045
Elsewhere, playing	Week	-0.1974	[-1.35 0.95]	0.7786
	Group	12.31	[2.69 21.89]	0.099
	Week.Group	-0.0794	[-2.45 2.21]	0.3228
Elsewhere, not playing	Week	-0.973	[-2.42 0.50]	0.2819
	Group	10.19	[2.30 18.16]	0.0980
	Week.Group	-3.543	[-6.33 -0.79]	0.0535

Random Intercept Linear Mixed Effect Model for participant focus during session (%). 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.

5.3 NRI-SPV

Analysis of results did not find any significant factors between groups or over time for positive (POS) or negative (NEG) scores or for any of the individual relationship provisions (AFF,ALL,WOR,CON,COM,ANT,DIS,AID,NUR) [15].

6. DISCUSSION

Investigating changes in a range of quantitative interaction measures and participant focus of attention between experimental conditions and over time, we found several results of interest. Some show an effect of the control condition, demonstrating the difference that introducing head poses and facial expressions can make, whilst others displayed a change as the the study progressed, allowing us to draw more general conclusions about the use of music as a platform for developing human robot relationships.

Primarily, and most crucially, we found that those in Condition A spent more time voluntarily playing with the robot over the course of the study. Demonstrated clearly by Figure 4, they also actually increased the time they spent as the study continued. Given Bickmore's definition of engagement

as the degree of and regularity users choose to have with the robot [2], we confidently take this as a sign of the positive effect of including the nonverbal behaviour. Further, its inclusion has not only avoided the novelty effect but reversed it, with users seemingly becoming more engaged with the robot over time.

Beyond this, we examined the way the participants used the system in both musical and social contexts during the sessions. With regards to the latter, we found that participants used the tablet to interrupt the robot less in the Condition A overall, implying a greater social presence with a robot utilising nonverbal behaviour. Participants were less willing to curtail the talking and move on as they would if using a computer program or instrument. Moreover, for both groups this decreased over time, suggesting that social presence grew as the trial progressed. These positive results were not mirrored for the musical equivalent of an interruption, the button stop, where we found no significant differences.

As musicians often use head movements as cues during performance, especially during improvisation, we predicted nonverbal behaviour would aid the fluency of the music played, reducing frustration and aiding long engagement. However, we found longer tracks within the session for the group as a whole as time progressed, showing more engaged, uninterrupted playing. This suggests learning over time was a more important factor than the inclusion of nonverbal behaviour. Further, the finding that, regardless of group, participants spent less time playing and more time interacting socially as time passed shows that although music is the main focus of the sessions, users increasingly explored *Mortimer's* social faculties as well.

Gaze can have a large effect on the dynamics of dyadic social interaction. Mutual gaze is thought to be revelatory about the interpersonal relationship between participants, for example, as a display of immediacy [1]. This would suggest reduced social presence in Condition A and run counter to results from the quantitative interactional data. However, Gratier does claim that mutual gaze serves less of a purpose for grounding musical interactions than it does in conversation [16] so it may only be the findings of reduced focus towards the robot whilst not playing that cause concern. This being said, there is also evidence to suggest that mututal gaze occurs less as a relationship develops in social situations [29], so it may be that the reduced focus is in fact a signifier of a closer relationship.

We suggest the indeterminate results from the NRI-SPV demonstrate that in our case surveys lack the required sensitivity to examine human-robot relationships as it failed to find differences between the groups or over time when the behavioural metrics showed clear effects. This strengthens our resolve that the use behavioural metrics is the favourable approach for our interests.

7. CONCLUSION

Taking a novel methodological approach of automated behavioural metrics, in place of the common practice of self report questionnaires, we uncovered several results which lead us to believe improvised musical interaction is a solid grounding for building long term, sustainable and positive relationships between humans and robots. Our hypothesis that this is aided by the the inclusion of appropriate head poses and facial expressions in both musical and social contexts is supported by our quantitative interaction data. However, this interpretation is somewhat less categorical in relation to participant gaze.

Future work will focus on the inclusion of more social modalities alongside the musical improvisation to see if improvements continue.

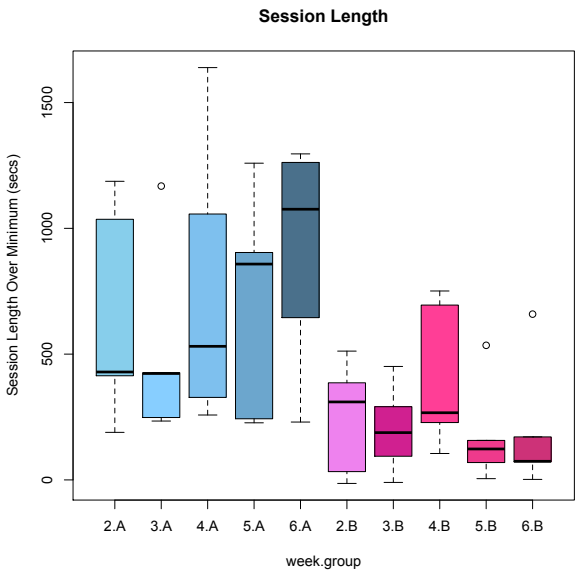

Figure 4: Session Length

8. REFERENCES

[1] A. Abele. Functions of gaze in social interaction: Communication and monitoring. *Journal of Nonverbal Behavior*, 10(2):83–101, 1986.

[2] T. Bickmore, D. Schulman, and L. Yin. Maintaining engagement in long-term interventions with relational agents. *Applied Artificial Intelligence*, 24(6):648–666, 2010.

[3] 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*, 12(5):456–480, Oct. 2003.

[4] C. Breazeal and P. Fitzpatrick. That Certain Look: Social Amplification of Animate Vision . In *Proc AAAI Fall Symposium*, pages 3–5, North Falmouth, MA, 2000.

[5] P. Brennand and C. Gordon. Automatic Face Replacement for Humanoid Robot with 3D Face Shaped Display. In *Proc. 2012 Social Robotics Conf.*, pages 469–474, Chengdu, 2012.

[6] P. Briñol and R. E. Petty. Overt head movements and persuasion: A self-validation analysis. *Journal of Personality and Social Psychology*, 84(6):1123, June 2003.

[7] N. Chovil. Facing others: A social communicative perspective of facial displays. In J. M. Fernández-Dols and J. A. Russell, editors, *The Psychology of Facial Expressions*. Cambrige University Press, Cambridge, 1997.

[8] F. Delaunay, J. de Greeff, and T. Belpaeme. Towards retro-projected robot faces: An alternative to mechatronic and android faces. In *Proc 2009 Robot and Human Interactive Communication Conf.*, pages 306–311, Toyama, 2009.

[9] P. Ekman and M. O'Sullivan. Facial Expression: Methods, Means and Moues. In R. S. Feldman and B. Rime, editors, *Fundamentals of Nonverbal Behaviour*. Cambridge University Press, Cambridge, 1991.

[10] 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.*, pages 39–48, Barcelona, 2010.

[11] J. M. Fernández-Dols. Spontaneous facial behaviour during intense emotional episodes: Artistic truth and optical truth. In J. M. Fernández-Dols and J. A. Russell, editors, *The Psychology of Facial Expressions*. Cambride University Press, Cambridge, 1997.

[12] J. M. Fernández-Dols and J. M. Carrol. Is the meaning perceived in facial expression independent of its context? In J. M. Fernández-Dols and J. A. Russell, editors, *The Psychology of Facial Expression*. Cambridge University Press, 1997.

[13] T. Fong, I. Nourbakhsh, and K. Dautenhahn. A survey of socially interactive robots. *Robotics and Autonomous Systems*, 42:143–166, 2003.

[14] A. Fridlund and J. A. Russell. The functions of facial expressions: What's in a face? In V. Manusov and M. L. Patterson, editors, *The SAGE Handbook of Nonverbal Communication*. SAGE publications, Thousand Oaks, CA, 2006.

[15] W. Furman and D. Buhrmester. Children's Perceptions of the Personal Relationships in Their Social Networks. *Developmental psychology*, 21(6):1016–1024, 1985.

[16] M. Gratier. Grounding in musical interaction: Evidence from jazz performances. *Musicae Scientiae*, 12(1 suppl):71–110, 2008.

[17] L. K. Guerrero and K. Floyd. *Nonverbal Communication in Close Relationships*. Lawrence Erlbaum Associates, Mahwah, NJ, 2006.

[18] Y. Iwano, S. Kageyama, E. Morikawa, S. Nakazato, and K. Shirai. Analysis of head movements and its role in spoken dialogue. In *Proc 1996 Spoken Language Conf.*, pages 2167–2170, Tokyo, 1996.

[19] R. E. Kraut and R. E. Johnston. Social and emotional messages of smiling. *Journal of Personality and Social Psychology*, 37(9):1539–1553, 1979.

[20] M. LaFrance. What's in a robot's smile? The many meanings of positive facial display. In R. Aylett and L. Canamero, editors, *Animating Expressive Characters for Social Interaction*. John Benjamins Publishing, Amsterdam, 2008.

[21] 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*, pages 695–704, 2012.

[22] K. Macdorman and S. Cowley. Long-term relationships as a benchmark for robot personhood. In *Proc. 2006 Robot and Human Interactive Communication Conf.*, pages 378–383, Hertfordshire, 2006.

[23] L. McCallum and P. W. McOwan. Shut Up and Play: A Musical Approach to Engagement and Social Presence in Human Robot Interaction. In *Proc 2014 Robot and Human Interactive Communication Conf.*, pages 949–954, Edinburgh, 2014.

[24] E. Z. McClave. Linguistic functions of head movements in the context of speech. *Journal of Pragmatics*, 32(7):855–878, 2000.

[25] M. Michalowski, S. Sabanovic, and R. Simmons. A spatial model of engagement for a social robot. In *Proc 2006 Advanced Motion Control Workshop*, pages 762–767, Istanbul, 2006.

[26] P. Noller. Nonverbal Communication in Close Relationships. In M. L. Patterson and V. Manusov, editors, *The SAGE Handbook of Nonverbal Communication*. SAGE Publications, Thousand Oaks, CA, 2006.

[27] T. Ribeiro and A. Paiva. The illusion of robotic life: Principles and practices of animation for robots. In *Proc 2012 Human Robot Interaction Conf.*, pages 383–390, New York, NY, 2012.

[28] W. E. Rinn. Neuropsychology of Facial Expressions. In R. S. Feldman and B. Rime, editors, *Fundamentals of Nonverbal Behaviour*. Cambridge University Press, Cambridge, 1991.

[29] D. Schulman. *Embodied Agents for Long-Term Discourse*. PhD thesis, Northeastern University, Boston, MA, Jan. 2013.

[30] C. A. Smith and J. M. Fernández-Dols. Componential Approach. In J. M. Fernández-Dols and J. A. Russell, editors, *The Psychology of Facial Expression*, page 400. Cambridge University Press, Mar. 1997.

[31] H. Soyel and P. W. McOwan. Towards an affect sensitive interactive companion. *Computers & Electrical Engineering*, 39(4):1312–1319, May 2013.

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

[33] W. Thompson and P. Graham. Seeing music performance: Visual influences on perception and experience. *Semiotica*, (156):203–227, 2005.

[34] L. Tickle-Degnen. Nonverbal behaviour and its functions in the ecosystem of rapport. In V. Manusov and M. L. Patterson, editors, *The SAGE Handbook of Nonverbal Communiation*. SAGE Publications, Thousand Oaks, CA, 2006.

[35] B. W. Vines, C. L. Krumhansl, M. M. Wanderley, and D. J. Levitin. Cross-modal interactions in the perception of musical performance. *Cognition*, 101(1):80–113, 2006.

[36] G. Weinberg and S. Driscoll. Robot-human interaction with an anthropomorphic percussionist. In *Proc 2006 Human Factors in Computing Systems Conf*, pages 1229–1232, Montreal, QU, 2006.

[37] G. Weinberg, A. Raman, and T. Mallikarjuna. Interactive jamming with Shimon: a social robotic musician. In *Proc. 2009 Human Robot Interaction Conf.*, pages 233–234, San Diego, CA, 2009.