Non-Linguistic Utterances Should be Used Alongside Language, Rather than on their Own or as a Replacement

Robin Read
Plymouth University
Centre for Robotics and Neural Systems
Drake Circus, Plymouth, PL4 8AA, UK
robin.read@plymouth.ac.uk

Tony Belpaeme
Plymouth University
Centre for Robotics and Neural Systems
Drake Circus, Plymouth, PL4 8AA, UK
tony.belpaeme@plymouth.ac.uk

ABSTRACT

This paper presents the results of a small experiment aimed at determining whether people are comfortable with a social robot that uses robotic Non-Linguistic Utterances alongside Natural Language, rather than as a replacement. The results suggest that while people have the most preference for a robot that uses only natural language, a robot that combines NLUs and natural language is seen as more preferable than a robot that only employes NLUs. This suggests that there is potential for NLUs to be used in combination with natural language. In light of this, potential utilities and motivations for using NLUs in such a manner are outlined.

Categories and Subject Descriptors

H.5.2 [**User Interfaces**]: Auditory (non-speech) feedback; H.5.5 [**Sound and Music Computing**]: Systems; I.2.9 [**Robotics**]: Operator interfaces

General Terms

Design, Human Factors, Experimentation, Theory

Keywords

Non-Linguistic Utterances; Social HRI; Natural Language

1. INTRODUCTION

Non-Linguistic Utterances (NLUs) are robotic sounds made by synthetic social agents, rather than utterances that are designed to resemble natural speech, such as a artificial languages [1] or gibberish speech [3]. While NLUs have been used to great effect within the world on animation and popular culture/media as a *proto-language*, very little is understood about how NLUs can be applied to real world Human-Robot Interaction (HRI). In previous work [2] we have shown that situational context is an important aspect of social HRI that directs how NLUs are interpreted by people. The work outlined in this paper makes an initial investigation into how people respond to a robot (an Aldebaran Nao) that uses NLUs alongside natural language (NL) as NL is another rich source of mood and context within social HRI. Assuming that people indeed respond positively to a robot that uses both NLUs and NL in combination, there are a number of potential utilities that NLUs can then provide for social HRI (as we see in section 4). This experiment initially probes the validity of this assumption.

HRI'14, March 3–6, 2014, Bielefeld, Germany.
ACM 978-1-4503-2658-2/14/03.
http://dx.doi.org/10.1145/2559636.2559836 .

2. EXPERIMENTAL SETUP

The experiment tests the following hypotheses: H_1: A robot that uses only NL will be rated as more preferable then a robot that uses NLUs alongside language. H_2: A robot using only NLUs will be rated as less preferable than a robot using NLUs alongside language. H_3: The rating of preference will be influenced by how NLUs and NL are combined.

To test these hypothesis, four videos were created each showing a robot playing a guessing game with a human, with the type of utterance (NL, NLUs, or a combination of the two) being varied across the four videos. The game was based upon the "Cups and Balls" game (figure 1). In this game, an object is hidden under one of three cups. The cups are then shuffled and presented to the player (the robot), and the player needs to be guess which cup the object is under. This scenario was chosen as it facilitates the robot making a variety of different vocalisations throughout the scenario, ranging from linguistic comments, conversational fillers, and reactive and expressive vocalisations. For the purpose of this experiment the unfolding of the game was scripted such that the robot's physical behaviour was always the same, and that the robot always made the same (incorrect) guess. The four video conditions[1] were as follows: V_1: the robot using only NL. V_2: the robot using only NLUs. V_3: the robot using a combination of NLUs and NL. V_4: the robot using an *inverted* combination of NLUs and NL.

This experiment was conducted using the online crowd sourcing facility, *CrowdFlower*. Subjects were rewarded $0.3 USD for their participation, and subjects were limited to the United States alone. They were asked to provide their age and gender, and then presented with the four videos in a random order. After viewing all of the videos, subjects were asked to provide a preference rating for each video individually. The rating was captured using a 9 point Likert scale (1 = Least Preferred, 9 = Most Preferred). Finally, subjects were asked whether they were familiar with the Nao robot.

[1]The videos can be viewed here: `http://www.youtube.com/playlist?list=PLJwvAKNW4DupVHtrTaHSTxnOWYHhobG49`

(a) Ball being presented.

(b) Cups being shuffled.

(c) Robot choosing a cup.

(d) Ball location revealed.

Figure 1: Images depicting the game unfolding.

3. RESULTS

In total, 270 subjects responded. 89 males (mean age = 36.74, std = 10.8) and 181 females (mean age = 37.74, std = 11.0). 166 subjects self reported that they had seen the robot before (57 males, 109 females).

A 3-way repeated measures ANOVA (4x2x2 design) was performed using the video condition as the within subjects factor, and subject gender and robot familiarity as the between subjects factors. It was found that there were main effects due to the video condition ($F(3, 798) = 67.4$, $MSE = 145.264$, $p < 0.001$) and a subject's familiarity with the Nao robot ($F(1, 266) = 6.959$, $MSE = 53.002$, $p = 0.009$). These results are shown in figure 2.

With respect to the video condition main effect, post-hoc tests showed that V_1 had the highest rating (mean = 6.845, 95% CI = [6.645 7.063]) and was significantly higher than all the other conditions ($p < 0.001$). V_2 was found to have the lowest rating (mean = 4.969, 95% CI = [4.659 5.319]) and was significantly different than all the other conditions ($p < 0.001$). V_4 had the second highest rating (mean = 6.470, 95% CI = [6.256 6.684]) and V_3 the third (mean = 6.260, 95% CI = [6.044 6.475]), with no significant difference found between these two conditions. With respect to the main effect due to robot familiarity, post-hoc tests revealed that subjects who had seen the robot before provided higher ratings (mean = 6.387, 95% CI = [6.165 6.609]) than subjects who had not seen the robot before (mean = 5.899, 95% CI = [5.611 6.188]).

4. DISCUSSION & CONCLUSIONS

The results provide support for both hypotheses H_1 and H_2 in that subjects showed the highest preference for the condition where the robot only used NL (V_1) and showed the lowest preference for the robot that used only NLUs (V_2). H_3 was not supported as there was no significant difference found between the preference ratings for the two videos showing the robot using a combination of NLUs and NL.

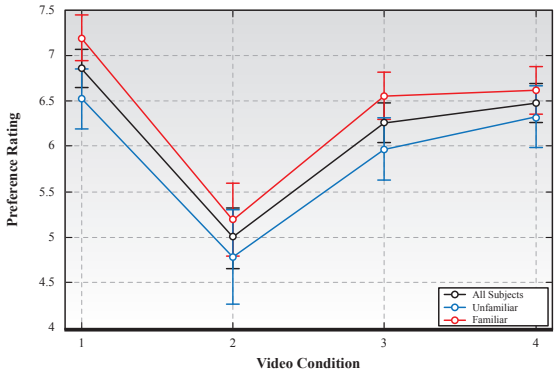

Figure 2: Plot of the Preference Ratings across the four video conditions.

The notion that there is indeed potential for NLUs to be used alongside language opens a number of potential uses of NLUs during HRI. For example, a toy idea could be to use NLUs to indicate a robot swearing, where the robot is able to suggest the mood of an utterance without actually being explicitly offensive. Another potentially fruitful utility is to use NLUs as a means to provide conversational fillers or for back-channelling, both of which are essential cues during social interaction. In this context, NLUs can provide a supportive role to Natural Language Processing (NLP) during linguistic exchanges. Furthermore, if used in this manner, NLUs may also be used as a way to *camouflage* situations in which NLP may perform sub-optimally (e.g. if the processing of user input is taking too long, or if the confidence of speech recognition falls below an acceptable threshold). In this utility, NLUs could be used with the aim of mitigating any damage to to an interaction should NLP fail to perform at a desired level. It should be noted that the experimental results and discussion presented here represents research ideas that are currently very much in their infancy and require further, more detailed scientific exploration. Addressing this is the aim of our future work.

5. ACKNOWLEDGMENTS

This work is (partially) funded by the EU FP7 ALIZ-E project (grant 248116).

6. REFERENCES

[1] O. Mubin, C. Bartneck, L. Leijs, H. Hooft van Huysduynen, J. Hu, and J. Muelver. Improving Speech Recognition with the Robot Interaction Language. *Disruptive Science and Technology*, 1(2):79–88, Dec. 2012.

[2] R. Read and T. Belpaeme. Situational Context Directs How People Affectively Interpret Robotic Non-Linguistic Utterances. In *Proceedings of the 9th International Conference on Human-Robot Interaction (HRI'14)*, Bielefeld, Germany, 2014. ACM/IEEE.

[3] S. Yilmazyildiz, D. Henderickx, B. Vanderborght, W. Verhelst, E. Soetens, and D. Lefeber. Multi-modal emotion expression for affective human-robot interaction. In *Proceedings of the Workshop on Affective Social Speech Signals (WASSS 2013)*, Grenoble, France, 2013.