Visual Priming to Improve Keyword Detection in Free Speech Dialogue

Chao Qu[a] **Willem-Paul Brinkman**[a] **Pascal Wiggers**[a] **Ingrid Heynderickx**[a,b]
a. Delft University of Technology, Mekelweg 4, 2628 CD Delft, the Netherlands
b. Philips Research Laboratories, High Tech Campus 34, 5656 AE Eindhoven, the Netherlands
{C.Qu, W.P.Brinkman, P.Wiggers}@tudelft.nl Ingrid.Heynderickx@philips.com

ABSTRACT

Motivation – Talking out loud with synthetic characters in a virtual world is currently considered as a treatment for social phobic patients. The use of keyword detection, instead of full speech recognition will make the system more robust. Important therefore is the need to increase the chance that users use specific keywords during their conversation.

Research approach – A two by two experiment, in which participants (n = 20) were asked to answer a number of open questions. Prior to the session participants watched priming videos or unrelated videos. Furthermore, during the session they could see priming pictures or unrelated pictures on a whiteboard behind the person who asked the questions.

Findings/Design – Initial results suggest that participants more often mention specific keywords in their answers when they see priming pictures or videos instead of unrelated pictures or videos.

Research limitations/Implications – If visual priming in the background can increase the chance that people use specific keywords in their discussion with a dialogue partner, it might be possible to create dialogues in a virtual environment which users perceive as natural.

Take away message – Visual priming might be able to steer people's answers in a dialogue.

Keywords

priming, speech recognition, dialogue, keyword detection, social phobia

INTRODUCTION

Social phobia is an anxiety disorder, where people have a strong fear of social situations, such as talking in public, entering a room with other people, ordering food in a restaurant etc. It is one of the most often occurring anxiety disorders, estimated to affect 13.3% of the US population (Kessler, et al., 1994). Virtual Reality Exposure Therapy (VRET) currently receives considerable research attention as a treatment for anxiety disorders, and a recent meta-analysis (Powers and Emmelkamp, 2008) indicates that VRET is as effective as exposure in vivo. VRET for social phobia may include, for example, a social situation in which a patient has to talk with virtual characters in a virtual word. Essential in this treatment is that the communication with the avatars is experienced by the patient as natural. The behaviour, such as the speech, of the avatars is often triggered by a therapist, who has the dual task of delivering the treatment and controlling the simulation environment. Semi-automating the control of the simulation environment would therefore reduce the therapist's workload, and potentially open the option for treating multiple patients simultaneously. In this context, speech recognition technology might enable appropriate avatar behaviour, such as giving an appropriate reply on something the patient says. Computer recognition of completely free speech, however, seems currently too ambitious to realise (Jurafsky, Martin, 2008). An alternative therefore, is to search for a set of predefined keywords in patients' utterances and fall back on general, predefined scripts if these keywords are not found. Since the goal of the system is to provide the experience of a conversation rather than specific information exchange, this approach is expected to be sufficient. However, as in principle the patient may say anything, the flow of the dialogue improves if the patient actually uses the predefined keywords in his discussion with the avatars. Thus, the main motivation of the work presented here is to search for a mechanism that enhances the chance that patients use specific keywords in their answers, and that gives them the feeling of not being limited in their verbal expression and of having a natural conversation with an avatar.

Priming theory

In VRET, speech recognition is used to evoke the anxiety of having an actual social interaction with another person and not to capture the meaning of what a person tries to communicate. It seems therefore less relevant that a person provides an unbiased opinion. Priming people to give a specific verbal response can be an appropriate mechanism to bias them in favour of giving responses that include specific keywords. Priming can be seen as the incidental activation of a person's knowledge structure, which can lead to specific behaviours and attitudes (Bargh, Chen, Burrows, 1996).

An extensive amount of work has been conducted on priming, however to our knowledge not in the context of supporting question-answer dialogues in virtual reality, or indeed in reality. Therefore, before studying priming in virtual reality, this study first addresses the fundamental question whether priming pictures or videos can increase the chance that people use specific keywords in their answers. In this paper we therefore report on a small experiment to examine this research question, which is a pre-condition before extending the study into a virtual reality environment.

METHOD

In the experiment, participants sat in a room with a person who asked them to answer seven questions on four general topics (democracy, dogs, France, and penguins). The experiment had a two by two within-subjects design, with two independent variables: 1) priming video vs. unrelated video, and 2) priming pictures vs. unrelated pictures. Prior to each session participants saw two videos in a room next door, each with a length of around one and a half minute. In the condition where they saw two priming videos, the videos featured relevant objects or events related to questions they got later on. For example, for the questions on France, one priming video showed the liberation of Paris in the Second World War. This was expected to trigger the answer to a question about historical events that happened in Paris. The unrelated videos were selected with the intent not to prime specific keywords.

Figure 1: Room with pictures on whiteboard

In the condition with priming pictures, seven related pictures were placed on a whiteboard (Figure 1) including also seven unrelated pictures as a diversion. In the condition with unrelated pictures, all 14 pictures on the whiteboard were unrelated. The whiteboard was positioned behind the person that asked the questions, making this a double-blind experiment as the person could not see the pictures on the whiteboard. Each time the participant went to the other room to watch the videos, the pictures on the whiteboard were changed by a second experimenter. The order, in which the four conditions were presented, was counterbalanced, while the topic in each condition was assigned randomly. Furthermore, participants were not informed about the priming aspect of the videos or the pictures. After the experiment the participants were asked to complete a questionnaire. The twenty people that participated in the experiment were all students or staff members of the university.

RESULTS

The number of questions, in which a participant mentioned a specific keyword in his or her answer was counted, resulting in a keyword hit index ranging from zero to seven. To examine the effect of priming with videos or pictures, an MANOVA was conducted with hit index as dependent variable. The results showed a significant main effect for priming with pictures ($F(1,19) = 9.00$, $p. = 0.007$), and also for priming with videos ($F(1,19) = 14.34$, $p. = 0.001$). No significant two-way interaction between pictures and videos ($F(1,19) = 3.59$, $p. = 0.074$) was found.

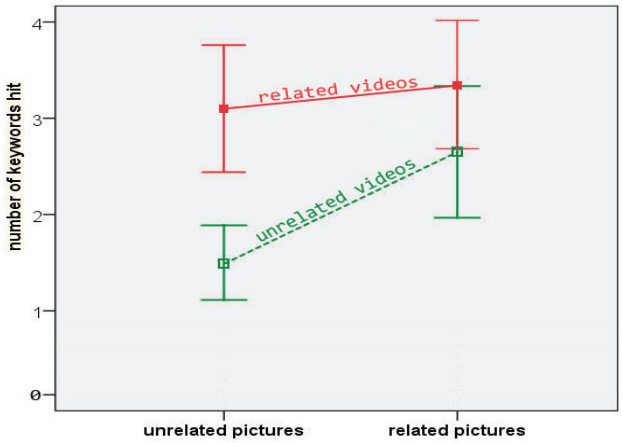

Figure 2: Average keywords hits in each condition

Examining Figure 2 shows that on average more keywords were mentioned in the condition with the priming pictures or videos than in the condition with unrelated pictures or videos. Furthermore, the questionnaire indicated that all participants had noticed that the videos and pictures were related to the topic during the experiment. They liked the priming and most of them thought it would help their conversation (i.e. 12 participants felt the videos helpful and 15 out of 20 felt the pictures helpful).

DISCUSSION AND FURTHER RESEARCH

Our findings already suggest that priming with videos or pictures can result in answers with a specific keyword. The next step of research is to re-run the experiment in a virtual environment; initially with non-phobic participants, and later on, if possible, with people with social phobia.

REFERENCES

Bargh, J.A., Chen, M., and Burrows, L. (1996). Automaticity of social behaviour: direct effects of trait construct and stereotype activation on action. *Journal of personality and social psychology,71*, 230 – 244.

Kessler, R.C., McGonagle, K.A., Zhao, S., Nelsen, C.B., Hughes, M., Eshleman, S., Wittchen, H.U., and Kendler, K.S. (1994). Lifetime and 12-month prevalence of DSM-III-R psychiatric disorders in the United States. Results from the National Comorbidity Survey, *Arch Gen Psychiatry, 51*, 8-19.

Powers, M.B., and Emmelkamp, P.M., (2008). Virtual reality exposure therapy for anxiety disorders: A meta-analysis, *Journal of Anxiety Disorders, 22*, 561-9.

Jurafsky, D., and Martin, J. H. (2008) Speech and Language Processing, second edition, Pearson