Comparing Virtual Patients with synthesized and natural speech

Alexandre HEITZ[a], Andreas DÜNSER[a], Philippa SEATON[b], Lesley SEATON[b], Arindam BASU[c]

[a] *The HIT Lab NZ, University of Canterbury, NZ*
[b] *School of Nursing and Human Services, Christchurch Polytechnic Institute of Technology, NZ*
[c] *Health Sciences Centre, University of Canterbury, NZ*

Abstract. Virtual Patient (VP) simulations are often designed to use pre-recorded speech in order to provide more realism and immersion. However, using actors for recording these utterances has certain downsides. It can add to the cost during implementation, can take considerable time especially when a large number of VPs have to be created, and is not very flexible for example when sentences or words have to be added frequently. This study aims to explore the use of synthesized speech as an alternative to pre-recorded speech for VPs. Two medical scenarios have been prepared for this study, and both have been implemented using a VP with natural language or with synthesized speech. In a pilot study we explored students' retention rates of the symptoms reported by the VP under both conditions to investigate whether synthesized speech can serve as a good enough alternative.

Introduction

VPs are embodied agents used in computer simulations to represent patients. They allow real time interaction with a patient that manifests symptoms, gives immediate feedback, and can be subjected to treatments depending on the aim of the medical simulation. Previous research shows that VPs can effectively be implemented as an alternative to Simulated Patients (SPs), who are trained actors, in various types of scenarios [1, 2].

Using VPs in healthcare education has the potential to improve clinical practice of trainees. However, implementing VPs is a time consuming process, especially for medical history taking.

VPs' speech interaction is commonly implemented through iterative cycles. Initially simple interactions scripts are implemented including a limited number of answers and questions to which they can respond. Over time these scripts and the vocabulary can become more elaborate by integrating the results of actual user interaction [3] When a system uses speech recorded by actors, this can create extra costs for each iteration and can be very time consuming, as the same actor has to be used in order to ensure consistency. In this study we explored the influence of the type of speech used in a VP system on students' ability to retain information provided by the patients.

1. Method

For this pilot study a sample of 16 nursing students was recruited (Bachelor of Nursing). Two medical scenarios were implemented, each with a VP using synthesized speech and with a VP using an actor's voice. The MARY Text-to-Speech System [4] was used to generate the VPs' synthesized speech. Participants experienced both scenarios, one with natural speech and one with synthesized speech. Scenario and speech type were counterbalanced and participants randomly assigned. The study was conducted online. Participants were presented with videos of the VPs describing their condition. After this the participants completed a questionnaire about the VP's realism and made a written report on diagnosis critical information. Retention rate was assessed by the percentage of critical key items the participants did report, such as shortness of breath (SOB), exacerbation, or the medication the patients were taking.

2. Results and Conclusion

The students' retention rates were scored and compared between the two speech conditions. Students could retain 42% (SD = 23.12) for recorded speech and 38% (SD = 18.38) for synthesized speech. While this indicates that both speech conditions are comparable, results of statistical analyses are inconclusive and show no evidence for the two speech conditions to be either statistically different (paired t-test; $t = 0.41$, $df = 15$, $p = .69$) or statistically equivalent. Equivalence was tested using a confidence interval approach [5] (90% CI for score difference (-10.05,16.67) is not contained within the equivalency interval (±4.18; $\pm10\%$ of mean for natural speech). This can be attributed to to the small number of participants used in this pilot study.

3. Conclusion

This pilot study is part of a larger project and was aimed at informing us whether using synthesized speech when implementing VPs for medical training is a valid alternative to recorded speech. Using synthesized speech allows more flexibility during implementation and could reduce costs as well as implementation time. In a follow up study with a larger sample we hope to get more conclusive results and determine whether synthesized speech is a valid option to be used for VPs.

References

[1] Conradi, E., Kavia, S., Burden, D., Rice, A., Woodham, L., Beaumont, C., Savin-Baden, M., et al. (2009). Virtual patients in a virtual world: Training paramedic students for practice. Medical Teacher, 31(8), 713-720. Taylor & Francis Ltd.

[2] Tan, Z. S., Mulhausen, P. L., Smith, S. R., & Ruiz, J. G. (2010). Virtual Patients in Geriatric Education. Gerontology & Geriatrics Education, 31(2), 163-173. Routledge. doi:10.1080/02701961003795813

[3] Rossen, B., Cendan, J., & Lok, B. (2010). Using Virtual Humans to Bootstrap the Creation of Other Virtual Humans, 392-398.

[4] The MARY Text-to-Speech System. Retrieved February 28, 2012, from http://mary.dfki.de/

[5] Rogers, J. L., Howard, K. I., & Vessey, J. T. (1993). Using significance tests to evaluate equivalence between two experimental groups. Psychological Bulletin, 113(3), 553-565.