The Accuracy of an Audio Interface Designed for Value Elicitation

Elvira Voorneveld
E.Voorneveld@student.tudelft.nl

Supervisor: Pei-Yu Chen
Responsible Professor: Catholijn Jonker

0. Background

Behaviour Support Applications (BSA) are used to provide personalized support to its users.

The brain behind a BSA is an **intelligent agent** that must understand the user's preferences, values, and context in order to function.

User models are created and then used to capture the relationship between users' desired behaviours and their values [1,2]

However, the preferences of a user may change over time, requiring **real-time** updates to the user model.

Collecting information through a **conversational** agent is a new but effective method [4]. Therefore, this audio interface is built using a conversational style.

References

[1] Kließ, M. S., Stoelinga, M., and van Riemsdijk, M. B. (2019). From Good Intentions to Behaviour Change: Probabilistic Feature Diagrams for Behaviour Support Agents. *PRIMA 2019: Principles and Practice of Multi-Agent Systems.*
[2] Tielman, M. L., Jonker, C. M., and van Riemsdijk, M. B. (2018). What Should I Do? Deriving Norms from Actions, Values and Context. *Modelling and Reasoning in Context.*
[3] Berka, J., Jonker, C. M., Mikovec, Z., van Riemsdijk, M. B., and Tielman, M. L. (2022). Misalignment in Semantic User Model Elicitation via Conversational Agents: A Case Study in Navigation Support for Visually Impaired People. *International Journal of Human-Computer Interaction.*
[4] Ponathil, A., Ozkan, F., Welch, B., Bertrand, J., and Madathil, K. C. (2020). Family health history collected by virtual conversational agents: An empirical study to investigate the efficacy of this approach. *Journal of Genetic Counselling*, 29(6).

1. Research Question

What is the efficacy of an **audio interface** that **elicits** values-related information using **isolated** questions?

2. Methodology

1. Create interface
   - Use a Text-to-Speech (TTS) system
   - Follow a dialogue to imitate the intelligent agent

2. Test it with a user study
   - Scenarios that require an update to the user model
   - Context of a party influences the user's values (Fig. 1)

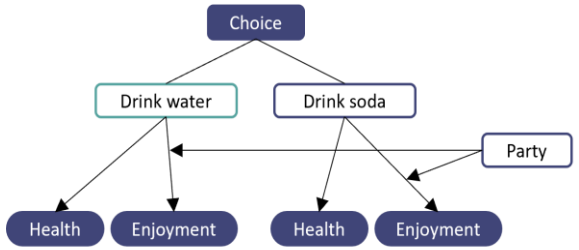

Fig. 1: Scenario that includes part of a user model in the context of a party

- System Usability Scale (SUS) survey for interface's **usability**
- Participants judge resulting user models on **accuracy**
- Open-ended interview questions for **feedback**

3. Analyse results of surveys and accuracy measure

3. Results

- Overall SUS score is 76.7, average worldwide is 68

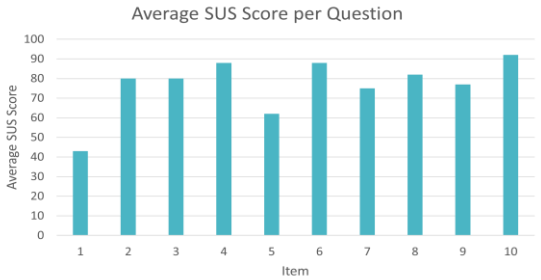

Fig. 2: Results of the SUS sore for each of the items

- Hamming distance mean is 3.6 with standard deviation 6.2
- Value difference mean is 13.5 with standard deviation 15.6

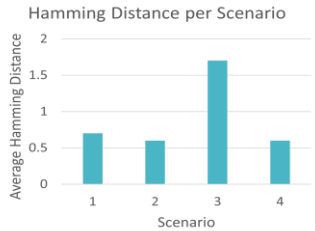

Fig. 3: Average Hamming distance of the behaviour trees per scenario

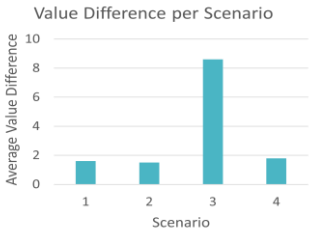

Fig. 4: Average value difference of the behaviour trees per scenario

4. Conclusions & Limitations

- Performs above average in terms of usability
- Most models required minimal changes, except one
- Feedback indicates participants would prefer visual additions
- Using audio, it takes long to elicit values, especially in isolation

- No actual intelligent agent
- No voice recognition system (Wizard-of-Oz setup)
- Real-world situations are more nuanced than simplified models
- Values chosen by researchers, not all backed by research