Effects of text-to-speech on user satisfaction and efficiency

CHRISTOS IOANNIDIS, Utrecht Science Park, Utrecht University, The Netherlands (NL)

JORT KOEDIJK, Utrecht Science Park, Utrecht University, The Netherlands (NL)

LUC MINNEE, Utrecht Science Park, Utrecht University, The Netherlands (NL)

JULIUS BIJKERK, Utrecht Science Park, Utrecht University, The Netherlands (NL)

Additional Key Words and Phrases: Methods, Artificial Intelligence (AI), Research, Machine Learning (ML), Classification, Labeling, Learning, Dialogue System, Chatbot

1 INTRODUCTION

Text-to-speech is a very important utility for a chatbot, especially considering the groups of people that benefit from such a feature. More precisely, people with vision impairments, people with illiteracy, as well as people with some form of dyslexia would face difficulty trying to interact with a standard chatbot without this functionality. That is why most chatbots nowadays do, in fact, support text-to-speech. It is commonly observed that this feature is rarely enabled by default; instead, users typically have to manually activate it.

This particular fact has led the authors of this report to wonder *why text-to-speech*, a feature that is a facilitation designed to aid accessibility, *is mostly rejected by users who do not absolutely require it.* Consequently, the purpose of this research is to identify the user reaction to the same chatbot with or without text-to-speech and to determine the ways it impacts user satisfaction. For clarification, this particular research does not aim to test the aforementioned groups that count on text-to-speech to interact with chatbots. Rather, *it targets typical users and examines how text-to-speech affects their chatbot experience.*

2 EXPERIMENTAL STRUCTURE

2.1 Participants

To conduct an experimental procedure to investigate the mentioned questions, a satisfactory number of participants has to be gathered to interact with the system. For that reason, relatives, friends, acquaintances of the authors and other students will be asked to take part in the survey. While paying attention to every participant's privacy, it will be ensured that all of them belong to the 'conventional user' category.

2.2 Experimental Design

Provided that an adequate number of participants is gathered, the experiment will follow the *within-subjects design*. More precisely, a portion of the participants will be asked to interact with the chatbot with the text-to-speech option deactivated (Group A), while for the rest of them the option will be toggled on (Group B). This decision was taken after discussion with the project supervisor, in order to discard possible result distortion any order effect which will be involved with a within-subjects design, in which all participants interact with the system both with, and without text-to-speech.

The participants of each group will be asked to interact with the system three times, in order to get a complete sense of the experience. The questionnaire will gather general demographic information and include various types of user-satisfaction questions. These will consist of yes/no questions, rating-scale questions (e.g., "On a scale from 1 to 5, how satisfied were you with the experience?"), and open-ended questions that allow participants to provide short written responses (e.g., 1-2 sentences).

2.3 Measures and Metrics

• **Time to complete the interaction**: The duration taken by participants to finish their interaction with the system will be recorded for both groups.

• **User satisfaction**: Participants will provide subjective feedback through the questionnaire, with questions measuring their satisfaction (e.g., "How satisfied were you with the interaction?" on a rating scale from 1 to 5). User satisfaction will be treated as a dependent variable.

• **Other metrics**: Additional metrics may include the number of turns in the conversation, the frequency of back-and-forth exchanges, and any observable delays in response times that might be attributed to the use of text-to-speech. Also demographics will be taken into account.

3 CONSIDERATIONS

Firstly, in the current implementation participants are unable to type while the system's voice is being played. This limitation could impact the results, as it may slow down interactions or frustrate users. Moreover, this is often not the case with conventional chatbot models. Options to modify this are currently being explored—either allowing typing during speech playback or introducing a hotkey to stop the voice playback. If this is resolved, the impact of the change will be evaluated. For instance, if the hotkey option is implemented, the frequency with which participants stop the speech will be tracked. Additionally, correlations between the stopping frequency and the length of the system's utterance could be explored to determine any significant patterns.

Furthermore, only one voice is currently available for the system's utterances. The potential effects of voice gender (e.g., comparing female and male voices) on user satisfaction and comfort could also be explored.