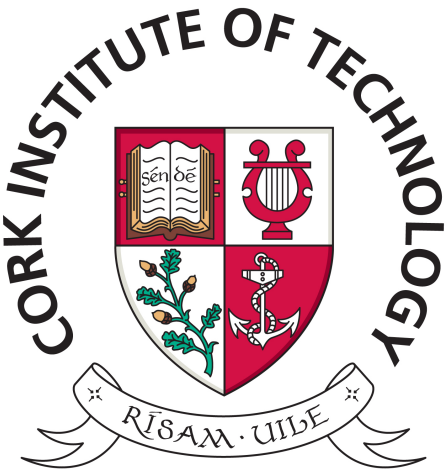

Exploring and designing speech modality in conversational agents

by

Nitesh Gupta

This thesis has been submitted in partial fulfillment for the degree of Master of Science in MSc in Artificial intelligence

in the
Faculty of Engineering and Science
Department of Computer Science

November 2020

CORK INSTITUTE OF TECHNOLOGY

Abstract

Faculty of Engineering and Science
Department of Computer Science

Master of Artificial intelligence

by Nitesh Gupta

Conversational agents are the systems which are designed to simulate conversation with a human using artificial intelligence and developed for customer query assistance, laboratory assistance etc. Agents come in many modes, one high-level classification is between voice and text. These agents are also commonly known as chatbots. Text chatbot responds based on a predefined set of conditions for extracting intents from the user's inputs. Voice chatbot gives the output in form of speech by reading the retrieved text. The output can be managed with common back-end infrastructure and a text-to-speech module to produce the speech modality. Moreover, the speech output should not just be the text spoken and there is further processing (e.g. changing the pattern of stress and vocal intensity that produces a different kind of emotions) before responding to the user. Furthermore, few interventions can be applied to the text even before reaching to the voice processing step to improve the quality of the interaction and applying Speech Synthesis Markup Languages (SSML) that can customize speech by changing the volume of the voice, pitch, pausing, speaking rate, specifying pronunciations etc. Chatbot complies with a predefined persona which is filled in slots, that responds every answer differently by using sentiment analysis. for example – respond annoyingly if the user asks the same question repeatedly, encouraging if the user is making progress etc. It is envisaged to develop this on a cloud platform that is provided by AWS. In current academia, there are ERA chatbots already being studied/developed which can be used as a case-study or a new chatbot can be designed and developed.

0.1 Research Context and Contribution to the Research Field

In today's era, there are a plethora of web-based services like a business, entertainment, virtual assistance and many more. It is a very user-friendly approach to avail everything to doorstep[1]. There is one of customer service available like live chatbots.
Chatbots are a user-friendly platform where user can efficiently interact with our system using natural language. These interactions can happen in two ways: voice or text. A chatbot should be smart enough to understand the user's entire context which is also known as an input modality. They are used by several web clients to negotiate access for information or learning.
Chatbot explores convenient ways to extract text from academic manuals, storing them in a database, and provides the front-end UI for the user to query and fetch stored information by using powerful Natural language processing (NLP) algorithms, Dialogflow etc[2]. Nevertheless, despite adapting artificial intelligence in chatbot technology, it seems that they are still in early-stage and not yet reached the utmost of their capabilities in customer satisfaction. Most of the current text-based chatbots are a stupid and present user with pre-configured responses. Likewise, voice-based chatbots are just a wrapper over those just reciting the output. These chatbots are not impactful enough and are unable to connect with users emotionally.[3]
The motivation of this research project is to fully utilise the opportunities of the speech modality beyond its simple text to speech conversion. Since chatbots are in the pioneer phase in the R&D world, There are plenty of unsolved questions that are needed to be solved, therefore, research on enhancements of chatbot like developing chatbot persona etc. can make a cause of evolution in the chatbot world.
The proposed approach is to use technologies such as Speech Synthesis Markup Language (SSML), Artificial Intelligence Markup Language (AIML), Latent Semantic Analysis (LSA) etc. for designing an appropriate model for text and voice chatbot that can interact with humans and provide a response with voice expression based on the sentiment of context. This model would be able to understand the variety of utterance by accurately extracting intents and entities from the user's dialogue input[4]. SSML changes the patterns of stress and vocal intonations in speech and gives output with different kinds of emotions. For example, vocal intensity and speech rate increases for anger and decreases for sadness. The principle of a common (speech and text) backend infrastructure would be preserved during the designing of the chatbot.
This chatbot will provide an efficient and accurate answer for any query using Artificial Intelligence Markup Language (AIML) and Latent Semantic Analysis (LSA). General

questions like welcome, greetings etc. will be responded using AIML while other service-based questions use LSA to respond timely, increasing user satisfaction[5]. It can be used by any University to answer FAQs to curious students in an interactive fashion.
Keywords— NLP (Natural language processing), SSML (Speech Synthesis Markup Language), AIML (Artificial Intelligence Markup Language), LSA (Latent Semantic Analysis) etc. [3][4]

0.2 Research Aim

This research project aims to improve the speech output of chatbot using SSML while still preserving the principle of a common (speech and text) backend infrastructure..

0.3 Research Objectives

• Design and develop a bimodal chatbot where users can interact either via text or voice. It can answer domain-specific questions as well as respond to general interactions such as FAQs
• To improve the speech output so as it complies with predefined chatbot persona while preserving the principle of a common (speech and text) backend infrastructure
• Evaluate the performance of chatbot based on effectiveness, efficiency, and user satisfaction by receiving feedback from them

Bibliography

[1] A. Shaha, B. Jain, and B. Agrawal, "Problem solving chatbot for data structures."

[2] H. Hien and P.-N. Cuong, "Intelligent assistants in higher-education environments: The fit-ebot, a chatbot for administrative and learning support," *International Conference Proceeding Series (ICPS)*, December 2018. [Online]. Available: https://doi.org/10.1145/3287921.3287937

[3] Kovalevskyi and V., "Own chatbot based on recurrent neural network. retrieved," December 2016. [Online]. Available: https://blog.kovalevskyi.com/rnn-based-chatbot-for-6-hours-b847d2d92c43#.zi22qcx5z

[4] R. Sutoyoa, A. Chowandaa, A. Kurniatia, and R. Wongsoa, "Designing an emotionally realistic chatbot framework to enhance its believability with aiml and information state," *4th International Conference on Computer Science and Computational Intelligence 2019*, September 2019.

[5] Bhavika, Nidhi, and S. Singh, "Chatbot for university related faqs," *Review of Scientific Instruments*, 2017.