Development of a natural language speech dialogue system for an AR-based, adaptive mobility agent

I. Kraljevski, M. Fischer, A. Gjoreski, D. Hirschfeld
voice INTER connect GmbH, Ammonstraße 35, D-01067 Dresden, Germany
{kraljevski,fischer,gjoreski,hirschfeld}@voiceinterconnect.de

In this paper we are presenting an insight into the design and development of a module responsible for speech interaction in an AR-based, adaptive mobility agent. Such smart mobility companion should support the users in various situations they are facing while travelling from one to another specific location. Tasks of the agent include outdoor and indoor routing and navigation as well as search and information retrieval for point of interests, time schedule notifications, current traffic and weather conditions, changing means of transportation, unexpected changes of the conditions on a planned route (delays, accidents, traffic jams, cancelled flights, etc.).

In order to achieve the above mentioned functionality in real time, such system should take advantage of local and remote services which are providing simple and complex input data and derive higher-order knowledge to notify the user via audio-visual feedback or adapt the current behaviour of the system. The architecture of the agent is combining separate services, including handsfree interaction via AR-glasses and speech dialogue, synchronized by a central management component. The interaction in the system is based on communication protocols like in Internet of Things, where many elementary sensors/services are providing multi-modal data in real time using lightweight message transfer (like Message Queuing Telemetry Transport – MQTT). Based on this modular architecture, the system can be easily adapted and extended by activating/deactivating modules which are not necessary or permitted in some particular situations (visual input at airport security checks) or are not available at the moment (no remote connectivity).

One of the core system components is the Speech User Interface (SUI) providing hands-free natural language interaction. Implementing such functionality and quality interaction on a mobile device is a challenge due to non permanent connection to remote speech services. To provide acceptable user experience the SUI has to have robust hybrid recognizer with possibility of identifying semantic concepts locally with the same performance as remotely.

In the assumed scenarios the system will be used in open and closed spaces with a significant presence of background noise, where the user might be under stress and interacting using spontaneous speech (Lombard effect, hyper-articulation, hesitations, breathing, etc.). Therefore, a suitable dialogue manager has to be able to handle possible miscommunication by employing adaptive management strategies, using confirmations and error recovery. Also, it has to be able to resolve deictic expressions, pronouns, other anaphora and ellipsis by using other available modalities. In order to achieve system adaptability, a User Modelling (UM) component as a modality provides additional informations about the user`s preferences and habits, where they could be predefined or derived statistically from the recent interactions or long term history.

In order to collect data and simulate the human-computer interaction, Wizard of Oz (WOz) experiments were carried out in a VR environment providing valuable insights how speakers interact in the predefined scenarios while covering complete system's functionality. The preliminary investigations using online questionnaires yielded general directions about preparation of the WOz experiments from the aspect of the users' expectations.

The scenarios were based on a predefined list of system's functions including specific semantic entities like locations, time information, mean of transport, points of interest, etc. The experiments also provided significant amount of multi-modal data, which in the case of the speech interface component is used for creation and optimization of the recognition resources – language models and semantic parsers. The WOz experiments also provided knowledge which was used as a basis for the initialisation of the User Model and the final dialogue design, particularly focusing on prompting and recovery strategies.