

RESIVOZ: Dialogue System for Voice-based Information Registration in Eldercare

RESIVOZ: Sistema de Diálogo para el Registro de Información de Cuidado de Mayores mediante Voz

Laura García-Sardiña, Manex Serras, Arantza del Pozo, Mikel D. Fernández-Bhogal

Vicomtech Foundation, Basque Research and Technology Alliance (BRTA)

Mikeletegi 57, 20009 Donostia-San Sebastián (Spain), +[34] 943 30 92 30

{lgarcias, mserras, adelpozo, mfernandez}@vicomtech.org

Abstract: RESIVOZ is a spoken dialogue system aimed at helping geriatric nurses easily register resident caring information. Compared to the traditional use of computers installed at specific control points for information recording, RESIVOZ's hands-free and mobile nature allows nurses to enter their activities in a natural way, when and where needed. Besides the core spoken dialogue component, the presented prototype system also includes an administration panel and a mobile phone App designed to visualise and edit resident caring information.

Keywords: Gerontechnology, Spoken Dialogue Systems, Eldercare, Information Registration

Resumen: RESIVOZ es un sistema de diálogo orientado a ayudar a gerontólogos a registrar fácilmente información sobre sus cuidados a los residentes. En comparación con el uso tradicional de ordenadores instalados en puntos de control específicos para registrar la información, la naturaleza manos-libres y móvil de RESIVOZ permite al personal gerontológico registrar sus actividades de forma natural, donde y cuando se necesite. Además del principal componente de sistema de diálogo hablado, el prototipo de sistema también incluye un panel de administración y una aplicación móvil diseñada para visualizar y editar la información de cuidados a residentes.

Palabras clave: Gerontotecnología, Sistemas de Diálogo Hablado, Cuidado de Mayores, Registro de Información

1 Introduction

Spoken Dialogue Systems (SDS) are increasingly being integrated and deployed into a wide range of devices we can find in our daily lives. Many smart phones, speakers, and cars come with conversational assistants (e.g. Alexa, Siri, Google Assistant, Cortana) which can provide us with useful information (e.g. weather forecasts) and help us with various simple tasks (e.g. playing music). Because of their popularity for infotainment purposes, the application of SDS is also in demand and being explored in professional fields like healthcare (Laranjo et al., 2018). In particular, the use of conversational agents to support different kinds of patient populations (e.g. the elderly, the chronically ill) with health tasks is an emerging field of research with several projects underway, like

EMPATHIC¹, MENHIR² or SHAPES³.

The use case presented in this paper is also linked to healthcare and aims to facilitate the registration of resident information by geriatric nurses throughout a work shift. The implemented SDS provides them with a hands-free, natural, and flexible interface to register resident information in a nursing home management system, allowing them to devote more time to practice people-based care.

The paper is structured as follows: Section 2 presents our application use case; Section 3 describes the system's components; Section 4 displays a sample dialogue showing different interaction situations; and Section 5 gives some conclusions and future work directions.

¹www.empathic-project.eu/

²www.menhir-project.eu/

³cordis.europa.eu/project/id/857159

2 Use Case

The application scenario is that of an elderly care residence where geriatric nurses need to register information regarding the activities they carry out with each resident into a management system. Such activities include: oral hygiene, shaving, grooming, changing diapers, showering, checking overall appearance, cutting and checking fingernails, performing postural changes, checking bowel movements, and checking how much residents eat and drink in each meal. Sometimes, nurses also need to add observations to these activities and follow-up comments for each resident, so that other carers in following shifts are informed about particular incidents.

One of the issues with the traditional computer-based systems is that there is usually just one shared computer per floor or unit. Geriatric nurses usually do not have the time to stop by and register activities as they are completed, so they tend to wait until the end of their shift to enter all the activities carried out throughout their working day. This entails two main problems: on the one hand, queues are generated, since all geriatric nurses want to register all the information simultaneously at the end of their shift; on the other hand, the information introduced may not be entirely correct due to the difficulty of remembering all activities carried out with each resident throughout a shift.

The proposed solution integrates a mobile spoken dialogue system that allows geriatric nurses to introduce the information in an easy, hands-free way anywhere and anytime throughout their working day. Given the nature of the geriatric nurses' work, the solution's selected hardware needs to be: (i) **ergonomic** it needs to be lightweight, wireless and wearable, since it needs to be carried throughout the whole shift; (ii) **safe**, some wearable devices like smartwatches can cause abrasion injuries on residents' skin and should be avoided; and (iii) **discreet**, audio capturing and playing should not be invasive nor interfere with social interactions.

3 System Architecture

Figure 1 shows a diagram of RESIVOZ's architecture. To comply with the identified requirements, the selected hardware devices include (i) a single-earphone plus short microphone Bluetooth headset, and (ii) a lightweight, small smartphone. The micro-

phone has soft noise cancelling and mute option, so geriatric nurses can disable audio recording when talking to someone other than the system. The smartphone allows mobile audio capturing, App usage, and response generation using its built-in Text-To-Speech (TTS) functionalities.

As for software, the developed prototype system consists of four main components: 1) a Spoken Dialogue System that allows user-system interaction, 2) a Mobile Phone App for information visualisation and edition, 3) a Control Panel Interface to manage users, shifts, and resident caring activities, and 4) a Domain Database where all the necessary information is stored.

3.1 Spoken Dialogue System

This is RESIVOZ's core component, acting as a voice-based interaction bridge between the final users and the rest of the system. The SDS receives the information to save and lets the users know whether the it has been correctly logged in or there has been some problem in the process: there is no identified resident, the information is incomplete or invalid for the current shift, there has been a database connection issue, or the system could not parse the user's input. The implemented SDS is a pipeline of four submodules:

(1) Automatic Speech Recognition (ASR) In this module, the audio stream captured by the headset goes to an energy-based Voice Activity Detector, which separates speech from non-speech segments and sends the former to a Speech-To-Text (STT) service. Here, Google Cloud's STT API service was integrated after introducing some custom domain phrases, words, and names as *contexts* for domain adaptation⁴.

(2) Spoken Language Understanding (SLU) For this module, given that real domain data was not available, Phoenix grammars (Ward, 1991) were developed to parse the transcribed geriatric nurses' input into semantic codifications of *act-slot-value* sets. A hierarchical taxonomy was defined covering the domain. It includes four acts (*identify*, *inform*, *deny*, *delete*), 34 slots, and 28 fixed values (as named entities, resident names are considered variable values and not included in the taxonomy). Apart from parsing

⁴cloud.google.com/speech-to-text/docs/context-strength. Data sharing options were disabled in compliance with the GDPR

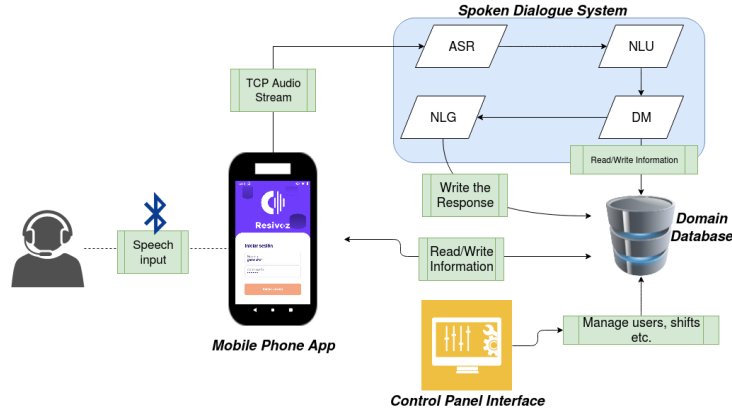


Figure 1: Architecture of the RESIVOZ system

the semantic actions from users’ inputs, the NLU module also identifies named entities – resident names in this case – from users’ utterances, using a fuzzy string-matching method and their registered full names. Residents are identified with their name and first surname⁵, while their second surname is used for disambiguation when needed.

(3) Dialogue Manager (DM) Decision making and interaction planning is based on Attributed Probabilistic Finite State Bi-Automata (Torres, 2013), which store the latest user input, system output, and attributes (i.e. dialogue memory). The latter store task-sensitive information such as whether a resident is identified, or if there is any ambiguity with the provided information (Serras, Torres, and del Pozo, 2019). The DM is also responsible for reading and writing information in the Domain Database and for modifying attributes of the dialogue state. To leverage the lack of data, the next action is selected using *if this, then that* rules.

(4) Natural Language Generation (NLG) Natural language text templates were generated for each possible system action so as to transmit an adequate response to users in a human-interpretable manner.

3.2 Mobile Phone App

An Android App was developed to serve as a front-end and access point for the geriatric nurses. This is meant to be used at sporadic times, since nurses need to have both hands free while on their caring duties. When they log in, the information on their shift is shown: their assigned resident group, daily

tasks, previous follow-up comments on residents, and so on. Also, a TCP connection is established with the server, starting the audio stream. The application has three purposes: 1) to be a gateway to easily display and modify the information that is being logged into the database; 2) to handle the socket connection and send the audio stream to the server where the back-end is located; and 3) to play the text responses given by the SDS using the phone’s built-in TTS.

3.3 Control Panel Interface

A control panel web interface was developed to allow managers to easily administer shift-related data. From this panel, shifts and nurse and resident profiles can be created and deleted in the database in a user-friendly way.

3.4 Domain Database

The domain database is where all the information is registered and read from: resident and user profiles, the shifts scheduled by the manager, and the caring activities, observations and follow-up comments introduced by the geriatric nurses for each resident. It also stores the users’ sessions during their assigned shifts. A series of customised RESTful API calls were implemented to communicate the database with the SDS, the Mobile Phone App, and the Control Panel.

4 Sample Dialogue

This section presents a simulated dialogue⁶ between a geriatric nurse (U) and the system (S), showing some of the different situations that may happen during interaction.

⁵In Spain, unlike in other countries, people have two surnames.

⁶The names are made up and do not correspond to specific real residents. The dialogue is presented in English although RESIVOZ uses Spanish.

- (1) U: Mouth cleaning okay
 S: There is no resident identified
 U: Ana Atxa had her teeth brushed
 S: Oral hygiene done for resident Ana Atxa Cruz

In the first interaction, presented in example 1, no resident is identified, so the system will not register the information until one is. Then, the system is able to identify a resident given their name and first surname.

- (2) U: Resident Pilar García
 S: There are several residents with that name
 U: Pilar García Zubia
 S: Resident Pilar García Zubia has been identified

In example 2, the user tries to identify a resident by name and first surname, but there are several residents with that name. By using the second surname, the target resident is correctly disambiguated.

- (3) U: She had half her meal for lunch
 S: Intake of solids at lunchtime 50%
 U: Add observation to intake of solids at lunchtime
 S: Dictate
 U: Pilar said she was too full and refused to eat more, she didn't want dessert either, *end of annotation*
 S: Observations to solid intake at lunchtime saved

Although the user is not saying a name explicitly, in example 3 the identified resident is known from the session memory. In this example, the user adds an observation, i.e. a free text comment, to a registered control. Dictation ends when a pre-established command is uttered: “end of annotation”.

- (4) U: All my residents but Ander Goñi have been showered
 S: Shower registered for all your residents but Ander Goñi Uria

RESIVOZ incorporates shortcuts to allow the selection of several residents and/or controls at the same time. Example 4 shows the use of a shortcut where all the residents in the nurse's assigned group but one are selected.

- (5) U: Delete shower for Ander Goñi
 S: No completed control with that name has been found

In example 5, the user tries to delete a control for a resident, but this control was

registered before for all residents in the group except for the now identified. The system informs that it cannot delete a control that has not been saved as complete.

5 Conclusions and Future Work

We have developed and integrated a Spoken Dialogue System that helps geriatric nurses in an elderly care centre easily register resident caring information at any time and any place during their work shifts. Using RESIVOZ over the traditional system, only available in a few shared computers at specific control points, may lend geriatric nurses more quality time for their elderly caring activities. For future work, an evaluation by geriatric nurses is to be performed in their workplace, tackling different aspects such as system usability, points of improvement, difficulties found, adoption period, and other feedback regarding the system's usability and applicability in the real work environment.

Acknowledgements

This work was supported by the Adinberri Programme of the Provincial Council of Gipuzkoa, via the RESIVOZ project.

References

- Laranjo, L., A. G. Dunn, H. L. Tong, A. B. Kocaballi, J. Chen, R. Bashir, D. Surian, B. Gallego, F. Magrabi, A. Y. Lau, and E. Coiera. 2018. Conversational agents in healthcare: a systematic review. *Journal of the American Medical Informatics Association*, 25(9):1248–1258.
- Serras, M., M. I. Torres, and A. del Pozo. 2019. Goal-conditioned user modeling for dialogue systems using stochastic bi-automata. In *Proceedings of the 8th International Conference on Pattern Recognition Applications and Methods*, pages 128–134.
- Torres, M. I. 2013. Stochastic bi-languages to model dialogs. In *Proceedings of the 11th International Conference on Finite State Methods and Natural Language Processing*, pages 9–17.
- Ward, W. 1991. Understanding spontaneous speech: The phoenix system. In *Proceedings of the 1991 International Conference on Acoustics, Speech, and Signal Processing*, pages 365–367.