

Spoken Dialog System: Direction Guide for Lahore City



Aneef Izhar ul Haq¹, Aneef Anwar¹, Aitzaz Ahmad¹, Tania Habib¹,
Sarmad Hussain¹, Shafiq-ur-Rahman²



Introduction

A Spoken Dialog System aims to provide services, such as Weather Information and Direction Guidance, to common people by interacting with them through speech – user records his query and the system fetches and plays back the response accordingly. This research paper focuses on the architecture and framework of a prototype Spoken Dialog System, used for providing the direction guidance to, and from the 49 locations of Lahore.

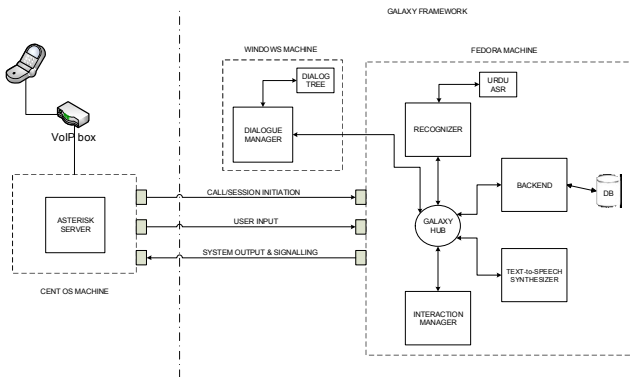


Figure 1. System Architecture

RavenClaw Dialog Manager

It is the central component of the whole system and controls the flow of the dialogue. It is the first module that is triggered after Asterisk receives a call. It is a generic module borrowed from the Olympus Dialog System [5] and is programmed using the Dialog Task Tree. The whole dialogue consists of prompts, either from the user or the system. Dialog manager executes each of the prompt sequentially, and either gets the user query or generates the appropriate response for the corresponding query [6].

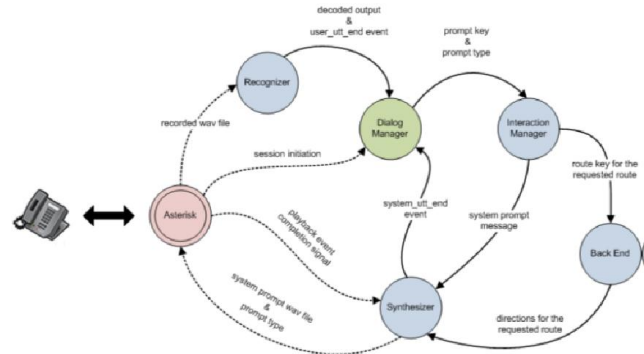


Figure 2. Process/Call Flow

System Architecture

The Direction Guidance Spoken Dialog System is based on the distributed architecture of Galaxy Framework [1] [2], which consists of a central entity called HUB, and several other independent modules. All communication between the modules takes place via HUB [3] and the Galaxy Framework only provides the interface between the HUB and the other modules. Figure 1 provides the high-level architecture of the complete dialog system, along with the individual subcomponents.

- **Telephony Framework/Asterisk Server** – It is the only non-Galaxy compliant module in the whole system and communicates through socket connections. It is responsible for initiating the calls, getting the user inputs and playing back the system response to the user.
- **Automatic Speech Recognizer (ASR)** – The Urdu speech recognizer has a vocabulary of 49 words, the names of various locations of Lahore, which were chosen due to their relatively large training data which in turn results in good recognition rate.
- **Backend Server** – The backend server interacts with the routes database and fetches the route directions for the corresponding source, destination pair. The path directions from the source location to the destination are retrieved from The Google Directions API [4].
- **Text-to-Speech Synthesizer (TTS)** – Path directions received from the Backend Server are converted from text to speech using the Festival speech synthesizer. Since Urdu TTS system has not been made functional yet, the system response is generated in English.
- **Interaction Manager** – Its sole purpose is to parse the messages sent by the Dialog Manager and re-route them to the desired module.

Process Flow

Figure 2 shows the process flow diagram of the dialog system and interaction between the various components. The call flow goes as:

- Asterisk server receives the phone call.
- It connects the caller with the GALAXY framework.
- A new session initiates in the Dialog Manager
- Greeting response is generated through TTS & played back via Asterisk
- User is prompted to speak the source and destination locations for which the directions are required
- Recorded wavfiles are sent to ASR for decoding
- ASR decodes the spoken utterances and returns the source, destination route key
- Route key is passed to the Backend server which fetches the directions for the requested route
- System response is generated through TTS and played back to the user via Asterisk

Sample Dialog

System: Welcome to the CLE Direction Guide.

System: What is your current location?

User: لاهور

System: Where are you headed to?

User: لوالہنڈی

System: Just a minute let me check that for you.

System: Total distance is 2.3 km. Head southwest towards Katchery Rd. After 77m ... onto Flemming Rd. After 0.1 km, Gawal Mandi, Lahore.

System: Thank you for using the CLE Direction Guide. Goodbye!

Challenges

- **Misrecognition** – Location spoken is misrecognized as some other location in the vocabulary list. e.g. گوالہنڈی is misrecognized as لوالہنڈی.
- **Non-recognition** – Location spoken is not recognized at all.

Contact

¹Centre for Language Engineering,
Al-Khwarizmi Institute of Computer Science, UET, Lahore, Pakistan
aneef.izhar@kics.edu.pk, firstname.lastname@kics.edu.pk

²FAST NUCES, Lahore, Pakistan
shafiq.rahman@nu.edu.pk

References

- [1] S. Seneff, E. Hurley, R. Lau, C. Pao, P. Schmid and V. Zue, "GALAXY-II: A Reference Architecture for Conversational System Development," in *Proc. ICSLP'98*, Sydney, Australia, 30th November - 4th December, 1998, pp. 931-934.
- [2] J. Polifroni and S. Seneff, "GALAXY-II as an Architecture for Spoken Dialogue Evaluation," in *Proc. Second International Conference on Language Resources and Evaluation (LREC)*, Athens, Greece, May 31- June 2, 2000.
- [3] "GALAXY Communicator Documentation," [Online] <http://communicator.sourceforge.net/sites/MITRE/distributions/GalaxyCommunicator/docs/manual/>
- [4] Google, [Online]. <https://developers.google.com/maps/documentation/directions/>. [Accessed 13 November 2013].
- [5] D. Bohus, A. Raux, T. K. Harris, M. Eskenazi and A. I. Rudnicky, "Olympus: an open-source framework for conversational spoken language interface research," in *Proc. HLT-NAACL 2007*, Rochester, New York, USA, April, 2007.
- [6] D. Bohus and A. I. Rudnicky, "The RavenClaw dialog management framework: Architecture and systems," *Computer Speech and Language*, vol. 23, January, 2009.