



Designing an Architecture for Delivering Multilingual Interactive Voice Information Services

Honelet Endale Mulugeta¹, Eshetu Deresu²,

¹Department of Computer Science, Institute of Technology, Ambo University, Oromia-Ethiopia
hoenmune@gmail.com

²Department of Computer Science, Institute of Technology, Ambo University, Oromia-Ethiopia
eshetu.deresu@ambo.edu.et

ABSTRACT

In this paper, we address the issue of how to design multilingual market information service delivery interactive voice response system (MMI-IVR) architecture to serve hundreds of thousands of low literacy target groups to access market information simply and independently using their mobile and fixed line telephones network for three mostly spoken Ethiopian local languages Amharic, Afaan Oromo and Tigrigna. The proposed system architecture uses three-tier architecture model, which segments into three tiers of services. The top layer of the presentation logic is service delivery which is mainly used as an interface to farmers and traders, which are located at distant locations, via the telephone network. The second level of the architecture is middle tiers that accepts telephone digit as inputs from the user over the telephone network and provide required information back to the caller via voice messages. The data tier is the database containing market price information. When market information is requested by the caller, the presentation logic reads the information from the database and uses its template files to construct on voice message. To accomplish this, it uses the pre-recorded template files, and dynamically generated information which basically done in this tier by text to speech engine. To illustrate the breadth of the MMI-IVR architecture prototype is developed using PHP programming language PHPAGI script to interface with PBX as software (Asterisk server) and emulated by telephone X-lite soft phone as phone interface.

Key words: Asterisk Gateway Interface. Asterisk server, Dial plan, Interactive voice response system.

1. INTRODUCTION

In Ethiopia, agriculture is the largest livelihood provider (more than 80% being farmers). Like other developing countries, farmers and intermediaries have continued to rely on market information supplied and verified through traditional word-of-mouth approach [15]. The existing solutions mainly target upper class stakeholders In contrast, the majority of smallholder farmers produce just consumable

commodities and selling these commodities are the only ways to get money for purchasing input for the next farming, besides using it to cover their living expenses. In addition, some of the commodity exchange systems Ethiopian Commodity Exchange (ECX) force a farmer to have at least a certain predefined amount of a commodity before entering to the market competition. With respect to this, the existing commodity exchange services are observed to disseminate only little commodity market information. Clearly, the abundance of the mentioned challenges in the existing solutions and their increasing popularity make a strong case for the need to look for interventions at the level of processes, logistics and services, delivery and presentation with IT support for the benefit of all stakeholders. In other words, an appropriate ICT-based information system for agriculture can be an enabler to improve the current trading process, provide fine grain logistic support, and bring a transparent and fair trading process. Formerly a preliminary project work is being done by the Ethiopian Commodity Exchange to deliver Market Information. This work will help to provide market information for the lower stakeholders. This project work is not intended to replace or duplicate the work that is being done by the Ethiopian Commodity Exchange but rather its outcome can help to provide additional market information to what is already being done by the Ethiopian Commodity Exchange. Multilingual market information interactive voice response system will be used to serve hundreds of thousands of farmers to access market information simply and independently using their mobile and fixed line telephones network. Additionally, the result of this project work can be used by farm households, farmers' organizations, cooperatives and cooperative unions, traders, brokers, provider, advisors (e.g., agricultural extension), authorities, suppliers, and service providers. In support with paper on [13], mentioned that the current mobile phones penetration in developing world answers our choice to be mobile-based market information service delivery. In fact with different levels of sophistication, low end terminal mobile devices are targeted for the delivery of market information services to make accessible by marginalized rural area user in Ethiopia. To date, there have been some research activities to explore

the feasibility of IVR interfaces for low literacy users [1-6]. There have been a number of preliminary findings and practical lessons learnt but much still needs to be done in terms of fully exploiting the possibilities of IVR interfaces in a context very different from the developed world [7, 8]. In light of the foregoing, the research question addressed in this paper is how we should design architecture for delivering multilingual Market information Interactive Voice Information Services for low literacy users in Ethiopia. The main objective of this research is designing architecture for delivering multilingual Market information Interactive Voice Information Services for illiterate or semiliterate farmers: The specific objectives include; (1) Design multilingual market information IVR architecture (2) Develop multilingual market information IVR system. The scope of this research lies on low end terminal as a means of service delivery of multilingual Market information services for illiterate or semiliterate individual farmers and transporters target group peoples using interactive voice response system via Dual Tone Multi-Frequency (DTMF) signal input. The telephone user can access the information from anywhere at any time simply by dialing a specified number and following an on-line instruction when a connection has been established. The IVR system uses pre-recorded or computer generated voice responses to provide information in response to an input from a telephone caller. The input may be given by means of touch-tone or Dual Tone Multi-Frequency (DTMF) signal, which is generated when a caller presses a key of his/her telephone set, and the sequence of messages to be played is determined dynamically according to an internal menu structure (maintained within the IVR application program) and the user input [9-12].

The proposed multilingual IVR system consists of the speech entity of the market information IVR system which offers a multilingual environment in Text To Speech Engine which supports Amharic, Afaan Oromo and Tigrigna languages. Text-to-speech (TTS) software will be developed to automate its outgoing messages. Instead of recording all of the possible responses to users query, the computer can generate customized text-like responses and read it back to the customer using an automated voice.

2. PROPOSED SYSTEM ARCHITECTURE

In the following part of the report we discuss the all aspects of the proposed architecture starting with the application tier and then moving towards the data tier. Fig. 1 shows the architecture for a multilingual market information interactive voice response system. The proposed multilingual market information interactive voice response system uses three-tier architecture model, which is the fundamental framework for the logical design model, which segments an application's components into three tiers of services namely, the presentation tier, the middle tier, and the data tier. These tiers do not necessarily correspond to physical locations on various computers on a network, but rather to logical layers of the application. In the proposed architecture the top layer of the presentation logic is service delivery which is mainly

used as an interface to farmers and traders, which are located at distant locations, via the telephone network. The second level of the architecture is middle tiers that accepts inputs from the user and provide required information back to the caller via voice messages. The third tier of the architecture is data tier which consists of voice database and General information service for Market Price database. From the developmental view point the three tiers: The presentation tier or more appropriately, user services layer, gives a user access to the application. This layer presents data to the user and permits data manipulation and data entry. The middle tier, or business services layer, consists of business and data rules of the proposed Market Information System. Also referred to as the business logic tier, it is here where all business problems and logic is implemented.

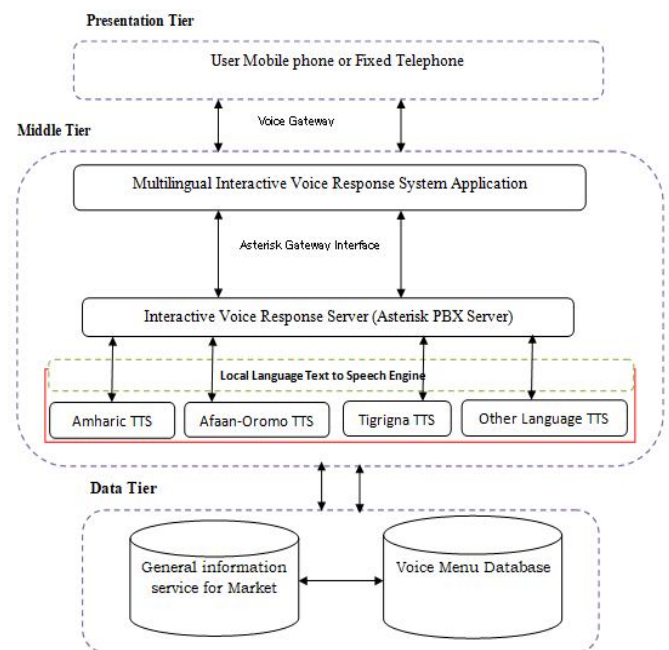


Figure 1: Architecture of Multilingual Market information interactive voice response System

The data tier, or data services layer, interacts with persistent data stored at the General information service for Market Price database. In fact, this is the actual DBMS access layer. This layer can be accessed through the middle tier. During an application's life cycle, the three-tier approach provides benefits such as reusability, flexibility, manageability, maintainability, and scalability.

2.1. Presentation Tier

The presentation tier provides interactive communication between human interface on user touch tone or DTMF and computer. This tier accepts telephone digits from the user over the telephone network. For example, when a user presses one on the telephone keypad, telephone digit of one will be sent to the application logic via the telephone network. The application logic detects the frequency and identifies the key. Then, depending on the voice menu level,

the key pressed will be interpreted to provide the service by having callers select from a list of menu options using their touch-tone keypad (i.e. “press 1 for Amharic market information price, press 2 for Afaan-Oromo market information price, press 3 for Tigrigna market information price, press * to return to back menu and press # to end call”). VoIP (Voice over Internet Protocol) is a booming technology in the recent period and has achieved appreciation in the industries like educational and professional institutions. In [14, 15], there is a detailed description of SIP proxy server performance while using VoIP services. The main protocols for the implementation of this technology are SIP. They have also analyzed and evaluated most of the features of SIP. The Asterisk voice engine is open source and is available free of charge. Asterisk-based IVR systems that connect over VoIP require no special hardware or software licenses. In the proposed architecture the phone which are being used for inter user communication are IPsoft phones known as X-lite which is a proprietary VoIP soft phone that uses Session Initiation Protocol (SIP).

2.2.Middle Tier

This is the application logic of the system which interacts with the presentation tier via voice gateway. Generally, the application logic is based on one central server known as asterisk server dedicated to service aggregation and with multiple local language engine handling (local) text to speech and localized knowledge bases. The central server will be dedicated to the aggregation of these services and to the aggregation of information coming from all the other nodes. Ideally, the central server has national scope. In the local language TTS engine (Amharic language TTS, Afaan Oromo language TTS, Tigrigna language and other local language TTS) we particularly seek to build a text to speech engine by structuring, integrating and standardizing data and information from market information domains. Thus, each local TTS engine provides market information services that are relevant to the community in which they are deployed, namely to allow all the system stakeholders in the area to communicate freely and a localized knowledge base providing information relevant to the area in the local language. The local TTS engines are connected with all the other TTS engines and the central server. The important aspect of localization is divided between the application logic and the user interfaces, where the application logic is mainly responsible for the localization of data management and the user interfaces are responsible for the actual content delivery management per users of the system. Moreover, security should be handled throughout the entire system, though certain aspects of security are best implemented at specific layers of the architecture. The three sub components of this tier are multilingual market information interactive voice response application, asterisk server (Asterisk PBX server) and other local languages text to speech engine.

A. Multilingual market information interactive voice response application

This subsystem is mainly responsible to manage issues related to voice menus handling. The subcomponent have some general pre-recorded messages recorded based on IVR menu structure that will be concatenated with synthesized local language speeches that it plays while interacting with caller. Specifically, the application logic has several pre-recorded vocabulary files. Numbers, names of weeks, names of months, common words used for communication and application specific voice messages that will be recorded and saved at a predefined location with defined file name. When market information is requested by the caller, the application logic reads the information from the database and uses its template files to construct on voice message. To accomplish this, it uses the pre-recorded template files, and dynamically generated information which basically done in this tier by text to speech engine. By playing voice files we can completely read the market information. All messages are constructed in this manner dynamically. Hence, the first step in developing Interactive Voice Response system is recording template and building text to speech engine. More than three languages can be implemented in the system because a user belongs to different region and languages and thus an IVR System must support multiple languages. For our demonstration purpose we used three mostly spoken Ethiopian languages Amharic, Afaan Oromo and Tigrigna. This makes the service friendlier to the users because some inhabitants of rural communities in isolated areas cannot speak English or Amharic very well. The information stored in the IVR server is related to market information in order to stimulate the rural community to increase economic growth in areas such as agriculture. This sub component is designed using AGI scripts in PHP programming language i.e. PHP AGI.

B. Asterisk server (Asterisk PBX server)

The proposed architecture used Asterisk 1.8.10.1 used to turn an ordinary computer into a communications server which is configured on Linux (Ubuntu 10.4) operating system. The Asterisk server comprises of different configuration files; “sip.conf” file that contains the SIP user registration for Inter Asterisk Communication for VoIP networks located on /etc/asterisk/sip.conf. The “extensions.conf” file that contains the “dial plan” the heart of any Asterisk system defines how Asterisk handles incoming or outgoing calls and controls execution flow for all of its operations.

C. Local language text to speech engine

It is a software component that enables IVR system to support multiple languages. In this sub component text to speech engine for mostly spoken Ethiopian local language belongs to different region is developed. Fedora8 operating system is used to develop the proposed system Afaan Oromo, Amharic and Tigrigna Languages text to speech (TTS) engine using Festival. Afaan Oromo, Amharic and Tigrigna Languages TTS engine is developed by unit selection speech synthesis prin

ciple by segmenting Afaan Oromo, Amharic and Tigrigna Language diaphones from each language database with recorded utterances and then concatenated with each other in the process of synthesis. To integrate Festival server with asterisk server festival.scm module is used. The configuration file “festival.conf” is configured for controlling how Asterisk connects to and interacts with the Festival server. The proposed system can stimulate market activity in rural areas efficiently. The availability of market information to stakeholders is expected to increase production output. Sometimes rural communities who trade their output to the market center via brokers cannot sell their output optimally because they do not know the list price of their goods in the market. This problem will be reduced by making the price list information available on the proposed Market Information System.

2.1 Data Tier

The third tier of the architecture is data tier which consists of voice database and General information service for market database. Since concatenate synthesis uses actual short segments of recorded speech that were cut from recordings and stored in an inventory (“voice database”), either as “waveforms” (Unicode), or encoded by a suitable speech coding method, we prepared voice database for the proposed system. The database contains collection of information about the market price information (which will be updated online, that is stored by any existing (commercial) commodity exchange initiatives. This database also contains information on services' registry which we require for efficient service discovery as a web service.

3. RESULT AND DISCUSSION

Here, the implemented system is described. How the user interacts with the system and some of the results of interaction with the system along with the screen shots of the prototype. In this system user interact by using telephone keypad with the asterisk server and beside, the asterisk server have dial plan which is used to define extension number (i.e. the specific IVR system number). When messages are received on this channel to asterisk server, the message is sent to the MMI-IVR application for processing using Asterisk's Application Gateway Interface (AGI). Finally, the MMI-IVR will interact with the user on the channel based on predefined IVR process tree defined on MMI-IVR application. Implementation detail for each component is described as follows:

3.1 X-Lite Soft phone

This application used as phone interface to communicate with MMI-IVR system. The soft phone installed and configured on windows operating system computer and connected with the asterisk server that is installed and configured on Ubuntu operating system. In the implemented system private range IP address has been set for the two computers. The two computers are uses UDP port 5060 and TCP port 10000:20000.

3.2 Asterisk server

Asterisk is a complete PBX in software. It runs on Linux (ubuntu) operating system and provides all of the features you would expect from a PBX and more. In the implementation part, numbers of installations and configurations have been performed in order to make the prototype system ready. First of all, the asterisk based multilingual Market information server is configured and made available in a one computer. For installing and configuring asterisk, several numbers of associated and dependent packages are installed and configured beforehand. Once the asterisk server is ready, we made server capable for handling and establishing VoIP connections through some installations and configurations. There are a few common configuration files that are used during prototype implementation.

The major configuration task is performed on extensions.conf file to configure the dialplan. Dialplan / extensions.conf, this component is the heart of Asterisk system. In the dialplan we defined how calls flow into and out of the system. It is a form of scripting language and contains instructions that asterisk follows in response to external triggers. The Asterisk dialplan is specified in the configuration file named extensions.conf. The dialplan is made up of four main concepts: contexts, extensions, priorities, and applications. In our system it located on /etc/asterisk/extensions.confpath. The configuration file "extensions.conf" contains the "dial plan" of Asterisk, the master plan of control or execution flow for all of its operations. It controls how incoming and outgoing calls are handled and routed. This is where you configured the behavior of all connections through our asterisk server. The content of "extensions.conf" is organized in sections, which can be either for static settings and definitions or for executable dialplan components in which case they are referred to as contexts. The settings sections are general and global and the names of contexts are entirely defined by the system administrator. A special type of contexts are macros, label by a user defined name prefixed with macro. These are reusable execution patterns, like procedures in a programming language. Every section in extensions.conf starts with the name of the section contained within square brackets.

3.3 Multilingual Market Information Interactive Response (MMI-IVR) application

This is a major component of the system developed using PHP. The system call flow is implemented based on the design of the multilingual IVR tree structure. For example: in Amharic language: “ይህ የኢትዮጵያ ምርት ገበያ የገበያ ዋጋ የመረጃ መስመር ነው ለአማርኛ አንድን ይጫኑ”,

in Afaan Oromo language “kuuni seereroo odefenno omis haa geeti gaabai Ethiopiat Afaan Oromo fi lemma geedii qabaa” and in Tigrigna language: “እዚ ናይ ኢትዮጵያ ምርት ሰደጋ ናይ መረጃ መስመር እዩ ንናይ ትግርኛ 3 ቁፀሪ ይፅፍኑ”. With the introduction of the Asterisk Open Source PBX, it was

required to establish a method for third party programs to interact with Asterisk, while not changing the Asterisk core – hence the AGI interface. Asterisk Gateway Interface (AGI) enables the development of Asterisk enabled applications without the need of modifying the Asterisk core. Asterisk communicates with MMI-IVR AGI scripts via STDIN/STDOUT. When a AGI script is invoked from within the Asterisk dial-plan, the following steps always happen: Asterisk forks out and runs the MMI-IVR application in its own user space. All channel variables that were available to the Asterisk dial-plan, prior to executing the MMI-IVR AGI script are available to the AGI script. Asterisk sends out a bunch of information that must be handled before the MMI-IVR AGI script actually starts running. Finally the MMI-IVR logic runs at this point. During the implementation of MMI-IVR application we have used an AGI wrapper classes PHPAGI. PHPAGI is built from three different classes for programming AGI scripts, `phpagi.php`, `phpagi-asmanager.php` and `phpagi-fastagi.php`.

4. CONCLUSION

The proposed multilingual market information interactive voice response system uses three-tier architecture model, which is the fundamental framework for the logical design model, which segments an application's components into three tiers of services namely, the presentation tier, the middle tier or application logic, and the data tier. The top layer of the presentation logic is service delivery which is mainly used as an interface to farmers and traders, which are located at distant locations, via the telephone network. The second level of the architecture is middle tiers that accepts inputs from the user and provide required information back to the caller via voice messages. This tier accepts telephone digits from the user over the telephone network. The application logic has several pre-recorded template files. Numbers, names of weeks, names of months and common words used for communication, and application specific voice messages will be recorded and saved at a predefined location having a predefined file name. When market information is requested by the caller, the application logic reads the information from the database and uses its template files to construct on voice message. To accomplish this, it uses the pre-recorded template files, and dynamically generated voice information which basically done in this tier by text to speech engine. For our demonstration purpose we use three mostly spoken Ethiopian languages Amharic, Afaan Oromo and Tigrigna languages.

During implementation this project has shown how easy it is for services to be created in Asterisk. It has also illustrated that these services can be extended so that they are accessible from any interface or device, integrated with text to speech engine and that they can be expanded so that their functionality reaches deep into the system, allowing the users total control over their environment.

In addition, it has investigated and documented various open source systems that were used in achieving these goals, and shown that they are mature enough and stable enough to be used in the deployment of a production telephony system. The flow of MMI-IVR system looks like when the user interact by using telephone keypad with the asterisk server and then the asterisk server has dial plan which is used to define how calls flow into and out of the system. When messages are received on this channel to asterisk server, the message is sent to the MMI-IVR application for processing using Asterisk's Application Gateway Interface (AGI).

Finally, the MMI-IVR will interact with the user on the channel based on predefined IVR process tree on MMI-IVR application. Speech synthesis system built for the three languages in application logic tier. Based on prompt design, we have recorded directly to computer files. Most commonly we use a laptop (not connected to the mains power) in a quiet room to reduce background noise. All the prompts are recorded by female native speaker with the objective to meet the desired style for the synthesizer. We have used 108 sample utterances for Amharic speech synthesis in market information domain, 108 for Afaan Oromo language speech synthesis in market information domain and 108 for Tigrigna language speech synthesis in market information domain. The third tier of the architecture is data tier which consists of voice database and General information service for market database. Since concatenate synthesis uses actual short segments of recorded speech that were cut from recordings and stored in an inventory ("voice database"), either as "waveforms" (uncoded), or encoded by a suitable speech coding method, we prepared voice database for the proposed system.

In the future, it would be good to be able to give different information for different regions. Market information is one obvious topic that should be regionalized, but also perhaps which crops are suitable to grow in a specific area or what fertilizer should be used, which varies between different regions and climatic conditions. Regional information like this could support the farmers in enterprise selection.

REFERENCES

1. M.P. Plauche, U. Nallasamy, J. Pal, C. Wooters & D. Ramachandran. **Speech Recognition for Illiterate Access to Information and Technology**. In *proc. IEEE/ACM International Conference on Information and Communication Technologies and Development*. USA 2006.
2. M.P. Plauche & U Nallasamy. **Speech Interface for Equitable Access to Information Technology**. *Information Technologies and International Development*. 2007, pp. 69-86.
3. A. Sharma Grover, M. P. Plauche, C. Kuun, & E. Barnard, **HIV health information Access using spoken dialogue systems: Touchtone vs. Speech**, In *Proc. IEEE/ACM 3rd Int. Conf. on ICTD*. Doha, Qatar, Apr, 2009.

4. J. Sherwani, S. Paliyo, S. Mirza, T. Ahmed, N. Ali, & R. Rosenfeld. **Speech vs. touchtone: Telephony interfaces for information access by low literate users.** *In Proc. IEEE/ACM 3rd Int. Conf. on ICTD.* Doha, Qatar, Apr, 2009.
5. N. Patel, S. Agarwal, N. Rajput, A. Nanavati, P. Dave & T. Parikh. **A Comparative Study of Speech and Dialed Input Voice Interfaces in Rural India.** *ACM CHI*, Boston, Massachusetts, USA, April 4 - 9, 2009.
6. P. Nasfors. **Efficient Voice Information Services for Developing Countries**, MS. Department of Information technology. Uppsala University, Sweden, 2007.
7. E. Barnard, M. Plauche & M. Davel. **The Utility of Spoken Dialog Systems**, *In Proc. Spoken Language Technology for Development workshop, 2nd IEEE SLT workshop.* Goa, India, 2008.
8. F. Weber, K. Bali, R. Rosenfeld & K. Toyama. **Unexplored Directions in Spoken Language technology for Development**, *Proc. Spoken Language Technology for Development 2nd IEEE SLT workshop.* Goa. India, 2008.
9. K. M. Lee & J. Lai. **Speech versus Touch: A Comparative Study of the Use of Speech and DTMF Keypad for Navigation.** *International Journal of Human Computer Interaction*, 19 (3), 2005.
10. C. Delogu, A.D. Carlo, P. Rotundi & D. Sartori, **Usability evaluation of IVR systems With DTMF and ASR.** *In Proc. International Conference on Spoken Language Processing.* Australia, 1998.
11. M. Franzke, A.N. Marx., T.L. Roberts & G.E. Engelbeck, **Is Speech Recognition Usable? An Exploration of the Usability of a Speech-Based Voice Mail Interface.** *ACM SIGCHI Bulletin*, 25(3), 1993.
12. B. Suhm, J. Bers, D. McCarthy, B. Freeman, D. Getty, K. Godfrey & P. Peterson. **A Comparative Study of Speech in the Call Center: Natural Language Call Routing vs. Touch-Tone Menus.** *In Proc. 2002 Conference on Human Factors in Computing Systems*, (CHI-02), 2002.
13. TU. **International Telecommunication Union. ICT Development Index (IDI).**
14. Prasad, J.K., Kumar, B.A. **Analysis of SIP and realization of advanced IP-PBX features**, Vol 6, *IEEE*, 2011
15. **Towards Designing an Architecture for Delivering Distributed Agricultural Information Services for Developing Countries**, *In Proceedings of ICTD*, Atlanta, GA, USA, ACM, 2012