Design, Development and Implementation of an Automated IVR System with feature based TTS using Open Source Tools.

[1]Anil Kumar, [2]S. Niranjan
[1]P.hD scholar, CMJ University, Shillong, [2]Professor, PDM College of Engg, Bahadurgarh

Abstract

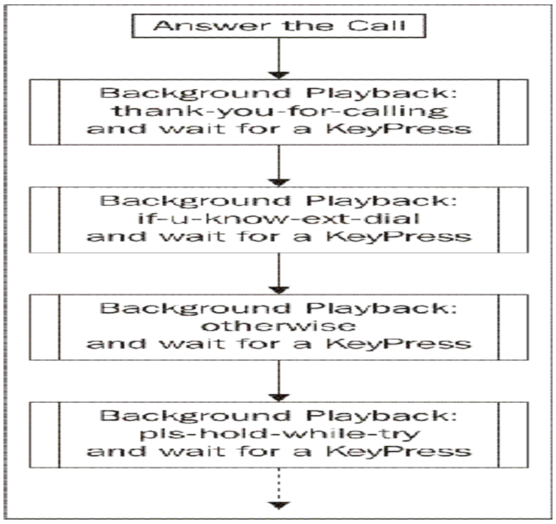
The Paper describes the concept of Interactive Voice Response System with feature based Text to Speech System using Open Source Software &Tools. The text to speech system converts the input text into the desired speech, the system is using for , to reading the user input text, speak aloud in various languages.

Interactive Voice Response (IVR) systems are complex network elements providing a comprehensive set of features and functionality in order to complete or even substitute human call agents. The IVR system is very often the only contact a caller has with a company when he requests a service, such as reserving a ticket for a movie. It is therefore very important that the IVR system provides high quality, in terms of robustness, stability, correctness of the menu branches and quality of the voice announcements. This White Paper discusses key objectives and requirements for the efficient and comprehensive testing of Interactive Voice Response (IVR) systems.

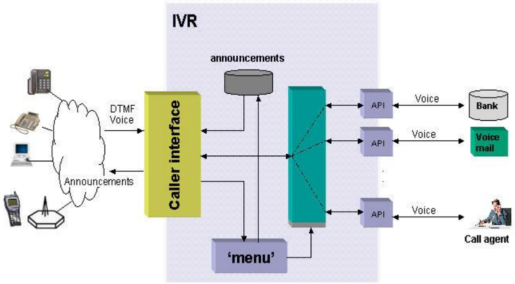
Introduction IVRs & TTS

IVRs (Interactive Voice Response systems) represent a powerful means for automating business and customer-facing processes. It is an automated telephony technology that is used to interact to the clients or customer through phone keypad or voice commands. . IVR systems process phone calls, play pre-recorded messages, provide callers with real-time data from any number of databases and potentially route calls to service agents. IVR technology requires virtually no human interaction over the telephone, as the user's interaction with the database is predetermined by what the IVR system will allow the user access to. For example, banks and credit card companies use IVR systems so that their customers can receive up-to-date account information instantly and easily without having to wait to speak with someone directly. Mostly it is used 24*365 days in the BPO, KPO, and Banking sector as well as in-house and offices to save money and employee resources where any emergency inquiry or service is not required. IVR system is very useful where some limited inquiry is asked like checking bank balances, managing credit cards, checking the timing of store hours or locations, or ordering the prescript medicine etc IVR systems can combine touch-tone input, speech recognition and text-to-speech capabilities, resulting in high customer satisfaction and operational effectiveness.

Text to speech System is most widely used system in speech technology. We have various text to speech synthesizer systems available like Festival, Multilingual and Flite etc. A Text-To- Speech (TTS) systems a computer-based system that should be able to read any text aloud, whether it was directly introduced in the computer by an operator or scanned and submitted to an Optical Character Recognition (OCR) system. Speech synthesis is a process where verbal communication is replicated through an artificial device. A computer that converts text to speech is one kind of speech synthesizer. In the business world, such situations are very common, especially for telephone transactions. Without text-to-speech (TTS) alternatives, business owners would have to spend money hiring even more customer service personnel. Synthesized solutions avoid this problem, since everything is done by computer, not a human being. Depending on the level of sophistication of the individual device, the sounds produced may be somewhat stilted and artificial sounding, or sound very much like the voice of a real person. The concept of speech synthesis has been


Answer the Call
Background Playback: thank-you-for-calling and wait for a KeyPress
Background Playback: if-u-know-ext-dial and wait for a KeyPress
Background Playback: otherwise and wait for a KeyPress
Background Playback: pls-hold-while-try and wait for a KeyPress

around for centuries, but only in recent decades has the process become available to the general public. It is the digitized audio rendering of computer text into speech.



TTS software can read text from a document, Web page or e-Book, generating synthesized speech through a computer's speakers. It can be used for variety of applications such as Email reader, text reader etc. using text to speech system. Text to speech systems are increasingly becoming an essential component of different type of computing systems. Also known as an artificial voice synthesizer, a text to speech system can produce human voice artificially based on a given string. Developing a text to speech system for a language that can support inputs in other languages can be helpful not only to the users know that language but are not familiar with its relative keyboard layout but also to international users that do not know that language at all and can hence type in that language using their local language keyboard layout.

Text To Speech Technology is a branch of Artificial Intelligence. Text To Speech Synthesis is a voice/ speech technology in which raw text is converted into audible speech. Text To Speech (TTS) is a process through which input text is analyzed, processed, and understood and then the text is rendered as digital audio and then spoken. The TTS are consisting of two major components:

- Natural Language Processing (NLP)
- Digital Signal Processing (DSP).

The process of TTS conversion allows the transformation of a string of phonetic and prosodic symbols into a synthetic speech signal. The quality of the result produced by a TTS synthesizer is a function of the quality of the string, as well as of the quality of the generation process. The most important qualities of a speech synthesis system are naturalness and intelligibility. Naturalness describes how closely the output sounds like human speech, while intelligibility is the ease with which the output is understood. The ideal speech synthesizer is both natural and intelligible. Speech synthesis systems usually try to maximize both characteristics. The two primary technologies for generating synthetic speech waveforms are concatenative synthesis and formant synthesis. Each technology has strengths and weaknesses, and the intended uses of a synthesis system will typically determine which approach is used. The basic types of synthesis system the following are:

- Formant, Concatenated & Prerecorded

Free & Open Source Software (FOSS) for IVR

Open-source software (**OSS**) is computer software that is available in source code form: the source code and certain other rights normally reserved for copyright holders are provided under a software license that permits users to study, change, improve and at times also to distribute the software, and which includes a license allowing anyone to modify and redistribute the software. Source code is the actual instructions which programmers write to create a piece of software, the "recipe" for the program. Once a program has been "compiled" into a form which can be installed and run on a computer, its source code is irretrievable.

It is practically impossible to make changes to a program without having a copy of its source code. If a program's license includes the right to modify the program, this right is meaningless unless the source code is readily available.

Following Tools/Applications required to setting up an IVR system?

S. No.	Basic Requirement	Optional Requirements
1.	Centos Linux operating system	Apache web server
2.	Asterisk	GUI Scripts
3.	Festival / eSpeak Text to Speech	IPtables / PHP
4.	MySQL database	Send Mail

Linux is a perfect operating system for Computer Telephony because:

1. It is freely available and easily accessible. Just download from the internet your favorite distribution!
2. It provides a reliable server platform suitable for background processes usually associated with telephony.
3. It is open source. You can recompile the Linux kernel or optimize it to run on a small footprint....useful for embedded applications!
4. It has excellent support for secure remote administration.
5. It has amassed extraordinary momentum in the software developer community and enjoys greater hardware vendor support.

For these reasons many enterprise-scale telephony systems and large PBXs run some form of UNIX based operating system like Linux.

Text To Speech (TTS)

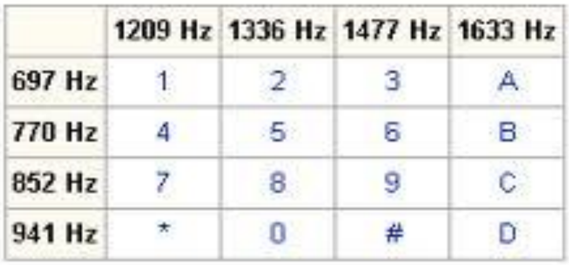
ESpeak : ESpeak is a compact open source software speech synthesizer for English and other languages, for Linux. eSpeak uses a "formant synthesis" method. This allows many languages to be provided in a small size. The speech is clear, and can be used at high speeds, but is not as natural or smooth as larger synthesizers which are based on human speech recordings. It includes different Voices, whose characteristics can be altered and can produce speech output as a WAV file with SSML (Speech Synthesis Markup Language) is supported (not complete), and also HTML.

Festival : Festival offers a general framework for building speech synthesis systems as well as including examples of various modules. As a whole it offers full text to speech through a number APIs: from shell level, though a Scheme command interpreter, as a C++ library, from Java, and an Emacs interface. Festival is multi-lingual (currently English (British and American), and Spanish) though English is the most advanced. Other groups release new languages for the system.

DTMF DECODER

This circuit detects the dial tone from a telephone line and decodes the keypad pressed on the remote telephone. The dial tone we heard when we pick up the phone set is call Dual Tone Multi-Frequency, DTMF in short. The name was given because the tone that we heard over the phone is actually make up of two distinct frequency tone, hence the name dual tone. The DTMF tone is a form of one way communication between the dialer and the telephone exchange. A complete communication consists of the tone generator and the tone decoder.

As technology matures, pulse/dial tone method was inverted for telephony communication. It uses electronics and computer to assist in the phone line connection. Basically on the caller side, it is a dial tone generator. In DTMF there are 16 distinct tones. Each tone is the sum of two frequencies: one from a low and one from a high frequency group. There are four different frequencies in each group. A normal telephone only uses 12 of the possible 16 tones. There are 4 rows and 4 columns. The rows and columns select frequencies from the low and high frequency group respectively. When a key is being pressed on the matrix keypad, it generates a unique tone consisting of two audible tone frequencies. For example, if the key '1' is being press on the phone, the tone you hear is actually consisting of a 697hz & 1209hz sine signal. Pressing key '9' will generate the tone form by 852hz & 1477hz. The frequency use in the dial tone system is of audible range suitable for transmission over the telephone cable.



	1209 Hz	1336 Hz	1477 Hz	1633 Hz
697 Hz	1	2	3	A
770 Hz	4	5	6	B
852 Hz	7	8	9	C
941 Hz	*	0	#	D

The exact values of the frequencies are listed below:

FIGURE: DTMF Table of Frequency Combinations

The tone frequency associated with a particular key is deciphered as follows. Each key is specified by its row and column locations. For example the "2" key is row 0 and column 1. Thus using the above table, "2" has a frequency of 770 + 1336 = 2106 Hz The "9" is row 2 (R3) and column 2 (C3) and has a frequency of 852 + 1477 = 2329 Hz. On the telephone exchange side, it has a decoder circuit to decode the tone to digital code. For example, the tone of 941hz + 1336hz will be decoded as binary '1010' as the output. This digital output will be read in by a computer, which will then act as a operator to connect the caller's telephone line to the designated phone line. The telephone exchange center will generate a high voltage signal to the receiving telephone, so as to ring the telephone bell, to notify the receiving user that there is an incoming call.

Conclusion

Interactive Voice Response Systems are an interesting technology for supporting mobility in modern IT projects. But using this technology needs additional thoughts about its restrictions. Not every use case can be usefuly expressed by voice dialogs. Voice processing and telephony technology is more expensive than traditional other technologies. But one main advantage remains: all you need is a phone. An IVR system is a powerful tool for increasing customer satisfaction, and it can help reduce the overall cost of a any office/call center etc while maintaining or even increasing the number of incoming calls. As we have seen that developers/testers, and in particular those that are new to telephony domain, have experienced difficulties while making IVR applications especially for above mentioned functionality of IVR applications. By using the right strategies and proven best practices as described using Open Source Tools in the paper, organizations can avoid difficulties that can result in financial losses and customer dissatisfaction.

Tentative Production Cost

By using Open Source software, telephony systems can be built for the price of a telephony card, a PC and a little effort. This can give your company a tremendous cost advantage over traditional business models that charge largely for their proprietary software. End-users can use this cost advantage to build and maintain their own low cost, high quality telephony systems.

Component	**Cost**
PC	20000
CTI Card	12000
Operating System (Linux)	Free
Application Software (Database, Dialer etc.)	Free
Total	**32000**

References

[1] Black et al., 2001 Black A, Taylor P, Caley R (2001), "The Festival speech synthesis system: system documentation". University of Edinburgh [Online] Available : http://www.cstr.ed.ac. ukprojects/festival/.

[2] open source telephony project Asterisk: http://www.asterisk.org

[3] Chopra D. , "Gayatri – A Fast Hindi Text To Speech System with Input Support For English Language", International Journal of Information Technology and Knowledge Management January-June 2011, Volume 4, No. 1, pp. 139-141

[4] Dutoit, "An Introduction to Text-to-Speech Synthesis," First edition, Kluwer Academic Publishers, 1996

[5] Ganapathiraju M., Balakrishnan M., Balakrishnan N., Reddy R., "Om: One tool for many (Indian) languages," Journal of Zhejiang University Science, vol. 6A, no. 11, pp. 1348–1353, 2005.

[6] Google Online Hindi to English Transliteration Tool, [Online] Available : www. google.com/transliteration.

[7] Free TTS engine: http://freetts.sourceforge.net/docs/index.php.

[8] K. Bali, A. G. Ramakrishnan, P. P. Talukdar, S. K. Nemela, "Tools for the development of a Hindi speech synthesis system," in 5th ISCA Speech Synthesis Workshop, Pittsburgh, USA, 2004.

[9] Balentine, B., and Morgan, D. P. (1999), How to Build a Speech Recognition application. Enterprise Integration Group, 59, 186, 244.

[10] Mukhopadhyay, A. Chakraborty, S. Choudhury, M. Lahiri, A. Dey, S. Basu, A., "Shruti- an Embedded Text-to-speech System for Indian Languages" Software, IEEE Proceedings, 153, Issue 2, April 2006, Page(s) 75–79.

Anil Kumar completed his Master degrees as MCA & MBA and currently doing his P.hD from CMJ University, has done also professional certification like MCSE, MCDBA, CCNA, RHCE(Tr.), presently working as System Engineer at PDM Engineering College, Bahadur garh, (Haryana). He has 10 years experience as Technical Support in **heterogeneous network** and involved in multiple research areas such as Open Source Implementations, Speech processing and various computing technology.