

Simmortel Voice Vocalised Services!

Using voice recognition technology to cater to customers speaking regional languages was, until now, an unexplored concept. Simmortel Voice offers a pioneering service.

From interactive voice response technology, the world has moved to voice recognition technology. Earlier, the user pressed the necessary phone keys and the relevant computerised voice response track played. The system did not require or register a user's speech. Voice recognition technology, however, picks up spoken keywords, which the computer can match, enabling it to access the requested information accordingly. Simmortel Voice has launched a platform that incorporates technologies that can recognise a caller's language and accent.

The company has developed a voice recognition engine, and voice recognition language models for Indian languages and accents. The engine and models developed by Simmortel Voice are meant to be used in mobile phone and telephone applications. The voice is carried over to the computer server connected to the telephone. Simmortel Voice hosts the server. It converts the voice signal to words, and then performs some action like database look-up, and plays back a response.

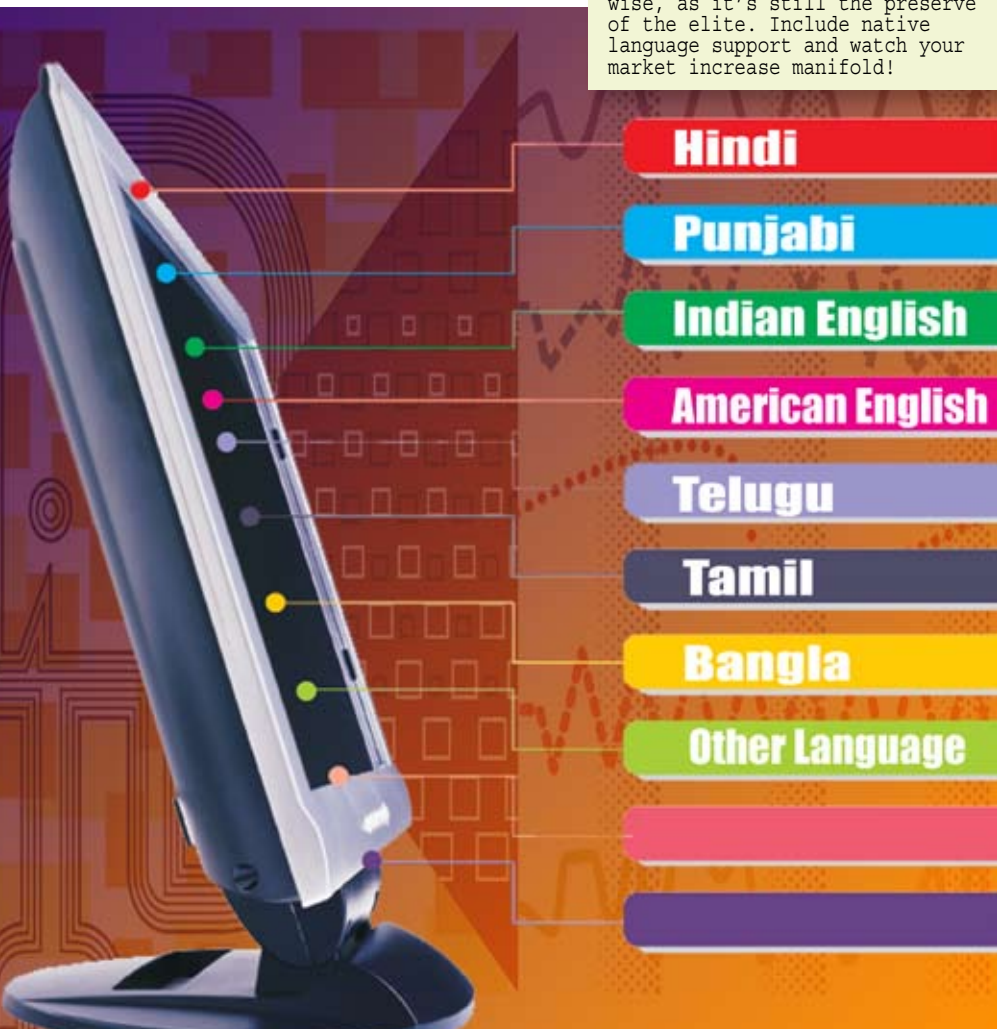
The technology

Voice recognition technology that converts voice signals into words using digital signal processing like Fourier analysis and statistical modelling, backs this platform. Another important aspect of this platform



The fortune cookie

If you want to tap a really large market, relying only on the English language might not be wise, as it's still the preserve of the elite. Include native language support and watch your market increase manifold!



“Doing everything from scratch for new languages is a technological barrier, which gives us a solid foundation for our business.”

Abhishek Singh, CEO and co-founder, Simmortel Voice

is the telecom grade server. It is a highly available server for processing millions of phone calls in a multi-tenant architecture using commodity class servers.

How does it work for Indian languages?

There are two aspects to speech: language and acoustics. The acoustics is studied with the help of Fourier analysis, statistical signal processing and hidden Markov models. The language part is all about pronunciation and phonetics.

Given a signal in the time domain, it is converted into its frequency domain representation using Fourier analysis and related transforms. The reason for transforming the signal into the frequency domain is for delineation of the distinguishing speech features. Information about pitch and intonation, which are speaker dependent, is filtered out while features rich in acoustics are retained. Thereafter, the features are further transformed to neutralise any remaining effects of pitch.

However, just having these features is not enough to apply a simple pattern-matching technique, because speech is a stochastic signal. Speech features vary a lot for the

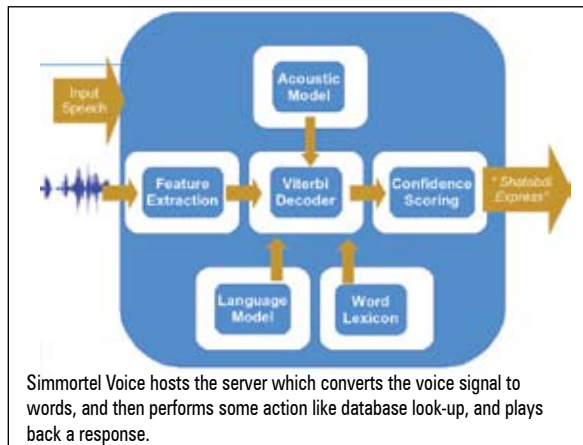
same speaker and the same speech. Therefore, a statistical model is built using these features from training data. This model is called the hidden Markov model, using which, we can predict if a given signal belongs to a particular model.

To make it work for Indian languages, one has to train the process for Indian languages. Acoustic models are built using

the phonetic knowledge of Indian languages, and language models are built using linguistic and statistical knowledge. Experts from linguistics, phonetics, acoustics, statistics, computer science and electrical engineering collaborate to make this entire process work.

“We study different ways in which people pronounce the same words; also, different phones are used. All this knowledge goes into making more robust acoustic models. This effort has to be repeated for each new language. That said, we are still not done with applications for dictation, which is truly language dependent. This is still in the R&D stage. Applications

that we are focusing on are actually not tied to a particular language. They are tied to more than one language. For example, names of trains, stations, people, places, businesses, movies, songs, etc, in India belong to many languages. We are making such applications possible,” confirms Abhishek Singh, CEO and co-founder, Simmortel Voice.



Call via the Simmortel voice engine

An example of how the voice portal will communicate with a farmer:

Computer: Sat Sri Akal. Je aaya noon. Jedi fasal da mandi rate chahida he, uhda naam bolo. (Greetings! Let me know the name of the crop for which you want to know the market price.)

Caller: Kanak (Wheat)

Computer: Mandi vich kanak da rate hai 1,200 quintal (The market price for wheat is Rs 1,200 per quintal.)

Computer: Tussi aane waale dina vich kehdi fasal beejan waale ho? (Which seeds are you planning to sow in the coming days?)

Caller: Makka (Maize)

Computer: Tuhanu ais fasal di jankari assi call karke dawaange. Thanyavaad. (We will call to let you know the information related to maize. Thank you!)

"An entrepreneur is always inspired by figures... We sensed the potential and the reach this device [mobile phone] has, and thus centred our innovation around this."

Abhishek Singh

"What we are doing has been done for American English around eight years ago. However, doing everything from scratch for new languages is a technological barrier, which gives us a solid foundation for our business," he adds.

The platform recognises your voice!

"Simmortel Voice is closely working with a client in the agricultural services industry, which has deployed a voice portal that enables the farmers to speak the name of a crop and get information about it over the phone. This application is language-specific and is currently available in Punjabi. The beauty of this technology is that it can be used from any landline or mobile phone. All the necessary computation is performed on the server side, which is hosted by Simmortel," says Singh.

For language-specific applications, Simmortel Voice can currently support Hindi, Punjabi, Indian English and American English (developed for US customers). Languages under development are Telugu, Tamil and Bangla.

Not every application is specific to a particular language. Examples of applications, which are neutral to languages and accents, are railway queries, travel queries, SMS dictation, local city search, directory listings, etc. These applications are the hardest

to build because of the variation in phonetics and accents.

For an example of an application that is not language-specific, think of the popular helpline-number 131 or 1331 of the Indian Railways. This number is currently limited by a key-press interface. It asks you to punch in the number of the train. However, you may not remember the number of the train. Or you may want to know which trains go from, say, Kolkata to New Delhi. Such a query is not possible with most key-press interfaces. Therefore, we need voice recognition.

With the help of the Simmortel Voice platform, a caller can make queries about 'Rajdhani Express', 'Kolkata to New Delhi', 'Status of Shatabdi Express today', etc. These queries are converted to words on the computer server, which proceeds to perform an appropriate action. This application is not language-specific, as it depends on the names of the trains that are common in all languages.

The founders shared that OnMobile could be a possible rival, though they are the first to offer such a service in India. TellMe and Free411 are other such services available in the USA.

Acknowledging innovation

There are many innovative aspects of this platform, both

technological and business-related. The first technological innovation is that it is India's first multi-grammar voice recognition technology based on a considerable vocabulary, to the credit of Simmortel R&D. It supports vocabularies as large as 5,000 words for Indian languages and various accents. The second innovation is that it includes India's first multi-tenant architecture for voice recognition and the IVR platform hosted on computing class servers to provide telecom-grade availability.

The business innovation lies in providing a hosted platform for content providers instead of asking them to invest upfront in buying infrastructure. And also in hosting voice portals on numbers that can be called at local rates by end users.

They get to voice their opinion

Simmortel's direct customers are those who want to set up a voice portal. Currently, setting up a voice portal will cost above Rs 2 crore for the necessary hardware and software infrastructure—computer servers, telephony hardware, telecommunication signalling equipment, operating systems, application software, IVR software and voice recognition software. With the Simmortel Voice platform, there is no initial set-up cost because the company has already set up the necessary infrastructure. Businesses employing this platform can program the behaviour of this platform over a Web-based interface using VoiceXML standards. Simmortel Voice, for a small set-up fee, gives businesses an easy-to-

remember number that customers/clients can dial up at local tariffs to know about the services offered. While a caller dials in to listen to the business's audio content (information), Simmortel Voice plays an advertisement and shares the revenues with the business/institution.

Typical target customers include:

- Media houses like television channels, radio, newspaper, magazines that want their viewers to connect with them.
- Providers of local search information and directory listings.
- Those who provide weather forecasts, stock quotes, cricket scores, music, ring tones and caller tunes over the phone.
- Travel services like airlines, taxi services, the Indian Railways.
- Absolutely anybody who wants to set up a voice portal where people can call them up to ask for and share information.
- Directory assistance and tourist information services providers.
- Call centres with small traffic.

In collaboration with the Indian Railways, Simmortel has launched the beta version of the railway enquiry service platform that will inform passengers about the trains that ply a route, their availability, how late a train is running, etc.

The groundwork of the company

Abhishek Singh, an alumnus of IIT Kanpur, and S. Umesh, professor of electrical engineering, IIT Kanpur, founded the company in September 2007. The professor is a world-renowned expert in voice

A simple scenario of someone making an enquiry for a train

Caller: (Rings up the number)

Phone: (Ring Ring) "Welcome." "What is your source station?"

Caller: New Delhi

Phone: "What is your destination station?"

Caller: Kolkata

Phone: "For which date do you require a booking?"

Caller: 28th July

Phone: Please confirm. From Kolkata to New Delhi on 28th July? Say Yes or No?

Caller: Yes

Phone: You have seats available in Rajdhani Express, Shatabdi Express etc.

recognition technology, with over 15 years of experience in the field. Abhishek Singh is his student. They began working together on this platform in 2006.

They have one R&D office at IIT Kanpur and one development centre in Noida. Three patent applications have been filed by the founders. "We use open source software only, and contribute back whenever possible. Lack of technological manpower and resources were the major challenges that the firm faced during the initial phase," reveals Singh.

"We have had tremendous success with our voice calls for farmers in Punjab who receive information about weather forecasts and weather insurance, daily. We have recorded more than 30,000 calls received per month. Even the response to the railway service enquiry platform, which is being beta-tested in Kanpur, has been quite significant. On

an average, we receive more than 10,000 calls per month," explains Singh.

A great future

"An entrepreneur is always inspired by figures. The facts and figures regarding the number of phone users in India motivated the team to innovate in this field. 250 million is a good number and this accounts for the number of phone users in India. We sensed the potential and the reach this device has, and thus centred our innovation around this," shares Singh.

The entire process has been very challenging for the team. Getting the platform to recognise different accents, languages, vocabularies, and using the telecom grade platform and commodity class servers were all tough jobs.

The team firmly believes in innovating and bringing forth every possible technological innovation to the common people so that it pervades everyday lives and solves real problems. The R&D is in full swing at Simmortel and the future road map is to make this platform capable of supporting ten Indian languages and five international languages. The team is also working towards enabling this platform to support large vocabulary dictation of SMS and voicemails. **IT**

Cholena Deb

i.t.' Bureau