

# 3G and Speech: Connect the Masses to the Web

Nuance India.

**Abstract—** This document provides a brief on how new technologies and speech interfaces can change the way of life for the masses. The use of speech as the primary interface to access the web through mobile phones and handheld devices, in combination with 3G in India, can bring a new phase in the way people will access services, content and social networking sites. Moreover, e-Governance can ride on speech and other multi-modal interfaces as India has a significant proportion of people who are illiterate, semi-literate, and the impatient ‘tweeting’ younger generation. Speech interfaces will accelerate, enrich and simplify in connecting these masses to the web.

**Keywords—** Speech Interface, ASR, TTS, Verification, Nuance

## I. MOTIVATION

India’s broadband penetration is less than 1% where as the mobile penetration is 45%<sup>1</sup>. India has the largest illiterate population of any nation on earth<sup>2</sup>. Bubbly, voice based twitter gained 5lakh users in just four week of launch<sup>3</sup>. Laptops and desktops will soon be rarity and mobile devices that can do e-mail, browsing, gaming and social networking will be the norm. Indian Government is adding new ways of efficiently and effectively delivering the citizen centric services using ICT. Speech, being the most natural and user friendly interface, along with 3G services, can kick start connecting a very large population to the web.

## II. TECHNOLOGY AND THE STANDARDS

There are three components of the speech technology. Automatic Speech Recognition (ASR), Text-To-Speech (TTS), and Speaker Verification (SV). 3G communication technology is largely new to India. 3G allows simultaneous use of speech and data services. The data rate is up to 14Mbps/s downlink and 5.8Mbps/s uplink. W3C provides several standards for interoperability and easier application development – VoiceXML for voice only and X+V profile for Multimodal applications. IETF provides MRCP (Media Resource Control Protocol) for accessing speech technology resources seamlessly in a distributed network. In addition Nuance provides several other complementary components like SDKs to create multimodal mobile applications of small footprint that use speech resources in the network. Nuance also provides the ability to leverage both embedded speech and speech resources in the network allowing a seamlessly handoff between device computing and cloud computing. One such ability is through the use of Nuance Mobile Speech Platform (NMSP) VoiceXML, ASR, TTS and MRCP v1 are already in the Indian market for several years. But, the full capabilities of these technologies are still in their infancy. The combination of Speech technology and 3G can act as catalyst in bringing the web to the masses.

## III. OPENING THE FLOODGATE

The applications of the technology and standards mentioned above are unlimited. The advantage of NMSP is that it leverages the creativity of the thousands of active application developers community who keep churning out dynamic and contextually relevant applications.

Citizen information services can be provided using speech interfaces as part of e-Governance.

Voice verification can be used for unique identification and recording attendance. Voice verification can eliminate the cases of fake attendances in social schemes like NREGA.

Speech interfaces can be used to leave voice tweet messages using either data or voice channel. The backend systems use ASR to transcribe the voice tweet message to text message and send it as an SMS to their friends and followers.

Multimodal interfaces can be used to search and download the entertainment content like music, trailers and videos on the phone with the 3G. Web search, YouTube search and driving directions are just one click away using the speech, with the help of ASR and TTS. 3G enables these services by reducing the latency in downloading the content.

People can use speech to dictate their messages in social networking sites. Others can listen to it with the help of TTS. Users can listen to their friends messages and postings handsfree while they are driving.

Speech is the most effective interface for blind people. They can easily browse on their mobile by just speaking to it and listening to the output with the help of TTS.

## IV. CONCLUSION

It is quite apparent that the mobiles and handheld devices will be the choice of future communication. 3G will be universally available in the Indian market. W3C standards will ensure that all the devices and network resources are interoperable. Speech will work over all these to ensure that the web reaches to all the sections of the population.

## REFERENCES

- [1] <http://economictimes.indiatimes.com/infotech/internet/DoT-wants-govt-cos-to-drive-wireless-broadband-play/articleshow/5720844.cms>
- [2] [http://en.wikipedia.org/wiki/Literacy\\_in\\_India](http://en.wikipedia.org/wiki/Literacy_in_India)
- [3] <http://digital.venturebeat.com/2010/03/09/bubbly-bubble-motion/>