# SPEECH PROCESSING

Prayank Sehgal[1], Rahul Kumar Jain[2]
[1]Dept. of Electronics and Communication, Truba Institute of Engineering and Information Technology, RGPV, Bhopal (M.P.), India
prayanks26@gmail.com
[2]Dept. of Mechanical Engineering, Oriental College of Tech., R.G.P.V, Bhopal (M.P), India
rj.1105@hotmail.com

## ABSTRACT

*Speech processing is the study of speech signals and the processing methods of these signals and can be regarded as a special case of digital signal processing, applied to speech signals. The greatest step towards this goal is voice recognition technology, where humans can communicate in a way that is natural to them through speech. People with disabilities can benefit from speech processing programs. For individuals that are deaf, dumb or disabled, speech recognition and synthesis software is used to ease communication with machines as well as humans. When a person speaks, he creates vibrations in the air. The analog-to-digital converter (ADC) digitizes this sound by taking precise measurements of the wave at frequent intervals. The system filters the digitized sound to remove unwanted noise, and sometimes to separate it into different bands of frequency. It also adjusts it to a constant volume level (Normalizing). People don't always speak at the same speed, so the sound must be adjusted to match the speed of the template sound samples already stored in the system's memory. Then the signal is divided into small segments. The program then matches these segments to known phonemes in the user known language. A universal translator is still far into the future, a system has not yet been built that combines automatic translation with voice activation technology. One problem is making a system that can flawlessly handle roadblocks like slang, dialects, accents and background noise. The different grammatical structures used by languages can also pose a problem. Advancement in speech processing with parallel developments in image processing, touch technology and augmented reality can also lead to development of robust Natural User Interfacing (NUI) systems which can help humans communicate much more easily to present day complex machinery.*

## I. INTRODUCTION

Speech processing is the study of speech signals and processing methods of these signals. The signals are usually processed in a digital representation, so speech processing can be regarded as a special case of digital signal processing, applied to speech signal. The various aspects of speech processing include the acquisition, manipulation, storage, transfer and output of digital speech signals.

It is also closely tied to natural language processing (NLP), as its input can come from or output can go to NLP applications. For example, a computer may accept a command or inputs verbally from the user; or synthesize speech as output, instead of graphics. The main applications of speech processing are recognition, synthesis and compression of human speech.

Speech processing includes the following domains:

- Speech recognition, which deals with analysis of the linguistic content of a speech signal and its conversion into a computer-readable format, which is like understanding the speech for the computer.
- Speaker recognition, where the aim is to recognize the identity of the speaker.
- Speech coding, a specialized form of data compression usually used in telecommunication.
- Speech synthesis: the artificial synthesis of speech, which usually means computer-generated speech. Advances in this area improve the computer's usability.

## II. SPEECH RECOGNITION

Speech recognition is the process of translation of spoken words into computer readable format, which clearly means into the binary language. To convert speech to on-screen text or a computer command, a computer has to go through several complex steps. When a person speaks, he creates vibrations in the air, which are analog in nature. An analog-to-digital converter (ADC) translates this analog wave into digital data that the computer can understand. To do this, it samples or digitizes the sound by taking precise measurements of the wave at frequent intervals. In order to process speech, the system filters the digitized sound to remove unwanted noise, and sometimes to separate it into different bands of frequency. It also normalizes the sound, or adjusts it to a constant volume level. It may also have to be temporally aligned.

People not always speak at the same speed, so the sound must be adjusted to match the speed of the template sound samples already stored in the system's memory. The signal is, thus, further divided into small segments, as short as a few centiseconds, or even milliseconds in the case of plosive consonant sounds. Moreover, the software examines phonemes in the context of the other phonemes around them. It runs the contextual phoneme plot through a complex statistical model in the software library and compares them to a large library of known words, phrases and sentences. The program then determines what the user was probably saying and either outputs it as text or issues a computer command.

## III. SPEAKER RECOGNITION

Once the speech signals are fed into the system, they can be used for various purposes. One such purpose is to recognize the voice or the speaker. This can be achieved by comparing parameters like pitch which is indicated by frequency with the frequencies stored in the system. Similar is the case word recognition, where the words spoken by the speaker are compared with the already stored data.

Such systems are usually used as security systems or identification systems where actual recognition of people is a necessity. Such systems can also be used as an aid to other speech processing operations.

## IV. SPEECH CODING

Speech coding is the process of data compression of digital audio signals containing speech. Speech coding uses speech-specific parameter estimation using audio signal processing techniques to model the speech signal, combined with similar data compression algorithms to represent the resulting modeled parameters in a compact bit-stream. Speech coding is mainly used in telecommunication processes to strengthen signal transmission and reception and improve signal to noise ratio.

## V. SPEECH SYNTHESIS

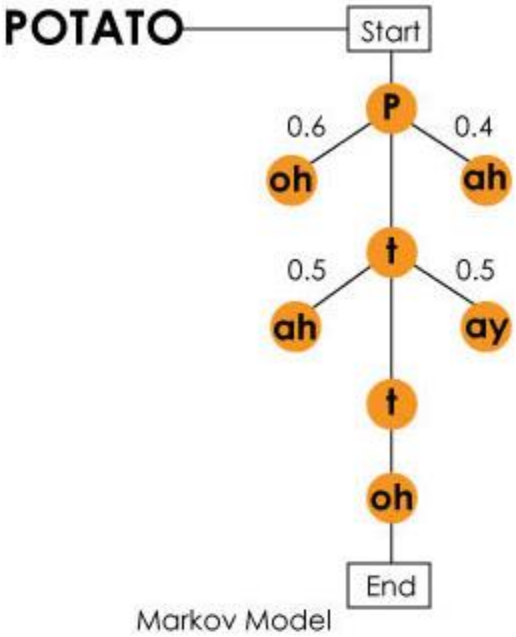
Speech synthesis is the artificial production of human speech. A computer system used for this purpose is called a speech synthesizer, and can be implemented in software or hardware. A text-to-speech system converts symbolic linguistic representations like phonetic transcriptions into a complete speech.

Synthesized speech can be created by concatenating pieces of recorded speech that are stored in a database. Systems differ in the size of the stored speech units; a system that stores phones or diaphones provides the largest output range, but may lack clarity. For specific usage domains, the storage of entire words or sentences allows for high-quality output. Alternatively, a synthesizer can incorporate a model of the vocal tract and other human voice characteristics to create a completely "synthetic" voice output.

## VI. STATISTICAL MODELING AND USER TRAINING

Today's speech recognition systems use powerful and complicated statistical modeling systems. These systems use probability and mathematical functions to determine the most likely outcome. According to John Garofolo, Speech Group Manager at the Information Technology Laboratory of the National

Institute of Standards and Technology, Maryland; the two models that dominate the field today are the Hidden Markov Model and neural networks. These methods involve complex mathematical functions, but essentially, they take the information known to the system to figure out the information hidden from it.



However, this is a highly complicated process. If a program has a vocabulary of 60,000 words (common in today's programs), a sequence of three words could be any of the 216 trillion possibilities. Even the most powerful computer cannot search through all of them without some help. That help comes in the form of program training both by users as well as users.

Statistical systems need lots of exemplary training data to reach their optimal performance. These training data are used to create acoustic models of words, word lists, and multi-word probability networks. While the software developers who set up the system's initial vocabulary perform much of this training, the end users must also spend some time training it. They must also train the system to recognize terms and acronyms particular to their organization. This makes the system much less prone to error while in usage.

## VII. CURRENT CHALLENGES TO SPEECH PROCESSING SYSTEMS

In the signal processing section, the challenges faced by a speech processing program are low signal to noise ration, intensive usage of computer power identification of speech by one person from various on-going voices.

But even bigger challenges are in the linguistic section. If the words spoken fit into a certain set of rules, the program could determine what the words were. However, human language has numerous exceptions to its own rules, even when it's spoken consistently. Accents, dialects and mannerisms can vastly change the way certain words or phrases are spoken.

**A classical example of such a situation is:**

**r eh k ao g n ay z s p iy ch**
**"recognize speech"**
**r eh k ay n ay s b iy ch**
**"wreck a nice beach"**

A program may easily get confused between these two above mentioned phrases; when the same words are spoken by two different people. The two phrases have completely different meanings for humans.

## VIII. FUTURE OF SPEECH PROCESSING

### *a) A Universal Translator:*

The Defense Advanced Research Projects Agency (DARPA), Virginia is already funding a research and development effort called TRANSTAC to enable soldiers to communicate more effectively with civilian populations in non-English speaking countries.

If such a system can be developed for all languages, one can easily communicate to anyone in the world with the language barrier removed. One day, one might speak on the phone in English to a person who understands only French because the message will processed to be delivered in French.

However, a universal translator is still far into the future. The problem faced is making a system that can flawlessly handle roadblocks like slang, dialects, accents and background noise. The different grammatical structures used in various languages can also pose a problem.

### *b) Natural User Interfacing:*

A Natural User Interface, or an NUI, is the common phrase used by designers and developers of human-machine interfaces to refer to a user interface that is effectively invisible, or becomes invisible with successive learned interactions, to its users, and is based on nature or natural elements. The word natural is used because most computer interfaces use artificial control devices whose operation has to be learned. An NUI relies on a user being able to quickly transition from novice to expert.

Being a natural mean of interaction with the environment for humans makes speech processing an integral part of human computer interactions. With furthermore advancement in speech processing technology, combined with parallel developments in touch technology, holographic touch, image processing and development of more robust, powerful and fast computer systems; the way in which humans interact with computers can be revolutionised.

A user interface can be formed, which would make technology so invisible that it would not even register to the conscious mind while under usage. At its peak, such an interface would be so natural for humans, that many would not even realise that it is not a natural part of our environment.

## ACKNOWLEDGEMENT

I would like to acknowledge Mr. Nelesh Gupta, Head Electronics And Communication Department, TrubaInstitute of Engineering And information Technology, Bhopal (M.P.) for his continuous guidance and support throughout this Research.

## REFERENCES

[1]. Fundamentals of Speech Recognition; Lawrence Rabiner & Biing-Hwang Juang Englewood Cliffs NJ: PTR Prentice Hall (Signal Processing Series), c1993, ISBN 0-13-015157-2

[2]. Hidden Markov models for speech recognition; X.D. Huang, Y. Ariki, M.A. Jack. Edinburgh: Edinburgh University Press, c1990

[3]. Electronic speech recognition: techniques, technology and applications, edited by Geoff Bristow, London: Collins, 1986.

[4]. Roe, David B., and Jay G. Wilpon. Whither speech recognition: the next 25 years. IEEE communications magazine, v. 31, Nov. 1993: 54-62.

[5]. Rudnicky, Alexander I., Alexander G. Hauptmann, and Kai-Fu Lee. Survey of current speech technology. Communications of the ACM, v. 37, Mar. 1994: 52-57.

[6]. Baker, J. (2005, August 30). Milestones in speech technology - past and future!. Speech Technology Magazine.

[7]. Flanagan, J. L. (2005) "Perspectives on the Evolution of Speech Technology", in Eurospeech 2005 - Interspeech 2005. Proceedings of the 9th European Conference on Speech Communication and Technology. 4-8 September, 2005. Lisbon, Portugal.

## AUTHOR'S BIOGRAPHY

**Prayank Sehgal**, pursuing Bachelor of engineering in Electronics and Communication fromTruba Institute of Engineering And information Technology affiliated to R.G.P.V Bhopal (M.P.) He has got past experience of research and has got International Research Papers Published. He has worked in depth in the field of bionics in engineering and concluded essential output with his research. His area of interest is in information system. **prayanks26@gmail.com**

**Rahul Kumar Jain**, pursuing Bachelor of Engineering in Mechanical engineering from Oriental College of Technology, affiliated to R.G.P.V, Bhopal (M.P). He has worked in the application of theories of operation management in different disciplines and has got published International research paper on the same. His area of Interest is to ameliorate the efficiency of an organization firm provide goods or services via management in all the aspects.
**rj.1105@hotmail.com**