International Journal of Electrical and Computer Engineering (IJECE)
Vol. 9, No. 1, February 2019, pp. 281~288
ISSN: 2088-8708, DOI: 10.11591/ijece.v9i1.pp281-288

The role of speech technology in biometrics, forensics and man-machine interface

Satyanand Singh
School of Electrical and Electronics Engineering, Fiji National University, Republic of Fiji

Article Info

Article history:

Received Apr 13, 2018
Revised Jul 16, 2018
Accepted Aug 19, 2018

Keywords:

Artificial intelligence (AI)
Gaussian mixing model (GMM)
Man-machine interface (MMI)
Natural language processing (NLP)
Natural language understanding system (NLU)
Universal background model (UBM)

ABSTRACT

Day by day Optimism is growing that in the near future our society will witness the Man-Machine Interface (MMI) using voice technology. Computer manufacturers are building voice recognition sub-systems in their new product lines. Although, speech technology based MMI technique is widely used before, needs to gather and apply the deep knowledge of spoken language and performance during the electronic machine-based interaction. Biometric recognition refers to a system that is able to identify individuals based on their own behavior and biological characteristics. Fingerprint success in forensic science and law enforcement applications with growing concerns relating to border control, banking access fraud, machine access control and IT security, there has been great interest in the use of fingerprints and other biological symptoms for the automatic recognition. It is not surprising to see that the application of biometric systems is playing an important role in all areas of our society. Biometric applications include access to smartphone security, mobile payment, the international border, national citizen register and reserve facilities. The use of MMI by speech technology, which includes automated speech/speaker recognition and natural language processing, has the significant impact on all existing businesses based on personal computer applications. With the help of powerful and affordable microprocessors and artificial intelligence algorithms, the human being can talk to the machine to drive and control all computer-based applications. Today's applications show a small preview of a rich future for MMI based on voice technology, which will ultimately replace the keyboard and mouse with the microphone for easy access and make the machine more intelligent.

Corresponding Author:

Satyanand Singh,
School of Electrical and Electronics Engineering,
Fiji National University, Fiji Island, Republic of Fiji.
Email: yogitechno@gmail.com

1. INTRODUCTION

There are many convincing arguments to support the continuous development of voice-based man-machine interface. Most of the protagonists cite the intrinsic "naturalness" of the voice interfaces in which the skills of the spoken language acquired by the users as Infants can easily be recruited to understand the information provided by the text output to the speech synthesizer, to control the equipment by talking to an automatic speech recognition system, or to access information by conversing with a spoken language dialogue system [1], [2], [3]. Even those who question the naturalness of such interactions still admit that the voice channel has the potential to offer real application benefits in operating environments without hands and without sight, where even a wrong man-machine interface can improve transfer rates information as competing for interface technologies.

However, in recent years there has been a significant convergence of the methods and techniques used to develop the man-machine interaction based on the word and the data statistical modeling paradigm (such as HMM-based acoustic modeling, n-gram-based language modeling, concatenative speech synthesis) dominated the research agenda. Of course, this convergence of modeling paradigms has emerged because of the real improvements in the quality and performance of the system that these approaches have provided over a period of nearly three decades. The principle of defining a model, estimating its parameters from the sample data, then implementing this model as a mechanism of generalization in unprecedented situations is irreproachable and the use of statistical methods represents one of the most powerful and effective tools available for the scientific community for such modeling [4], [5], [6]. The only problem is that the amount of training speech data needed to improve state-of-the-art speaker recognition systems seems to grow exponentially (despite the relatively low complexity of the underlying models) and system performance appears to be asymptotic at a level that may be inadequate for many real-world MMI applications [7], [8]. Furthermore, the current speech technology is quite fragile, even on a fairly positive day conditions; not only contemporary automatic speech/speaker recognition is so scarce to recognize and understand highly accented or colloquial speech, but the speech generated by the machine lacks individuality, expression and the communicative intent and the dialogue systems of the spoken language are rigid and inflexible.

Forensic and ASR research communities have developed several methods for at least seven decades independently. In contrast, native recognition is the natural ability of human beings which is always very effective and accurate. Recent research on brain imaging has shown many details that how a human being does cognitive-based speakers recognition, which can motivate new directions for both automated and forensic system [9], [10].

Voice interface technology, which includes automatic speech recognition, synthesized speech, and natural language processing, includes the knowledge areas required for the man to machine communication. In the near future, man-machine communication applications will surely grow with only voice-based, increasing the need for natural language processing technology to enhance speech interpretation. Automatic speech recognition is the power of machines that interpret the speech to execute commands or generate text. An important related area to make machine smarter is automatic speaker recognition, which is the ability of machines to identify an individual based on the voices.

2. SPEECH TECHNOLOGY BACKGROUND

During early 1970, many attempts were made to invoke knowledge of the structure and behavior of spoken language in order to develop practical systems of human-machine interaction. It was the era of the "Human speech analysis system" and it was assumed that the classical principles of phonetics and linguistics could be used to interface machine with human being to make electronic system more reliable. Practical results were almost universally disappointing with the best system that used less phonetic and linguistic knowledge. Since then, the perceived value of every intuition in the human process has greatly diminished.

ASR systems and synthetic speech technology often require the use of high speed computer hardware resources, ASR technology is essentially software based. Advanced digital signal processors are used by all smartphones and tablets, but some speech systems only use analog / digital converters and general purpose computer hardware. As reported in [11], [12] voice recognition is the ability to identify the words and phrases of an electronic machine or program, spoken language and converting them into a machine-readable form. The basic characteristics of a speech recognition software-enabled system is that it has a limited vocabulary and can only be read and execute when someone speaks very clearly. More sophisticated artificial intelligence ASR system has the ability to accept natural spoken voice of an individual. Speech recognition applications include voice search, call routing, voice dialing and speech-to-text, speaker verification, speaker recognition. There are three broad categories of services used for speech recognition application: (a) Automated serving (b) Routing of incoming call (c) Value added services. The accuracy of the speech recognition system depends on the language and the voice model [13], which are mainly produced, i.e., these models need to analyze parallelism with spoken voice samples. In the same way, the speaker recognition system is necessary to create a large selection of words and phrases while creating and refining the current language and acoustics of the model [14].

2.1. Uses of speech technology functionality in smartphone devices

Although there is no clear definition of what a smart phone device is, it can be said that a smart phone is a device which increases the capabilities of traditional mobile terminal devices. A smartphone is expected to have a more powerful CPU, more storage space, more RAM, faster connectivity options and larger screen than a regular cell phone. New smartphones are equipped with innovative sensors such as accelerometer and gyroscopes. Accelerators provide a screen display in portrait and landscape mode, while

the gyroscope makes smartphones for games to support motion-based navigation. Five major features of smart electronic systems are intelligent sensing, automation, remote accessibility, awareness and learning. Google uses artificial intelligence algorithms to identify a spoken sentence, store anonymously for the analysis of voice data, and uses cross-match data with written queries on the server. The problems with computational power, information availability and the management of large amounts of information are making use of Android speech recognizer Intent package [15]. The current smartphone is using the client app and the user wants to log in using Google speech recognition. Google server receives audio data as input for processing and text is sent back to the client. Input text is transmitted to Natural Language Processing (NLP) server for processing using HTTP (HperText Transfer Protocol) POST. Figure 1 shows that the steps of data flow diagram in the speech recognition system NLP as (i) Lexical analysis converts character sequence into token sequence. (ii) Morphology analysis defines, analyzes, and describes the structure of language units of a particular language. (iii) Syntactic analysis analyzes the text made from a series of markers to determine grammar structures. (iv) Semantic Analysis relates syntactic structures from the levels of phrases and sentences to their language-independent meanings.

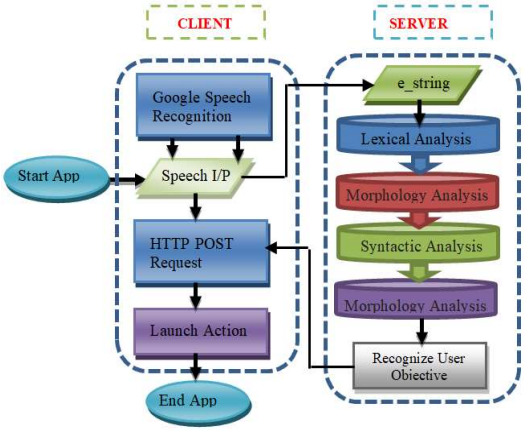

Figure 1. Natural language processing data flow diagram in man-machine interface

2.2. Future man-machine interface (MMI) through voice technology

MMI with speech technology have been a dream of technologists for several decades. But in recent years, due to some noticeable advances in machine learning, voice control has become very practical. By speech enhancement and noise suppression technique no longer limited to just a small set of predetermined voice commands, it now works even in a noisy environment you feel that speaking across a room. Virtual operating voice assistants such as Apple's Siri, Microsoft's Coratana and Google now are bundled with the largest number of smartphones, and it is an easy way to look at information in new gadgets like Amazon's Alexa, to sing songs and their build lists of spending with the voice. Smartphones are more common than desktops or laptops, yet surfing the web, sending messages and doing other activities can make the pain slow and frustrating. Andrew NG says, “This is a challenge and there is a chance, in 2008, under MIT Technology review innovators, was nominated for work in artificial intelligence (AI) and robotics at Stanford. “Instead of being able to train people by desktop computers for new behaviors suitable for mobile computers, many of them can learn the best ways to start a mobile device from the beginning”. It is believed that the voice can soon be reliable enough to interact with all types of devices. For example, robots or smart electronic devices can be easily managed by MMI.

Jim Glass, a senior MIT scientist who has worked on vocal technology, believes that time can finally be right for voice control. They say, the speech technology has reached a turn in our society. In my experience, when people can talk with the device instead of a remote control, they want to do it. In future, I want to talk to all of our devices and understand them. I hope that one day you can say “Hello” to your microwave oven; you will get a reply “Hi” what do you like to have?. After the advent of artificial intelligence, voice and more commonly language based technologies like Chatbot, Siri and Amazon Echo, MMI is the best possibility of becoming the next important technical platform after mobile devices. There are many promises in the field of MMI conversation that how human beings interact with technology, thanks to such trends: Increased contact with mobile devices, which are small screens in nature which can make graphic elements difficult to display. Demand for abolishing friction as a way to obtain consumer demand and/or to gain profit more quickly and easily. Increasing messaging applications for real-time communication

between multiple users. Once the evolving technologies like speech recognition, the understanding of natural language, intent and expression synthesis is getting more refined and more than being planted in production.

2.3. Future man-machine interface (MMI) through voice technology

There are some key features that make MMI applications based on effective speech technology. (i) It should be really colloquial -A good interactive MMI uses a natural language that is human and shares conversation control. It means not only answering questions, but using machine learning, give appropriate suggestions. It should be done individually as a conversation on one to one. The voice of the interactive user interface should be both personal and private. Directing a user by name, for example using the language that passes through the analysis of emotion to match the emotional state of the user. (ii) It should be right sympathetic-MMI should show individual personal sympathy, how the user can feel the information presented. Understand the situation and respond accordingly. For example, a status update that "Your current account has been canceled" is not indicated in a bright and happy voice. (iii) It should maintain context and story-A strong interactive MMI refers to the conversation and is able to take lead or answer on the basis of previous questions where you are?, who are you?, what are you doing? etc. It should be transferred from one request to another and customized as needed. (iv) It should be accurate and consistent to gain confidence-Along with human contacts, a level of trust between the user and the interactive user interface should be established. A good interactive user interface is accurate and consistent, not only on the information provided, but also at the level of understanding displayed by the interactive user interface response, but also a level increase in confidence with the user.

Growing, vocal engines that "give machines a human voice" are integrated with ASR System and software for understanding human language which is called the Natural Language Understanding system (NLU). Together, it make complex circuits that allow humans to interact with machines in natural language is shown in Figure 2.

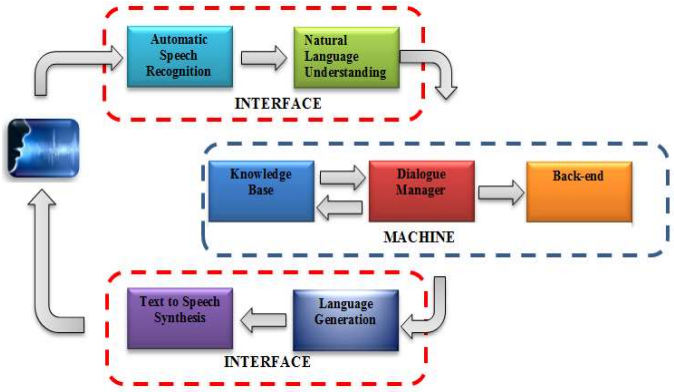

Figure 2. Block diagram representation of speech synthesis and man-machine interface

3. FEATURE EXTRACTION AND MODELING ALGORITHMS FOR MMI APPLICATION

ASR is a mathematical algorithm based computer system designed to recognise the voice of a speaker operated independently with minimum human intervention. The ASR system admin can adjust algorithm parameters, but to compare between speech segments, all users have to provide speech signal to the ASR system. In this paper, we concentrate our attention on the text-independent ASR system and the speaker verification. As mentioned earlier, humans are good in differentiating voiced and non-voiced signal that is the important part in auditory forensic speaker recognition. Obviously, in ASR it is desirable that the speaker-specific feature can only be extracted from the voiced speech signal by voice activity detection (VAD) [16]. Detection and feature extraction from speech segment is important when considering the condition of excessive noise/degraded speech signal. Recently used VAD algorithm is explained in although more accurate unsupervised solution has emerged as successful in various ASR applications in diverse audio condition [17].

Short-term speaker specific feature in ASR application shows the parameters extracted from the short segment of speech signal within 20-25 ms. In ASR application the most popular short-term acoustic features reported are the Mel-frequency cepstral coefficients (MFCCs) [18] and linear predictive coding (LPC) based features [19]. Steps involved in to obtain MFCC feature from speech signal are (i) Divide

speech signal into short overlapping form (25 ms). (ii) Multiplication of these segments with Hamming and Hanning window function to get Fourier power spectrum (iii) Apply logarithm of the spectrum (iv) Apply nonlinear Mel-space filter-bank to obtain spectral energy in each channel (24 channel filter bank) (v) Apply discrete cosine transform (DCT) to obtain MFCC. As previously indicated, the specific speaker feature is the desirable qualities of the acoustic feature are robustness to degradation. The features normalization is one of the desirable characteristics of an ideal feature parameter [20].

When there is no prior knowledge of speech content in text-independent speaker recognition tasks, it has been found that Gaussian Mixuture Model (GMM) applications are more effective for acoustic modeling to shape short-term functionality. The average behavior of this is expected short-term spectral features are more dependent on speakers than being influenced by the temporary features. Therefore, even when the test data of ASR has a different acoustic situation, then due to GMM being a potential model it may be related to better data than the more restrictive Vector Quantization (VQ) model. A GMM is a mixture of Gaussian probability density functions (PDFs), parameterized by a number of mean vectors, covariance matrices, and weights of the individual mixture components. The template is a weighted sum of individual PDFs. The density of the Gaussian mixture is the weighted sum of M component densities and it represented mathematically:

$$p(\vec{x}|\lambda) = \sum_{i=1}^{M} p_i b_i(\vec{x}) \quad (1)$$

Where $\vec{x}$ represents D-dimension random vectors, component densities $b_i(\vec{x}), i = 1,..,M$, and mixture weight represented by p_i. Each component density is a D voriate Gaussian function of the form

$$b_i(\vec{x}) = \frac{1}{(2\pi)^{D/2}\ |\sum_i|^{1/2}} \exp\left\{-\frac{1}{2}(\vec{x} - \vec{\mu}_i)' \sum_i^{-1} (\vec{x} - \vec{\mu}_i)\right\} \quad (2)$$

$\vec{\mu}_i$ represents mean vector, $\sum_i$ represents covariance matrix. The complete density of the Gaussian mixture is parameterized by the mean vector, covariance matrix and mixture components of all density. These parameters are represented collectively by signaling

$$\lambda = \{p_i, \vec{\mu}_i, \sum_i\}\ i = 1,..,M \quad (3)$$

For ASR system, each speaker is represented by one by the GMM and is referred to by his/her model λ. The size of GMM may vary depending on the choice of covariance matrix. The GMM model can be evaluated using the probability of a vector attribute in (1).

An SVM is a binary classifier that makes its decisions by constructing a linear decision boundary or hyperplane that optimally separates the two classes. Depending on its position in relation to Hyperplane, the model can be used to predict the class of unknown observation. Let us consider training vector and labels as (x_n, y_n) , $x_n \in \Re^d$, $y_n \in \{-1, +1\}$, $n \in \{1, ...T\}$ the optimal hyperplane is chosen according to the maximum margin criterion then target of SVM can be learn the function $f: \Re^d \to \Re$ so that the class labels of any unknown vector x can be expected as $I(x) = \text{sign}(f(x))$.

For linearly separable data labeled [21], hyperplane H can be obtained from $x^T x + b = 0$, which separates the two class of data, so that $y_n(w^t x_n + b) \geq 1$, $n ....T$. An optimal linear divider H provides maximum margins between classes, i.e. the distance between H and the training of two different sections is highest in the data estimates. The maximum margin is found in the form of $\frac{2}{\|w\|}$ and data points x_n for which $y_n(w^t x_n + b) \geq 1$ that the margin is known as super vectors. When ASR training data is not linearly separable, then speaker specific features can be mapped to a higher dimensional space, in which kernel functions are linearly divided.

The purpose of the FA is to describe variability in high dimensional observable data vector using less number of unobservable/hidden variables. For ASR application, the idea of explaining s peaker and channel-dependent variability in the GMM supervector space, FA has been used in [22]. Many forms of FA methods have been employed since, which ultimately brought the current state of the art i-vector approach. In a linear distortion model, a speaker-dependent GMM supervisor m_s is generally considered as four component which are linear in nature.

$$m_{s,h} = m_0 + m_{spk} + m_{ch} + m_{res} \quad (4)$$

Where m_0 speaker, channel, environment-independent component is, m_{spk} is speaker dependant component, m_{ch} is channel environment dependant component and m_{res} is residual.

The joint FA (JFA) model is prepared in conjunction with eigenvoice and eigenchannel, which is achieved with a MAP optimization for a model. The sub-spaces are aligned by V and U matrix, as the first model recommends for an informal choice of speakers s and sessions h, mean supervector of GMM can be represented by

$$m_{s,h} = m_0 + U_{Xh} + V_{ys} + D_{Zs,h} \quad (5)$$

So now this is the only model, which we are considering all the four components of linear distortion model we discussed earlier. In fact, JFA has been shown to overcome other current method.

4. FUTURE ADVANCES IN SPEECH AND SPEAKER RECOGNITION FOR MMI

At present, robots working in Japan and the United States are android projects; Facial expression or mirroring, it is very popular for the target human that interacts with the system to create emotional bond with the machine. Speech recognition systems that teach body language and facial expression can also be used to evaluate the danger, for example the replacement of human workers at the airport, border crossings and such places or obstacles.

4.1. Body language facial expression and voice recognition

Speech recognition systems that are capable to read body language and facial expression can also be used to evaluate the danger, for example the airport, border crossings and the replacement of human workers in such places or obstacles. If you smile on the robot Android and smile at you, then you are talking, it enhances the sentimental value of interaction with humans. Perhaps the system can start praising you, if you have been convinced by the system, it will probably reflect the answer or the anger would have to be repaired or the work to spread the situation, obviously it all depends on its programming, But you can see progress, potential applications and future trends

If you remember Hell in a well-known science-fiction computer, then he said, "I declare hostility in your Dave voice." Probably once it was in science fiction job, today's human scientists are trying to make it this way. Right now, with this technique, speech recognition software can see sentiment, hesitation, aggression, hostility, anger etc. So, within five years we will see these features in more and more applications.

Haptics is another field of science, which lends fusion to well between emotional recognition of facial recognition and facial features. Perhaps the future robots will look human and imitate their characteristics, a robot that joins a strong hand and feels a firm grip with the voice of a person's self-confidence with a soulful ego, by a stepping stone or two can pick up the aspect.

4.2. Emulation of emotion and empathy

Imagination and empathy is coming now. At present, most artificial call centers intelligent customer feedback system advisors recommend that the sound from the other side, if coming from the machine, should be easily identified by humans who call the system, because the computer with speech recognition functions Humans do not like to cheat, when they find out, it annoys them, of course, emotional emulation or sympathy It is possible with the passage and now we have the ability to do this.

In fact, artificial intelligent computers are used to go online and participate in forums and can take up to 15 threads or more without detection. In speech recognition, if the voice sounds legitimate, then the entire conversation may continue for a time, without the person knowing that he is talking to a machine.

A call center system that manages the complaint, an IT system can be a part of the client and can hear it and even say it; "I know how you feel, I'm sorry that it happened, let me see what I can do"; "Yes, I think it is very important, I will talk to you with my supervisor" So the customer should send it to a real human system or maybe someone else, with a more official voice? On the second line, the client never knows whether to talk to a computer or a computers, in fact, it does not go very well with many industries, but it is a place where speech recognition software professionals are thinking and now discussing, of course, you can see the application for it.

4.3. Smart enough to understand humor and respond

Artificial Intelligence (AI) is always improving, soon, AI software engineer will create fun recognition systems, in which the computer will be able to understand the irony and when the human is saying fun, then repay with a joke, maybe making a joke, jokes for scratches For human interaction in all

cultures, the system should be pre-loaded with all the common jokes. He will be able to select the one who cannot be heard most by the man working with that time; it also remembers that this person has been asked by the person so that he does not repeat it.

Wow, This is becoming slightly complicated, it is not like that, and that's why it's not fully realized. Humor is a major obstacle for human speech recognition and artificial intelligence systems, but it is a talent for some people, however, they are working on this challenge and we will see it in 5-10 years, people of artificial intelligent software Licking will be a problem. This means the progress for long-term space flight for the human partner means helping with rehabilitation and reducing the stress of humans working with colleagues or robot assistants, such as the transition of robots and human workers. Because robots will work with humans and will help humans, it will be necessary to maintain peace to promote cooperation.

4.4. Vocal cord vibration recognition and current voice recognition system

At present, there is an advanced search in the US military that allows you to read the vocal cord, without sound or voice, these systems are now working; it is done with a device near the signaling Gathers, which is connected to a transmitter to send. Any other member of the receiver or special force has a small earring so that he can listen to that speech, all those silent surrounding which are within six inches using the system. It is very close to copying the idea of transfer, but in short it is a form of speech recognition, which is connected to a communication device. These systems will be better and soon the secret services members, Special Forces, SWAT teams will now have small strings not coming out of their ears, but they communicate without warning. Vibrational flirting of the Larrynx can be increased within the “clip tie” and no one will be sensible. If you think about it then there are many applications for it.

5. MMI APPLICATION POSSIBILITIES WITH SPEECH TECHNOLOGY

The availability of computer processing power and network connectivity in cars and mobile terminal devices is the result of for the explosion of applications and services available to users. One of the potential services using a mobile device while driving, though the voice recognition function is used. Automotive environment for speech recognition is one of the toughest environments. It is important to reduce driver's view and physical commitment due to possible intervention in those cases such as car occupants and their conversation, background music or similar background noise, wind, noise of windshield wiper etc. For these and other reasons, cars and equipment manufacturers invest in improving and optimizing voice recognition applications suited to the specific environment of the car. Looking at the above, high quality microphones have been installed, as well as a technique which reduces the noise. Applications are improved using specific acoustic environments for the automotive environment [23]. Voice is one of the natural methods of MMI [24]. Speech recognition skills are rapidly developed and used in the automotive industry. It is not surprising that the competitiveness of the modern car market depends on their technical characteristics and innovations.

There are following areas where we can see more development of MMI based on speech recognition based technology in near future. Access of mobile terminal devices with MMI by speech technology, Access of navigation system with MMI by speech technology, Access and control of Car on-board system with MMI by speech technology, Operation and control of mechanical machine with MMI by speech technology

Smart terminal devices have become increasingly popular with the development of hardware segments and with the new features generated using the increasing number of sensors. In any case, an important smartphone app is likely to have voice recognition and processing of such information/orders. There are many possibilities for the development of applications for modern intelligent terminal devices due to the specificity of the individual mobile operating system, different applications that allow at least some speech functions to be recognized for greater or lesser extent have developed. The purpose of these solutions is to develop software that provides all the tasks that speech can be used only interface for input and output data for machine.

6. CONCLUSION

This paper gives an overview of what MMI has to offer and showed a glimpse of what the future might hold. One thing is certain technologies are starting to converge, devices combine functionality, new levels of sensor fusions are created and all of this for one purpose, to improve our interaction with human machines. The technology involved in MMI is quite incredible. However, MMI still has a long way to go, for example, Nanotechnology has provided a new exemption from progress, but these still need to be fully used in MMI, nanotechnology has an important future role to play. The nano-machines and super-batteries have not completely functional, so we have something to look forward to MMI application. There is also the potential for Quantum Computing which will release a new processor level, with incredible speeds. MMI technology is impressive now, but there will not be anything like it in the future. No matter who you are,

what language you speak or what your disability is, the variety of technology will satisfy everyone. In the near future, we will see prostheses with higher functions, more interfaces for brain computers, speech recognition and recognition of the most used camera gestures. Although this is not exactly the death of the mouse and keyboard every day, we will certainly begin to see new types of technologies incorporated into our daily lives. Portable devices are becoming smaller and more complex, so we should start seeing growth in portable interfaces. The robots and the way we interact with them are already starting to change, we are in the computer age, but soon we will be in the age of robotics.

REFERENCES

[1] S.Singh, "Forensic and Automatic Speaker Recognition System," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 8, 2804-2811, October 2018.

[2] S.Singh and Dr. E.G. Rajan., "Vector Quantization Approach for Speaker Recognition Using MFCC and Inverted MFCC," *International Journal of Computer Application*, vol. 17, pp. 1-7, March 2011.

[3] S.Singh and Dr. E.G. Rajan, "MFCC VQ Based Speaker Recognition and Its Accuracy Affecting Factors," *International Journal of Computer Application*, vol. 21, pp. 1-6, May 2011.S.Singh and Ajeet Singh., "Accuracy Comparison using Different Modeling Techniques Under Limited Speech Data of Speaker Recognition Systems," *Mathematics and Decision Sciences*, vol. 16, pp.1-17, 2016.F. Jelinek, "Five Speculations (and a Divertimento) on the Themes of H. Bourlard, H. Hermansky, and N. Morgan," *J. Speech Comm*, vol. 18, pp. 242-246, 1996.

[6] S.Singh and Dr. E.G. Rajan., "Application of Different Filters In Mel Frequency Cepstral Coefficients Feature Extraction And Fuzzy Vector Quantization Approach In Speaker Recognition," *International Journal of Engineering Research & Technology*, vol. 2, pp. 3171- 3182, June 2013.

[7] E. Keller., "Towards Greater Naturalness: Future Directions of Research in Speech Synthesis," Improvements in Speech Synthesis, E. Keller, G. Bailly, A. Monaghan, J. Terken, and M. Huckvale, eds., *John Wiley & Sons*, 2001.

[8] Fergyanto E. Gunawan, Kanyadian Idananta, "Predicting the Level of Emotion by Means of Indonesian Speech Signal," *Telecommunication Computing Electronics and Control* (*TELKOMNIKA),* vol.15, pp. 665-670, June 2017.

[9] Eriksson, "Tutorial on Forensic Speech Science," *in Proc. European Conf. Speech Communication and Technology*, pp. 4-8,2005.

[10] P. Belin, R. J. Zatorre, P. Lafaille, P. Ahad, and B. Pike, "Voice-selective Areas in Human Auditory Cortex," *Nature*, vol. 403, pp. 309-312, Jan. 2000.

[11] Prather, M, "Understanding Speech Recognition Technology," *SpeechRec 101: Colla Voice Consulting*, San Francisco, CA, United States of America, 2012.

[12] S.Singh, Mansour. H. Assaf and Abhay Kumar, "A Novel Algorithm of Sparse Representations for Speech Compression/Enhancement and Its Application in Speaker Recognition System," *International Journal of Computational and Applied Mathematics*, vol. 11, pp. 89-104, 2016.

[13] S. Singh, Abhay Kumar, David Raju Kolluri, "Efficient Modelling Technique based Speaker Recognition under Limited Speech Data, " *International Journal of Image, Graphics and Signal Processing(IJIGSP)*, vol. 8, pp.41-48, 2016.

[14] Sukmawati Nur Endah , Satriyo Adhy , Sutikno, "Comparison of Feature Extraction MFCC and LPC in Automatic Speech Recognition for Indonesian," *Telecommunication Computing Electronics and Control* (*TELKOMNIKA)*, vol. 15, pp. 292-298, March 2017.

[15] Agarwal, A., Wardhan, K., Mehta, P, "A Natural Language Processing Application for Android," *JEEVES* http://www.slideshare.net, 2012

[16] F. Beritelli and A. Spadaccini, "The Role of Voice Activity Detection In Forensic Speaker Verification," *in Proc. Digital Signal Processing*, pp. 1–6, 2011.

[17] S. O. Sadjadi and J. H. L. Hansen, "Unsupervised Speech Activity Detection Using Voicing Measures And Perceptual Spectral Flux," *IEEE Signal Processing Letters*, vol. 20, pp. 197-200, March 2013.

[18] S.Singh , Assaf Mansour H, Abhay Kumar and Nitin Agrawal, "Speaker Recognition System for Limited Speech Data Using High-Level Speaker Specific Features and Support Vector Machines," *International Journal of Applied Engineering Research (IJAER)*, vol. 12, pp. 8026-8033, 2017.

[19] H. Hermansky, "Perceptual Linear Predictive (PLP) Analysis of Speech," *J. Acoust. Soc. Amer*, vol. 87, pp. 1738-1752, April 1990.

[20] Douglas Reynolds, *et al.*, "The Super SID project: Exploiting high-level information for high-accuracy speaker recognition," *in Proc. IEEE Acoustics, Speech, and Signal Processing*, pp. 784-787, 2003.

[21] S.V.S.Prasad, T. Satya Savithri, Iyyanki V. Murali Krishna, "Comparison of Accuracy Measures for RS Image Classification using SVM and ANN Classifiers," *International Journal of Electrical and Computer Engineering (IJECE),* vol. 7, pp. 1180-1187, 2017.

[22] P. Kenny and P. Dumouchel, "Disentangling Speaker and Channel Effects In Speaker Verification," *in Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, pp. 37-40, 2004.

[23] S.Singh, "Support Vector Machine Based Approaches For Real Time Automatic Speaker Recognition System," *International Journal of Applied Engineering Research*, vol. 13, pp. 8561-8567, 2018.

[24] Koolagudi, S. G., Rao, K. S, "Emotion Recognition From Speech: A Review," *International Journal of Speech Technology* 15, pp. 99-117, 2012.