ISSN:2321-1156

International Journal of Innovative Research in Technology & Science(IJIRTS)

Development of English Language Isolated Words' Corpus and Views on Recognition Methods

[1]V. K. Kale, Department of Computer Science & IT, Dr. Babasaheb Ambedkar Marathwada University, Aurangabad (MS) India

[2]R.R.Deshmukh, Department of Computer Science & IT, Dr. Babasaheb Ambedkar Marathwada University, Aurangabad (MS) India

[3]G. B. Janvale, Symbiosis Centre for Information Technology, Symbiosis International University, Pune (MS) India

Abstract:

This paper reviews the English language isolated words databases which are collected from native and non - native speakers of English language throughout the world. It also reports the various methods that are used in English speech recognition system. The second part of the paper discusses the design and development of English isolated speech database. The data base contains the English language isolated spoken words corpuses which were recorded through PRAAT software from 20 regional Non –native speaker of English language including males and females from Marathwada region of Maharashtra state in India.

Keywords: **English language accent and phonetics, Speech Data base and Speech Feature.**

1. Introduction

English language is used in an education, business and science [1]. So, it has global demand. The English speech recognition is also become a popular and active area of research [2]. The speech recognition is used to translate the words spoken by humans so as to make them computer recognizable [3]. The recognition usually involves the extraction of patterns from digitized speech samples and representing them using an appropriate data model [4]. Most of the speech-recognition systems are classified as isolated or continuous speech. Isolated words recognition requires a brief pause between each spoken word, whereas continuous speech recognition does not. So that continuous speech recognition is very difficult. It can be further classified as speaker- dependent and independent. These patterns are subsequently compared to each other using mathematical operation to determine their contents [5]. The main challenges in speech recognition involve modeling the variation of the same word as spoken by different native and non native speakers of English language. The recognition rate is also varied in speaking styles, accents, regional and social dialects, gender, voice patterns etc. The paper is organized as follows. Section 2 describes the history of English language. The section 3 explains the speech database from non-native Speaker of English language. Section 4 describes the speech features used to study the English language. Different future extraction techniques are presented in section 5. Finally, section 6 and 7 give development of speech corpus database from non – native speakers of English.

2. English Language

English is a Germanic language which belongs to the Ind-European languages. The question of the original home of the Ind-Europeans has been much debated, but nowadays most scholars agree that the original group of people that spoke Porto-Ind-European, the language which would later split into a number of branches, including the Germanic branch, live somewhere between the Black Sea and the Caspian Sea some 6000 years ago[6]. Most scholars believe that this population then expanded eastward, westward and northward thereby came to inhabit most of Europe and parts of Western Asia.

2.1 Accents

Accent is one of the most important characteristic of speakers. It is manner of pronunciation of a language. In case of the speakers are non native, carry the intonation phonological process and pronounces rules from their mother tongue in their English speech [7]. The word is used in two different senses. The way, accent refers to importance given to a language unit, usually by the use of pitch. For example, in the word 'potato' the middle syllable is the most salient; if you say the word on its own you will probably produce a fall in pitch on the middle language unit, making that syllable accented. In this sense, accent is distinguished from the more general term stress, which is more often used to refer to all sorts of importance or to refer to the effort made by the speaker in producing a stressed language unit. And second, accent also refers to a particular way of pronouncing for example, you might find a number of

English speakers who all share the same grammar and vocabulary, but pronounce what they say with different accents such as European country [8].

2.2 English Phonetics

A fundamental distinctive unit of a language is a phoneme. Different languages contain different phoneme sets. Syllables contain one or more phonemes, while words are formed with one or more syllables, concatenated to form phrases and sentences. One broad phoneme classification for English is in terms of vowels, consonants, diphthongs, and semi vowels. Phonetics is the scientific study of speech. The central concerns in phonetics are the discovery of how speech sounds are produced, how they are used in spoken language, how we can record speech sounds with written symbols and how we hear and recognize different sounds[9][10].

3. Speech Database

Agnes Stephenson et at. Have presented a paper in which they compared the performance of English language learners (Non-Native of English speakers) and native speakers of English on the basis of scores of an English language proficiency test. The experiments were conducted on 200 students. Random samples were collected from theses students. It is found that the primary and elementary test level of Stanford English learning proficiency was not reliable according to discriminate analysis method [11]. Yemi olagbaj et at. Have collected the speech samples from ten Mandarin and ten Hindi speakers from the University of Bridgeport without having speech and hearing problem, to study formant frequencies in speech. All speakers were age of matched with the rage of 22 -25 years. They were given eleventh words of English the speech corpus to pronounce [12]. Laura and Tokyoite collected the sample from 12 native speakers of Japanese; all of them had lived in the United States for less than 1 year. All had studied English for a minimum 8 years in Japan. The all speakers were, between the age of 20 and 40, and educated. The recording was done in a quiet room with close talking microphone. They had been given reading tasks. The first task consisted of transcriptions of utterance produced by native and non – native speakers. The database was used to study the linguistic properties of non native speech [13]. Qingqing Zhang et at. Have also developed database from Non English speakers. All the data were recorded and digitized at 16 KHz sampling rate 16 bits resolutions. The data base was developed to study different variations of non –native pronunciations variations of English language [14]. Bishnu Prasad Das and Ranjan Parekh have designed and developed the data set consists of 280 speech samples recorded by 28 speakers each uttering the name of 10 digits, from 0 to 9, in English. Out of 28 speakers 14 were male and 14 female. The speech samples are recorded directly over microphone in a controlled environment. All the audio signals are stored in the WAV format with sample rate of 22050 Hz, bit rate of 16 bits and in mono format. They have used different features extraction techniques to study the accent and phonology of English isolated spoken words [15]. Tanjin Taher Tomal and Abu Hasnat have developed the system for recognition of English language for Bengali regional speakers. The accents of English have been examined both from the orientation of native and non Native of language. Various research results show that non-native speakers of English language produce certain speech identifying which are special in native speakers' speech. This is because non-native speakers do not produce the same tongue movement as native speakers. Here, they have detected a different speech diagnostic in which two acoustic features i.e. pitch and formants have been utilized to develop the steelmaker – independent and stands on template based approach.

4. Speech Feature

4.1. Pitch

Pitch refers to the sensed fundamental frequency of complex speech signal. It is produced due to the vibration of the vocal folds. It depends on the tension of the vocal folds and the sub speech organ air pressure when speech is generated. Pitch in human voice is dependent on the length and thickness of the vocal cord as well as the tightening and relaxation of the muscles surrounding them [16]. Since women possess shorter vocal cord than that of the men, they generally have higher pitch value than men do. Hence, pitch-contour of an utterance is very useful for gender identification.

4.2. Formants

The term Formant [17] [18] [19] refers to peaks in the harmonic spectrum of a complex sound. In speech science and phonetics, formants are referred to acoustic resonance of a human vocal tract. The spectrum of a speech signal may consist of several formants, but the first three have great significance in speech recognition. Formants are estimated mainly when pronouncing a vowel, because recognition of vowels based on them, is easier and gives better result. These formants are quite similar for the same vowel different fundamental frequencies, enabling it to be recognized regardless of the pitch. In contrast, they are completely different for dissimilar vowel sounds. Hence, words containing dissimilar vowel sounds can be easily detected based on the difference of formant values.

4.1 Mel Frequency Cepstral Coefficients (MFCC)

The speech signal is often assumed to be the output of a system. It is the convectional of the output and in pulse response. The most useful information for phone detection is filters i.e. exact position and shape of vocal tract. The separate source of filter efficient mathematical way is spectrum. The cepstrum is defined as the Descript Fourier Transform (DFT) of the log magnitude of DFT of signal. The cepstral coefficients have the extremely useful property that the variance of different coefficient tends to be uncorrelated. The first step in MFCC feature extraction [20] [21] [22] [23] is to boost the amount of energy in the high frequencies. It turns out that if we look at the spectrum for voice segment like vowels. There is more energy at the lower frequencies than the higher frequencies. This drop in energy across frequencies which is called spectral tilt is caused by nature of the glottal pulse

5. Feature Extraction Techniques

5.1 Hidden Markov Model

A Hidden Markov Model [24] [25] [26] (HMM) is a type of random model appropriate for non stationary random sequences, with statistical properties that undergo distinct random transitions among a set of different stationary processes. The HMMs are suitable for the classification from one or two dimensional signals and can be used when the information is incomplete or uncertain. To use a HMM, need a training phase and a test phase.

5.2 Dynamic Time Warping (DTW) Method

DTW [27] [28] algorithm is most important algorithmic in speech recognition. For different speech samples of given word will have somewhat different durations in time series. This problem can be eliminated by simply normalizing the templates and the unknown speech so that they all have an equal duration in time series. Another problem is that the optimal alignment between a template and the speech sample may be nonlinear. DTW is an efficient method for finding this optimal nonlinear alignment. DTW is an instant of general class of algorithm known as dynamic programming. Its time and space complexity is merely linear in the duration of the speech sample and vocabulary size. The algorithm make single pass through a matrix of frame scores while computing locally optimize segment of the global alignment path.

5.3 Linear Predictive Coding (LPC)

LPC [29] [30] is tool used for analysis of speech signals. In speech coding, the success of LPC have been explained by the fact that all pole model is a reasonable approximation for transfer function of vocal tract. All pole model is also appropriate in them of human hearing, because the ear is more sensitive to spectral peaks than spectral valley, Hence an all pole model is useful not only because it may be physical model for signal. The advantages of LPC are that the way in which LPC is applied to the analysis of speech signal leads a reasonable source vocal Tract separation and LPC is analytically tractable model.

6. Development of Speech Corpus form Non native speaker of English Language

For development of a Speech database, the basic requirement is the grammatically correct Text corpus which would be recorded from various speakers. The text corpus should be correct in terms of composition and grammar.

6.1. Speaker Selection

All the speakers were the college students, including 10 males and 10 females between the age groups of 18 to 25 from Aurangabad city of Maharashtra state. They were given some English words to pronounce. Before that, they were trained and comfortable with speaking the English language with datasets.

6.2. Data Collection:

The isolated English words were recorded through normal microphone. The isolated words were selected as a dataset which are frequently used in communication. The 'PRAAT' software was used for recording database with the sampling rate of 44 kHz and 16 bit. Each sample was recorded ten times. The distance between mouth and microphone was adjusted nearly 30 cm. The recorded data were stored in WAV file format, the corpus details is shown in table 1. Recording was done in noisy environment.

Table 1: English corpus collected from Regional people

Sr. No	Set of words	Occurrences	Male Speakers	Female Speakers
1	Hello	10	10	10
2	Please	10	10	10
3	Help	10	10	10
4	Okay	10	10	10
5	Come	10	10	10
6	By	10	10	10
7	Yes	10	10	10

8	Sad	10	10	10
9	Happy	10	10	10
10	New	10	10	10
			1000	1000
Total Number of Words				2000

7. Conclusion:

Initially, we have studied the basic of English language and different accents and phonology, and have reviewed the literature of different databases which are collected from various non - native speakers of English. We have also explained the development of corpus spoken by Marathi speakers from Marathwada region of Maharashtra state. The recorded corpus will be used to study of phonetics and accents of native and non native speakers of English. The third phase of this paper we have study the recognition of English isolated word technique like DTW and HMM, also some method we have discuses for feature extraction. This complete paper shows the ideas regarding the database collection and method used for feature extraction and recognition. Our futures work is to recognize the English isolated word and find out the variation between native and non -native speaker.

8. Acknowledgment:

This work is supported by University Grants Commission on Major Research Project. The author would like to thank the University Authorities and Department of Computer Science and IT for providing the infrastructure to carry out the research.

9. References:

[1] Yen – Minkhaw and Tien-Ping Tan, "Pronunciation Modeling for Malaysian English", International Conference on Asian Language Processing, DOI 10.1109/IALP 2012.72, 978-0-7695-4886-9/12. 2012.

[2] Chitralekha Bhat, K.L. Srinivas, and Preeti Rao, "Pronunciation Scoring for Indian English Learners using a phone recognition system", ACM 978-1-4503-0408-5/10/12, 2012.

[3] Liu Xiao-Feng, Zhang Xue-ying, and Duan Ji-Kang , "Speech Recognition Based on Support Vector Machine and Error Correcting Output Codes", IEEE Computer Society, 978-0-7695-4180-8/10, (2010) DOI 10.1109./PCSSPA 2010.

[4] Sheguo Wang, Xuxiong Ling, Fuliang Zhang, and Jianing Tong, "Speech Emotion Recognition Based on Principal Component Analysis and Back Propagation Neural Network", IEEE Computer Society, 978-0-7695-3962-1/10, DOI 10.1109/ICMTMA 2010.

[5] Tristan Kleinschmidt, Michael Tason, Eddie and Sridha Sridharan, "The Australian English Speech Corpus for In-car Speech Processing", IEEE ICASSP 2009, 978-1-4244-2354-5/2009

[6] Albert Croll Baugh and Thomas Cable, "A History of the English Language", Rowledge, ISBN 0415093791, 9780415093798, 1993.

[7] Joshi M., Iyer M. and Gupta N., "Effect of Accent on Speech Intelligibility in Multiple Speaker Environment with Sound specialization", IEEE Computer Society, ISBN 978-0-7695-3984-3/10 2010.

[8] Shamalee Deshpande, Sharat Chikkerur, and Venu Govindaraju, "Accents Classification in Speech", IEEE Computer Society, 0-7695-2475-3/05 2005.

[9] Elizabeth Hume and Keith Johnson, "The Impact of Partial Phonological Contrast on Speech Perception", Proceeding in the 15th International Congress of Phonetic Science 2003.

[10] Agenes Stepheson, Hong Jiao and Nathan Wall, "A Performance Comparison of Native and Non – Native Speakers of English on an English language Proficiency Test", Pearson Technical Report, August 2004.

[11] Yemi Olagbaju, Buket D. Barkana, Navarun Gupta, "English Vowel Production by Native Mandarin and Hindi Speakers", IEEE Computer Society, ISBN 978-0-7695-3984-3, 2010

[12] Ren Wenxia, Zhaug Huili, and Lv Wenzhe, "Realization of Isolated – Words Speech Recognition System", IEEE Computer Society, 978-0-7695-3614-9/09, 2009, DOI 10.1109/PACCS 2009.

[13] Laura Tomokiyo, "Linguistic Properties of Non- Native Speech", ISBN 0-7803- 6293- 4, 1335-1338, Vol. 3, June 2000.

[14] Qingqing Zhang, Ta Li, Jielin Pan and Yonghong Yan, "Non native Speech Recognition based on Stste – Level Bikingual Model Modification", IEEE Computer Society, ISBN 970-07695-3407-7/08, 2008

[15] Bishnu Prasad Das, Ranjan Parekh," Recognition of Isolated Words using Features based on LPC, MFCC, ZCR and STE, with Neural Network Classifiers". International Journal of Modern Engineering Research, Vol.2, Issue.3, pp-854-858 ISSN: 2249-6645, 20.

[16] Jhing-Fa Wang and Shi-Huang Chen, "Wavelet Transforms for Speech Signal Processing", Journal of the Chinese Institute of Engineers Vol.22 No. 5, PP. 549-560.

[17] Liu Wai and Pascule Fung, "Fast Accent Identification and accented Speech Recognition", 0-7803-5041-3/99. IEEE 1999

[18] Pierre-Yves Oudeyer, "The Production and Recognition of emotion in speech: Features and algorithms", International Journal of human computer studies Elsevier Science doi 10.1016/S1071-5819(02)00141-6.\ 59(2003) 157-189, 2002

[19] Xia Mao, Lijiang Chen and Bing Zhang, "Mandarin Speech Emotion Recognition based on a hybrid of HMM/ANN", International Journal of Computers. Issue 4, Volume 1, 2007.

[20] Iosif Mporas, Todor Ganchev, Mihalis Siafarikas, and Nikos Fakotakis, "Comparison of Speech Features on the Speech Recognition Task", Journal of Computer Science 3 (8): 608-616, , ISSN 1549-3636, 2007

[21] Hossan M.A., "A Novel Approach for MFCC Feature Extraction", 4th International Conference on Signal Processing and Communication System pp 1-5, December, 2010.

[22] Ganesh B. Janvale, Vishal B. Waghmore, Vijay Kale, and Ajit S. Ghodke, "Recognition of Marathi Isolated Spoken Words Using Interpolation and DTW techniques", ICT and critical Infrastructure: Proceeding of the 48th Annual of computer society of India Vol I. Advances in Intelligent system 3-319-03107-3-3, Print ISBN 978-3-319-031066, Online ISBN 978-3-319-03107-1.,2013

[23] Vishal B. Waghmare, Ratnadeep R. Deshmukh, Pukhraj P. Shrishrimal, and Ganesh B. Janvale, "Emotion Recognition System from Artificial Marathi Speech using MFCC and LDA Techniques", Proceeding of International Conference on Advances in Communication, Network, and Computing. 21th to 22nd February, 2014

[24] Mihalis Siafarikas, Iosif Mporas, Todor Ganchev and Nikos Fakotakis, "Speech Recognition using Wavelet Packet", Journal of Wavelet Theory and Applications, ISSN 0973-6336 Volume 2 No.1, 2008

[25] Felix Weninger, Jarek Krajewski, Anton Batliner, and Bjorn Schuller, "The Voice of Leadership: Models and Performances of Automatic Analysis in Online Speeches", IEEE Transactions on Affective Computing, Vol. 3, No. 4, October –Dumber 2012.

[26] James K. Tamgno, Etienne Barnard, Claude Lishou, and Morgan Richomme, "Wolof Speech Recognition Model of Digits and Limited-Vocabulary Based on HMM and ToolKit", IEEE Computer Society, 978-0-7695-4682-7/12, DOI 10.1109/ 2009.

[27] Zhao Lishuang and Han Zhiyan, "Speech Recognition System Based on Integrating feature and HMM", IEEE Computer Society, 978-0-7695-3962-1/10, DOI 10.1109/2009

[28] Rupayan Chakraborty and Utpal Garain, "Role of Synthetically Generated Samples on Speech Recognition in a Resource-Scarce Language", IEEE Computer Society, 1051-4651/10, DOI 10.1109/2009

[29] Clarence Goh Kok Leon, "Robust Computer Voice Recognition Using Improved MFCC Algorithm", IEEE Computer Society, 978-0-7695-3687-3/09, DOI 10.1109/2009

[30] A.Revathi and Y. Venkataramani, "Perceptual Features based Isolated Digit and Continuous Speech Recognition using Iterative Clustering Approach", IEEE Computer Society, 978-0-7695-3924-9/09, DOI 10.1109/2009.

Biographies

Mr.V.K.Kale: Rceived the B.Sc.degree in Computer Science from the Dr.Babasaheb Ambedkar Marathwada University Aurangabad, (MS) India in 2006. Also M.Sc. degree in Computer Science from the Dr.Babasaheb Ambedkar Marathwada University Aurangabad (MS) India in 2008. Respectively Currently, He is an M.Phil Computer Science Students in Department of Computer Science and IT, Dr.Babasaheb Ambedkar Marathwada University, Aurangabad (MS) India.
vijaykale1685@gmail.com

Dr.R.R.Deshmukh: M.E. (CSE), M.Sc.(CSE) Ph.D .FIETE, Presently working as Professor in Department of Computer Science and Information Technology, Dr. Babasaheb Ambedkar Marathwada University, Aurangabad, MS-India. He is a Member of Management Council of Dr. Babasaheb Ambedkar Marathwada University, Aurangabad, MS-India. He is Chairman of Ad-Hoc Board of Studies in Computer Science and IT and Bioinformatics, Dr. Babasaheb Ambedkar Marathwada University, Aurangabad, MS-India. He is Member of Ad--Hoc Board of Studies in Computer Engineering of Mumbai University, Mumbai. He is Member of Board of Study Computer Science, Solapur University; Solapur.He is a Fellow of IETE, Senior member of IEEE, Life member of ISCA, CSI, ISTE, ACEEE, CSTA and IDES. He has edited of nine books and published more than 70 research papers in reputed Journals, National and international conferences. He is reviewer and editor of several jour-

nals at national & international level. His areas of specialization are Human Computer Interaction (HCI), Data Mining, Data Warehousing, Image Processing, Pattern Recognition, Artificial Intelligence; Computational Auditory Scene Analysis (CASA) Neural Networks etc.He has organized several workshops and conferences. He is nominated as a subject expert on various academic & professional bodies at national level government bodies. He is Faculty member for Engineering, Science & Management Faculty & Member of various committees at University level.
ratnadeep_deshmukh@yahoo.co.in

Dr. G. B. Janvale: Received post graduate and doctoral degree in Computer science from Dr. Babasaheb Ambedkar Marathwada University, Aurangabad. His research interests include Human Computer Interfacing, Speech Signal Processing and Neuroscience (EEG and fMRI) and electromagnetism. He has published numerous research papers in various national and international journals. Currently, he is working as an Assistant Professor at Symbiosis Centre for Information Technology, Symbiosis International University, Pune India.
ganeshjanvale@gmail.com