International Journal of Engineering Technology and Applied Science

(ISSN: 2395 3853), Vol. 3 Issue 3 March 2017

Analysis of Speech Recognition and Voice Characterization

[1]MD. Jiyaurrhaman Ansari
[2]Sonu Lal
[3]Ashish Raghuwanshi
[1]M-Tech Scholar, [2]Associate Professor, [3]Assistant Professor
[123]Department of Electronics & Communication Engineering, ICOT, Bhopal

Abstract
Recognizing the speaker can simplify the task of translating speech in systems that have been trained on specific person's voices or it can be used to authenticate or verify the identity of a speaker as part of a security process. This work discusses the Implementation of an Enhanced Speaker Recognition system using MFCC and LBG Algorithm. MFCC has been used extensively for purposes of Speaker Recognition. This work has augmented the existing work by using Vector Quantization and Classification using the Linde Buzo Gray Algorithm. A complete test system has been developed in MATLAB which can be used for real time testing as it can take inputs directly from the Microphone. Therefore, the design can be translated into a Hardware having the necessary real time processing Prerequisites. The system has been tested using the VID TIMIT Database and using the Performance metrics of False Acceptance Rate (FAR), True Acceptance Rate (TAR) and False Rejection Rate(FRR). The system has been found to perform better than the existing systems under moderately noisy conditions.

KEYWORDS: Voice recognition, MFCC, LBG

I INTRODUCTION

Sound is created by the vibration of any medium, for instance, the vibrations of the pc speaker or just of air molecules or pressure within the air. These vibrations area unit typically modelled as 2 forms of layers, interleaved, motion along through the medium; air mass layers (molecules compressed over normal) and low layers (molecules relaxed over normal). The vibrations have an effect on the ears which is that the underlying theory behind listening. In fact, sound will be perceived as a signal; the amplitude of that corresponds to the pressure modification and also the length of that corresponds to the space between 2 consecutive high (and 2 consecutive low) pressure layers.

Human speech is one kind of sound; which individuals have developed through time to hold valuable info for communication, like thoughts and feelings. However, it additionally carries alternative derived characteristics like the speaker's identity, language, diction, dialect, gender and mood. It based mostly upon the mix of lexical and names that are drawn from terribly giant info (usually concerning 10,000 completely different words). Every vocal is made out of the phonetic combination of a restricted set of vowels and consonant language unit units. These vocabularies, the syntax that structures them, and their set of language unit units disagree, inflicting the existence of the many thousands of various forms of reciprocally unintelligible human languages. Most human speakers are able to communicate in 2 or a lot of the languages, therefore known as polyglots. The vocal talents that change humans to {provide} speech additionally provide humans with the power to sing. Speech is researched in terms of the utterance and sound perception of the sounds employed in vocal language. Alternative analysis topics concern repetition of speech, the power to relate detected and spoken words into the vocalizations required to recreate. This plays a key role within the vocabulary enlargement in youngsters and speech errors. Many educational disciplines study these as well as acoustics, psychology, speech pathology, linguistics, scientific discipline, communication studies, and otology and computing. Another space of analysis is however the human brain in its completely different areas like the center and nerve center underlies speech forgotten or lost.

Speech is extremely basic manner for humans to convey info. The most objective of Speech is communication. Speech is outlined because the response of vocal track to 1 or a lot of excitation signal. Large quantity of information transmission is extremely tough each in terms of transmission and storage. Speech Compression could be a methodology to convert human speech into an encoded kind in such the way that it will later be decoded to induce back the first signal.

Compression is essentially to get rid of redundancy between neighboring samples and between adjacent cycles. Major objective of speech compression is to

represent signal with lesser variety of bits. The reduction of information ought to be drained such the way that there's acceptable loss of quality.

II RECENT STATES OF ART

Speaker popularity has been an energetic studies place and a number of the extra current works inside the field have been cited here. "together with human know-how in speaker reputation systems: report on a pilot assessment" with the aid of CS Greenberg, AF Martin, IEEE journal of Speech and sign, 2012[50] discusses Speaker reputation assessment (SRE10) included a test of Human Assisted Speaker recognition (HASR) wherein systems primarily based in entire or in element on human know-how were evaluated on confined sets of trials. "supply-normalized-and-weighted LDA for robust speaker reputation using i-vectors" by way of M McLaren, D Van Leeuwen - Acoustics, Speech and signal, 2012[51]. The currently advanced i-vector framework for speaker reputation has set a new performance standard in the studies area. An i-vector is a compact illustration of a speaker utterance extracted from a low-dimensional general variability subspace Multi-variability speech database for sturdy speaker popularity through BC Haris, G Pradhan, A Misra, S Shukla NCC), 2011 national, 2013[52]. On this paper, the authors have presented an initial look at with the lately collected speech database for developing robust speaker popularity structures in Indian context. The database incorporates the speech records amassed across extraordinary sensors, languages, talking styles and so on. Inside the paper performance evaluation of Speaker reputation using Vector Quantization by LBG and KFCG by using H. B. Kekre and Vaishali Kulkarni [53], techniques for speaker recognition primarily based on Vector quantization are proposed and their performances are in comparison. Vector Quantization (VQ) is used for characteristic extraction in both the education and checking out stages. Methods for codebook technology have been used. Inside the 1st method, codebooks are generated from the speech samples by way of using the Linde-Buzo-gray (LBG) algorithm. In the second approach, the codebooks are generated using the Kekre"s rapid Codebook generation (KFCG) set of rules. Inside the paper Speaker recognition the usage of Vector Quantization by using MFCC and KMCG clustering set of rules by using HB Kekre, VA Bharadi, AR Sawant IEEE, 2014[54] authors have applied a speaker reputation system using a combination of Mel Frequency Capstral Coefficients (MFCC) &Kekre's MCG clustering set of rules. As is obvious from the listed works within the bankruptcy, it can pretty genuinely be seen that the field has been a lively research vicinity inside the beyond few years by means of the no of manuscripts that have been published on the topic. However, the number one metrics on which the system is evaluated for the performance i.e. TAR, some distance and FRR have been discovered to be one-of-a-kind in extraordinary works.

III DOWNSIDE IDENTIFICATION

In this technology for the speech recognition task we've got varied process task as given below

- Speaker identification
  - Speaker detection
  - Speaker verification
  - Speaker segmentation
  - Speaker clump
  - Speaker diarizatio
- Speaker Recognition
  - Conventional identification
  - Speaker verification
- Text conversion
  - Text freelance recognition
  - Text dependent recognition

In this work, we found for the identification and speaker recognition the accuracy of on top of process task isn't sensible or not fascinating. In this technology within the yelling setting it's not appropriate or it'll not method properly.

IV PROPOSED WORK

We like to represent the speech signal parametrically for the speaker recognition task. Typically or this can be often named because the signal-processing forepart. During this work Mel-Frequency Cepstrum Coefficients (MFCC) has been used for feature extraction. Mel frequency cepstral coefficients (MFCC) is perhaps the simplest renowned and most generally used technique for each speech and speaker recognition. A Mel may be a unit of live supported human ear's perceived frequency. The Mel scale has roughly linear frequency spacing below one thousand cycles per second and an exponent spacing on top of 1000Hz. The approximation of Mel from frequency will be expressed as-

$$Mel(f)= 2595*\log10 (1+f/700) \quad \text{……Equation 1}$$

Where f denotes the important frequency and Mel(f) denotes the perceived frequency.

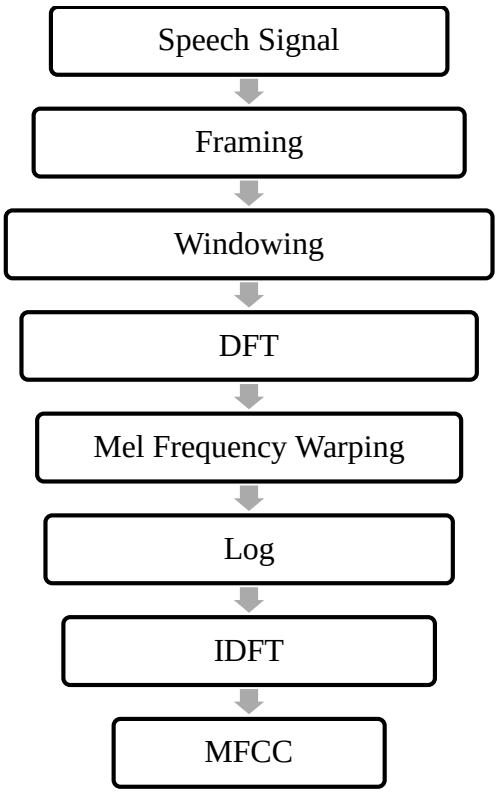

Figure 1 MFCC Extraction

MFCCs are obtained as follows as shown above in the figure 1.

1. Take the Fourier transform of (a windowed excerpt of) a signal.
2. Map the powers of the spectrum obtained higher than onto the Mel scale, exploitation triangular overlapping windows.
3. Take the logs of the powers at every of the Mel frequencies.
4. Take the distinct cosine remodel of the list of Mel log powers, as if it were a signal.
5. The MFCCs are the amplitudes of the ensuing spectrum.

Speech signals are unremarkably pre-processed before options are extracted to boost the accuracy and potency of the extraction processes. Speech signal preprocessing covers digital filtering and speech reception. Filtering includes pre-emphasis filter and filtering out any close noise.

V SIMULATION RESULTS

Performance Metrics

Speaker recognition system performance is measured using various metrics such as recognition or acceptance rate and rejection rate. Recognition rate deals with the number of genuine speakers correctly identified, whereas rejection rate corresponds to the number of imposters (people falsifying other‟s identity) being rejected.

Along with these performance metrics there are some performance measures and tradeoffs one need to consider while designing speaker recognition systems. Some of the performance measures generally used in the evaluation of these systems include:

1. **False Acceptance Rate (FAR)** - the rate at which an imposter is accepted as a legitimate speaker,
2. **True Acceptance Rate (TAR)** - the rate at which a legitimate speaker is accepted
3. **False Rejection Rate (FRR)** - The rate at which a legitimate speaker is rejected (FRR=1-TAR)

FAR Results have been shown below in figure 2. As we can see that the false acceptance rate over a test set ranging from 44 to 100 users is roughly 9%.

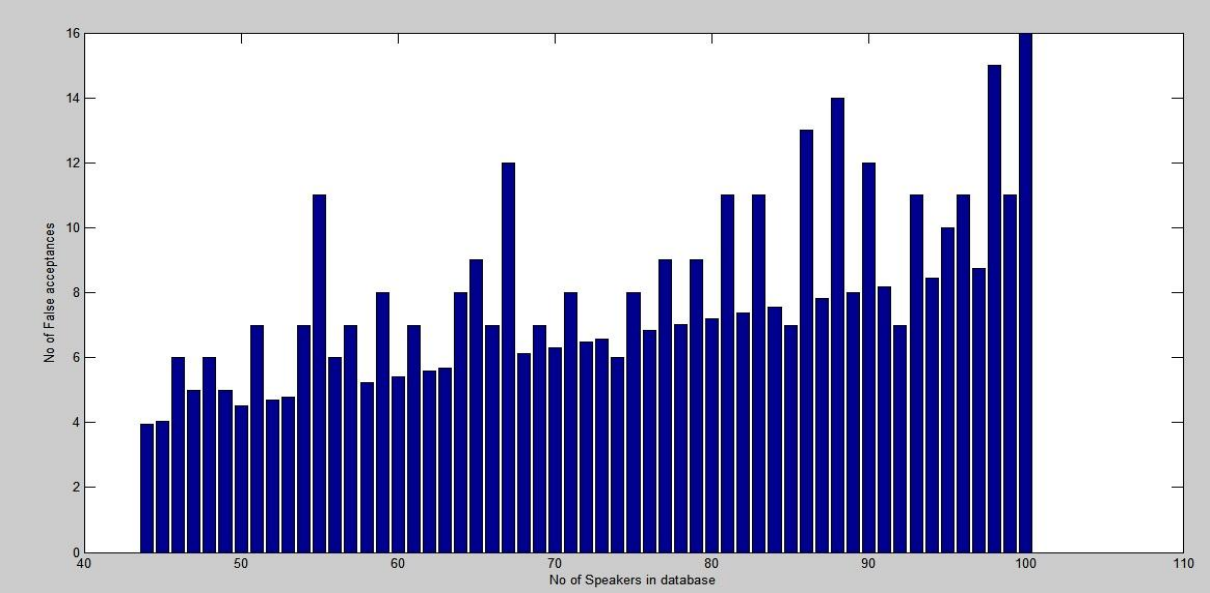

Figure 2 FAR for 44 to 100 user database

Figure 3 shows false rejection rate which is roughly 15% where legitimate speakers have been rejected as Imposters. The dataset has again been varied from 44 to 100 users.

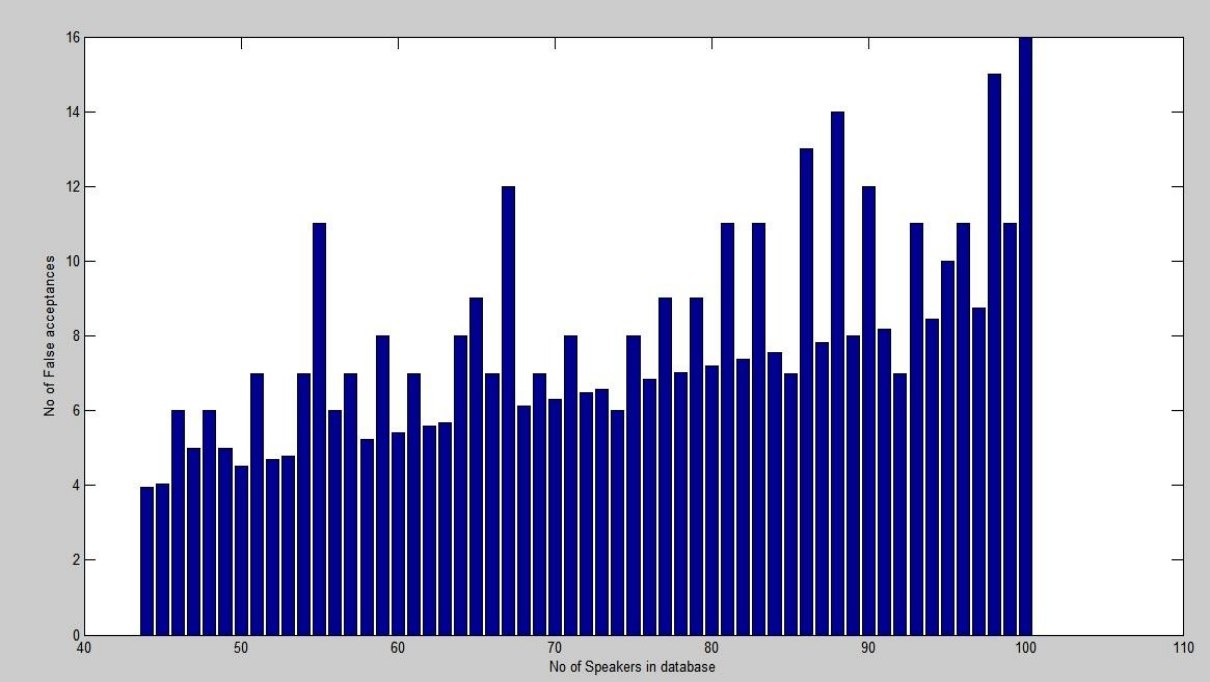

Figure 3 False rejection rate results

Figure 4 shows the True acceptance rate which has been tested on a similar database and has been found to have an accuracy of roughly 95%. However the acceptance rate shows a decline if the no of user inputs are from microphone owing to noisy environments.

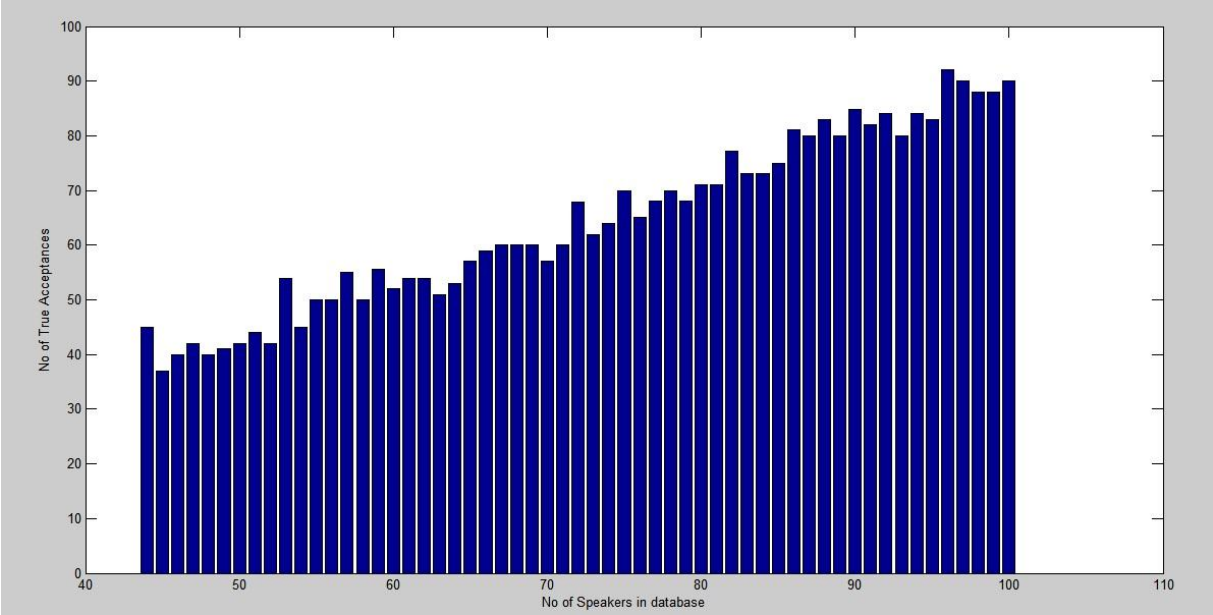

Figure 4 True acceptance rate results

VI Conclusion

This work has presented an enhanced mechanism of Speaker Recognition using a combination of the well-known MFCC algorithm as well as LBG algorithm for generating the vector code-words. The training and testing was done on the VID TIMIT database and the system was found to perform efficiently as is visible from the **False Acceptance Rate (FAR) True Acceptance Rate (TAR)**, **False Rejection Rate (FRR)** results. Out of the total database of 43 user and 20 self-generated inputs, we were able to achieve a True acceptance rate of 96%, false acceptance rate of less than 9% and false rejection rates of less than 7 %. The system has been found to perform satisfactorily under noisy conditions as well however has been found prone to increase in FAR if user inputs are from microphone under noisy conditions. The testing has been done by using standard Microphones in acoustically silent environments and then additional hum has been added for noise simulations. The GUI developed for the purpose has capabilities of real time speaker recognition, making it a significant contribution to the work.

REFERENCES

1) Grenander, Ulf (1959). Probability and Statistics: The Harald Cramér Volume. Wiley. "The Nyquist frequency is that frequency whose period is two sampling intervals."
2) Harry L. Stiltz (1961). Aerospace Telemetry. Prentice-Hall. "the existence of power in the continuous signal spectrum at frequencies higher than the Nyquist frequency is the cause of aliasing error"
3) Thomas Zawistowski, Paras Shah. "An Introduction to Sampling Theory". Retrieved 17 April 2010. "Frequencies "fold" around half the sampling frequency - which is why [the Nyquist] frequency is often referred to as the folding frequency."
4) Campbell Jr., J.P., 1997. Speaker recognition: A tutorial. Proceedings of the IEEE, 85(9), pp.1437–1462
5) Daniel Jurafsky & James H. Martin "Automatic Speech Recognition" Speech and Language Processing: An introduction to natural language processing, computational linguistics, and speech recognition, 2007

6) He´bert, M., 2008. Text-dependent speaker recognition. In: Benesty, J.,Sondhi, M., Huang, Y. (Eds.), Springer Handbook of Speech Processing. Springer-Verlag, Heidelberg, pp. 743–762

7) Douglas A Reynolds "An Overview of Speaker Recognition Technology", MIT Lincoln Laboratory, MA 2002

8) Speaker segmentation and clustering (2008) M Kotti, V Moschou, C Kotropoulos., 2008"

9) Digital Speech Processing: Synthesis, and Recognition, Second Edition, Sadaoki Furui, 2000 CRC Press

10) An Automatic Speaker Recognition System. P. Chakraborty, F. Ahmed, Md. Monirul Kabir, M. Shahjahan, and Kazuyuki Murase. ICONIP 1, volume 4984 of Lecture Notes in Computer Science, page 517-526. Springer, (2007).

11) . R. Rabiner,B.-H. Juang, C.-H. Lee ,An Overview of Automatic Speech Recognition Automatic Speech and Speaker Recognition, The Kluwer International Series in Engineering and Computer Science Volume 355, 1996, pp 1-30

12) Phaophak Sirisuk, Fearghal Morgan, Tarek El-Ghazawi, Hideharu Amano Reconfigurable Computing: Architectures, Tools and Applications: 6th Edition page 359, Bangkok, Thailand, March 17-19, 2010, Proceedings

13) From Frequency to Quefrency: A History of the Cepstrum Alan V. Oppenheim and Ronald W. Schafer, IEEE Signal Processing Magazine Reprinted 2004

14) Atal, B. S., and Hanauer, S. L. (1971) "Speech analysis and synthesis by linear prediction of the speech wave," J. Acousto Soco Am. 50, 637-655..

15) . R. Sambur A, . E. RosenbergL, . R. Rabinera, nd C. A. McGonegal "On reducing the buzz in LPC synthesis", M Bell Laboratories

16) Davis, S. ; Mermelstein, P Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences Acoustics, Speech and Sig Processing, IEEE Transactions on (Volume:28 , Issue: 4)Aug 1980, Page(s): 357 - 366

17) Todor Ganchev , Nikos Fakotakis , George Kokkinakis Comparative evaluation of various MFCC implementations on the speaker verification task (2005),

18) Soong, F. ;; Rosenberg, A. ; Rabiner, L. ; Juang, B.H ,A vector quantization approach to speaker recognition. Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP '85. (Volume:10) Apr 1985, Page(s):387 - 390

19) Singh, G Panda, A. ; Bhattacharyya, S. ; Srikanthan, T "Vector quantization techniques for GMM based speaker verification". ; Indian Inst. of Technol., Kanpur, India Acoustics, Speech, and Signal Processing, 2003. Proceedings International Conference on. (ICASSP '03). 2003 IEEE (Volume:2) 6-10 April 2003 Page(s):II - 65-8 vol.2 ;.

20) Sheeraz Memon, Automatic Speaker Recognition: Modelling,Feature Extraction and Effects of Clinical Environment, Research RMIT, 2010.

21) Bimbot, F., Bonastre, J.-F., Fredouille, C., Gravier, G. Magrin-Chagnolleau, I., Meignier, S., Merlin, T., Ortega-Garcia, J., Petrovska-Delacretaz, D., and Reynolds, D. A tutorial on text-independent speaker verication. EURASIP Journal on Applied Signal Processing 2004, 4 (2004), 430-451.

22) Muller, C., Ed. Speaker Classication I: Fundamentals, Features, and Methods (2007), vol. 4343 of Lecture Notes in Computer Science, Springer.