

Co-Variance based Real Time Speaker Recognition and Gender Identification by Rectangular Window and Low Pass Filter

Keyur G. Kulkarni

School of Electronics and Communication Engineering,
KLE Technological University
Hubballi, India

Abstract—Speech Processing is one of the emerging technologies of Signal Processing. Some areas of Speech Processing are Speaker Recognition(SR), Gender Identification(GI), age prediction, etc. Speaker Recognition is a prime research area of signal processing. Speech Recognition means identifying the right speaker based on his/her voice sample. Gender Identification is an extension of Speaker Recognition that identifies the gender of the speaker. For SR and GI, two things are essential. One is coaching part and another is testing part. Feature extraction is that the extraction of tiny info from the obtainable audio wave signal. The signal is in .wav format. Feature Extraction is done various methods like FFT(Fast Fourier Transform), LPC(Linear Predictive Coefficients), MFCC(Mel Frequency Cepstral Coefficients) and many more. SR in this paper is achieved based on the co-variance technique as it is more accurate than FFT.GI is achieved through Low Pass Filtering and Rectangular Windowing techniques.

Index Terms—Co-variance,Low Pass Filter,Rectangular Window,Speaker Recognition,Gender Identification

I. INTRODUCTION

Speech is that the basic mode of communication for personalities. Computer's usage has become inevitable in modern era. Ex : Amendment of data between Humans became a natural complication so speech recognition system comes in light-weight . The recognition system converts words spoken by Humans into a kind during which the computer will perceive and respond consequently. There are 2 main parts in a speech recognition system. One is Coaching part and another is Testing part.It's not possible to acknowledge all the words by the system. The speech recognition system use strategies like DTW, HMM.The speech recognition system starts with changing human voice into digital signal. Co-variance is employed to extract vector features from the voice signal.Co-variance algorithmic rule is based mostly on human hearing perceptions. Gender Identification is based on Low pass filter and Rectangular Windowing.

II. LITERATURE REVIEW

A. Speaker Identification Using GMM with MFCC Tahira Mahaboob, Memoona Khan, Malik Sikandar Hayat Khoyal, Ruqia Bibi

Speaker identification comes underneath the sphere of Digital Signal process. The primary step in any voice recognition system is for the user to offer an input by speaking a word

or a phrase into a microphone. Then an analog to digital device converts the electrical signal to digitized type and stores it within the memory. The pc then tries to work out the means of a voice sample by matching it with a example that encompasses a legendary that means. This paper targets the implementation of MFCC with GMM techniques so as to spot a speaker. The developed system is consisting of three processes, 1. options Extracting 2. Training 3. Matching In initial method, the developed system can result as computed options of human voice. These options are voice options that are taken from the persons. These options are extracted by exploitation of MFCC technique. The steps concerned in MFCC are, 1. Pre-emphasizing 2. Framing 3. Windowing 4. quick Fourier Transform 5. Mel Filter 6. Frequency Wrapping 7. discrete cos Transform In coaching method, the extracted options are trained exploiting the Gaussian Mixture Modelling. Expectation Maximization (EM) formula is employed to coach the extracted options of human voice in system so finally accustomed to store as information. Steps concerned in GMM coaching are: 1. Clustering 2. Expectation 3. Maximization The matching is finished on by finding the log probability of voice sample. The planned system during this paper shows the most effective potency up to 87.5

B. Real Time Speaker Recognition System using MFCC and Vector Quantization Technique Roma Bharti, Priyanka Bansal

This paper represents a really sturdy mathematical algorithmic rule for automatic Speaker Recognition (ASR) system exploiting MFCC and vector division technique within the digital world. ASR could also be a variety of bio-metric that uses associate individual's voice for recognition technique.

The target of the speaker recognition system is to convert the acoustic audio signal into computer code kind. The human speech is then processed by the machine relying upon the two factors 1. Feature Extraction 2. Feature Matching In this paper MFCC is employed for feature extraction because it is standard and most desirable technique, supporting the far-famed variation of the human ear's crucial information measure frequency. Vector quantisation Technique for feature extraction is employed during this paper and it is based-on pattern recognition and LBG (Linde, Buzo and Gray) algorithmic program. Vector quantisation maps the vectors from an outsized set into clusters

and these clusters are termed as codebook. For each listed speaker a codebook is generated, and through checking the Euclidean distance between the acoustic vector of test signal and therefore the mapped codebook is calculated. The speaker having the tiniest Euclidean distance is chosen. The real time speaker recognition system during this paper is developed exploiting MATLAB. The result from this paper was ended exploiting 120 speakers with TIDIGIT information resulting in 91 traditional condition at 20db SNR.

C. Speaker Identification and Clustering Using Convolutional Neural Networks Yanick Lukic Carlo Vogt Thilo Stadelman

Convolutional Neural Networks a sort of Deep Learning has sparked substantial enhancements in pc vision and related fields in recent times. This paper proposes a unique answer to boost the speaker recognition pipeline exploiting Convolutional Neural Networks. CNNs are applied on spectrograms so as to be able to learn speaker specific options from a rich acoustic supply illustration. This paper works with the procedure of talker identification and speaker clump using CNNs. The clump technique is advanced and results in several errors. This paper worries with the advancement of pure speaker recognition so as to the shut the preceding gap. The identifying and prime options of this technique of talker identification are, 1. Speaker Discriminating options are sought-after. 2. Pitch info is recorded. 3. Specific voice connected characteristics' of the speech signal are exploited. The paper additionally elaborates the way to transfer network trained for talker identification to speaker clump. From spectrograms by victimization of CNNs trained for talker identification options are learnt then one amongst the post convolutional layers is employed for feature illustration. The input signal is remodelled to a exposure victimisation python libraries. The input is processed with python libraries like 'Librosa', 'Keras' to create a Mel exposure with 128 components in frequency direction for every knowledge. A Mel exposure is Associate in Nursing architecture of $F \times T$ with MFCC. X-axis=time Y-axis=frequency The dark bands represent quantity of energy at a specific frequency. additionally this exposure is max-pooled employing a smaller window Eg: 2×2 . Max-pooling may be a sample-based discretization method. the target is to down sample an input illustration reducing its spatiality and leaving assumptions to be created regarding options contained within the sub-regions binned. All the minute details of speaker are obtained by max-pooling. The paper depicts the experiments a number of talker identification from TIMIT dataset. To ensure enough coaching knowledge 8 out of 10 sentences were used for every speaker testing. An accuracy of 97

III. PROPOSED METHOD

On a median variance offers the effective (efficient) identification accuracy within the clean setting. we have a tendency to propose a feature extraction technique with the changed design of Co-variance. So, our projected technique could also be a variant of variance which is for the real-world setting to boot as for the clean environment. variance or the method

of data according to mathematics principles has adult into a robust tool for structure elucidation, signal assignments and identification of mixture constituents. Experimental process by variance can either replace or accompany the traditional Fourier transformation. variance take a glance at structures are of nice profit in many domains of mathematics analysis and so the development of compressed detection matrices could also be a significant disadvantage in signal method. driven by these applications, Cai, in an article, studies the laws limiting the coherence of a random matrix NXP among the dimensional framework where p area unit typically much larger than n, most so as that every the law of huge numbers and distribution limit are determined. We've a bent to then study how to take a glance at the variance matrix of a dimensional Gaussian distribution that options independence tests as a particular case. The limiting laws of the coherence of the data matrix play a vital role in building the take a glance at.

- I.Co-variance:

$C = \text{cov}(A)$ returns the co-variance.

If A could be a vector of observations, C is that the scalar-valued variance.

If A could be a matrix whose columns represent random variables and whose rows represent observations, C is that the variance matrix with the corresponding column variances on the diagonal.

C is normalized by the quantity of observations-1.

If there's just one observation, it's normalized by one.

If A could be a scalar, $\text{cov}(A)$ returns zero. If A is an empty array, $\text{cov}(A)$ returns NaN.

- soundsc:

Scale data and play as sound

Syntax:

`soundsc(y,Fs)`

`soundsc(y)`

`soundsc(y,Fs,bits)`

`soundsc(y,...,slim)`

Description:

1) `soundsc(y,Fs)` sends the signal in vector y (with sample frequency Fs) to the speaker on laptop and most OS platforms. The signal y is scaled to the vary before it's vie, leading to a sound that's vie as loud as potential while not clipping.

2) `soundsc(y)` plays the sound at the default sample rate or 8192 cycle.

3) `soundsc(y,Fs,bits)` plays the sound victimisation bits range of bits/sample if potential. Most platforms support bits = eight or bits = sixteen.

4) `soundsc(y,...,slim)`, wherever slim = [slow shigh], maps the values in y between slow and shigh to the total sound vary. The default price is slim = $[\min(y) \max(y)]$.

- abs(Absolute Value):

Absolute value and complex magnitude

Syntax:

$Y = \text{abs}(X)$

Description:

example $Y = \text{abs}(X)$ returns the absolute value of each element in array X .

If X is complex, $\text{abs}(X)$ returns the complex magnitude.

- Low Pass Filter:

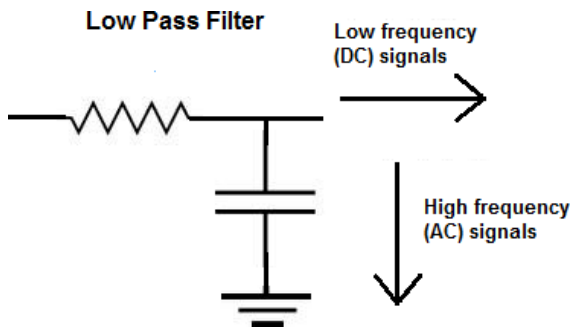


Fig. 1. Low Pass Filter Block Diagram

A Low Pass Filter may be a circuit which will be designed to change, reshape or reject all unwanted high frequencies of associate electrical signal and settle for or pass solely those signals needed by the circuits designer. The Low Pass Filter – the low pass filter solely permits low-frequency signals from 0Hz to its cut-off frequency, f_c purpose to pass whereas interference those any higher. Simple First-order passive filters (1st order) will be created by connecting along one resistance associated one electrical device nonparallel across an input, (V_{IN}) with the output of the filter, (V_{OUT}) taken from the junction of those 2 parts. Depending on that means around we have a tendency to connect the resistance and therefore the electrical device with regards to the sign determines the sort of filter construction leading to either a coffee Pass Filter or a High Pass Filter. As the perform of any filter is to permit signals of a given band of frequencies to pass in-situ whereas attenuating or weakening all others that aren't needed.

- Rectangular Window:

Definition (M odd):

$$w_R(n) = \begin{cases} 1, & |n| \leq \frac{M-1}{2} \\ 0, & \text{otherwise} \end{cases}$$

$$W_R(\omega) = M \cdot \text{sinc}_M(\omega) = \frac{\sum_{n=0}^{M-1} \sin(\omega n)}{\sin(\omega)}$$

- Gender Identification:

There are 11 data (.wav files) which is being used for classification. In this few of the sample are male and few are females. The gender is classified by taking the characteristics such as pitch, short period energy, number of zero crossing. The Characfeatures.m takes the features taking in the input data and finding the features with it. The identification.m takes all the inputs and creates a data mat variable which has all the features of all the data set.

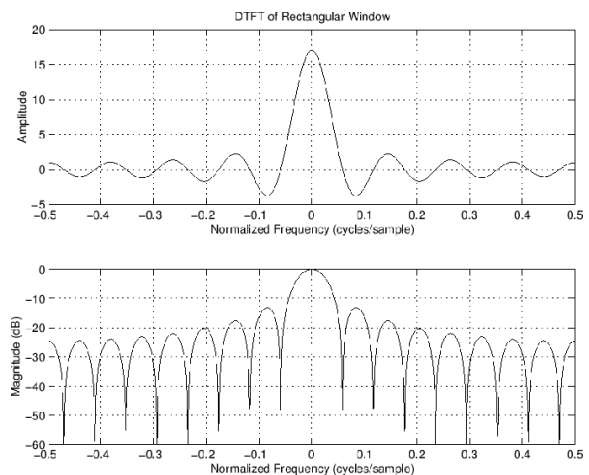


Fig. 2. DTFT of a Rectangular Window

Manually threshold is set and basic classifier is made. Finally it successfully classifies any input as being male or female.

IV. EXPERIMENTS AND RESULTS

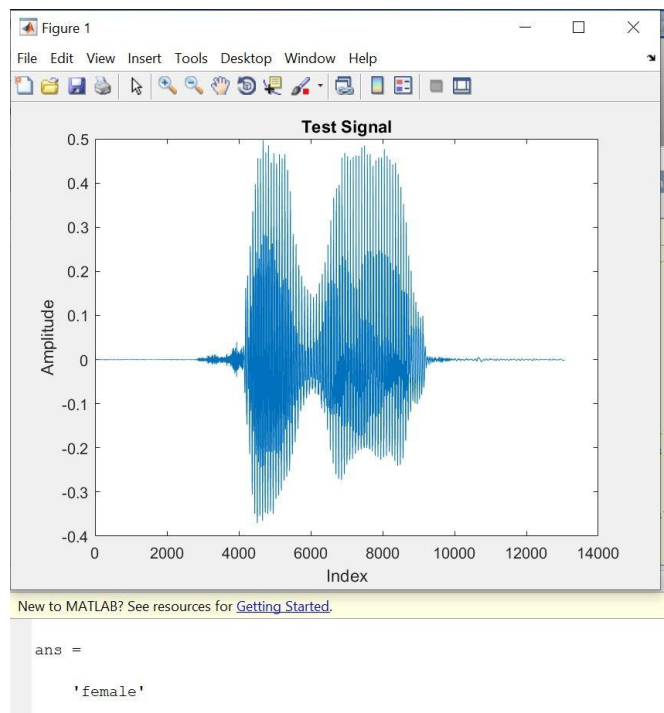


Fig. 3. Gender Identification : FEMALE

The results are tested against the specified objectives of the proposed system. The developed system is tested by taking 2 speech samples from each speaker with sampling frequency of 7kHz. Speaker Recognition is done on the basis of Covariance which is more efficient and accurate than FFT techniques. The test was conducted on a total of 5 speakers with an accuracy rate of 80 percentage

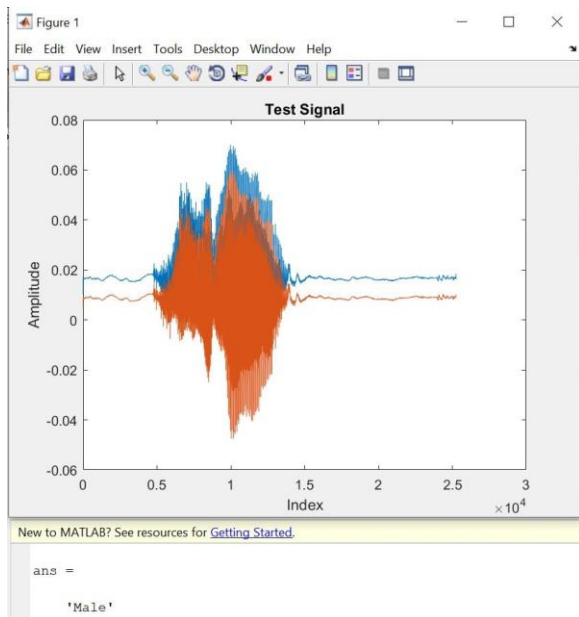


Fig. 4. Gender Identification : MALE

Accuracy Rate= number of correctly identified test samples/total number of test samples $= (4/5) * 100 = 80$ percentage

Gender Identification was done on the basis of character extractions like pitch, number of zero crossing and short period energy. The test was conducted on a total of 11 speakers (9 female and 2 male) with an accuracy rate of 100 percentage

Accuracy Rate= number of correctly identified test samples/total number of test samples $= (11/11) * 100 = 100$ percent- age

REFERENCES

- [1] B. Paresh M Chauhan and Nikita P Desai" Mel Frequency Cepstral Coefficients(MFCC) based Speaker Identification In Noisy Environment using Wiener Filter", 10.1109/ICGCCEE.2014.6921394 IEEE .
- [2] Roma Bharti and Priyanka Bansal, "Real Time Speaker Reconition Using MFCC and Vector Quantization Technique ",10.5120/20520-2361,2015.
- [3] Tahira Mahaboob, Memoona Khan, Malik Sikandar Hayat Khiyal, Ruqia Bibi" Speaker Identification using GMM with MFCC" 10.1109/ICASID.2010.5551341.
- [4] Yanick Lukic Carlo Vogt Thilo Stadelman;"Speaker Identification and Clustering Using Convolutional Neural Networks" 978-1-5090-0746-2/16,IEEE,ITALY
- [5] Kevin R. Farrellet. al, "Speaker Identification Using Neural Tree Networks", CAIE' Center, Rutgers University Piscataway, IEEE 1994.
- [6] Longbiao Wang, Kazue Minami, "Speaker Identification By Combining Mfcc And Phase Information In Noisy Environments", Toyohashi University of Technology, Japan - 978-1-4244-4296-6/10, IEEE 2010.
- [7] Martinez, J. Et. all "Speaker recognition using Mel frequency Cepstral Coefficients (MFCC) and Vector quantization (VQ) techniques" in Elec- trical Communications and Computers (CONIELECOMP), IEEE2012.