Benchmark for Speaker Identification using MEL Frequency Cepstral Coefficients on Vowels Following the Nasal Continuants in Kannada

Suman Suresh

Register No: 14FST004

An Independent Project Submitted in Part Fulfillment of PG Diploma in

Forensic Speech Science and Technology (PGDFSS&T)

University of Mysore

Mysuru

ALL INDIA INSTITUTE OF SPEECH AND HEARING

MANASAGANGOTHRI, MYSURU-570006

JULY, 2015

CERTIFICATE

This is to certify that this independent project titled **"Benchmark for Speaker Identification using MEL Frequency Cepstral Coefficients on Vowels Following the Nasal Continuants in Kannada"** is the bonafide work submitted in part fulfilment for the Post Graduate Diploma in Forensic Speech Science and Technology by the student (Registration No. 14FST004). This has been carried out under the guidance of a faculty of this institute and has not been submitted earlier to any other University for the award of any other Diploma or Degree.

Mysuru
July, 2015

Dr. S. R. Savithri
Director
All India Institute of Speech & Hearing
Manasagangothri, Mysuru - 570 006

CERTIFICATE

This is to certify that this independent project titled **"Benchmark for Speaker Identification using MEL Frequency Cepstral Coefficients on Vowels Following the Nasal Continuants in Kannada"** has been prepared under my supervision and guidance. It is also certified that this has not been submitted earlier in any other University for the award of any Diploma or Degree.

Mysuru
July, 2015

Dr. Hema N

Lecturer in Speech Sciences
Department of Speech-Language Sciences
All India Institute of Speech and Hearing
Mysuru - 570006

DECLARATION

This is to certify that this independent project titled **"Benchmark for Speaker Identification using MEL Frequency Cepstral Coefficients on Vowels Following the Nasal Continuants in Kannada"** is the result of my own study under the guidance of Dr. Hema N., Lecturer, Department of Speech-Language Sciences, All India Institute of Speech and Hearing, Mysuru, and has not been submitted earlier in any other university for the award of any diploma or degree.

Mysuru
July, 2015

Register No. 14FST004

Contents

CHAPTER I

INTRODUCTION

Identifying the speakers from their voices is an ability of the human listeners that has long been known (Atal, 1972). Voice is the very emblem of the speaker, indelibly woven into the fabric of speech, to elaborate; our utterances of spoken languages carries not only its own message, but through accent, tone of voice and habitual voice quality it is at the same time an audible declaration of our membership of a particular social regional groups, of our individual physical and psychological identity, and our momentary mood" (Lavner, 1994).

Among the biometric features verification of individuals identity based on voice has significant advantages and practical utilizations because speech is the most natural to produce and compelling biometric where it does not require a specialized input device, therefore the user acceptance of the system would be high. Recent advancement in speech technologies have produced new tools that can be used to improve the performance and flexibility of speaker recognition. While there are few degrees of freedom or alternative methods when using fingerprint or iris identification techniques, speech offers much more flexibility and different levels to perform recognition: the system can force the user to speak in a particular manner, different for each attempt to enter. Also, with voice input, the system has other degrees of freedom, such as the use of knowledge/codes that only the user knows, or dialectical/semantical traits that are difficult to counterfeit. Thus, apart from speaker identification, these methods can also be employed in forensic scenarios.

Forensic Speaker Identification is seeking an expert opinion in the legal process as to whether two or more speech samples are of the same person. According to Rose (1992), Fururi (1994) and Nolan (1997) speaker recognition can be speaker identification and speaker verification. Speaker recognition is the process of automatically recognizing the speaker based on the information included in speakers' voice. Hecker (1971) describes it as any decision making process that uses speaker dependent features of speech signal.

The main goal is to identify the speaker by extraction, characterization and recognition of the speaker-specific information contained in the speech signal.

Speaker verification is a common task in speaker recognition. Nolan (1983) describes it as a process where ‘an identity claim from an individual is accepted or rejected by comparing a sample of his speech against a stored reference sample by the individual whose identity he is claiming’. Speaker identification aims ‘to identify an unknown voice as one or none of a set of known speakers on comparison (Nolan, 1983; Naik, 1994).

Bricker and Pruzansky(1976) classified Speaker Identification as:

1. Speaker identification by listening.
2. Speaker identification by visual method.
3. Speaker identification by machine.
   a) Semi-automatic speaker identification.
   b) Automatic speaker identification.

Among the three available methods of speaker identification semi – automatic method is the most accepted and used one (Hollien, 1990; Kuwabara & Sagisaka, 1995; Fakotakis, Anastasios & Kokkinakis, 1993; Atal, 1974; Reyond, 1995; Rabiner & Juang, 1993). The distinction between identification and verification depends on the type of question that is asked and secondly on the nature of decision making task involved to answer the question.

The performance of the Speaker Verification and Identification tasks are determined by the type of speech material used to claim its identity. A text dependent system (2-3sec of speech sample) uses a predetermined text and thus requires a high degree of user cooperation, whereas text-independent systems (10-30sec of speech for training and 5-10sec for verification/testing) accept speech from unrestricted text.

The crime rates of all sorts are increasing at a world-wide scale. The usage of mobile phones has increased exponentially and the rate of its usage in committing crimes has also dramatically increased. When a crime is committed through telecommunication, **voice** is the only

evidence available for analysis. Therefore, there is a pressing need on the part of police and the magistrate for establishment of legal proof of identity from measurements of voice. And there is a tendency to disguise one's voice to conceal their identities especially while making threatening calls, kidnapping or extortion. The most frequently opted mode of disguising includes falsetto, whisper, change in the speaking rate, imitation, pinched nostrils and object in the mouth (Ramya, 2013). Therefore expert opinion is always being sought to establish whether two or more recordings are from the same speaker. This has brought the field of Forensic Speaker Identification into limelight.

Speech is a complex acoustic signal produced as a result of numerous transformations occurring at several different levels such as semantic, linguistic, articulatory and acoustic. Differences in the acoustic properties of the speech signal appear due to the differences in these transformations. Anatomical differences in the vocal tract and learned speaking habits of the individuals result in the speaker related differences. These differences can be used to discriminate between speakers. Vocal tract shape can be estimated from the formant location and spectral tilt of the voice signal. Vocal tract resonances are termed formants. Features derived from the vocal tract reveal the speaker related information.

Speaker recognition system use features generally derived only from the vocal tract. The excitation source of the human vocal system also contains speaker specific information. The excitation is generated by the airflow from the lungs, which thereafter passes through the trachea and then through the vocal folds. The excitation is classified as phonation, whispering, frication, compression, vibration or combination of these. There have been several studies on the choice of acoustic features in the speech recognition tasks. In these methods first and second formant frequencies (Stevens, 1971; Atal, 1972; Nolan, 1983; Hollien, 1990; Kuwabara & Sagisaka, 1995; Lakshmi & Savithri, 2009) and higher formants (Wolf, 1972) have been used in the past.

Vowels, nasals and fricatives (in decreasing order) are commonly recommended for voice recognition because they are relatively easy to identify in speech signals and their spectra contain features that reliably distinguish speakers.

Nasals have been of particular interest because the nasal cavities of different speakers are distinctive and not easily modified (except via colds).

One study found nasal co articulation between /m/ and an ensuing vowel to be more useful than spectra during nasals themselves (Su, Li, & Fu, 1974).

Kannada is a Dravidian language spoken predominantly by people in the South India in the state of Karnataka (40 million native speakers). Also spoken by the people of AndraPradesh, Telangana, Tamil Nadu, Maharashtra, Kerala and Goa. It is the 8th most spoken language in India and 33rd in the world. Kannada language consists of 49 characters in its alphasyllabary and is phonemic. As different characters can be combined to form compound characters (ottaksharas), the number of written symbols however is far more than the 49 characters. The characters divided into three groups: swaras (vowels), vyanjanas (consonants) and yogavaahakas (part vowel, part consonant). Two types of consonants have been identified in Kannada script, the structured consonants and the unstructured consonants. According to Sreedevi (2013) the most frequently occurring consonant in Mysuru dialect of conversational Kannada language is nasals and /n/ being the highest. The mean percentage and standard deviation of frequency of phonemes /n/, /m/ and /ṇ/ is 7.59% (0.31), 2.8% (0.26) and 0.3% (0.1) respectively.

The present study is focused on the category of vowels (/a/, /i/, /u/) of the Kannada script. The mean percentage and standard deviation of frequency of vowels /a/, /i/ and /u/ is 14.6% (1.3), 6.7% (0.44) and 4.3% (0.47) respectively in Mysuru dialect of conversational Kannada Sreedevi (2012). These vowels are speech sounds produced by voiced excitation of the open vocal tract. In the production of a vowel, the vocal tract normally maintains a relatively stable shape and offers minimal obstruction to the airflow. The energy produced can be radiated through the mouth or nasal cavity without audible friction or stoppage. Vowels are described in terms of the relative position of the constriction of the tongue in the oral cavity (front, central and back), the relative height of the tongue (high, mid and low), the relative position of the lips (spread, rounded and unrounded), the position of the soft palate (closed and open), the phonemic length of the vowel (short and long), the tenseness of the articulator (lax and tense), and the relative pitch of the vowel (high, mid and low). Acoustically vowels are characterized by formant pattern, spectrum, duration and fundamental frequency.

Nasal consonants are considered to be voiced. They are produced by lowering the velum so that the air flows through the nasal tract and is radiated through the nostrils. Nasalized vowels are produced in a similar manner to nasal consonants with the exception being that the oral cavity is not blocked, thereby allowing air flow through both oral and nasal cavities.

Many studies that review effective disguise for speaker identification state that nasal disguise and slow rate of speech are the least effective disguises. Therefore, nasal continuants would be the best speech sounds to investigate speaker identification.

Studies suggest that the nasal consonants can have a greater effect on the neighboring vowels. Following the release of a nasal consonant, the initial portion of a following vowel will be nasalized during the time interval that the velum is closing and the same holds true for the final portion of the vowel preceding the nasal consonant. The major characteristics of a nasalized vowel were a weakened and broadened first formant and an overall weaker vowel level than in a non-nasalized vowel (House & Stevens, 1956), presence of a dull resonance around 250Hz and an anti-resonance at about 500Hz (Hattori, Yamamoto & Fujimura, 1958). These acoustical features can act as unique cues in manual method of speaker verification.

For example, Ali et al (1971) reported an experiment indicating that his participants were able to predict the presence of a nasal consonant from the preceding vowel. They hypothesized that listeners use the anticipatory nasalization feature, common for nasal production in English, to help lighten the phoneme processing load.

The difficulty of identifying a speaker from his speech signal for example is a complex and confounding one which includes many aspects, levels and parameters to be considered (Bolt et al, 1979; Gruber & Poza, 1995; Nolan, 1997). The present study is concerned with third method of Hecker (1971), where machines can be used for speaker identification in semi – automatic or fully automatic manner (objective).

In Semi - automatic Speaker Identification (SAUSI), the known and the unknown samples from the speaker are selected by the examiner and are processed by the computer program for exact parameters such as first and second formants (Stevens, 1971; Atal, 1972; Nolan, 1983; Hollien, 1990; Kuwabara & Sagisaka, 1995; Lakshmi & Savithri, 2009),

higher formants (Wolf, 1972), fundamental frequency (Atkinson, 1976), fundamental frequency contours (Atal, 1972), Linear prediction coefficients (Markel & Davis, 1979; Soong, Rosenberg, Rabiner & Juang, 1985), Cepstral coefficients and Mel frequency Cepstral coefficients (Atal, 1973; Fakotakis, Anastasios & Kokkinakis, 1993; Reyond, 1995; Rabiner & Juang, 1995), Long term average spectrum (Kiukaanniemi, Siponen & Matilla, 1982) and interpretations are made by the examiner. In fully automatic method of speaker identification, majority of the work is done by the computer and examiners' role is minimal. For the purpose of automatic identification specially designed algorithms are used which differ based on phonetic context. This method is used very often in forensic science and can be easily affected by factors such as noise and distortions. The above mentioned methods have their own advantages and disadvantages and studies have shown varying efficiencies (McGhee, 1937; Thompson, 1987). However, the Cepstral Coefficients and the Mel Frequency Cepstral Coefficients have been found to be more effective in speaker identification compared to other features. Hence, the present study is focused on usefulness of Mel frequency cepstral coefficients (MFCC) on speaker recognition.

Mel Frequency Cepstrum Coefficient (MFCC) modeled on human auditory system has been used as a standard acoustic feature set for speech related applications. Psychophysical studies of the frequency resolving power of the human ear has motivated modeling the non-linear sensitivity of human ear to different frequencies. The selective frequency response of the basilar membrane (hair spacing) acts as a bank of band pass filters equally spaced in the Bark scale. Figure 1 shows the linear spacing between 100 Hz to 1 kHz and the logarithmic spacing above 1 kHz further reduces dimensionality of frame/vector of speech. The low-frequency components of the magnitude spectrum are ignored and the useful frequency band lies between 64 Hz and half of the actual sampling frequency. This band is divided into 23 channels equidistant in Mel frequency domain. MFCC's are based on the known variation of the human ears critical bandwidths with frequency, filters spaced linearly at low frequencies and logarithmically at high frequencies. In addition, MFCC's are shown to be less susceptible to the variation of the speaker's voice and surrounding environment. Initially, Fast Fourier Transformation (FFT) of a speech sample is extracted which is converted to Mel frequency. Cepstral coefficients are extracted on Mel frequencies.

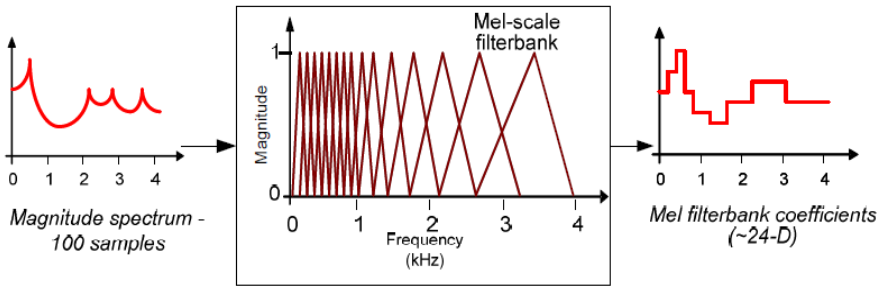

Figure 1: Mel filtering (Taken from Milner, 2003)

Mel frequency cepstrum is actually a cepstrum with its spectrum mapped onto the Mel-Scale before log and inverse fourier transform is taken. As such, the scaling in Mel-frequency cepstrum mimics the human perception of distance in frequency and its coefficients are known as the MFCC. The main difference between computation of the MFCC and the cepstral coefficients is the inclusion of Mel- Scale filter banks. MFCC are now widely used for speaker recognition tasks and has been shown to yield excellent results. Physiological studies of the frequency resolving power of the human ear has motivated modeling the non-linear sensitivity of human ear to different frequencies. MFCC's are based on the known variation of the human ears critical bandwidths with frequency, filters spaced linearly at low frequencies and logarithmically at high frequencies. In addition, MFCC's are shown to be less susceptible to the variation of the speaker's voice and surrounding environment. Initially, Fast Fourier Transformation (FFT) of a speech sample is extracted which is converted to Mel Frequency. Cepstral coefficients are extracted on Mel frequencies.

Study done on speaker identification by Hasan, Jamil, Rabbani &Rahman (2004) using MFCCs for feature extraction and vector quantization in security system based on speaker identification showed that MFCCs outperforms normal cepstral coefficients for speaker identification. The system has been implemented in Matlab 6.1 on windows XP platform. Results showed 57.14% speaker identification for code book size of 1, 100% speaker identification for code book size of 16.

CHAPTER II

REVIEW OF LITERATURE

2.1. Review of literature

There is a rapid increase in crime rate mainly through telecommunication means especially due to increase in technological advancements and usage of mobile phones. In such situations **voice** is the only source available for analysis. Therefore, there is a pressing need on the part of police and the magistrate for establishment of legal proof of identity from measurements of voice. Among the three available methods of speaker identification semi – automatic method is the most accepted and used one (Hollien, 1990; Kuwabara & Sagisaka, 1995; Fakotakis, Anastasios & Kokkinakis, 1993; Atal, 1974; Reyond, 1995; Rabiner & Juang, 1993).

The voice identification was first adopted by the Michigan State Police in 1996 and introduced it in the American court. Thus, "Forensic Voice identification is a legal process to decide whether two or more recordings of speech are spoken by the same speaker" (Rose, 2002).

Several studies have been reported on speaker identification using the listening method. According to a study done by McGehee (1937) 5 male voices were given to the listeners and were asked to identify the speaker. This was done with delays ranging from 1day to 5 months. The identification accuracy declined from 83% to 13% (after 1 day to after 5 months).

Kersta (1962) analyzed the spectrograms of five clue words spoken in isolation using 12 talkers and closed set identification. With 5 days of training the participants were asked to identify the spectrograms on the basis of 'unique acoustic cues'. The results showed that identification accuracy was inversely proportional to the number. In support to Kersta (1962), the study done by Glenn and Kleiner (1967) results show that the power spectrum of acoustic radiation produced during the nasal phonation provides a strong clue to speaker identity. Recognition accuracies were 97% for a population of 10 speakers and 93% for a population of 30 speakers.

Acoustical analysis of any speech samples can be done at three principal variables the frequency, amplitude or energy and temporal related parameters. Atal (1972) examined the temporal variations of pitch in speech as a speaker identification characteristic using 60 utterances spoken by 10 speakers consisting of 6 repetitions and found 97% correct identification. Atal (1974) examined several different parameters using linear prediction model for their effectiveness for automatic recognition of speakers from their voices. He determined 12 predictor coefficients approximately once every 50 msec from speech sampled at 10 kHz. The predictor coefficients, as the impulse response function, the autocorrelation function, the area function and the cepstrum function were used as input to an automatic speaker recognition system. The speech data consisted of 60 utterances, consisting of 6 repetition of the same sentence spoken by 10 speakers. The identification decision was based on the distance of the test sample vector from reference vector for different speakers in the population; the speaker corresponding to reference vector with minimal distance was judged to be the unknown speaker. In verification, the speaker was verified if the distance between the test sample vector and the reference vector for the claimed speaker was less than a fixed threshold. He reported that cepstrum was found to be the most effective parameter, providing an identification accuracy of 70% for speech 50 msec in duration, which increased to more than 98% for duration of 0.5sec. Using the same speech data, verification accuracy was found to be approximately 83% for duration of 50 msec increasing to 95% for duration of 1sec.

Su, Li and Fu (1974) found that a speaker-dependent characteristic, the co-articulation between /m/ and the following vowel context can be used as an acoustic clue for identifying speakers which is more reliable than nasal spectra and also because it concerns a rapid event, it is not likely to be consciously modified in natural speech. Power spectrum of nasal consonants and co-articulated nasal spectra provide strong cues for the machine matching of speakers. Glass (1984) has found that nasal consonants can be detected 88% of the times, while a vowel adjacent to a nasal consonant can be detected 74% of the times.

Glenn & Kleiner (1968) described an experiment using automatic method of speaker identification based on the spectrum of nasal sounds in different environments. Their experimental group of 30 speakers was divided into 3 groups (10 male speakers, 10 female

speakers and an additional 10 male speakers). For each speaker, all 10 samples of the spectrum of /n/ from the test set were averaged to form a test vector. The test vectors were compared with the stored reference vectors respectively. If only one speaker was correlated with the thirty reference vectors, an identification rate of 43% was got. This increased to 93% when the average of 10 speaker samples was used for correlation and further increased to 97% when the relevant population of speakers was reduced to 10. The results indicated that quite accurate speaker identification can be achieved on the basis of spectral information taken from individual segments of an utterance, in this case nasal phonemes.

Apart from traditional acoustic analysis, MFCC was found to be a beneficial approach for speech recognition according to Davis and Mermelstein (1980). Kinnunen (2003) indicated that MFCC is the most evident example of a feature set that is extensively used in speaker recognition. (A Cepstrum is the result of taking the Inverse Fourier Transform (IFT) of the logarithm of the estimated spectrum of a signal. It was first adopted as a tool for automatic pitch detection by Noll, 1964).

According to few Indian studies, Saravanan (1998) studied the effect of telephone transmission by measuring the temporal and spectral parameters and found significant differences between the 2 recordings.

Pamela (2002) investigated the reliability of voiceprints by extracting acoustic parameters in speech samples. Six normal Hindi speaking male participants in the age range of 20-25years participated in the study. Twenty nine bisyllabic meaningful Hindi words with 16 plosives, 5 nasals, 4 affricates and 4 fricatives in the word-medial position formed the speech material. Subject read the words five times. The results indicated no significant difference in F2, onset of burst and frication noise, F3 transition duration between subjects. However, the results indicated high amount of intra-subject variability. High intra-subject variability for F2 transition duration, onset of burst, closure duration, retroflex and F2 of high vowels were observed. Low inter-subject variability and intra-subject variability for phoneme duration was observed indicating that this could be considered as one of the parameters for speaker verification. The results indicated that greater than 67% of the measures were different across subjects and 61% of the measures were different within subjects. It was suggested that two speech samples can be

considered to belong to the same speaker when not more than 61% of the measures are different and two speech samples can be considered to be from different speakers when more than 67% of the measures are different.

According to Jakkar (2009), the benchmark for speaker identification using cepstrum was 88.33% (live Vs live), 81.67% (mobile Vs mobile) among 20 Hindi speakers. Srividya (2010) indicated higher percent correct identification for /u:/ 70% and at chance identification (50% identification each) for vowel /a:/ and /i:/.

Medha (2010) percent correct identification for females in /a:/ 40 %, /i:/ 40%, /u:/ 20 % and for males /a:/ 80 %, /i:/ 80 % and /u:/ 20 %. High vowels /i:/ and /u:/ had higher percent correct identification compared to vowel /a:/. Vowels /u:/ and /i:/ had highest and lowest mean normalized quefrency in direct and mobile recording and are identified better than vowel /a:/ and quefrency is inversely proportional to F0 and high vowels have higher F0 compared to low vowels.

Chandrika (2010) compared the performance of speaker verification system using MFCCs while recording with mobile handsets over a cellular network against a digital recording using long vowels /a:/, /i:/ and /u:/. Ten subjects participated in the study and they were provided with words containing the long vowels (/a: /, /i: / and /u:/). Speakers were given CDMA handset (Reliance, LG). MFCC values were extracted and the results revealed that the overall performance of the speaker verification system was about 80%. The overall performance of speaker recognition was 90% to 95% for /i:/. The accuracy of performance for vowel /i:/ was marginally better than vowel /a:/ and /u:/.

Ramya (2011) used MFCCs for speaker identification and the results indicated that the percent correct identification was above chance level electronic vocal disguise for females. Interestingly vowel /u: / had higher percent identification (96.66%) than vowels /a:/ 93.33% and /i:/ 93.33%.

Bhattacharjee (2013) did a comparative study of LPCC and MFCC features for the recognition of Assamese phonemes. He found that the performance of the system degrades considerably with the change in the training and testing conditions.

It has been observed that under the same environmental condition, when different set of speakers are used for training and testing the system, LPCC gave a recognition accuracy of 94.13%, whereas MFCC gave 89.14%. Thus LPCC appears to give a better representation of speaker independent contents of the speech signal whereas; MFCC captures some of the speaker dependent properties. However, in noisy conditions it has been observed that MFCC based system gave a relatively robust performance compared to LPCC. At 20dB SNR MFCC based system gave 97.03% recognition accuracy whereas LPCC based system gave 73.76% recognition accuracy. Rana and Miglani (2014) found that MFCC used in Automatic speech recognition system provides 80% accuracy whereas, LPCC used in Automatic speech recognition gave 60% accuracy.

A study on "Benchmark for speaker identification using nasal continuants in Hindi in direct mobile and network recording" was conducted by Rida (2014). The aim was to establish benchmark for speaker identification for nasal continuants in Hindi using MFCC. The objective of her study was to provide benchmarks for MFCC for Hindi nasal continuants in mobile and network condition. Ten participants between the age range of 20-40 years with at least ten years of exposure to Hindi language as a mode of oral communication were included in the study. Materials included six Hindi sentences with bilabial, dental and velar nasals embedded in words in all positions. Participants were instructed to speak the sentences under two conditions- directly into the recording mobile (live) and through another mobile into the recording mobile phone (network)- three times at an interval of one minute. The network used for making the calls was Vodafone (GSM 900/ GSM 1800 MHz frequency) and the receiving network was also Vodafone on a sony Ericson xperia pro mobile phone. Analysis of the data was carried out using SSL work bench (Voice & Speech Systems, Bangalore, India) to extract Euclidean distances. A speaker was presumed to be identified correctly when the Euclidean distance between training and test samples was the least. Results indicated that the percent correct speaker identification was 100%, 90% and 100% for /m/, /n/ and /ŋ/ respectively when live recording was compared with live recording using MFCC. Results indicated that the percent correct speaker identification was 50%, 80% and 90% for /m/, /n/ and /ŋ/ respectively when network recording was compared with network recording using MFCC. Results indicated that the percent correct speaker identification was 80%, 70% and 50% for /m/, /n/ and /ŋ/ respectively when live recording was compared with

network recording under telephone equalized condition using MFCC. Results indicated that the percent correct speaker identification was 90%, 90% and 30% for /m/, /n/ and /ŋ/ respectively when live recording was compared with network recording under telephone not equalized condition using MFCC. Results indicated that nasal continuant /ŋ/ had the best percent correct speaker identification among the nasals except under telephone equalized and not equalized conditions

It is evident from the review that MFCCs is, perhaps, the best parameter for speaker identification and less susceptible to variation of the speaker's voice and surrounding environment (noise). Also, the vowels may be the most suitable, among speech sounds, for speaker identification. However, till date there are limited studies on vowels as strong phonemes for speaker identification using semi-automatic methods. Scientific testimony impresses any court of law in whichever country that might be. However for any result to be called scientific, it has to be measured, quantified and reproducible if and when the need arises. Therefore, a method to carry out these analyses becomes a must. In this context, the present study is planned.

2.2 Need for the study

Recent researchers have used Cepstral Coefficients (Jakkhar, 2009; Medha, 2010; Sreevidya, 2010) and Mel Frequency Cepstral coefficients (Plumpe, Quateri & Reynolds, 1999; Hassan, Jamil & Rahman, 2004; Chandrika, 2010; Tiwari, 2010) to identify speaker. Mel-frequency Cepstral Coefficients (MFCCs) is a spectral feature extensively used in practical speaker identification systems. MFCCs are computed by warping the frequency domain of the speech signal to the Melody (Mel) scale (Reynolds, 1995; Beigi, 2001; Kinnunen & Li, 2009) with the aid of a psycho-acoustically motivated filter bank, followed by logarithmic compression and discrete cosine transform (DCT) (Kinnunen & Li, 2009). MFCC parameter have been widely used for speaker identification but there are dearth of methods and studies which make use of MFCC on vowels for the purpose of speaker identification on same individuals. Hence there is a need for instigate as to what percent matching would indicate similarity/dissimilarity of speaker or various features for speaker identification using WORKBENCH.

2.3 Aim

The aim of the present study was to obtain the percentage of correct speaker identification among Kannada speaking individuals and thus establish a benchmark for speaker identification using Mel frequency Cepstral coefficients (MFCC) for the vowels following the nasal continuants in Kannada language.

2.4 Objective

Establish a benchmark for speaker identification using Mel frequency Cepstral coefficients (MFCC) for the vowels following the nasal continuants in Kannada language.

CHAPTER III

METHOD

3.1 Participants

Kannada speaking neuro-typical adult males from Mysuru is considered as participants. They had minimum of ten years of formal education with Kannada as one of the subjects and all the participants spoke the Mysuru dialect of Kannada language and were drawn from the work/residential place in and around Mysuru, Karnataka, India. A total of 20 male participants in the age range of 20-30 years were considered for the study. The inclusion criteria for the participating speakers were no history of speech, language and hearing problem, no associated psychological or neurological problems, and no reasonable cold or respiratory conditions at the time of recording and normal oral structure. Hearing was screened using Ling's sound test. Kannada Diagnostic Picture Articulation Test (KDPAT- Appendix A) (Deepa & Savithri, 2010) was administered by a Speech Language Pathologist to rule out any misarticulations to be present in the speech.

3.2 Material

Commonly occurring hypothetical Kannada meaningful words with long vowels /a:/, /i:/, /u:/ following the nasal continuants /m/ and /n/ embedded in twenty eight sentences formed the stimulus material (Appendix B). Among these sentences a total of 30 words with vowels following nasal continuants were only considered for the present study and are listed below.

/su**ma:**ru/
/ma:ta:ḍiḍ̯anu/
/ma:ṭ̯re/
/ma:va/
/ma:suṭ̯aḍ̯e/
/mi:se/

/mi:sala:giṭ̠a/

/mi:ri/

/sa**mi:**paviḍ̠e/

/ʃa:**mi:**lagiḍ̠a:ne/

/mu:r̈ka/

/mu:rṭ̠i/

/mu:ɭe/

/mu:da/

/mu:ru/

/na:tja/

/na:lige/

/na:nu/

/na:vu/

/na:jaka/

/ni:ṭ̠i/

/ni:tʃa/

/ni:ru/

/ni:lagiri/

/ni:du/

/nu:kida/

/nu:liga/

/nu:ru/

/nu:lu/

/nu:ṭ̠ana/

Out of four Trails (Trail I, II, III, and IV), vowels occurring consecutively two times in the Trial II and Trial III were selected for analysis The written materials were provided to the participants and were made familiarized before recording begins in a laboratory condition for each participant individually.

3.3 Procedure

Speech samples of participants were recorded individually. Participants were informed about the nature of the study and written consent (Appendix C) was taken from all the participants. The sentences were presented visually and participants were instructed to read the sentences in a normal modal voice. Recordings were done under two conditions, a) mobile (network) recording and b) direct recording. Direct (live) recording of maximum of four repetitions of these sentences were taken for the present study. The distance between the mouth and the dynamic microphone (Shure) was kept constant at approximately 10 cm. In the first recording the participants were given a mobile phone (Nokia) and a call was made to Gionee S5.5 smart phone. The network used for making the calls was Vodafone/Airtel and the receiving network was also Vodafone/Airtel on a mobile phone. A speaker participating in an experiment was given a mobile phone with network of Vodafone. A call was made to the participants' handset from another Vodafone/Airtel mobile phone with recording option held by the experimenter's Gionee S5.5 smart phone. Speech signal was recorded as the speaker utters the test sentences. All the mobile recordings were done at different places according to the participants' convenience and the noise was controlled as much as possible at that place. The recordings at the receiving end were saved by the experimenter in the microchip of the smart phone. Later the recorded sentences were uploaded to a computer for further analysis. The mobile recordings were done one week after the live recordings were carried out (contemporary and non-contemporary speech samples).

The mobile recordings were done in the first sitting and after two week of gap the direct (live) recordings was carried out (contemporary and non-contemporary speech samples). The Live recordings was done using Computerized Speech Lab (CSL 4500 model; Kay PENTAX, New Jersey, USA) in a laboratory condition where computer memory used a desired (16) Bit (analog-digital) converter at a required sampling frequency of 16 KHz. These files were stored in ***.wav format***. The mobile (network) recordings were converted into ***.wav files*** using adobe audition software so that analysis was carried out in an effective manner on a computer. All the files were opened in PRAAT software (Boersma & Weenink, 2009) and down sampled to 8 kHz.

Of the four recordings, the first recording was not to be analyzed as the material was novel to the participant and the second and third recordings was used for analysis and comparison. If any of the second/third recordings were not lucid, then the fourth recording was used. From the down sampled speech material the vowels (/a/, /i/, /u/) followed by nasal continuants /m/ and /n/ in initial, medial and final position were truncated manually from the samples depicted in the wide band bar type of spectrograms and were stored in folders in the name of the participant for the convenience of analysis using the PRAAT (Boersma & Weenink, 2009) software program.. Three complete cycles (approximately 300ms) of the vowel following the nasal continuant /m/ or /n/ was segmented (Figure 2) and pasted onto a particular file name convention. For Ex: For speaker 1, first sample, first session, first occurrence was given the file name as ***"(speakers name)_call_1m.wav*** and saved in a folder with the name ***spk1***. There were sample files (2 nasals * 3 vowels * 5 occurrences * 2 repetitions * 2 conditions= 120). Out of four repetitions, 2nd and 3rd repetitions were only considered for the present study. Similar pattern was followed for other participants. Converted samples were stored in separate folders for each participant. These were stored separately in two main folders by the name 'live' and 'mobile' recordings.

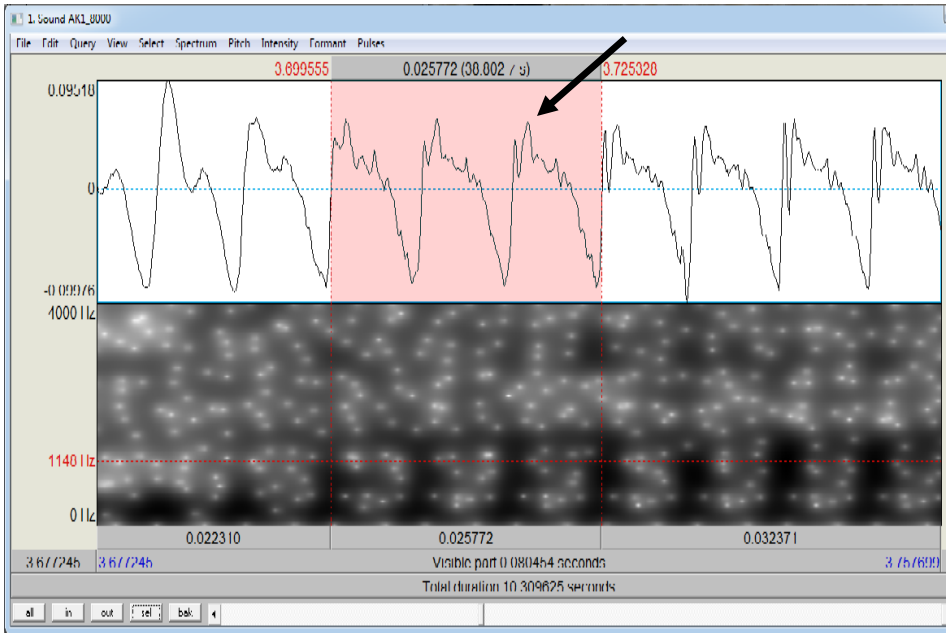

Figure 2: A segment of speech signal

3.4 Analysis

Speech Science Lab (SSL) Work bench, (Voice and Speech Systems, Bangalore, India) a Semi-Automatic vocabulary dependent speaker recognition software was used to extract Mel-Frequency Cepstral Coefficients (MFCC) for the truncated vowels following the nasal continuants.

The trail/repetitions and utterances of each recording were randomized by the software automatically and were considered as test set and training set on equal distribution. Seven samples for training and three samples for testing were taken. Thus, the SSL Pro.V4 software was used to test the performance of distance based, semiautomatic speaker recognition system, which is vocabulary dependent. Initially the file was specified using notepad in Workbench software (Figure 3) and .dbs file (Figure. 4), the extension of notepad file was created as mentioned below.

- Label: the phoneme or sound being analyzed (/a/, /i/, /u/)
- Number of speakers: the number of participants in the study (20)
- Number of occurrences of the label: the frequency of occurrence of a sound in a particular stimulus (/na/-5, /ni/-5 , /nu/-5) & (/ma/-5 , /mi/-5 , /mu/-5)
- Number of sessions: number of repetitions of the stimulus (four trails)

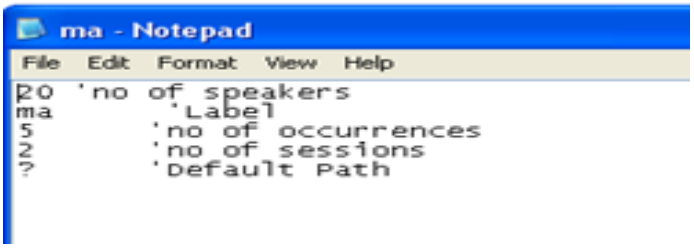

Figure 3: Illustration of the note pad

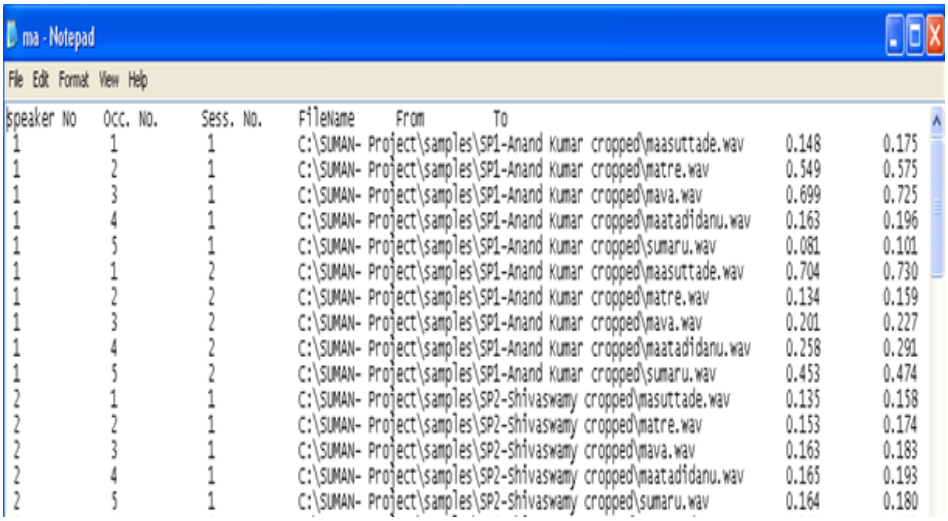

Figure 4: Illustration of .dbs file, the extension of notepad file

3.5 Segmentation

Followed by this, samples for analysis were segmented to the workbench software. To do this, the speaker number, session number and occurrence number were specified because averaging and comparison takes place between the same samples at different sessions. Figure 5 illustrate the speaker number being selected for segmentation and Figure 6 illustrate the session and occurrence number respectively.

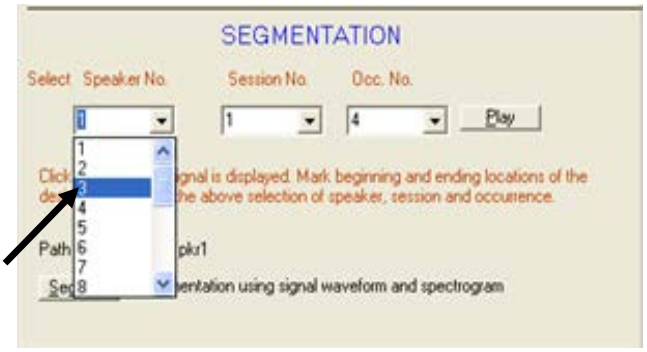

Figure 5: Illustration of speaker number being selected for segmentation.

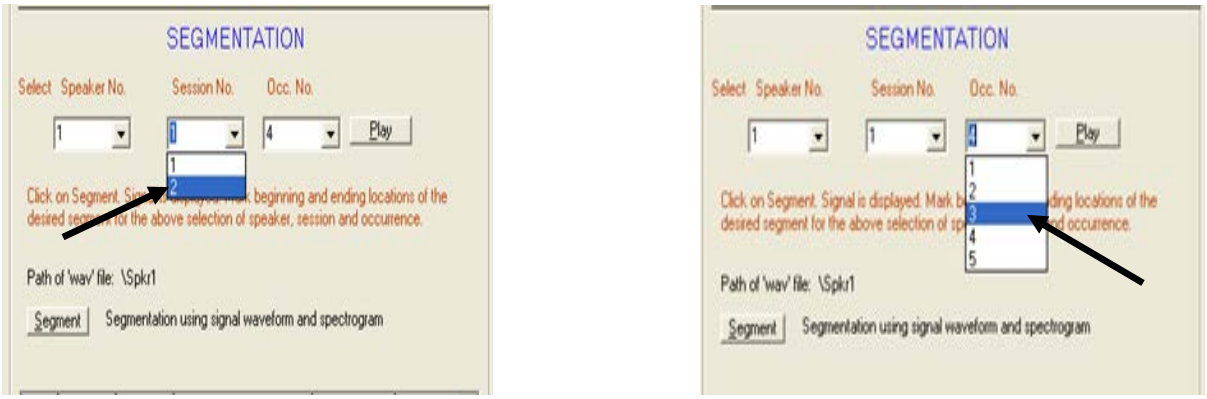

Figure 6- Illustration of selecting the session number and occurrence number

This required segmented file was selected and the option of 'assign highlighted' was selected from the 'Edit' option. Following this the assigned segment had to be confirmed (Figure 7).

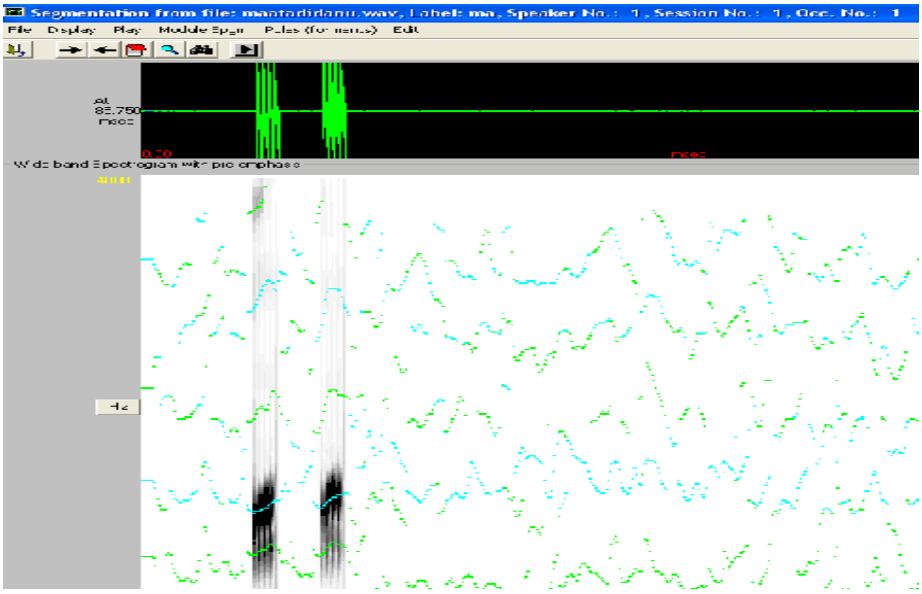

Figure 7: Depiction of segmentation window showing 2 sessions of vowel for a speaker

The segment of the file required was selected and the option of ‘assign highlighted’was selected from the ‘Edit’ menu. After this, confirmation was done. Figure 8, shows the dialogue box asking for confirmation of the highlighted segment in the file.

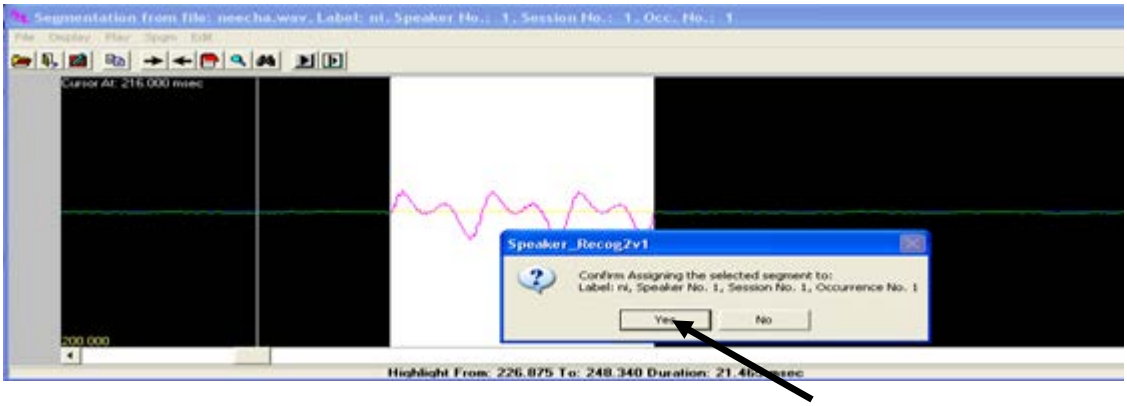

Figure 8: Showing dialogue boc asking for confirmation of the highlighted segment in the file.

Once when all the files were segmented for all the speakers we had to go to the ‘save segmentation’ option (Figure 9) and save the segmented files which were saved onto a .dbs file created as an extension of the notepad file.

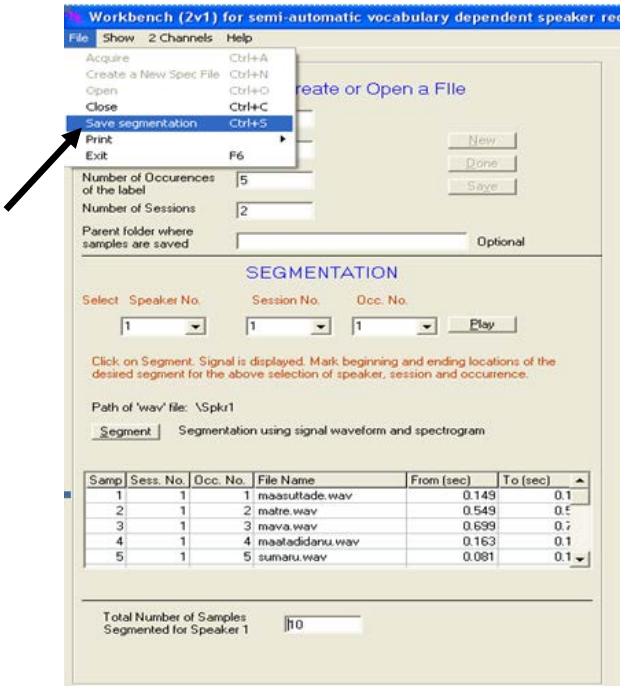

Figure 9- Illustration of ‘save segmentation’ option in workbench software.

Thus, as soon as all files were segmented the workbench software opens another window to train the samples randomly. The trail/repetitions and utterances of each recording were randomized by the software and were considered as test set and training set on 3:7 distribution (Figure 10). Training sample number was specified to be '3' and the rest '7'were automatically selected as test samples. Following this, 'compute' option was clicked on. This checked all the samples and compared them grossly and gave a qualitative analysis of each speaker. Later the 'testing' button was clicked on. After training, 13 MFCC were selected and the sample for identification was tested. Thus, the SSL Pro.V4 software was used to test the performance of distance based, semiautomatic speaker recognition system, which is vocabulary dependent.

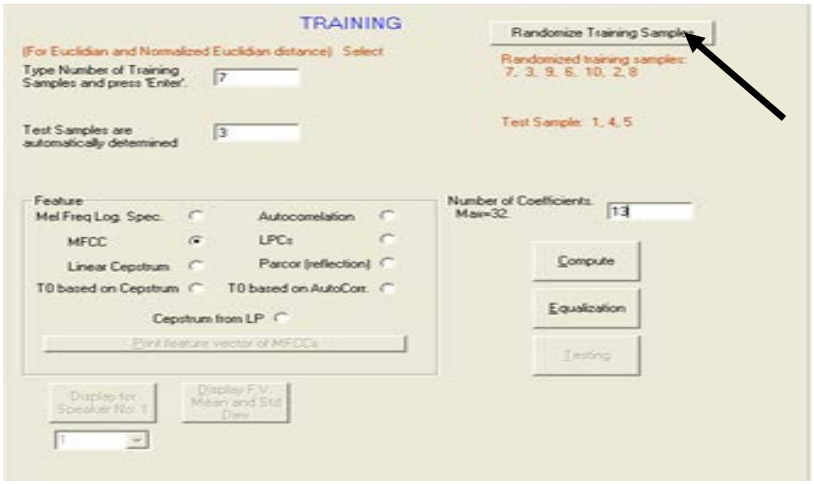

Figure 10: Analysis window of SSL workbench

Finally the software automatically generates the speaker identification threshold in terms of Euclidian Distance. Thus, the correct percentage of speaker identification was depicted after selecting the option of 'compute score for identification' as illustrated in Figure 11. The diagonal matrix in the lower half of the window and a final percentage for correct speaker identification was depicted. The same was selected for print and saved as .text file as illustrated in Figure 12. Thus, the data was stored and the same procedure was repeated at least for 30 times by randomizing the training and testing samples and the speaker identification thresholds was noted for the highest score and the lowest score.

The Euclidian distance of the samples were averaged by the software separately for the test and reference sample of the same speaker and were then compared against all the speakers.

The one with minimum displacement from reference was identified as the test speaker. If the test and the reference speakers were the same then it was considered as correct identification and if not it was considered as incorrect identification. Percentage corect identification was calculated by the formula Number of correct identification/Total number of speakers * 100. In this study, all the speech samples are contemporary, as all the recordings (live and mobile) of the participants were carried out in the different sessions. Closed set speaker identification tasks was performed, in which the examiner was aware that the 'unknown speaker' is one among the 'known' speakers.

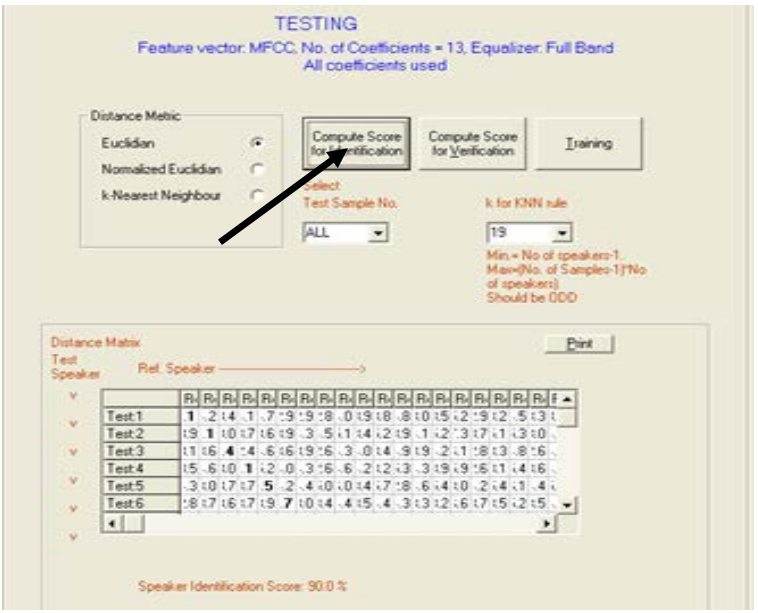

Figure 11: Analysis window of SSL workbench showing diagonal matrix and the final speaker identification score

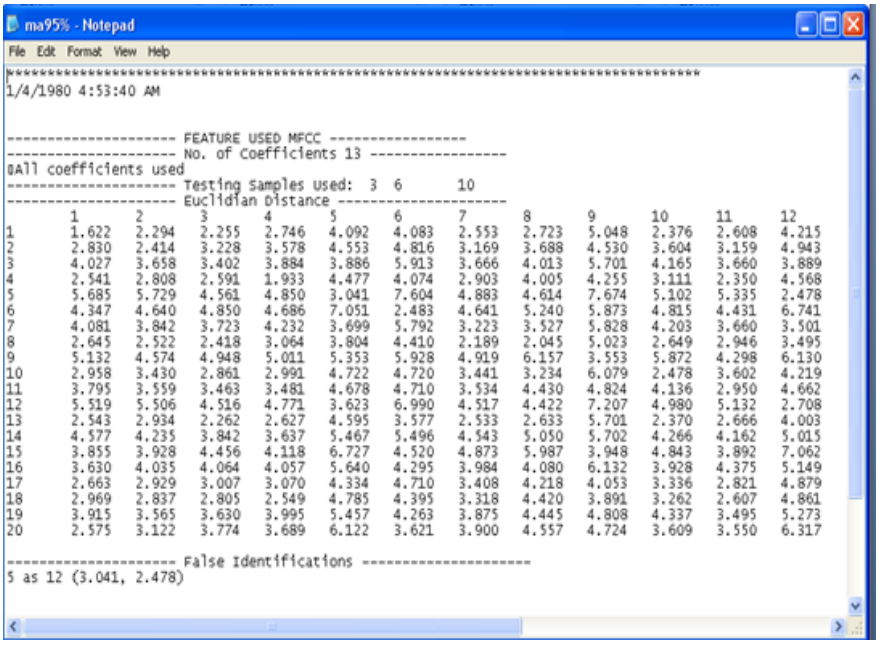

Figure 12- Speaker identification depicted in .dbs file

CHAPTER IV

RESULTS

The aim of the study was to establish a benchmark for speaker identification in Kannada using MFFCs derived from the vowels following the nasal continuants. The Euclidean distance of the samples for the reference and test samples of each speaker were averaged separately by the workbench software. This was then tabulated as a distance matrix comparing all the speakers. The one with the minimum distance from the reference was identified as test speaker. A distance matrix was computed by the software, for different combinations of test and reference speakers chosen. In this case, both the reference and test speakers were chosen from the live recordings. 30 combinations of 7 references and 3 tests (5 occurrences * 3 vowels /a:/, /i:/, /u:/ following the * 2 nasals /m/ & /n/ for each speaker) were chosen. Percentages of correct identification were calculated for the three categories (live verses live, mobile verses mobile and live verses mobile) and results of the study are discussed under the following sections:

Section A: Comparison of MFCC of the speakers- live recording vs. live recording for nasal continuants /m/ and /n/.

Section B: Comparison of MFCC of the speakers- mobile recording vs. mobile recording for nasal continuants /m/ and /n/.

Section C: Comparison of MFCC of the speakers- live recording vs. mobile recording for nasal continuants /m/ and /n/.

4.1. Section A: Comparison of MFCC of the speakers- live recording vs. live recording for nasal continuants /m/ and /n/.

Results indicating correct percent identification score for /ma:/, /mi:/, /mu:/, /na:/, /ni:/ and /nu:/ was noted to be 95%, 100%, 90%, 100%, 95% and 90% respectively. Table 1, 2, 3, 4, 5, 6 (Appendix D) depicts the highest correct speaker identification scores obtained out of thirty trials for the vowels following the nasal continuants like /ma:/, /mi:/, /mu:/, /na:/, /ni:/ and /nu:/ respectively when Live recording was compared with Live recording. The test sample was taken along the column and the reference average was taken along the row. The Euclidian distance of the samples were averaged by the workbench software separately for the test sample and the reference sample of the same speaker. These were then compared against all the speakers. As mentioned in the method section, the one with the minimum displacement from the reference was identified as the test speaker. The green color in the table indicates the correct identification of speaker sample as belonging to the same speaker as the reference sample. The red color in the table indicates the error identification of test sample as belonging to a different reference speaker. Table 7 and 8 depicts the speaker identification scores obtained for all thirty trials for the vowels following the nasal continuants /m/ and /n/ respectively. On comparison among the three vowels following the nasal continuant /m/, /i: / is better followed by /a: / and /u:/. Whereas for the nasal continuant /n/ the vowel /a: / is better followed by /i: / and /u: /. On an average of percentage of correct speaker identification of three vowels compared between the two nasal continuant /m/ and /n/, the vowels following the nasal /n/ (90%) and /m/ (90%) was similar.

Sl. No. of trails	Test samples from randomization	/ma:/	/mi:/	/mu:/
		Percentage Of correct identification	Percentage Of correct identification	Percentage Of correct identification
1.	3, 6, 8	45%	45%	45%
2.	2, 3, 8	50%	55%	50%
3.	5, 9, 10	55%	60%	55%
4.	2, 5, 9	60%	65%	60%
5.	2, 4, 6	65%	70%	65%
6.	1, 3, 9	70%	75%	70%
7.	2, 8, 9	75%	80%	75%
8.	2, 7, 10	80%	85%	80%
9.	5, 6, 10	85%	90%	**90%**
10.	4, 6, 9	90%	**100%**	50%
11.	1, 3, 7	**95%**	55%	55%
12.	2, 6, 7	50%	60%	60%
13.	3, 4, 8	60%	65%	65%
14	4, 8, 10	65%	70%	75%
15.	3, 7, 10	70%	75%	75%
16.	3, 6, 7	75%	80%	80%
17.	5, 6, 8	80%	85%	90%
18.	1, 3, 5	85%	90%	50%
19.	5, 6, 7	50%	60%	55%
20.	2, 4, 8	60%	65%	60%
21.	6, 7, 9	65%	70%	65%
22.	5, 6, 9	70%	75%	70%
23.	1, 5, 6	75%	80%	75%
24.	1, 3, 4	80%	90%	80%
25.	3, 6, 7	85%	65%	50%
26.	4, 5, 6	90%	70%	55%
27.	3, 4, 9	65%	75%	60%
28.	4, 6, 8	75%	80%	65%
29.	3, 5, 9	80%	75%	70%
30.	3, 6, 9	85%	80%	75%

Table 7- Speaker identification scores for thirty trials of vowels following the nasal continuants /m/ in live verse live recording.

Sl. No. of trails	Test samples from randomization	/na:/	/ni:/	/nu:/
		Percentage Of correct identification	Percentage Of correct identification	Percentage Of correct identification
1.	3, 6, 8	60%	60%	40%
2.	2, 3, 8	65%	65%	50%
3.	5, 9, 10	70%	70%	55%
4.	2, 5, 9	75%	75%	60%
5.	2, 4, 6	**100%**	**95%**	80%
6.	1, 3, 9	90%	90%	75%
7.	2, 8, 9	85%	85%	70%
8.	2, 7, 10	80%	80%	65%
9.	5, 6, 10	65%	70%	85%
10.	4, 6, 9	70%	75%	**90%**
11.	1, 3, 7	75%	80%	55%
12.	2, 6, 7	80%	85%	60%
13.	3, 4, 8	70%	75%	80%
14.	4, 8, 10	95%	70%	75%
15.	3, 7, 10	90%	95%	70%
16.	3, 6, 7	85%	90%	65%
17.	5, 6, 8	75%	80%	90%
18.	1, 3, 5	80%	85%	55%
19.	5, 6, 7	85%	90%	60%
20.	2, 4, 8	90%	95%	65%
21.	6, 7, 9	80%	85%	60%
22.	5, 6, 9	75%	80%	80%
23.	1, 5, 6	70%	75%	75%
24.	1, 3, 4	95%	70%	70%
25.	3, 6, 7	85%	90%	70%
26.	4, 5, 6	90%	80%	75%
27.	3, 4, 9	85%	85%	70%
28.	4, 6, 8	90%	90%	75%
29.	3, 5, 9	80%	85%	70%
30.	3, 6, 9	85%	90%	75%

Table 8- Speaker identification scores for thirty trials of vowels following the nasal continuants /n/ in live verse live recording.

4.2. Section B: Comparison of MFCC of the speakers- mobile recording vs. mobile recording for nasal continuants /m/ and /n/.

Results indicating correct percent identification score for /ma:/, /mi:/, /mu:/, /na:/, /ni:/ and /nu:/ was noted to be 90%, 80%, 70%, 90%, 85% and 90% respectively. Table 9, 10, 11, 12, 13, 14 (Appendix E) depicts the highest correct speaker identification scores obtained out of thirty trials for the vowels following the nasal continuants like /ma:/, /mi:/, /mu:/, /na:/, /ni:/ and /nu:/ respectively when mobile recording was compared with mobile recording. The test sample was taken along the column and the reference average was taken along the row. The Euclidian distance of the samples were averaged by the workbench software separately for the test sample and the reference sample of the same speaker. These were then compared against all the speakers. As mentioned in the method section, the one with the minimum displacement from the reference was identified as the test speaker. The green color in the table indicates the correct identification of speaker sample as belonging to the same speaker as the reference sample. The red color in the table indicates the error identification of test sample as belonging to a different reference speaker. Table 15 and 16 depicts the speaker identification scores obtained for all thirty trials for the vowels following the nasal continuants /m/ and /n/ respectively. On comparison among the three vowels following the nasal continuant /m/, /a: / is better followed by /i: / and /u: /. Similarly, for the nasal continuant /n/ the vowel /a: / and /u: / are better followed by /i: /. On an average of percentage of correct speaker identification of three vowels compared between the two nasal continuant /m/ and /n/, the vowels following the nasal /n/ (88.33%) was better compared to /m/ (80%).

Sl. No. of trails	Test samples from randomization	/ma:/	/mi:/	/mu:/
		Percentage Of correct identification	Percentage Of correct identification	Percentage Of correct identification
1.	2, 3, 7	60%	70%	65%
2.	2, 4, 10	70%	75%	60%
3.	4, 5, 9	80%	80%	65
4.	5, 7, 8	**90%**	50%	**70%**
5.	3, 9, 10	60%	65%	40%
6.	2, 6, 8	65%	55%	40%
7.	2, 3, 4	75%	65%	40%
8.	7, 8, 9	65%	70%	60%
9.	1, 8, 9	50%	**80%**	40%
10.	3, 6, 10	70%	70%	50%
11.	3, 8, 10	50%	65%	35%
12.	3, 7, 9	70%	65%	50%
13.	1, 3, 5	70%	70%	40%
14.	2, 5, 6	60%	60%	45%
15.	2, 3, 9	65%	80%	20%
16.	3, 8, 9	70%	65%	60%
17.	1, 2, 3	70%	70%	50%
18.	1, 4, 10	85%	45%	50%
19.	1, 3, 9	75%	40%	45%
20.	3, 6, 7	55%	70%	25%
21.	2, 6, 9	75%	65%	60%
22.	2, 6, 7	70%	80%	55%
23.	3, 7, 10	55%	75%	50%
24.	2, 3, 6	75%	55%	45%
25.	3, 5, 7	50%	80%	60%
26.	6, 8, 10	70%	60%	50%
27.	1, 6, 9	60%	70%	55%
28.	6, 7, 10	65%	60%	70%
29.	2, 4, 8	80%	70%	40%
30.	5, 7, 9	70%	75%	15%

Table 15- Speaker identification scores for thirty trials of vowels following the nasal continuants /m/ in Mobile vs Mobile recording.

Sl. No. of trails	Test samples from randomization	/na:/	/ni:/	/nu:/
		Percentage Of correct identification	Percentage Of correct identification	Percentage Of correct identification
1.	2, 3, 7	70%	50%	75%
2.	2, 4, 10	65%	65%	70%
3.	4, 5, 9	85%	55%	50%
4.	5, 7, 8	70%	75%	70%
5.	3, 9, 10	70%	60%	60%
6.	2, 6, 8	**90%**	60%	65%
7.	2, 3, 4	80%	65%	80%
8.	7, 8, 9	60%	55%	55%
9.	1, 8, 9	70%	**85%**	**90%**
10.	3, 6, 10	80%	60%	50%
11.	3, 8, 10	65%	45%	60%
12.	3, 7, 9	80%	45%	70%
13.	1, 3, 5	60%	55%	55%
14.	2, 5, 6	75%	75%	50%
15.	2, 3, 9	75%	40%	65%
16.	3, 8, 9	75%	55%	60%
17.	1, 2, 3	75%	75%	80%
18.	1, 4, 10	80%	60%	75%
19.	1, 3, 9	85%	65%	75%
20.	3, 6, 7	85%	75%	85%
21.	2, 6, 9	80%	70%	65%
22.	2, 6, 7	80%	45%	60%
23.	3, 7, 10	70%	70%	75%
24.	2, 3, 6	80%	75%	70%
25.	3, 5, 7	75%	75%	70%
26.	6, 8, 10	80%	60%	60%
27.	1, 6, 9	65%	80%	80%
28.	6, 7, 10	70%	65%	55%
29.	2, 4, 8	75%	55%	70%
30.	5, 7, 9	70%	65%	55%

Table 16- Speaker identification scores for thirty trials of vowels following the nasal continuants /n/ in Mobile vs Mobile recording.

4.3. Section C: Comparison of MFCC of the speakers- live recording vs. mobile recording for nasal continuants /m/ and /n/.

Results indicating correct percent identification score for /ma:/, /mi:/, /mu:/, /na:/, /ni:/ and /nu:/ was noted to be 55%, 60%, 40%, 60%, 65% and 65% respectively. Table 17, 18, 19, 20, 21, 22 (Appendix F) depicts the highest correct speaker identification scores obtained out of thirty trials for the vowels following the nasal continuants like /ma:/, /mi:/, /mu:/, /na:/, /ni:/ and /nu:/ respectively when mobile recording was compared with mobile recording. The test sample was taken along the column and the reference average was taken along the row. The Euclidian distance of the samples were averaged by the workbench software separately for the test sample and the reference sample of the same speaker. These were then compared against all the speakers. As mentioned in the method section, the one with the minimum displacement from the reference was identified as the test speaker. The green color in the table indicates the correct identification of speaker sample as belonging to the same speaker as the reference sample. The red color in the table indicates the error identification of test sample as belonging to a different reference speaker. Table 23 and 24 depicts the speaker identification scores obtained for all thirty trials for the vowels following the nasal continuants /m/ and /n/ respectively. On comparison among the three vowels following the nasal continuant /m/, /i: / is better followed by /a: / and /u: /. Whereas, for the nasal continuant /n/ the vowel /i: / and /u: / are better followed by /a: /. On an average of percentage of correct speaker identification of three vowels compared between the two nasal continuant /m/ and /n/, the vowels following the nasal /n/ (63.33%) was better compared to /m/ (51.66%).

Sl. No. of trails	Test samples from randomization	/ma:/	/mi:/	/mu:/
		Percentage Of correct identification	Percentage Of correct identification	Percentage Of correct identification
1.	2, 3, 7	25%	30%	10%
2.	2, 4, 10	15%	30%	15%
3.	4, 5, 9	20%	35%	20%
4.	5, 7, 8	35%	55%	30%
5.	3, 9, 10	40%	50%	30%
6.	2, 6, 8	35%	50%	30%
7.	2, 3, 4	10%	15%	5%
8.	7, 8, 9	15%	15%	15%
9.	1, 8, 9	40%	45%	**40%**
10.	3, 6, 10	30%	50%	40%
11.	3, 8, 10	35%	35%	25%
12.	3, 7, 9	40%	40%	40%
13.	1, 3, 5	30%	25%	20%
14.	2, 5, 6	20%	25%	35%
15.	2, 3, 9	**55%**	40%	5%
16.	3, 8, 9	30%	40%	30%
17.	1, 2, 3	20%	15%	5%
18.	1, 4, 10	20%	30%	20%
19.	1, 3, 9	20%	45%	5%
20.	3, 6, 7	35%	35%	30%
21.	2, 6, 9	35%	60%	40%
22.	2, 6, 7	35%	50%	30%
23.	3, 7, 10	45%	**60%**	30%
24.	2, 3, 6	25%	30%	20%
25.	3, 5, 7	40%	50%	0%
26.	6, 8, 10	10%	20%	10%
27.	1, 6, 9	40%	50%	35%
28.	6, 7, 10	15%	15%	10%
29.	2, 4, 8	25%	30%	20%
30.	5, 7, 9	40%	50%	30%

Table 23- Speaker identification scores for thirty trials of vowels following the nasal continuants /m/ in Live vs Mobile recording.

Sl. No. of trails	Test samples from randomization	/na:/	/ni:/	/nu:/
		Percentage Of correct identification	Percentage Of correct identification	Percentage Of correct identification
1.	2, 3, 7	25%	30%	30%
2.	2, 4, 10	15%	35%	15%
3.	4, 5, 9	35%	35%	25%
4.	5, 7, 8	45%	45%	50%
5.	3, 9, 10	**60%**	40%	50%
6.	2, 6, 8	50%	55%	50%
7.	2, 3, 4	10%	10%	10%
8.	7, 8, 9	10%	15%	20%
9.	1, 8, 9	40%	60%	**65%**
10.	3, 6, 10	55%	45%	45%
11.	3, 8, 10	35%	50%	50%
12.	3, 7, 9	35%	35%	35%
13.	1, 3, 5	30%	40%	25%
14.	2, 5, 6	30%	35%	25%
15.	2, 3, 9	30%	35%	15%
16.	3, 8, 9	35%	55%	35%
17.	1, 2, 3	10%	10%	5%
18.	1, 4, 10	30%	25%	15%
19.	1, 3, 9	25%	25%	15%
20.	3, 6, 7	40%	40%	35%
21.	2, 6, 9	20%	50%	45%
22.	2, 6, 7	50%	50%	50%
23.	3, 7, 10	55%	50%	35%
24.	2, 3, 6	25%	40%	25%
25.	3, 5, 7	30%	40%	25%
26.	6, 8, 10	15%	30%	25%
27.	1, 6, 9	25%	**65%**	50%
28.	6, 7, 10	10%	25%	20%
29.	2, 4, 8	35%	30%	15%
30.	5, 7, 9	40%	50%	45%

Table 24- Speaker identification scores for thirty trials of vowels following the nasal continuants /n/ in Live vs Mobile recording.

To summarize, the percent correct identification for /ma:/, /mi:/, /mu:/, /na:/, /ni:/ and /nu:/ was noted to be 95%, 100%, 90%, 100%, 95% and 90% respectively for live recording compared with live recording. For mobile verses mobile recording the percent correct identification was 90%, 80%, 70%, 90%, 85% and 90% respectively. And for live recording compared with mobile recording the percentage correct identification was 55%, 60%, 40%, 60%, 65% and 65% respectively. The same is represented graphically in Figure 13 and 14.

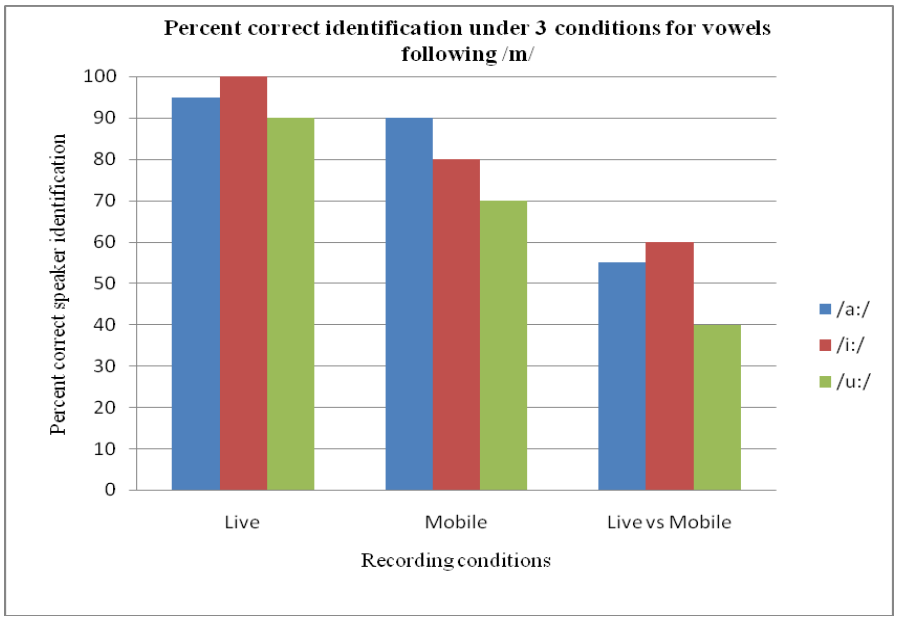

Figure 13: Percent correct identification in 3 conditions for vowel /a:/, /i:/ and /u:/ following /m/

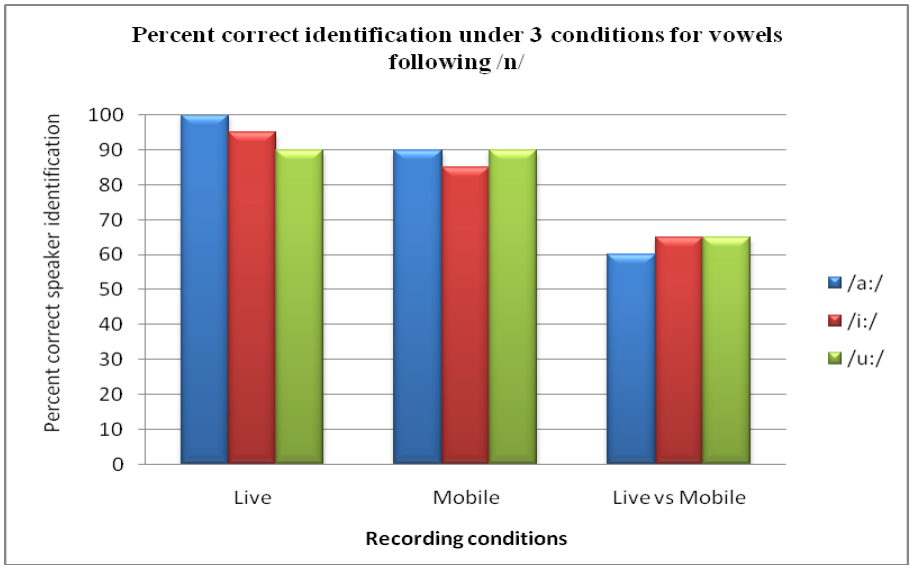

Figure 14: Percent correct identification in 3 conditions for vowel /a:/, /i:/ and /u:/ following /n/

The results indicated that the nasal /n/ had the best percentage of correct speaker identification in both mobile verse mobile (condition II) and live verses mobile (condition III) when compared to /m/. In Figure 15, the graphical representation depicts the difference between the nasal continuant /ma:/ verses /na:/ to be 5% for condition I and III and no difference for condition II. In Figure 16, /mi:/ verses /ni:/ the difference is 5% for all the three conditions (I, II, III) and finally in Figure 17, /mu:/ verses /nu:/, there was no difference for condition I and difference of 20% for condition II and 25% for condition III which is relatively higher. Thus, /n/ had the relatively best percent correct identification compared to /m/ nasal continuant.

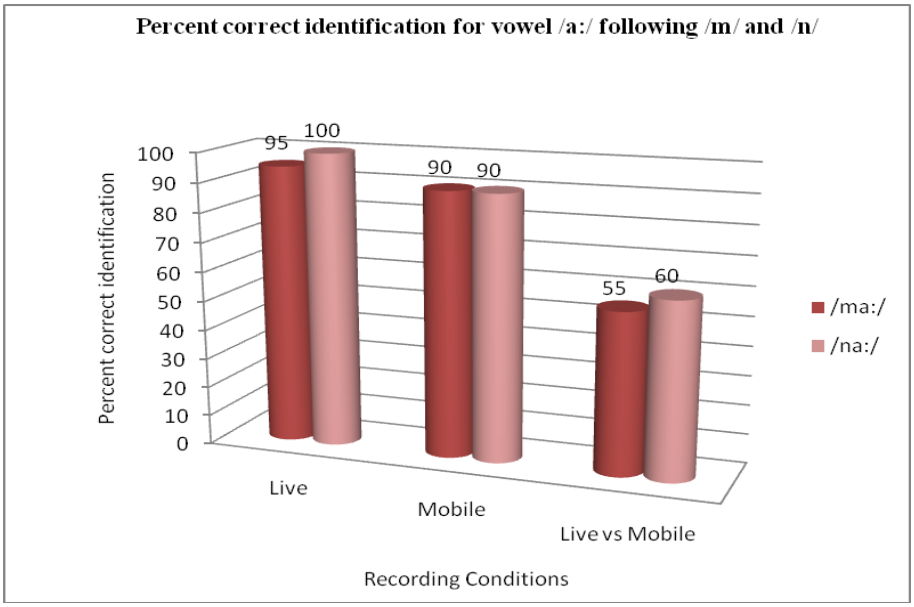

Figure 15: Difference in percent correct identification of nasal continuant /ma:/ verses /na:/

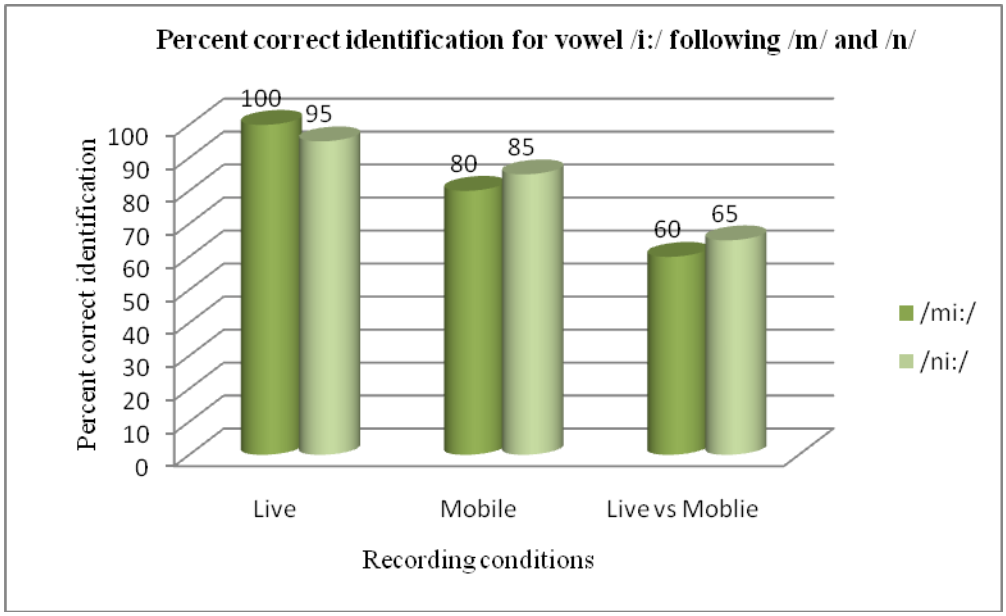

Figure 16: Difference in percent correct identification of nasal continuant /mi:/ verses /ni:/

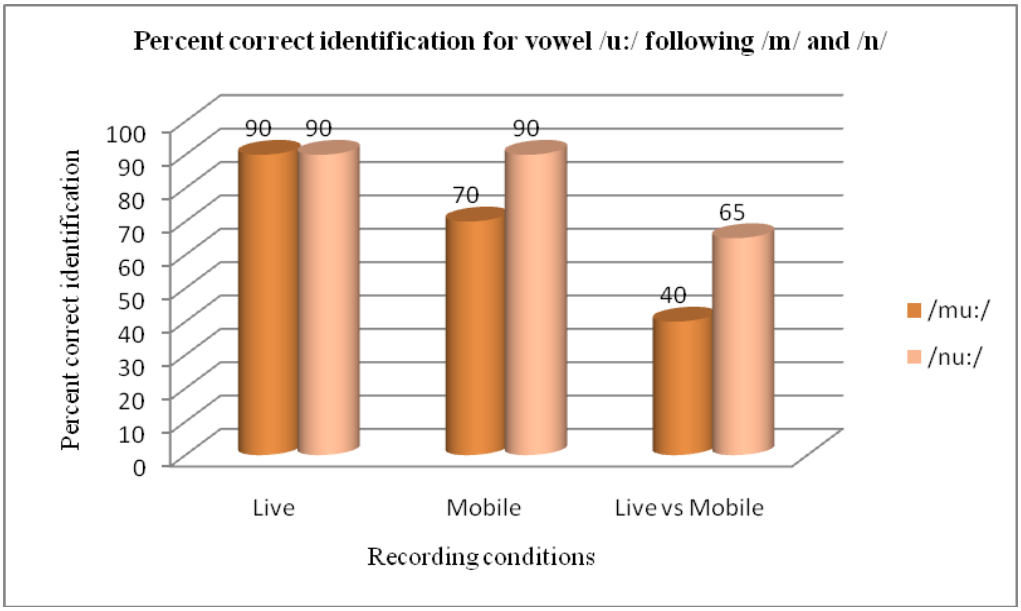

Figure 17: Difference in percent correct identification of nasal continuant /mu:/ verses /nu:/

CHAPTER V

DISCUSSION

The present study aimed at establishing a benchmark for speaker identification using MFCCs extracted from the vowel following the nasal continuants in Kannada language in both Live using Computerized Speech lab (CSL) and Mobile recordings and a comparison between the two. Initially in the comparison I, the live recordings speaker identification scores were found to be 95%, 100% and 90% for /ma:/, /mi:/ and /mu:/ respectively and 100%, 95% and 90% for /na:/, /ni:/ and /nu:/ respectively. Later, in the comparison II, the mobile recordings the speaker identification scores were 90%, 80% and 70% for /ma:/, /mi:/ and /mu:/ respectively and 90%, 85% and 90% for /na:/, /ni:/ and /nu:/ respectively. Finally, in comparison III, the mobile verse live recordings the speaker identification scores were 55%, 60% and 40% for /ma:/, /mi:/ and /mu:/ respectively and 60%, 65% and 65% for /na:/, /ni:/ and /nu:/ respectively.

The **identification scores between /m/ and /n/** were found to be the same in live recording condition but the score of /n/ was found to be better in both mobile recording and mobile vs live recording conditions. The accuracy scores decreased drastically in the mobile network condition when compared to the live recording condition. The scores decreased by around 5%, 20%, 20%, 10%, 10% and 0% for /ma:/, /mi:/, /mu:/, /na:/, /ni:/ and /nu:/ respectively from live to mobile recording. Using Cepstral measures Amino et al (2006) reported that coronal nasals were better in identifying a speaker than bilabial nasals. The studies done by Amino and Arai (2008) showed that the coronal nasals /n/ were more useful in identifying a speaker, when compared to a bilabial nasal /m/, in Japanese. They explained that this could be due to larger intra-speaker variability encountered in a bilabial nasal. To support in consonance, perceptual studies conducted by Amino and Arai (2009) state that coronal nasals were more reliable in identifying a speaker. Lakshmi (2011) conducted a study on Telugu nasal continuants using formant and bandwidth measures, which showed that nasals /ɳ/ and /ɳ̥/ were better for speaker identification compared to other nasals. The percent correct identification in the present study, interestingly, is very high in live recording.

This could be attributed to the characteristics of nasal continuants. Nasal continuants require two movements for its correct articulation, first the movement of tongue or lips to occlude the oral tract and second is the lowering of the velum. This contributes a unique quality to the spectrum produced (Pickett, 1980).

In live recording, **on comparison among the three vowels** following the nasal continuant /m/, /i: / is better followed by /a: / and /u:/. Whereas for the nasal continuant /n/ the vowel /a: / is better followed by /i: / and /u: /. In mobile recording, on comparison among the three vowels following the nasal continuant /m/, /a: / is better followed by /i: / and /u: /. Similarly, for the nasal continuant /n/ the vowel /a: / and /u: / are better followed by /i: /. In live verses mobile recording, on comparison among the three vowels following the nasal continuant /m/, /i: / is better followed by /a: / and /u: /. Whereas, for the nasal continuant /n/ the vowel /i: / and /u: / are better followed by /a: /. There are some studies which are partially in consonance with the present study. Chandrika (2010) reported that the overall accuracy using MFCCs extracted from long vowels /a:/, /i:/ and /u:/ was about 80% and the performance accuracy using vowel /i/ was 90% to 95%.

Ramya (2011), in her study reported an accuracy of 93.3%, 93.3% and 96.6% for the vowels /a:/, /i:/ and /u:/ respectively. The higher percentage of speaker identification using certain vowels in the above studies, might be attributed to the fact that the study was conducted in a controlled, laboratory environment, and the stimuli used were read out in a formal manner. However, the current study was carried out in a natural environment with some amount of ambient noise (Mobile recording) though the samples were read out by the participants.

On the other hand, Amino et al. (2006) compared the performance of nasal and oral sounds in speaker identification, using perceptual and acoustic analysis methods, reported greater inter-speaker distances while using nasals. Pruthi and Espy-Wilson (2007) extended Glass's and Zue (1995) work on detecting nasalized vowels in American English and selected a set of 9 knowledge based features for classifying vowel segments into oral and nasal categories automatically. The effectiveness of the nasals in speaker identification can be explained by the uniqueness of the morphology of the resonators. It is reported that the shapes of the nasal cavity

and paranasal sinuses are different among individuals (Dang & Honda, 1996). Also, the shapes of these resonators cannot be altered voluntarily.

Also, studies based on cepstral coefficients conducted by Amino and Osanai (2013), concluded that on an average, vowels were more efficient at identifying a speaker when compared to nasals. According to earlier studies the nasal regions of speech are an effective cue for speaker identification, because the nasal cavity is both speakers specific and fixed. Various acoustic features have been suggested to detect nasality.

On **comparison between conditions**, the comparison III (Live verses Mobile), the percent correct speaker identification is lower compared to comparison I (Live verse Live) and II (Mobile verse Mobile). The reason could be during the transmission of voice signals through communication channels, the signals are reproduced with errors caused by distortions from the microphone and channel, and acoustical, electromagnetic interferences and noises affecting the transmitting signal. Since, the network used in the present study is Vodafone and Airtel (GSM 900/GSM 1800 MHz). In general, GSM (Global System for Mobile Communications) is the pan-European cellular mobile standard. Where speech coding algorithms that are part of GSM compress speech signal before transmission, reduce the number of bits in digital representation but at the same time, maintain acceptable quality. Since this process modifies the speech signal, it can have an influence on speaker recognition performance along with perturbations introduced by the mobile cellular network (channel errors, background noise) (Barinov, Koval, Ignatov & Stolbov, 2010).

These distortions change the formant's energy and position which are crucial for speaker identification. Barinov, Koval, Ignatov and Stolbov conducted a study in 2010 to examine the characteristics of speech transmitted over a mobile network. They concluded that the non-linearity of the GSM channel's frequency response in the range 750-2000 Hz might cause a change in the energy distribution and affect 2^{nd} and 3^{rd} formants (F2 and F3). They also reported a fall-off in the channel's frequency response at 3500 Hz which led to the shifting of the fourth formant (F4) which might affect the MFCC.

Ridha (2014) reported similar results when mobile network recording was compared with mobile network recording i.e., the scores dropped drastically by about 50% for /m/, 10% for /n/ and 10% for /ɳ /. She also reported scores of 50%, 80% and 90% for the nasals /m/, /n/ and /ɳ/. This could be due to the loss of information over the network frequency bandwidth (900/1800 in Vodafone). This limitation might have masked the characteristics of nasals useful in identifying a speaker.

Overall, the speaker identification scores obtained in the Live vs Live condition was better than the scores obtained for the Mobile recording vs Mobile recording and Live vs Mobile recording condition. The mobile recordings were done in a natural environment, without controlling parameters such as background noise. This might be the reason for not achieving 100% percent correct speaker identification in this present study.

Limitations and future directions

The study was conducted with a limited number of speakers and considering only the male participants. The commonly occurring nasal continuants /m/ and /n/ were only chosen for the present study. Future studies on other nasals in Kannada with large number of speakers in comparison with other Indian languages and more number of occurrences of nasal continuants can be experimented for speaker identification.

CHAPTER VI

SUMMARY AND CONCLUSION

Biometrics refers to the identification of a person's identity based on his/her traits. Among all the biometric features verification of individuals identity based on voice has significant advantages and practical utilizations because speech is the most naturally produced and compelling biometric where it does not require a specialized input device, therefore the user acceptance of the system would be high. Forensic Speaker Identification is seeking an expert opinion in the legal process as to whether two or more speech samples are of the same person.

Rose (1992), Fururi (1994) and Nolan (1997) categorized speaker recognition as speaker identification and speaker verification. Speaker recognition is the process of automatically recognizing the speaker based on the information included in speakers' voice. Hecker (1971) describes it as any decision making process that uses speaker dependent features of speech signal. The main goal is to identify the speaker by extraction, characterization and recognition of the speaker-specific information contained in the speech signal.

The usage of mobile phones has increased exponentially and the rate of its usage in committing crimes has also dramatically increased. When a crime is committed through telecommunication, voice is the only evidence available for analysis. Hence there is a tendency to disguise one's voice to conceal their identities especially while making threatening calls, kidnapping or extortion. (Ramya, 2013)

Vowels, nasals and fricatives (in decreasing order) are commonly recommended for voice recognition because they are relatively easy to identify in speech signals and their spectra contain features that reliably distinguish speakers. The most frequently occurring consonant in Mysuru dialect of conversational Kannada language is nasals (Sreedevi, 2013). Studies have found that the nasal consonants can have a greater effect on the neighboring vowels be it preceding or following.

According to Su, Li and Fu (1974), the co-articulation between /m/ and the following vowel context can be used as an acoustic clue for identifying speakers which is more reliable than nasal spectra and also because it concerns a rapid event, it is not likely to be consciously modified in natural speech. Power spectrum of nasal consonants and co-articulated nasal spectra provide strong cues for the machine matching of speakers. Glass (1984) found that nasal consonants can be detected 88% of the times, while a vowel adjacent to a nasal consonant can be detected 74% of the times.

Researchers have used Cepstral Coefficients (Jakkhar, 2009; Medha, 2010; Sreevidya, 2010) and Mel Frequency Cepstral coefficients (Plumpe, Quateri & Reynolds, 1999; Hassan, Jamil & Rahman, 2004; Chandrika, 2010; Tiwari, 2010) to identify speaker. MFCC parameters have been widely used for speaker identification. So the current study aimed to instigate the percentage of speaker identification among Kannada speaking individuals and thus establish a benchmark for speaker identification using Mel frequency Cepstral coefficients (MFCC) for the vowels following the nasal continuants in Kannada language. The dearth of methods and studies that make use of MFCC on vowels for the purpose of speaker identification on same individuals validated the need for the study.

Twenty male participants between the age of 20 and 30 years were chosen for this study. They were native speakers of Kannada and had no history of speech, language or hearing difficulties. Twenty eight hypothetical meaningful Kannada sentences were chosen for the study. These sentences consisted of vowels /a: /, /i: / and /u: / following the nasal continuants /m/ and /n/ in the initial, medial and final positions of words in the sentences. 5 occurrences of each vowel following the nasals were considered for the study. The participants were asked to repeat each sentence four times.

Live recording was done using computerized speech lab (CSL) and Mobile recordings were done using Nokia 101 and Gionee S 5.5. A call was made from the Nokia to Gionee phone where it was recorded and saved. The recorded samples were transferred to computer memory and analyzed using SSL- Workbench for Semi-Automatic vocabulary dependent speaker recognition (Voice and Speech Systems, Bangalore, India) software. From the stimuli, the vowel portions (following the nasal continuants, >30msec) were segmented and stored.

Every speaker was represented by a total of ten occurrences of each vowel following the nasal continuant /m/ and /n/ for each condition (live, mobile and live vs mobile). The analysis performed was of three types:

- Live vs Live
- Mobile vs Mobile
- Live vs Mobile

In the live vs live condition, the reference and the test sample were obtained from the live recording. For the mobile network vs mobile network condition, the reference and test samples were obtained from the mobile recordings and for Live vs Mobile recording the reference and test samples were obtained from both live and mobile recordings. MFCCs derived from the vowels following the nasal continuants were used to compute the Euclidean distance between the test and reference samples. For the present study, the feature vector chosen was MFCC with 13 coefficients. Upon choosing the feature vector, the system computed a measure of distance (Euclidean distance) and displayed the summarized distance matrix for the selected test and reference sample. From the distance matrix, the total percentage of correct speaker identification score was displayed.

The highest percentage of correct speaker identification was obtained out of thirty trials for the vowels following the nasal continuants when Live recordings were compared with Live recordings. Results indicated a correct percentage of identification of 95%, 100%, 90%, 100%, 95% and 90% for /ma:/, /mi:/, /mu:/, /na:/, /ni:/ and /nu:/ respectively. When mobile recordings were compared with mobile recordings the results were 90%, 80%, 70%, 90%, 85% and 90% for /ma:/, /mi:/, /mu:/, /na:/, /ni:/ and /nu:/ respectively and finally when live recordings were compared with mobile recordings, the results were 55%, 60%, 40%, 60%, 65% and 65% for /ma:/, /mi:/, /mu:/, /na:/, /ni:/ and /nu:/ respectively.

The current study shows that the vowels following both the nasals /m/ and /n/ were reliable for speaker identification when live recordings were compared with live recordings. Whereas, when mobile recordings were compared with mobile recordings and live recordings were compared with mobile recordings vowels following the nasal /n/ was found to be better

than the vowels following the nasal /m/. This can be attributed to the study done by Amino et al (2006) which states coronal nasals are better in identifying a speaker than bilabial nasals, using cepstral measures. Also, perceptual studies conducted by Amino and Arai (2009) state that coronal nasals were more reliable in identifying a speaker. The effectiveness of the nasals in speaker identification can be explained by the uniqueness of the morphology of the resonators. It is reported that the shapes of the nasal cavity and paranasal sinuses are different among individuals (Dang and Honda, 1996). Also, the shapes of these resonators cannot be altered voluntarily.

On comparison among the three vowels, there are some studies which are partially in consonance with the present study. Chandrika (2010) reported that the overall accuracy using MFCCs extracted from long vowels /a:/, /i:/ and /u:/ was about 80% and the performance accuracy using vowel /i/ was 90% to 95%. Ramya (2011), in her study reported an accuracy of 93.3%, 93.3% and 96.6% for the vowels /a:/, /i:/ and /u:/ respectively. The higher percentage of speaker identification using certain vowels in the above studies, might be attributed to the fact that the study was conducted in a controlled, laboratory environment, and the stimuli used were read out in a formal manner. However, the current study was carried out in a natural environment with some amount of ambient noise (Mobile recording) though the samples were read out by the participants.

Also, studies based on cepstral coefficients conducted by Amino and Osanai (2013), concluded that on an average, vowels were more efficient at identifying a speaker when compared to nasals. According to earlier studies the nasal regions of speech are an effective cue for speaker identification, because the nasal cavity is both speakers specific and fixed. Various acoustic features have been suggested to detect nasality.

On comparison between conditions, the comparison III (Live verses Mobile), the percent correct speaker identification is lower compared to comparison I (Live verse Live) and II (Mobile verse Mobile). The reason could be during the transmission of voice signals through communication channels, the signals are reproduced with errors caused by distortions from the microphone and channel, and acoustical, electromagnetic interferences and noises affecting the transmitting signal. The network used in the present study was Vodafone and Airtel

(GSM 900/GSM 1800 MHz). Since the speech coding algorithms process modifies the speech signal, it can have an influence on speaker recognition performance along with perturbations introduced by the mobile cellular network (channel errors, background noise) (Barinov, Koval, Ignatov & Stolbov, 2010).

This is an initial attempt towards speaker identification using MFCC for the vowels following the nasal continuants in Kannada language with only limited number of speakers and thus it would be generalized to lab condition. The results of the study might show a relative good benchmark for speaker identification using MFCC for a following vowel of nasal continuants. The present study is in consensus with the hypothesis of proving or disproving several reports. Thus, the variables like vowel, its position in a word, the co-articulatory effect with the following nasal consonant influence the MFCC in speaker identification and these variables related to stimulus acts as a cue for correct speaker identification.

References:

Amino, K., & Arai, T (2007). Effect of stimulus contents and speaker familiarity on perceptual speaker identification. *Journal of Acoustical Society of Japan, 28* (2), 128-130.

Amino, K. & Arai,T (2009). Speaker dependent characteristics of the nasals. *Forensic Science International, 158* (1), 21- 28.

Atal, B. S. (1972), Automatic speaker recognition based on pitch contours. *The Journal of the Acoustical Society of America, 52,* 1687-1697.

Atal, B. S. (1974). Effectiveness of linear prediction characteristics of the speech wave for automatic speaker identification and verification. *Journal the Acoustic Society of America, 55* (6), 1304- 1312.

Bhattacharjee, U. (2013). A comparative study of LPCC and MFCC features for the recognition of Assamese Phonemes. *International Journal of Engineering Research & Technology, 2*(1), 2278-0181.

Bricker, P.S., & Pruzansky, S. (1976). Speaker recognition: Experimental Phonetics. London: Academic press.

Campbell, R. A. (1977). Speaker Recognition: A tutorial, Proceedings of IEEE, 85 (9):1437-1462.

Chandrika. (2010). The influence of hand sets and cellular networks on the performance of a speaker verification system. Project of PGDFSST, University of Mysuru.

Flege, J. E (1988). Anticipatory and carry over nasal co-articulation in the speech of children and adults. *Journal of Speech and Hearing Research, 31*, 525- 536.

Fururi. S. (1994). An overview of speaker recognition technology. Proceeding of ESCA Workshop on Automatic Speaker Recognition, Identification and Verification, 1-8.

Glass, J. B. (1984). Nasal Consonants and Nasalized Vowels: An Acoustic Study and Recognition Experiment. Submitted in Partial Fulfillment of the Requirements for the Degrees of Master of Science and Electrical Engineering (Massachusetts Institute of Technology).

Hasan. R., Jamil, M., Rabbani, G., & Rahman, S. (2004). Speaker Identification using MelFrequency Cepstral Co-efficient. 3rd International Conference on Electrical and Computer Engineering.

Hecker, M. H. L. (1971). Speaker recognition: Basic considerations and Methodology, *The Journal of Acoustical Society of America.,* 49,138.

Hollien, (2002). Forensic Voice Identification. San Diego, CA: Academic Press.

Glenn. J. W. & Kleiner. N. (1968). Speaker Identification Based on Nasal Phonation. *Journal of Acoustic Society of America, 43*(2), 368-372.

Kersta. L. G. (1962). Voice Identification, Nature, 196, 1253-1257. In Nolan, 1983(ed), The Phonetic Bases of Speaker Identification. Cambridge: Cambridge University press

Kinnunen. T. (2003). Spectral features for automatic text independent speaker recognition. Unpublished Thesis, University of Joensuu, Department of Computer Sciences, Finland.

Lavner, J. M. D. (1994). Principles of Phonetics, Cambridge: Cambridge University Press.

Loong, J. L. C., Subari, K. S., Abdulla, M. K., Ahmad, N. N. & Besar, R. (2010). Comparison of MFCC and Cepstral Coefficients as a feature set for PCG biometric systems. World academy of Science, Engineering and Technology, 754- 758.

Lo-Soun Su, K. P. Li., & K. S. Fu (1974). Identification of speakers by use of nasal co-articulation. *Journal of the Acoustic Society of America, 56* (6), 1876-1882.

Mc Dermott, M. C., Owen & Mc Dermott, F. M. (1996). Voice identification: the aural spectrographic method. In P. Rose, 2002,(ed.), Forensic Speaker Identification. Taylor and Francis, London.

McGehee, F.(1937). Reliability of Identification of Human Voices. *The Journal of General Psychology, 17,* 249-271.

Milner B., Shao X. (2007).Prediction of fundamental frequency and voicing from mel-frequency cepstral coefficients for unconstrained speech reconstruction. *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 15, no.14, pp. 24-33,

Mukesh, R., Saloni, M. (2014) Performance Analysis of MFCC and LPCC Techniques in Automatic Speech Recognition. *International Journal of Engineering and Computer Science, 3*(8), 7727-7732.

Naik, J. (1994). Speaker Verification over the telephone network: database, algorithms and performance, assessment, *Proc. ESCA Workshop Automatic Speaker Recognition Identification Verification,* 31-38.

Nolan, F. (1983). *Phonetic bases of speaker recognition*, Cambridge: Cambridge University.

Nolan, F. (1997). Speaker recognition and forensic phonetics, in Hard castle and Laver (eds) (1997): 744-67.

Pamela.S. (2002). Reliability of voice prints. Unpublished dissertation of AIISH, University of Mysore

Patrica, L. L & Hamlet, S. L (1987). Co- articulation effects on the nasalization of vowels using nasal/ voice amplitude ratio instrumentation. *Journal of cleft palate, 24,* (4), 286- 290.

Ramya. B. M. (2013). Bench mark for speaker identification under electronic vocal disguise using Mel Frequency Cepstral Coefficients. Unpublished project of Post graduate Diploma in Forensic Speech Science and Technology submitted to University of Mysore, Mysuru.

Rose, P. (2002). Forensic Speaker Identification. Taylor and Francis, London.

Samuel, M. C. (2003). Labial coarticulation in Malayalam, Unpublished dissertation of AIISH, UOM.

Saravanan, E & Nataraja, N. P. (1998). Study of effect of transmission system on speech- A variable in Speaker Identification. Unpublished master dissertation submitted to University of Mysore.

Sreedevi, N. (2013). Frequency of occurrence of Phonemes in Kannada. Project funded by AIISH Research Fund (ARF).

Stevens, K. N. (1968). Speaker authentication and identification: A comparison of spectrographic and auditory presentations of speech material. *The Journal of the Acoustic Society of America, 44*, 1596-1607.

Stevens, K. N. (1971). Sources of inter and intra speaker variability in the acoustic properties of speech sounds, Proceedings 7th International Congress. Phonetic Science. Montreal, 206-227.

Volkmann, J, Stevens, S. S., & Newmann. (1937). A scale for the measurement of the psychological magnitude pitch, *The Journal of the Acoustical Society of America, 8* (3), 208.

Yingyong, Qi., & Robert, A. F. (1992). Analysis of nasal consonants using perceptual linear prediction. *Journal of the Acoustic Society of America, 91*, 3.

Zakia, R. A. (2014). Benchmark for speaker identification using Nasal Continuants in Hindi in Direct Mobile and Network Recording. Unpublished Dissertation of AIISH. Submitted to The University of Mysore.