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

JAIISH(2015)
Vol 34, pp. 63-75

Suman Suresh and Hema N.

Affiliations

[a]All India Institute of Speech and Hearing, Manasagangothri, Mysore, 570 006

Corresponding Author

Hema N.
Lecturer in Speech Sciences
Department of Speech-Language Sciences, All India Institute of Speech and Hearing, Mysore-06
E-mail:hema_chari2@yahoo.com

Key Words

Hypothetical
Mobile Transmission
Semi-Automatic
Electromagnetic Interference

Abstract

Aim was to obtain the benchmark for speaker identification using Mel Frequency Cepstral Coefficients (MFCC) on vowels following the nasal continuants in Kannada language. Participants chosen were twenty Kannada speaking male neuro-typical adults, in the age range of 20-30 years. Kannada meaningful words (30) with long vowels /a:/, /i:/, /u:/ following the nasal continuants /m/ and /n/ formed the material. Speech Science Lab Work bench, a Semi-Automatic vocabulary dependent speaker recognition software was used to extract MFCC for the truncated (PRAAT) long vowels. Results indicated higher percent correct identification for Condition I (live verse live recording). On comparison among the three vowels following the nasal continuant /m/, /i:/ is better followed by /a:/ and /u:/. Whereas for /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 nasal continuant, the vowels following the nasal /n/ (90%) and /m/ (90%) was similar. Condition II (Mobile verse Mobile) and Condition III (Mobile verse Live) was comparatively poorer than Condition I, thus the benchmark was obtained. Discussion concludes that 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 affect the transmitting signal. Where speech coding algorithms that are part of Global System for Mobile compress speech signal before transmission, reduce the number of bits in digital representation but at the same time, maintain acceptable quality.

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. Among the various 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.

But in forensic sciences, 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 based on speaker recognition method. Any decision making process that uses speaker dependent features of speech signal is speaker recognition according to Hecker (1971) and it can be speaker verification and speaker identification according to Rose (1992), Fururi (1994) and Nolan (1997). The main goal of speaker recognition method is to identify the speaker by extraction, characterization and recognition of the speaker-specific information contained in the speech signal (Reynold, 2002). Speaker verification is 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' (Nolan, 1983). Speaker identification aims to identify an unknown voice as one or none of a set of known speakers on comparison (Nolan, 1983, Naik, 1994). Speaker Identification can be done in three ways that is by listening, on visual method and the machine method which includes semi-automatic speaker identification and automatic speaker identification according to Bricker and Pruzansky (1976). Among these three available methods of speaker identification semi automatic method is the most accepted and used one.

In the field of forensic sciences, 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, speech is the only evidence available for analysis. Forensic voice samples usually differ in their recording mode and conditions, affecting the findings due to the variation among the acoustical parameters in terms of frequency, pitch and energy (Vasan, Mathur & Dahiya 2015). Therefore, there is a pressing need on the part of police and the magistrate for establishment of legal proof of identity from measurements of speech. 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.

The speech 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). In this process of voice identification, the perceptual and acoustical parameters of speech of the speakers are subjected to various short and long term acoustical analysis techniques and thus find a significant cue for speaker identification. Researchers have come up with several acoustic correlates and parameters that can be used to characterize nasalized vowels for analysis, synthesis, perception and recognition. Fant (1960) reviewed the acoustic characteristics of nasalization pointed out in the literature and from his own observations confirmed the reduction in the amplitude of the first formant due to an increase in its bandwidth, and the rise in the first formant frequency. In addition to this feature, Fujimura and Lindqvist (1971) observed a shift in the frequency of the first formant towards higher frequencies and the introduction of pole-zero pairs in the first (often below the first formant) and third formant regions.

To list other few, a study by Glenn and Kleiner (1967), showed that the power spectrum of acoustic radiation produced during the nasal phonation provides a strong clue to speaker identity. A study by 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.

A study by Glass and Zue (1985) has recommended six acoustic measures of nasalized vowels to be considered for recognition experiment. They are, reduction in the first formant amplitude (A1), the relationship between reduction in the first formant amplitude and the amplitude of the first harmonic and the difference between A1 and amplitude of the extra nasal poles P0 (one below the first formant at around 250-450 Hz) and P1 (the one above the first formant at around 1000 Hz). These are the acoustical parameter which is extensively used in recognition experiments.

The logarithm of the estimated acoustic signal in a spectrum can undergo Inverse Fourier Transform (IFT) and the resultant will be the called as Cepstrum. This Cepstrum is also a parameter used for speaker identification. According to Jakhar (2009), the benchmark for speaker identification using Cepstrum of three long vowels (/a:/, /i:/, /u:/) both live and telephone recording conditions. The results show 88.33% (live recording Vs live recording means the number of speakers selected will be the total number of participants and number of sessions will be two i.e. trail one and trail two live recording sample of the same speaker), 81.67% (mobile recording Vs mobile recording means the number of speakers selected will be the total number of participants and number of sessions will be two i.e. trail one and trail two mobile recording sample of the same speaker) among 20 Hindi speakers. The same method on Kannada language, Srividya (2010) indicated higher percent correct identification for /u:/ to be 70% and at chance identification of 50% for the vowel /a:/ and /i:/.

Other study by Medha (2010) on speaker identification using Cepstrum measurement in text independent data revealed various results like 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 where the quefrency is inversely proportional to F0 and high vowels have higher F0 compared to low vowels.

From the above literature review, for instance until now the studies where on listening and visual methods. But there are also a number of studies using the machine method which includes semi-automatic speaker identification and automatic speaker identification. In this machine method the major parameters assessed are the LPCC and MFCC. To list few, a study by Bhattacharjee (2013) on comparison between the 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.

Another study on Benchmark for speaker identification using nasal continuants in Hindi in direct mobile and network recording' was conducted by Rida (2014). Results indicated that the percent correct speaker identification was 100%, 90% and 100% for /m/, /n/ and /ɳ/ 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 /ɳ/ 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 /ɳ/ 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 /ɳ/ respectively when live recording was compared with network recording under telephone not equalized condition using MFCC. Results indicated that nasal continuant /ɳ/ 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. Since 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. Acoustically vowels are characterized by formant pattern, spectrum, duration and fundamental frequency. In the same study, the most frequently occurring consonant is nasals and /n/ being the highest. The mean percentage and standard deviation of frequency of phonemes /n/, /m/ and /ɳ/ is 7.59% (0.31), 2.8% (0.26) and 0.3% (0.1) respectively.

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.

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.

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. However, till date there are limited studies on the vowels following the nasal continuants as strong phonemes for speaker identification using semi-automatic methods. The present study is a text dependent (since the vowels adjacent to a nasal consonant is only considered for the study) and non-contemporary is where the recordings (test sample and reference sample) were not done at the same time.

Scientific authentication impresses any court of law in whichever country that might be. However for any result to be called scientific, it has to be measured first, 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. Researchers have used different parameters in a hierarchy to process the speech signal by the computer program for correct speaker identification. The exact parameters as such are the 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 (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 a speaker.

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 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 following nasal continuants for the purpose of closed set speaker identification. Hence there is a need to instigate as to what percent matching would indicate similarity/dissimilarity of speaker or various features for speaker identification using semi-automatic speaker recognizer. Thus the aim of the present study was to obtain the percentage of correct speaker identification among Kannada speaking individuals and hence establish a benchmark for speaker identification using Mel frequency Cepstral coefficients (MFCC) on the vowels following the nasal continuants in Kannada language.

Method

Participants: The participants chosen for the study were twenty Kannada speaking neuro-typical adults. These total twenty participants were in the age range of 20-30 years with a minimum of ten years of formal education in Kannada and were graduates and belonged to the same dialect of Kannada language usage (Mysuru dialect). These participants were selected from the work/residential place in and around Mysuru, Karnataka, India and were included in the study only on fulfilling certain specific inclusion criteria. The inclusion criteria of participants were no history of speech, language and hearing problem, normal oral structures and no other associated psychological and neurological problems. They were reasonably free from cold or other respiratory illness during recording. Hearing was screened using Ling's sound test administered by an Audiologist/Speech-Language Pathologist. Kannada Diagnostic Photo Articulation Test (KDPAT) (Deepa & Savithri, 2010) was administered by a Speech-Language Pathologist to rule out any misarticulations to be present in their speech.

Material : The material used was commonly occurring hypothetical Kannada meaningful words with long vowels /a:/, /i:/, /u:/ following the nasal continuants /m/ and /n/ embedded in twenty eight sentences (Appendix-A) formed the stimulus material. Among these sentences a total of 30 words with vowels following nasal continuants were considered for the present study. The same is listed in Table 1.

Table 1: List of words used as a stimulus material

Stimulus	
/suma:ru/	/na:tja/
/ma:ta:d̪id̪anu/	/na:lige/
/ma:t̪re/	/na:nu/
/ma:va/	/na:vu/
/ma:sut̪ad̪e/	/na:jaka/
/mi:se/	/ni:t̪i/
/mi:sala:git̪a/	/ni:t̪a/
/mi:ri/	/ni:ru/
/sami:pavid̪e/	/ni:lagiri/
/ʆa:mi:lagid̪a:ne/	/ni:du/
/mu:rk˃a/	/nu:kida/
/mu:rt̪i/	/nu:liga/
/mu:ɭe/	/nu:ru/
/mu:da/	/nu:lu/
/mu:ru/	/nu:t̪ana/

Procedure

Recording Software Speech samples of participants were recorded individually. Participants were informed about the nature of the study and written consent 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 recording and b) direct recording. A maximum of four repetitions of these sentences were taken for both live and mobile recordings. 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 Airtel and the receiving network was Vodafone 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 (Airtel network) mobile phone with recording option held by the experimenter's Gionee S5.5 smart phone. Speech signal was recorded as the speaker uttered the sentences. All the mobile recordings were done at different places according to the participants' convenience with some amount of ambient noise. 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. These recordings were in .amr format and the sampling frequency was 44100 Hz. The live recordings were done two weeks after the mobile 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 were 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.

Analysis Software All the files recorded in Computerized Speech Lab (to obtain the best quality of recording) were stored in .wav format. The mobile recordings were converted into .wav files using adobe audition 3.0 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 as that is the sampling frequency used in the WORKBENCH software for speaker identification.

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/ were segmented and

Table 2: Speaker identification scores for thirty trials of vowels following the nasal continuants /m/ in live verse live recording.

Sl. No. of trials	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 3: 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%

pasted onto a particular file name as per convention.

Speech Science Lab (SSL) WORKBENCH, (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 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. Initially the file was specified using notepad in Workbench software and .dbs file, the extension of notepad file was created by specifying the phoneme, speaker, number of sessions and occurrences and was then segmented. Once all the files were segmented for all the speakers they were saved and as soon as all the files were segmented the workbench software opens another window to train the samples randomly. The repetitions and utterances of each recording were randomized by the software and were considered as test set and training set on 3:7 distribution as mentioned earlier.

Training sample number was specified to be '3' and the remaining '7' were automatically selected as test samples. After training, 13 MFCCs were selected since the sampling frequency is 8 kHz and therefore the analysis can be done up to 4 KHz (frequency distribution of an individual's speech frequency ranges till 4 KHz), with in 4 KHz only 13 Mel-frequency cepstral co-efficient can be computed efficiently. Following this the sample for identification was tested and the results were computed. The software automatically generates the speaker identification threshold in terms of Euclidian Distance and thus, the correct percentage of speaker identification was calculated. The data was stored and the same procedure was repeated for 30 times by randomizing the training and testing samples and the speaker identification thresholds were noted for the highest score and the lowest score as shown in Table 2, 3, 4, 5, 6 and 7.

Table 4: 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	/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%

was calculated. The data was stored and the same procedure was repeated for 30 times by randomizing the training and testing samples and the speaker identification thresholds were noted for the highest score and the lowest score as shown in Table 2, 3, 4, 5, 6 and 7.

The Euclidian distance of the samples were averaged by the software for the test and reference sample of the same speaker and were then compared against all the speakers. One with minimum displacement from reference was identified as the test speaker. Closed set speaker identification tasks was performed, in which the examiner was aware that the 'unknown speaker' is one among the 'known' speakers.

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 and was then tabulated as a distance matrix comparing all the speakers. 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 three sections. Condition I is the comparison of MFCC of the speakers- live recording verses live recording for nasal continuants /m/ and /n/. Condition II is the comparison of MFCC of the speakers- mobile recording verses mobile recording for nasal continuants /m/ and /n/. The final Condition III is the comparison of MFCC of the speakers- live recording verses mobile recording for nasal continuants /m/ and /n/.

Condition I: 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. 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.

Condition II: 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. 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%).

Table 5: Speaker identification scores for thirty trials of vowels following the nasal continuants /n/ 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 6: 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	/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%

Condition III: 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. 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%). As a summary the results discussed above of these three sections are graphically represented in Figure 1 and Figure 2.

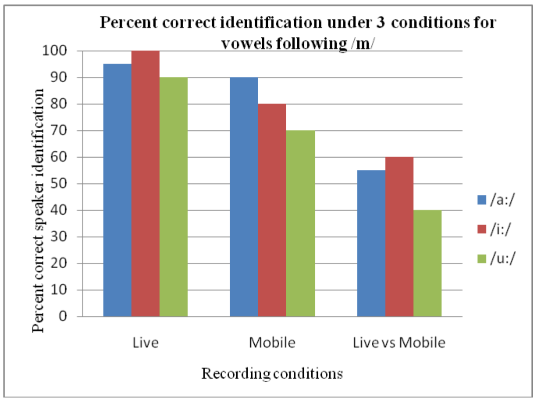

Figure 1: Percent correct identification in 3 conditions for vowel /a:/, /i:/ and /u:/ following.

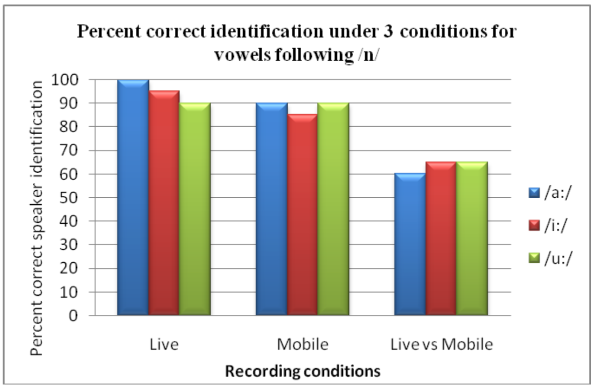

Figure 2: Percent correct identification in 3 conditions for vowel /a:/, /i:/ and /u:/ following.

Table 7: Speaker identification scores for thirty trials of vowels following the nasal continuants /n/ in Live vs Mobile recording.

Sl. No. of trials	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%

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 3, 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 4, /mi:/ verses /ni:/ the difference is 5% for all the three Conditions (I, II, III) and finally in Figure 5, /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.

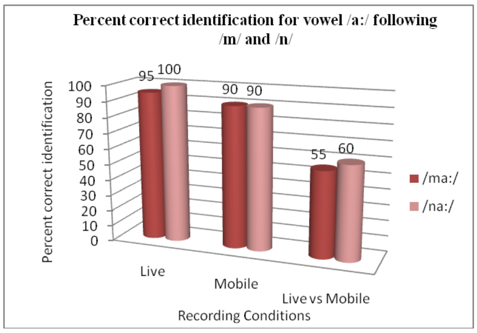

Figure 3: Percent correct identification in 3 conditions for vowel /a:/, /i:/ and /u:/ following.

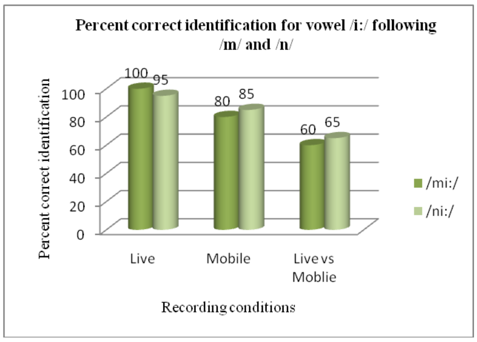

Figure 4: Percent correct identification in 3 conditions for vowel /a:/, /i:/ and /u:/ following.

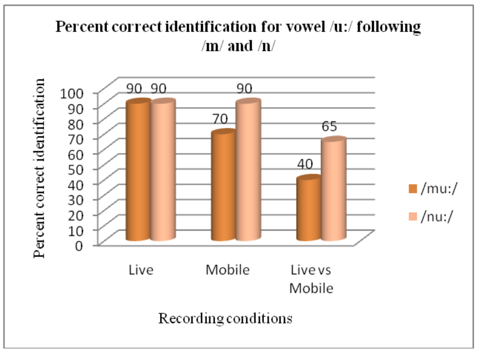

Figure 5: Percent correct identification in 3 conditions for vowel /a:/, /i:/ and /u:/ following.

Discussion

The identification scores between /m/ and /n/ were found to be the same in live recording condition (I) but the score of /n/ was found to be better in both mobile recording (II) and mobile vs live recording conditions (III). 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.

The initial point to support the present results is that coronal nasals were better in identifying a speaker than bilabial nasals according to the Cepstral measures reported by Amino et al (2006). The later studies done by Amino and Arai (2007) showed that the coronal nasals /n/ were more useful in identifying a speaker, when compared to a bilabial nasal /m/, in Japanese language. They explained that this could be due to larger intra-speaker variability encountered in a bilabial nasal. To support further and to be in consonance with the previous results, perceptual studies conducted by Amino and Arai (2009) also stated that coronal nasals were more reliable in identifying a speaker. Various nasal continuants in Telugu language were studied using formant and bandwidth measures by Lakshmi (2011). The results showed that nasals /n/ and /ɳ/ were better for speaker identification compared to other nasals. The percent correct identification in the present study, interestingly, is very high in live recording.

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:/.

The foremost point is 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.

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%. The 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. To add on, Amino et al. (2006) compared the performance of nasal and oral sounds in speaker identification using perceptual and acoustic analysis methods. They 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.

On comparison between conditions, the Condition III (Live verses Mobile), the percent correct speaker identification is lower compared to Condition 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 distribu-

Table 8: Benchmark for speaker identification using MFCC on the vowels following nasal continuants.

Nasals	/m/			/n/		
Vowels	/a:/	/i:/	/u:/	/a:/	/i:/	/u:/
Live vs Live (Condition I)	71.16%	73%	65.66%	77.83%	81.33%	68.83%
Mobile vs Mobile (Condition II)	68%	67%	48.33%	75%	63%	67%
Live vs Mobile (Condition III)	29%	37%	23%	32%	38%	32%

tion and affect 2nd and 3rd 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.

A study by 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/ when compared to live recording condition. 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. In conditions like mobile recording, application of noise reduction algorithms using definite standardized noise reduction software will be the future need and implication from this present research on speaker identification in forensic sciences. The resultant of the present study is the benchmark for speaker identification using MFCCs on vowels following the nasal continuants in Kannada as reported in Table 8.

Conclusions

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. 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.

On comparison between conditions, the Condition III (Live verses Mobile), the percent correct speaker identification is lower compared to Condition 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.

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 present study show a need to obtain a relative good benchmark for speaker identification using MFCC for a following vowel of nasal continuants. 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., Sugarwa, T., & Arai, T. (2006). Idiosyncrasy of nasal sounds in human speaker identification and their acoustic properties. *Acoustic Science and Technology, 27,* 233-235.

Amino, K., Sugarwa, T., & Arai, T. (2006). Effects of the syllable structure on perceptual speaker identification. *The Journal of the Institute of Electronics, Information and Communication Engineers (IEICE), 105,* 109-114.

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.

Amino, K., & Osanai, T. (2012). Speaker characteristics that appear in vowel nasalization and their change over time. *Acoustical Science and Technology, 33*(2), 96-105.

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

Atkinson, J. E. (1976). Inter and intra speaker variability in fundamental voice frequency. *Journal of the Acoustical Society of America, 60*(2), 440-445.

Beigi, H. (2011). *Fundamentals of Speaker Recognition.* Springer, New York. ISBN: 978-0-387-77591-3.

Bhattacharjee, U. (2013). A comparative study of Linear Prediction Cepstral Co-efficient (LPCC) and Mel- Frequency Cepstral Co-efficient (MFCC) features for the recognition of Assamese Phonemes. *International Journal of Engineering Research and Technology, 2*(1), 2278-0181.

Boersma & Weenink, D. (2009). PRAAT S.1.14 software, restricted from http://www.goofull.com/au/program/142\35/speedytunes.html.

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

Chandrika. (2010). *The influence of hand sets and cellular networks on the performance of a speaker verification system.* Project of Post Graduate Diploma in Forensic Speech Science Technology, University of Mysore.

Dang, J., & Honda, K. (1996). Acoustical modeling of the vocal tract based on morphological reality: Incorporation of the paranasal sinuses and the piriform fossa. Proceedings of 4th Speech Production Seminar, 49-52, Grenoble.

Deepa, A., & Savithri, S. R. (2010). Re-standardization of Kannada articulation test. *Student research at All India Institute of Speech and Hearing (Articles based on dissertation done at AIISH), 8,* 53-55.

Fant, G. (1960). *Acoustic Theory of Speech Production.* Netherlands: Mouton and Co., 's- Gravenhage, ISBN: 9027916004.

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

Fujimura, O., & Lindqvist, J. (1971). Sweep-Tone measurements of the vocal tract characteristics. *Journal of the Acoustical Society of America, 49*(2), 541-548.

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).

Glass, J. R., & Zue, V. W. (1985). Detection of nasalized vowels in American English. In: Proceedings of International Conference on Acoustics, Speech, and Signal Processing (ICASSP), 15691572.

Glass, J. R., & Zue, V. W. (1995). Detection of nasalized vowels in American English. In the Proceedings of International Conference on Acoustics, Speech and Signal Processing, 15691572.

Glenn. J. W. & Kleiner. N. (1967). Speaker Identification Based on Nasal Phonation. *The Journal of the Acoustical Society of America, 43*(2), 368-372.

Hasan, R., Jamil, M., Rabbani, G., & Rahman, S. (2004). Speaker Identification using Mel Frequency Cepstral Coefficient. 3rd International Conference on Electrical and Computer Engineering.

Hattori.S., Yamamoto.K., & Fujimura.O. (1958). Nasalization of vowels in relation to nasals. *Journal of the Acoustical Society of America, 30,* 267-274.

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

Hollien, H. (1990). *The Acoustics of Crime: The New Science of Forensic Phonetics.* New York. Plenum Press.

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

House, A. S., & Stevens, K. N. (1956). Analog studies of the nasalization of vowels. *Journal of Speech and Hearing Disorders, 22*(2), 218-232.

Jakhar, S. S. (2009). *Benchmark for speaker identification using Cepstrum.* Unpublished project of Post Graduate Diploma in Forensic Speech Science and technology, submitted to University of Mysore, Mysore.

Kinnunen, T. (2009). *Spectral features for automatic text independent speaker recognition.* Unpublished Thesis, University of Joensuu, Department of Computer Sciences, Finland.

Kawabara, H. & Sagisaks, Y., (1995). Acoustic characteristics of speaker individuality: control and conversion. *Journal of Speech Communication, 16,* 165-173.

Lakshmi, P., & Savithri. S. R. (2009). Benchmark for speaker Identification using Vector F1 & F2. *Proceedings of the International Symposium, Frontiers of Research on Speech and Music,* 38-41.

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

Markel, J. & Davis, S. (1979). Test independent speaker recognition from a large linguistically unconstrained time-spaced data base. *IEEE (Institute of Electronics and Electronic Engineers) Transactions on Acoustics, Speech, and Signal Processing, 27*(1), 74-82.

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

Medha, S. (2010). *Benchmark for Speaker Identification using Cepstrum measurement using Text-independent data.* Unpublished project of Post graduate Diploma in Forensic Speech Science and Technology submitted to University of Mysore, Mysore.

Naik, J. (1994). Speaker Verification over the telephone network: database, algorithms and performance, assessment,*Proceedings of ESCA (European Speech Communication Association) 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), 744-67.

Pruthi, T., & Espy-Wilson, C. (2006). An MRI based study of the acoustic effects of sinus cavities and its application to speaker recognition, *Proceedings of Interspeech,* Pittsburgh, 21102113.

Ramya. B. M. (2011). 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, Mysore.

Rana, M. & Miglani, S. (2014). Performance Analysis of Mel Frequency Cepstral Co-efficient and Linear Prediction Cepstral Co-efficient Techniques in Automatic Speech Recognition. *International Journal of Engineering and*

Computer Science, 3(8), 7727-7732.

Reynolds. D. A & Rose. R. (1995). Robust text-independent speaker identification using Gaussian Mixture speaker models. *IEEE (Institute of Electronics and Electronic Engineers) Transactions on Speech and Audio Processing,* 3, 72-83.

Reynolds. D. A. (2002). An Overview of Automatic Speaker Recognition Technology. *Proceedings in IEEE (Institute of Electronics and Electronic Engineers),* 4072-4075.

Ridha Z.A. (2014). Benchmark for speaker identification using Nasal Continuants in Hindi in Direct Mobile and Network Recording. Unpublished Dissertation of AIISH (All India Institute of Speech and Hearing). Submitted to The University of Mysore.

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

Soong. F., Rosenberg. A., Rabiner. L., & Juang. B. H. (1985). A vector quantization approach to speaker recognition. *Proceedings in the International Conference on Acoustic Signal Processing,* 387-390.

Sreedevi, N. (2012). Frequency of occurrence of Phonemes in Kannada. Project funded by AIISH (All India Institute of Speech and Hearing) Research Fund (ARF).

Sreevidya (2010). Speaker Identification using Cepstrum in Kannada Language. Unpublished project of Post Graduate Diploma in Forensic Speech Science and Technology submitted to University of Mysore, Mysore.

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

Stevens, K. N. (1956). 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.

Thompson, C. (1987). Voice Identification: Speaker Identificability and correction of records regarding sex effects. *Human Learning,* 4, 19-27.

Tiwari. V. (2010). Mel- Frequency Cepstral Co-efficient and its applications in speaker recognition Dept. of Electronics Engineering, Gyan Ganga Institute of Technology and Management, Bhopal, (MP) India.

Vasan, M., Mathur, S., & Dahiya, M. S. (2015). Effect of different recording devices on forensic speaker recognition system. Paper presented in 23rd All India Forensic Science Conference, Bhopal.

Wolf, J. J. (1972). Efficient acoustic parameter for speaker recognition. *The Journal of the Acoustical Society of America,* 20442056.

Appendix-A

HYPOTHETICAL SPEECH SAMPLE (Stimulus material)

1) /navilina naatya balu sundara/
2) /nanu muru idli tinde/
3) /maatre togo/
4) /maava nale bartare/
5) /batte haleyadadare maasuttade/
6) /sumaaru nalkaidu dina agirabeku/
7) /sarkara samaanyara kashta nashtagalige spandisabeku/
8) /halligalalli mooda nambikegalu hecchu/
9) /naalige moole illada anga/
10) /neeru jeeva jala/
11) /avanige batte needu/
12) /naavu oorige hogtidivi/
13) /naayi niyattina prani/
14) /innu ninna naataka mugitu/
15) /veerappange dodda meese ittu/
16) /reeshme noolu dubari/
17) /nanage nooru rupayi beku/
18) /naveena shaaleya noothana nayaka/
19) /avanu nannannu nookida/
20) /vishavallada hasiru haavannu nooliga annuttare/
21) /neelagiri parvatha karnatakadallide/
22) /avaladu neecha buddi/
23) /ajja neethi kathe heltare/
24) /murthy shatru sainyadodane shameelagiddane/
25) /ramana mane nanna maneya sameepavide/
26) /avanu mithi meeri maatadidanu/
27) /idu avanige meesalagitta kelasa/
28) /avanu obba moorkha/