Research Article

Closed Set Speaker Identification using Mel Frequency Cepstral Coefficients on Vowels Preceding Nasal Continuants in Kannada

Arjun M. Shivakumar and Rajasudhakar R.

JAIISH(2015)
Vol 34, pp. 76-84

Affiliations

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

Corresponding Author

Arjun. M. S.
Department of Speech-Language Sciences, All India Institute of Speech and Hearing, Mysore-06
Email: arjun.aiish@gmail.com

Key Words

Mel Frequency
Cepstral Coefficients
Speaker Identification
Semi-Automatic Method
Forensic Science
Nasal Continuant

Abstract

The aim of the present study was to obtain the percentage of speaker identification using vowels preceding nasal continuants in Kannada speaking adult individuals using semi-automatic method. The participants were twenty Kannada speaking adult males in the age range of 21-32 years constituted as Group I. This was further sub grouped as Group II constituting ten speakers. The material was meaningful Kannada words containing long vowels /a:/, /i:/ and /u:/ preceding nasal continuants /m/ and /n/ embedded in Kannada sentences. The participants read the material four times each under two conditions (a) live recording and (b) mobile network recording. The target words were truncated using the PRAAT software. Each vowel preceding nasal was subjected for extraction of Mel Frequency Cepstral Coefficients (MFCCs) using Speech Science lab Workbench for Semi-automatic speaker recognition software. The study was compared under three conditions: (a) Live vs live recording, (b) Mobile network vs mobile network recording and (c) Live vs mobile network recoding. The same was found across the three conditions when the participants reduced from twenty to ten in number. The results of the present study indicated quite high percent of correct speaker identification using MFCCs in Live vs Live and Mobile network vs Mobile network conditions compared to Live vs mobile network condition. The obtained outcome would serve as potential measure in the forensic scenario for identification of speakers using vowels preceding nasal continuants in Kannada.

Introduction

"As each one of the ridges of your fingers or on the palm of your hand differ from each other, so do all of the other parts of your body. They are unique to you, including your Voice Mechanisms" is a quote by Hollien (1990).

Forensic science is the scientific method of gathering and investigating information about the past which is then used in a court of law. It can also be defined more broadly as that scientific discipline which is directed to the detection or recognition, identification, individualization, and evaluation of physical evidence by the application of the principles and methods of natural sciences for the purpose of administration of criminal justice. One of the branches of forensic sciences is forensic speaker identification/recognition.

The most natural and common mode to communicate information by humans is through speech and the speech signal conveys several types of information. From the speech production point of view, the speech signal conveys linguistic information (example- message and language) and speaker information (example- emotional, regional, and physiological characteristics). Most of us are aware of the fact that voices of different individuals do not sound alike. This important property of speech of being speaker dependent is what enables us to recognize a friend over a telephone. The ability of recognizing a person solely from his voice (perceptually) is known as speaker recognition.

The need to establish the identity for identifying a person from his/her voice is important because of the legal ramifications and forensic involvements. In the present era of widely used telephone, mobile phone, radio and tape recorder communication, the only information available to investigators may consist of a single voice recording, generally made during a telephone or mobile phone conversation. In the legal process, forensic speaker identification is seeking an expert opinion to take a decision as to whether two or more speech recordings are of same person (Rose, 2002). Identification of speaker in forensic perspective is generally about comparing voices. The serious problem in forensic speaker identification is to recognize an unfamiliar speaker whose voice has been recorded

during an offense, for example ransom demand, a bomb threat, sexual abuse, hoax emergency call or drug deal. The experts compare the incriminating recording of speech samples from a suspect and make a decision to identify the person behind or eliminate the suspect. Speaker identification is deciding if a speaker belongs to group of known speaker population. Speaker verification is verifying the identity claim of the speaker.If the system is forced to choose one of the enrolled speakers then it is called a closed set identification system. If the system has the flexibility to make a choice none from the specified group then it is called an open set identification system. Based on the content used for speaker identification or verification, the tasks can further be classified as text dependent, where the speaker's identity is dependent on the text uttered, and text independent, where no constraints are placed on the text uttered (Hollien, 2002). Furthermore, the speech samples used for speaker identification or verification can be contemporary (recordings from same time period) and non-contemporary (recordings from different time period). The task of speaker recognition or speaker identification becomes very important in our digital world. Most of the law enforcement organizations use either automatic or manual speaker identification tools for investigation processes. In any case, before carrying out the identification analysis, they usually need to record a voice sample from the suspect either for one to one comparison or to fill in the database. So, the effect of recording media or voice sample recording for forensic speaker identification is very imperative (Barinov, 2010; Margi, Surbhi & Dahiya, 2015).

The present study focuses on the third method of speaker recognition semi automatic method which involves computer analysis. Here the voice analysis has been facilitated by the advent of computers installed with specific software (Software used for the present study was SSL Workbench version 2.1, Voice and Speech Systems, Bangalore).

Vowels, nasals and fricatives (in decreasing order) are generally suggested 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).

The present study is focused on vowels (/a:/, /i:/,/u:/) preceding nasal continuants (/m/ and /n/) which fall under the category of structured consonants of the Kannada script. The mean percentage and standard deviation of frequency of occurrence of vowels /a/, /i/ and /u/ is 14.6% (1.3), 6.7% (0.44) and 4.3% (0.47) respectively, and frequency of occurrence of phonemes /m/ and /n/ is 2.8% (0.26) and 7.6% (0.31) respectively in Mysuru dialect of conversional Kannada (Sreedevi Vikas, 2012).

The nasalization of the acoustic signal applies not only to the nasal consonants but also to certain surrounding sounds, particularly vowels. In general, vowels preceding or following nasal consonants tend to be nasalized to some degree. The present study is focused on bilabial (/m/) and dental (/n/) place of articulation and the vowels (/a:/, /i:/ & /u:/) preceding nasal continuants. Effects on influence of co-articulation can be of three types; (a) forward effect, (b) backward effect or (c) both. According to Carney and Moll (1971), there are anticipatory and/or carryover co-articulatory effects of vowel on the production acoustic realization of a neighboring consonant.The majority of the studies have found greater backward effect than forward effect (Ohde & Sharf, 1975). Thus, the nasal phonemes have been identified as being more reliable as a speaker cue because nasal cavity is both speaker specific and fixed so as its volume and shape cannot be changed (Arai, Amino & Sugawara, 2006). Larson and Hamlet (1987) investigated on the phonetic contextual details of nasal co-articulation using nasal voice amplitude ratio instrumentations. Nasalization was greater for vowels between two nasal consonants than for vowels between a nasal consonant and a fricative or stop. Results reported by authors were greater nasalization for pre-nasal vowels than post nasal vowels.

Mel Frequency Cepstrum Coefficients (MFCCs) modelled on human auditory system has been used as a standard acoustic feature set for speech related applications. 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 MFCCs. The main difference between computation of the MFCCs and the cepstral coefficients is the inclusion of Mel- Scale filter banks. MFCCs are now widely used for speaker recognition tasks and have been shown to yield excellent results.

In the past, researchers have used formant frequencies, fundamental frequency, F0 contour, liner prediction coefficients (Atal, 1974; Imperl, Kacic & Horvat, 1997), Cepstral Coefficients (Jakkhar, 2009; Medha, 2010; Sreevidya, 2010) and Mel frequency cepstral coefficients (Plumpe, Quatieri & Reynolds, 1999; Hasan, Jamil, & Rahman, 2004; Chandrika, 2010; Tiwari et al., 2010; Ramya, 2011; Singh & Rajan, 2011; Jyotsna, 2011; Rida, 2014; Suman, 2015) to identify speaker. The studies conducted by Jyotsna (2011), Rida (2014)

and Suman (2015) on speaker identification using MFCCs across different conditions like live recording and mobile recordings have proved the usefulness of vowels and nasals (>80%) towards speaker identification task. Hence, the Mel Frequency Cepstral Coefficients have been found to be more effective in speaker identification compared to other parameters and hence the present study is focusing on usefulness of MFCCs on vowels preceding nasal continuants in Kannada. There is no empirical data to establish the speaker identification scores for vowels preceding nasal continuants in Kannada. To prove that the suspect is the criminal, it needs to be verified beyond reasonable doubt that the voice of the criminal and the voice of the suspect are the same. So in order to overcome this problem, a semi automatic and reliable speaker identification system is desired. However, there are studies on benchmarking of nasals and nasal co-articulation in other languages.In this context, the present study was planned. The aim of the study was to establish Speaker identification scores using mel frequency cepstral coefficients on vowels preceding nasal continuants in Kannada. The objectives of the study were to establish speaker identification scores using MFCCs on vowels preceding nasal continuants in Kannada in live recording, mobile network recording and comparison between them.

Method

Participants

Twenty Kannada speaking neuro-typical adult males constituted as Group I were chosen to participate in the study. This was further sub grouped as Group II constituting ten speakers. The participants were in the age range of 21-32 years (Mean age = 25 years, SD= 3.4) and were graduates with Kannada as one of the subject and all the participants belonged to the Mysuru dialect of Kannada and were drawn from the work/residential place in and around Mysuru, Karnataka, India. Participants were included in the study only on fulfilling certain criteria. The inclusion criteria of subjects were - no history of speech, language, hearing and communication problems, normal oral structures, no other associated social or psychological or neurological problems and reasonably free from cold or other respiratory illness at the time of recording.Informed written consent was taken from the participants after explaining about the aim and objectives of the study. Hearing was screened using Ling's sound test. Kannada Diagnostic Picture Articulation Test (KDPAT) (Deepa, 2010) was administered by a Speech Language Pathologist to rule out any misarticulations in speech.

Materials

The material used was thirty commonly occurring, meaningful Kannada words (Target words) containing the nasal continuant /m/ (Bilabial) and /n/ (Dental) that are shown in Appendix, and embedded in seventeen sentences (text independent). These sentences consisted of words with three basic vowels (/a:/, /i:/, /u:/) preceding two places of nasal consonants (/m/ and /n/) and were embedded in 3-6 word meaningful sentences to maintain the naturalness of speech. The vowels preceding nasals continuants were added in the initial and medial positions. There were five occurrences for each vowel preceding nasal continuants (/a:m/, /i:m/, /u:m/, /a:n/, /i:n/ and /u:n/).

Recording Procedure

Speech samples of participants were recorded individually. Sentences were written on a card that was presented to the participants visually for familiarization. Participants were instructed to read the sentences four times in a natural way at normal rate of speech under two conditions- (a) mobile network recording and (b) live recording at a sampling frequency of 16 kHz. (a) Mobile network recording was done first and the network used for making the calls was a common network on a NOKIA 101 and the receiving network was also another common network on a Gionee S5.5 mobile phone. A participant participating in an experiment was given a NOKIA 101 handset. A call was made from the participant's handset to the experimenters' handset with recording option held by the experimenter. Speech signal was recorded as the participant uttered the test sentences. All the mobile network recordings were done at different places according to the participant's convenience with some amount of ambient noise (40db-60db). The noise level was mild to moderate as the mobile network recording was done in a natural setting. The recordings at the receiving end were saved by the experimenter in a microchip or memory SD card of that mobile phone. Later, the recorded sentences were uploaded to a computer memory for further analysis. (b) The live recordings was carried out after two weeks using Computerized Speech Lab (CSL 4500 model; Kay PENTAX, New Jersey, USA) in Forensic Speech Laboratory at the Department of Speech-Language Sciences, All India Institute of Speech and Hearing, Mysuru, and the files were stored in .wav format. The distance between the mouth and the dynamic microphone was kept constant at approximately 10 cm. The mobile network recordings were converted into .wav files using adobe audition software so that analysis can be compared between the conditions. Of the four recordings, the first recording was not analyzed as the material is novel to the subject and the second and third recordings were subjected to analysis and used for comparison. If any of the second/third recordings were not lucid, then the fourth recording was used.

Down Sampling

SSL Workbench version 2.1 software employs sampling frequency of 8 kHz and hence all the live and mobile network recordings were opened in PRAAT software (Boersma & Weenink, 2009) and down sampled to 8 kHz. All the recorded speech samples were stored separately for each speaker onto the computer memory at mono channel, 16 bit format having sampling frequency of 8 kHz.

Segmentation

The down sampled speech material was segmented (approximately 300ms) manually using PRAAT software to obtain the vowels preceding nasal continuants in initial and medial positions of the target words.

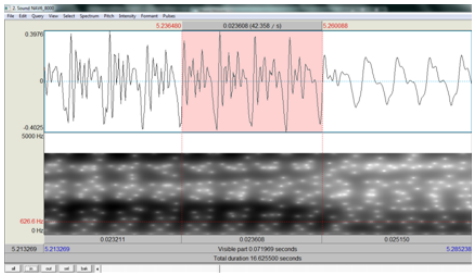

Figure 1: A segment of vowel preceding nasal continuant from a speech signal.

Analyses

Data analyses was carried out using Speech Science Lab (SSL) Workbench version 2.1 (Voice and Speech Systems, Bangalore, India) - a semi-automatic speaker recognition software. The segmented vowels preceding nasal continuants were analyzed at a sampling frequency of 8 kHz, to extract and compare its Mel Frequency Cepstral Coefficients (MFCCs).

Two repetitions and five occurrences for each vowel preceding nasal continuants were randomized by the software and considered as test set and training set in the ratio of 3:7. The file was specified initially using a notepad and .dbs file (extension of notepad file) was created automatically. Followed by this samples for analysis were segmented. As soon as all files were segmented the software opens another window to train the samples. After training, MFCCs were selected and the sample for identification was tested. Finally the software automatically generated the speaker identification threshold in terms of Euclidian Distance. This data was stored and the same procedure was repeated at least for 30 times by randomizing the training samples and the speaker identification thresholds were noted for the highest score and the lowest score. MFCCs derived from the vowels preceding nasal continuants were used to compute the Euclidian distance between the test and reference samples. For the present study, the feature vector chosen was MFCCs with 13 coefficients. Upon choosing the feature vector, the system computes a measure of distance (Euclidian distance) and displays the summarized distance matrix for the selected test and reference sample. The reference sample was taken along the row and the test sample wast taken along the column.From the distance matrix, the total percentage of correct speaker identification score was displayed. The percent correct identification (PCI) was calculated using the following formula:

$$PCI = \frac{No.\ of\ correct\ identification}{No.\ of\ total\ possible\ identification} * 100$$

In this study, closed set speaker identification tasks were performed, in which the examiner was aware that 'unknown speaker' was one among the 'known speaker'. Here, since the mobile network recordings for all speakers were carried out initially in the same session and live recordings for all speakers were done after two weeks in the same session, it can be stated that contemporary speech samples (live vs live & mobile network vs mobile network) and non-contemporary speech samples (mobile network vs live) were used for analyses. Analysis was done and correct percentage of speaker identification was calculated for the vowels preceding nasals (/a:m/, /i:m/, /u:m/, /a:n/, /i:n/ and /u:n/).

Results

The aim of the present study was to establish speaker identification scores in Kannada using MFCCs derived from vowels preceding nasals. The results of the study are explained under two sections; Section A and Section B with reference to the live recording, mobile recording and comparison between them.

Speaker Identification among Twenty Speakers

Condition I - Speaker identification scores for live recording: In this condition, contemporary speech samples were used where the live recording (test) was compared with live recording (reference). An average percentage of correct identification for 30 trials was obtained for each vowel preceding nasal continuant and results showed an average correct identification score of 92%, 80%, 80%, 93%, 78% and 80% for /a:m/, /i:m/, / u:m/, /a:n/, /i:n/ and /u:n/, respectively. Results showed an average correct identification score of 93%, 79%, 80%, 84% and 84% for /a:/, /i:/, /u:/, /m/ and /n/, respectively across vowels and nasals (Table 1).

Condition II - Speaker identification scores for mobile network recording: In this condition, contemporary speech samples were used where

Table 1: Average (AVG) percentage of correct identification scores across vowels and nasals for twenty speakers

Condition I				Condition II				Condition III			
	/m/	/n/	AVG		/m/	/n/	AVG		/m/	/n/	AVG
/a:/	92	93	93	/a:/	75	72	74	/a:/	38	39	39
/i:/	80	78	79	/i:/	58	49	54	/i:/	36	36	36
/u:/	80	80	80	/u:/	51	53	52	/u:/	34	39	37
AVG	84	84	-	AVG	61	58	-	AVG	36	38	-

*Condition I - Live vs Live, Condition II- Mobile network vs Mobile network, Condition III- Mobile network vs Live

both the reference and test speakers were chosen from the mobile network recordings. An average percentage of correct identification for 30 trials was obtained for each vowel preceding nasal continuant and results showed an average correct identification score of 75%, 58%, 51%, 72%, 49%, and 53% for/a:m/, /i:m/, / u:m/, /a:n/, /i:n/ and /u:n/, respectively. Results showed an average correct identification score of 74%, 54%, 52%, 61% and 58% for /a:/, /i:/, /u:/, /m/ and /n/, respectively across vowels and nasals (Table 1).

Condition III - Comparison of speaker identification scores between mobile network and live recording: In this condition, non-contemporary speech samples were used where the reference speakers were chosen from live recordings and test speakers were chosen from the mobile network recordings. An average percentage of correct identification for 30 trials was obtained for each vowel preceding nasal continuant and results showed an average correct identification score of 38%, 36%, 34%, 39%, 36% and 39% for /a:m/, /i:m/, /u:m/, /a:n/, /i:n/ and /u:n/, respectively. Results showed an average correct identification score of 39%, 36%, 37%, 36% and 38% for /a:/, /i:/, /u:/, /m/ and /n/, respectively across vowels and nasals (Table 1).

Speaker Identification among Ten Speakers

Condition I - Speaker Identification Scores for Live Recording: In this condition, contemporary speech samples were used where the live recording (test) was compared with live recording (reference). An average percentage of correct identification for 30 trials was obtained for each vowel preceding nasal continuant and results showed an average correct identification score of 92%, 85%, 86%, 95%, 81% and 89% for/a:m/, /i:m/, /u:m/, /a:n/, /i:n/ and /u:n/, respectively. Results showed an average correct identification score of 94%, 83%, 88%, 88% and 88% for /a:/, /i:/, /u:/, /m/ and /n/, respectively across vowels and nasals (Table 2).

Condition II - Speaker Identification Scores for Mobile Network Recording: In this condition, contemporary speech samples were used where both the reference and test speakers were chosen from the mobile network recordings. An average percentage of correct identification for 30 trials was obtained for each vowel preceding nasal continuant and results showed an average correct identification score of 80%, 68%, 60%, 85%, 58% and 69% for/a:m/, /i:m/, /u:m/, /a:n/, /i:n/ and /u:n/, respectively. Results showed an average correct identification score of 83%, 63%, 65%, 69% and 71% for /a:/, /i:/, /u:/, /m/ and /n/, respectively across vowels and nasals (Table 2).

Condition III - Comparison of Speaker Identification Scores Between Mobile Network and Live Recording: In this condition, non-contemporary speech samples were used where the reference speakers were chosen from live recordings and test speakers were chosen from the mobile network recordings. An average percentage of correct identification for 30 trials was obtained for each vowel preceding nasal continuant and results showed an average correct identification score of 47%, 51%, 50%, 50%, 53% and 46% for/a:m/, /i:m/, /u:m/, /a:n/, /i:n/ and /u:n/, respectively. Results showed an average correct identification score of 49%, 52%, 48%, 49% and 50% for /a:/, /i:/, /u:/, /m/ and /n/, respectively across vowels and nasals (Table 2).

Discussion

The results obtained from this study revealed several points of interest. The result obtained from condition I of section A were in consonance with those of the other previous studies using MFCCs with Hasan et al., (2004), Singh and Rajan (2011), Tiwari et al., (2010) and Chandrika (2010) where an identification accuracy of 80% - 100% were reported. Rajsekhar (2008) reported 75% identification in MFCCs using the word 'zero'. Chandrika (2010) reported overall performance of speaker verification system using MFCC as about 80% and overall performance of speaker recognition is about 90%-95% for vowel /i/. Tiwari et al., (2010) used

Table 2: Average (AVG) percentage of correct identification scores across vowels and nasals for ten speakers

Condition I				Condition II				Condition III			
	/m/	/n/	AVG		/m/	/n/	AVG		/m/	/n/	AVG
/a:/	92	95	94	/a:/	80	85	83	/a:/	47	50	49
/i:/	85	81	83	/i:/	68	58	63	/i:/	51	53	52
/u:/	86	89	88	/u:/	60	69	65	/u:/	50	46	48
AVG	88	88	-	AVG	69	71	-	AVG	49	50	-

*Condition I - Live vs Live, Condition II- Mobile network vs Mobile network, Condition III- Mobile network vs Live

MFCCs for designing a text dependent speaker identification system and reported progress in percent correct speaker identification with increase in number of filters in MFCCs with 85% for 32 filters. Jyotsna (2011) reported similar results on speaker identification using MFCCs in Malayalam speaking individuals and results of her study indicated 93.3% of correct identification for all vowels preceding nasals and vowel /a/ performed better compared to /i/ and /u/ using MFCCs as feature vector. Ramya (2011) studied the speaker identification under electronic vocal disguise using MFCCs where the results indicated the percent correct identification was above chance level for electronic vocal disguise for females and, interestingly vowel /u:/ had 96.66%, both /a:/ and /i:/ had 93.33%. Patel and Prasad (2013) used MFCCs and reported 13% error rate for the word 'hello'. Pickett (1980) reported nasalization effect stays for 100ms preceding and following the nasal continuant leading to maintenance of nasal characteristics for a longer duration than any other speech sounds.

The results obtained from condition II of section A showed that the percentage of speaker identification for mobile network recording was significantly lower compared to live recording. GSM (Global System for Mobile Communications) is the pan-European cellular mobile standard. Speech coding algorithms that are part of GSM compress speech signal before transmission, reducing 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). During transmission of speech 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 could have led to poorer scores in the mobile network condition in comparison with live recording.

The results obtained from condition III of section A showed that the percentage of speaker identification for mobile network recording versus live recording was highly lowered compared to condition I (live vs live) and II (mobile network vs mobile network). Here, the speech samples were non contemporary. Mobile network recordings were done initially and the live recordings were done after two weeks. The test speakers were chosen from mobile network recordings and the reference speakers were chosen from live recordings. Scores were poorer because speaker's emotional state during mobile network recording and live recording plays an important role and can affect speaker identification scores. Speaker's emotional state cannot be same during mobile network recording and live recording after two weeks whereas this is the condition in most of the forensic cases. The crime sample will be obtained from mobile whereas the suspect's (reference) sample will be extracted after a week or so in a police station or a recording room and the criminal's emotional state will not be the same under both the circumstances. Also, the environment in which both the recordings were done also influence the findings. Mobile network recording was done in a natural field condition and the live recording was done in a laboratory (noise free) condition. Ghiurcau, Rusu and Astola (2011) used MFCCs and support vector machines (SVM) in text independent speaker identification and reported that when emotions alter the human voice, the performances of the speaker recognition system decrease significantly. Devi, Srinivas and Nandyala (2014) reported that when the emotional state of speaker differs in the testing phase the recognition rate decreased drastically and the outcome showed that the accuracy rate of speaker recognition has been significantly increased when compared to the recognition rate where emotional state of the speaker was not considered.

It is also observed that the percent correct identification scores increase as the number of participants decreased. This was observed among all three vowels and among two nasal continuants. This result contradicts the findings of Hollien (2002) that decrease in error rate with increase in number of participants. But, it is in consonance with the results of Glenn and Kleiner (1968), where they described a text dependent method of automatic speaker identification based on spectra produced during nasal phonation showed better performance

Table 3: Speaker identification scores using MFCCs on vowels preceding nasal continuants in Kannada considering twenty speakers

Nasals	/m/			/n/		
vowels	/a:/	/i:/	/u:/	/a:/	/i:/	/u:/
I	92%	80%	80%	93%	78%	80%
II	75%	58%	51%	72%	49%	53%
III	38%	36%	34%	39%	36%	39%

*Condition I - Live vs Live, Condition II- Mobile network vs Mobile network, Condition III- Mobile network vs Live

Table 4: Speaker identification scores using MFCCs on vowels preceding nasal continuants in Kannada considering ten speakers

Nasals	/m/			/n/		
vowels	/a:/	/i:/	/u:/	/a:/	/i:/	/u:/
I	92%	85%	86%	95%	81%	89%
II	80%	68%	60%	85%	58%	69%
III	47%	51%	50%	50%	53%	46%

*Condition I - Live vs Live, Condition II- Mobile network vs Mobile network, Condition III- Mobile network vs Live

when the subjects reduced from 30 to 20 in number. Characteristically the presentation of a text- independent speaker verification system is poorer than a text-dependent system (Doddington, 1998; Boves & Den Oves, 1998) whereas in the present study, text independent procedure was established.

The results of the present study were in agreement with the findings of the power spectra of nasal consonants (Glenn & Kleiner, 1968) and co-articulated nasal spectra (Su, Li & Fu, 1974) provide strong cues for the machine matching of speakers. Results of the present study were consistent with the studies conducted by Larson and Hamlet (1987) in which they investigated on the phonetic contextual details of nasal co-articulation using nasal voice amplitude ratio instrumentations. Nasalization was greater for vowels between two nasal consonants than for vowels between a nasal consonant and a fricative or stop. Results revealed greater nasalization for pre-nasal vowels than post nasal vowels. The results of present study can be compared with that of Mili (2003) which indicated strong anticipatory co-articulation compared to carry over co-articulation. Also,most of the studies have found greater backward effect than forward effect (Ohde & Sharf, 1975). Also, this study can be compared with a similar study conducted by Suman (2015) in which vowels following nasals were considered. The present study focused on backward effect i.e., effect of nasals on preceding vowels thus providing good speaker identification scores. Table 3 depicts the speaker identification scores using MFCCs on vowels preceding nasal continuants in Kannada when twenty speakers were considered and Table 4 depicts the same when the number of speakers was reduced to ten.

Conclusions

Finally, to conclude, based on three conditions, vowel /a:/ preceding the two nasals /m/ and /n/ was reliable for speaker identification compared to other vowels. Hence, it would facilitate for the better identification. The poor scores between mobile network recording and live recording conditions could be attributed to the transmission characteristics of the network. The current study was a text-independent study conducted in a natural environment with background noise. These factors could have contributed to further reduction in accuracy of speaker identification.The current study indicated speaker identification scores using MFCCs on vowels preceding nasal continuants in Kannada and this outcome can be utilized in forensic speaker identification task. In general, it could be accomplished that vowels preceding nasal continuants also add good percent of correct identification among Kannada speakers on semi automatic machine technique of analysis in Forensic Sciences. To add on, when number of speakers were reduced, there is an increase in the performance of speaker identification by the system. Further research is warranted in the area of semi automatic and automatic methods by considering other forensic conditions like distortion, disguises, and so on.

References

Arai, T., Amino, K., & Sugawara, T. (2006). Idiosyncrasy of nasal sounds in human speaker identification and their acoustic properties, *Science and Technology, 27*(4).233-

235.

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

Barinov, A. (2010, November). *Voice samples recording and speech quality assessment for forensic and automatic speaker identification.* In Audio Engineering Society Convention 129. Audio Engineering Society.

Barinov, A., Koval, S., Ignatov, P. & Stolbov, M. (2010). Channel compensation for forensic speaker identification using inverse processing. *Proceedings of Audio Engineering Society 39th International Conference,* 53-58.

Boersma & Weenink, D. (2009). PRAAT S.1.14 software, restricted from http://www.goofull.com/au/program/14235/speedytunes.html.

Boves, L. and den Os, E. (1998). Speaker recognition in telecom applications. *Proceedings IEEE IVTTA-98, Torino,* 203-208.

Carney, P. J., & Moll, K. L. (1971). A cinefluorographic investigation of fricative consonant-vowel coarticulation. *Phonetica, 23*(4), 193-202.

Chandrika, (2010). *The influence of handsets and cellular networks on the performance of a speaker verification system.* Unpublished project of Post Graduate Diploma in Forensic Speech Science and Technology submitted to University of Mysore, Mysuru.

Deepa, A. (2010). *Re-standardization of Kannada Articulation Test.* A Dissertation submitted in part fulfillment of final year M.Sc Speech-Language Pathology, University of Mysore, Mysuru.

Devi, J. S., Srinivas, Y., & Nandyala, S. P. (2014). Automatic Speech Emotion and Speaker Recognition based on Hybrid GMM and FFBNN. *International Journal on Computational Sciences & Applications (IJCSA), 4*(1), 35-42.

Doddington, G. (1998). Speaker recognition evaluation methodology- an overview and perspective. *Proceedings for RLA2C Workshop on Speaker Recognition and its Commercial and Forensic Applications, Avignon, France,* 60-66.

Flege, J. E. (1988). Anticipatory and carry-over nasal coarticulation in the speech of children and adults. *Journal of Speech, Language, and Hearing Research, 31*(4), 525-536.

Ghiurcau, M. V., Rusu, C., & Astola, J. (2011).Speaker recognition in an emotional environment. *Proceeding Signal Processing and Applied Mathematics for Electronics and Communications,* 81-84.

Glenn, J. W., & Kleiner, N. (1968). Speaker identification based on nasal phonation. *The Journal of the Acoustical Society of America, 43*(2), 368-372.

Hasan, M. R., Jamil, M., & Rahman, M. G. R. M. S. (2004). Speaker identification using Mel frequency cepstral coefficients variations, 1, 4.

Hollien, H. F. (1990). *The acoustics of crime.* Springer Science & Business Media.

Hollien, H. F. (2002). *Forensic voice identification.* Academic Press.

Imperl, B., Kacic, Z. & Horvat, B. (1997). A study of harmonic features for the speaker recognition. *Speech communication, 22,* 385-402.

Jakkar, 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, Mysuru.

Jyotsna, (2011). *Speaker Identification using Cepstral Coefficients and Mel-Frequency Cepstral Coefficients in Malayalam Nasal Co-articulation.* Unpublished project of Post Graduate Diploma in Forensic Speech Science and Technology submitted to University of Mysore, Mysuru.

Larson, P. L., & Hamlet, S. L. (1987). Coarticulation effects on the nasalization of vowels using nasal/voice amplitude ratio instrumentation. *The Cleft palate journal, 24*(4), 286-290.

Vasan, M., Mathur, S., & Dahiya, M. S (2015). Effect of different recording devices on forensic speaker recognition system. In Proceedings of 23rd All India Forensic Science Conference 2015, Bhopal, Madhya Pradesh. 1.

Medha, S. (2010). *Benchmark for speaker identification by cepstral measurement using text-independent data.* Unpublished project of Post Graduate Diploma in Forensic Speech Science and Technology submitted to University of Mysore, Mysuru.

Mili, C. S. (2003). *Labial coarticulation in Malayalam.* Unpublished dissertation of Master of Science in Speech Language Pathology submitted to University of Mysore, Mysuru.

Ohde, R. N., & Sharf, D. J. (1975). Coarticulatory effects of voiced stops on the reduction of acoustic vowel targets. *The Journal of the Acoustical Society of America, 58*(4), 923-927.

Patel, K., & Prasad, R. K. (2013). Speech recognition and verification using MFCC & VQ. *International Journal of Emerging Science and Engineering, 1* (7), 33-7.

Pickett, J. M. (1980). *The sounds of speech communication.* Baltimore, MD: University Park.

Plumpe, M. D., Quatieri, T. F., & Reynolds, D. (1999). Modeling of the glottal flow derivative waveform with application to speaker identification. *Speech and Audio Processing, IEEE Transactions on, 7*(5), 569-586.

Rajsekar, A. (2008). Real time speaker recognition using MFCC and VQ.Thesis submitted in fulfillment of Master of Technology degree in Electronics and Communication Engineering to National Institute of Technology, Rourkela. Downloaded from http://ethesis.nitrkl.ac.in/4151/1/2.pdf

Ramya, B. M. (2011). *Benchmark 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.

Rida, Z. A. (2014). *Benchmark for speaker identification using nasal continuants in Hindi in direct and mobile network recording.* Unpublished dissertation of Master of Science in Speech Language Pathology submitted to University of Mysore, Mysuru.

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

Singh, S., & Rajan, E. G. (2011). Vector quantization approach for speaker recognition using MFCC and inverted MFCC. *International Journal of Computer Applications, 17*(1), 1-7.

Sreedevi, N. & Vikas, M. D. (2012). *Frequency of occurrence of Phonemes in Kannada.* Project funded by AIISH Research Fund (ARF).

Sreevidya, M.S (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, Mysuru.

Su, L. S., Li, K. P., & Fu, K. S. (1974). Identification of speakers by use of nasal coarticulation. *The Journal of the Acoustical Society of America, 56*(6), 1876-1883.

Suman, S. (2015). *Benchmark for speaker identification using MFCC on vowels following nasal continuants in Kannada.* Unpublished project of Post-Graduate diploma in Forensic Speech-Science and Technology submitted to University of Mysore, Mysuru.

Tiwari, R., Mehra, A., Kumawat, M., Ranjan, R., Pandey, B., Ranjan, S. et al. (2010). Expert system for speaker identification using lip features with PCA. In Intelligent Systems and Applications (ISA), 2nd International Workshop on (pp. 1-4). IEEE.

Appendix-A
Target words used in the study

IPA Transcription	Vowel Preceding Nasal Continuant
ba:da:mija	/a:m/
gra:ma	/a:m/
ra:manigɛ	/a:m/
sa:ma:njava:gi	/a:m/
t̪ a:mra	/a:m/
b^{h} i:ma	/i:m/
ʧi:ma:ri	/i:m/
d^{h}i:mənt̪a	/i:m/
si:ma	/i:m/
si:mɛjənɛ	/i:m/
b^{h}u:məndəla	/u:m/
b^{h}u:mijənnu	/u:m/
ʧ^{h}u:ənt̪ra	/u:m/
d^{h}u:mapa:na	/u:m/
hu:ma:lɛ	/u:m/
b^{h}a:nuva:ra	/a:n/
ha:nikara	/a:n/
ʤa:napəḍa	/a:n/
ʤa:nuva:ru	/a:n/
ka:rk^{h}a:nɛ	/a:n/
ḍi:ərigɛ	/i:n/
hi:na:jəva:gi	/i:n/
ki:na	/i:n/
nəvi:nənigɛ	/i:n/
t̪əlli:naɭa:gut̪t̪ a:ɭɛ	/i:n/
gu:nu	/u:n/
ku:na	/u:n/
məgu:na:	/u:n/
u:na	/u:n/
ʃu:nja	/u:n/