Paper7

by

1   **Abstract**

2   Speech is always accompanied by noise when the speaker is talking in the environment.

3   To improve the intelligibility of speech signal, noise should be reduced. The process of noise

4   reduction technique plays a major role in current scenario and is available in different forms of

5   software. From the existing software the aim of the present study was to examine the effect of

6   noise and noise reduction technique on speaker identification using Mel Frequency Cepstral Co-

7   Efficient (MFCC) on the long vowels in Kannada language. Ten Kannada speaking neuro-typical

8   adults in the age range of 20-35 years (5 males and 5 females) participated in the study.

9   Commonly occurring Kannada meaningful sentences with long vowels /a: /, /i: /, /u: / was used

10   for reading task. The same was recorded in two different conditions: Laboratory condition and

11   Traffic Field condition. These recorded samples were analyzed under two phases: Before noise

12   reduction and After noise reduction, using Sound Cleaner- Universal Noise Cancellation

13   Software. Speech Science Lab Work bench, a Semi-Automatic vocabulary dependent speaker

14   recognition software was used to extract Mel-Frequency Cepstral Coefficients for the truncated

15   (PRAAT software) vowels. Results of the study revealed that in *Lab condition, Traffic*

16   *condition, Traffic condition compared across traffic* condition, *Lab condition compared across*

17   *traffic* and in *Lab condition compared across traffic condition* , the vowel /i:/ is found to be

18   better followed by /a:/ and /u:/ in the average percentage of correct speaker identification of the

19   vowels. Overall results revealed vowel /i: / is better for speaker identification. Hence, the 'sound

20   cleaner' has a significant effect on percent speaker identification by reducing the influence of

21   noise without majorly affecting the acoustical parameter of certain vowel considered for the

22   present study.

Key Words: Sound cleaner, Semi-automatic, hypothetical, distortion, truncate

INTRODUCTION

The most natural and common way used to communicate information by humans is through speech. The speech signal conveys several types of information. From the speech production point of view, the speech signal conveys linguistic information (example- language and message) and speaker information (example- regional, emotional, and physiological characteristics). Most of us are aware of the fact that voices of different individuals do not sound alike. Like how a friend over a telephone is recognized, it is due to the important property of speech of being speaker dependent. The ability of recognizing a person exclusively from his voice (perceptually) is known as speaker recognition and known since long period (Atal, 1972). In recent days, crime rate is increasing drastically mainly through telecommunication especially due to increased usage of mobile phones which is misused for many crime related activities such as bomb threat, ransom demand, sexual abuse and hoax emergency call, in these conditions voice is the only evidence available for analysis. Hence there is a necessary need in the measurement of the voice for the establishment of legal proof by police and magistrates.

Among the biometric features and unique feature as speech or handwriting, 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. In recent advances to improve the performance and flexibility of speaker recognition, new tools have been produced in speech technologies. The method called speaker identification aims 'to identify an unknown voice as one or none of a set of known speakers on comparison' (Nolan, 1983; Naik,

2

1994). Speaker verification is the other common task in speaker recognition. Here '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). Hence, 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. Thus, according to Fururi (1994), Nolan (1997), Rabiner and Juang (1993) and Rose (2002), speaker recognition can be studied as a) speaker identification and b) speaker verification.

Various factors affect speaker recognition, background noise is one among them. As we know speech is always accompanied by noise when the speaker is talking in the environment. Most of the time speech is not heard clearly to the listener due to the surrounded noise. Background noise also plays a major role in forensic speaker identification. Most of the software will have difficulty in identifying speech signal when it is accompanied by background noise. To overcome this problem, the noise has to be filtered so that the required speech signals will be heard clearly. Various approaches have been implemented to improve the noise robustness of speaker recognition. The following are the techniques which can be listed in general:

Kalman filtering (Fingscheid, Suhadi, & Stan, 2003) or spectral subtraction (Garcia & Rodriguez, 1996) can be used to filter noise from speech, based on the prior knowledge of the noise characteristics. It is also possible to extract noise-robust features.

Relative spectral features (Hermansky & Morgan, 1994) from speech signal instead of removing the background noise. It is also possible to ignore the parts of speech corrupted by background noise using the following technology that is missing feature theory.

Missing feature theory (Bonastre, Besacier, & Fredouille, 2000). The above approaches are used in statistical speakers' models (e.g Gaussian Mixture Models (GMMs). A Gaussian

3

Mixture Model (GMM) is defined as a parametric probability density function which is represented as a weighted sum of Gaussian component densities. GMMs are commonly used in biometric systems, such as vocal-tract related spectral features in speaker recognition systems though it has the capability of symbolizing a large class of sample distributions.

According to a study by Yeldener and Rieser (2000), normally low bit rate of voice coding system will not have its own mechanism to reduce background noises from the target voice signal. This is due to constraint in the scope of voice coding systems and the complexity of the target signal which is a voice signal. Hence, it is very essential to include a method specific for removal of background noises. This process will happen by passing the voice signal to a pretreatment process. Here the background noises which disturb the identification processing of the signals will be removed. Since the presence of random noises will degrade the voice signals. In a study by Rozeha and Adib (2008), the removal of any unwanted signal is done by passing it through Digital Filter Design block of MATLAB (SIMULINK) software which serves as digital infinite-impulse response (IIR) bank-pass filter.

Spectral subtraction method is the other procedure which is used to remove background noise from voice recognition signal. According to Udrea and Coichina (2003) this method involves the basic principles of spectral subtraction method, from the signal the short term spectral magnitude of noise is subtracted. There will be an attempt made to estimate the average signal and average noise. Following this the same will be subtracted from each other. This results in improvement of signal to noise ratio.

In very general the speech signal can be improved by passing the signal through low-pass filter and use Fourier method for processing the signal. The processing of digital signal can be divided into FIR (finite-impulse response) and IIR (infinite-impulse response). FIR

4

is a non-recursive filter where it has an output that is a function of input samples and is not a function of previous output samples. A recursive filter has an output that is a function of an input samples and previous output samples. In general, FIR filter have better performances in analyzing the signal but execute slower because the process of Fast Fourier Transform takes a longer time. The IIR filter on the other hand executes faster but has low performances (Gold, Morgan & Ellis, 2011). Commonly, filters are designed to be low-pass (passing frequencies below some cut-off point).

Darren (2001) in unpublished thesis on Design of Speaker Recognition has proposed a method in removing background signal. First, signal will be converted to frequency domain through the use of a shifted FFT. Then, using 3rd order Butterworth low-pass filter which is also an IIR filer, the higher frequency signal will be removed. The cutoff frequency is chose to remove as much of noise signal while still preserving the original shape of the signal.

From these different methods of noise reduction techniques it can be assumed that for example, the acoustical parameter of the voice signal that is if we consider the voice signal to be a phoneme will be definitely altered with reference to the frequency and its range in correspondence with the filters setting used.

However, as a global leader in speech technologies SpeechPro Inc. has been developing specialized tools for efficient noise reduction and text transcription of low quality recordings for over 20 years. Various studies on the perception of poor audio recordings and noisy speech signals carried out by SpeechPro have resulted in the formation of the unique sound filtering algorithms that are now presented in the software and hardware products like Sound Cleaner, ANF II and The Denoiser Box. In the present study Sound Cleaner Signal Enhancement Program

Model 5142 (Noise Cancellation Software) is used to reduce the background noise and also to see its effect after noise reduction method.

In the present study, speaker identification will be carried out using machine method where semi-automatic speaker identification process will be used. This has been selected from the classification of Hecker (1971) and Bricker and Pruzansky (1976) speaker identification as: (i). Speaker identification by listening, (ii). Speaker identification by visual method & (iii) Speaker identification by machine which is subdivided into (a) Semi-automatic speaker identification and (b) Automatic speaker identification.

The present study focuses on the Semi - automatic Speaker Identification (SAUSI), where the known and the unknown samples from the speaker are selected by the examiner and are processed by the computer program for exact parameters such as first and second formants (Stevens, 1971; Atal, 1972; Nolan, 1983; Hollien, 1990; Kuwabara & Sagisaka, 1995; Lakshmi & Savithri, 2009), higher formants (Wolf, 1972), fundamental frequency (Atkinson, 1976), fundamental frequency contours (Atal, 1972), Linear prediction coefficients (Markel & Davis, 1979; Soong, Rosenberg, Rabiner & Juang, 1985), Cepstral coefficients and Mel frequency Cepstral coefficients (Atal, 1974; Fakotakis, Anastasios & Kokkinakis, 1993; Reyond & Rose, 1995; Rabiner & Juang, 1993), Long term average spectrum (Kiukaanniemi, Siponen & Matilla, 1982) and interpretations are made by the examiner.

Among these short and long term acoustical parameters Mel frequency Cepstral coefficients (MFCCs) are extensively used in present era for speaker identification tasks and has

6

been shown to yield tremendous results. Mel frequency cepstrum is a cepstrum with its spectrum mapped onto the Mel- Scale before log and inverse fourier transform is taken. MFCCs are derived from the known variation of the human ear's critical bandwidths with frequency (Hansen & Proakis, 2000). The two main filters used in MFCCs are linearly spaced filters and logarithmically spaced filters. To incorporate the phonetically essential characteristics of speech, MFCC will be used in speech signal. A series of calculation will takes place which uses cepstrum with a nonlinear frequency axis following mel scale. To get mel cepstrum, the speech signal will be windowed first using analysis window and then Discrete Fourier Transform will be computed. The main rationale behind MFCC is to mimic the human ears behavior. As such, the scaling in Mel-frequency cepstrum mimics the human perception of distance in frequency and its coefficients are known as the MFCC. The present study will be focusing on usefulness of Mel frequency cepstral coefficients (MFCC) on speaker identification.

REVIEW

To list out few reviews, a study using Mel-Frequency Cesptral Coefficients for feature extraction and vector quantization in security system based on speaker identification was conducted by Hasan, Jamil, Rabbani and Rahman (2004). Total of 21 speakers participated in the study. During framing in linear frequency scale different types of windows were used such as triangular, rectangular and hamming window. Among them hamming window yielded a better results. Hamming window is the sum of rectangle and hanning window. It is amplitude weighting of the time signal which is used with gated continuous signals which gives a slow onset and cut-off in turn to decrease the making of side lobes in their frequency spectrum. This window has similar properties to the Hanning window with the supplementary feature which

suppresses the first side lobe which gives the best results for large signal. The study revealed that when codebook size is 1 speaker identification score was 57.14% as codebook size increased to 16, the speaker identification increased to 100%. Hence it was concluded that the combination of Mel-Frequency and Hamming windows gives the best results.

Among speech sounds vowels, nasals and fricatives (in decreasing order) provide better speaker recognition compared to plosives. This is because they are comparatively easy to identify in speech signals and their spectra contain features that reliably distinguish speakers. (Douglas O' Shaughnessy, 1987; Sigmund, 2003). Vowels have proven to be effective for characterizing individual speakers and been widely used for speaker recognition and in forensic analysis.

For example an Indian review, Jakhar (2009) studied the benchmark for text dependent speaker identification in Hindi language using cesptrum. Live and telephonic recordings were done. For five speakers, the results showed 83.33%, 81.67% and 78.33% for /a:/, /i:/ and /u:/ respectively. For ten speakers, the results were 81.67%, 68.33% and 68.33% for a:/, /i:/ and /u:/ respectively. Whereas for twenty speakers the results were 60%, 50% and 43.33% for a:/, /i:/ and /u:/ respectively for the conditions such as live v/s live, mobile v/s mobile and live v/s mobile respectively. The results indicate as the number of speakers increases the percentage of correct speaker identification decreases and also scores are better when conditions are similar. Among /a:/, /i:/ and /u:/, /a:/ yielded better results in live recording and vowel /i:/ in mobile recording condition.

Medha (2010) studied the benchmarks for speaker identification of three long vowels /a:/, /i:/ and /u:/ using cepstral coefficients on text-independent data in Hindi language. Total of 20 Hindi speakers participated in the study, among them 10 were males and 10 were females. For

females the percent correct speaker identification scores were 40%, 40% and 20% for /a:/, /i:/ and /u:/ respectively. Whereas for males it was 80%, 80% and 20% for /a:/, /i:/ and /u:/ respectively. Therefore the benchmarking for female speakers was below chance level whereas for male speakers it was 80% for the vowels /a:/ and /i:/. Hence the study concluded saying that in text-independent condition the extraction of cepstral coefficient quefrency and amplitude is useful in speaker identification for vowels /a:/ and /i:/ in males only.

Sreevidya (2010) conducted a study to check the benchmark in Kannada language by text independent speaker identification method using cepstrum in both direct and mobile recording conditions. The results of the study showed in direct speech and reading, vowel /u:/ had highest score (70 and 80%) and for vowel /i:/ had highest score as (70 and 67%). Also quotes that for both the direct verse mobile recordings, for all vowels and for groups of speakers the results were below chance level.

Chandrika (2010) compared the efficacy of speaker verification system using MFCCs. Speakers participated in the study were 10 in number. Material consists of long vowels (/a:/, /i:/, and /u:/) in medial position occurring in five target Kannada words embedded in sentences (text-dependent). Speech recording was carried out in two conditions: mobile network and digital recording. MFCCs values were extracted for all the long vowels and the results indicated an overall verification of 80%. The overall performance of speaker recognition was 90% to 95% for vowel /i:/ where the accuracy of performance of vowel /i:/ was marginally better than /a:/ and /u:/.

Tiwari (2010) used MFCCs to extract, characterize and recognize the information about speaker identity. During mel-frequency wrapping the subjective spectrum is stimulated using filter bank. The author used different number of filter settings (12, 22, 32 and 42) to check its

effectiveness. Out of these, the results showed 85% effectiveness using MFCCs with 32 filters in speaker recognition task.

Jyotsna (2011) studied speaker identification using cepstral coefficients and MFCCs in Malayalam nasal coarticulation. Results showed using cepstral coefficients, the benchmark for speaker identification was 80% and for using MFCCs it was 90% for nasal co-articulation in Malayalam.

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

Ridha (2014) studied the benchmark for speaker identification using nasal continuants in Hindi speakers. Nasals like /m/, /n/ and /ŋ/ were chosen which were embedded in words in all positions. Results revealed percentage of correct identification obtained when live recording was compared with live recording were 100%, 90% and 100% for /m/, /n/ and /ŋ/. Meanwhile when mobile network recordings were compared with mobile network recordings the percent correct identification was 50%, 80% and 90% respectively. Among /m/, /n/ and /ŋ/, /ŋ/ had best percent correct speaker identification except under telephone equalized/ not equalized conditions. Under these conditions /m/ had best percent correct speaker identification. Similar study was conducted by Ayesha (2016) where the percent correct speaker identification score for /m/, /n/ and / ŋ/ was 70, 80 and 100, respectively when direct recordings were compared with direct recordings using MFCC. The percent correct speaker identification score for /m/, /n/ and /ŋ/ was 60, 70 and 60, respectively when network recordings were compared with network recordings using MFCC. The percent correct speaker identification scores decreased drastically when network recordings

were compared with network recordings. Overall, the results revealed that the velar nasal continuant /n/ had the best percent correct speaker identification in this study.

It is evident from these review that MFCCs is, perhaps, the best parameter for speaker identification and less susceptible to variation of the speaker's voice and surrounding environment (noise). Also, the vowels may be the most suitable, among speech sounds, for speaker identification. However, till date there are limited studies on vowels as strong phonemes for speaker identification using semi-automatic methods in presence and absence of noisy situations (conditions) and after the application of speech signal to any noise reduction techniques. Here an attempt is made to use the Sound Cleaner software to reduce the noise and study the effect of this on speaker identification. Since the scientific testimony impresses any court of law in whichever country that might be. And however for any result to be called scientific, it has to be measured, quantified and reproducible if and when the need arises. Therefore, a method to carry out these analyses becomes a must. In this context, the present study was conducted.

Aim:

The aim of the present study is to investigate the effect of noise and noise reduction technique on speaker identification using MFCC on the long vowels in Kannada language.

Objectives:

1) To evaluate the percent correct speaker identification using MFCC on the long vowels in Kannada for lab recording conditions and traffic recording (with noise) before the application of noise reduction technique.

2) To evaluate the percent correct Speaker Identification using MFCC on the long vowels in Kannada for traffic recording (without noise) after the application of noise reduction technique.

3) To compare between speaker identification using MFCC on long vowels in Kannada for lab recording condition verses traffic recording (with noise) before the application of noise reduction technique.

4) To compare the percent correct Speaker Identification using MFCC on the long vowels in Kannada for lab recording verses traffic recording (without noise) after the application of noise reduction technique.

METHOD

Participants

A total of 10 native Kannada speakers with 5 males and 5 females in the age range of 20-35 years were considered for the study. These participants had 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). The inclusion criteria for the participating speakers were no history of speech, language, hearing problem, no associated psychological or neurological problems, and no reasonable cold or respiratory conditions at the time of recording and normal oral structure. Hearing was screened using Ling's sound test. Kannada Diagnostic Picture Articulation Test (KDPAT) (Deepa & Savithri, 2010) was administered by a Speech Language Pathologist to rule out any misarticulation to be present in the speech.

Procedure

(a). Material

Commonly occurring hypothetical Kannada meaningful sentences with long vowels /a:/, /i:/, /u:/ formed the material for a reading task. The vowels were embedded in fifteen words within nineteen sentences. These target words formed the material for the present study and the same is listed in Table 1 of Appendix A.

(b). Recording software

Recording was done for three trails (Trail I, II and III). Vowels occurring consecutively five times in the sentences of Trial II and III only were selected for analysis out of three Trails. Where, Trial I acted as a model setter for the following two trails. The written material was provided to the participants and was made familiarized before recording begins in a laboratory condition and field condition for each participant individually. The same was recorded in two different conditions: *Condition I-* Laboratory recording and *Condition II-* Traffic Field recording. The time gap between these two conditions was two weeks. For lab recording, Computerized Speech Lab (CSL 4500 model; Kay PENTAX, New Jersey, USA) was used where the computer memory used was a desired 16 Bit (analog-digital) converter at a required sampling frequency of 8 kHz. The distance between the mouth and the dynamic microphone (Shure) was kept constant at approximately 10 cm. These recordings were stored in *.wav format*. For field condition, Olympus digital voice recorder (LS100) with attached dynamic microphone (Shure) was used for recording with the background noise of around 80 dB (A) (Kalaiselvi & Ramachandraiah, 2010). The field recording samples were transferred from Olympus digital voice recorder to a computer and the samples were stored in *.wav files* using an USB cable so that the analysis can be carried out in an effective manner in a computer.

1   (c) Analysis Software

2   1. Sound Cleaner: These individually recorded samples were analyzed under two phases:

3   Phase I, the audio files of condition I and condition II was not subjected to any noise reduction [20]

4   algorithm. In Phase II all the audio samples were subjected to noise reduction algorithm using [20]

5   Sound Cleaner Signal Enhancement Program Model 5142 (Noise Cancellation Software). 'Street

6   Noise' scheme, one of the modules in Sound Cleaner software was only used for the present

7   study. 'Street noise' scheme consists of sub-modules such as, 'Input', 'Waveform-input', 'Broad

8   band Filter', 'Dynamic Filter', 'Output/file' and 'Speaker'. 'Broad band Filter' was set at its

9   default settings and for 'Dynamic Filter', which has options such as 'strong signal' and 'weak

10   signal', where 'strong signal' was remained as 'strong' and weak signal was 'weaken' and the

11   threshold was kept at 4kHz. Data flows from the starting 'Input' process module (.wav file) to

12   the final one (.wav file) through an intermediate modules such as 'Broad band Filter' and

13   'Dynamic Filter' and will be processed and saved.

14   2. PRAAT: Thus, the samples of Phase I and Phase II stored in a separate folder will be

15   opened in PRAAT software (Boersma & Weenink, 2009) and down sampled to 8 kHz since the

16   analysis can be done up to 4 KHz (frequency distribution of an individual's speech frequency

17   ranges till 4 KHz). Of the three recordings, the first recording was not analyzed as the material is

18   novel to the participant and the second and third recordings were only used for analysis and

19   comparison as mentioned in the previous section. [33] From the down sampled speech material the

20   long vowels /a:/, /i:/ and /u:/ in medial position of the target words was truncated from the wide [11]

21   band bar type of spectrograms using PRAAT software program and was stored in different

22   folders for each participant for the convenience of further analysis. Three complete cycles

23   (approximately 300 ms) of the long vowels was segmented and pasted onto a particular file name

convent to the investigator. For Ex: Condition, speaker 1, first occurrence, vowel and first session was given the file name as "*LB_ SPM1_ (thupaaki)_(a)_2.wav*" and saved in a folder with the name *SPM1*.

3. Speech Science Lab (SSL) Work bench: This is Semi-Automatic vocabulary dependent speaker recognition software used in the present study to extract Mel-Frequency Cepstral Coefficients (MFCC) for the truncated (PRAAT software) vowels. 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. Followed by this, the truncated samples for analysis were segmented to the workbench software. As soon as all files are segmented the software train the samples randomly. The trail/repetitions and utterances of each recording will be randomized on 5:5 distribution by the software and will be considered as test set and training set on equal distribution. Thus, the SSL Pro.V4 software was used to test the performance of distance based, semiautomatic speaker recognition system, which is vocabulary dependent. After training, 13 MFCC was selected and the samples for identification were tested. Finally the software automatically generated the speaker identification threshold in terms of Euclidian Distance and thus, the correct percentage of speaker identification was calculated. This data was stored and the same procedure was repeated at least 15 times by randomizing the training and testing samples and the speaker identification thresholds was noted for the highest score and the lowest score. All the speech samples were non-contemporary, as all the recordings of the same person were carried out in two different conditions. Closed set speaker identification tasks were performed, in which the examiner was aware that the 'unknown speaker' is one among the 'known' speakers.

RESULTS

The aim of the present study was to examine the effect of noise and noise reduction technique on speaker identification using Mel Frequency Cepstral Co-Efficient (MFCC) on the long vowels in Kannada language. The Euclidean distance of the samples for the test and reference samples of each speaker were calculated and was then tabulated as a distance matrix comparing all the speakers. Following this the correct percentage of speaker identification scores were obtained and the same was randomized for 15 times to obtain the highest correct percentage of speaker identification (HPI). In the present study the HPI ranged from 70% to 100%. Thus, the speaker identification results of the study are discussed under five sections. 1) Lab condition. 2) Traffic condition. 3) Traffic condition followed with noise reduction technique. 4) Lab recording verses traffic recording preceding noise reduction technique. 5) Lab recording verses traffic recording followed with noise reduction technique.
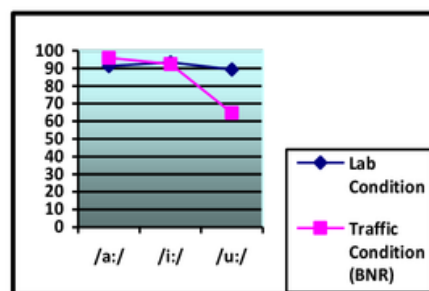
1) Lab condition

The task under this section was to evaluate the percent correct Speaker Identification using MFCC on the long vowels in Kannada language for lab recording (test sample) verses lab recording (reference sample). Here the results revealed that the highest percent correct identification (HPI) was 100% for /a:/, /i:/ and /u:/ vowels. The frequency of the occurrence of HPI for the three vowels was 6, 11 and 6 respectively. On an average of 15 time randomization score of the percent correct speaker identification of three vowels /a:/, /i:/ and /u:/ was 91.33% (SD: 8.33), 93.33% (SD: 11.75) and 89.33% (SD: 13.34) respectively. This indicates /i:/ to be better followed by /a:/ and /u:/. The descriptive data of speaker identification scores obtained for all fifteen randomized trials for vowels is depicted in Table 1 of Appendix B.

16

1   2) Traffic condition

2   The task under this section was to evaluate the percent correct Speaker Identification

3   using MFCC on the long vowels in Kannada language for traffic recording (test sample) verses

4   traffic recording (reference sample). Where, these samples contain some amount of traffic noise

5   embedded in it during analysis. Here the results revealed that the highest percent correct

6   identification (HPI) was 100% for /a:/ and /i:/ and 90% for /u:/ vowel. The frequency of the

7   occurrence of HPI for the three vowels was 9, 7 and 2 respectively. On an average of 15 time

8   randomization score of the percent correct speaker identification of three vowels /a:/, /i:/ and /u:/

9   was 96% (SD: 4.07), 92.33% (SD: 7.98) and 64.66% (SD: 20.30) respectively. This indicates /i:/

10   to be better followed by /a:/ and /u:/ which is similar to the previous lab condition. The

11   descriptive data of speaker identification scores obtained for all fifteen randomized trials for

12   vowels is depicted in Table 2 of Appendix B.

13   Comparison on observation among the average percent correct speaker identification

14   score for lab verses traffic recording condition, the differences was seen only for the vowel /u:/

15   when compared to /a:/ and /i:/. The same is represented graphically in figure (1).

16   Figure 1: Percent correct speaker identification score for vowels of lab verse traffic condition

3) Traffic recording followed with noise reduction technique.

The task under this section was to evaluate the percent correct Speaker Identification using MFCC on the long vowels in Kannada language for traffic recording conditions followed by noise reduction technique (test sample) verses traffic recording conditions followed by noise reduction technique (reference sample). Where, these samples had undergone noise reduction scheme in sound cleaner software. The results of this sample revealed that the high percent of correct speaker identification to be 100% for /a:/, /i:/ and /u:/ respectively. The frequency of the occurrence of HPI for the three vowels was 9, 8 and 1 respectively. On an average of 15 times randomization, the percent correct speaker identification of three vowels /a:/, /i:/ and /u:/ was 94% (SD: 8.28), 92.66% (SD: 10.99) and 66% (18.43) respectively. This indicates /i:/ to be better followed by /a:/ and /u:/. The descriptive data of speaker identification scores obtained for all fifteen randomized trials for vowels is depicted in Table 3 of Appendix B.

4) Lab recording verses traffic recording preceding noise reduction technique.

The task under this section was to evaluate the percent correct Speaker Identification using MFCC on the long vowels in Kannada language for lab recording (reference sample) verses traffic recording (test sample) preceding noise reduction technique. Here, the lab sample was absolutely speech and no noise, whereas the traffic samples contain some amount of traffic noise embedded in it during analysis. The results of this comparison revealed that the high percent of correct speaker identification for vowel /a:/, /i:/ and /u:/ to be 90%, 100% and 80% respectively. The frequency of the occurrence of HPI for the three vowels was 2, 1 and 1 respectively. On an average of 15 time randomization, the percent correct speaker identification of three vowels /a:/, /i:/ and /u:/ was 74.66% (SD: 12.45), 77.33% (SD: 15.79) and 58% (SD:

12.07) respectively. This indicates /i:/ to be better followed by /a:/ and /u:/. The descriptive data of speaker identification scores obtained for all fifteen randomized trials for vowels is depicted in Table 4 of Appendix B.

5) Lab recording condition verses traffic recording condition followed with noise reduction technique.

The task under this section was to evaluate the percent correct Speaker Identification using MFCC on the long vowels in Kannada language for lab recording (reference sample) verses traffic recording (test sample) following the application of noise reduction technique. Here, the lab sample was absolutely speech and no noise, whereas the traffic samples containing some amount of traffic noise embedding in it was removed with the sound cleaner software during analysis. The results of this comparison revealed that the high percent of correct speaker identification for vowel /a:/, /i:/ and /u:/ to be 80%, 100% and 70% respectively. The frequency of the occurrence of HPI for the three vowels was 6, 1 and 7 respectively. On an average of 15 time randomization the percent correct speaker identification of three vowels /a:/, /i:/ and /u:/ was 68.66% (SD: 13.55), 85.33% (SD: 8.33) and 57.33% (SD: 14.37) respectively. This indicates /i:/ to be better followed by /a:/ and /u:/. The descriptive data of speaker identification scores obtained for all fifteen randomized trials for vowels is depicted in Table 5 of Appendix B.

Comparison on observation among the percent correct speaker identification score for section 4 and 5, there is increment in the percent correct speaker identification scores in vowel /i:/ and slight decrement in vowel /a:/ after the application of noise reduction technique. The same is represented graphically in figure (2).

19
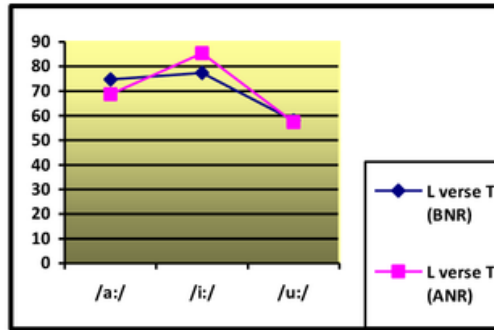
Figure 2: Percent correct speaker identification score for vowels of lab verse traffic condition

Summary of the results:

Table 1: Average and standard deviation (SD) of the percentage of speaker identification for condition I, II, III, IV and V

Average and standard deviation (SD) of the percentage of speaker identification						
Conditions/comparisons	**/a:/**		**/i:/**		**/u:/**	
	Mean	**SD**	**Mean**	**SD**	**Mean**	**SD**
1 Lab v/s Lab	91.33	8.33	93.33	11.75	89.33	13.34
2 Traffic (BNR) v/s Traffic (BNR)	96	4.07	92.33	7.98	64.66	20.30
3 Traffic (ANR) v/s Traffic (ANR)	94	8.28	92.66	10.99	66	18.43
4 Lab v/s Traffic (BNR)	74.66	12.45	77.33	15.79	58	12.07
5 Lab v/s Traffic (ANR)	68.66	13.55	85.33	8.33	57.33	14.37

Note* SD= Standard deviation, BNR= Before noise reduction, ANR= After noise reduction

DISCUSSION

The aim of the present study was to examine the effect of noise and noise reduction technique on speaker identification using Mel Frequency Cepstral Co-Efficient (MFCC) on the long vowels in Kannada language. Results of the study revealed that for (1). *lab condition*, HPI

20

is 100% for /a:/, /i:/ and /u:/ vowels. On an average the percent correct speaker identification of the vowels, the vowel /i:/ is found to be better followed by /a:/ and /u:/. For (2) *Traffic condition* (BNR- Phase I) HPI is 100% for /a:/ and /i:/ and 90% for vowel /u:/. On an average the percent correct speaker identification of the vowels, the vowel /i:/ is found to be better followed by /a:/ and /u:/. For the next (3). *Traffic condition compared across traffic* condition (ANR Phase II) the results of traffic condition revealed HPI to be 100% for /a:/, /i:/ and /u:/ vowels. On an average the percent correct speaker identification of the vowels, the vowel /i:/ is found to be better followed by /a:/ and /u:/. Following was (4). *Lab condition compared across traffic* (BNR), the results revealed HPI for /a:/, /i:/ and /u:/ vowels are 90%, 100% and 80% respectively. On an average the percent correct speaker identification of the vowels, the vowel /i:/ is found to be better followed by /a:/ and /u:/. The final was (5). *Lab condition compared across traffic condition* (ANR), the results revealed HPI for /a:/, /i:/ and /u:/ vowels are 80%, 100% and 70% respectively. On an average the percent correct speaker identification of the vowels, the vowel /i:/ is found to be better followed by /a:/ and /u:/.

From all the above conditions the vowel /i:/ is found to be better compared to vowel /a:/ and /u:/. The results of the present study are in support with the previous studies. To mention few a study by Jakhar (2009), who found that vowel /a:/ yielded better results in live recording and vowel /i:/ in mobile recording. Medha, (2010) found vowels /a:/ and /i:/ were useful in speaker identification in males. Chandrika, (2010) found better performance of vowel /i:/ compared to /a:/ and /u:/. In contrast to the present study Arjun, (2015) found vowel /a/ preceding nasals performed better compared to /i/ and /u/.

From the above discussion it was clear that speaker identification scores were poorer in condition (4) and (5) compared to condition (1), (2) and (3). This could be because of the sound cleaner contributing a significant affect in reducing the influence of noise without majorly affecting the acoustical parameter of certain vowel. Interestingly it was found that there is slight increment in the percent correct identification scores in vowel /i:/ after the application of noise reduction technique. From the present study, it is observed that in a semi-automatic method of speaker identification the vowel /i: / is considered to be the strongest phoneme which is not majorly affected by the influence of noise and the noise reduction technique. This was with reference to the sound cleaner software. In general the reasons for this difference in results could be as follows:

Reason (1) Different recording situations- During a real speech a person can recognize the surrounding sounds and concentrate on the speech of another person thus filtering the desired information out of various audio environments. Therefore the ability of a human to recognize and filter sounds significantly increases the intelligibility and comprehension of the speech even if a communication takes place in a noisy environment, situation or condition. This is not in the case of lab condition, where the individuals concentrate on their own speech with no task of filtering other audio environment since there will be complete silence in the lab.

However, in traffic condition it is a different situation. The recording equipment does focus on certain audio streams (specialized microphone) and impartially record everything that happens in the audio spectrum. As a product we receive a "flat picture" of all recorded sounds which often makes the speech partially unintelligible, quiet and buried in the noises.

Reason (2). The signal in the lab condition does not contain noise and is not subjected to undergone the removal of background noise from voice recognition signal for example using spectral subtraction method. Here, in this method the short term spectral magnitude of noise will be subtracted from the signal. That is the average noise and average signal is estimated and subtracted from each other (Udrea & Coichina, 2003).

Reason (3). The quality and accuracy of spectral picture is the most important factor for both experts and automatic systems (Kersta, 1962; Goldstein, 1975; Barinov, Koval, Ignatov, 2010). These authors describe only those parameters which affect instrumental identification analysis and which is one of the aim of the present study. Thus, each of these parameters, affecting spectrum, affects also the perceived quality of speech. The parameters listed are overloading, signal-to-noise ratio, reverberation, nonlinearity of frequency response and sampling frequency and bit rate. This might have contributed for poor percent correct identification score of traffic condition in the present study.

Therefore to conclude the study, the outcome after the application of noise reduction technique on speaker identification for traffic noise has not shown significant effect on the acoustical characteristics of the speech sounds. The speech sounds considered are being vowels only and the average percent correct speaker identification scores was better for vowel /iː/ followed by /aː/ and /uː/. Hence to conclude from the present study, the vowel /iː/ acts as a better cue for speaker identification. As an implication from the present study, there is also a future need to study the effect of other speech sounds in speaker identification under other noise reduction technologies.

23

CONCLUSION

The present study aimed to investigate the effect of noise and noise reduction technique of speaker identification using MFCCs on the long vowels in Kannada language. Results revealed vowel /i:/ is better for speaker identification because there is slight increment in the percent correct identification scores in vowel /i:/ after the application of noise reduction technique. Which indicates that the phonemic cue has not been altered much after noise reduction for vowel /i:/. Whereas in vowel /a:/ and /u:/ there was changes observed. Therefore, the sound cleaner has a significant affect in reducing the influence of noise without majorly affecting the acoustical parameter of certain vowel. Though, this study is a preliminary study which stepped to see the effect of noise reduction technique using Sound Cleaner. Further studies has to be conducted to check the effect of Sound Cleaner or other technology related to noise reduction techniques in reducing the background noise with reference to certain variables like increased participants, stimulus in different languages and considering the same in different environmental noise.

REFERENCE

Arjun. M. S. (2015). Benchmark for speaker identification using MFCC on vowels preceding the nasal continuants in Kannada. Unpublished *project of Post graduate Diploma in forensic Speech Sciences and Technology submitted to University of Mysore*. Mysore, Karnataka, India.

ASHA Monographs Number 16 (American Speech and Hearing Association, Washington D.C.).

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

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

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

Ayesha (2016). Benchmarks for speaker identification for nasal continuants in Urdu in direct and mobile network recording. *Unpublished dissertation of Master of Science in Speech Language Pathology submitted to University of Mysore, Mysore.*

Barinov, A. S., Koval, S. L., & Ignatov, P.V. (2010). Forensic Speaker Identification based on the Formants Matching Approach. *Forensic science international Journal.*

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

Bonastre, J.F., Besacier, L., & Fredouille, C. (2000). "Localization and selection of speaker-specific information with statistical modeling." *Speech Communication, 31*, 89–106.

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

Chandrika, S. (2010). The influence of handsets and cellular networks on the performance of a speaker verification system. *Project of Post graduate Diploma in Forensic Speech Sciences and Technology submitted to University of Mysore* . Mysore, Karnataka, India.

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

Douglas O'Shaughnessy, D. (1987). Speaker Recognition. *Speech Communication: Human and Machine*. Universities Press.

Fakotakis, N., Anastasios, T., & Kokkinakis, G. (1993). A text-independent speaker recognition system based on vowel spotting. *Speech Communication, 12*(1), 57-68.

Fingscheidt, T ., Suhadi, C. B., & Stan, S. (2003). "An evaluation of vts and imm for speaker verification in noise," *EUROSPEECH*, 1669–1672.

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

Garcia, O. J., & Rodriguez, J . G. (1996). "Overview of speech enhancement techniques for automatic speaker recognition," *ICSLP* 96, 2,929–932.

Goldstein, U.G. (1975). Speaker-identifying features based on formant tracks. *Journal of Acoustical Society of America, 59(1)*, 176-182.

Gold, B., Morgan, N., & Ellis, D. (2011). Speech and audio signal processing: processing and perception of speech and music. 2nd rev. ed. Wiley-Blackwell.

Hasan,R., Jamil, M., Rabbani, G., & Rahman, S. (2004). Speaker identification using mel frequency Cepstral coefficients. *3rd international conference on electrical & computer engineering, 565-568.*

Hecker, M. H. L. (1971), *Speaker Recognition: An Interpretive Survey of the Literature,*

Hermansky, H., & Morgan, N. (1994). "Rasta processing of speech," *Speech and Audio Processing, 2,* (4), 578–589.

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

Jakhar, S. (2009). Benchmark for speaker identification using Cepstrum. *Project of Post graduate Diploma in Forensic Speech Sciences and Technology submitted to University of Mysore .* Mysore., Karnataka.

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

Kalaiselvi, R., & Ramachandraiah, A. (2010). Environmental noise mapping study for heterogeneous traffic conditions. *Proceedings of 20th International Congress on Acoustics, ICA,* Sydney, Australia.

Kersta, L. G. (1962). Voiceprint Identification. *Nature,* 196, 1253-1257.

Kiukaanniemi, H., Siponen, P. & Mattila, P. (1982). Individual differences in the Long-Term Speech Spectra. *Folia Phoniatrica, 34,* 21-28.

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

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

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

Medha, S. (2010). Benchmark for speaker identification by Cepstrum measurement using text-independent data. *Project of Post graduate Diploma in Forensic Speech Sciences and Technology submitted to University of Mysore, Mysore.*

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

26

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

Nolan, F. (1997). Speaker recognition and forensic phonetics, in Hardcastle and Laver (eds) (1997): 744-767.

Rabiner, L. & Juang, B. H. (1993). Fundamentals of Speech Recognition, Signal Processing, editor A. Oppenheim. Englewood Cliffs: Prentice-Hall.

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.*

Reyond. A. D. & Rose. R. (1995). Robust text-independent speaker identification using Gaussian Mixture speaker models. *IEEE Transaction Speech Audio Process, 3,* 72-83.

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

Sreevidya, M. S. (2010). Speaker identification using Cepstrum in Kannada language. *Project of Post Graduate Diploma in Forensic Speech Sciences and Technology submitted to University of Mysore,* Mysore.

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.

Tiwari, V. (2010). MFCC and its applications in speaker recognition. *International Journal on Emerging Technologies, 1*(1), 19-22.

Wolf, J. J. (1972). Efficient acoustic parameter for speaker recognition. *The Journal of the Acoustical Society of America,* 2044–2056.

Paper7

9   cslu.cse.ogi.edu
     Internet Source   <%1

10   Bagul, S G, and R. K. Shastri. "Text
     independent speaker recognition system using
     GMM", 2013 International Conference on   <%1
     Human Computer Interactions (ICHCI), 2013.
     Publication

11   Weirich, Melanie(Harrington, J. and Pompino-
     Marschall, B.). "The influence of Nature and
     Nurture on speaker-specific parameters in   <%1
     twins speech", Philosophische Fakultät II, 2012.
     Publication

12   Sharma, Punit Kumar, B. R. Lakshmikantha,
     and K. Shanmukha Sundar. "Real time control
     of DC motor drive using speech recognition",   <%1
     India International Conference on Power
     Electronics 2010 (IICPE2010), 2011.
     Publication

13   Lecture Notes in Computer Science, 2007.
     Publication   <%1

14   Submitted to University of Leeds
     Student Paper   <%1

15   www.iconceptpress.com
     Internet Source   <%1

16   thescipub.com
     Internet Source   <%1

17	infocomp.dcc.ufla.br Internet Source	<%1
18	ijcsit.com Internet Source	<%1
19	www.sonomacity.org Internet Source	<%1
20	Alexandra Parbery-Clark. "What subcortical-cortical relationships tell us about processing speech in noise : Neural correlates of speech in noise", European Journal of Neuroscience, 02/2011 Publication	<%1
21	Submitted to University of Hertfordshire Student Paper	<%1
22	researchdirect.uws.edu.au Internet Source	<%1
23	www.slideshare.net Internet Source	<%1
24	research.ijcaonline.org Internet Source	<%1
25	host.quid5.net Internet Source	<%1
26	ir.canterbury.ac.nz Internet Source	<%1

27 www.ee.cuhk.edu.hk
    Internet Source   <%1

28 etheses.whiterose.ac.uk
    Internet Source   <%1

29 Mahesh, Sangeetha and Geetha, Y. V..
    "Phonetic Context in Disfluencies of Children   <%1
    with Stuttering", Language in India, 2013.
    Publication

30 foodrisklabs.bfr.bund.de
    Internet Source   <%1

31 www.uef.fi
    Internet Source   <%1

32 uef.fi
    Internet Source   <%1

33 uhra.herts.ac.uk
    Internet Source   <%1

34 www.jnltech.com
    Internet Source   <%1

35 The Acoustics of Crime, 1990.
    Publication   <%1

BIBLIOGRAPHY