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

Arjun

Register No: 14FST001

An Independent Project Submitted in Part Fulfilment of Third Component

Post Graduate Diploma in Forensic Speech Science and Technology,

University of Mysore, Mysuru

ALL INDIA INSTITUTE OF SPEECH AND HEARING

MANASAGANGOTHRI

MYSURU - 570006

July, 2015

Dedicated

To

AMMA

CERTIFICATE

This is to certify that the independent project entitled ***"Benchmark for Speaker Identification using Mel Frequency Cepstral Coefficients on Vowels Preceding Nasal Continuants in Kannada"*** is the bonafide work submitted in part fulfilment for the Post Graduate Diploma in Forensic Speech Science and Technology of the student (Registration No: 14FST001). This has been carried out under the guidance of a faculty of this institute, and has not been submitted earlier to any other University for the award of any other Diploma or Degree.

Dr. S.R. Savithri
Director
All India Institute of Speech and Hearing
Manasagangothri, Mysuru - 570006

Mysuru
July, 2015

CERTIFICATE

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

Guide
Mr. R. Rajasudhakar
Lecturer in Speech Sciences
Department of Speech-Language Sciences
All India Institute of Speech and Hearing
Manasagangothri, Mysore - 570006

Mysuru,
July, 2015

DECLARATION

This is to certify that the project entitled ***"Benchmark for Speaker Identification using Mel Frequency Cepstral Coefficients on Vowels Preceding Nasal Continuants in Kannada"*** is the result of my own study under the supervision and guidance of Mr. R. Rajasudhakar, Lecturer in Speech Sciences, Department of Speech Language Sciences, All India Institute of Speech and Hearing, Mysuru, and has not been submitted earlier in any other University for the award of any Diploma or Degree.

Mysuru
July, 2015

Register No: 14FST001

ACKNOWLEDGEMENTS

It is an immense pleasure to acknowledge the assistance and contribution of a large number of individual to this effort. I am deeply indebted to each and every one of them for their constructive criticism and support.

First and foremost, I would like to thank "THE DIVINE MOTHER" for her blessings in finishing this humble effort. I thank her for the large amount of love and kindness showered on me, without which I would never have been able to complete this report. Thank you AMMA for leading me to the right path always.

I thank my dearest parents and grandparents for their love, affection, encouragement and financial support throughout my education.

Dear Rajith, words are not enough to express my affection and gratitude; I am thankful to you for guiding me throughout my life. Your love and prayers have helped me to go through all the difficult situations.

I am very thankful to Dr. S.R. Savithri, Director, AIISH, Mysuru, for granting me permission to conduct this study.

I extend my sincere thanks to my guide, Mr. R. Rajasudhakar, Lecturer in Speech Sciences, Department of Speech Language Sciences, AIISH, for his constant support, guidance and advice to complete the project on time.

I am grateful to Dr. N. Sreedevi, HOD and Dr. Y. V. Geetha, Professor, Dept of Speech-Language Sciences, AIISH for their timely encouragement and providing all facilities to develop this report.

I also have a great pleasure acknowledging a sense of gratitude to Dr. N. Hema and Dr. T. Jayakumar, Lecturers in Speech Sciences, AIISH for the keen interest and enthusiasm shown towards this effort.

Thanks to Ms Sahana, Clinical Assistant, Dept of Speech-Language Sciences, AIISH for her co-operation and support.

I would like to thank the Library staff, AIISH for their on time help.

Special thanks to my friends – Divyalaxmi, Sumukhee, Nithya, Suman and Ishu for their timely help.

Finally, thanks to all the participants. It would not have been possible without all your voices.

Table of Contents

Sl No.	Title	Page No.
01	Table of Contents	vii
02	List of Tables	viii
03	List of Figures	xi
04	List of Illustrations	xii
05	Abstract	01
06	Introduction	02
07	Review of Literature	15
08	Method	47
09	Results	60
10	Discussion	77
11	Summary and Conclusions	85
12	References	xiii
13	Appendix A	xxvi
14	Appendix B	xxvii
15	Appendix C	xxxi
16	Appendix D	xxxii

List of Tables

Sl. No.	Title	Page No.
1	Details of participants	48
2	Target words used for the study	49
Speaker identification scores for live recording considering twenty speakers		
3	Average percentage of correct identification across vowels and nasals	61
4	Average and standard deviation of speaker identification percentage along with test samples	62
5	Euclidian Distance Matrix for /a:m/	xxxii
6	Euclidian Distance Matrix for /a:n/	xxxii
7	Euclidian Distance Matrix for /i:m/	xxxiii
8	Euclidian Distance Matrix for /i:n/	xxxiii
9	Euclidian Distance Matrix for /u:m/	xxxiv
10	Euclidian Distance Matrix for /u:n/	xxxiv
Speaker identification scores for mobile network recording considering twenty speakers		
11	Average percentage of correct identification across vowels and nasals	63
12	Average and standard deviation of speaker identification percentage along with test samples	64
13	Euclidian Distance Matrix for /a:m/	xxxv
14	Euclidian Distance Matrix for /a:n/	xxxv
15	Euclidian Distance Matrix for /i:m/	xxxvi
16	Euclidian Distance Matrix for /i:n/	xxxvi
17	Euclidian Distance Matrix for /u:m/	xxxvii
18	Euclidian Distance Matrix for /u:n/	xxxvii
Comparison of speaker identification scores between mobile network and live recording considering twenty speakers		
19	Average percentage of correct identification across vowels and nasals	65
20	Average and standard deviation of speaker identification percentage along with test samples	66

21	Euclidian Distance Matrix for /a:m/	xxxviii
22	Euclidian Distance Matrix for /a:n/	xxxviii
23	Euclidian Distance Matrix for /i:m/	xxxix
24	Euclidian Distance Matrix for /i:n/	xxxix
25	Euclidian Distance Matrix for /u:m/	xl
26	Euclidian Distance Matrix for /u:n/	xl
27	Average and Standard deviation of speaker identification percentage for condition I, II and III	67
Speaker identification scores for live recording considering ten speakers		
28	Average percentage of correct identification across vowels and nasals	69
29	Average and standard deviation of speaker identification percentage along with test samples	70
30	Euclidian Distance Matrix for /a:m/	xli
31	Euclidian Distance Matrix for /a:n/	xli
32	Euclidian Distance Matrix for /i:m/	xli
33	Euclidian Distance Matrix for /i:n/	xlii
34	Euclidian Distance Matrix for /u:m/	xlii
35	Euclidian Distance Matrix for /u:n/	xlii
Speaker identification scores for mobile network recording considering ten speakers		
36	Average percentage of correct identification across vowels and nasals	71
37	Average and standard deviation of speaker identification percentage along with test samples	72
38	Euclidian Distance Matrix for /a:m/	xliii
39	Euclidian Distance Matrix for /a:n/	xliii
40	Euclidian Distance Matrix for /i:m/	xliii
41	Euclidian Distance Matrix for /i:n/	xliv
42	Euclidian Distance Matrix for /u:m/	xliv
43	Euclidian Distance Matrix for /u:n/	xliv

Comparison of speaker identification scores between mobile network and live recording considering ten speakers		
44	Average percentage of correct identification across vowels and nasals	73
45	Average and standard deviation of speaker identification percentage along with test samples	74
46	Euclidian Distance Matrix for /a:m/	xlv
47	Euclidian Distance Matrix for /a:n/	xlv
48	Euclidian Distance Matrix for /i:m/	xlv
49	Euclidian Distance Matrix for /i:n/	xlvi
50	Euclidian Distance Matrix for /u:m/	xlvi
51	Euclidian Distance Matrix for /u:n/	xlvi
52	Average and Standard deviation of speaker identification percentage for condition I, II and III	75
53	Speaker identification percentage with different number of participants	76

List of Figures

Sl No.	Title	Page No.
1	Log spectrum of a vowel segment in Hz scale and in Mel scale	40
2(a)	Magnitude squared spectrum in mel scale multiplied by a triangular bandpass filter response	41
2(b)	Filters equally spaced in the mel frequency scale	41
3	Spectrogram depicting a segment of vowel preceding nasal continuant from a speech signal	51
4	Design specification window of SSL Workbench version 2.1	52
5	Input specification defined using NOTEPAD (*.txt file*) and *.dbs* file	53
6	SSL Workbench window depicting speaker selection.	53
7	SSL Workbench window depicting session selection	53
8	SSL Workbench window depicting occurrence selection	54
9	Segmentation window depicting two sessions of an occurrence for a speaker	54
10	Dialogue box depicting confirmation of a highlighted segment	55
11	Save segmentation' option of SSL Workbench window	55
12	.dbs file depicting particulars of segmentation for each speaker	56
13	Training frame of SSL Workbench	57
14	Testing frame of SSL Workbench depicting speaker identification score	58
15	Notepad file of Euclidian distance matrix	58

List of Illustrations

Sl No.	Title	Page No.
1	Schematic representation of speaker recognition	17
2	Schematic representation of speaker verification	18
3	Schematic representation of speaker identification	19
4	Classification of errors	21

ABSTRACT

Identification of speakers in forensic context is generally about comparing voices. In forensic speaker identification, the serious problem is to identify an unfamiliar speaker whose voice has been recorded at some stage in the committing of a crime. Vowels, nasals and fricatives (in decreasing order) are usually suggested for voice recognition because they are somewhat easy to identify in speech signals and their spectra contain features that reliably differentiate speakers based on semi-automatic methods. In this context, the aim of the present study was to obtain the percentage of speaker identification using vowels preceding nasal continuants in Kannada speaking individuals using semi-automatic method. The participants chosen for the study were twenty Kannada speaking adult males in the age range of 21-32 years constituted as Group I. This was further sub grouped (participants reduced) as Group II constituting ten speakers. The material was meaningful mono-, bi-, and/or multisyllabic 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 which were stored into the computer memory. The target words were truncated using the PRAAT software. Each vowel preceding nasal was subjected for Mel Frequency Cepstral Coefficients (MFCCs) using Speech Science lab Workbench for Semi-automatic speaker recognition (vocabulary dependent) software. The same was found across the three conditions when the participants reduced from twenty to ten in number. 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 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. Thus, the present study provided some proof to look at the efficiency of semi-automatic method using MFCC which helps in speaker identification. The obtained outcome would serve as potential measure in the forensic scenario for identification of speakers using vowels preceding nasal continuants in Kannada.

CHAPTER I

INTRODUCTION

"Evidence can vary depending on the circumstances, the weather, and how long it has been hanging around" - Pat Brown

The term 'Forensic' is derived from the Latin word *"forēnsis"* which means belonging to *courts of justice or courts of law or public discussion and debate*. 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.* Forensic science embraces all branches of physical and natural sciences. Over the years it has developed its own branches which are more or less the exclusive domain of forensic sciences. Anthropometry, footprints, fingerprints, ballistics, documents, serology, odontology, speaker identification/recognition were essentially developed to aid the criminal justice system (Nabar, 2001).

It is a general understanding that criminals, terrorists and extremists use telephones to threaten bombing of vital installations or aeroplanes, make obscene phone calls, claim ransom money for release of kidnapped victim's etc. Besides, now-a-days anti-corruption agencies have been making use of tape recorders to record conversation at the time of trapping corrupt public servants. In all such cases identity of the speaker's voice to enable apprehension or prosecution of the culprits become a very important task of the police or enforcement agencies. Courts in India have also made tape recorded voice or recorded voice as admissible as evidence provided the voice is properly recognised. In the lack of any scientific technique available, aural recognition of the recorded voice samples was the only means available for so long. However, with the developments in electronics, acoustics and computer knowledge, subjective and objective methods of voice recognition has become a reality. After substantial research and development, it has been possible to build up an instrument, which is skilled of analysing speaker's voice.

“The voice is the very emblem of the speaker, indelibly woven into the fabric of speech. In this sense, each of our utterances of spoken languages carries not only its own message, but through accent, tone of voice and habitual voice quality and at the same time an audible declaration of our membership of particular social/regional groups, of our individual physical and psychological identity, and of our momentary mood” (Laver, 1994).

Voice recognition is a well-known practice for most people like identifying a voice of a friend over the phone, recognizing the voice of a known personality on the radio or TV, hearing the voice of a colleague/classmate shouts out from the back, in recognizing voice in mobile calls. A type of problem in attempting to distinguish a speaker is that each individual’s tone of voice can fluctuate very much. We alter our voices depending on whom we are talking to, how formal the circumstances are, the feeling we desire to express and whether there is background noise. Voice of a speaker also changes if they are drunk, tired or have a sore throat or a cold. So, a voice is to a great extent is more complex to capture and difficult to analyze due to its variations than a fingerprint, which is permanent, fixed feature of a human being.

The most natural and common way used 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. Fingerprinting, photographic and anthropometric techniques are the most commonly used methods of identification. 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. The voice identification technique was adopted by the Michigan State in 1966 and introduced in the American court in the mid 1960's. Such method was used widely in different states including California, Florida and New York since then. However, different admission standards and interpretation methods were used among courts resulting in a lack of consistency (McDermott & Owen, 1996). Speaker identification has been used in a variety of criminal cases, including murder, rape, extortion, drug smuggling, wagering-gambling investigations, political corruption, money-laundering, tax evasion, burglary, bomb threats, terrorist activities and organized crime activities. Forensic acoustic analysis also involves tape filtering and enhancement, tape authentication, gunshot acoustics, reconstruction of conversation and analysis of any other questioned acoustic event. In every case, the features were compared with the previously stored features for the person whose identity is being claimed. If the comparison is favourable, based on decision criterion, then the claimed identity is verified. Among these methods, identity verification based on a person's voice has special advantages for practical deployment. Since personal identity verification is an essential requirement for controlling access to protected resources. Personal identity is usually claimed by presenting a unique personal possession such as a key, a badge, or a password. However, these can be lost or stolen. Further, a simple identity claim is not sufficient if the potential for loss is great and the penalty for false identification is severe. Hence, verification of that claimed identity is necessary. This can be attempted by examining individuals' biometric features, such as fingerprints, hand geometry, or retinal pattern, or by examining certain features derived from individuals' unique activity such as speech or hand writing. Since speech is the most natural means of communication and therefore user acceptance of the system would be very high.

An air freight cargo handler, Paul Prinzivalli, in Los Angeles, stood trial for telephoning bomb threats to his employer, Pan Am. He was suspected because he was known to be disgruntled employee, and because some Pan Am executives thought that the offender's voice sounded like his. Defence was able to demonstrate with the help of forensic-phonetic analysis that the offender's voice samples contained features typical of a New England accent, whereas Prinzivalli's accent was unmistakably from New York. To untrained West Coast ears, the differences between New York and Boston accents are not very salient; to ears trained in linguistic and phonetic analysis, the recordings contained shibboleths galore. Reasonable doubt was established and Prinzivalli was acquitted (Labov and Harris, 1994).

In a case of kidnapping and murder of an 11 – year-old German girl (Künzel, 1987), considerable agreement was found between the voice samples of a suspect and that of the kidnapper and on the basis of this and other evidence, the suspect was arrested. Subsequently, a more intensive comparison between offender and suspect voice samples yielded yet more similarities. The man confessed during his trial.

In another American case involving a telephoned bomb threat (Hollien, 1990), the defendant had been identified by his voice. However, it was clear to forensic phoneticians even from an auditory comparison that the voices of the defendant and the offender were very different. For example, the offender's voice had features typical of someone who spoke English as a second language. The case was dismissed.

In the late 1990s in Australia, in a case concerning illegal drug trafficking, police intercepted 15 incriminating telephone conversations containing 31 voice samples of three different speakers. Since the police, but not the analyst, knew the identity of some of the samples, not only could two of the speakers be identified, but accuracy of the identification could also be checked. Another Australian case involved the interception of telephone conversation between two brothers, one of whom was charged with drug-related matters. Defence claimed that their voices were similar that the incriminating recordings could not be attributed to the suspect. A forensic-phonetic analysis was able to show that the brothers' voices were distinguishable. Although their voices were indeed acoustically similar in many respects, they still differed in others, and in particular they both had different ways of saying their 'r' sound.

According to Hecker (1971), speaker recognition is defined as any decision making process that uses the speaker dependent features of the speech signal. In everyday life the identification of people by their voices is a common practice. We identify persons by listening to their voices, over a radio, phone line, among other devices. If the person is familiar to us, we can identify her/him by the style of speaking, the tone of the voice and so on. This imperative property of speech of being speaker-dependent is what enables us to recognize a friend over a telephone. If we do not know the individual, we can still infer some characteristics like gender, emotional state, age and language among others.

Speech set are usually continuous and a speaker never says exactly the same thing the same way twice and there is always variation between speech samples. However, differences in the speech of two persons appear due to the differences in anatomical differences in the vocal tract and learned speaking habits of the individuals. These differences can be used to discriminate between speakers. The speech sub-systems include the nervous system, the respiratory system, the phonatory system, the articulatory system and the resonatory system. The phonatory, articulatory and resonatory systems contribute to the differences in the voice and speech of individuals and this forms the basis of speaker identification. Also, variations between speech samples of same speaker are less compared to variations between speech samples of different speakers.

The three major methods of speaker recognition suggested by Hecker (1971) and Bricker and Pruzansky (1976) are – (i) Listening (Aural/perceptual method), here a person hears a voice and then attempted to match it to a particular individual, i.e., the one whose speech they heard. Here the type of listener involved, characteristics exhibited by the speaker and nature of environment in which utterances is produced are also taken into consideration. Using this method, the speaker gender identification judgment were 96% correct in one context, 91% correct in filtered speech context and 75% correct in whispered speech context according to Lass (1976). This result revealed that the laryngeal fundamental frequency appears to be more important acoustic cue in speaker gender identification.

(ii) The second method of speaker identification is by visual inspection of spectrograms i.e., comparing two sets of spectrograms, it is a three dimensional (time, amplitude and frequency) display of speech sounds. Spectrograms were used in attempts to identify unknown speakers by matching their speech/voice patterns with those of known speakers (or suspects). These were used in attempts to identify speakers by visually matching their speech/voice patterns. In the mid 1940s, the scientist of the Bell Telephone Laboratories in USA developed the first sound spectrograph (the Sonagraph), a visual record of speech including frequency, intensity and time (McDermott & Owen, 1996). In 1950s, Lawrence Kersta, an engineer from the Bell Telephone Laboratories, developed “voiceprint identification” (Hollien, 2002). In USA studies using the spectrograph were carried out in the 1950s and 1960s. According to Kersta (1960) voice pattern is unique to individuals and speaker can be identified with error rate of 1% using spectrogram. Stevens (1968) compared aural with visual examination of spectrograms using a set of eight talkers and found that error rate is 6% for listening and is 21% for visual. These scores depended upon the talker, phonetic context and duration of speech material. Speaker recognition by this method is coming into use in criminology, but the validity of this method is still in question according to Hecker (1971).

(iii) Machine methods, here the speaker dependent features from the speech signal are extracted and are analyzed by machines. This includes semi automatic speaker identification and automatic speaker identification. In the *semi automatic speaker identification (SAUSI)* the examiner selects unknown and known samples (similar phonemes, syllables, words and phrase) from speech samples, which have to be compared, here the computer process these samples, extracts parameters and analyze them according to a particular program. The interpretation is made by the examiner, especially to decide whether samples are good enough or affected by factor such as noise, co articulation, rate of speech etc to select comparable parts of speech samples for computerized acoustic analysis and to evaluate the results that the computer provides. This is typically encountered in for forensic where an interaction of computer and investigator takes place. Computer automatically extracts comparable data from forensic samples after they have been selected by the investigator. In the *Automatic speaker identification (AUSI)*, the computer does all the work and the participation of the examiner is minimal. As reported by Kunzel (1994), human involvement is to provide speech materials. For the purpose of automatic identification, special algorithms are used which differ based on the phonetic context and statistical techniques on speech acoustics is used to identify the speech. This method is used very often in forensic sciences but factors such as noise and distortion factors of voice and other samples need to be controlled. In such case a combination of subjective and objective methods would be used. High level of objectivity is claimed for this kind of a procedure. The Automatic Speaker recognition (ASR) system involves three components, the feature extraction, feature comparison and normalization.

Forensic speaker identification involves verifying a speaker from speech recorded under less than ideal conditions typical in forensics. A comparison may have to be made of a disguised voice sample recorded over a telephone channel with the voice sample recorded under laboratory conditions. In forensic applications, it is common to first perform a speaker identification process to create a list of "best matches" and then perform a series of verification processes to arrive at a conclusive match.

Speaker recognition is the process of automatically recognizing who is speaking on the basis of individual information included in speech signal. This process makes it possible to use the speaker's voice to verify his/her identity and thereby control access to services such as voice dialing, banking by telephone, telephone shopping, database access services, information services, voice mail, security control for confidential information regions, remote access to computers etc. Another important application of speaker recognition technology is for forensic purposes.

Speaker recognition can be of two types - (a) ***speaker identification*** and (b) ***speaker verification***. Speaker identification can be thought of as the task of determining who is talking from a set of known voices of speakers. It is the process of determining who has provided a given utterance based on the information contained in speech waves. The unknown voice comes from a fixed set of known speakers, thus the task is referred to as closed set identifications. In other words, speaker identification is the task of determining an unknown speaker's identity out of a database of N speakers. It is a 1:N match where the voice is compared against N templates. For example, a police officer comparing the sketch of an assailant to disguise against a database of previously documented criminals to find the closest match. This can be implemented covertly without the user's knowledge to identify talkers in a discussion, alert automated systems of speaker changes, check if a user is already enrolled in a system, etc. In speaker verification process, the speaker claims to be of certain identity. The voice is used to verify the claim. It is 1:1 match where one speaker's voice is matched to only the claimant's template. For example, verifying the passport at the border control. The agent compares the face to the picture in the document. Verification process is usually employed as a "gatekeeper" in order to provide access to a secure system (telephone banking). These systems operate with the user's knowledge and typically require their co-operation. Speaker verification, in other words, is the process of accepting or rejecting the speaker claiming to be the actual one. Since it is assumed that imposters (those who fake as valid users) are not known to the system, this is referred to as the open set task. Adding none of the above option to the closed set identification task would enable merging of two tasks, and is called open set identification. Error that can occur in speaker identification is the false identification of speaker and the errors in speaker verification can be classified into the following two categories: (a) ***False rejections***: a true speaker is rejected as an imposter, and (b***) False acceptances***: a false speaker is accepted as a true one. Speaker recognition methods can also be divided into text dependent or fixed text systems and text independent methods or text free systems. In case of text dependent method a speaker is required to utter a predetermined set of words or sentences (example- a password). Features of voice are extracted from the same utterance. In case of text independent methods, there is no predetermined set of words or sentences and the speakers may not even be aware that they are being tested.

All forensic identifications aim at reaching individuality. It is possible in some types of evidence. In others individuality has not been achieved. In voice evaluations so far individualization has not been achieved, where it may be said that the disputed utterances are of Mr. X (example) only and of no other person. In most of the cases it has not been possible to generate sufficient data to reach certainty. The evidence can only indicate the possible or the probable identity of the speaker. In some cases fairly high probability has been achieved.

The present study focuses on the third method of speaker recognition 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).

Language is the expression of thought by means of speech sounds. It is a system of communication by sound that is through the organs of speech and hearing using vocal symbols possessing arbitrary conventional meanings (Pei & Gaynor, 1954). Language is a systematic verbal symbolism which makes use of verbal elements such as sounds, words and phrases which are arranged in certain ways to make sentences.

Kannada is one of the major languages of India, spoken by over 40 million people particularly in the state of Karnataka located in southern part of India. Kannada is one of the 15 languages listed in the eighth schedule of the Indian Constitution (Sridhar, 2007). Kannada is one of the four major literacy languages of the Dravidian family, the other three being Telugu, Tamil and Malayalam. Kannada is one of the traditional languages among Dravidian family with a fine grammatical tradition. Kannada has a very complex range of regional, social and stylistic variations: the Mysuru/Bengaluru dialect, the coastal dialect (Mangalore), the northern dialect (Dharwar) and Kalaburgi dialect (Upadhyaya, 1976). The Kannada lexicon has been enriched by uninhibited borrowing from several sources, majorly from Sanskrit, Hindi-Urdu, and English. The Kannada script which is evolved from the 5th century Kadamba script is used to write the Kannada language. The history of Kannada can be conventionally divided in three periods; initial old Kannada (halegannada) flourished in 6th century Ganga dynasty from 450-1200 A.D., middle Kannada (nadugannada) from 1200-1700 A.D., and modern Kannada from 1700 to present.

Kannada language consists of 49 characters in its alphasyllabary and is phonemic. As different characters can be combined to form compound characters (ottaksharas), the number of written symbols however is far more than the 49 characters. The characters divided into three groups: swaras (vowels), vyanjanas (consonants) and yogavaahakas (part vowel, part consonant). Two types of cconsonants have been identified in Kannada script, the structured consonants and the unstructured consonants. 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 (Vikas & Sreedevi, 2012).

These vowels are speech sounds produced by voiced excitation of the open vocal tract. During vowel production, the vocal tract normally maintains a relatively stable shape and offers minimal obstruction to the airflow. The energy produced can be radiated through the mouth or nasal cavity without audible friction or stoppage. Vowels are described in terms of the relative position of the constriction of the tongue in the oral cavity (front, central and back), the relative height of the tongue (high, mid and low), the relative position of the lips (spread, rounded and unrounded), the position of the soft palate (closed and open), the phonemic length of the vowel (short and long), the tenseness of the articulator (lax and tense), and the relative pitch of the vowel (high, mid and low). Acoustically vowels are characterized by formant pattern, spectrum, duration and fundamental frequency.

The nasal consonants or nasal continuants are produced with closure of the oral cavity and radiation of the sound through nasal cavity while the oral obstruction is maintained. The oral cavity acts as a shunt or side-branch resonator that is even though the oral cavity is closed at some point; it nevertheless contributes to the resonant qualities of the nasal consonants. If it did not, then it would be impossible to distinguish the nasals in sustained, isolated productions. Although the nasal consonants are not always easily distinguished in such productions, they do not sound exactly alike. Acoustically nasal continuants are characterized by nasal murmur, F1 at around 300 Hz, damped formants, wide bandwidths and formant transitions. The articulatory feature of velopharyngeal opening accompanied by oral cavity obstruction is linked to an acoustic feature of nasal murmur. The murmur is the acoustic segment associated with an exclusively nasal radiation of sound energy. Formants are evident as the bands of energy on the spectrogram. There are also regions of greatly reduced energy labelled as antiformants. Unlike orally radiated vowels, which theoretically possess only formants in their transfer function, the nasal consonants possess both formants and antiformants. The antiformants can be thought of interfering with, or preventing, the transmission of energy in the frequency range. Antiformants, like formants, can be described with two numbers, the center frequency and the bandwidth.

The interaction of formants and antiformants in the spectrum of a nasal sound is not a simple matter of assigning formants to spectral peaks and antiformants to spectral valleys. Although such a result may occur, other spectral consequences may occur as well. For example, if a formants and antiformants have exactly the same center frequency and bandwidth the result of an interaction is a mutual cancellation. In fact, formants and antiformants often occur in pairs. When the members of a pair have the same frequencies and bandwidths, they cancel, but when the formants and antiformants diverge in these values, a particular spectral consequence will be seen.

The murmur is similar to the vowel in having a number of spectral peaks but only one of these, the low frequency nasal formant, has amplitude comparable to that of the vowel formants. The reduced amplitude of the other spectral peaks in the nasal murmur means that the nasal will have less overall energy than the vowel. Thus, the murmur portion of a nasal consonant has a dominant low frequency resonance, the nasal formant, accompanied by a number of much weaker resonances at higher frequencies.

The nasal formant is associated with a rather long tube extending from the larynx up through the opening of the nose. Because of this and because of the mucous filled nasal cavity the formants tend to be highly damped. That is, they have large bandwidths reflecting a rapid rate of absorption of sound energy. Therefore, the nasal cavity is not a sharply tuned resonator. Nasals can occur in several places of articulation namely labial, dental, alveolar, retroflex, palatal and velar. The formant transitions can be interpreted according to the place of articulation and like that for their homorganic stops. Therefore, similar formants transitions are observed for */b/-/m/*, */d/-/n/*, */dz/-/ñ/ and /g/-/ṅ/*. This similarity is not surprising given that the F2 transition relates to the place of articulation of the oral cavity. In many respects, the nasal consonants can be regarded as nasalized stops; that is, they share some fundamental properties with the stop consonants. The major differences between stops and nasals are explained by the effects of nasalization. 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.

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 has been shown to yield excellent results.

In the past, researchers have used formant frequencies, fundamental frequency, F_0 contour, liner prediction coefficients (Atal, 1974; Imperl, Kacic & Horvat, 1997), Cepstral Coefficients (Jakkhar, 2009; Medha, 2010; Sreevidya, 2010) and Mel frequency cepstral coefficients (Plumpe, Quateri & Reynolds, 1999; Hassan, Jamil, Rabbani & Rahman, 2004; Chandrika, 2010; Tiwari et al., 2010; Ramya, 2011; Singh & Rajan, 2011; Jyotsna, 2011; Rida, 2014) to identify speaker. However, 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.

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 (1997), there are anticipatory and/or carryover co-articulatory effects of vowel on the production acoustic realization of a neighbouring 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 et al, 2006).

Most of the studies (Reich & Duke, 1979; Reich, Moll & Curtis, 1976) elucidate the effective disguise for speaker identification. It is reported that nasal disguise and slow rate of speech are the least effective disguises. The correct speaker identification is degraded by noise, different transmission channels, emotional states and so on. If the disguises are more deliberate, then identification becomes more difficult (Ramya, 2013). It is necessary to study the effect of disguise on speaker identification. Especially if the speaker identification will focus on speech sounds with less association with oral cavity as the perpetrators focus on changing the characteristics of this cavity to disguise voice. The nasal cavity is a somewhat tougher option when it comes to manipulation (Lei, Lopez-Gonzalo, 2009). Therefore, nasal continuants would be the most appropriate speech sounds to investigate speaker identification under disguise. Therefore, the present study aims to study the effect of nasal continuant on preceding vowels using MFCCs in speaker identification tasks in Kannada and this supports the need for the present study.

Given the physiological production of nasal continuant one requires to explore on the possibility of using effect of vowels preceding nasal continuants for SPID. It is necessary to examine if vowels preceding nasal continuants provide high percent of identification when two or more speech samples are compared. However, till date there are limited studies on vowels preceding nasals as strong phonemes for speaker identification. Scientific testimony impresses any court of law irrespective of place that might be. However for any results to be called scientific, it has to be measured, quantified and reproducible if and when the need arises. There must be a technique to carry out these analyses and in this context the present study was planned.

The term benchmark can be defined as a set or point of references. However, there are studies on benchmarking of nasals and nasal co-articulation in other languages, the present study was planned to establish benchmark for speaker identification using Mel frequency cepstral coefficients in vowels preceding nasal continuants in Kannada. The aim of the study was to establish a ***Benchmark for speaker identification using Mel Frequency Cepstral Coefficients on Vowels Preceding Nasal Continuants in Kannada.*** The objectives of the study were three fold:

1. To establish benchmark for speaker identification using MFCCs on vowels preceding nasal continuants in Kannada in direct recording.
2. To establish benchmark for speaker identification using MFCCs on vowels preceding nasal continuants in Kannada in mobile network recording.
3. To compare speaker identification scores/performances using MFCCs on vowels preceding nasal continuants in Kannada between live and mobile network recording.

Since speaker identification or voice analysis has been, and is, a helpful procedure in the fight against crime. It has to be kept updated and sharpened all the way through research. This course of action will be needed more regularly in the upcoming times when other types of proof will be becoming scarcer.

CHAPTER II

REVIEW OF LITERATURE

The review of literature is organized under the following headings and sub-headings;

2.1. Speaker verification and identification

2.1.1. Speaker Verification

2.1.2. Speaker Identification

2.1.2.1. Open and Closed Set Identification

2.1.2.2. Text Dependent and Text Independent Speaker Recognition

2.1.2.3. Types of Decision

2.1.2.4. Types of Errors in Speaker Identification

2.2. Concerns in Speaker Recognition

2.2.1. System Distortion

2.2.2. Speaker Distortion

2.3. Factors Affecting Speaker Identification and Verification

2.4. Methods of Speaker Identification

2.4.1. Speaker Identification by Listening

2.4.2. Speaker Identification by Visual Examination of Spectrograms

2.4.3. Speaker Identification by Machine Method

“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).

In the legal process expert opinion is increasingly being sought as to whether two or more recordings of speech are from the same speaker or not. This is usually termed as Forensic Speaker Identification or Forensic Speaker Recognition. Forensic speaker identification can be very effective, contributing to both conviction and elimination of suspects. Forensic speaker identification can often be found classified as a kind of speaker recognition (Nolan, 1997). The aim of speaker identification is, not surprisingly, identification but ‘to identify an unknown voice as one or none of a set of known voice’ (Naik, 1994).

The importance of voice identification was first noted during the period of World War II related to the assassination of Adolf Hitler. This occurred on July 21, 1944 at Wolf’s Lair, his field headquarters in East Prussia. At that time, no one knew if he had been killed or just escaped out of Germany. There were still some speeches said to be from Hitler but their authentication was questioned. Fortunately, some of his past speeches were recorded and stored. Groups of scientists comprising phoneticians and engineers then decided to compare the old and new recordings. A series of analysis led to the conclusion that Adolf Hitler was still alive (Hollien, 2002). Since the anatomical structure of the vocal tract is unique for every person and hence the voice information available in the speech signal can be used to identify the speaker. Recognizing a person by his/her voice is known as speaker recognition. Since differences in the anatomical structure are an intrinsic property of the speaker, voice comes under the category of biometric identity. Using voice for identity has several advantages. One of the major advantages is remote person authentication.

Recognition of an individual from a forensic quality recording can prove to be an extremely challenging task. The method used for the analysis can be automatic, semi-automatic or human based. The material of the recording may also differ significantly ranging from a yelling on the telephone to a whisper recording, recording under stress, drugs, sickness or disguise, and recording in the presence or absence of noise. These unknown and known variables make the discrimination between speakers a complicated and daunting task.

The application of speaker recognition in India is not yet in wide spread use. Accuracy of speaker recognition system is important in the world of biometrics and must be accepted through forensic technology which should be robust in nature.

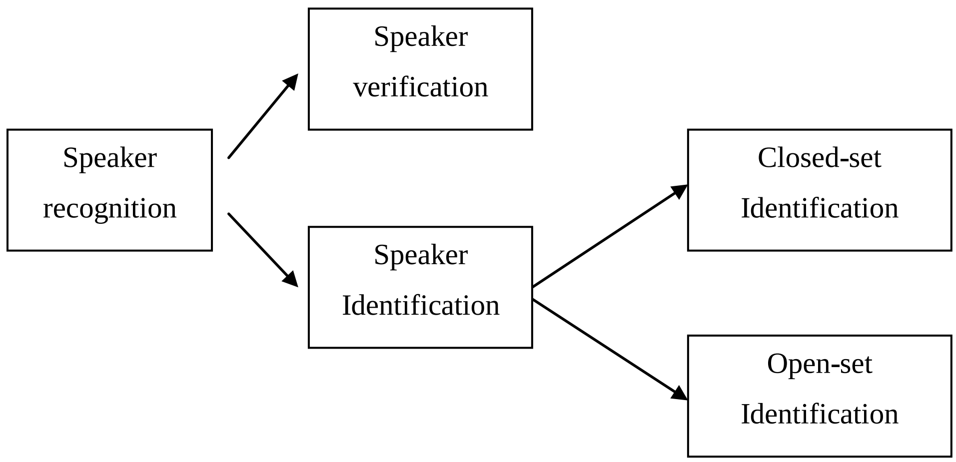


Illustration 1: Schematic representation of speaker recognition

2.1. Speaker verification and identification

As it is shown in the Illustration 1, there are two main classes of speaker recognition task, the identification and verification (Furui, 1996; & Nolan, 1997). The distinction between them rests firstly on the type of question that is asked and secondly on nature of the decision-making task involved to answer that question.

2.1.1. Speaker Verification

Speaker verification is one of the most common task in speaker recognition. This is where 'an identity claim from an individual is accepted or rejected by comparing a sample of his speech against a stored reference sample by the individual whose identity he is claiming' (Nolan, 1980). The schematic representation of speaker verification is shown in Illustration 2. The speaker D wants to access and verified. The system has samples of speaker D's voice in storage, which it retrieves and compares with that of the sample rendered by speaker D. If the two voice samples are judged similar enough, speaker D's claim is verified and he is given access.

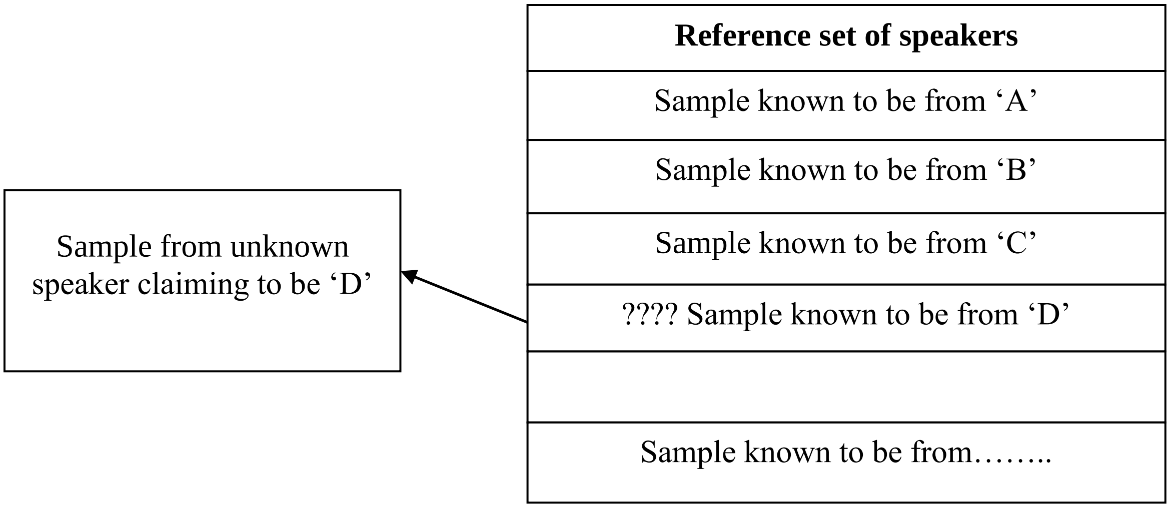


Illustration 2: Schematic representation of speaker verification.

2.1.2. Speaker Identification

The aim of speaker identification is, not surprisingly, identification. It is 'to identify an unknown voice as one or none of a set of known voice' (Naik, 1994). One has a speech sample from an unknown speaker, and a set of speech samples from different speakers the identity of whom is known. The task is to compare the sample of unknown speaker from the known set of samples, and determine whether it was produced by any of the known speaker (Nolan, 1980).

Illustration 3 shows schematic representation of simple speaker identification. The speaker identification experiment is represented with a reference set of 30 known speaker samples. In Illustration 3, the unknown sample on the left is compared with that of known speaker 1(A), then known speaker 2 (B), and so on. The question mark represents the question as are these two speech sample from the same speaker? If it is decided that the unknown sample is the same as one of the known speaker, say known speaker 4, then that identifies the speaker of the unknown sample as D.

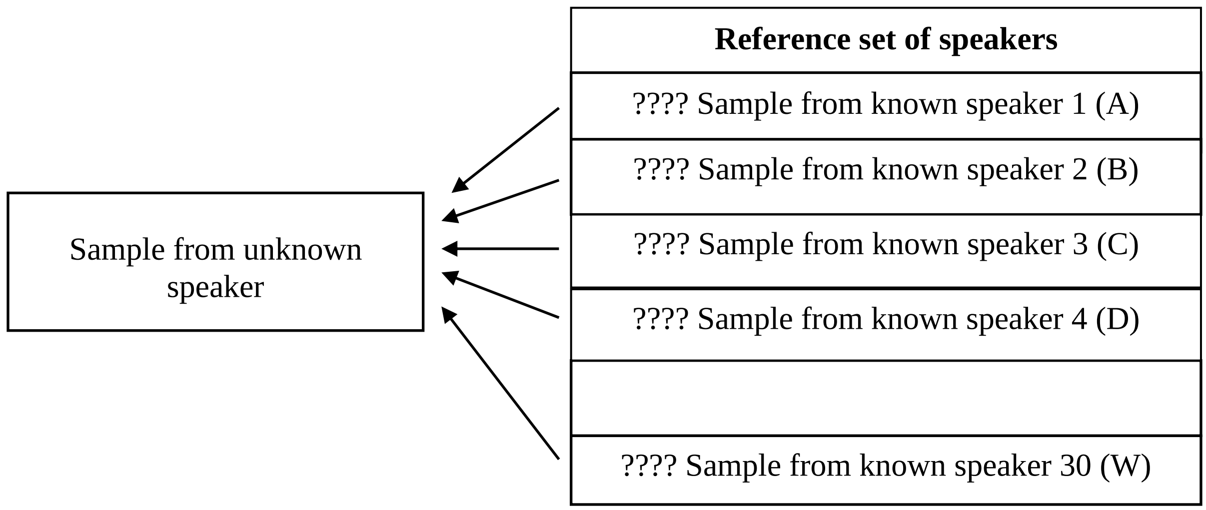


Illustration 3: Schematic representation of speaker identification.

2.1.2.1. Open and Closed Set Identification

The reference set of known speakers in speaker identification can be of two types: closed or open. A closed reference set means that the owner of the unknown voice is one of the known speakers. An open set means that it is not known whether the owner of the unknown voice is present in the reference set or not. Closed set identification is usually a much easier task than open set identification, since it is known that the unknown speaker is one of the reference set. The closed set identification task lies in (1) Estimating the distance between the unknown speaker and each of the known reference speakers, and (2) Picking the known speaker that is separated by the smallest distance from the unknown speaker. The pair of sample separated by the smallest distance is then assumed to be from the same speaker (Nolan, 1980). Because the nearest known speaker is automatically selected in a closed set identification, no threshold is needed. Both closed and open sets can occur in forensic case-work, although the latter, where we do not know if the putative offender is among the suspect or not, is usually far more common. Since the task usually becomes simpler with a closed set, the distinction between open and closed set tasks is an important one in forensic speaker identification.

The distinction between closed and open set does not have any meaning for verification. For the process to make any sense, it must be assumed that the individual whose identity is being claimed is among the reference set of speakers. In typical verification scenario, the reference set of speakers is unknown and the questioned voice is usually one of a set whose membership is not known. When speaker recognition is performed by untrained observer in real life conditions, it is termed as naïve speaker recognition. When an expert does speaker recognition using analytic techniques, speaker recognition is termed as technical speaker recognition and can be divided into text dependent and text-independent speaker recognition.

2.1.2.2. Text Dependent and text Independent Speaker Recognition

In text dependent speaker recognition, same text, i.e. same words or utterances, is used for both training the recognition device and testing it. In text independent, speaker recognition is attempted on utterances that are not controlled for lexical content (Furui, 1994). Text-dependent speaker recognition performs better than text independent speaker recognition (Nakasone and Beck, 2001). This is because some of the overall differences between speakers will be contributed by different speech sounds. The acoustic differences will more directly reflect differences between the speakers and can be used to recognize the speakers by use of text dependency.

2.1.2.3. Types of Decision

In identification, only two types of decision are possible. Either the unknown sample is correctly identified or it is not. Verification is more complicated, with four types of decision. The decision can be correct in two ways: the speaker is correctly identified as being who they say they are, or not being who they say they are. And it can be incorrect in two ways: the identity claim of the speaker can be incorrectly rejected (the speaker is who they say they are but rejected), or incorrectly accepted (the speaker is an important but is nevertheless accepted).

2.1.2.4. Types of Errors in Speaker Identification

In the open set speaker identification task- three types of errors are possible and in closed set only one type of error is possible. Illustration 4 shows the schematic representation of classification of errors.

(1) Error A: In the open set speaker identification, match did exist but the examiner selected the wrong one (False Identification)

(2) Error B: In the open set speaker identification, match did exist but the examiner failed to recognize it (False Elimination).

(3) Error C: In the open set speaker identification, match did not exist although the examiner selected one (False Identification).

(4) Error D: In the closed set speaker identification, since a match always existed, only one kind of error was possible, false identification or wrong identification. This error from closed set identification is labelled Error D.

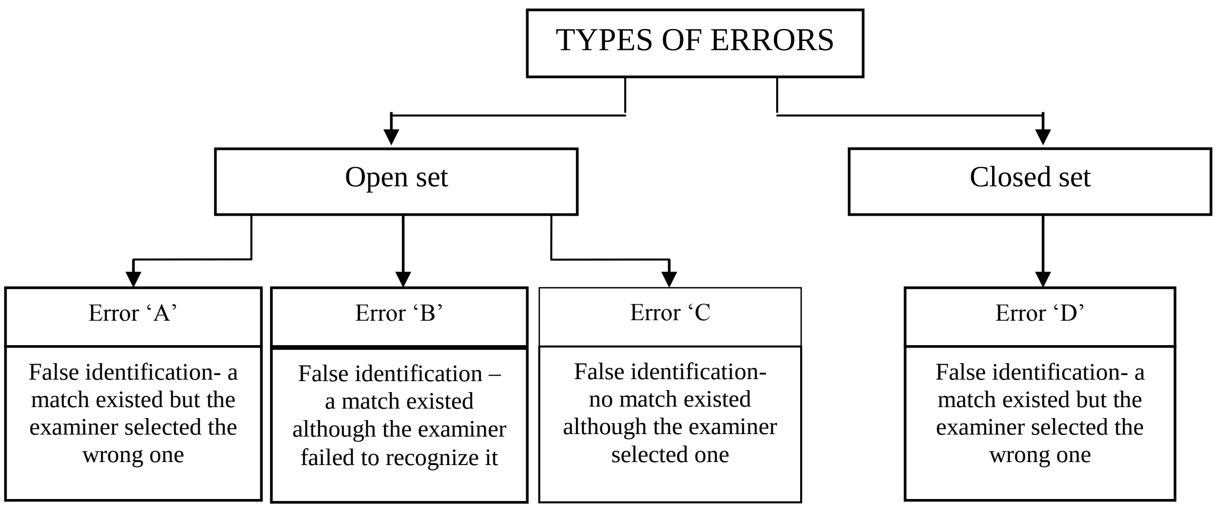


Illustration 4: Classification of Errors.

2.2. Concerns in Speaker Recognition

There are many problems in carrying out a speaker identification task and some of them are uniqueness and distortion. In uniqueness for example, the identification task might involve an open set of trials. Specifically, the unknown must be detected from a large to very large population of 'possibilities'. But this can be overcome to some extent that we can reduce the number of possibilities by taking into consideration, the gender, dialect, language, some common phrases used and style of speaking by the speaker as unique concern.

The other is distortion, for example it becomes very difficult to identify a speaker by his/her voice, especially when they are talking in an environment which distorts or masks their utterances (channel distortions) or when they are excited or stressed (speech distortions). The distortions are broadly classified into two types, (1) System Distortion and (2) Speaker Distortion.

2.2.1. System Distortion

This category includes several kinds of signal degradation. Some of the limitations in the system related issues are; (i) reduced frequency response through telephone line or mobile phone, (ii) poor quality tape recorders and (iii) reduced dynamic range and/or frequency response of microphones. In these cases, the important information about the talker is lost and these elements are not usually retrievable. Such limited signal pass band can reduce the number of helpful speaker specific acoustic factors. Furthermore, noise can cause a particularly debilitating type of system distortion as it tends to make the talker's voice and, therefore, can obscure elements needed for identification. Examples of possible noise are wind, motors, fans, automobile movement and clothing friction. The noise itself may be intermittent or steady state sawtooth or thermal and so on. Also, another type of noise is frequency or harmonic distortion that can make the task of identification more difficult. Examples include intermittent short circuits, variable frequency response, and harmonic distortion and so on.

2.2.2. Speaker Distortion

The speaker themselves can be a source of many types of distortions. Fear, anxiety or stress like emotion can occur when the perpetrator is speaking during the commission of crime. They often will degrade identification as the speech directly triggered by these emotions which subsequently changed one or more parameters of speech signal. The effects of ingested drugs or alcohol; and even a temporary health state such as a cold can affect the speech. Also, the suspect may sometimes attempt to disguise their voice. All these affect the speaker identification process in a troublesome way.

2.3. Factors Affecting Speaker Identification and Verification

In order to recognize a speaker, a set of feature characterizing the speaker's identity must be available to the listener. These features should contain idiosyncratic information that is specific for a speaker. Hollien (2002) presented a list of features used perceptually by listeners to identify a speaker. The list includes pitch, articulation, voice, quality, prosody, vocal intensity and speech characteristics. Also, the other factors are gender, regional dialect, foreign accent, age, voice distinctiveness, vocal disguise, emotions, retention interval, sample duration, voice quality and speaker familiarity affect the speaker identification and verification.

A range of factors have been correlated or found to be important in speaker recognition that are related to the original set of indices defined by Abercrombie (1967). The features presented include the speaker's gender, age and regional or foreign accent. In addition, other factors not related to the voice production have impact on the listener's ability to detect speaker's identity. These includes retention interval, sample duration and speaker familiarity.

The ability to identify a person on the basis of their voice has long been investigated. For many years law enforcement agencies have tried to use forensic speaker identification to incriminate or confirm innocence or guilt of a suspect.

2.4. Methods of Speaker Identification

A person's voice is a complex acoustic signal which encodes various kinds of information that reflects some of the anatomy and physiology of the speaker (Corsi, 1982). The problem of identifying individuals from their speech is a complex one exhibiting many measures, levels and parameters. With respect to the current state of the art, even the selection of a particular method for speaker identification is difficult and often confusing process. Hecker (1971) classifies the methods of speaker identification into three general categories like Speaker Identification by Listening (Subjective Method), Speaker Identification by Visual Examination of Spectrograms (Subjective Method) and Speaker Identification by Machine (Objective Method).

The above three methods have demonstrated some success in the laboratory condition but none have been particularly successful under field conditions. In the first method, the expert hears or perceives the voices and decides whether two voices belong to the same person or not.

In the second method, spectrograms of two speech samples are visually matched to identify a speaker. Of these approaches, the third method (semi automatic and automatic methods) appears to be the most promising for the future, primarily because (1) specific parameters within the speech signal can be selected and analyzed serially or simultaneously, (2) the selected vectors may be used in various combinations and (3) elimination of subjective analysis by humans. In all three methods used, the hypothesis is that "inter-speaker variability" is always greater or different than "intra-speaker variability" regardless of the features involved in these variabilities. Since the features derived from the variabilities are not well determined and quantified, at the present time the only way to prove scientifically that "inter-speaker variability" is always greater or different than "intra-speaker variability", is by inference. However, an inference thus derive might be affected by both from the speakers and methods of identification used.

2.4.1. Speaker Identification by Listening (Subjective Method)

The listening method is also known as aural-perceptual speaker identification (AP-SPID). The main focus here is the listener. However, the type of listener involved, characteristics exhibited by speaker and nature of the environment in which utterance is produced are also taken into considerations.

Some studies were reported earlier on speaker identification by listening method. Hecker (1971) reported that speaker recognition by listening appears to be the most accurate and reliable method at that time. Smrkovski (1976) reported 100% accuracy using trained and experienced examiners for aural analysis of 38 to 54 words of fixed context. Rothman (1977) performed aural study of 4-2 second speech segments of random context by 30 listeners and concluded that "aural method is clearly superior to the spectrographic or 'voiceprint' method. Reich and Duke (1979) reported 92% of correct identifications for undisguised condition of nine words spoken in fixed context by 40 speaker population.

Speaker authentication and identification were examined by Stevens et al., (1968), for two different methods of presentation of speech material: (1) speech samples presented aurally through headphones, and (2) speech samples presented visually as conventional intensity-frequency-time patterns, or spectrograms. They carried out two kinds of experiments: (1) a series of closed tests in which there was a library of samples from eight speakers, and test utterances were known to be produced by one of the speakers; and (2) a series of open tests in which the same library of eight speakers was used, but test utterances may or may not have been produced by one of the speakers. They reported that aural identification of talkers based on utterances of single words or phrases is more accurate than identification from the spectrograms and average error rate obtained by listening is 6-8% than visual 31-47% for the closed set identification. The authors reported that the scores were depended upon the talker, the subject, the phonetic content and duration of the speech material. For the open visual tests, appreciable numbers of false acceptance (incorrect authentications) were made. The results suggested procedures that might be used to minimize error scores in practical situations.

In a study by McGehee (1937), listeners attempted to select a single target voice from a set of five male voices after delays that ranged from day one to five months. The correct identification scores were declined from 83% after day one to 80.8% after one week, 68.5% after 2 weeks, 57% after one month, and to 13% after five months. Thompson (1985) used male voices in a six-voice line up task in which listeners rated each voices as to whether it was the voice they had heard one week previously. They could also respond that the voice heard previously was not in the lineup or that they were not sure whether it was in the lineup. However, the listener were not given the option of saying the voice heard previously was in the lineup more than once. Thus, from the viewpoint of the listeners, the experiment was an open-set task, but not an independent-judgement task. Such a task can be considered an open-set, multiple-choice task with a decision threshold by the listener. The results of McGehee (1937) were 62.1% correct identifications, 22.1% incorrect identifications, and 15.8% “not in lineup” or “not sure if in lineup” response.

There are studies designed on identification of speakers’ sex from isolated production of voiceless fricatives. One such study by Schwartz (1968), designed to investigate the ability of listeners to identify speakers’ gender from isolated productions of /s/, /ʃ/, /f/ and /ɵ/. Among the total of 18 participants (nine males and nine females) four fricatives in isolation were recorded. The stimuli were randomized and presented through auditory mode via loudspeaker to ten listeners for gender identification. The obtained results indicated that, only from the isolated productions of /s/ and /ʃ/ the listeners identified the sex of the speakers, but did not from the /f/ and /ɵ/ productions. The consequential spectrographic analysis of the /s/ and /ʃ/ stimuli revealed that the male spectra tended generally to be lower in frequency than the female. Ingeman (1968) supported the above results and reported that listeners are often able to identify the sex of a speaker from hearing voiceless fricatives in isolation and authors concluded that the gender was better identified on fricative /h/.

Schwartz and Ingemann (1968) employed isolated voiceless fricatives as auditory stimuli and found that listeners accurately identified speakers' gender from these stimuli, especially from /h/, /s/ and /ʃ/. The laryngeal fundamental was not available to the listeners because of the voiceless condition of the consonants. These findings indicated that accurate speaker gender identification is possible from vocal tract resonance information alone.

Schwartz and Rine (1968) investigated the listeners' ability to identify speaker's gender from two whispered vowels /i/ and /α/. They found 100% correct identification for /α/ and 95% correct identification for /i/, despite the absence of the laryngeal fundamental.

Coleman (1971) employed /i/, /u/ and a prose passage to study the speakers' gender identification abilities of his subjects. All stimuli were produced at the same vocal fundamental frequency (85 Hz) by means of an electro larynx. Coleman discovered that the listeners are capable of accurately recognizing the gender of the speaker, even when the fundamental frequency remained constant for all speakers. In a later experiment, Coleman (1973) presented recordings of 40 speakers' normal (voiced) productions of a prose passage to a group of listeners, and he found that "majority of listeners were basing their judgments of the degree of maleness or femaleness in the voice on the frequency of the laryngeal fundamental".

Many researchers have done studies related to identification of speakers' sex using fricative sounds. LaRiviere (1974) investigated speaker identification from turbulent portions of fricatives. In this experiment, 8 male speakers were asked to produce fricatives and 12 listeners were exposed to isolated turbulence portions of fricatives, as produced by those male speakers, and were asked to judge speaker identity. Results indicated that speakers were identified from such fricative stimuli, but the addition of laryngeal source resulted in higher identification levels and as place of articulation moves posterior the performance tends to improve.

Lass, Hughes, Bower and Bourne (1976) investigated the relative importance of the speaker's laryngeal fundamental frequency and vocal tract resonance characteristics in speaker sex identification tasks. Six sustained isolated vowels *(/i/, /ɛ/, /æ/, /α/, /o/, and /u/)* were recorded by

20 speakers, 10 males and 10 females, in a normal and whispered manner. A total of three master tapes (voiced, whispered and filtered) were constructed from these recordings. The filtered tape involved 255 Hz low-pass filtering of the voiced tape. The tapes were played to 15 listeners for speaker gender identification judgments and confidence ratings of their evaluations. Results of their judgments indicated that, of the 1800 identifications made for each tape (20 speakers * 6 vowels * 15 listeners), 96% were correct for the voiced tape, 91% were correct for the filtered tape, and 75% were correct for the whispered tape. Moreover, the listeners were most confident in their judgments on the voices tape, followed by the filtered tape, and showed the least amount of confidence on the whispered tape. These findings indicate that the laryngeal fundamental frequency appears to be a more important acoustic cue in speakers' sex identification tasks than the resonance characteristics of the speaker.

Another experiment involving the effects of selected vocal disguise upon speaker identification by listening was conducted by Reich et al. (1976). The results of this experiment suggested that certain vocal disguises markedly interfere with speaker identification by listening. The reduction in speaker identification performance by vocal disguise ranged from naïve listeners was 22% (slow rate) to 32.9% (nasal) and sophisticated listeners was 11.3% (hoarse) to 20.3% (nasal). In general, results of this experiment show that nasal disguise (naïve and sophisticated listeners) was the most effective, while slow rate disguise (naïve listeners) and hoarse disguise (sophisticated listeners) were the least disguises on speaker identification by listening.

Whiteside (1998) conducted an experimental study to test whether three phonetically naïve listeners were able to identify the speaker's sex from brief (30msec to 100 msec) voiceless fricative segments. All speech segments were extracted from sentences spoken by members of a group of 3 men and 3 women with a British general Northern accent. The consonants segments were significantly identified by the listeners with an accuracy of 64.4 %. A sample of the fricative segments was chosen to investigate acoustic and phonetic differences related to a speaker's sex, using spectrographic analysis. Analysis showed that on the average the frication of the women's voiceless fricatives was significantly higher in frequency than that of men.

A study was done by Amino (2004) to examine the effects of stimulus content and speaker familiarity on perceptual speaker identification with 16 unfamiliar speakers who were naive speakers of Japanese. Results of the identification test with naive listeners showed that nasal sounds */na/ and /nja/* obtained higher score followed by nasal sound */ma/.*

Amino and Arai (2009) conducted perceptual speaker identification experiment in order to examine the reproducibility of the nasal effectiveness, and to see the effects of the following vowels. Coronal nasals (nasals produced with raising the tip or blade of the tongue) were shown to be effective despite the different speaker set or the following vowels, and the stimuli containing a nasal were significantly better than those without it. The contours of the nasals appeared to be stable within a speaker, compared to other type of sounds.

Speaker identification by listening only, one of the methods discussed is, far from being 100% accurate. However, aural or perceptual method is an entirely subjective method; an expert witness using only this method would be unable to justify conclusions in the court of law.

2.4.2 Speaker Identification by Visual Examination of Spectrograms

The second method of speaker identification is based upon the visual examination by comparison of spectrograms. Spectrogram is a three dimensional (time, amplitude and frequency) representation of speech sounds. These were used in attempts to identify speakers by visually matching their speech/voice patterns. In the mid 1940s, the scientist of the Bell Telephone Laboratories in USA developed the first sound spectrograph (the Sonagraph), a visual record of speech including frequency, intensity and time (McDermott & Owen, 1996). In 1950s, Lawrence Kersta, an engineer from the Bell Telephone Laboratories, developed “voiceprint identification” and studies using the spectrograph were carried out in the 1950s and 1960s in USA (Hollien, 2002).

A set of reviewer suggested of using the spectrogram as a method for speaker recognition (Kopp & Green, 1946; Potter, 1964). They portray talker dependent features in addition to the phonetic variations. The term Voiceprint was introduced by Lawrence Kersta (1962) and he studied the patterns on sonograms and reported that these exhibited features could be used to identify speakers. He published a paper on “Voice Identification” in which he initiated an erroneous idea that there is a close relationship between fingerprint and voice print. Kersta’s identification method where the human observers visually match the spectrograms and to duplicate his investigation is what we believe is methodological and analytical improvements.

Kersta (1962) examined the “voiceprint” using spectrograms taken from five clue words spoken in isolation using 12 talkers and closed test identification. High school girls were trained for 5 days to identify talkers from spectrograms on the basis of eight “unique acoustic cues”. A 5X4, 9X4, or 12X4 matrices of spectrograms was presented to the observer whose task was to group the spectrogram in piles representing the individual talkers. Results of the study shown high rate of identification accuracy that were inversely related to the number of talkers. For 5, 9 and 12 talkers, identification rate were 99.6%, 99.2% and 90.0% respectively and for word spoken in isolation the correct rates were higher for the “bar prints” than for the “contour prints”. Voiceprint can be considered as an acoustic equivalent of a fingerprint, and many people claim to be able to identify speakers from spectrograms. This is, however, a controversial standpoint and, thus, it is of some interest to try to estimate the amount of influence on acoustic characteristics of recording situation, storage medium and related factors.

Kersta (1962) has examined spectrograms and reported that more than 99% of correct identification. However, similar results are not obtained by other researches. The correct identification scores reported by Kersta are outstandingly high, 99% - 100%, for short words spoken either in isolation or in context, as compared to (a) 84% - 92%, for short words spoken in isolation, reported by Pollack, Pickett and Sumby (1954), (b) 89% for short words taken from context, reported by Pruzansky (1963) and (c) 81% - 87%, for short words spoken in isolation, reported by Bricker and Pruzansky (1966). Some methods yielded virtually 100% correct identification rates when the test stimuli were identical sentence and somewhat lower rates when the test stimuli were short or monosyllabic words spoken in isolation. Young and Campbell (1967) reported that the correct identification rate for word in different context was 37%, and word in isolation was 78% using three words spoken by five speakers. The results were interpreted to indicate that different contexts decrease the identification ability of observers. This is because the shorter stimulus durations of word in context decreases the amount of acoustic information available for matching and the different spectrographic portrayals introduced by different phonetic contexts outweighs any intra-talker consistency.

Further, the duration of the speech sample required for speaker identification is not known. Stevens, Williams, Carbonell and Woods (1968) compared aural with the visual examination of spectrograms using a set of eight talkers and a series of identification tests was carried out. The average error rate for listening is 65% and for visual is 21%. They investigated and observed that mean error rate decreased from approximately 33.0% to 18.0% as the duration of speech samples increased from monosyllabic words to phrases and sentences. They also concluded that for visual identification, longer utterances increase the probability of correct identification.

Bolt, Cooper, David, Denes, Picket and Stevens (1970) reported that speech spectrograms, when used for voice identification, are not analogous to fingerprints, primarily because of fundamental differences in the source of patterns and differences in their interpretation. To assess reliability of voice identification under practical conditions, whether by experts or explicit procedures are not yet been made, and requirements for such studies are not outlined. Hecker (1971) reported that speaker recognition by visual comparison of spectrograms is coming into use in criminology, but the validity of this method is still in question.

A survey of 200 voice identification comparisons made by Federal Bureau of Investigation (FBI) examiners (Koenig, 1968) was used to determine the observed error rate of the spectrographic voice identification technique under actual forensic conditions. The survey revealed that decisions were made in 34.8% of the comparisons with 0.31% false identification error rate and 0.53% false elimination error rate. Hence, some scientists supported speaker identification by visual examination of spectrograms and others did not support this technique for speaker identification.

Bolt, Cooper, David, Denes, Picket and Stevens (1973) wrote a letter to the editor, reviewed recent research on speaker identification by comparisons of speech spectrograms by human observers. Various factors affecting the reliability of identification are discussed, particularly those that would be present in practical forensic situations. The interpretation of the new data leads us to reiterate previous conclusion: that the degree of reliability of identification under practical conditions has not been scientifically established.

Black et al. (1973) replied the above letter and highlighted their opinion on voice identification as produced in real-life cases by specially trained professional examiners using a combined aural and visual examination of speaker samples. The opinion of these authors is not based on personal experience or even on a direct observation of these examinations. In addition, they disregard crucial facts that strongly interact with the reliability of those positive decisions produced by professional, full-time examiner, such as their special training and responsibility, the five possible decisions they are entitled to produce after each examination, the number of samples, and the length of time used to perform each examination. It is our contention that opinions based on feelings other than in actual forensic experience are of little value, irrespective of the scientific authority of those who produce such an opinion.

Findings of a large scale study by Tosi (1972) were published in which attempts were made to more closely imitate law enforcement conditions, but only spectral comparisons were made (no aural). The "forensic model" included open set trials, non-contemporary samples, trained examiners and high confidence decisions. This resulted in an approximate error rate of 2% for false identification and 5% for false elimination. Hollien (1974) commented that spectrographic speaker identification may differ in actual forensic conditions, since most of the experiments conducted are in laboratory condition.

A two-year research on voice identification through visual examination of spectrograms was performed with two fold goal of inspection. Kersta's (1962) assert in this topic was testing models as well as variables connected to forensic tasks. The 250 speakers used in this testing were arbitrarily selected from a homogeneous population of 25000 males students speaking general American English belonging to Michigan State University. A total of 34996 test trials of identification were performed by 29 skilled examiners. Ten to forty known voices were involved in each trial, in diverse conditions, with open and closed trials, non-contemporary and contemporary spectrograms, six or nine clue word spoken in isolation, in a fixed context and in a random context, etc.

The examiners were forced to reach a positive decision (identification or elimination) in each instance, taking an average time of 15 minutes. Their decisions were based solely on inspection of spectrograms, listening to the identification by voices was excluded from this experiment. The examiners graded their self-confidence in their judgments on a 4-point scale (1 and 2, uncertain; 3 and 4, certain). Results of this experiment confirmed Kersta's experimental data, which involved only closed trials of contemporary spectrograms and clue words spoken in isolation. Experimental trials of this study, correlated with forensic models (open trials, fixed and random contexts, non contemporary spectrograms), yielded an error of approximately 6% false identifications and approximately 13% false eliminations.

The examiners judged approximately 60% of their wrong answers and 20% of their right answers as “uncertain”. This suggests that if the examiners had been able to express no opinion when in doubt, only 74% of the total number of tasks would have had a positive answer, with approximately 2% errors of false identification and 5% errors of false elimination. Main differences of conditions that could exist between models and real cases are as follows:

(1) Population of known voices: In forensic cases, the catalog of known voices could theoretically include millions of samples. In the present practical situations that police must handle. In these cases the catalog of known voices is open, true, but limited to a few suspected persons. Therefore, it seems reasonable to disregard size of the population of known voices as a differential characteristic that could hamper extrapolation of results from the present experiment to real cases.

(2) Availability of time and responsibility of the examiners: In real cases, a professional examiner may devote all the time necessary to reach a conclusion. In addition, he is aware of the consequences that a wrong decision could mean to his professional status as well as the consequences to the speaker whom he might erroneously identify. Availability of time and responsibility between experimental and professional examiners that might help to improve the accuracy of the professional examiners.

(3) Type of decision examiners are urged to reach in each trial: In the statistical models, the examiners were forced to reach a positive conclusion in each trial, even if they were uncertain of the correct response. In real forensic cases, the professional examiner is permitted to make the following alternative decision (a) Positive identification; (b) Positive elimination; (c) Possibility that the unknown speaker is one of the suspected persons, but more evidence is necessary in order to reach a positive identification; (d) Possibility that the unknown speaker is none of the available suspected persons, but more evidence is necessary to reach a positive elimination; (e) Unable to reach any conclusion with the available voice samples. The possibilities of alternative decisions could confer an extremely high reliability to the positive identifications or eliminations.

(4) Availability of clues: In the experimental models of this study, only spectrograms of nine or six clue words were available to the examiners for visual inspection. Rather, a

professional examiner is entitled to request as many samples as he deems necessary to reach a positive conclusion. In real forensic cases the professional examiner must necessarily listen first to the unknown and known voices while processing the spectrograms for visual comparison. A combination of methods of voice recognition by listening and by visual enhances the accuracy of voice identifications.

In summary, the discussions above suggest, in opinion of Tosi (1972) that the conditions a professional examiner encounters performing voice identifications wioul tend to decrease rather than increase the percentage of error observed in the present experiment.

Hazen (1973) reported that for reduced population, error rates were higher for closed tests (12.86% and 57.14%) than for open tests (11.91% and 52.38%), but were almost five times as great for the different context condition (57.14% and 52.38%) than for the same context condition (12.86% and 11.91%). Hollien (1974) commented on spectrographic speaker identification, it now appears that the controversy about "voiceprints" is doing the judicial system and the relevant scientific community a considerable disservice. Final perspective of the letter is to urge responsible investigators interested in the problem to focus their research activities on the development of methods. That will proide efficient and objective ways to identify individuals from their speech, especially in the forensic situation. All these may be possible under undisguised voice. However, with vocal disguise the situation may be different.

Some authors have experimented on speaker identification in vocal disguises. Reich, Moll and Curtis (1976) described an experiment involving the effects of certain vocal disguises upon speaker identification using spectrograms. The results of this test suggested that certain vocal disguises markedly interfere with spectrographic speaker identification. The reduction in speaker identification performance ranged from 14.17% (slow rate) to 35.00% (free disguise). These experimental data obviously contradicted Kersta's (1962) claim that speaker identification using spectrograms was in actual fact not affected by attempts at disguising one's voice. The mean performance level (56.67% correct) on the undisguised task was considerably poorer than the data for similar experimental conditions (approximately 80%) by Tosi, Oyer, Lashbrook, Pedrey, Nichol and Nash (1972).

The nasal disguise, for example, was the most effective disguise in speaker identification by listening experiment (Reich & Duke, 1979). In contrast, the nasal disguise was the least effective in a previous spectrographic matching experiment (Reich et al., 1976). The power spectra of nasal consonants (Glenn and Kleiner, 1968) and co-articulated nasal spectra (Su, Li and Fu, 1974) provide strong cues for the machine matching of speakers. It is interesting to know the listeners in the present study were unable to successfully utilize these seemingly speaking dependent cues. The free (i.e., extemporaneous) disguise proved to be very effective in both the spectrographic matching experiment (Reich et al., 1976) and the present listening experiment.

There are a few disguise, but first it is important to determine if the talker is attempting to alter, or not alter, his or her speaking mode. Reich (1981) examined the capability of naïve and sophisticated listeners to identify extemporaneous disguise in the male voice. Both naïve and sophisticated listeners were able to spot the existence of selected disguises with an elevated level of correctness and consistency.

Thus, the effects of certain vocal disguises markedly interfere with spectrographic speaker identification as well as speaker identification by listening. The nasal and slow rate were the least effective disguises, while free disguise was the most effective disguise upon the spectrographic speaker identification, and nasal disguise (naïve and sophisticated listeners) was the most effective, while slow rate disguise (naïve listeners) and hoarse disguise (sophisticated listeners) were the least effective disguises upon the speaker identification by listening. Both naïve and sophisticated listeners were able to detect the presence of selected vocal disguise with a high degree of correctness and consistency.

Hecker et al. (1968) conducted an experiment to establish the outcome of induced stress on speaker identification and reported that listeners identified the stressful response of some subjects with better than 90% accuracy and of other only at chance level. The test phrases from contrasting responses were analysed with respect to level and fundamental frequency, and spectrograms of these test phrases were examined. The results of this experiment indicated that task-induced stress can produce a number of characteristic changes in the acoustic speech signal. Most of these changes are attributable to modification in the amplitude, frequency and detailed waveform of the glottal pulses. The authors also concluded that the other changes noticed were resulted from differences in articulation.

To list out few Indian studies, Pamela and Savithri (2002) investigated the reliability of voiceprints by extracting acoustic parameters in the speech samples using wideband spectrograms. Twenty-nine bisyllabic meaningful Hindi words with 16 plosives, five nasals, four affricates and four fricatives in the word-medial position formed the material. Percent of time a parameter was the same within and between subjects was documented. The results indicated no significant difference in F2, onset of burst and frication noise, F3 transition duration, closure duration, and phoneme duration between participants. However, the results indicated high intra-subject variability. High intra-subject variability for F2 transition duration, onset of burst, closer duration, retroflex and F2 of high vowels was observed. Low inter-subject variability and high intra-subject variability for phoneme duration was observed indicating that this could be considered as one of the parameters for speaker verification. The results indicated that more than 67% of measures were different across subjects and 61% of measures were different within subjects. It was suggested that two speech samples can be considered to be of the *same speaker* when not more than 61% of the measures are different and two speech samples can be considered to be from different speakers when more than 67% of the measures are different. Probably this was the first study in Indian context, an attempt to establish benchmarking was done.

Ranganathan and Savithri (2003) studied speaker identification in disguise speech. In this study, they investigated the effect of vocal disguise upon accuracy of spectrographic speaker identification in disguised and a comparison between two conditions. The results indicated no significance difference between the disguised and undisguised speech for both statement and command upon spectrographic analysis.

Arjun (2014) conducted a preliminary study on “Speaker identification using fricatives in Kannada speaking individuals” which included spectrographic analysis of speech samples such as fricative duration, fricative amplitude and centre frequency of frication. This preliminary study on fricatives in speaker identification showed relatively positive results on few specific combinations of acoustic parameters. This study explored the need to create benchmark for speaker identification using fricative duration and fricative amplitude (T2), the best combination among the selected acoustical parameters of the fricative /s/ and /ʃ/ of Kannada speaking individuals.

2.4.3 Speaker Identification by Machine Method (Objective Method)

Voice processing technology became quite popular in the years following identification by the aural mode. The simplest approach used was to generate and examine amplitude and frequency, time matrices of speech samples. The other approach used was to extract speaker dependent parameter from the signals and analyse them by machines. The objective methods include ***semi-automatic method and automatic method***. In the semi-automatic method, there is extensive involvement of the human with the computer, whereas in the automatic method, this contact is limited.

There are two phases in automatics speaker recognition, training phase and testing phase. Initially, during the training phase, a large number of exemplars or tokens or enrolment samples are collected for each speaker. For each of these samples, the feature is computed and saved in memory along with the speaker's label or identity. This is referred to as the training database. In case of testing phase, an utterance of the speaker is fed to the system and the speaker recognition system compares it with the stored database to determine the identity of speaker or verify speaker's identity. The training database itself is divided into two sets: (i) Training set (about 70%) and Test set (about 30%).

The automatic speaker recognition system goes through the subsequent steps in order to reach at a judgment. They are feature extraction, pattern matching and classification. Certain measurements are made on the training samples to derive what are called 'features' or 'templates'. A number of alternatives are available for feature selection. Some of the desirable characteristics for the features are: (i) They must be highly discriminable across speakers (ii) variability for the same speaker from session to session must be minimal (iii) Either there must be a robust method for extracting the features or the features must be robust to extraneous factors such as recording conditions and (iv) It must be difficult to impersonate.

Automatic speaker verification was accomplished by Luck (1969) using cepstral measurement. The phrase 'My code is…..' was used to characterize short segments in each of the first two vowels. Additional parameters were also assessed like the speaker's pitch and the duration of the word 'my'. Like identification, verification also presents as having a black and white decision- the claimant is the authorized speaker or not. A comparison of the reference data with the authorized speaker was carried out. This shows that if the reference data was collected over a period of time, say many days, then verification can be done as late as two months after the collection of the sample, whereas, if reference data was collected at one sitting, verification would be very inaccurate as little as one hour later. Four authorized speakers and 30 imposters were examined, with error rates obtained from 6% to 13%. When individuals tried to deceive the system by acting as impostors of the authorized speaker, they could not do so. It has been observed by many who have seen the system in operation that greater accuracy would be obtained if a final decision were based on a series of two or three repetitions of the test phrase. This is to say that increased accuracy depends on increasing the information available to the decision mechanisms.

Wolf (1972) suggested associations connecting the voice signals, vocal tract shapes and gestures as an efficient approach in selecting parameters. It is desirable to use acoustic parameters for mechanical recognition of speakers as that are very much connected to characteristics of voice that discriminate speakers. Instead of measuring the entire utterance and giving general parameters, only major significant characteristics of certain segments are used. Speech events can be manually located within the utterance after feeding it into a simulated speaker recognition system and then measuring the parameters at these locations to classify the speakers. Helpful parameters found were word duration, feature of vowel and spectra of nasal consonant, fundamental frequency (F_o), assessment of voice onset time (VOT) and glottal source spectrum slope.

Atal (1972) reported characteristic cue of speaker identification by examining the temporal variations of pitch in speech. Sixty utterances spoken by ten speakers, consisting of six repetitions of the similar sentences were recorded for analysis of the pitch. The pitch contours were linearly changed so that the ratio of inter-speaker to intra-speaker variance in the altered space was maximized. Again, the speaker sample with the least distance in the reference vector was taken as the correct identification. 97% of correct identification was reported which led to temporal variations of pitch being suggested as a good and effective parameter for automatic speaker recognition.

Atal (1974) examined a number of different parameters using model of linear prediction for their efficiency for automatic recognition of speakers from their voices. Twelve predictor coefficients were determined approximately once in every 50msec from speech sampled at 10 kHz. The predictor coefficients, as the impulse response function, the area function, the autocorrelation function and the cepstrum function were used as input to an automatic speaker recognition system. The speech information consisted of 60 utterances, consisting of 6 repetition of the same sentence spoken by 10 speakers. The decision for identification was based on the distance of the test sample vector from reference vector for different speakers in the population; the speaker corresponding to reference vector with minimal distance was judged to be the unknown speaker. In verification, the speaker was verified if the distance between the test sample vector and the reference vector for the claimed speaker was less than a fixed threshold. He reported that cepstrum was found to be the most useful parameter, providing an identification accurateness of 70% for speech (50msec in duration), which increased to more than 98% with a duration of 0.5sec. Same speech data was used to find the accuracy of verification which in turn was found to be increased approximately by 83% with a duration of 50msec to 95% with a duration of one second.

The results of the research suggest several conclusions. Firstly, it may be concluded that n-dimensional Euclidian distance among long-term speech spectra may be quite successfully utilized as criteria for speaker identification, at least under laboratory conditions. Moreover, this method displays a number of advantages: (a) It is relatively simple to carry out; (b) it eliminates such crucial factors as the time-alignment problem; (c) the data generated for the identification does not depend on the overall power level of the speech samples used; and (d) the process does not depend on a human and, hence, subjective judgements. Finally, it appears that distortions created by limited pass band and stress as these two factors are defined in these experiments have only minimal effects on the sensitivity of the LTS (least trimmed squares) vector as a speaker identification cue.

Study done by Glenn and Kleiner (1968) aimed to find out the validity of the hypothesis that each speaker produces a unique and identifiable power spectra during nasal phonation in recognition experiments. The result showed 97% of identification accuracy with nasal sound */n/* for the entire experiment. The study on technique of automatic speaker identification was described based on the physiology of the vocal apparatus and in actual fact not dependent of the oral communication. The power spectra formed are changed and statistically matched during nasal phonation. Initially, the population of 30 speakers was divided into three subclasses, each containing 10 speakers. Subclass 1 contained 10 male speakers, subclass 2 contained 10 female speakers and subclass 3 contained an additional 10 male speakers. For each speaker, all 10 samples of the spectrum of */n/* from the test set were averaged to form a test vector. The test vectors were compared, with the stored speaker reference vectors for the appropriate subclass. The values of the cosine of the angle between the reference and the test vectors were correlation values between the test vector for a given speaker and the reference vector for each speaker in the subclass. The maximum correlation value for each test vector was used and 97% over all correct identification was attained. Next, the effect of a larger population was tested by correlating each speaker's average test data with the reference vectors for all 30 speakers and an average identification accuracy of 93% was reached. Finally, the effect of averaging speaker samples was tested as follows. The same speaker reference vectors based on all 10 training samples were used. However, the test data were subjected to varying degrees of averaging. First, single-speaker samples were correlated with the 30 speaker reference vectors. The average identification accuracy for all 300 such samples (10 per speaker) was 43%. Then, average of two speaker samples from the test data were taken as test vectors. The average identification accuracy for 150 such vectors was 62%. Next averages of five speaker samples from the test data were taken as test vectors. The average identification accuracy for 60 such vectors was 82%. The procedure developed to exploit this information provides a basis for automatic speaker identification without detailed knowledge of the message spoken. This study was focused on the nasal phonation using power spectra, used reference and test vectors. This is the only study used spectra of nasal continuants.

Mel Frequency Cepstral Coefficients (MFCCs)

The model tradition is to represent the log spectrum with frequency axis in Hz scale. It is likely to compute the log spectrum with the frequency axis in mel-scale (or in bark-scale or in logarithmic scale). Figure 1 shows the schematic representation of log spectrum of a vowel segment in Hz and Mel scale. Note that the low frequency components in the mel scale are widely separated where as the high frequencies components are packed in. There are standard formulae for converting frequency in Hz to frequency in mel (or bark) and vice-versa. Discrete fourier transform (DFT) of log spectrum with the frequency axis in mel scale gives mel-scale cepstrum.

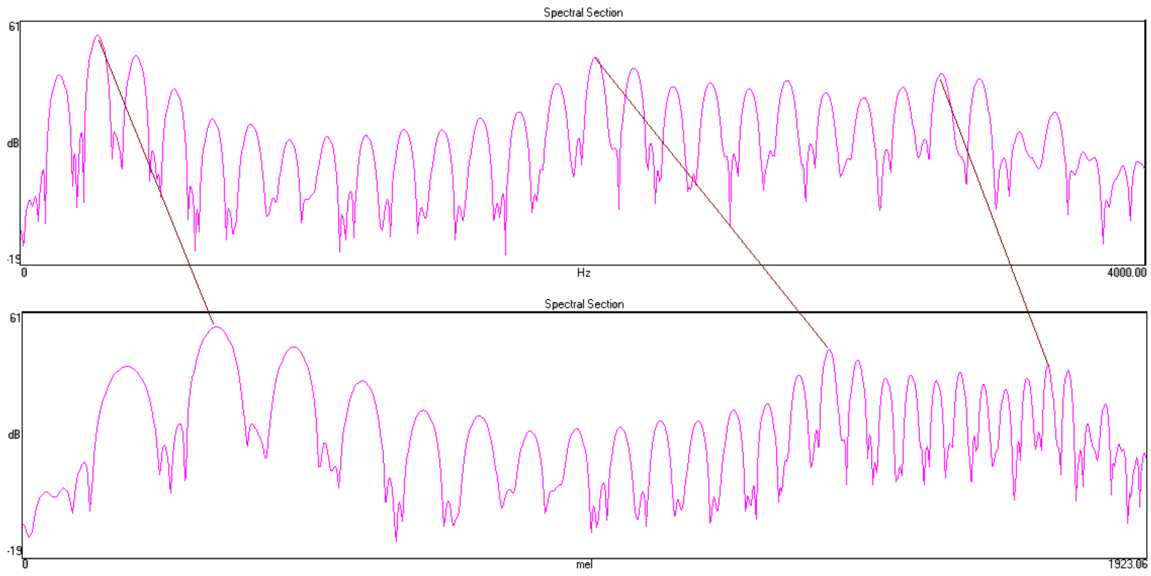


Figure 1: Log spectrum of a vowel segment in Hz scale (top) and in Mel scale (bottom).

Experimentally researchers have established that of all the features, Mel frequency **c**epstral **c**oefficients (MFCCs) give the finest performance for automatic speech recognition, speaker recognition, language identification etc. Although, in principle MFCCs are alike to the mel-scale frequency cepstrum described above but the differences arise in two distinct ways: (a) A filter bank to obtain the log spectral components with frequency axis in mel scale (b) Instead of a DFT, a transformation called Discrete Cosine Transform (DCT) is used.

Using a filter bank: The magnitude squared spectrum is computed in the standard manner using DFT applied on the windowed pre-emphasized speech segment. The magnitude squared spectrum is warped so that the frequency scale is in mel. Then the mel scale magnitude squared spectrum is divided into M bands of equal width by means of overlapping triangular windows, where M is a pre-decided number of coefficients (Figure 2). The magnitude squared spectrum is multiplied by a triangular function which retains frequency components of only one band (Figure 2(a)). The total energy in the band is the sum of all the non-zero spectral components of that band. The energy is then represented in dB (or log) scale to obtain the filter bank outputs. This procedure gives M filter bank outputs.

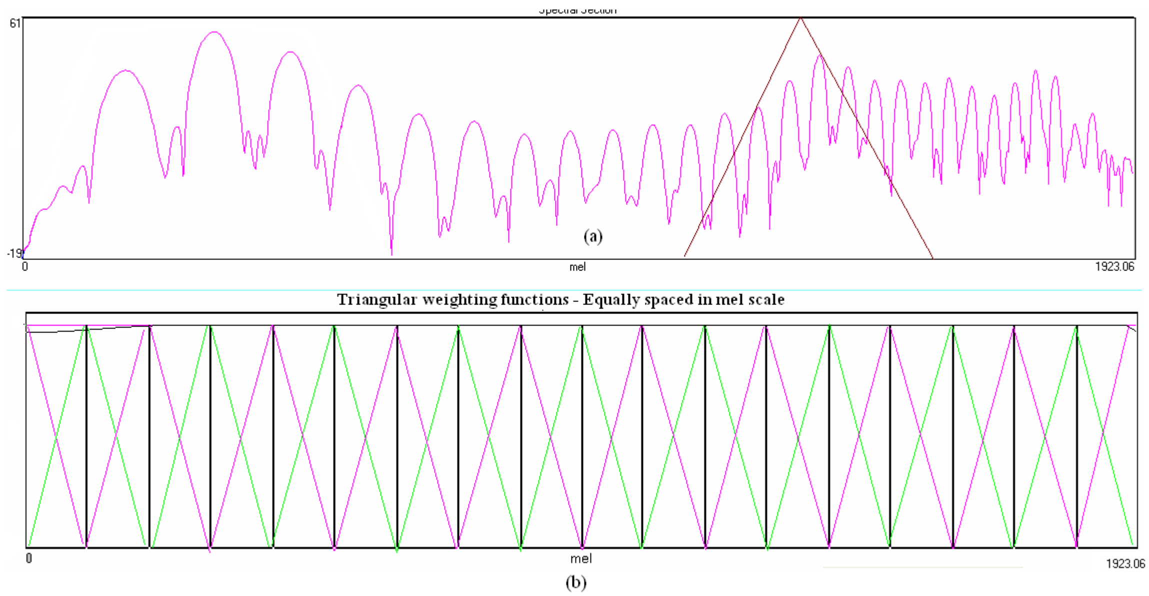


Figure 2: (a) Magnitude squared spectrum in mel scale is multiplied by a triangular bandpass filter response. (b) A number of such triangular filters constitute a filter bank. The filters are equally spaced in the mel frequency scale.

Study conducted by Kinnunen (2003) indicated that the MFCCs are the most apparent example of a feature set that is comprehensively used in speaker recognition. By using MFCCs feature extractor, one makes a postulation that the human hearing mechanism is the finest speaker recognizer. Authors aimed to find the critical parameters that affect the performance and also tried to give some general guidelines about the analysis parameters. They conducted experiments on two speech corpora using vector quantization (VQ) speaker modeling. The corpora were a 100 speaker subset of the American English TIMIT corpus, and a Finnish corpus consisting of 110 speakers. Although noise robustness is an important issue in real applications, however, it was outside the scope of the thesis. The author's main attempt was to gain at least some understanding of what is individual in the speech spectrum. The results indicated that in addition to the smooth spectral shape, a significant amount of speaker information is included in the spectral details, as opposed to speech recognition where the smooth spectral shape plays more important role.

Hasan, Jamil, Rabbani and Rahman (2004) used MFCCs for feature extraction and vector quantization in security system based on speaker identification. Database consists of 21 speakers, which included 13 male speakers and 8 female speakers. The system has been implemented in

Matlab 6.1 on windows XP platform. Results showed 57.14% speaker identification for code book size of 1, 100% speaker identification for code book size of 16. Study revealed MFCCs method has been applied for speaker identification.

Linear Predictive Coding (LPC) parameter and Mel Frequency Cepstral Coefficients (MFCCs) were used for speaker identification by Mao, Cao, Murat and Tong (2006). The text-dependent recognition rate of 50 speakers increased from 42% to 80% and the text-independent recognition rate of 50 speakers increased from 60% to 72%.

A study by Pruthi and Epsy-Wilson (2007) was conducted and they extracted acoustic parameters from nasalized vowels for automatic detection and reported accuracies of 96.28%, 77.9% and 69.58% using StoryDB, TIMIT and WS 96/97 databases respectively.

Wang, Ohtsuka and Nakagawa (2009) used a method that integrated the phase information with MFCCs on a speaker identification task. The speech database consisted of normal, slow and fast speaking modes. The anticipated new phase information was more robust than the original phase information for all speaking modes. By integrating the new phase information with the MFCCs, the speaker identification error rate was remarkably reduced for normal, slow and fast speaking rates in comparison with a standard MFCCs-based method. The experiments showed that the phase information is also very helpful for the speaker verification.

Tiwari (2010) studied using MFCCs to extract, characterize and recognize the information about speaker identity with different number of filters. The results showed 85% effectiveness using MFCCs with 32 filters in speaker recognition task.

Chandrika (2010) compared the influence of handsets and cellular networks on the performance of a speaker verification system using MFCCs when recording is done with mobile handsets over a cellular network as against digital recording. Ten Kannada speaking normal subjects were included in the study. The subjects ranged in age from 20 to 40 years. The spoken materials used was long vowels */a:/, /i:/ and /u:/* in medial position occurring in five target Kannada words embedded in sentences (text-dependent). A speaker participating in an experiment was given a CDMA handset (Reliance, LG) and call was made to the speakers' handset from another CDMA Reliance LG handset with recording option held by the experimenter. Speech signal was recorded as the speaker uttered the test sentences and similar recording procedure was used with GSM (Reliance, Nokia) handsets. Simultaneously the recording was also done directly in to the digital recorder which was kept at a distance of 10 cm from the speaker's mouth. The recording was repeated with same procedure another day for reliability check. The average MFCCs vector over the entire segment was extracted using MATLAB coding. The overall performance of speaker verification system using MFCCs was about 80% for the data base considered and overall performance of speaker recognition is about 90% to 95% for vowel */i/.* The above study used limited number of participants and it considered contemporary samples only.

Ramya (2011) conducted study on speaker identification under electronic vocal disguise using MFCCs. The results indicated the percent correct identification was above chance level for electronic vocal disguise for females. Interestingly vowel */u:/* had 96.66%, vowel */a:/* had 93.33% and vowel */i:/* had 93.33%.

Singh and Rajan (2011) studied on vector quantization approach for speaker recognition using MFCCs and inverted MFCCs due to high accuracy and its simplicity. Efficiency of 98.57% with duration of two seconds was the result obtained for speaker recognition.

"Speaker identification using cepstral coefficients and MFCCs in Malayalam nasal co-articulation" was studied by Jyotsna (2011). Speaker identification was found to be better when MFCCs was used as the parameter compared to cepstral coefficients. This result was consistent with the study done by Hasan, Jamil, Rabbani and Rahman (2004). The dynamic range of Euclidian values was more for MFCCs compared to cepstral coefficients. Vowels and nasals also showed better identification using MFCCs compared to cepstral coefficients. The study revealed that the benchmark for speaker identification using cepstral coefficients was above 80% and using MFCCs, it was above 90% for nasal co-articulation in Malayalam.

A study on "Benchmark for speaker identification using nasal continuants in Hindi in direct mobile and network recording" was conducted by Rida (2014). The aim was to establish benchmark for speaker identification for nasal continuants in Hindi using MFCCs. The objective of her study was to provide benchmarks for MFCCs for Hindi nasal continuants in mobile and network conditions. Ten participants between the age range of 20-40 years with at least ten years of exposure to Hindi language as a mode of oral communication were included in the study.

Materials included six Hindi sentences with bilabial, dental and velar nasals embedded in words in all positions. Participants were instructed to speak the sentences under two conditions- directly into the recording mobile (live) and through another mobile into the recording mobile phone (network)- three times at an interval of one minute. The network used for making the calls was Vodafone (GSM 900/ GSM 1800 MHz frequency) and the receiving network was also Vodafone on a sony Ericson xperia pro mobile phone. Analysis of the data was carried out using SSL Workbench (Voice & Speech Systems, Bangalore, India) to extract Euclidian distances. A speaker was presumed to be identified correctly when the Euclidian distance between training and test samples was the least. Results indicated that the percent correct speaker identification was 100%, 90% and 100% for */m/, /n/ and /ŋ/* respectively when live recording was compared with live recording using MFCCs. Results indicated that the percent correct speaker identification was 50%, 80% and 90% for */m/, /n/ and /ŋ/* respectively when network recording was compared with network recording using MFCCs. Results indicated that the percent correct speaker identification was 80%, 70% and 50% for */m/, /n/ and /ŋ/* respectively when live recording was compared with network recording under telephone equalized condition using MFCCs. Results indicated that the percent correct speaker identification was 90%, 90% and 30% for */m/, /n/ and /ŋ/* respectively when live recording was compared with network recording under telephone not equalized condition using MFCCs. Results indicated that nasal continuant /ŋ/ had the best percent correct speaker identification among the nasals except under telephone equalized and not equalized conditions.

The review states that in the semi automatic methods of speaker identification the following parameters addressed were, ***First and Second Formant Frequencies*** (Stevens, 1971; Atal, 1972; Nolan, 1983; Hollien, 1990; Kuwabara & Sagisaka, 1995; Lakshmi & Savithri, 2009), ***Higher Formants*** (Wolf, 1972), ***Fundamental Frequency (F_0)*** (Atkinson, 1976), F_0-contour (Atal, 1972), ***Linear Prediction Coefficients*** (Markel & Davis, 1979; Soong, Rosenberg, Rabiner & Juang, 1987), ***Long Term Average Spectrum*** (Kiukaanniemi, Siponen & Mattila, 1982), ***Cepstral Coefficients*** (Luck, 1969; Atal, 1974; Furui, 1981; Li & Wrench, 1983; Higgins & Wohlford, 1986; Che & Lin, 1995; Jakkhar, 2009; Medha, 2010; Sreevidya, 2010) and ***Mel Frequency Cepstral Coefficients*** (Plumpe, Quateri & Reynolds, 1999; Hassan, Jamil, Rabbani & Rahman, 2004; Chandrika, 2010; Tiwari et al., 2010; Ramya, 2011; Singh & Rajan, 2011; Jyotsna, 2011; Rida, 2014), have been used in the past.

Previous studies suggests that nasal regions of speech are an effective speaker cue, because the nasal cavity is both speaker specific, and fixed in the sense that one cannot change its volume or shape. Various acoustic features have been proposed for detecting nasality. Glass and Zue (1985) used six features for detecting nasalized vowels in American English. Pruthi and Epsy-Wilson (2007) extended Glass's word and selected a set of nine knowledge-based features for classifying vowel segments into oral and nasal categories automatically.

Flege (1988) conducted a study to examine the difference in the temporal extent of anticipatory nasal co-articulation in two different age groups. Results showed lack of significant difference between groups.

Mili (2003) examined labial co-articulation (anticipatory and carryover effect) of round vowels */u/ and /o/* with fricatives, laterals and trills in Malayalam. The results indicated strong anticipatory co-articulation compared to carry over co-articulation.

Speech articulators do not function individually and independently. The speech sounds are modified by the influence of contiguous phoneme. The acoustic properties of certain sounds are changed under the influence of adjacent sounds that is called co-articulation. These articulatory effects are categorized as anticipatory or forward co-articulation and carryover or backward co-articulation. Anticipatory refers to the influence of the articulatory characteristics of one sound on the production of the preceding sound. Carryover refers to the influence of the articulatory characteristics of one sound on the production of the succeeding sound. Most of the studies have found greater backward effect than forward effect (Ohde & Sharf, 1975). However, there are few reports of forward effects being greater than the backward effect (Butcher & Weiner, 1976). The power spectra of nasal consonants (Glenn and Kleiner, 1968) and co-articulated nasal spectra (Su, Li and Fu, 1974) provide strong cues for the automatic speaker identification by machines. In a study by Atal (1976), an acoustic measurement of nasal co-articulation in a consonant or vowel context was obtained by measuring spectral differences between the mean square spectrum of nasal consonants followed by a front vowel and that of the same consonant followed by a back vowel.

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.

The review provided the usefulness of Mel Frequency Cepstral Coefficients and the effectiveness of co-articulation of nasal continuants on preceding vowels in the field of speaker identification. There is no empirical data to establish thebenchmark 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. In this context, the present study was planned to establish the benchmarking in speaker identification using Mel Frequency Cepstral Coefficients on Vowels Preceding Nasal Continuants in Kannada. The objectives of the present study were:

1. To establish benchmark for speaker identification using MFCCs on vowels preceding nasal continuants in Kannada in direct recording.
2. To establish benchmark for speaker identification using MFCCs on vowels preceding nasal continuants in Kannada in mobile network recording.
3. To compare speaker identification scores/performances using MFCCs on vowels preceding nasal continuants in Kannada between live and mobile network recording.

CHAPTER III

METHOD

3.1. Participants

Twenty Kannada speaking neuro-typical adult males constituted as Group I (Mean age = 25 years, SD= 3.4) 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. They had a minimum of ten years of formal education with Kannada as one of the subject and all the participants belonged to the Mysuru dialect of Kannada language and were drawn from the work/residential place in and around Mysuru, Karnataka, India. The present study was carried out at the Forensic Speech Laboratory at the Department of Speech-Language Sciences, All India Institute of Speech and Hearing, Mysuru. Participants were included in the study only on fulfilling certain criteria. The inclusion criteria of subjects were as follows;

- No history of speech, language, hearing and communication problems
- Normal oral structures
- No other associated social or psychological or neurological problems
- Reasonably free from cold or other respiratory illness at the time of recording.

Written consent was taken from all the participants (Appendix-A) 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) (Appendix-B) was administered by a Speech Language Pathologist to rule out any misarticulations in the speech. Table A shows the details of the participants.

Table A: *Details of Participants*

Participants	Age in Years	Gender	Mother Tongue
1	22	M	Kannada
2	23	M	Kannada
3	23	M	Kannada
4	24	M	Kannada
5	30	M	Kannada
6	22	M	Kannada
7	30	M	Kannada
8	25	M	Kannada
9	32	M	Kannada
10	25	M	Kannada
11	23	M	Kannada
12	32	M	Kannada
13	21	M	Kannada
14	26	M	Kannada
15	26	M	Kannada
16	22	M	Kannada
17	23	M	Kannada
18	25	M	Kannada
19	26	M	Kannada
20	22	M	Kannada

3.2. Materials

The material used was thirty commonly occurring, meaningful Kannada words (Target words) containing the nasal continuant */m/* (Bilabial) and */n/* (Dental) and embedded in seventeen sentences (text-dependent). These sentences consisted of words with three basic vowels *(/a:/, /i:/, /u:/)* preceding to the 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/)*. The sentences are given in Appendix-C. Table B shows the target words from the sentences.

Table B: *Target words used in the study*

Target Words (Kannada)	IPA Transcription	Vowel Preceding Nasal Continuant
ಬಾದಾಮಿಯ	ba:da:miȷa	*/a:m/*
ಗ್ರಾಮ	gra:ma	*/a:m/*
ರಾಮನಿಗೆ	ra:manigɛ	*/a:m/*
ಸಾಮಾನ್ಯವಾಗಿ	sa:ma:nȷava:gi	*/a:m/*
ತಾಮ್ರ	t̪a:mra	*/a:m/*
ಭೀಮ	b^{h}i:ma	*/i:m/*
ಛೀಮಾರಿ	ʧi:ma:ri	*/i:m/*
ಧೀಮಂತ	d^{h}i:mənt̪a	*/i:m/*
ಸೀಮ	si:ma	*/i:m/*
ಸೀಮೆಯೆಣ್ಣೆ	si:mɛȷənɛ	*/i:m/*
ಭೂಮಂಡಲ	b^{h}u:məndəla	*/u:m/*
ಭೂಮಿಯನ್ನು	b^{h}u:miȷənnu	*/u:m/*
ಛೂಮಂತ್ರ	ʧ^{h}u:mənt̪ra	*/u:m/*
ಧೂಮಪಾನ	d^{h}u:mapa:na	*/u:m/*
ಹೂಮಾಲೆ	hu:ma:lɛ	*/u:m/*
ಭಾನುವಾರ	b^{h}a:nuva:ra	*/a:n/*
ಹಾನಿಕರ	ha:nikara	*/a:n/*
ಜಾನಪದ	ʤa:napəd̪a	*/a:n/*
ಜಾನುವಾರುಗಳ	ʤa:nuva:ru	*/a:n/*
ಕಾರ್ಖಾನೆಯ	ka:rk^{h}a:nɛ	*/a:n/*
ದೀನರಿಗೆ	d̪i:nərigɛ	*/i:n/*
ಹೀನಾಯವಾಗಿ	hi:na:ȷəva:gi	*/i:n/*
ಕೀನವಾಗಿತ್ತು	ki:na	*/i:n/*
ನವೀನನಿಗೆ	nəvi:nənigɛ	*/i:n/*
ತಲೀನಳಾಗುತ್ತಾಳೆ	t̪əlli:naɭa:gut̪t̪a:ɭɛ	*/i:n/*
ಗೂನು	gu:nu	*/u:n/*
ಕೂನ	ku:na	*/u:n/*
ಮಗೂನಾ	məgu:na:	*/u:n/*
ಊನ	u:na	*/u:n/*
ಶೂನ್ಯ	ʃu:nȷa	*/u:n/*

3.3. Recording Procedure

Speech samples of participants were recorded individually. Sentences were written on a card that was presented to the participants visually to familiarize themselves to utter the sentences at a normal rate of speech. 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. (a) Mobile network recording was done first and the network used for making the calls was Airtel on a NOKIA 101 and the receiving network was Vodafone on a Gionee S5.5 mobile phone. A participant participating in an experiment was given a NOKIA 101 handset (Airtel network). A call was made from the participants' handset to the experimenters' handset (Vodafone network) 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. 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 a laboratory 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.

3.4. 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 and Weenink, 2009) and down sampled to 8 kHz. Hence, 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.

3.5 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 3 shows a segment of vowel preceding nasal continuant from speech signal. The segmented vowels preceding nasal continuants were saved using a particular file name convention. For Ex: For live recording, speaker 1, first sample, first session, first occurrence was given the file name as ***LR_sp1_s1_aam1.wav and saved in a folder with name speaker1***. There were 240 sample files (3 preceding vowels * 2 nasals * 5 occurrences * 2 repetitions * 2 conditions = 120) for each speaker. Similar pattern was followed for all the other participants. Converted samples were stored in separate folders for each participant and separate folders for each repeated recording. These were stored in two main folders by the name 'live' and 'mobile network' recordings.

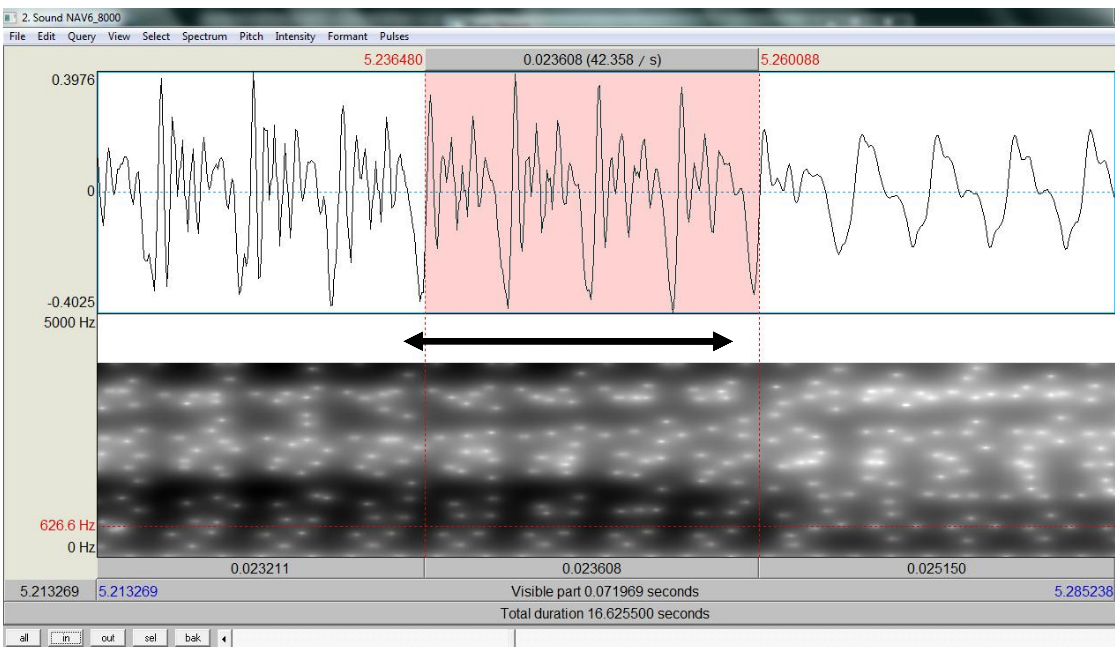


Figure 3: A segment of vowel preceding nasal continuant from a speech signal.

3.6 Analyses

Data analyses was carried out using Speech Science Lab (SSL) Workbench version 2.1 (Voice and Speech Systems, Bangalore, India) - a semi-automatic vocabulary dependent 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).

Design specification screen was used to enter the details and the following particulars were entered as shown in Figure 4

- ***Label***: the phoneme of interest for analysis needs to be entered (*/a:m/,/i:m/,/u:m/,/a:n/,/i:n/ & /u:n/* in the present study)
- ***Number of speakers***: total number of participants in the study (20 speakers in the present study)
- ***Number of occurrences of the label***: frequency of occurrence of a sound in a particular stimulus (5 occurrences in the present study)
- ***Number of sessions***: number of repetitions of the stimulus (2 sessions in the present study)

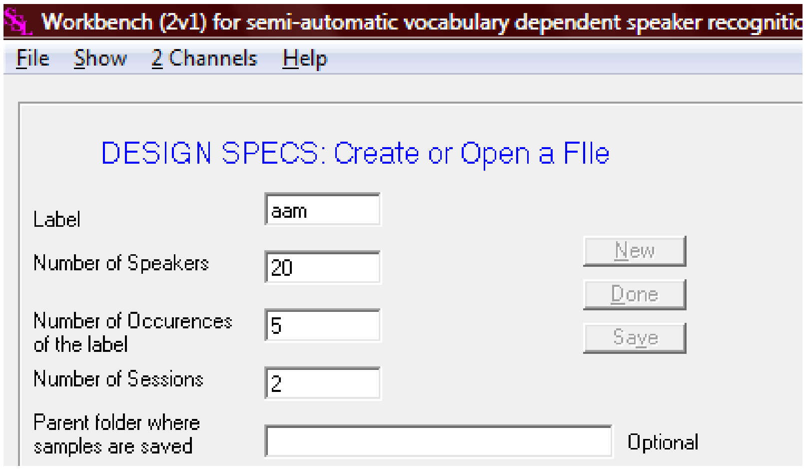


Figure 4: Design specification window of SSL Workbench version 2.1

The above information in sequence was entered in the system and an input specification file was defined using a NOTEPAD ***(.txt file)*** which in term automatically creates ***.dbs file*** that is the extension of NOTEPAD ***(.txt file)*** (Figure 5). The label of the file (Fig.4) was kept as 'aam' because the file was made for the long vowel */a:/* preceding nasal continuant */m/*. This was changed according to the vowel preceding nasal continuant for which the file was being made. The 'number of occurrences' and 'session number' was specified according to the vowel preceding nasal continuant being studied again.

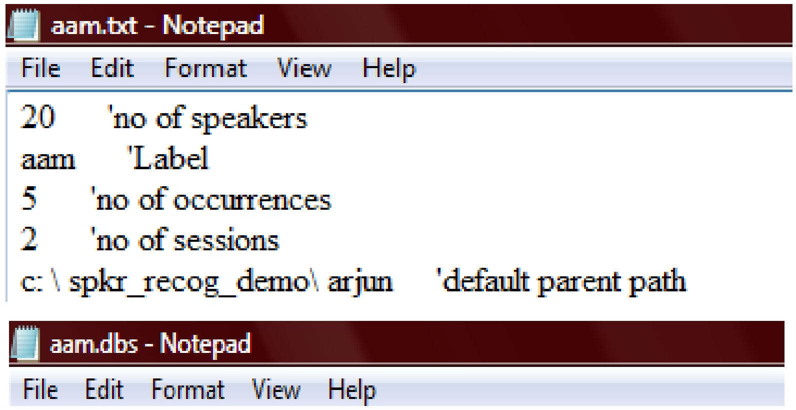


Figure 5: Input specification defined using NOTEPAD ***(.txt file)*** and ***.dbs file***

Further, segmentation of truncated speech samples of all twenty speakers were done by choosing the speaker number, session and occurrence number (Figure 6-8) because averaging and comparison between the same samples takes place at different sessions. Figure 6, 7 and 8 depict the schematic representation of selection of speaker, session and occurrence respectively in the software.

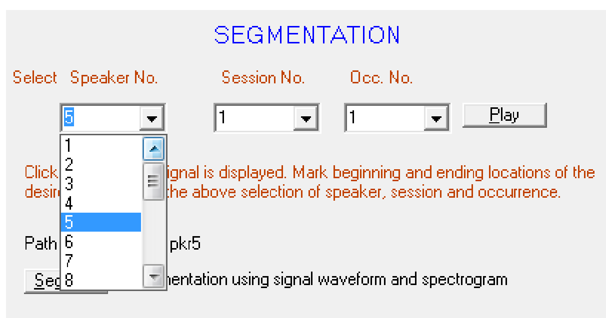


Figure 6: Speaker selection.

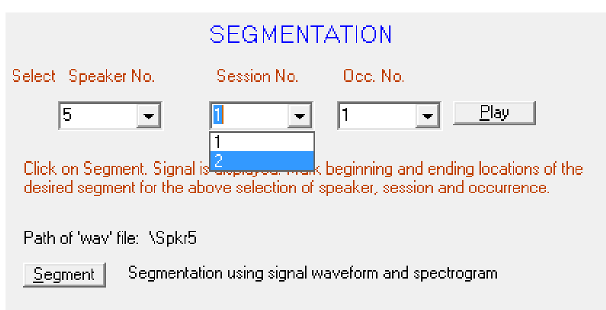


Figure 7: Session selection.

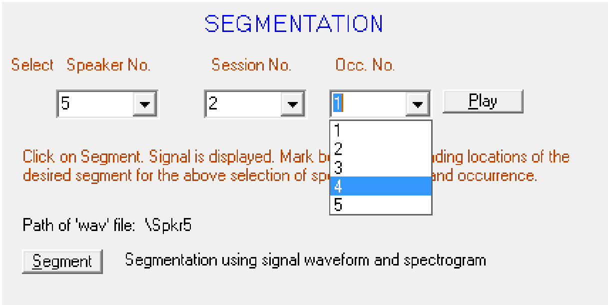


Figure 8: Occurrence selection.

Once the above segmentations were made, the 'segment' option was chosen to unlock the dialogue box for selecting the file from the particular parent path. Then the required truncated segments (Figure 9) were displayed. Figure 9 shows the beginning and ending locations of the desired segment was marked and the option 'Assign Highlighted' was chosen from the 'Edit' menu option (Figure 10) to confirm the segmentation. Figure 10 shows the dialogue box depicting confirmation of a highlighted segment.

Figure 9: Segmentation window depicting two sessions of an occurrence for a speaker.

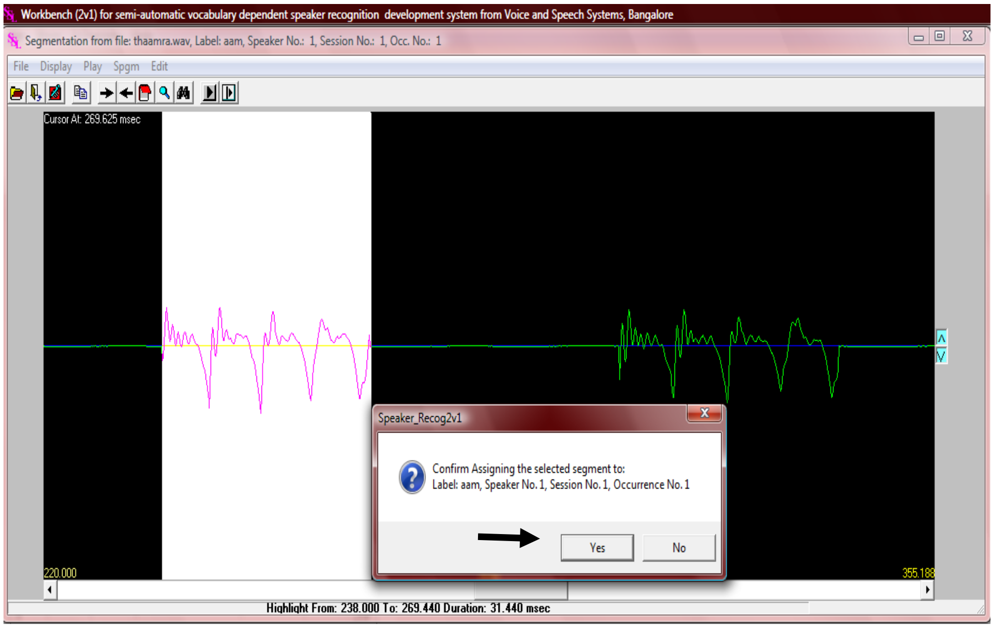


Figure 10: Dialogue box depicting confirmation of a highlighted segment.

Figure 11 and 12 shows segmentation information for all speakers were automatically stored in ***.dbs*** file (Figure 12) once the segmentation was saved by choosing the option 'File'-'Save segmentation' (Figure 11). Segmentation was not done in a single run. It was done in different runs because the previously saved segmentation information was used using the NOTEPAD ***(.txt file)***.

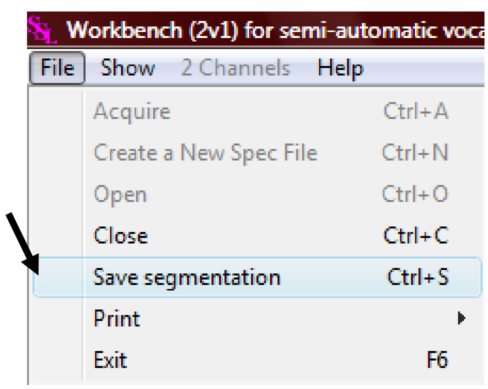


Figure 11: 'Save segmentation' option of SSL Workbench version 2.1

aam.dbs - Notepad

File Edit Format View Help

Speaker No	Occ. No.	Sess. No.	FileName	From	To
1	1	1	H:\Arjun-PROJECT\Live Recording - Cropped Files\SP1- Anand Kumar 8kHz\aam\baadaami.wav	0.067	0.103
1	2	1	H:\Arjun-PROJECT\Live Recording - Cropped Files\SP1- Anand Kumar 8kHz\aam\graama.wav	0.070	0.107
1	3	1	H:\Arjun-PROJECT\Live Recording - Cropped Files\SP1- Anand Kumar 8kHz\aam\raamanige.wav	0.062	0.095
1	4	1	H:\Arjun-PROJECT\Live Recording - Cropped Files\SP1- Anand Kumar 8kHz\aam\saamaanya.wav	0.072	0.103
1	5	1	H:\Arjun-PROJECT\Live Recording - Cropped Files\SP1- Anand Kumar 8kHz\aam\thaamra.wav	0.238	0.269
1	1	2	H:\Arjun-PROJECT\Live Recording - Cropped Files\SP1- Anand Kumar 8kHz\aam\baadaami.wav	0.141	0.176
1	2	2	H:\Arjun-PROJECT\Live Recording - Cropped Files\SP1- Anand Kumar 8kHz\aam\graama.wav	0.142	0.179
1	3	2	H:\Arjun-PROJECT\Live Recording - Cropped Files\SP1- Anand Kumar 8kHz\aam\raamanige.wav	0.138	0.171
1	4	2	H:\Arjun-PROJECT\Live Recording - Cropped Files\SP1- Anand Kumar 8kHz\aam\saamaanya.wav	0.141	0.173
1	5	2	H:\Arjun-PROJECT\Live Recording - Cropped Files\SP1- Anand Kumar 8kHz\aam\thaamra.wav	0.308	0.341
2	1	1	H:\Arjun-PROJECT\Live Recording - Cropped Files\SP2- Shivaswamy 8kHz\aam\baadaami.wav	0.093	0.118
2	2	1	H:\Arjun-PROJECT\Live Recording - Cropped Files\SP2- Shivaswamy 8kHz\aam\graama.wav	0.071	0.100
2	3	1	H:\Arjun-PROJECT\Live Recording - Cropped Files\SP2- Shivaswamy 8kHz\aam\raamanige.wav	0.076	0.102
2	4	1	H:\Arjun-PROJECT\Live Recording - Cropped Files\SP2- Shivaswamy 8kHz\aam\saamaanya.wav	0.066	0.093
2	5	1	H:\Arjun-PROJECT\Live Recording - Cropped Files\SP2- Shivaswamy 8kHz\aam\thaamra.wav	0.066	0.091
2	1	2	H:\Arjun-PROJECT\Live Recording - Cropped Files\SP2- Shivaswamy 8kHz\aam\baadaami.wav	0.152	0.178
2	2	2	H:\Arjun-PROJECT\Live Recording - Cropped Files\SP2- Shivaswamy 8kHz\aam\graama.wav	0.124	0.154
2	3	2	H:\Arjun-PROJECT\Live Recording - Cropped Files\SP2- Shivaswamy 8kHz\aam\raamanige.wav	0.116	0.143

Figure 12: The ***.dbs*** file with particulars of segmentation for each speaker.

The details of the above figure 12 are as follows:

- ***Speaker number***: Represents the number of speakers selected (1,2,3,4,...20)
- ***Occurrence number:*** Five occurrences for each session
- ***Session number:*** Two sessions were used in the present study
- ***File name:*** Location or path of the file matching to the speech segment of the speaker chosen
- ***From and To:*** The beginning and ending duration of the desired segment.

The training frame appeared once the segmentation for all the twenty speakers was done. The software needs a set of training samples, representing a speaker in order to identify the speaker correctly when provided with a test sample. Here, the number of training samples was assigned as seven and testing samples were automatically assigned as three (Figure 13). The software at random assigns the occurrences as reference and test (For example, Reference: 1,4,5,6,8,9,10 and Test: 3,5,7). In this study, 30 random combinations of 7 references and 3 tests samples were chosen. These combinations was varied at random by clicking on the 'Randomize Training Samples' button in the training frame. The system has choice of selecting a number of feature vectors. MFCC (Mel frequency cepstral coefficients are discrete cosine transform coefficients of MF Log spectral coefficients) with 13 coefficients was selected as the desired feature vector out of nine different features available. Then the option 'Compute' was selected which checked and compared the samples grossly and a qualitative analysis of each speaker was given. The option 'Testing' was selected soon after that. Whenever the number of training samples changes or a new randomization was done, then the features was recomputed by selecting the option ‘Compute’ followed by ‘Testing’.

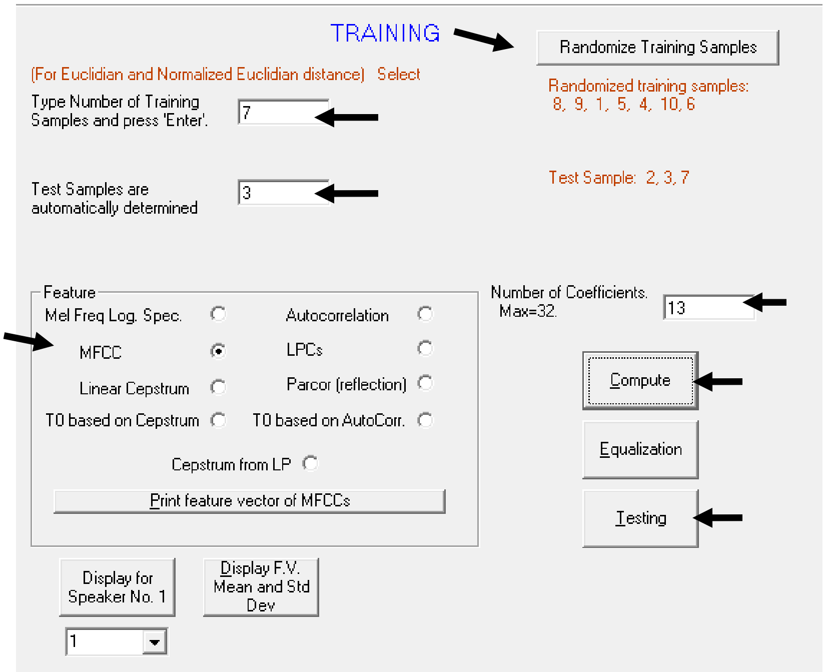


Figure 13: Training frame of SSL Workbench version 2.1

The testing frame was shown (Figure 14) where the distance metric was selected as ‘Euclidian’ followed by the option ‘Compute Score for Identification’. Figure 14 show the reference sample was taken along the row and the test sample was taken along the column. The Euclidian distance of the reference samples and test samples for each speaker were averaged and were obtained separately by the SSL Workbench software. Now a distance matrix as well as the score was shown for all 20 speakers i.e., test speaker versus reference speaker. Euclidian distance (ED) is an ordinary distance between two points and is a measure of similarity or dissimilarity. Euclidian distance within and between participants was saved by choosing command button ‘Print’ and this prepared a file of the distance matrix (Figure 15). If the ED distance between the test sample and corresponding reference sample is least, then the identification was considered as ***correct identification/same speaker.*** Anything above the least distance was considered as a ***different speaker.*** The percent correct identification was calculated using the following formula:

$$\text{Percent correct identification} = \frac{\text{Number of correct identification}}{\text{Number of total possible identications}} \times 100$$

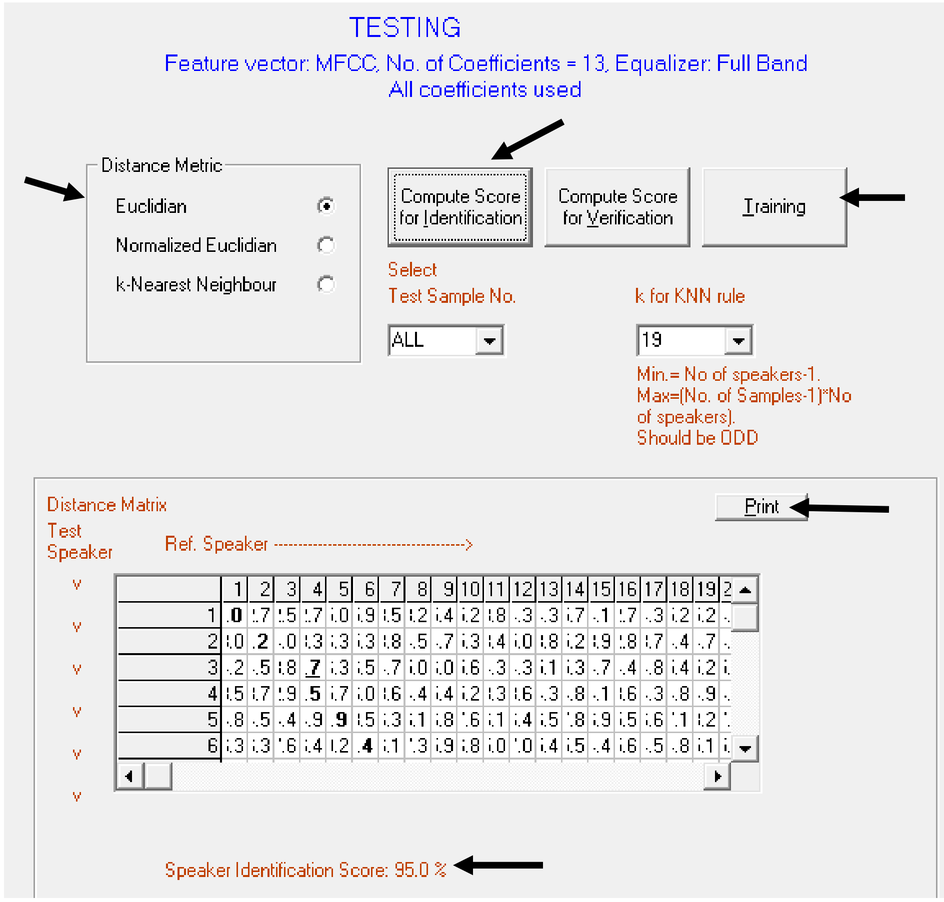
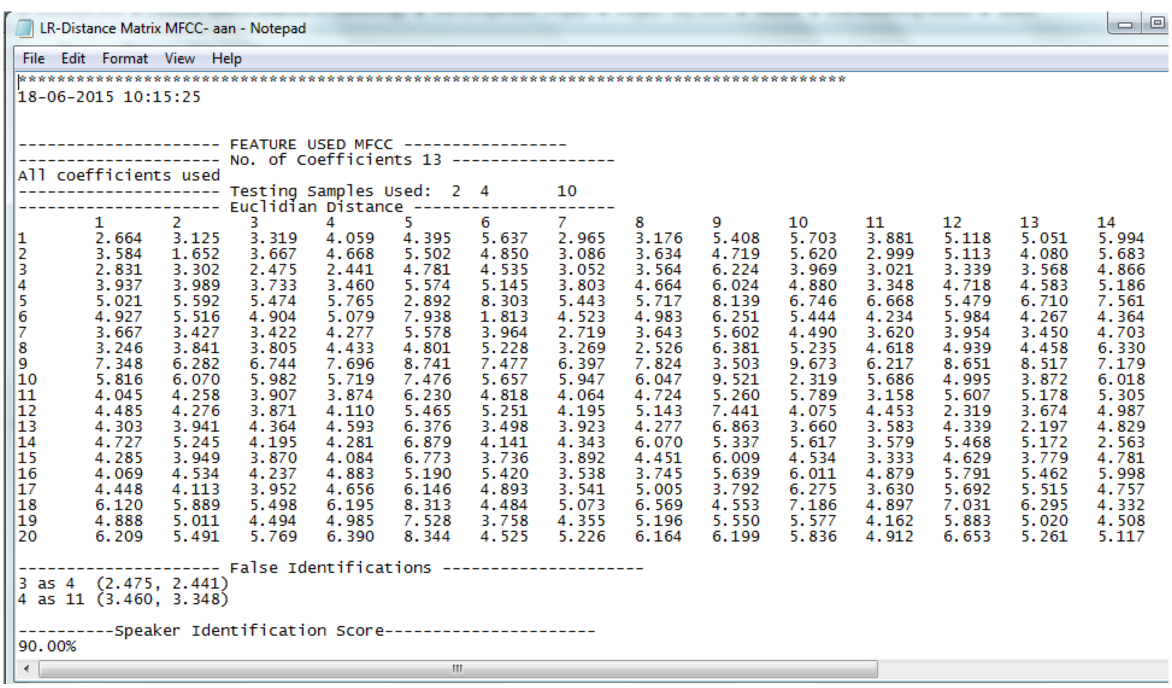


Figure 14: Testing frame depicting speaker identification score.



LR-Distance Matrix MFCC- aan - Notepad

File Edit Format View Help

```
*************************************************************************************
18-06-2015 10:15:25


--------------------- FEATURE USED MFCC -----------------
--------------------- No. of Coefficients 13 -----------------
All coefficients used
--------------------- Testing Samples Used:  2  4      10
--------------------- Euclidian Distance --------------------
```

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	2.664	3.125	3.319	4.059	4.395	5.637	2.965	3.176	5.408	5.703	3.881	5.118	5.051	5.994
2	3.584	1.652	3.667	4.668	5.502	4.850	3.086	3.634	4.719	5.620	2.999	5.113	4.080	5.683
3	2.831	3.302	2.475	2.441	4.781	4.535	3.052	3.564	6.224	3.969	3.021	3.339	3.568	4.866
4	3.937	3.989	3.733	3.460	5.574	5.145	3.803	4.664	6.024	4.880	3.348	4.718	4.583	5.186
5	5.021	5.592	5.474	5.765	2.892	8.303	5.443	5.717	8.139	6.746	6.668	5.479	6.710	7.561
6	4.927	5.516	4.904	5.079	7.938	1.813	4.523	4.983	6.251	5.444	4.234	5.984	4.267	4.364
7	3.667	3.427	3.422	4.277	5.578	3.964	2.719	3.643	5.602	4.490	3.620	3.954	3.450	4.703
8	3.246	3.841	3.805	4.433	4.801	5.228	3.269	2.526	6.381	5.235	4.618	4.939	4.458	6.330
9	7.348	6.282	6.744	7.696	8.741	7.477	6.397	7.824	3.503	9.673	6.217	8.651	8.517	7.179
10	5.816	6.070	5.982	5.719	7.476	5.657	5.947	6.047	9.521	2.319	5.686	4.995	3.872	6.018
11	4.045	4.258	3.907	3.874	6.230	4.818	4.064	4.724	5.260	5.789	3.158	5.607	5.178	5.305
12	4.485	4.276	3.871	4.110	5.465	5.251	4.195	5.143	7.441	4.075	4.453	2.319	3.674	4.987
13	4.303	3.941	4.364	4.593	6.376	3.498	3.923	4.277	6.863	3.660	3.583	4.339	2.197	4.829
14	4.727	5.245	4.195	4.281	6.879	4.141	4.343	6.070	5.337	5.617	3.579	5.468	5.172	2.563
15	4.285	3.949	3.870	4.084	6.773	3.736	3.892	4.451	6.009	4.534	3.333	4.629	3.779	4.781
16	4.069	4.534	4.237	4.883	5.190	5.420	3.538	3.745	5.639	6.011	4.879	5.791	5.462	5.998
17	4.448	4.113	3.952	4.656	6.146	4.893	3.541	5.005	3.792	6.275	3.630	5.692	5.515	4.757
18	6.120	5.889	5.498	6.195	8.313	4.484	5.073	6.569	4.553	7.186	4.897	7.031	6.295	4.332
19	4.888	5.011	4.494	4.985	7.528	3.758	4.355	5.196	5.550	5.577	4.162	5.883	5.020	4.508
20	6.209	5.491	5.769	6.390	8.344	4.525	5.226	6.164	6.199	5.836	4.912	6.653	5.261	5.117

```
--------------------- False Identifications ---------------------
3 as 4  (2.475, 2.441)
4 as 11 (3.460, 3.348)

----------Speaker Identification Score----------------------
90.00%
```

Figure 15: File of Euclidian distance matrix.

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. The variables measured are the effects of different vowels preceding nasal continuants on speaker identification. 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/)* on the following ways:

- Live recording (test) was compared with live recording (reference)
- Mobile network recording (test) was compared with mobile network recording (reference)
- Mobile network recording (test) was compared with live recording (reference)

CHAPTER IV

RESULTS

The aim of the present study was to establish ***"Benchmark for speaker identification in Kannada using MFCCs derived from vowels preceding nasals".*** Results of the study are explained under two sections – Section A and Section B with reference to the following three conditions:

I. Speaker identification scores using MFCCs on vowels preceding nasal continuants for live recording.
II. Speaker identification scores using MFCCs on vowels preceding nasal continuants for mobile network recording.
III. Speaker identification scores using MFCCs on vowels preceding nasal continuants for mobile network recording compared with live recording.

4.1. Section A

Group I consisted of twenty speakers for speaker identification. The results are discussed with reference to the following three conditions.

4.1.1. 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 was obtained for each vowel preceding nasal continuant. *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*. Table D depicts the speaker identification scores obtained for all thirty randomized trials for the vowels preceding nasals, along with the test sample combinations. Results showed an average correct identification score of 93%, 79%, 80%, 84% and 84% for */a:/, /i:/, /u:/, /m/ and /n/,* respectively. Table C depicts average percentage of correct identification across vowels and nasals. One sample of a distance matrix for each vowel preceding nasal continuant from Table D is shown in Appendix D. Tables 1-6 in Appendix D depicts the Euclidian distance when live recording was compared with live recording and the green color in the tables indicates the correct identification of speaker sample as belonging to the same speaker as the reference sample whereas the red indicates the wrong identification of test sample as belonging to a different reference speaker.

Table C: *Average (AVG) percentage of correct identification across vowels and nasals for twenty speakers in condition I*

Live Vs Live			
Condition I - 20 Speakers			
	/m/	***/n/***	**AVG**
/a:/	92	93	**93**
/i:/	80	78	**79**
/u:/	80	80	**80**
AVG	**84**	**84**	

Table D: *Average (AVG) and standard deviation (STD) of speaker identification percentage along with test samples for twenty speakers in condition I.*

Live Vs Live							
Trial	**Test Sample**	**Percentage**					
		/a:m/	***/i:m/***	***/u:m/***	***/a:n/***	***/ i:n/***	***/u:n/***
1	2,3,7	75	60	65	65	65	70
2	2,4,10	95	90	90	90	90	80
3	4,5,9	80	80	65	85	70	70
4	5,7,8	100	95	85	90	90	75
5	3,9,10	85	80	80	100	85	80
6	2,6,8	95	90	80	95	85	80
7	2,3,4	95	90	75	100	85	80
8	7,8,9	95	80	80	100	65	65
9	1,8,9	100	80	85	95	90	95
10	3,6,10	100	80	90	100	85	75
11	3,8,10	80	80	80	95	60	75
12	3,7,9	90	80	80	95	70	80
13	7,8,9	95	80	80	100	65	65
14	2,5,6	90	85	80	95	95	80
15	7,8,9	95	80	80	100	65	65
16	2,3,9	90	80	80	100	70	70
17	3,8,9	75	65	75	85	65	75
18	1,2,3	95	95	85	95	80	90
19	1,4,10	95	85	80	100	85	75
20	1,8,9	100	80	85	95	90	95
21	1,3,9	95	75	85	95	80	90
22	3,6,7	90	75	75	85	75	90
23	3,7,9	90	80	80	95	70	80
24	2,6,9	95	80	85	100	75	85
25	2,6,7	85	55	70	75	75	80
26	3,8,9	75	65	75	85	65	75
27	1,8,9	100	80	85	95	90	95
28	3,7,10	95	80	75	95	80	85
29	2,3,6	95	90	85	90	80	90
30	3,5,7	100	90	90	90	95	75
Average		**92**	**80**	**80**	**93**	**78**	**80**
STD		**7.8**	**9.3**	**6.4**	**8.1**	**10.5**	**8.8**

4.1.2. 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 was obtained for each vowel preceding nasal continuant. 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. Table F depicts the speaker identification scores obtained for all thirty randomized trials for the vowels preceding nasals, along with the test sample combinations. Results showed an average correct identification score of 74%, 54%, 52%, 61% and 58% for */a:/, /i:/, /u:/, /m/ and /n/,* respectively. Table E depicts average percentage of correct identification across vowels and nasals. One sample of a distance matrix for each vowel preceding nasal continuant from Table F is shown in Appendix D. Tables 7-12 in Appendix D depict the Euclidian distance when mobile network recording was compared with mobile network recording and the green color in the tables indicates the correct identification of speaker sample as belonging to the same speaker as the reference sample whereas the red color indicates the wrong identification of test sample as belonging to a different reference speaker.

Table E: *Average (AVG) percentage of correct identification across vowels and nasals for twenty speakers in condition II*

Mobile Network Vs Mobile Network			
Condition II - 20 Speakers			
	/m/	***/n/***	**AVG**
/a:/	75	72	**74**
/i:/	58	49	**54**
/u:/	51	53	**52**
AVG	**61**	**58**	

Table F: *Average (AVG) and standard deviation (STD) of speaker identification percentage along with test samples for twenty speakers in condition II.*

Mobile Network Vs Mobile Network							
Trial	**Test Sample**	**Percentage**					
		/a:m/	***/i:m/***	***/u:m/***	***/a:n/***	***/i:n/***	***/u:n/***
1	2,3,7	70	35	50	40	30	30
2	2,4,10	80	70	65	85	60	70
3	4,5,9	50	45	35	75	15	40
4	5,7,8	65	60	55	75	45	65
5	3,9,10	55	70	50	80	55	75
6	2,6,8	70	45	60	80	50	55
7	2,3,4	70	50	50	75	65	60
8	7,8,9	80	60	40	65	55	35
9	1,8,9	85	80	60	85	55	65
10	3,6,10	80	60	60	75	55	65
11	3,8,10	70	45	40	50	20	40
12	3,7,9	70	55	45	70	45	50
13	7,8,9	80	60	40	65	55	35
14	2,5,6	80	65	45	85	55	65
15	7,8,9	80	60	40	65	55	35
16	2,3,9	75	50	60	80	70	35
17	3,8,9	70	45	35	55	40	45
18	1,2,3	75	45	55	75	50	60
19	1,4,10	90	75	60	65	60	40
20	1,8,9	85	80	60	85	55	65
21	1,3,9	80	80	70	90	50	55
22	3,6,7	85	50	65	55	50	55
23	3,7,9	70	55	45	70	45	50
24	2,6,9	85	80	45	85	50	65
25	2,6,7	50	45	35	55	40	40
26	3,8,9	70	45	35	55	40	45
27	1,8,9	85	80	60	85	55	65
28	3,7,10	75	60	65	75	55	50
29	2,3,6	80	40	70	85	45	65
30	3,5,7	90	60	45	70	50	60
Average		**75**	**58**	**51**	**72**	**49**	**53**
STD		**10.3**	**13.5**	**11.1**	**12.6**	**11.7**	**12.8**

4.1.3. 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 was obtained for each vowel preceding nasal continuant. 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. Table H depicts the speaker identification scores obtained for all thirty randomized trials for the vowels preceding nasals, along with the test sample combinations. The obtained percent correct identification scores were lesser than the chance factor (<50%). Results showed an average correct identification score of 39%, 36%, 37%, 36% and 38% for */a:/, /i:/, /u:/, /m/ and /n/*, respectively. Table G depicts average percentage of correct identification across vowels and nasals. One sample of a distance matrix for each vowel preceding nasal continuant from Table H is shown in Appendix D. Tables 13-18 in Appendix D depict the Euclidian distance when mobile network recording was compared with live recording and the green color in the tables indicates the correct identification of speaker sample as belonging to the same speaker as the reference sample whereas the red color indicates the wrong identification of test sample as belonging to a different reference speaker.

Table G: *Average (AVG) percentage of correct identification across vowels and nasals for twenty speakers in condition III*

Mobile Network Vs Live			
Condition III - 20 Speakers			
	/m/	***/n/***	**AVG**
/a:/	38	39	**39**
/i:/	36	36	**36**
/u:/	34	39	**37**
AVG	**36**	**38**	

Table H: *Average (AVG) and standard deviation (STD) of speaker identification percentage along with test samples for twenty speakers in condition III.*

Mobile Network Vs Live							
Trial	**Test Sample**	**Percentage**					
		/a:m/	***/i:m/***	***/u:m/***	***/a:n/***	***/i:n/***	***/u:n/***
1	2,3,7	40	55	45	50	25	40
2	2,4,10	35	40	30	25	35	40
3	4,5,9	35	55	30	40	30	35
4	5,7,8	50	30	40	35	40	35
5	3,9,10	45	45	45	70	45	50
6	2,6,8	50	25	45	30	25	35
7	2,3,4	15	5	10	5	10	10
8	7,8,9	25	15	20	20	30	25
9	1,8,9	50	20	30	55	55	40
10	3,6,10	45	50	45	50	35	40
11	3,8,10	35	40	45	55	50	55
12	3,7,9	45	60	45	30	45	55
13	7,8,9	25	15	20	20	30	25
14	2,5,6	30	45	40	25	25	35
15	7,8,9	25	15	20	20	30	25
16	2,3,9	30	30	35	60	20	40
17	3,8,9	40	45	25	45	50	50
18	1,2,3	15	5	15	5	15	15
19	1,4,10	40	55	35	35	55	35
20	1,8,9	50	20	30	55	55	40
21	1,3,9	25	45	40	50	45	30
22	3,6,7	40	50	25	35	35	45
23	3,7,9	45	60	45	30	45	55
24	2,6,9	45	25	30	60	45	50
25	2,6,7	35	25	35	40	25	45
26	3,8,9	40	45	25	45	50	50
27	1,8,9	50	20	30	55	55	40
28	3,7,10	50	55	35	60	40	35
29	2,3,6	35	30	50	35	10	45
30	3,5,7	40	50	40	40	35	40
Average		**38**	**36**	**34**	**39**	**36**	**39**
STD		**10.2**	**16.7**	**10.4**	**16.4**	**13.3**	**11.1**

Table I: *Average (AVG) and Standard deviation (STD) of speaker identification percentage for condition I, II and III when twenty speakers were considered*

Speaker Identification Percentage for 20 Speakers								
		/m/			***/n/***		
		/a:/	***/i:/***	***/u:/***	***/a:/***	***/i:/***	***/u:/***
Live Vs Live (Condition I)	**AVG**	92	80	80	93	78	80
	STD	7.8	9.3	6.4	8.1	10.5	8.8
Mobile Network Vs Mobile Network (Condition II)	**AVG**	75	58	51	72	49	53
	STD	10.3	13.5	11.1	12.6	11.7	12.8
Mobile Network Vs Live (Condition III)	**AVG**	38	36	34	39	36	39
	STD	10.2	16.7	10.4	16.4	13.3	11.1

Table I shows the average and standard deviation of percent correct identification scores across three conditions, three vowels and two nasal continuants. To summarize the results in section A;

- The percent correct identification scores were 93% for */a:/;* 80% for */u:/* and 79% for the vowel */i:/* (Table C).
- The vowel */a:/* had the highest identification score followed by */u:/* and then */i:/*.
- The percent correct identification scores for vowels preceding nasals */m/ and /n/* resulted in similar results of 84% in live vs live comparison condition (Table C).
- The percent correct identification was the highest for vowel */a:/* followed by */i:/* and then */u:/* in condition II (Table E).

- The identification scores were better for vowels preceding nasal consonant */m/* when compared to nasal */n/* which was just above chance factor (58%) (Table E).
- The percent correct identification scores was the highest for vowel */a:/* followed by vowel */u:/* and then vowel */i:/* in condition III (Table G).
- The identification scores were better for vowels preceding nasal continuant */n/* when compared to nasal */m/*.
- In general, condition III had identification scores much lesser than the chance factors for the three vowels preceding nasal continuants.

4.2 Section B

Group II consisted of ten speakers for speaker identification in order to check the identification accuracy with less number of subjects. The results are discussed with reference to the following three conditions.

4.2.1. 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 was obtained for each vowel preceding nasal continuant. 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. The percent correct identification scores were better and higher when the number of speakers reduced from 20 to 10 compared to condition 1 of section A. Table K depicts the speaker identification scores obtained for all thirty randomized trials for the vowels preceding nasals, along with the test sample combinations. Results showed an average correct identification score of 94%, 83%, 88%, 88% and 88% for */a:/, /i:/, /u:/, /m/ and /n/*, respectively. Table J depicts average percentage of correct identification across vowels and nasals. One sample of a distance matrix for each vowel preceding nasal continuant from Table K is shown in Appendix D. Tables 19-24 in Appendix D depict the Euclidian distance when live recording was compared with live recording and the green color in the tables indicates the correct identification of speaker sample as belonging to the same speaker as the reference sample whereas the red color indicates the wrong identification of test sample as belonging to a different reference speaker.

Table J: *Average (AVG) percentage of correct identification across vowels and nasals for ten speakers in condition I*

Live Vs Live			
Condition I - 10 Speakers			
	/m/	***/n/***	**AVG**
/a:/	92	95	**94**
/i:/	85	81	**83**
/u:/	86	89	**88**
AVG	**88**	**88**	

Table K: *Average (AVG) and standard deviation (STD) of speaker identification percentage along with test samples for ten speakers in condition I*

Live Vs Live							
Trial	**Test Sample**	**Percentage**					
		/a:m/	***/i:m/***	***/u:m/***	***/a:n/***	***/i:n/***	***/u:n/***
1	2,3,7	80	60	70	80	60	70
2	2,4,10	90	90	80	90	100	80
3	4,5,9	80	90	80	100	70	90
4	5,7,8	100	100	90	90	80	100
5	3,9,10	90	80	80	100	80	90
6	2,6,8	100	100	80	100	90	90
7	2,3,4	90	90	70	100	90	90
8	7,8,9	100	80	90	100	90	80
9	1,8,9	100	90	90	100	90	100
10	3,6,10	100	80	100	100	90	90
11	3,8,10	70	90	100	90	60	80
12	3,7,9	90	90	100	90	80	90
13	7,8,9	100	80	90	100	90	80
14	2,5,6	100	80	80	100	90	90
15	7,8,9	100	80	90	100	90	80
16	2,3,9	90	100	80	100	70	80
17	3,8,9	100	80	90	90	70	90
18	1,2,3	90	100	80	100	80	100
19	1,4,10	90	80	80	100	100	70
20	1,8,9	100	90	90	100	90	100
21	1,3,9	90	80	90	100	90	100
22	3,6,7	100	90	70	80	70	100
23	3,7,9	90	90	100	90	80	90
24	2,6,9	90	70	90	100	70	90
25	2,6,7	90	60	70	80	60	70
26	3,8,9	70	80	90	90	70	90
27	1,8,9	100	90	90	100	90	100
28	3,7,10	90	80	90	100	70	100
29	2,3,6	90	100	90	90	70	100
30	3,5,7	100	90	90	90	90	90
Average		**92**	**85**	**86**	**95**	**81**	**89**
STD		**8.6**	**10.4**	**8.9**	**6.8**	**11.7**	**9.6**

4.2.2. 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 was obtained for each vowel preceding nasal continuant. 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. Table M depicts the speaker identification scores obtained for all thirty randomized trials for the vowels preceding nasals, along with the test sample combinations. Results showed an average correct identification score of 83%, 63%, 65%, 69% and 71% for */a:/, /i:/, /u:/, /m/ and /n/,* respectively. Table L depicts average percentage of correct identification across vowels and nasals. The identification scores were better in section B, condition I when compared to section A, condition I. The machine matches the test sample with the reference sample more accurately when the number of speakers reduced in the comparison data. One sample of a distance matrix for each vowel preceding nasal continuant from Table M is shown in Appendix D. Tables 25-30 in Appendix D depict the Euclidean distance when mobile network recording was compared with mobile network recording and the green color in the tables indicates the correct identification of speaker sample as belonging to the same speaker as the reference sample whereas the red color indicates the wrong identification of test sample as belonging to a different reference speaker.

Table L: *Average (AVG) percentage of correct identification across vowels and nasals for ten speakers in condition II*

Mobile Network Vs Mobile Network			
Condition II - 10 Speakers			
	/m/	***/n/***	**AVG**
/a:/	80	85	**83**
/i:/	68	58	**63**
/u:/	60	69	**65**
AVG	**69**	**71**	

Table M: *Average (AVG) and standard deviation (STD) of speaker identification percentage along with test samples for ten speakers in condition II*

Mobile Network Vs Mobile Network							
Trial	**Test Sample**	**Percentage**					
		/a:m/	***/i:m/***	***/u:m/***	***/a:n/***	***/ i:n/***	***/u:n/***
1	2,3,7	90	40	50	90	30	80
2	2,4,10	100	70	70	90	50	50
3	4,5,9	50	40	50	90	20	70
4	5,7,8	70	60	80	80	60	60
5	3,9,10	70	90	70	100	70	60
6	2,6,8	80	50	70	80	80	70
7	2,3,4	80	60	60	80	70	90
8	7,8,9	80	70	70	80	70	60
9	1,8,9	70	90	60	10	80	70
10	3,6,10	90	70	60	90	70	80
11	3,8,10	80	70	60	100	40	50
12	3,7,9	70	70	80	80	40	80
13	7,8,9	80	70	70	80	70	60
14	2,5,6	90	60	40	90	50	80
15	7,8,9	80	70	70	80	70	60
16	2,3,9	80	80	60	90	60	60
17	3,8,9	80	60	50	90	50	60
18	1,2,3	80	60	60	80	60	80
19	1,4,10	90	80	70	90	50	60
20	1,8,9	70	90	60	100	80	70
21	1,3,9	70	80	70	100	50	60
22	3,6,7	100	50	80	90	60	80
23	3,7,9	70	70	8	80	40	80
24	2,6,9	100	80	60	90	60	80
25	2,6,7	70	50	40	70	50	70
26	3,8,9	80	60	50	90	50	60
27	1,8,9	70	90	60	100	80	70
28	3,7,10	80	80	70	80	70	80
29	2,3,6	90	50	60	90	50	80
30	3,5,7	80	70	50	80	50	70
Average		**80**	**68**	**60**	**85**	**58**	**69**
STD		**11.0**	**14.3**	**14.5**	**16.1**	**15.2**	**10.5**

4.2.3 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 was obtained for each vowel preceding nasal continuant. 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. Table O depicts the speaker identification scores obtained for all thirty randomized trials for the vowels preceding nasals, along with the test sample combinations. Results showed an average correct identification score of 49%, 52%, 48%, 49% and 50% for */a:/, /i:/, /u:/, /m/ and /n/*, respectively. Table N depicts average percentage of correct identification across vowels and nasals. The percent correct identification scores were above the chance factor except for the vowels */a:/ and /u:/* preceding nasal continuants */m/ and /n/* respectively. One sample of a distance matrix for each vowel preceding nasal continuant from Table O is shown in Appendix D. Tables 31-36 in Appendix D depict the Euclidian distance when mobile network recording was compared with live recording and the green color in the tables indicates the correct identification of speaker sample as belonging to the same speaker as the reference sample whereas the red color the wrong identification of test sample as belonging to a different reference speaker.

Table N: *Average (AVG) percentage of correct identification across vowels and nasals for ten speakers in condition III*

Mobile Network Vs Live			
Condition III - 10 Speakers			
	/m/	***/n/***	**AVG**
/a:/	47	50	**49**
/i:/	51	53	**52**
/u:/	50	46	**48**
AVG	**49**	**50**	

Table O: *Average (AVG) and standard deviation (STD) of speaker identification percentage along with test samples for ten speakers in condition III*

Mobile Network Vs Live							
Trial	**Test Sample**	**Percentage**					
		/a:m/	***/i:m/***	***/u:m/***	***/a:n/***	***/i:n/***	***/u:n/***
1	2,3,7	60	70	70	70	50	60
2	2,4,10	40	50	50	30	60	30
3	4,5,9	40	80	40	50	30	50
4	5,7,8	70	60	60	40	60	40
5	3,9,10	30	70	60	80	60	50
6	2,6,8	90	40	40	60	40	40
7	2,3,4	30	5	20	30	20	30
8	7,8,9	10	20	30	10	50	30
9	1,8,9	50	50	40	60	80	40
10	3,6,10	40	70	70	80	40	60
11	3,8,10	40	70	80	60	60	80
12	3,7,9	50	70	70	40	40	50
13	7,8,9	10	20	30	10	50	30
14	2,5,6	60	40	60	40	30	30
15	7,8,9	10	20	30	10	50	30
16	2,3,9	60	40	50	80	40	40
17	3,8,9	30	60	40	60	70	60
18	1,2,3	30	5	20	30	20	20
19	1,4,10	60	70	60	30	70	30
20	1,8,9	50	50	40	60	80	40
21	1,3,9	80	50	60	70	80	40
22	3,6,7	60	70	60	40	60	60
23	3,7,9	50	70	70	40	40	50
24	2,6,9	50	50	60	70	70	70
25	2,6,7	40	30	40	50	50	50
26	3,8,9	30	60	40	60	70	60
27	1,8,9	50	50	40	60	80	40
28	3,7,10	60	80	60	80	80	60
29	2,3,6	60	40	70	60	20	40
30	3,5,7	80	60	50	50	50	60
Average		**47**	**51**	**50**	**50**	**53**	**46**
STD		**20.0**	**21.1**	**16.1**	**20.8**	**18.8**	**14.3**

Table P: *Average (AVG) and Standard deviation (STD) of speaker identification percentage for condition I, II and III when ten speakers were considered*

Speaker Identification Percentage for 10 Speakers							
		/m/			***/n/***		
		/a:/	***/i:/***	***/u:/***	***/a:/***	***/i:/***	***/u:/***
Live Vs Live	**AVG**	92	85	86	95	81	89
	STD	8.6	10.4	8.9	6.8	11.7	9.6
Mobile Network Vs Mobile Network	**AVG**	80	68	60	85	58	69
	STD	11.0	14.3	14.5	16.1	15.2	10.5
Mobile Network Vs Live	**AVG**	47	51	50	50	53	46
	STD	20.0	21.1	16.1	20.8	18.8	14.3

Table P shows the average and standard deviation of percent correct identification scores across three conditions, three vowels and two nasal continuants. To summarize the results in section B;

- The percent correct identification scores were 94% for */a:/;* 88% for */u:/* and 83% for the vowel */i:/* (Table J).
- The vowel */a:/* had the highest identification score followed by */u:/* and then */i:/*.
- The percent correct identification scores for vowels preceding nasals */m/ and /n/* resulted in similar results of 84% in live vs live comparison condition (Table J).
- The percent correct identification was the highest for vowel */a:/* followed by */u:/* and then */i:/* in condition II (Table L)

- The identification scores were better for vowels preceding nasal consonant */m/* when compared to nasal */n/* (Table L).
- The percent correct identification scores was the highest for vowel */i:/* followed by vowel */a:/* and then vowel */u:/* in condition III (Table N).
- The identification scores were better for vowels preceding nasal continuant */n/* when compared to nasal */m/* (Table N).
- In general, condition III had identification scores much lesser than the chance factors for the three vowels preceding nasal continuants.

Table Q: *Speaker identification percentage with different number of participants*

		20 Speakers			**10 Speakers**		
		I	**II**	**III**	**I**	**II**	**III**
Vowels	***/a:/***	93	74	39	94	83	49
	/i:/	79	54	36	83	63	52
	/u:/	80	52	37	88	65	48
Nasals	***/m/***	84	61	36	88	69	49
	/n/	84	58	38	88	71	50

Note: I –Live vs live recording, II – Mobile network vs mobile network recording, III – Mobile network vs live recording.

Table Q shows the grand average correct percentage of speaker identification scores across three conditions, three vowels and two nasal continuants when number of participants was twenty and ten.

CHAPTER V

DISCUSSION

The present study aimed at establishing a benchmark for speaker identification using Mel frequency cepstral coefficients extracted from vowels preceding nasal continuants of Kannada language in live recording, mobile network recording and a comparison between them.

The results obtained from this study revealed several interesting points of interest;

First, the percentage of correct speaker identification scores for twenty speakers ranged from 78% to 93% when live recording was compared with live recording for the vowels preceding nasal continuants */a:m/, /i:m/, /u:m/, /a:n/, /i:n/ and /u:n/* and can be stated that these scores were drawn for contemporary speech samples (Table D). Here, the reference and the test speakers were derived from the live recordings. Speaker identification scores vary among the vowels preceding nasals and also when the samples were randomized (Table D). Average speaker identification percentage for */a:m/, /i:m/, /u:m/, /a:n/, /i:n/ and /u:n/* was *92%, 80%, 80%, 93%, 78% and 80%*, respectively. Standard deviation for */a:m/, /i:m/, /u:m/, /a:n/, /i:n/ and /u:n/* was 7.8, 9.3, 6.4, 8.1, 10.5 and 8.8, respectively (Table D). In this condition, performance of the vowel */a:/* preceding nasal continuant */n/* (*/a:n/*) was better than vowel */a:/* preceding nasal continuant */m/* (*/a:m/*). The performance of the vowel */i:/* preceding nasal continuant */m/* (*/i:m/*) was better than vowel */i:/* preceding nasal continuant */n/* (*/i:n/*). The performance of the vowel */u:/* preceding nasal continuants */m/* (*/u:m/*) *and /n/* (*/u:n/*) were one and the same. Across vowels and nasals, results showed an average correct identification score of 93%, 79%, 80%, 84% and 84% for */a:/, /i:/, /u:/, /m/ and /n/,* respectively. The overall performance of vowel */a:/* preceding nasal continuants */n/* and */m/* (*/a:n/ & /a:m/*) was better compared to the vowels */i:/ and /u:/* preceding nasal continuants */n/ and /m/* (*/i:n/, /i:m/, /u:n/ & /u:m/*). Vowel */a:/* with 93% has the highest correct identification scores than vowels */i:/ and /u:/.* Across nasals */m/ and /n/,* both obtained similar scores with 84% (Table Q).

This result is compatible with those of the other previous studies using MFCCs with Hasan et al (2004), Singh and Rajan (2009), Tiwari (2010) and Chandrika (2010) where an identification accuracy of 80% - 100% were obtained. Rajshekhar (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, (2010) used 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) conducted study on 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%, vowel */a:/* had 93.33% and vowel */i:/* had 93.33%. Patel and Prasad (2013) used MFCCs and reported 13% error rate for the word “hello”. Pickett (1980) says 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. Also Fledge (1988) obtained significant nasalance effect on vowel */i/* compared to */u/*.

Second, the percentage of speaker identification scores for twenty speakers ranged from 49% to 75% when mobile network recording was compared with mobile network recording for the vowels preceding nasal continuants */a:m/, /i:m/, /u:m/, /a:n/, /i:n/ and /u:n/* and can be stated that these scores were drawn for contemporary speech samples (Table F). Here, the reference and the test speakers were derived from the mobile network recordings. Speaker identification scores vary among the vowels preceding nasals and also when the samples were randomized (Table F). Average speaker identification percentage for */a:m/, /i:m/, /u:m/, /a:n/, /i:n/ and /u:n/* was 75%, 58%, 51%, 72%, 49%, and 53%, respectively. Standard deviation for */a:m/, /i:m/, /u:m/, /a:n/, /i:n/ and /u:n/* was 10.3, 13.5 11.1, 12.6, 11.7 and 12.8, respectively (Table F). In this condition, performance of the vowel */a:/* preceding nasal continuant */m/* (*/a:m/*) was better than vowel */a:/* preceding nasal continuant */n/* (*/a:n/*). The performance of the vowel */i:/* preceding nasal continuant */m/* (*/i:m/*) was better than vowel */i:/* preceding nasal continuant */n/* (*/i:n/*). The performance of the vowel */u:/* preceding nasal continuant */n/* (*/u:n/*) was better than the vowel */u:/* preceding nasal continuant */m/* (*/u:m/*). Across vowels and nasals, results showed an average correct identification score of 74%, 54%, 52%, 61% and 58% for */a:/, /i:/, /u:/, /m/ and /n/*

respectively. The overall performance of vowel */a:/* preceding nasal continuants */n/* and */m/* (*/a:n/ & /a:m/*) was better compared to the vowels */i:/* and */u:/* preceding nasal continuants */n/* and */m/* (*/i:n/, /i:m/, /u:n/ & /u:m/*). Vowel */a:/* with 74% has highest correct identification scores than vowels */i:/ and /u:/.* Across nasals, */m/* with 61% has highest correct identification scores (Table Q).

The results showed that the percentage of speaker identification for mobile network recording was drastically lowered compared to live recording. For this condition, speaker identification scores were not as good as scores obtained for live recording because of the recording characteristics of mobile network.

Third, the percentage of speaker identification scores for twenty speakers ranged from 34% to 39% when mobile network recording was compared with live recording for the vowels preceding nasal continuants and can be stated that these scores were drawn for non contemporary speech samples (Table H). Here, the reference and the test speakers were derived from the live and mobile network recordings, respectively. Speaker identification scores vary among the vowels preceding nasals and also when the samples were randomized (Table H). Average speaker identification percentage for */a:m/, /i:m/, /u:m/, /a:n/, /i:n/ and /u:n/* was 38%, 36%, 34%, 39%, 36% and 39%, respectively. Standard deviation for */a:m/, /i:m/, /u:m/, /a:n/, /i:n/ and /u:n/* was 10.2, 16.7, 10.4, 16.4, 13.3 and 11.1, respectively (Table H). In this condition, performance of the vowel */a:/* preceding nasal continuant */n/* (*/a:n/*) was better than vowel */a:/* preceding nasal continuant */m/* (*/a:m/*). The performance of the vowel */i:/* preceding nasal continuant */n/* (*/i:n/*) *and /m/* (*/i:m/*) was one and the same. The performance of the vowel */u:/* preceding nasal continuant */n/* (*/u:n/*) was better than the vowel */u:/* preceding nasal continuant */m/* (*/u:m/*). The overall performance of vowel */a:/* preceding nasal continuants */n/* and (*/a:n/*) and the vowel */u:/* preceding nasal continuant */n/* (*/a:n/*) was better compared to the vowel */i:/* preceding nasal continuants */n/* and */m/* (*/i:n/ & /i:m/*) and to the vowels */a:/* and */u:/* preceding nasal continuant */m/* (*/a:m/ and /u:m/*). Across vowels and nasals, results showed an average correct identification score of 39%, 36%, 37%, 36% and 38% for */a:/, /i:/, /u:/, /m/ and /n/* respectively. The obtained percent correct identification scores for three vowels and two nasals were lesser than the chance factor (<50%) (Table Q).

The results showed that the percentage of speaker identification for mobile network recording versus live recording was hugely lowered compared to live 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 was considered. 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.

Fourth, the percentage of speaker identification scores for ten speakers ranged from 85% to 92% when live recording was compared with live recording for the vowels preceding nasal continuants */a:m/, /i:m/ and /u:m/*. The percentage of speaker identification scores for ten speakers ranged from 81% to 95% when live recording was compared with live recording for the vowels preceding nasal continuants */a:n/, /i:n/ and /u:n/.* This can be stated that these scores were drawn for contemporary speech samples (Table K). Here, the reference and the test speakers were derived from the live recordings. Speaker identification scores vary among the vowels preceding nasals and also when the samples were randomized (Table K). Average speaker identification percentage for */a:m/, /i:m/, /u:m/, /a:n/, /i:n/ and /u:n/* was 92%, 85%, 86%, 95%, 81% and 89%, respectively. Standard deviation for */a:m/, /i:m/, /u:m/, /a:n/, /i:n/ and*

/u:n/ was 8.6, 10.4, 8.9, 6.8, 11.7 and 9.6, respectively (Table K). In this condition, performance of the vowel */a:/* preceding nasal continuant */n/* (*/a:n/*) was better than vowel */a:/* preceding nasal continuant */m/* (*/a:m/*). The performance of the vowel */i:/* preceding nasal continuant */m/* (*/i:m/*) was better than vowel */i:/* preceding nasal continuant */n/* (*/i:n/*). The performance of the vowel */u:/* preceding nasal continuant */n/* (*/u:n/*) was better than the vowel */u:/* preceding nasal continuant */m/* (*/u:m/*). Among the vowels, the overall performance of vowel */a:/* preceding nasal continuants */n/* and */m/* (*/a:n/ & /a:m/*) was better compared to the vowels */i:/* and */u:/* preceding nasal continuants /*n/ and /m/* (*/i:n/, /i:m/, /u:n/ & /u:m/*). Across vowels and nasals, results showed an average correct identification score of 94%, 83%, 88%, 88% and 88% for */a:/, /i:/, /u:/, /m/ and /n/,* respectively. Hence, vowel */a:/* with 94% and both the nasals */m/ and /n/* with 88% has the highest correct identification scores (Table Q).

Fifth, the percentage of speaker identification scores for ten speakers ranged from 60% to 80% when mobile network recording was compared with mobile network recording for the vowels preceding nasal continuants */a:m/, /i:m/ and /u:m/.* The percentage of speaker identification scores for ten speakers ranged from 58% to 85% when mobile network recording was compared with mobile network recording for the vowels preceding nasal continuant */a:n/, /i:n/ and /u:n/.* It can be stated that these scores were drawn for contemporary speech samples (Table M). Here, the reference and the test speakers were derived from the mobile network recordings. Speaker identification scores vary among the vowels preceding nasals and also when the samples were randomized (Table M). Average speaker identification percentage for / for */a:m/, /i:m/, /u:m/, /a:n/, /i:n/ and /u:n/* was 80%, 68%, 60%, 85%, 58% and 69%, respectively. Standard deviation for */a:m/, /i:m/, /u:m/, /a:n/, /i:n/ and /u:n/* was 11, 14.3, 14.5, 16.1, 15.2 and 10.5, respectively (Table M). In this condition, performance of the vowel */a:/* preceding nasal continuant */n/* (*/a:n/*) was better than vowel */a:/* preceding nasal continuant */m/* (*/a:m/*). The performance of the vowel */i:/* preceding nasal continuant */m/* (*/i:m/*) was better than vowel */i:/* preceding nasal continuant */n/* (*/i:n/*). The performance of the vowel */u:/* preceding nasal continuant */n/* (*/u:n/*) was better than the vowel */u:/* preceding nasal continuant */m/* (*/u:m/*). The overall performance of vowel */a:/* preceding nasal continuants */n/* and */m/* (*/a:n/ & /a:m/*) was better compared to the vowels /i:/ and /u:/ preceding nasal continuants */n/ and /m/* (*/i:n/, /i:m/, /u:n/ & /u:m/*). Across vowels and nasals, results showed an average correct identification score of 83%, 63%, 65%, 69% and 71% for */a:/, /i:/, /u:/, /m/ and /n/,* respectively. Hence, vowel */a:/* with 83% and the nasal */n/* with 71% has the highest correct identification scores (Table Q).

Sixth, the percentage of speaker identification scores for ten speakers ranged from 47% to 51% when mobile network recording was compared with live recording for the vowels preceding nasal continuants */a:m/, /i:m/ and /u:m/.* The percentage of speaker identification scores for ten speakers ranged from 46% to 53% when mobile network recording was compared with live recording for the vowels preceding nasal continuant */a:n/, /i:n/ and /u:n/* and can be stated that these scores were drawn for non contemporary speech samples (Table O). Average speaker identification percentage for */a:m/, /i:m/, /u:m/, /a:n/, /i:n/ and /u:n/* was 47%, 51%, 50%, 50%, 53% and 46%, respectively. Standard deviation for */a:m/, /i:m/, /u:m/, /a:n/, /i:n/ and /u:n/* was 20, 21.1, 16.1, 20.8, 18.8 and 14.3, respectively (Table O). In this condition, performance of the vowel */a:/* preceding nasal continuant */n/* (*/a:n/*) was better than vowel */a:/* preceding nasal continuant */m/* (*/a:m/*). The performance of the vowel */i:/* preceding nasal continuant */n/* (*/i:n/*) was better than vowel */i:/* preceding nasal continuant */m/* (*/i:m/*). The performance of the vowel */u:/* preceding nasal continuant */m/* (*/u:m/*) was better than the vowel */u:/* preceding nasal continuant */n/* (*/u:n/*). The overall performance of vowel */i:/* preceding nasal continuants */n/* and */m/* (*/i:n/ & /i:m/*) was better compared to the vowels */a:/* and */u:/* preceding nasal continuants */n/* and */m/* (*/a:n/, /a:m/, /u:n/ & /u:m/*). Results showed an average correct identification score of 49%, 52%, 48%, 49% and 50% for */a:/, /i:/, /u:/, /m/ and /n/,* respectively. The above results imply that this condition does not support the feature vector as a good cue for speaker identification compared to the other condition I and II of this section due to lesser or close to chance factor (Table Q).

From these results it is 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 decreae 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 which are transformed and statistically matched where these authors had divided the thirty speakers (20 males and 10 females) into three subclasses with ten speakers and obtained an identification accuracy of 97% at subclasses and 93% with thirty speakers.

Characteristically the presentation of a text- independent speaker verification system is poorer than a text-dependent system (Doddington, 1998; Boves and Den Oves, 1998) whereas in the present study, text dependent procedure was established. In general, the speaker identification scores obtained for condition I (live recording versus live recording) of section A and B was better than speaker identification scores obtained for condition II and III (mobile network recording versus mobile network recording & mobile network recording versus live recording) of section A and B. The speaker identification scores obtained for condition II (mobile network recording versus mobile network recording) of section A and B was better than speaker identification scores obtained for condition III (mobile network recording versus live recording) of section A and B. The speaker identification scores for condition III (mobile network recording versus live recording) of section A and B were extremely less and hence cannot be considered as a good cue for speaker identification in forensic scenario. The live recordings in this study were carried out in laboratory condition whereas the mobile network recordings were carried out in a field condition (natural environment) with moderate background noise. This may be one of the causes to obtain very less scores for condition III compared to condition I and II of section A and B.

The results of the study were in agreement with the findings of the power spectra of nasal consonants (Glenn and Kleiner, 1968) and co-articulated nasal spectra (Su, Li and 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). The present study focused on backward effect i.e., on effect of nasals on preceding vowels thus providing good speaker identification scores.

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. The current study indicated benchmarking for speaker identification 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.

Limitations and Future directions

The present study data is restricted to only twenty male speakers and concerned with Mysuru dialect Kannada language. Female participants were not considered in the present study. The results of the present study cannot be generalized to entire population of Kannada speaking individuals as number of participants taken were less. Similarly, the study can be further extended to larger population of Kannada speaking individuals and other Indian languages. Only two nasals */m/* and */n/* following the vowels were considered in the present study and co-articulation effect of other nasal sounds in Kannada can be considered for analysis. In this study, semi automatic speaker identification system was used whereas speaker identification scores can be compared using automatic speaker identification 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.

CHAPTER VI

SUMMARY AND CONCLUSIONS

The term 'Forensic' is derived from the Latin word *"forēnsis"* which means belonging to *courts of justice or courts of law or public discussion and debate*.

Biometrics refers to the identification of a person's identity based on his/her traits. Such traits may vary from simple factors such as height, weight, build, facial complexion, color of the eyes, etc. to the more sophisticated factors such as finger prints, DNA etc. Identification of a person's through his/her speech is called speaker recognition. Speech as a biometric has gained popularity due to the extensive use of speech in man-machine communication.

The need to establish the identity for identifying a person from his/her voice is important because of the legal ramifications and forensic involvements. Fingerprinting, photographic and anthropometric techniques are the most commonly used methods of identification. 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.

Speaker recognition/identification is defined as any decision making process that uses the speaker dependent features of the speech signal (Hecker, 1971). Speaker identification by machines has become popular since the invention of telephone and computers. It can be classified into semi-automatic method, where human interference is required in the decision making process, and automatic method, where the entire procedure of speaker identification/verification is carried out by the computer program. Typically, a speaker verification system extracts feature vectors from the speech sample, does a pattern matching with the available set of database or references and finally classifies the speaker as the true speaker or impostor.

Several authors in the past have used feature vectors such as Linear Prediction Coeffients (Atal, 1974), Cepstral coefficients (Jakkar, 2009), Mel-Frequency Cepstral Coefficients (Plumpe, Quateri and Reynolds, 1999) etc. MFCCs have found to be the most efficient feature vectors in classifying a speaker. Glenn and Kleiner (1968) conducted an experiment on nasal continuants using automatic speaker verification methods, which yielded them a result of 93% accuracy in identifying speakers. Jyotsna (2011) conducted a study which revealed that the benchmark for speaker identification using cepstral coefficients was above 80% and using MFCCs, it was above 90% for nasal co-articulation in Malayalam. Ridha (2014) conducted a study using MFCCs derived from Hindi nasal continuants and achieved scores of 100%, 90% and 100% for the nasals */m/, /n/ and /ŋ/* respectively on speaker identification.

The current study aimed at establishing a benchmark for speaker identification using Kannada vowels preceding nasal continuants in live recording, mobile network recording and comparison of the live and mobile network recordings. A lack of research in the area of speaker identification in Kannada language using vowels preceding nasal continuants has validated the need to carry out this study.

Twenty male participants between the age of 21 and 32 years were chosen for this study. They were native speakers of Kannada (Mysuru dialect) and had no history of speech, language or hearing difficulties. Meaningful Kannada sentences were chosen for the study. These sentences consisted of words with three basic vowels *(/a:/, /i:/, /u:/)* preceding to the two places of nasal consonants *(/m/ and /n/)* and were embedded in 3-6 word meaningful sentences to maintain the naturalness of speech and the sentences were derived based on the colloquial/informal spoken language. There were five occurrences for each vowel preceding nasal continuants *(/a:m/, /i:m/, / u:m/, /a:n/, /i:n/ and /u:n/)*. The participants were asked to repeat each sentence four times at habitual pitch, loudness and rate. They were specifically instructed not to adopt a strict reading style, instead asked to adopt a casual conversational style while reading out the sentences under live and mobile network recordings.

Mobile network recording was done first and the network used for making the calls was Airtel on a NOKIA 101 handset and the receiving network was Vodafone on a Gionee S5.5 mobile phone. A participant participating in an experiment was given a NOKIA 101 handset (Airtel network). A call was made from the participants’ handset to the experimenters’ handset (Vodafone network) 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. 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. The live recordings were carried out after two weeks using Computerized Speech Lab (CSL) in a laboratory and the files were stored in ***.wav format***. The distance between the mouth and the dynamic microphone was kept constant at approximately 10 cm. 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.

The recorded samples were transferred to computer memory, down sampled to 8 kHz and was segmented (approximately 300ms) manually using PRAAT software to obtain the vowels preceding nasal continuants in initial and medial positions of the target words. Later, SSL Workbench for Semi-Automatic vocabulary dependent speaker recognition (Voice and Speech Systems, Bangalore, India) software was used for analyses. The analyses were performed separately for live vs live condition, mobile network vs mobile network condition and mobile network vs live condition. In the live vs live condition, the reference and the test sample were obtained from the live recording. For the mobile network vs mobile network condition, the reference and test samples were obtained from the mobile network recordings where as for mobile network vs live condition, the reference and test samples were obtained from live and mobile network recordings, respectively.

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. From the distance matrix, the total percentage of correct speaker identification score was displayed. MFCC's were also used to compute Euclidian distance considering ten speakers (reducing the total number of speakers from 20 to 10). The percent correct identification increased as the number of participants decreased.

To summarize, vowel */a:/* with 93% had highest correct identification scores than vowels */i:/ and /u:/* and across nasals */m/ and /n/,* both shown similar results with 84% when live recording was compared with live recording considering 20 speakers. Vowel */a:/* with 74% had highest correct identification scores than vowels */i:/ and /u:/,* and across nasals, */m/* with 61% had highest correct identification scores when mobile network was compared with mobile network recordings considering 20 speakers. The obtained percent correct identification scores for three vowels and two nasals were lesser than the chance factor (<50%) when mobile network recording was compared with live recording considering 20 speakers.

Vowel */a:/* with 94% and both the nasals */m/ and /n/* with 88% had the highest correct identification scores when live recording was compared with live recording considering 10 speakers. Vowel */a:/* with 83% and the nasal */n/* with 71% had the highest correct identification scores when mobile network was compared with mobile network recordings considering 10 speakers. The results, obtained when mobile network recording was compared with live recording considering 10 speakers, does not support the feature vector as a good cue for speaker identification due to lesser or close to chance factor.

The poor scores in the mobile recording condition could be attributed to the transmission characteristics of the network. The current study was a text-dependent study conducted in a natural environment with some ambient noise. Both contemporary and non contemporary samples were used in the present study. These factors could have contributed to further reduction in accuracy of speaker identification.

The poor scores were obtained when mobile recording condition was compared with live recording condition, could be attributed because of non-contemporary speech samples recording at different sessions (approximately two weeks). It is also important to consider the environment in which both the recordings were done. Mobile network recording was done in a natural field condition and the live recording was done in a laboratory (noise free) condition. Also, emotional state of participants varies in both the conditions. When emotions alter the human voice, the performances of the speaker recognition system decrease significantly.

To conclude, the current study found that the vowel */a:/* preceding both the nasals */m/ and /n/* was reliable for speaker identification compared to other vowels. Also both the nasal continuants */m/ and /n/* has similar effect in speaker identification scores. Furthermore, the number of speaker considerably alters the performance of semi-automatic speaker recognition system.

REFERENCES

Abercrombie, D. (1967). *Elements of general phonetics* (Vol. 203). Edinburgh: Edinburgh University Press.

Amino, K. (2004). Properties of the Japanese phonemes in aural speaker identification. *Tech. Rep. IEICE*, *104*(149), 49-54.

Amino, K., & Arai, T. (2009). Effect of linguistic contents on perceptual speaker identification: Comparison of familiar and unknown speaker identifications. *The Acoustical Society of Japan*, 30, 2-4.

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

Amino, K., & Arai, T. (2009). Speaker dependent characteristics of the nasals. *Forensic Science International*, Vol.185, Issues 1-3, 21-28.

Amino, K., Sugawara, T., & Arai, T. (2006). Idiosyncrasy of nasal sounds in human speaker identification and their acoustic properties, Acoustic Science and Technology, Vol. 27 (4). 233-235.

Arjun, M. S. (2014). Speaker identification using fricatives in Kannada speaking individuals: A preliminary study. Unpublished dissertation of Master of Science in Forensic Science submitted to University of Mysore, Mysuru.

Atal, B. S. (1972). Automatic speaker recognition based on pitch contours. *The Journal of the Acoustical Society of America*, *52*(6B), 1687-1697.

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.

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.

Atal, B. S. (1976). Automatic recognition of speakers from their voices. *Proceedings of the IEEE*, *64*(4), 460-475.

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

Black, J. W., Lashbrook, W., Nash, E., Oyer, H. I., Pedrey, C., Tosi, O. I., & Truby, H. (1973). Reply to "Speaker identification by speech spectrograms: some further observations"[RH Bolt, FS Cooper, EE David, PB Denes, JM Pickett, and KN Stevens, J. Acoust. Soc. Am. 54, 531–534 (1973)]. *The Journal of the Acoustical Society of America*, *54*(2), 535-537.

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

Bolt, R. H., Cooper, F. S., David Jr, E. E., Denes, P. B., Pickett, J. M., & Stevens, K. N. (1970). Speaker identification by speech spectrograms: A scientists' view of its reliability for legal purposes. *The Journal of the Acoustical Society of America*, *47*(2B), 597-612.

Bolt, R. H., Cooper, F. S., David Jr, E. E., Denes, P. B., Pickett, J. M., & Stevens, K. N. (1973). Speaker identification by speech spectrograms: some further observations. *The Journal of the Acoustical Society of America*, *54*(2), 531-534.

Bricker, P. D., & Pruzansky, S. (1966). Effects of stimulus content and duration on talker identification.*The Journal of the Acoustical Society of America*,*40*(6), 1441-1449.

Bricker, P. D., & Pruzansky, S. (1976). Speaker recognition. *Contemporary issues in experimental phonetics*, 295-326.

Butcher, A., & Weiher, E. (1976). An electropalatographic investigation of coarticulation in VCV sequences. *Journal of Phonetics*, *4*(1), 59-74.

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.

Che, C., & Lin, Q. (1995). Speaker recognition using HMM with experiments on the YOHO database. In*Fourth European Conference on Speech Communication and Technology*.

Coleman, R. O. (1971). Male and female voice quality and its relationship to vowel formant frequencies. *Journal of Speech, Language, and Hearing Research*, *14*(3), 565-577.

Coleman, R. O. (1973). Speaker identification in the absence of inter-subject differences in glottal source characteristics. *The Journal of the Acoustical Society of America*, *53*(6), 1741-1743.

Corsi, P. (1982). Speaker recognition: A survey in *Automatic Speech Analysis and Recognition* (pp. 277-308). Springer Netherlands.

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.

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.

Furui, S. (1981). Cepstral analysis technique for automatic speaker verification. *Acoustics, Speech and Signal Processing, IEEE Transactions on*, *29*(2), 254-272.

Furui, S. (1996). An overview of speaker recognition technology. In *Automatic speech and speaker recognition* (pp. 31-56). Springer US.

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

Glass, J. R., & Zue, V. W. (1985, April). Detection of nasalized vowels in American English. In *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP'85*. (Vol. 10, pp. 1569-1572). IEEE.

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.

Hazen, B. (1973). Effects of differing phonetic contexts on spectrographic speaker identification.*The Journal of the Acoustical Society of America*,*54*(3), 650-660.

Hecker, M. H. (1971). Speaker Recognition: Basic Considerations and Methodology. *The Journal of the Acoustical Society of America*, *49*(1A), 138-138.

Hecker, M. H. (1971). *Speaker recognition: An interpretive survey of the literature*. American Speech and Hearing Association.

Hecker, M. H., Stevens, K. N., von Bismarck, G., & Williams, C. E. (1968). Manifestations of Task-Induced Stress in the Acoustic Speech Signal. *The Journal of the Acoustical Society of America*, *44*(4), 993-1001.

Higgins, A. L., & Wohlford, R. (1986, April). A new method of text-independent speaker recognition. In*Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP'86.* (Vol. 11, pp. 869-872). IEEE.

Hollien, H. (1974). Peculiar case of “voiceprints”. *The Journal of the Acoustical Society of America*, *56*(1), 210-213.

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.

Ingemann, F. (1968). Identification of the speaker's sex from voiceless fricatives. *The Journal of the Acoustical Society of America*, *44*(4), 1142-1144.

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

Kent, R. D., & Lass, N. (1976). Models of speech production. *Contemporary issues in experimental phonetics*, *13*.

Kersta, L. G. (1962). Voiceprint identification. *The Journal of the Acoustical Society of America*, *34*(5), 725-725.

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

Kiukaanniemi, H., Siponen, P., & Mattila, P. (1982). Individual differences in the long-term speech spectrum. *Folia Phoniatrica et Logopaedica*, *34*(1), 21-28.

Koenig, B. E. (1986). Spectrographic voice identification: a forensic survey. *The Journal of the Acoustical Society of America*, *79*(6), 2088-2090.

Künzel, H. J. (1987). *Sprechererkennung: Grundzüge forensischer Sprachverarbeitung*. Kriminalistik-Verlag.

Künzel, H. J. (1994). Current approaches to forensic speaker recognition. In *Automatic Speaker Recognition, Identification and Verification*.Kuwabara, H., & Sagisak, Y. (1995). Acoustic characteristics of speaker individuality: Control and conversion. *Speech communication*, *16*(2), 165-173.

La Riviere, C. (1974). Speaker identification from turbulent portions of fricatives. *Phonetica*, *29*(4), 246-252.

Labov, W., & Harris, W. A. (1994). Addressing social issues through linguistic evidence. *Language and the Law*, 265-305.

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

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.

Lass, N. J., Hughes, K. R., Bowyer, M. D., Waters, L. T., & Bourne, V. T. (1976). Speaker sex identification from voiced, whispered, and filtered isolated vowels. *The Journal of the Acoustical Society of America*, *59*(3), 675-678.

Laver, J. (1994). *Principles of phonetics*. Cambridge University Press.

Lei, H., & Gonzalo, E. L. (2009). Importance of nasality measures for speaker recognition data selection and performance prediction. In*INTERSPEECH* (pp. 888-891).

Li, K. P., & Wrench Jr, E. H. (1982). Text-independent speaker recognition with short utterances. *The Journal of the Acoustical Society of America*, *72*(S1), S29-S30.

Luck, J. E. (1969). Automatic speaker verification using cepstral measurements. *The Journal of the Acoustical Society of America*, *46*(4B), 1026-1032.

Mao, D., Cao, H., Murat, H., & Tong, Q. (2006). [Speaker identification based on Mel frequency cepstrum coefficient and complexity measure].*Sheng wu yi xue gong cheng xue za zhi= Journal of biomedical engineering= Shengwu yixue gongchengxue zazhi*, *23*(4), 882-886.

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

McGehee, F. (1937). The reliability of the identification of the human voice. *The Journal of General Psychology*, *17*(2), 249-271.

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.

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

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

Nabar, B. S. (2001). *Forensic Science in crime investigation*. Asia Law House.

Naik, J. (1994). Speaker verification over the telephone network: databases, algorithms and performance assessment. In *Automatic Speaker Recognition, Identification and Verification*.

Nakasone, H., & Beck, S. D. (2001). Forensic automatic speaker recognition. In *2001: A Speaker Odyssey-The Speaker Recognition Workshop*.

Nolan, F. (1997). Speaker recognition and forensic phonetics. *The handbook of phonetic sciences*, 744-767.

Nolan, F. J. D. (1980). *The phonetic bases of speaker recognition* (Doctoral dissertation, University of Cambridge).

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.

Owen, T., & McDermott, M. C. (1996). Voice Identification: The Aural/Spectrographic Method, Chapt. 6. *Lawyers & Judges Publishing Comapny, Inc*, *2*(0), 22.

Pamela, S. (2002). *Reliability of Voiceprints. M. Jayaram & SR Savithri* (Doctoral dissertation, Dissertation abstract: Research at AIISH).

Patel, K., & Prasad, R. K. (2013). Speech Recognition and Verification Using MFCC & VQ.*Interntional Journal*.

Pei, M., & Gaynor, F. (1954). *A dictionary of linguistics* (No. 177). Philosophical Library.

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.

Pollack, I., Pickett, J. M., & Sumby, W. H. (1954). On the identification of speakers by voice. *the Journal of the Acoustical Society of America*, *26*(3), 403-406.

Potter, R. K. (1946). Introduction to technical discussions of sound portrayal. *The Journal of the Acoustical Society of America*, *18*(1), 1-3.

Pruthi, T., & Espy-Wilson, C. Y. (2007, August). Acoustic parameters for the automatic detection of vowel nasalization. In *Interspeech* (pp. 1925-1928).

Pruzansky, S. (1963). Pattern Matching Procedure for Automatic Talker Recognition. *The Journal of the Acoustical Society of America*, *35*(3), 354-358.

Rajshekar, 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.

Ranganathan, M. (2003). Speaker identification in disguised speech. Unpublished dissertation of All India Institute of Speech and Hearing submitted to University of Mysore, Mysuru.

References

Reich, A. R. (1981). Detecting the presence of vocal disguise in the male voice. *The Journal of the Acoustical Society of America*, *69*(5), 1458-1461.

Reich, A. R., & Duke, J. E. (1979). Effects of selected vocal disguises upon speaker identification by listening. *The Journal of the Acoustical Society of America*, *66*(4), 1023-1028.

Reich, A. R., Moll, K. L., & Curtis, J. F. (1976). Effects of selected vocal disguises upon spectrographic speaker identification. *The Journal of the Acoustical Society of America*, *60*(4), 919-925.

Ridha, 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.

Rothman, H. B. (1977). A perceptual (aural) and spectrographic identification of talkers with similar sounding voices. In *International Conference on Crime Countermeasures-Science and Engineering*(pp. 37-42).

Schwartz, M. F. (1968). Identification of speaker sex from isolated, voiceless fricatives. *The Journal of the Acoustical Society of America*, *43*(5), 1178-1179.

Schwartz, M. F., & Rine, H. E. (1968). Identification of speaker sex from isolated, whispered vowels. *The Journal of the Acoustical Society of America*, *44*(6), 1736-1737.

Singh, S., & Rajan, E. G. (2011). Vector Quantization approach for speaker recognition using MFCC and inverted MFCC. *International Journal of Computer Applications*, *17*(1), 0975-8887.

Smrkovski, L. (1976). Study of speaker identification by aural and visual examination of non-contemporary speech samples. *JOURNAL OF THE ASSOCIATION OF OFFICIAL ANALYTICAL CHEMISTS*, *59*(4), 927-931.

Soong, F. K., Rosenberg, A. E., Juang, B. H., & Rabiner, L. R. (1987). Report: A vector quantization approach to speaker recognition. *AT&T technical journal*, *66*(2), 14-26.

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.

Sridhar, S. N. (2007). *Modern Kannada Grammar*. Manohar Publishers.

Stevens, K. N. (1971, August). Sources of inter-and intra-speaker variability in the acoustic properties of speech sounds. In *Proceedings of the 7th International Congress of Phonetic Sciences* (pp. 1596-1607).

Stevens, K. N., Williams, C. E., Carbonell, J. R., & Woods, B. (1968). Speaker authentication and identification: a comparison of spectrographic and auditory presentations of speech material. *The Journal of the Acoustical Society of America*, *44*(6), 1596-1607.

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.

Thompson, C. P. (1985). Voice identification: Speaker identifiability and a correction of the record regarding sex effects. *Human Learning: Journal of Practical Research & Applications*.

Tosi, O., Oyer, H., Lashbrook, W., Pedrey, C., Nicol, J., & Nash, E. (1972). Experiment on voice identification. *The Journal of the Acoustical Society of America*, *51*(6B), 2030-2043.

Wang, L., Ohtsuka, S., & Nakagawa, S. (2009, April). High improvement of speaker identification and verification by combining MFCC and phase information. In *Acoustics, Speech and Signal Processing, 2009. ICASSP 2009. IEEE International Conference on* (pp. 4529-4532). IEEE.

Whiteside, S. P. (1998). Identification of a speaker's sex: a fricative study. *Perceptual and motor skills*,*86*(2), 587-591.

Wolf, J. J. (1972). Efficient acoustic parameters for speaker recognition. *The Journal of the Acoustical Society of America*, *51*(6B), 2044-2056.

Young, M. A., & Campbell, R. A. (1967). Effects of context on talker identification. *The Journal of the Acoustical Society of America*, *42*(6), 1250-1254.

APPENDIX- A

All India Institute of Speech and Hearing, Naimisham Campus, Manasagangothri, Mysuru-570006

CONSENT FORM

Project on
"Benchmark for Speaker Identification using Mel Frequency Cepstral Coefficients on Vowels Preceding Nasal Continuants in Kannada"

Information to the participants

I, Mr. Arjun M.S. have undertaken the present study entitled **"Benchmark for Speaker Identification using Mel Frequency Cepstral Coefficients (MFCCs) on Vowels Preceding Nasal Continuants in Kannada"** under the guidance of Mr. R. Rajasudhakar, Lecturer, Department of Speech-Language Sciences, AIISH, Mysuru. The aim of this study is to establish benchmark for speaker identification using MFCCs on vowels preceding nasal continuants in Kannada in live recording, mobile network recording and to compare between live and mobile network recoding. I need to collect data from twenty Kannada speaking neuro-typical adult males. The participants should be in the age range of 20-40 years and all the participants should belong to the same dialect of language usage (Mysuru dialect). Information will be collected through an interview and audio recording. I assure you that this data will be kept confidential. There is no influence or pressure of any kind by me or the investigating institute to your participation. There is no risk involved to the participants but your co-operation in the study will go a long way in helping to establish benchmark for speaker identification using MFCCs on vowels preceding nasal continuants in Kannada.

Informed Consent

I have been informed about the aims, objectives and the procedure of the study. The possible risks-benefits of myself participation as human subject in the study are clearly understood by me. I understand that I have a right to refuse participation as participant or withdraw my consent at any time. I am also aware that by subjecting to this investigation, I will have to give more time for assessments by the investigating team and that these assessments may not result in any benefits to me.

I, ____, the undersigned, give my consent to be participant of this investigation/study/program.

Signature of participant
(Name and Address)

____

____

Signature of investigator
Date:

Appendix B

All India Institute of Speech & Hearing, Mysore

KANNADA DIAGNOSTIC PHOTO ARTICULATION TEST - SCORING SHEET

Name:
Age/Gender:
DOB:
Structural and functional abnormalities:
(OME, Speech, Language and Hearing):
Mode of testing: Repetition/ Picture naming:

Case No.
Date of Recording:
Mother tongue (other languages):

Score

Part I:
Part II:
Total:

Scoring key:

Score	Vowels & singleton consonants
1	Correct response
0.75	Distortion (D)
0.5	Substitution (S)
0	Omission (O)
	Other articulatory deviation (A0)

Part I

Age range (in years)	Phoneme Tested (Isolation)	Response	Position in Word	Stimuli (Word Level)		Response	SODA	Score
	A		I	ಅಜ್ಜಿ	/aʤʤI/			
	a:		I	ಆನೆ	/a:ne/			
	I		I	ಇಲಿ	/ili/			
	i:		I	ಈರುಳ್ಳಿ	/i:rUḷḷi/			
	U		I	ಉಂಗುರ	/uŋgura/			
	u:		I	ಊಟ	/u:ṭa/			
	E		I	ಎಲೆ	/ele/			
	e:		I	ಏಳು	/e:ḷu/			
	Ai		I	ಐದು	/aidu/_			
2.0-2.6	O		I	ಒಂಟೆ	/onṭe/			
	o:		I	ಓಲೆ	/o:le/			
	au		I	ಔಷಧ	/auʃada/_			
	k		I	ಕತ್ತರಿ	/kat̲t̲ari/			
	k		M	ಬೆಕ್ಕು	/bekku/			
	G		I	ಗಡಿಯಾರ	/gaḍIja:ra/			
	G		M	ಮೂಗು	/mu:gu/			
	t̲		I	ತಟ್ಟೆ	/t̲aṭṭe/			
	t̲		M	ಕೋತಿ	/ko:t̲i/			
	d̲		I	ದಾರ	/d̲a:ra/			
	d̲		M	ಕುದುರೆ	/kud̲ure/			

Note: Testing in isolation and spontaneous speech included for clinical purpose.

2.0-2.6	n		I	ನಲ್ಲಿ	/nalli/			
	N		M	ದೇವಸ್ಥಾನ	/devast̲a:na/			
	p		I	ಪೂರಿ	/pu:ri/			
	p		M	ಕಪ್ಪೆ	/kappe/			
	b		I	ಬಾಗಿಲು	/ba:gilu/			
	b		M	ಕಬ್ಬು	/kabbu/			
	m		I	ಮನೆ	/mane/			
	m		M	ಎಮ್ಮೆ	/jamme/			
	j		I	ಯಕ್ಷಗಾನ	/jakʃaga:na/			
	j		M	ತೆಂಗಿನಕಾಯಿ	/tɛnginakɔi/			
2.6–3.6	v		I	ವಿಮಾನ	/vɪma:na/			
	v		M	ಕಿವಿ	/kivi/			
	l		I	ಲೋಟ	/lo:ṭa/			
	l		M	ಹಲ್ಲು	/hallu/			
3.6–4.0	ʧ		I	ಚಪಾತಿ	/ʧapa:t̲i/			
	ʧ		M	ಬಾಚಣಿಗೆ	/ba:ʧaŋɪge/			
	ʤ		I	ಜಿಂಕೆ	/ʤɪnke/			
	ʤ		M	ಪೂಜಾರಿ	/pu:ʤa:rɪ/			
	ḍ		I	ಡಾಕ್ಟರ್	/ḍa:kṭar/			
	ḍ		M	ಅಂಗಡಿ	/aŋgaḍi/			
4.0–4.6	ṭ		I	ಟೋಪಿ	/ṭo:pɪ/			
	ṭ		M	ಚಿಟ್ಟೆ	/ʧiṭṭɛ/			
	ŋ		M	ಕಣ್ಣು	/kaŋŋu/			
4.6–5.0	r		I	ರೈಲು	/raɪlu/			
	r		M	ಮರ	/mara/			
	ḷ		M	ಬಳೆ	/baḷe/			
	ʃ		I	ಶರ್ಟು	/ʃarṭu/			
	ʃ		M	ಬ್ರಷ್	/braʃ/			
	s		I	ಸೇಬು	/se:bu/			
	s		M	ಮೀಸೆ	/mi:se/			
> 6.0	h		I	ಹಾವು	/ha:vu/			
	h		M	ಸಿಂಹ	/simha/			

Interpretation:

Clinician Supervisor

Part II

Age range (in years)	Phoneme Tested (Isolation)	Response	Position in Word	Stimuli (Word Level)		Response	SODA	Score
2.0-2.6	a		I	ಅಡಿಗೆಮನೆ	/aḍigemane/			
	a:		I	ಆಸ್ಪತ್ರೆ	/a:spat̲re/			
	i		I	ಇರುವೆ	/Iruve/			
	i:		I	ಈಳಿಗೆಮಣೆ	/i:ḷigemaɳe/			
	u		I	ಉಯ್ಯಾಲೆ	/ujja:le/			
	u:		I	ಊರು	/u:ru/			
	e		I	ಎರಡು	/eraḍu/			
	e:		I	ಏಣಿ	/e:ɳI/			
	ai		I	ಐವತ್ತು	/aivat̲t̲u/			
	o		I	ಒಂದು	/ond̲u/			
	o:		I	ಓಡು	/o:ḍu/			
	k		I	ಕಾರು	/ka:ru/			
	k		M	ಸೈಕಲ್ಲು	/saIkallu/			
	g		I	ಗಾಳಿಪಟ	/ga:ḷpaṭa/			
	g		M	ಕಾಗೆ	/ka:ge/			
	t̲		I	ತಬಲ	/t̲abala/			
	t̲		M	ಕತ್ತೆ	/kat̲t̲e/			
	d̲		I	ದೋಸೆ	/d̲o:se/			
	d̲		M	ಗೋದಿ	/go:d̲I/			
	n		I	ನಾಯಿ	/na:ji/			
	n		M	ಮೀನು	/mi:nu/			
	p		I	ಪೆನ್ನು	/pennu/			
	p		M	ಚಪ್ಪಲಿ	/ʧəppəli/			
	b		I	ಬಸ್ಸು	/bəssu/			
	b		M	ದಿಂಬು	/dImbu/			
	m		I	ಮೂರು	/mu:ru/			
	m		M	ಆಮೆ	/a:me/			
	j		I	ಯಂತ್ರ	/jantra/			
	j		M	ಕಡಲೆಕಾಯಿ	/kəḍəlekəI/			
2.6–3.6	v		I	ವೀಣೆ	/vi:ɳe/			
	v		M	ಹೂವು	/hu:vu/			
	l		I	ಲಂಗ	/langa/			
	l		M	ಗೋಲಿ	/go:lI/			

3.6–4.0	ʧ		I	ಚಕ್ಕುಲಿ	/ʧakkulI/			
	ʧ		M	ಮಂಚ	/manʧa/			
	ʤ		I	ಜಡೆ	/ʤaḍe/			
	ʤ		M	ಸೂಜಿ	/su:ʤI/			
	ḍ		I	ಡಬ್ಬಿ	/ḍəbbI/			
	ḍ		M	ಕನ್ನಡಕ	/kannaḍəka/			
4.0–4.6	ṭ		I	ಟೊಮೋಟೋ	/ṭomo:ṭo/			
	ṭ		M	ಕಿಟಕಿ	/kiṭΛkI/			
	ŋ		M	ಗಿಣಿ	/giŋi/			
4.6–5.0	r		I	ರಂಗೋಲಿ	/rango:lɪ/			
	r		M	ಸರ	/səra/			
	ḷ		M	ಕೋಳಿ	/ko:ḷI/			
	ʃ		I	ಶಂಖ	/ʃənkə/			
	ʃ		M	ಗಣೇಶ	/gəŋe:ʃa/			
	s		I	ಸೀರೆ	/si:rɛ/			
	s		M	ಹಸು	/hasu/			
	sk		I	ಸ್ಕೂಟರ್	/sku:ṭər/			
	sk		M	ಬಿಸ್ಕೆಟ್ಟು	/bIskeṭṭu/			
	kr		M	ಚಕ್ರ	/ʧakra/			
	bl		I	ಬ್ಲೇಡು	/ble:ḍu/			
	st		M	ಪೋಸ್ಟ್ ಬಾಕ್ಸ್	/po:sṭba:ks/			
5.0–5.6	st		I	ಸ್ಟಾಂಪು	/sṭa:mpu/			
	dr		M	ಚಂದ್ರ	/ʧ^nd̲ra/			
> 6.0	h		I	ಹುಲಿ	/hulI/			
	h		M	ಬಾಳೇಹಣ್ಣು	/ba:ḷehaŋŋu/			
	kʃ		M	ಆಟೋರಿಕ್ಷಾ	/a:ṭorikʃa/			
	dr		I	ದ್ರಾಕ್ಷಿ	/d̲rakʃI/			
	skr		I	ಸ್ಕ್ರೂ	/skru/			
	rʧi		M	ಕುರ್ಚಿ	/kurʧI/			

Spontaneous Speech Sample (Conversation/Narration):

Interpretation:

Clinician | Supervisor

APPENDIX-C

KANNADA SPEECH MATERIAL

1. **ಭೀಮ** ನಮ್ಮ **ಧೀಮಂತ** ನಾಯಕ
2. **ಧೂಮಪಾನ** ಆರೋಗ್ಯಕ್ಕೆ **ಹಾನಿಕರ.**
3. **ಸೀಮಯೆಣ್ಣೆ** ಬೆಲೆ ದುಬಾರಿಯಾಗಿದೆ.
4. **ನವೀನನಿಗೆ ಗೂನು** ಬೆನ್ನು.
5. **ಭಾನುವಾರ ಜಾನಪದ** ನಾಟಕಗಳ ಪ್ರದರ್ಶನವಿದೆ.
6. **ರಾಮನಿಗೆ ಹೀನಾಯವಾಗಿ** ಅವಮಾನ ಮಾಡಿ **ಛೀಮಾರಿ** ಹಾಕಿದರು.
7. ಮಂಡಲಾಕಾರವಾದ **ಭೂಮಿಯನ್ನು ಭೂಮಂಡಲ** ಎನ್ನುತ್ತಾರೆ.
8. ಪ್ರಸಿದ್ಧ **ಗೋಧಾಮ ಬಾದಾಮಿಯ** ಒಂದು ಪುಟ್ಟ **ಗ್ರಾಮದಲ್ಲಿದೆ.**
9. ಅವನು **ತಾಮ್ರ** ಮುಖದ ವಿಗ್ರಹಕ್ಕೆ **ಹೂಮಾಲೆ** ಹಾಕಿದನು.
10. ಸೀತಾಗೆ **ಶೂನ್ಯ** ಅಂಕ ಬಂದಿದೆ.
11. ಜಾದುಗಾರ **ಛೂಮಂತ್ರ** ಹಾಕಿ ಎಲ್ಲರನ್ನು ಮೋಡಿ ಮಾಡಿದ.
12. **ಸಾಮಾನ್ಯವಾಗಿ** ಸೀಮ ಹಾಡು ಕೇಳುತ್ತಾ **ತಲ್ಲೀನಳಾಗುತ್ತಾಳೆ.**
13. **ಜಾನುವಾರುಗಳ** ಸ್ಥಿತಿ ತುಂಬಾ **ಕೀನವಾಗಿತ್ತು.**
14. ಅವನು **ಊನ** ಮನಸ್ಸಿನ ವ್ಯಕ್ತಿ.
15. **ದೀನರಿಗೆ ಕಾರ್ಖಾನೆಯಲ್ಲಿ** ಕೆಲಸ ಕೊಡಬೇಕು.
16. ಅದು ನಿನ್ನ **ಮಗೂನಾ?**
17. **ಕೂನ** ಎಂದರೆ ಬಾಗಿರುವಿಕೆ.

APPENDIX-D

Section A: Euclidian distance Matrix for Group I consisting of 20 speakers

Condition I: Live vs live recording

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	**2.23**	2.91	3.13	3.25	5.09	5.03	3.13	3.16	5.87	4.25	3.21	3.77	3.07	5.54	4.06	2.50	3.52	4.50	4.57	4.23
2	3.15	**2.12**	3.47	3.22	5.09	5.69	4.05	4.37	4.61	5.81	3.44	4.99	3.80	5.48	4.61	3.47	3.47	4.26	4.82	4.79
3	2.95	4.45	**2.62**	2.77	3.81	6.97	4.60	4.29	5.71	6.28	4.38	4.26	5.69	5.85	5.26	4.15	5.21	5.73	5.70	6.50
4	3.73	4.46	3.35	**2.94**	5.83	5.91	3.63	4.51	5.59	5.58	3.31	3.99	4.68	4.92	4.28	3.95	4.59	4.90	4.73	5.27
5	4.93	5.79	5.03	5.32	**2.58**	9.26	6.80	6.28	7.99	7.54	6.92	6.52	7.22	8.47	8.09	6.22	7.11	7.75	8.54	8.63
6	6.21	6.39	6.67	6.10	9.30	**2.28**	5.55	6.01	6.24	6.64	4.52	5.84	5.28	5.20	3.50	4.61	5.26	5.16	4.27	5.35
7	4.08	4.71	4.53	4.38	7.02	5.14	**2.56**	3.77	7.09	3.45	3.58	3.72	3.02	5.73	4.17	3.48	4.51	5.18	4.66	3.72
8	3.43	4.69	4.26	4.65	6.03	6.22	3.42	**2.45**	7.48	4.68	4.86	4.43	4.49	7.25	4.76	3.46	5.64	6.67	5.49	5.23
9	6.45	5.50	6.35	5.37	7.85	6.03	7.12	7.73	**2.47**	9.34	5.30	7.62	6.99	5.11	5.89	6.31	5.15	4.70	6.08	7.27
10	5.22	5.93	5.46	5.68	7.91	5.93	4.19	4.81	8.53	**2.45**	4.55	4.12	3.66	6.28	5.23	4.64	5.41	5.81	4.99	3.82
11	3.74	4.04	3.56	2.86	6.18	4.48	3.74	4.75	4.79	4.87	**1.96**	3.80	3.38	3.48	3.65	3.52	3.33	3.12	3.60	4.07
12	4.89	6.11	4.33	4.44	6.80	6.88	4.16	5.12	7.90	4.22	4.40	**3.10**	5.02	5.68	5.18	4.86	5.54	5.94	5.36	5.39
13	4.41	4.31	4.75	4.47	7.13	4.14	3.91	4.65	6.29	4.15	3.05	4.29	**1.96**	5.02	3.84	3.71	3.51	3.86	4.02	3.13
14	5.82	5.74	5.19	4.64	8.29	4.74	5.17	6.58	5.26	6.07	3.41	4.64	5.06	**2.20**	4.03	5.11	3.88	3.31	3.48	4.76
15	4.32	4.69	4.31	3.76	7.33	3.31	3.88	4.56	5.06	5.41	2.55	3.75	3.78	3.98	**1.89**	3.32	3.37	3.66	3.03	3.88
16	3.41	4.16	3.91	3.87	5.60	5.23	3.71	3.71	6.22	5.03	3.92	4.12	4.51	5.82	4.22	**2.76**	4.87	5.38	4.84	5.32
17	4.59	4.42	4.41	3.94	6.95	4.03	4.66	5.47	4.76	5.51	2.98	4.12	3.97	3.39	3.51	4.00	**2.27**	2.59	3.71	4.03
18	5.23	5.07	4.86	4.15	7.46	4.33	5.36	6.29	4.46	6.06	2.94	4.79	4.50	2.43	4.16	4.72	3.15	**1.88**	3.68	4.37
19	5.31	5.52	5.07	5.02	8.46	4.61	4.43	5.23	6.47	5.15	3.60	4.63	4.32	4.37	3.41	4.47	4.72	4.75	**2.52**	3.85
20	4.24	4.29	4.38	4.42	7.51	4.32	3.43	4.14	6.39	3.74	2.96	3.75	2.67	4.76	3.28	3.50	3.33	3.98	3.04	**2.12**

Table 1: Euclidian distance matrix for */a:m/-100%*

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	**2.41**	3.24	2.91	3.28	4.42	4.52	2.83	2.97	5.61	5.20	3.43	4.32	4.25	5.12	3.77	2.92	3.84	5.89	4.52	5.58
2	3.66	**2.10**	3.20	4.11	5.60	4.54	3.26	3.96	5.49	5.03	3.02	3.98	3.19	5.72	3.03	3.81	3.95	5.88	4.24	4.62
3	3.02	4.17	**2.42**	2.49	4.77	4.58	3.30	3.99	6.21	5.01	3.61	3.50	4.56	4.70	3.59	3.68	4.09	5.98	4.13	5.85
4	3.81	4.60	3.35	**2.86**	5.57	5.14	4.14	4.68	6.42	5.41	3.76	4.39	5.05	5.33	3.88	4.20	4.33	6.37	4.67	6.09
5	4.72	5.62	5.23	5.36	**2.50**	7.86	5.33	5.30	8.13	6.82	6.48	5.09	6.68	7.38	6.72	5.65	6.56	8.94	7.60	8.71
6	5.72	6.03	5.21	5.09	8.34	**2.05**	4.75	5.68	7.06	5.39	4.61	6.04	4.14	4.53	4.42	4.83	5.35	4.42	4.28	3.94
7	2.76	3.08	2.46	3.29	4.98	3.99	**2.12**	2.88	5.43	4.59	3.26	3.39	3.40	4.99	3.05	2.69	3.30	5.32	3.92	4.73
8	3.39	4.10	3.66	4.28	5.56	4.55	3.53	**2.65**	6.85	5.25	4.52	4.62	3.90	6.48	4.17	3.21	4.96	6.80	5.06	5.70
9	6.45	5.53	6.06	6.63	7.97	6.52	5.58	6.97	**2.56**	9.22	5.29	7.53	7.67	6.29	5.77	6.07	4.10	5.02	5.42	6.73
10	5.62	6.22	5.17	4.93	7.06	5.49	5.53	5.74	9.41	**1.90**	5.57	4.75	3.66	6.09	4.96	5.35	6.62	7.39	5.77	5.14
11	3.78	4.12	3.45	3.42	5.97	4.17	3.86	4.59	5.55	5.35	**2.98**	4.72	4.41	4.60	3.59	4.16	3.84	5.36	4.02	5.24
12	5.15	5.54	4.23	4.27	6.13	5.06	4.53	5.79	8.09	4.04	5.18	**2.90**	4.03	5.09	4.48	5.31	5.73	6.59	5.23	5.62
13	4.38	4.50	3.87	3.92	6.13	3.45	3.95	4.58	7.53	3.21	3.90	3.80	**1.95**	4.90	3.77	4.39	5.27	6.07	4.48	4.49
14	4.82	5.13	4.12	4.01	6.42	4.27	4.05	5.79	6.12	5.00	4.00	4.38	4.50	**2.97**	4.09	4.92	4.35	4.40	4.17	4.99
15	4.25	4.19	3.20	3.44	6.73	2.78	3.45	4.51	6.04	4.33	3.15	3.95	3.15	4.33	**2.16**	3.54	3.80	4.49	2.80	3.33
16	3.20	4.00	3.15	3.58	5.13	4.10	3.06	2.90	5.94	5.13	4.00	4.48	4.29	5.36	3.49	**2.33**	3.97	5.62	4.01	5.03
17	3.94	3.86	3.23	3.57	6.15	4.04	3.11	4.53	4.17	5.72	2.79	4.62	4.64	4.21	2.69	3.41	**2.14**	4.00	2.95	4.18
18	6.24	5.93	5.40	5.66	8.50	3.84	4.79	6.60	4.81	6.91	4.65	6.50	5.78	3.52	4.57	5.28	4.08	**2.10**	3.61	4.03
19	5.67	5.75	4.92	5.04	8.14	3.57	4.90	6.08	6.26	5.41	4.42	5.85	4.89	3.87	3.95	4.97	4.55	3.75	**3.00**	3.54
20	6.38	5.67	5.44	5.88	8.60	3.77	4.89	6.29	6.01	5.84	4.73	6.23	4.75	4.65	4.07	4.92	4.48	3.21	3.50	**2.27**

Table 2: Euclidian distance matrix for */a:n/-100%*

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	**2.09**	3.06	3.01	3.16	3.17	3.90	3.51	3.03	5.04	5.76	6.07	2.74	5.08	6.15	4.70	3.72	7.14	5.73	3.72	4.83
2	2.54	**1.78**	3.82	3.87	2.89	4.43	4.26	4.20	5.17	5.79	6.56	2.81	5.48	6.69	5.80	5.49	8.04	5.05	4.10	3.37
3	2.82	3.09	2.72	3.35	**2.68**	3.36	3.17	3.47	4.65	4.98	5.36	2.86	4.49	5.40	4.99	4.85	6.82	4.71	4.28	4.15
4	3.07	3.79	2.62	**1.75**	2.91	4.04	2.54	2.79	5.27	4.79	5.04	2.51	4.54	4.75	4.67	3.91	6.34	5.72	4.60	5.23
5	4.08	4.21	4.14	4.13	**3.62**	5.27	4.55	4.56	6.26	5.78	6.46	3.67	5.83	6.37	6.23	5.91	7.89	6.12	5.28	5.21
6	3.46	4.54	3.51	4.36	4.61	**1.91**	3.06	2.96	3.04	3.89	3.89	3.78	2.82	4.43	3.35	4.57	4.84	3.99	3.83	4.53
7	4.87	5.99	4.13	4.31	5.33	4.30	**2.92**	3.04	4.89	4.32	3.62	4.43	3.78	4.08	3.40	4.13	4.33	5.67	4.77	6.40
8	4.85	6.00	4.06	4.20	5.24	4.90	3.47	**3.07**	5.87	4.92	4.44	4.52	4.48	4.14	4.34	4.16	5.10	6.48	5.18	6.97
9	5.73	6.90	5.41	6.36	6.82	3.86	5.02	4.66	**3.73**	4.70	3.93	5.76	4.20	4.84	4.05	5.88	3.98	4.76	5.21	6.35
10	6.95	7.71	6.42	6.91	7.32	5.82	5.62	5.43	5.94	**3.44**	3.82	6.42	4.15	4.14	5.47	7.39	3.85	5.67	6.53	7.40
11	6.61	7.76	5.76	6.89	7.27	4.35	5.15	5.11	3.87	4.37	**2.84**	6.17	4.19	4.34	3.83	6.35	2.87	4.57	5.43	6.59
12	3.89	4.37	3.76	3.93	3.78	4.23	3.21	3.70	4.92	4.48	4.65	**3.11**	4.37	5.10	4.76	5.31	6.18	4.83	4.45	4.16
13	5.66	6.69	5.29	5.75	6.41	4.25	4.42	4.08	4.72	3.04	3.33	5.43	**2.75**	3.55	4.32	6.13	3.62	5.10	5.60	6.51
14	7.04	7.96	5.99	6.38	7.36	5.32	5.58	5.30	5.56	3.74	3.21	6.40	4.18	**1.80**	5.61	7.08	3.62	5.61	6.72	7.53
15	6.36	7.57	5.99	6.71	7.39	4.75	5.29	4.96	4.38	5.24	4.32	6.20	4.80	5.55	**3.46**	5.57	4.01	5.58	5.07	7.08
16	3.97	5.21	4.00	4.12	5.03	4.60	3.59	3.24	5.32	5.81	5.55	4.16	5.15	6.03	3.66	**2.63**	6.05	6.43	4.12	6.62
17	7.57	8.88	6.83	7.43	8.46	5.63	5.90	5.47	5.37	4.59	3.29	7.21	4.42	4.14	4.25	6.63	**1.68**	6.29	6.54	8.38
18	5.32	5.77	5.22	6.51	5.86	4.02	5.26	5.35	3.71	5.22	4.87	5.23	5.02	6.11	4.94	6.65	5.86	**3.35**	4.56	4.15
19	4.73	5.36	4.86	5.89	5.59	4.52	4.75	4.44	4.63	5.49	5.34	4.86	5.02	6.15	4.08	5.16	5.86	4.99	**3.22**	5.38
20	4.32	3.77	5.15	5.48	4.10	5.21	5.27	5.29	5.72	5.81	6.59	4.06	5.78	6.90	6.86	7.36	8.35	4.85	5.30	**2.64**

Table 3: Euclidian distance matrix for */i:m/-95%*

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	**2.27**	4.22	3.33	3.37	4.55	5.59	3.90	3.35	5.56	3.62	4.95	5.25	4.62	7.03	3.24	3.77	5.57	7.61	4.90	6.99
2	3.83	**2.27**	4.02	4.72	3.07	3.97	5.13	5.20	4.72	4.31	5.43	3.19	3.97	7.33	4.55	6.59	6.26	6.09	4.91	4.64
3	3.97	4.36	**3.00**	3.55	4.15	5.22	3.43	3.95	4.78	4.52	3.92	4.73	4.68	5.83	3.87	4.71	5.06	6.14	4.49	5.89
4	2.76	3.33	2.64	**2.30**	3.64	4.28	3.33	3.68	4.17	3.52	3.87	4.11	3.75	5.95	2.94	4.06	4.86	6.16	4.11	5.33
5	4.77	3.57	4.83	5.12	**3.15**	4.56	6.24	6.41	5.99	5.50	6.65	4.13	5.45	8.74	5.66	6.98	7.56	7.46	6.40	6.18
6	3.70	**3.22**	3.95	4.74	4.09	4.07	4.54	4.33	4.26	3.63	4.76	4.42	3.64	6.52	3.90	5.97	5.03	5.89	4.94	5.55
7	3.45	4.51	3.11	3.42	4.96	5.15	**2.34**	2.92	3.88	3.61	2.77	4.78	3.56	4.40	3.01	4.37	3.66	5.46	4.00	5.18
8	3.73	4.69	4.19	4.56	5.42	5.34	4.16	**3.59**	5.42	4.40	4.97	5.76	4.82	6.77	3.81	4.64	5.34	7.03	5.01	6.59
9	5.18	3.95	4.38	4.67	5.11	3.15	4.93	5.32	**2.94**	4.76	4.11	5.08	3.85	5.01	4.51	6.35	4.53	4.00	3.94	3.91
10	3.61	4.48	4.33	4.55	5.04	5.41	4.42	4.38	4.18	**2.47**	4.12	4.51	3.02	5.61	4.17	5.82	4.19	6.25	5.12	5.58
11	4.94	5.53	4.41	4.65	6.08	5.59	4.00	4.50	4.22	4.58	**3.67**	5.87	4.38	4.80	4.32	5.45	4.20	5.39	4.90	5.60
12	5.95	5.37	6.05	6.72	4.99	6.38	6.00	6.53	5.81	5.46	5.95	**4.04**	4.81	7.44	6.58	8.60	6.51	6.89	7.10	5.21
13	3.97	4.08	4.29	4.46	5.30	4.00	3.75	3.90	2.43	2.84	3.05	4.58	**1.70**	4.27	3.49	5.97	2.95	4.53	3.98	3.60
14	5.99	6.31	5.18	5.11	6.99	5.97	4.27	5.16	3.42	5.17	3.05	6.56	4.36	**1.94**	4.92	6.15	3.20	4.52	4.66	5.07
15	4.23	4.96	3.54	3.74	6.02	4.83	3.32	3.26	3.92	4.43	3.37	6.24	4.35	4.88	**2.87**	3.88	3.80	5.34	3.42	5.78
16	4.39	6.17	4.50	3.91	6.44	6.60	4.92	4.59	6.19	5.52	5.56	7.36	6.25	7.05	4.23	**2.57**	6.05	8.07	5.55	8.12
17	4.96	5.93	4.69	4.85	6.79	5.64	3.45	3.60	3.31	4.22	2.53	6.50	3.75	3.12	3.72	5.10	**1.65**	4.79	4.33	5.38
18	6.72	5.90	5.31	6.29	6.97	4.52	5.15	5.60	3.48	6.01	4.08	6.69	5.04	4.56	5.29	7.22	4.13	**2.21**	4.43	4.57
19	4.30	3.76	3.65	3.98	5.13	4.70	4.85	4.66	4.78	5.01	5.06	5.75	4.92	6.41	3.97	4.89	5.90	6.14	**3.22**	5.86
20	7.24	4.93	6.64	7.35	5.79	4.67	7.27	7.64	5.65	7.03	6.87	5.19	5.82	8.18	7.05	9.31	7.58	5.45	6.26	**3.90**

Table 4: Euclidian distance matrix for */i:n/-95%*

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	**1.84**	3.25	2.42	3.46	2.67	2.96	3.29	2.70	5.68	3.83	3.00	3.36	3.03	4.88	3.43	3.32	4.00	5.15	3.31	4.72
2	3.93	**3.10**	3.48	4.10	3.87	3.36	4.32	4.58	4.17	4.71	3.28	4.10	3.39	5.55	3.17	5.13	3.88	3.78	3.66	3.49
3	2.96	3.29	**2.31**	3.78	2.76	3.88	3.01	2.46	6.09	4.03	2.99	3.14	4.21	4.09	3.98	2.69	3.57	5.04	3.14	5.48
4	3.48	4.07	2.79	**2.17**	3.34	4.07	4.37	3.77	5.51	3.34	2.64	3.97	3.40	5.22	3.75	4.66	3.36	5.38	3.79	4.59
5	3.72	2.92	2.88	3.58	2.92	3.13	3.51	4.65	3.88	4.79	2.64	3.30	3.32	5.86	**2.32**	5.55	3.98	3.22	3.86	2.65
6	2.90	2.91	3.05	3.65	3.73	**2.01**	4.07	4.18	3.90	4.49	2.82	4.15	2.53	5.24	3.20	4.87	3.59	3.92	3.55	4.14
7	3.82	4.00	3.75	5.23	3.40	4.50	**2.91**	3.58	7.05	5.04	4.30	3.25	5.14	5.38	4.22	3.75	5.38	5.45	4.19	5.69
8	3.84	4.69	3.92	5.66	3.83	4.96	3.65	**2.67**	7.56	5.31	4.67	4.13	5.46	4.93	5.24	3.12	5.38	6.18	4.38	6.54
9	5.88	5.07	5.48	5.43	6.46	4.36	6.86	6.99	**2.63**	6.78	4.89	6.84	4.54	6.64	5.37	7.42	4.39	4.58	5.54	5.63
10	4.97	5.53	5.14	4.76	4.88	6.05	5.91	5.07	7.94	**2.85**	5.06	5.28	4.76	6.90	5.62	6.00	5.92	7.40	5.15	5.83
11	4.12	3.78	3.69	4.02	4.07	3.75	4.68	4.99	4.52	4.70	**3.41**	4.52	3.65	6.54	3.68	5.77	4.49	4.49	4.41	3.82
12	5.27	5.11	4.90	5.52	4.55	5.71	4.63	5.36	7.05	5.70	5.11	**4.41**	5.79	6.82	4.87	5.72	6.05	6.00	5.34	5.54
13	4.36	4.60	4.52	4.26	4.70	4.62	5.74	5.47	6.28	4.21	4.27	5.44	**3.57**	7.31	4.85	6.30	5.43	6.46	5.28	4.98
14	4.48	5.02	4.12	4.61	4.41	5.01	4.70	4.12	6.13	4.75	4.37	4.46	5.03	**3.05**	4.77	4.15	3.80	5.40	3.57	6.11
15	2.36	3.07	2.52	3.27	2.35	3.25	2.46	3.12	5.73	3.54	2.90	**2.19**	3.38	4.57	2.61	3.59	4.04	4.63	2.90	4.25
16	3.97	5.13	3.98	5.03	3.89	5.49	4.39	2.88	7.95	4.44	4.76	4.45	5.34	4.82	5.50	**2.81**	5.16	6.98	4.28	6.86
17	4.84	4.71	4.15	3.93	5.03	4.56	5.55	4.82	4.71	4.99	3.86	5.19	4.22	4.81	4.79	5.48	**2.76**	4.87	4.01	5.66
18	5.52	4.01	4.53	5.65	5.00	4.10	4.86	5.95	3.40	6.51	4.09	4.99	4.93	6.18	4.13	6.62	4.33	**2.33**	4.57	4.38
19	4.03	3.71	3.48	4.27	4.14	3.95	4.64	4.01	5.00	4.35	3.48	4.39	3.80	4.16	4.19	4.46	2.95	4.45	**2.67**	5.25
20	4.96	4.52	4.98	5.08	4.39	5.33	4.91	5.87	6.67	4.94	4.93	4.19	4.81	7.72	3.96	6.93	6.33	5.62	5.19	**3.31**

Table 5: Euclidian distance matrix for */u:m/-90%*

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	**1.94**	4.36	2.72	2.89	2.98	3.28	3.06	3.21	4.23	3.76	3.32	3.93	2.97	4.42	4.20	2.20	2.36	5.29	4.59	6.45
2	4.76	**2.35**	3.97	4.67	4.76	4.47	4.32	3.86	4.34	5.34	4.04	4.95	4.63	6.26	4.74	4.75	5.45	5.60	5.28	5.36
3	3.46	4.75	**3.17**	3.23	3.56	3.72	3.51	4.07	3.99	4.36	3.37	4.45	3.61	4.55	4.04	3.53	3.69	4.90	4.77	6.17
4	2.84	4.09	2.93	**2.51**	2.87	3.80	2.99	3.96	4.67	3.52	3.31	3.21	3.35	4.80	4.47	3.35	3.52	5.67	4.85	6.33
5	2.89	4.34	2.97	2.56	**2.06**	3.70	3.13	3.72	4.49	3.37	3.20	3.38	2.86	4.29	3.99	2.97	3.58	5.30	4.98	6.16
6	3.42	4.74	3.12	3.54	3.75	**2.73**	3.73	4.15	4.01	4.66	3.61	3.93	3.28	4.71	3.67	4.01	4.16	3.91	3.63	5.19
7	3.47	4.29	3.39	3.61	3.44	3.97	**2.40**	3.30	5.21	4.69	3.93	3.00	4.19	5.18	4.65	3.83	3.69	5.72	4.59	5.83
8	3.91	5.10	4.14	4.66	4.30	4.77	3.59	**3.06**	5.46	5.76	4.95	4.79	4.58	4.83	4.91	3.71	3.77	6.37	4.98	6.96
9	4.56	4.03	3.51	4.08	5.05	3.51	4.76	4.74	**2.44**	5.04	3.02	5.54	4.24	5.78	3.65	4.55	5.14	3.53	4.69	4.70
10	4.21	5.36	4.60	4.55	4.16	4.38	4.86	5.66	5.72	**2.80**	4.48	4.84	4.05	5.91	5.51	5.10	5.53	5.97	6.10	6.61
11	2.86	4.11	2.42	2.48	3.29	2.89	3.14	3.69	3.15	3.72	**2.35**	3.64	2.97	3.92	2.92	3.06	3.45	4.29	3.84	5.31
12	4.52	5.18	4.59	4.24	3.85	5.38	3.65	4.54	6.38	5.33	5.08	**3.05**	4.97	5.39	5.55	4.72	4.69	7.07	5.44	6.90
13	3.41	5.47	3.33	3.43	3.98	3.17	4.27	4.80	3.72	4.51	3.77	4.55	**3.07**	3.80	3.65	3.93	3.96	4.14	3.72	5.99
14	5.43	7.61	5.56	5.64	5.49	6.29	5.90	6.08	6.43	6.89	6.62	6.17	5.39	**3.14**	5.62	5.49	5.42	7.65	5.23	8.95
15	4.58	5.19	3.71	3.99	4.81	3.75	4.49	4.70	3.57	5.46	3.56	4.78	4.30	4.37	**2.88**	4.45	4.85	3.87	3.85	4.70
16	2.76	5.49	3.15	2.98	3.50	3.76	3.12	3.60	4.56	5.03	3.91	3.78	3.72	4.04	4.01	2.68	**2.11**	5.27	4.16	6.57
17	3.90	6.08	4.02	3.97	4.67	4.75	4.05	4.35	4.98	5.96	4.60	5.05	4.79	4.88	4.81	3.53	**2.99**	5.94	4.97	7.35
18	5.34	5.18	4.06	4.64	5.21	3.14	4.82	5.06	3.52	6.05	3.85	5.07	4.53	5.71	3.04	5.48	5.89	**1.43**	3.89	3.14
19	5.43	5.78	4.71	5.21	5.51	4.15	5.38	5.62	4.78	6.41	5.24	5.12	4.73	4.96	4.00	5.88	6.07	4.30	**3.33**	5.16
20	6.46	4.92	5.43	5.95	6.20	4.64	5.50	5.59	5.42	6.59	4.83	5.37	5.92	7.33	4.75	6.61	7.17	4.28	5.50	**2.60**

Table 6: Euclidian distance matrix for */u:n/-95%*

Condition II: Mobile network vs mobile network recording

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**
1	**3.09**	8.77	5.94	7.05	6.64	4.08	9.78	4.64	6.17	8.37	7.59	4.69	3.97	8.43	3.56	4.93	6.79	5.83	8.04	6.57
2	9.28	**4.14**	9.72	9.94	9.68	7.15	8.44	8.11	7.18	8.55	8.70	9.49	8.09	12.19	8.96	7.58	7.76	9.00	9.40	7.80
3	5.51	9.23	3.07	3.41	**2.74**	5.74	6.75	4.61	3.95	6.19	6.36	3.98	6.04	4.76	4.80	5.19	4.10	4.08	5.01	7.15
4	7.98	11.04	4.27	**3.64**	4.05	8.00	6.51	6.11	5.80	6.32	7.28	6.05	8.56	4.85	7.15	7.12	4.74	5.31	4.71	8.70
5	8.21	11.76	4.56	4.99	**4.10**	8.79	7.93	7.55	6.18	7.70	7.76	6.18	9.08	5.33	7.72	7.95	6.05	5.65	5.84	9.25
6	4.61	7.11	6.43	7.34	6.92	**3.83**	8.70	4.93	5.55	7.76	6.89	5.88	4.46	9.29	4.78	4.68	6.16	6.00	7.06	5.69
7	8.75	9.43	6.08	5.68	5.98	8.10	**4.12**	6.86	5.84	4.41	6.93	6.61	8.77	6.73	7.72	7.27	4.47	5.82	5.82	8.04
8	7.04	9.13	6.26	5.45	6.60	6.17	7.61	**4.14**	6.70	6.96	8.69	6.75	6.76	7.60	6.12	6.40	5.60	6.85	7.21	8.46
9	8.10	8.88	6.32	6.27	5.73	7.53	7.47	7.18	**5.10**	7.81	7.30	7.00	8.23	7.55	7.73	6.97	5.98	6.21	6.31	7.93
10	9.15	9.53	6.67	5.89	6.96	8.29	4.72	6.65	7.17	**4.01**	8.37	7.22	8.62	6.89	7.73	8.23	5.27	7.11	7.30	9.01
11	7.35	8.94	5.68	6.71	5.78	7.08	7.28	7.10	5.10	6.56	**4.17**	6.33	8.01	8.14	7.26	6.05	5.36	5.30	5.03	6.38
12	5.44	9.57	4.53	4.92	4.65	5.94	7.52	5.04	5.19	6.86	7.25	**4.18**	5.96	5.45	4.67	5.77	5.37	4.93	6.52	7.56
13	3.88	7.85	5.69	6.43	6.02	3.84	8.64	4.41	5.36	7.61	7.61	4.60	3.74	7.80	**3.54**	4.99	6.18	5.55	7.50	6.78
14	8.84	12.79	5.95	5.38	5.65	9.49	7.84	7.97	7.37	7.73	9.06	6.45	9.62	**4.35**	8.00	8.91	6.94	6.40	7.14	10.30
15	4.24	9.59	6.13	6.77	6.76	5.07	9.25	5.22	6.58	8.03	8.52	4.51	4.65	7.41	**3.90**	6.12	7.03	5.97	8.38	7.57
16	6.97	8.54	6.47	7.31	6.75	6.33	7.66	6.38	5.88	7.84	6.37	6.87	7.46	8.98	6.97	**4.85**	6.08	5.95	6.03	5.73
17	7.86	8.90	6.17	5.84	6.37	7.23	6.40	6.09	6.26	6.27	7.75	6.73	7.79	7.49	7.07	7.00	**5.31**	6.54	6.79	8.30
18	6.43	9.03	4.91	6.17	5.02	6.54	6.79	6.66	4.19	6.61	4.39	5.17	7.54	7.24	6.60	4.88	4.83	**3.31**	4.30	4.84
19	9.90	11.58	6.43	6.04	5.87	9.36	7.22	7.77	7.08	7.74	8.00	8.31	10.17	7.05	9.05	8.06	6.31	7.22	**5.25**	9.59
20	6.12	7.77	6.94	8.93	7.58	5.70	8.90	7.35	5.92	8.20	5.13	6.66	6.83	10.49	6.87	4.32	6.72	5.53	6.56	**2.30**

Table 7: Euclidian distance matrix for */a:m/-90%*

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**
1	**2.57**	6.78	6.64	8.78	6.38	6.07	8.88	6.78	8.57	7.02	5.22	5.25	3.72	7.41	5.40	6.66	6.94	6.97	8.43	6.53
2	6.09	**5.04**	9.71	11.82	10.06	6.71	9.71	9.07	8.28	8.89	5.14	9.13	7.46	10.87	7.75	6.42	8.10	9.49	11.59	6.18
3	5.52	8.75	**2.82**	5.49	3.45	6.07	6.94	5.33	7.51	4.40	7.75	3.55	3.94	4.27	3.43	8.25	5.06	5.19	6.44	8.18
4	8.13	10.33	4.59	**2.84**	3.93	7.98	7.50	5.77	8.68	6.47	8.97	5.18	6.51	4.25	7.36	10.54	6.95	5.18	3.96	9.48
5	7.90	10.03	4.51	4.27	**3.64**	8.32	7.41	6.55	8.36	6.90	8.31	4.85	6.76	4.84	7.43	10.08	6.90	5.16	4.73	9.12
6	4.75	6.72	5.87	8.04	6.68	**3.88**	7.73	5.33	7.10	5.26	6.79	6.06	4.02	7.03	4.02	6.19	5.42	6.68	8.23	6.58
7	10.28	10.62	6.91	5.95	7.46	8.39	**5.24**	6.83	8.27	6.24	9.55	7.09	9.15	7.02	8.34	9.59	6.72	5.26	6.48	8.56
8	5.51	7.46	5.39	5.86	5.80	4.77	6.87	**3.02**	7.71	4.86	7.74	5.49	3.91	6.52	5.09	6.82	5.23	5.89	5.98	7.03
9	8.16	5.98	6.94	8.94	7.54	6.08	6.94	7.92	**4.06**	7.30	6.02	8.21	8.15	8.59	7.58	7.64	5.53	6.94	9.11	6.62
10	6.40	8.01	6.33	7.70	7.40	5.48	5.26	5.46	7.75	**3.68**	6.85	5.28	5.91	6.87	4.34	5.57	5.18	5.19	8.18	5.24
11	6.21	7.49	7.01	8.35	6.63	7.48	8.38	7.88	7.89	8.02	**5.04**	6.42	6.72	7.75	7.53	8.16	7.34	6.90	8.10	7.05
12	7.42	10.18	5.86	6.36	5.64	8.36	7.43	7.44	8.97	6.68	7.84	**4.68**	6.96	5.56	6.70	9.18	7.25	5.38	6.67	8.29
13	**4.19**	7.78	8.55	10.97	8.75	6.92	10.27	8.55	9.76	8.18	6.28	7.26	5.50	9.09	6.22	6.97	8.37	8.63	10.76	7.02
14	7.06	10.88	3.98	5.24	3.99	7.96	8.37	6.85	9.14	6.22	8.92	4.24	5.78	**3.22**	5.91	10.31	7.11	5.39	6.32	9.53
15	**3.62**	8.66	6.17	8.57	6.25	6.59	8.98	7.16	9.18	6.39	6.57	4.67	4.06	6.36	4.51	7.65	7.14	6.65	8.71	7.40
16	5.44	6.09	7.83	9.39	8.20	5.60	7.07	6.31	7.77	6.63	5.21	6.55	6.21	8.82	6.36	**3.25**	6.31	6.54	8.79	3.59
17	7.14	6.82	5.04	5.94	5.76	4.95	5.07	4.62	5.40	4.67	6.93	6.11	6.03	6.56	5.54	7.08	**3.72**	5.25	6.57	6.51
18	6.51	9.22	3.94	4.13	4.08	6.48	5.98	4.99	7.74	4.36	7.50	3.15	5.26	3.34	5.40	8.36	5.60	**3.04**	4.75	7.26
19	7.96	9.23	6.05	4.91	5.32	7.57	7.06	5.67	8.17	7.07	8.15	5.86	6.96	6.17	8.04	9.05	7.01	5.10	**3.88**	8.12
20	5.49	5.11	6.83	8.68	7.43	4.45	6.37	6.04	6.00	5.77	4.71	6.43	5.79	8.10	5.77	4.57	5.06	6.09	8.52	**3.91**

Table 8: Euclidian distance matrix for */a:n/-90%*

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	**4.37**	7.18	6.29	8.57	8.34	7.78	5.69	8.57	4.80	6.32	5.52	5.87	6.23	6.52	8.03	9.13	5.93	6.90	8.04	8.09
2	8.56	**5.65**	8.67	10.34	10.22	10.00	8.32	10.47	6.61	8.21	7.53	8.07	8.64	8.49	10.49	10.61	9.24	7.40	9.84	7.71
3	6.36	8.34	4.54	9.62	8.66	9.25	6.40	9.29	6.00	**4.41**	6.84	6.81	7.20	7.20	9.18	10.24	5.58	8.42	9.15	7.40
4	7.55	11.33	7.57	**3.16**	3.69	3.82	5.33	4.02	9.08	9.56	6.04	5.40	5.26	5.10	3.58	4.48	6.20	6.68	3.48	9.54
5	8.59	12.62	7.91	**3.15**	3.51	3.87	6.02	3.85	10.28	10.04	7.20	6.10	5.70	5.38	4.17	4.30	6.67	7.90	3.99	10.09
6	8.07	12.06	8.15	3.95	4.36	**3.21**	5.92	4.14	9.51	9.83	7.27	5.80	4.80	5.17	4.88	4.81	6.58	7.40	4.26	10.11
7	5.85	9.31	5.03	5.19	4.77	4.89	**3.87**	4.93	7.07	7.00	5.06	4.14	4.35	4.49	4.99	5.75	4.28	5.77	4.74	7.61
8	10.03	14.03	9.60	5.56	5.65	5.81	7.75	**5.00**	11.87	11.71	9.08	7.82	7.27	7.32	5.88	5.59	8.43	9.58	6.00	11.80
9	4.55	5.09	6.32	8.68	8.52	8.09	5.28	8.78	**3.84**	6.69	4.48	5.38	6.40	6.87	8.00	9.06	6.26	5.48	7.82	7.26
10	6.82	8.45	4.68	10.48	9.48	9.76	6.98	10.16	6.13	**3.97**	7.59	7.23	7.38	7.50	10.18	11.07	6.01	8.81	9.95	7.26
11	5.44	5.95	5.98	8.77	8.44	8.65	5.34	8.98	5.15	6.53	**4.70**	5.80	6.84	7.16	8.20	9.56	5.84	6.66	8.08	7.53
12	5.83	9.75	**4.40**	6.28	5.35	5.64	4.55	6.25	6.88	6.16	6.02	4.60	4.83	5.22	6.03	6.62	4.50	7.02	5.40	7.82
13	6.49	9.46	4.98	4.23	3.98	4.34	4.09	4.76	7.10	6.91	5.19	**3.92**	3.99	4.06	4.80	4.94	4.84	5.49	4.13	7.01
14	6.88	9.29	6.30	4.12	3.86	3.58	4.45	4.69	7.22	7.59	5.27	3.94	3.88	**3.36**	5.05	4.70	5.59	5.17	3.86	7.11
15	7.56	12.39	7.54	5.44	5.61	5.58	6.26	5.73	9.83	9.58	7.34	6.76	6.66	6.58	**5.04**	6.49	6.32	8.82	5.85	10.65
16	8.53	10.88	7.95	4.68	4.55	5.11	5.88	4.76	9.34	9.93	6.69	5.53	5.88	5.80	4.87	**3.91**	7.47	6.34	4.22	8.75
17	4.84	8.65	5.33	5.78	5.65	5.06	4.13	5.97	6.04	6.51	4.71	4.51	4.33	4.70	5.73	6.98	**3.96**	5.90	5.55	7.91
18	8.20	9.30	7.22	5.68	5.65	5.64	5.88	5.92	7.90	8.81	6.27	5.48	5.60	5.90	6.22	5.70	7.25	**4.56**	4.83	7.35
19	9.17	13.18	8.76	4.46	4.23	4.86	6.71	4.82	10.91	10.97	8.05	6.56	6.65	6.60	4.47	4.22	7.67	8.43	**4.11**	10.73
20	9.90	7.42	7.15	11.73	10.78	11.60	9.01	11.43	7.49	6.78	8.66	8.75	9.38	9.13	11.72	11.48	9.50	8.04	11.04	**5.03**

Table 9: Euclidian distance matrix for */i:m/-80%*

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	**4.41**	4.74	4.95	6.24	5.79	4.42	5.48	5.01	5.57	7.51	5.08	6.53	6.21	6.70	4.58	5.33	4.81	5.11	5.51	9.14
2	**6.20**	7.25	7.25	7.96	7.79	6.83	6.94	7.47	7.54	8.40	7.15	8.42	7.86	8.07	6.23	7.40	6.33	6.58	7.51	9.68
3	5.48	5.95	5.34	7.70	7.68	6.09	**4.95**	6.92	5.98	5.42	5.53	5.82	6.23	8.43	5.83	6.86	5.18	5.20	7.28	7.05
4	7.78	8.91	7.43	**5.30**	5.60	6.87	9.34	6.40	9.21	11.34	8.70	9.32	6.91	5.85	6.58	5.95	7.03	7.81	5.68	12.71
5	6.09	5.97	5.28	4.50	**4.07**	4.22	7.48	4.21	7.10	9.28	6.80	6.83	5.26	4.84	5.15	4.43	5.42	6.08	4.19	10.14
6	6.81	7.57	6.91	4.68	4.30	5.07	8.66	5.05	8.77	10.33	8.30	8.57	6.12	**4.18**	5.81	5.19	5.98	7.17	4.28	12.43
7	3.81	6.43	5.15	7.06	7.36	6.23	**3.56**	6.64	6.09	6.09	4.69	6.68	6.42	7.81	4.40	6.35	4.54	4.92	7.02	8.42
8	5.37	5.47	4.37	5.08	5.26	4.49	5.97	4.48	6.28	7.79	5.59	6.45	4.89	6.43	**4.29**	4.78	4.52	4.77	5.24	8.82
9	5.99	4.87	5.54	8.37	8.25	6.57	5.73	7.08	**3.90**	7.83	4.50	6.19	7.74	9.57	6.42	6.91	6.92	5.63	7.79	7.13
10	4.91	7.31	6.87	9.07	8.95	6.82	4.39	8.36	7.81	**3.43**	6.78	6.59	6.53	8.33	6.46	8.50	4.98	6.17	8.39	8.74
11	4.96	5.62	4.73	6.79	6.83	5.75	4.76	6.24	5.28	6.65	**4.70**	5.67	6.09	7.59	5.17	6.00	5.22	5.05	6.37	7.20
12	6.53	6.67	6.47	9.23	9.15	7.32	5.70	8.38	6.72	5.68	6.61	5.38	6.73	9.35	7.42	8.43	6.26	6.80	8.81	**5.27**
13	6.96	7.51	6.31	7.67	8.02	6.97	6.93	7.51	7.68	7.70	7.25	6.87	**6.05**	8.09	6.70	7.68	6.42	6.47	7.72	8.77
14	7.05	8.18	6.79	5.73	5.53	6.14	8.33	6.41	9.02	9.35	8.52	7.35	5.55	**4.62**	6.75	6.46	6.45	7.79	5.54	11.02
15	5.58	6.73	5.98	6.32	6.23	6.14	6.23	6.28	6.61	8.72	5.93	7.78	7.29	7.21	**5.44**	6.02	6.12	5.95	6.04	10.13
16	6.32	6.35	5.63	4.44	4.41	5.18	7.76	4.31	6.69	10.36	6.42	7.95	6.64	6.09	5.00	**4.02**	6.16	6.30	4.57	10.77
17	5.18	7.46	6.36	6.67	6.73	5.89	5.72	6.83	8.11	6.26	7.20	7.20	5.53	6.10	5.31	6.82	**4.33**	5.65	6.36	9.85
18	6.95	6.30	5.63	5.55	5.84	5.64	7.67	5.39	6.22	9.55	6.48	7.62	6.59	7.49	5.56	**5.30**	6.40	5.41	5.55	9.66
19	6.42	7.82	6.75	4.89	4.79	5.70	8.19	5.76	8.29	10.15	7.91	8.32	6.40	4.97	5.69	5.35	6.05	6.77	**4.57**	11.84
20	8.11	6.05	6.83	10.24	9.96	8.07	7.02	8.98	5.97	7.28	6.77	5.06	7.95	10.82	8.81	9.03	8.33	7.58	9.78	**3.06**

Table 10: Euclidian distance matrix for */i:n/-70%*

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	5.19	5.27	5.24	5.62	5.59	5.11	5.06	5.57	5.95	5.49	7.24	6.62	6.43	5.04	5.41	**4.72**	5.20	7.84	4.79	8.32
2	7.43	**3.94**	5.17	7.25	4.48	5.59	4.97	5.60	4.54	7.16	5.93	7.47	7.43	6.76	6.74	5.39	5.32	6.48	4.95	6.56
3	2.22	5.86	3.36	2.54	4.74	5.26	4.74	3.38	5.97	3.32	5.71	2.93	2.56	3.37	**2.03**	3.86	3.38	5.47	3.47	8.52
4	3.84	6.25	3.92	**3.04**	4.76	6.30	4.94	4.27	6.09	3.94	5.44	3.56	3.77	4.04	3.38	4.76	4.22	5.55	4.28	8.27
5	5.81	3.93	3.98	5.59	**3.38**	5.58	4.39	4.32	4.19	5.74	4.84	5.62	5.61	5.44	5.18	4.58	4.14	5.29	4.20	6.32
6	6.01	3.06	4.97	6.76	4.69	**2.95**	4.30	5.22	4.72	5.92	7.33	7.76	7.40	5.30	6.29	3.96	4.73	7.83	4.37	7.21
7	5.83	5.49	4.86	5.35	4.82	6.03	**4.74**	5.52	5.52	5.65	5.59	6.00	6.15	5.26	5.50	5.32	5.04	6.85	5.02	7.38
8	6.71	5.42	5.35	6.74	**5.10**	6.77	6.15	5.22	6.05	6.91	6.06	6.21	5.96	7.02	6.09	6.35	5.43	6.05	5.84	7.95
9	7.20	5.85	6.01	7.21	6.17	6.13	6.33	6.63	**5.21**	7.38	7.14	7.19	6.81	6.89	6.53	6.16	5.83	6.04	5.88	8.18
10	5.37	6.02	5.21	5.17	5.72	6.00	5.11	5.52	6.61	**3.87**	6.78	6.01	6.44	4.87	5.38	5.38	5.41	8.11	5.17	7.89
11	7.37	5.38	4.82	6.04	4.33	7.06	4.88	5.53	5.08	6.43	**3.83**	5.86	6.51	6.57	6.15	6.33	5.32	5.78	5.52	6.05
12	6.39	6.16	5.07	5.59	5.23	7.23	5.86	5.37	5.80	6.26	4.92	**4.54**	4.91	6.47	5.24	6.38	5.42	4.83	5.64	7.64
13	4.74	4.33	**3.87**	4.67	4.19	4.69	3.94	4.26	5.09	4.65	5.59	4.95	4.81	4.53	4.38	4.33	4.05	6.10	3.97	7.37
14	3.81	5.40	3.84	3.90	4.65	4.82	3.73	4.89	5.24	3.92	6.64	5.08	4.75	**2.95**	4.15	3.72	3.83	6.61	3.72	8.46
15	3.52	6.44	4.67	4.36	5.91	5.62	6.05	4.31	6.68	4.68	6.86	4.60	3.96	4.93	**3.41**	4.85	4.40	6.19	4.62	9.06
16	5.88	4.00	4.20	5.28	**3.65**	5.30	3.69	4.97	3.91	5.46	5.59	5.87	6.06	4.79	5.23	3.97	4.52	5.94	3.70	6.43
17	4.23	5.35	3.78	4.38	4.65	5.00	4.37	4.29	5.60	4.44	6.23	5.02	4.57	4.08	4.10	4.18	**3.72**	6.12	3.95	8.38
18	7.68	8.28	6.92	7.57	7.53	8.64	8.38	7.07	7.31	8.28	7.10	5.96	5.46	8.35	6.69	8.36	6.90	**4.25**	7.68	10.35
19	5.96	5.68	4.57	4.91	4.65	6.42	4.51	5.29	5.29	5.33	**4.37**	4.73	5.55	5.24	4.98	5.12	5.13	5.44	4.57	7.01
20	10.13	7.07	7.92	9.17	7.00	9.18	7.55	8.28	7.23	9.03	6.87	9.16	9.85	9.39	9.00	8.35	8.26	8.61	7.85	**4.92**

Table 11: Euclidian distance matrix for */u:m/-70%*

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	**4.83**	6.45	5.59	5.76	5.60	5.15	5.98	5.49	6.65	5.57	4.98	6.49	6.92	4.95	6.46	5.80	4.98	6.31	5.25	9.97
2	8.44	**6.80**	9.32	9.56	7.65	8.51	9.42	8.10	7.97	9.57	7.87	8.67	9.29	9.22	9.08	8.03	8.18	8.07	8.77	7.99
3	3.44	6.45	3.46	3.47	4.23	4.67	4.21	3.71	6.24	5.08	3.65	4.86	6.01	3.58	5.17	3.97	3.80	5.10	**3.00**	9.97
4	3.45	5.54	4.66	3.86	**2.57**	4.40	4.47	3.63	6.19	5.18	3.30	3.73	6.21	4.31	5.21	3.23	3.37	4.32	3.57	8.15
5	5.33	6.09	4.61	5.14	4.17	5.24	6.54	**4.00**	5.46	6.83	4.58	4.36	4.61	6.65	4.26	4.05	4.79	4.03	4.91	8.45
6	5.84	7.17	6.23	6.76	6.17	**4.72**	7.91	6.06	6.10	6.02	5.78	6.59	5.86	6.76	5.91	6.38	5.37	5.96	6.26	10.00
7	5.95	9.49	6.46	5.77	7.53	7.89	**4.67**	6.67	9.86	6.36	6.53	7.68	9.08	6.04	8.00	7.28	7.07	8.71	6.15	12.62
8	5.53	6.32	6.58	6.24	**5.21**	6.08	6.17	5.28	6.92	6.97	5.28	5.90	7.02	6.51	6.30	5.57	5.41	5.86	5.93	8.62
9	7.06	5.52	6.63	7.84	6.72	5.81	8.80	6.38	**4.43**	7.49	6.08	7.32	5.59	7.99	6.48	6.83	6.09	5.65	7.16	9.08
10	5.61	7.10	6.38	6.57	6.10	5.80	6.46	6.48	7.25	**3.62**	4.84	6.84	7.03	6.16	7.26	6.49	5.70	7.11	6.16	9.21
11	5.14	4.83	5.84	6.02	4.26	5.02	6.07	4.50	5.14	5.91	**3.98**	5.08	5.60	6.40	5.54	4.79	4.40	4.61	5.61	7.73
12	7.41	9.00	6.85	6.51	6.81	8.01	7.52	6.35	8.73	8.96	7.37	**6.00**	7.11	8.53	6.16	6.56	7.42	6.47	7.17	10.62
13	8.95	7.96	8.13	8.90	7.91	7.71	10.32	7.47	6.89	9.75	8.08	7.74	**6.01**	10.18	6.50	8.02	7.85	6.43	8.88	9.06
14	4.66	8.31	5.13	5.17	6.71	5.95	5.17	6.21	8.01	4.68	5.42	7.73	8.20	**3.39**	7.67	6.91	5.37	7.88	4.69	12.35
15	5.15	7.72	5.17	5.01	5.49	5.94	6.00	**4.76**	7.39	6.67	5.61	4.95	5.95	6.53	4.80	5.32	5.56	5.45	5.33	10.20
16	4.84	6.57	5.00	4.55	4.65	5.90	5.44	4.43	7.07	6.68	4.96	4.74	6.60	5.95	5.45	**4.22**	5.34	5.45	4.58	8.88
17	5.27	6.53	5.24	5.70	5.48	5.16	6.35	4.81	5.89	6.26	4.93	6.08	5.63	6.00	5.44	5.79	**4.67**	5.40	5.46	9.78
18	7.31	6.74	6.78	7.37	6.15	6.62	8.56	5.72	6.03	8.75	6.60	5.87	5.04	8.77	5.06	6.07	6.50	**4.71**	7.30	8.36
19	3.23	7.22	3.48	2.91	4.70	5.28	3.88	4.04	7.19	5.22	4.27	5.07	6.72	3.63	5.55	4.09	4.60	5.87	**2.56**	10.70
20	9.19	6.67	10.40	10.39	7.39	8.86	10.33	8.61	8.22	9.90	8.22	8.21	9.38	10.48	9.26	7.85	8.80	7.95	9.57	**3.68**

Table 12: Euclidian distance matrix for */u:n/-75%*

Condition III: Mobile network vs live recording

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**
1	8.10	8.99	8.52	8.76	9.82	7.87	8.40	8.26	9.28	8.63	8.09	8.11	7.63	8.69	**7.46**	8.22	8.09	7.95	8.23	7.83
2	9.05	**7.48**	9.40	9.89	10.37	8.44	8.85	9.03	9.18	9.60	8.68	9.76	8.11	10.74	9.05	8.62	8.48	8.90	9.46	8.31
3	7.94	9.96	7.64	7.43	8.14	8.98	8.10	8.31	9.33	7.87	8.12	7.41	8.24	**7.36**	8.10	8.91	8.02	8.13	7.94	8.98
4	7.42	9.34	7.21	**7.02**	8.25	8.38	7.56	7.60	8.82	7.52	7.84	7.03	7.69	7.07	7.36	8.34	7.60	7.77	7.61	8.55
5	8.81	9.61	8.18	8.32	8.54	9.16	8.82	9.42	8.59	9.20	7.85	8.41	8.79	**7.49**	8.67	9.18	8.13	7.69	7.60	8.78
6	8.71	10.25	9.63	9.87	11.48	7.90	9.64	8.89	10.49	9.30	9.07	8.67	7.82	9.71	**7.65**	8.87	9.36	8.91	9.11	8.53
7	7.85	7.78	7.79	7.92	8.48	8.58	6.98	8.00	8.56	7.15	7.39	7.94	7.72	8.44	8.45	7.84	**6.88**	7.65	8.03	7.50
8	7.69	9.20	8.30	8.32	9.41	7.80	8.06	7.60	9.70	7.87	8.32	7.70	**7.31**	8.76	7.32	8.16	8.12	8.28	8.46	8.05
9	8.28	8.41	8.10	8.56	9.41	7.75	8.57	8.74	7.31	9.27	7.12	8.27	7.95	7.77	7.66	8.00	7.58	**6.82**	6.99	7.39
10	7.89	8.98	8.36	8.33	9.17	8.36	7.37	8.06	9.77	**6.67**	7.74	7.76	7.36	8.30	8.20	8.29	7.56	7.75	8.19	7.69
11	7.33	8.14	7.26	7.71	8.11	7.12	7.87	8.20	6.95	8.50	**5.91**	7.21	7.45	7.07	7.29	6.97	6.71	5.96	6.22	6.35
12	8.08	9.98	8.10	8.34	9.40	8.46	8.55	8.82	9.41	8.54	7.97	**7.43**	7.96	7.70	7.45	8.59	8.12	7.69	7.94	8.04
13	8.06	8.92	8.91	9.27	10.29	7.67	8.60	8.35	10.03	8.50	8.40	8.22	**7.13**	9.74	7.52	8.16	8.44	8.50	8.94	7.86
14	8.24	10.58	7.89	7.90	9.11	9.13	8.62	8.86	9.24	8.39	8.28	7.37	8.55	**6.98**	7.83	9.18	8.41	7.97	7.80	9.11
15	7.70	9.60	8.32	8.58	10.02	7.39	8.69	8.14	9.40	8.60	8.04	7.53	7.23	8.51	**6.84**	8.04	8.29	7.92	8.09	7.79
16	7.04	8.87	7.48	7.53	8.53	7.55	7.52	7.31	9.00	7.36	7.54	**6.72**	6.97	7.84	6.73	7.58	7.48	7.54	7.69	7.47
17	6.14	7.94	5.98	5.97	7.07	6.89	6.22	6.50	7.16	6.23	5.92	5.76	6.36	**5.67**	6.27	6.78	5.90	5.98	5.95	6.77
18	6.92	8.09	6.84	7.32	8.20	6.68	7.46	7.84	6.68	7.83	5.64	6.42	6.63	6.07	6.24	6.71	6.41	**5.08**	5.52	5.58
19	5.45	7.89	5.55	5.57	7.10	6.13	6.12	5.87	6.90	6.15	5.72	**5.13**	5.81	5.29	5.39	6.13	5.88	5.55	5.39	6.31
20	6.25	6.59	6.91	7.38	8.45	6.06	6.52	6.43	7.59	6.80	6.21	6.59	5.63	7.82	6.08	5.93	6.15	6.11	6.67	**5.34**

Table 13: Euclidian distance matrix for */a:m/-50%*

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**
1	6.70	7.76	7.24	7.77	8.76	6.55	7.83	7.80	9.04	7.17	7.07	7.22	**6.54**	7.18	6.87	7.49	7.77	7.60	7.62	7.47
2	7.64	**6.22**	8.51	8.79	9.95	6.80	7.69	8.06	7.03	8.18	7.38	8.67	7.72	9.38	7.59	7.38	7.40	8.28	8.68	6.62
3	7.51	8.69	**6.75**	7.17	7.59	8.42	7.76	8.13	9.33	7.29	8.20	6.89	7.64	7.25	6.96	8.47	7.43	7.91	7.58	9.09
4	6.89	8.83	5.79	**5.75**	6.09	8.40	7.24	7.31	9.11	7.50	7.66	6.43	7.58	5.77	6.84	8.38	7.21	6.99	5.99	9.27
5	8.12	9.52	7.77	7.97	8.12	8.95	8.29	9.11	9.66	8.58	8.05	7.47	8.50	**7.01**	8.08	8.99	8.33	7.75	7.86	9.02
6	5.95	6.19	6.67	7.14	8.53	**5.37**	6.78	6.28	7.61	6.22	6.54	7.02	5.75	7.39	5.88	6.12	6.67	7.14	6.92	6.45
7	8.34	8.74	7.69	7.16	7.38	9.21	6.79	7.19	8.63	7.93	8.84	7.73	8.75	8.27	**8.08**	7.85	7.27	7.68	6.93	8.80
8	7.03	7.99	6.86	6.99	7.81	7.52	7.26	**6.36**	9.30	6.86	8.26	7.24	6.92	7.46	7.04	7.34	7.56	7.74	6.77	8.47
9	8.09	7.60	8.09	8.38	9.33	7.98	7.79	8.57	**7.36**	8.26	7.78	8.25	8.37	8.68	7.50	8.29	7.42	7.99	8.30	7.90
10	7.63	8.56	7.19	7.78	8.41	7.86	7.54	7.33	10.20	**5.85**	8.95	7.01	6.92	7.92	7.03	7.63	7.64	8.34	7.97	8.59
11	5.94	6.14	5.81	6.39	7.59	6.14	6.34	6.64	7.02	6.12	**6.06**	6.26	6.10	6.55	5.59	6.77	5.83	6.65	6.57	6.72
12	6.47	8.26	6.34	7.01	7.48	7.30	6.99	7.86	9.27	6.51	7.15	**5.49**	6.53	5.77	6.26	7.37	7.08	6.66	6.91	7.87
13	7.64	8.40	7.65	8.46	9.13	7.92	8.79	8.31	10.56	7.40	8.89	7.93	**7.03**	8.52	7.52	8.82	8.51	9.41	8.95	9.53
14	7.53	9.09	6.79	7.31	7.62	8.60	8.00	8.37	9.90	7.47	8.31	6.89	7.69	**6.74**	7.19	8.80	7.86	7.94	7.58	9.36
15	7.87	8.71	8.35	9.07	10.12	7.64	8.56	9.00	9.86	7.66	8.18	7.90	7.62	8.35	**7.59**	8.14	8.47	8.45	8.86	8.18
16	7.25	7.87	8.04	8.37	9.56	**6.58**	7.72	7.64	9.07	7.09	7.87	7.66	6.96	8.43	7.02	6.59	7.85	7.87	8.04	6.99
17	7.28	7.13	6.99	7.19	8.31	7.14	6.68	6.74	7.70	**6.33**	7.64	7.25	7.13	8.15	6.56	6.89	6.39	7.59	7.29	7.40
18	6.70	8.00	6.90	7.44	8.51	6.75	6.97	7.83	8.26	6.62	6.58	6.38	6.89	6.24	6.47	6.95	7.15	**6.14**	6.82	6.59
19	5.43	7.39	**4.97**	5.21	6.45	6.33	6.28	6.06	7.95	6.00	6.23	5.68	6.01	5.16	5.21	6.69	6.20	5.85	5.13	7.46
20	7.16	7.02	8.17	8.74	10.20	5.98	7.24	7.75	8.01	6.77	7.16	7.64	6.91	8.53	6.81	6.07	7.32	7.35	8.13	**5.68**

Table 14: Euclidian distance matrix for */a:n/-70%*

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**
1	**4.15**	5.22	4.82	5.76	6.50	5.23	4.67	4.73	4.93	5.56	4.65	4.70	4.74	5.37	4.60	5.53	4.78	5.41	5.84	6.88
2	7.33	7.18	7.54	7.46	7.79	7.29	7.49	7.14	7.48	8.31	7.48	6.89	7.60	8.36	7.56	7.85	8.55	**6.85**	7.41	7.35
3	7.56	7.66	**6.56**	6.92	6.82	7.87	7.02	6.94	8.62	7.87	7.48	6.77	7.18	7.13	7.68	7.82	8.10	7.82	7.46	8.24
4	7.21	8.72	7.39	6.35	6.34	7.18	7.12	6.72	9.02	9.45	7.70	7.12	7.75	7.71	6.75	6.54	8.43	8.52	**6.18**	9.59
5	6.13	7.16	6.21	4.83	**4.63**	5.55	6.14	5.40	7.70	8.38	6.71	5.56	6.59	6.89	5.87	5.67	8.04	6.48	4.79	7.35
6	6.57	7.87	7.22	6.13	6.26	**5.82**	6.81	6.35	7.77	8.79	7.19	6.61	7.14	7.73	6.32	6.40	8.08	7.05	5.97	8.18
7	5.81	6.43	5.63	6.07	6.31	6.28	**5.38**	5.66	6.86	7.30	5.86	5.55	6.13	6.54	5.71	6.12	6.62	6.41	5.86	7.45
8	7.75	9.30	7.86	7.01	**6.81**	7.75	7.68	7.10	9.79	10.32	8.30	7.73	8.66	8.50	7.20	7.02	9.08	9.18	6.83	9.98
9	7.03	6.49	6.89	8.94	9.63	7.93	6.89	7.75	**6.03**	6.24	6.26	7.09	6.83	7.02	7.37	8.42	6.17	6.66	8.77	7.61
10	8.30	7.94	7.03	9.64	10.09	9.45	7.65	8.63	7.98	**6.26**	7.38	8.01	7.31	6.68	8.79	9.82	6.57	8.55	10.07	9.14
11	8.13	7.37	7.86	9.89	10.58	9.37	7.63	8.85	7.02	7.39	6.90	8.07	8.16	8.16	8.19	9.22	**6.87**	7.85	9.51	8.58
12	6.87	6.90	6.28	6.47	6.47	7.30	6.43	6.50	7.69	7.87	6.83	**6.15**	6.98	7.14	6.86	6.95	7.78	6.92	6.48	7.55
13	6.27	6.83	6.54	6.29	6.43	6.05	6.35	6.18	6.82	7.38	6.35	6.07	6.43	6.81	6.20	6.57	7.02	6.24	**5.99**	7.09
14	6.98	6.90	6.93	6.97	7.33	6.35	6.79	6.82	6.76	7.64	6.51	6.45	7.01	6.97	6.80	7.10	7.49	**5.66**	6.84	6.12
15	7.81	8.83	7.87	8.27	8.47	8.41	7.49	7.95	8.77	9.33	7.72	8.00	8.41	8.30	**7.27**	7.51	7.89	8.98	7.62	10.22
16	7.03	7.94	7.47	6.29	6.23	6.59	6.99	6.58	8.31	9.64	7.42	6.78	7.98	8.36	6.36	5.88	8.76	7.37	**5.63**	8.16
17	5.96	6.77	5.91	6.68	7.28	6.45	**5.60**	6.25	6.38	6.48	5.70	6.08	5.88	6.17	5.96	6.75	5.65	6.65	6.87	8.04
18	7.14	6.76	7.33	7.42	7.69	6.78	6.93	7.23	6.61	8.04	6.70	6.69	7.40	8.02	6.97	7.39	7.98	**5.71**	6.76	6.35
19	6.15	7.38	6.69	5.50	5.39	5.74	6.32	5.86	7.52	8.71	6.78	5.97	7.01	7.30	5.74	5.55	8.06	6.79	**4.86**	7.91
20	8.45	7.02	7.96	8.54	8.59	8.51	8.28	8.13	8.10	8.47	8.14	7.44	8.32	8.87	8.82	9.18	9.45	6.95	8.44	**6.41**

Table 15: Euclidian distance matrix for */i:m/-60%*

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**
1	**5.45**	6.50	6.14	6.11	6.78	6.28	6.32	6.02	7.17	6.38	6.68	7.17	5.91	6.56	5.88	6.55	6.61	6.97	6.08	8.87
2	6.24	**4.93**	5.08	5.58	5.41	4.99	6.57	6.24	5.47	7.02	6.16	5.93	5.63	6.65	6.15	6.62	6.77	5.46	5.55	6.08
3	7.19	7.32	7.07	7.48	7.81	7.58	6.77	7.43	7.83	6.78	7.26	6.97	**6.65**	8.00	7.45	8.52	7.21	7.48	7.93	7.81
4	7.29	8.79	7.47	**6.91**	7.22	8.18	8.84	7.62	9.73	9.54	9.12	9.00	7.97	8.33	7.63	7.01	9.51	9.42	7.53	11.36
5	5.95	6.75	6.06	5.91	**5.59**	6.43	7.30	6.57	8.09	7.21	7.85	6.53	5.96	6.89	6.67	6.72	7.84	7.86	6.42	9.39
6	6.05	6.11	5.79	5.74	5.70	**5.57**	7.45	6.48	7.22	7.56	7.43	7.07	6.16	6.53	6.43	6.29	7.70	7.22	5.58	9.31
7	6.78	6.79	6.80	7.29	8.29	7.15	5.33	6.55	6.53	5.95	5.35	7.15	6.19	6.95	6.52	8.05	**5.22**	6.15	7.40	6.61
8	5.94	5.72	5.47	5.61	**5.36**	5.51	6.83	6.12	6.91	7.03	6.77	6.23	5.67	7.25	6.13	6.54	7.27	6.72	6.16	7.89
9	7.58	6.74	7.20	7.76	8.34	7.24	6.81	7.55	**6.42**	7.37	6.65	7.60	7.17	7.69	7.30	8.27	6.83	6.43	7.47	7.20
10	6.24	6.24	6.66	6.99	7.52	6.42	5.73	6.79	6.72	**4.21**	6.14	5.83	5.31	6.38	6.79	8.16	5.39	6.82	6.96	7.45
11	6.48	5.74	5.84	6.24	6.78	6.08	5.79	6.50	**5.52**	6.41	5.63	6.16	5.90	6.26	6.32	7.01	5.84	5.64	6.16	6.17
12	9.77	8.73	9.41	9.85	9.48	9.50	9.19	10.15	9.01	8.68	9.56	7.82	8.65	10.40	10.20	11.15	9.46	9.32	10.35	**7.45**
13	8.16	6.74	7.50	7.85	7.24	7.13	8.40	8.47	7.37	7.78	8.22	**6.49**	6.98	8.55	8.52	9.12	8.42	7.92	7.99	7.17
14	8.60	8.91	8.60	8.57	8.81	8.98	8.91	9.13	9.15	8.91	9.24	8.41	**8.17**	8.23	9.15	9.29	9.10	9.41	8.50	10.08
15	5.79	6.45	5.76	5.79	6.50	6.16	6.25	5.89	6.93	6.96	6.37	7.35	6.10	6.22	**5.72**	5.98	6.62	6.54	5.76	8.75
16	5.68	6.88	5.66	5.34	5.89	6.25	6.99	5.76	7.63	8.06	7.06	7.85	6.47	6.60	5.69	**5.21**	7.57	7.27	5.54	9.78
17	6.45	6.92	6.61	6.65	7.17	6.81	6.61	6.77	7.41	6.63	6.85	7.09	**6.27**	6.86	6.75	7.37	6.76	7.28	6.81	8.50
18	7.27	6.56	6.80	7.44	8.06	6.75	6.44	7.11	6.28	6.95	6.20	7.63	6.93	7.42	6.85	7.99	6.48	**5.93**	7.12	7.42
19	7.16	7.80	6.71	6.25	**6.21**	7.24	8.66	7.59	8.55	9.16	8.72	7.90	7.29	7.53	7.51	6.77	9.27	8.58	6.52	10.15
20	9.21	7.17	8.13	8.93	8.63	7.98	8.28	9.09	7.11	8.25	7.99	7.19	7.89	9.19	9.06	10.13	8.27	7.34	8.97	**5.91**

Table 16: Euclidian distance matrix for */i:n/-55%*

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	6.11	8.02	6.26	6.33	6.58	7.36	6.90	6.24	8.41	6.36	7.27	6.47	6.13	6.47	6.60	**6.04**	7.27	7.67	6.37	8.26
2	5.07	**4.27**	4.77	5.61	4.75	4.95	4.50	5.42	4.70	6.32	4.45	5.05	4.70	6.18	4.54	5.64	4.90	4.55	4.73	4.92
3	6.07	6.93	5.54	6.17	5.83	7.02	5.97	5.81	6.88	6.35	5.89	**5.49**	5.68	5.93	6.01	5.81	5.94	5.77	5.66	7.25
4	6.42	6.97	6.05	6.06	6.22	7.43	6.05	6.41	7.07	6.42	5.71	**5.64**	5.96	6.46	5.94	6.61	5.91	6.06	5.85	6.67
5	5.81	4.60	5.13	5.90	4.88	5.93	4.51	5.82	5.08	6.57	4.21	4.89	5.22	6.44	4.76	6.13	4.93	**4.12**	5.16	4.75
6	4.29	4.57	4.40	5.14	4.66	4.13	4.48	4.84	4.55	5.64	4.60	4.71	**4.04**	5.28	4.25	4.98	4.57	4.63	4.17	5.54
7	6.09	**5.44**	5.90	6.47	5.80	5.84	5.52	6.72	5.51	7.32	5.70	6.08	5.84	6.68	5.56	6.40	5.94	5.66	5.75	5.99
8	5.64	6.38	5.35	6.06	5.57	6.47	5.56	**5.24**	6.76	6.22	5.49	5.26	5.47	6.18	5.55	5.79	5.90	5.56	5.49	6.52
9	6.77	5.91	6.28	7.06	6.20	6.61	6.13	7.13	5.44	7.63	5.99	6.37	6.28	6.94	6.22	6.80	6.09	**5.34**	6.15	6.68
10	6.26	7.95	6.26	5.82	6.69	7.60	6.97	6.43	8.18	**5.07**	6.77	6.42	5.76	6.35	6.58	6.37	6.62	7.53	6.10	7.79
11	7.58	7.25	7.00	7.51	6.94	7.98	6.78	7.47	7.56	7.82	**6.67**	6.67	7.05	7.72	6.93	7.57	7.11	6.69	7.08	7.38
12	8.96	7.90	8.52	8.92	8.47	9.13	8.02	8.96	8.20	9.30	**7.51**	8.06	8.35	9.27	8.04	9.31	8.08	7.59	8.41	7.81
13	6.31	5.68	6.21	6.53	6.27	6.37	5.96	6.64	6.14	6.89	**5.49**	6.14	5.76	7.20	5.73	7.03	5.86	5.88	6.01	5.91
14	5.80	7.53	5.52	5.63	5.70	7.21	6.07	5.91	7.54	5.71	6.46	5.66	5.60	**5.35**	6.00	5.52	6.10	6.67	5.48	7.57
15	6.97	8.94	6.69	6.98	7.03	8.49	7.22	6.43	9.34	6.66	7.52	**6.36**	6.72	6.85	7.08	6.62	7.77	7.73	6.88	8.85
16	4.85	6.73	4.84	4.89	5.17	6.28	5.45	5.04	6.93	5.06	5.72	4.94	4.67	4.87	5.12	**4.65**	5.70	5.97	4.70	6.82
17	5.86	6.73	5.62	5.92	5.82	6.63	5.92	5.94	6.80	6.19	5.96	5.65	**5.49**	6.06	5.81	5.93	5.82	6.22	5.59	6.99
18	9.42	7.85	8.66	9.97	8.50	8.88	8.16	9.20	8.07	10.38	7.93	8.36	8.78	9.92	8.52	9.40	8.85	**7.10**	8.91	8.67
19	6.08	6.98	5.80	6.26	6.09	7.22	6.04	6.10	6.76	6.54	5.90	5.48	5.61	5.75	5.77	6.14	5.76	5.57	**5.42**	7.15
20	7.09	5.72	6.76	6.96	6.49	7.13	5.86	7.39	7.14	7.44	5.58	6.38	6.49	8.42	5.81	7.63	7.18	6.51	6.77	**4.90**

Table 17: Euclidian distance matrix for */u:m/-45%*

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	6.67	7.95	6.57	6.56	7.04	7.14	6.96	6.70	6.68	7.34	7.30	7.13	6.81	6.77	7.12	6.64	6.55	7.61	**6.34**	9.61
2	8.79	**7.41**	8.72	9.06	8.27	8.73	8.53	8.50	8.51	9.18	8.08	8.78	8.55	9.91	8.90	9.01	9.27	8.87	8.95	8.12
3	6.24	6.75	5.96	6.03	6.27	6.56	6.01	5.98	5.97	7.01	6.25	6.33	6.25	6.33	6.16	6.28	6.09	6.62	**5.79**	8.36
4	5.97	7.03	5.60	5.43	5.56	6.52	5.75	5.64	6.05	6.73	6.36	5.28	5.90	5.99	5.90	5.68	5.88	6.33	**5.20**	8.18
5	6.70	7.25	6.40	6.57	6.48	7.03	6.64	6.40	6.56	7.38	6.85	6.44	6.49	7.10	6.56	6.58	6.68	6.73	**6.32**	8.07
6	6.25	6.83	6.20	6.46	6.55	5.87	6.62	6.37	6.00	6.87	6.24	6.71	**5.81**	7.01	5.92	6.64	6.18	6.26	6.38	8.06
7	7.05	8.31	7.11	6.94	7.01	8.09	**6.75**	6.80	7.72	7.38	7.64	7.40	7.36	7.21	7.86	6.93	7.33	8.46	7.25	9.84
8	6.95	7.04	7.12	7.35	7.01	7.01	6.88	**6.67**	7.08	7.99	7.23	6.93	6.75	7.89	6.69	7.30	7.25	6.86	7.16	7.84
9	7.85	6.76	7.58	8.03	7.66	7.06	7.88	7.88	**6.75**	8.04	6.88	7.91	7.04	9.04	7.32	8.34	7.75	6.98	7.93	7.62
10	5.48	6.33	5.64	5.66	5.71	5.60	5.92	6.06	5.66	**4.77**	5.48	6.37	5.30	5.83	5.97	6.13	5.81	6.56	5.98	7.97
11	6.32	6.83	6.20	6.26	6.00	6.13	6.38	6.39	6.14	6.39	6.13	6.06	**5.62**	7.19	6.09	6.54	6.30	6.23	6.35	7.40
12	7.78	9.01	7.59	7.46	7.30	8.38	7.70	7.41	8.27	8.62	8.39	**7.03**	7.73	7.94	7.76	7.45	7.92	8.11	7.37	9.40
13	8.49	9.60	8.28	8.42	8.28	8.43	9.03	8.52	8.32	8.75	8.79	8.18	**7.79**	9.25	8.51	8.45	8.50	8.68	8.09	9.83
14	6.76	8.02	6.43	6.38	7.38	7.61	6.90	6.65	6.61	7.65	7.35	7.65	7.41	**5.65**	7.34	6.65	6.50	8.18	6.06	10.49
15	6.85	8.46	6.83	6.84	6.96	7.40	7.15	6.75	7.36	7.61	7.68	7.04	6.98	6.87	7.18	6.81	7.12	7.70	**6.57**	9.64
16	7.03	7.04	6.91	6.90	6.88	7.18	6.69	6.78	7.01	7.86	7.16	6.60	6.97	7.36	**6.57**	7.03	7.00	6.68	6.76	7.84
17	6.26	6.66	6.16	6.23	6.28	6.37	6.24	6.06	6.09	7.04	6.47	6.19	**5.97**	6.78	6.01	6.41	6.22	6.28	6.11	7.83
18	7.90	7.42	7.61	7.92	7.27	7.21	7.79	7.61	7.33	8.61	7.67	6.72	6.95	8.80	6.70	8.04	8.01	**6.24**	7.30	7.00
19	6.59	7.85	6.15	6.23	6.53	7.14	6.60	6.12	6.48	7.74	7.20	6.40	6.58	6.47	6.59	6.17	6.45	7.07	**5.62**	9.17
20	9.06	6.71	9.12	9.44	8.32	7.93	8.64	8.97	8.72	9.39	7.96	8.21	8.27	10.34	7.63	9.61	9.49	7.13	9.26	**5.32**

Table 18: Euclidian distance matrix for */u:n/-55%*

Section B: Euclidian distance Matrix for Group II consisting of 10 speakers

Condition I: Live vs live recording

	1	2	3	4	5	6	7	8	9	10
1	**2.23**	2.91	3.13	3.25	5.09	5.03	3.13	3.16	5.87	4.25
2	3.15	**2.12**	3.47	3.22	5.09	5.69	4.05	4.37	4.61	5.81
3	2.95	4.45	**2.62**	2.77	3.81	6.97	4.60	4.29	5.71	6.28
4	3.73	4.46	3.35	**2.94**	5.83	5.91	3.63	4.51	5.59	5.58
5	4.93	5.79	5.03	5.32	**2.58**	9.26	6.80	6.28	7.99	7.54
6	6.21	6.39	6.67	6.10	9.30	**2.28**	5.55	6.01	6.24	6.64
7	4.08	4.71	4.53	4.38	7.02	5.14	**2.56**	3.77	7.09	3.45
8	3.43	4.69	4.26	4.65	6.03	6.22	3.42	**2.45**	7.48	4.68
9	6.45	5.50	6.35	5.37	7.85	6.03	7.12	7.73	**2.47**	9.34
10	5.22	5.93	5.46	5.68	7.91	5.93	4.19	4.81	8.53	**2.45**

Table 19: Euclidian distance matrix for */a:m/-100%*

	1	2	3	4	5	6	7	8	9	10
1	**2.73**	3.71	3.25	3.75	4.17	5.36	3.43	3.46	6.41	5.79
2	3.14	**1.90**	3.33	3.92	5.62	4.79	2.83	3.92	4.95	5.84
3	2.71	3.48	**2.06**	2.33	5.04	4.21	2.51	3.81	5.46	5.38
4	3.62	4.72	3.25	**2.82**	5.53	5.17	4.18	4.61	6.68	5.83
5	4.26	4.75	4.46	4.84	**2.12**	7.33	4.64	4.66	7.37	6.72
6	5.30	5.68	5.14	4.97	8.34	**1.97**	4.44	5.92	6.75	5.60
7	2.80	3.58	2.53	3.06	4.90	4.20	**2.51**	3.40	6.35	4.42
8	3.16	3.66	3.60	4.03	5.66	3.72	2.94	**2.93**	7.04	4.45
9	5.84	5.63	5.97	6.21	7.97	6.48	5.08	6.54	**2.65**	9.31
10	4.75	5.38	4.34	4.13	6.32	5.19	5.12	5.58	8.82	**2.32**

Table 20: Euclidian distance matrix for */a:n/-100%*

	1	2	3	4	5	6	7	8	9	10
1	**2.40**	2.79	3.12	2.95	2.85	4.63	4.66	4.21	5.70	7.22
2	2.91	**2.17**	3.65	3.58	3.21	4.97	4.90	4.71	5.85	7.31
3	4.23	4.95	**4.17**	4.63	4.63	4.34	4.41	4.36	5.25	5.94
4	3.25	4.05	3.00	**2.09**	3.31	4.06	2.84	2.88	5.10	5.56
5	3.24	2.81	2.94	3.29	**2.44**	4.64	4.59	4.42	5.42	6.53
6	3.37	4.83	3.89	4.72	5.05	**2.22**	3.55	3.36	2.41	4.57
7	4.11	5.01	4.19	4.18	4.42	4.57	**3.40**	3.70	5.14	6.01
8	3.61	5.32	4.35	4.20	4.64	4.67	3.60	**3.12**	5.46	6.05
9	6.07	7.14	6.42	7.08	7.63	4.47	5.37	5.32	**3.98**	5.03
10	5.09	5.74	5.35	5.36	5.54	4.80	4.08	4.12	5.25	**3.27**

Table 21: Euclidian distance matrix for */i:m/-100%*

	1	2	3	4	5	6	7	8	9	10
1	**2.31**	4.11	3.74	3.37	4.15	5.00	4.46	3.74	5.50	4.05
2	4.06	**2.34**	4.28	4.57	3.11	3.55	4.99	4.57	4.10	4.49
3	2.89	3.19	**2.35**	2.69	3.72	3.36	3.13	3.05	3.67	3.87
4	2.60	3.49	2.86	**2.06**	3.86	3.89	3.51	3.32	4.12	3.98
5	5.13	4.10	4.98	5.26	**3.61**	4.74	5.56	5.48	5.19	5.49
6	5.45	5.03	5.88	6.10	5.86	**4.70**	5.86	5.36	5.21	5.47
7	3.43	5.01	3.22	3.20	5.63	5.00	**2.89**	3.53	4.60	3.90
8	3.99	6.17	4.49	4.67	6.91	6.16	4.00	**3.91**	6.15	5.16
9	5.41	4.99	4.85	4.66	6.28	3.83	4.41	4.88	**2.83**	4.61
10	2.43	3.99	3.83	3.56	4.90	4.39	4.02	3.39	4.28	**2.02**

Table 22: Euclidian distance matrix for */i:n/-100%*

	1	2	3	4	5	6	7	8	9	10
1	**1.94**	3.61	2.77	3.46	3.27	3.04	3.68	3.31	5.85	3.97
2	3.33	**2.45**	2.85	4.03	3.06	3.25	3.49	4.22	4.89	4.20
3	3.91	4.02	**3.50**	4.40	4.04	4.15	4.29	4.10	5.66	5.22
4	3.43	4.13	3.04	**2.16**	3.63	4.06	4.85	4.34	5.43	3.43
5	2.73	2.52	2.03	3.19	**1.63**	3.17	2.55	3.74	5.17	4.16
6	2.93	2.49	3.25	3.90	3.46	**1.91**	4.09	4.83	3.41	4.82
7	3.82	4.20	3.57	4.94	3.72	4.41	**3.20**	3.94	6.95	4.99
8	3.11	3.93	2.85	4.54	3.37	4.14	3.14	**2.20**	6.52	4.41
9	5.68	4.36	5.65	5.29	5.52	4.37	6.60	7.42	**2.21**	6.82
10	4.71	5.42	4.56	4.20	4.79	5.58	5.76	4.99	7.49	**2.47**

Table 23: Euclidian distance matrix for */u:m/-100%*

	1	2	3	4	5	6	7	8	9	10
1	**1.62**	4.66	2.83	2.70	3.06	3.11	3.47	3.32	4.54	3.80
2	4.52	**2.06**	3.74	4.40	4.37	4.18	4.26	3.84	4.02	5.07
3	3.61	5.27	**3.57**	3.74	3.89	3.61	4.35	4.36	4.44	4.69
4	2.95	4.41	2.69	**2.24**	3.07	3.38	3.10	3.91	4.27	4.13
5	3.08	5.03	3.35	2.91	**2.91**	4.14	3.49	4.04	5.45	4.22
6	3.49	4.95	3.45	3.69	3.74	**2.61**	3.89	4.42	4.29	4.56
7	2.90	4.42	3.02	3.02	3.25	3.31	**2.52**	3.32	5.01	4.38
8	3.58	4.67	3.93	4.32	4.23	4.24	3.12	**2.70**	5.73	5.66
9	3.97	4.30	3.14	3.90	4.47	3.31	4.45	4.21	**2.59**	4.89
10	3.79	5.37	4.31	3.88	3.71	4.20	4.68	5.43	5.73	**2.70**

Table 24: Euclidian distance matrix for */u:n/-100%*

Condition II: Mobile network vs Mobile network recording

	1	2	3	4	5	6	7	8	9	10
1	**4.04**	9.89	6.79	8.06	8.02	5.41	10.29	5.74	7.92	9.10
2	9.01	**4.72**	9.33	9.49	9.48	6.76	8.47	7.23	7.01	8.31
3	5.74	10.35	**3.61**	4.26	3.96	6.28	7.01	5.13	5.36	6.40
4	7.92	11.67	4.97	**4.38**	5.08	8.04	7.07	6.02	6.65	6.94
5	7.01	11.56	3.86	4.09	**3.79**	7.71	7.54	6.59	5.94	7.27
6	3.91	7.72	6.35	7.48	7.59	**3.71**	8.65	4.60	6.69	7.22
7	8.93	9.35	6.58	5.98	6.59	8.19	**3.85**	6.94	6.27	3.87
8	6.51	9.91	6.18	5.98	7.04	6.17	7.90	**4.32**	7.26	6.90
9	6.40	8.52	5.27	5.72	5.33	6.03	7.44	6.23	**4.65**	7.28
10	8.61	9.87	6.32	5.48	6.68	8.04	4.94	6.16	7.05	**3.38**

Table 25: Euclidian distance matrix for */a:m/-100%*

	1	2	3	4	5	6	7	8	9	10
1	**3.41**	6.81	8.01	9.19	7.29	6.95	9.47	7.32	8.95	7.91
2	6.16	**3.83**	9.35	10.62	9.16	6.24	9.17	8.16	7.11	8.59
3	4.26	8.32	**3.83**	5.78	4.24	5.68	7.05	5.19	7.57	4.50
4	8.79	11.41	5.19	**3.77**	4.38	8.90	8.33	6.16	9.36	7.52
5	7.37	10.56	4.43	4.60	**3.68**	8.59	8.01	6.69	8.69	7.22
6	4.27	5.87	6.18	7.58	6.55	**3.68**	7.41	5.18	6.57	5.34
7	10.07	10.72	6.96	5.79	7.23	8.51	**5.24**	6.85	8.11	6.47
8	6.32	7.98	6.38	6.57	6.71	5.17	6.88	**3.73**	8.01	4.91
9	8.72	6.89	7.54	9.14	7.88	6.89	7.79	8.42	**4.78**	8.27
10	6.12	7.04	6.11	6.73	6.75	4.84	4.51	5.04	6.53	**3.33**

Table 26: Euclidian distance matrix for */a:n/-100%*

	1	2	3	4	5	6	7	8	9	10
1	**3.27**	7.43	5.57	7.98	7.71	7.81	4.82	8.21	5.15	6.65
2	7.87	**4.23**	8.73	12.70	12.24	12.73	8.83	12.97	5.87	8.94
3	5.98	7.45	**3.72**	7.69	6.69	7.81	4.38	7.86	6.30	4.90
4	7.29	9.73	7.89	**3.11**	3.54	3.88	4.99	4.09	8.27	9.45
5	8.61	11.55	8.60	**3.56**	3.71	4.02	6.13	4.34	9.90	10.05
6	6.89	9.83	7.92	3.20	3.63	**3.06**	5.08	4.23	8.03	9.15
7	6.00	9.08	5.92	4.76	4.38	4.75	**3.98**	4.82	7.27	7.43
8	8.92	11.69	8.84	4.81	4.66	4.92	6.49	**4.11**	10.30	10.40
9	4.74	5.41	5.75	9.32	8.81	8.91	5.57	9.70	**4.12**	6.10
10	6.95	6.66	5.68	10.48	9.69	10.20	7.04	10.84	6.23	**5.24**

Table 27: Euclidian distance matrix for */i:m/-90%*

	1	2	3	4	5	6	7	8	9	10
1	**4.98**	5.29	5.09	6.02	5.33	5.29	5.75	5.14	6.22	7.94
2	6.36	5.47	5.62	6.31	5.26	5.16	7.03	**5.10**	6.33	8.89
3	5.01	5.29	**4.58**	7.38	7.02	6.74	4.66	5.99	5.92	5.70
4	7.51	8.03	6.54	**3.97**	4.49	5.43	8.13	5.36	9.22	10.72
5	7.33	7.78	7.02	**5.20**	5.54	5.20	8.17	6.07	9.66	9.95
6	7.24	6.87	6.96	7.05	6.61	**6.33**	7.88	6.57	7.88	9.13
7	4.59	5.53	5.81	8.79	8.65	8.00	**4.12**	7.31	6.26	4.80
8	6.02	5.83	4.53	4.46	4.35	4.71	6.56	**3.95**	7.18	8.72
9	7.43	6.81	6.73	8.53	8.03	8.26	7.38	7.49	**6.24**	9.38
10	5.64	6.15	6.97	9.69	9.30	7.92	5.70	8.12	7.61	**3.92**

Table 28: Euclidian distance matrix for */i:n/-80%*

	1	2	3	4	5	6	7	8	9	10
1	**4.61**	7.05	5.22	5.37	6.23	5.53	6.19	6.10	7.06	5.55
2	5.35	**4.79**	5.43	7.14	4.89	4.83	5.41	5.06	5.13	6.31
3	3.31	4.40	**2.99**	4.35	3.35	4.96	3.29	3.48	4.35	4.32
4	4.86	5.29	4.22	4.68	4.20	6.14	**3.70**	4.68	5.55	4.89
5	5.27	2.95	4.47	6.19	**2.29**	5.17	2.75	3.48	4.04	5.43
6	4.68	4.98	4.90	6.41	4.98	**3.88**	5.07	4.88	4.79	5.82
7	3.93	5.01	3.98	4.56	4.08	4.60	**3.35**	4.87	5.08	4.34
8	5.06	6.08	4.81	5.76	5.44	6.35	5.96	**4.66**	6.62	6.09
9	5.85	3.88	5.51	7.07	4.13	5.09	4.20	5.20	**3.41**	6.18
10	**4.21**	7.31	4.61	4.27	5.97	6.54	5.75	5.79	6.72	4.24

Table 29: Euclidian distance matrix for */u:m/-80%*

	1	2	3	4	5	6	7	8	9	10
1	**4.53**	6.09	5.25	5.24	5.32	5.23	5.68	5.11	6.29	5.53
2	9.54	**7.60**	9.83	9.94	9.11	9.68	10.08	9.41	8.65	10.36
3	4.42	6.80	**3.28**	4.07	4.05	5.47	5.93	4.10	5.93	5.49
4	4.50	7.20	4.05	**3.62**	4.49	6.18	4.97	4.70	7.52	6.03
5	5.58	5.08	6.07	5.12	**4.11**	5.08	6.51	5.24	5.94	6.73
6	5.40	7.19	5.46	5.86	5.72	**4.88**	7.30	5.60	6.05	5.36
7	6.45	10.06	5.91	6.00	7.26	8.99	**5.46**	6.57	10.24	7.84
8	6.07	6.11	6.18	5.62	**4.91**	6.56	6.61	5.34	6.92	7.68
9	7.98	6.03	7.52	8.08	6.58	6.70	9.75	6.94	**4.84**	8.13
10	5.74	7.47	6.50	6.56	6.77	6.63	5.55	6.69	7.45	**3.96**

Table 30: Euclidian distance matrix for */u:n/-90%*

Condition III: Mobile network vs live recording

	1	2	3	4	5	6	7	8	9	10
1	**6.48**	8.44	6.99	7.05	7.75	6.74	6.90	6.59	8.24	6.84
2	9.24	**8.33**	10.48	10.67	10.42	9.53	9.21	9.97	9.72	10.13
3	6.95	8.50	**6.60**	6.63	7.24	8.17	7.08	7.21	7.72	7.37
4	6.60	8.32	6.14	**6.10**	6.90	8.10	6.69	6.64	7.42	6.82
5	8.31	9.16	7.12	7.19	**7.07**	10.10	7.87	8.57	7.85	8.45
6	7.18	8.43	8.12	8.18	8.90	7.71	7.51	**7.14**	9.22	7.41
7	8.08	8.22	8.02	7.90	8.12	10.01	**7.09**	8.07	8.77	7.57
8	6.24	7.64	7.41	7.39	8.47	6.93	6.49	**6.13**	8.62	6.55
9	7.75	8.10	8.24	8.43	8.24	8.01	8.15	8.69	**6.75**	9.47
10	8.03	8.82	8.55	8.56	8.92	9.32	6.92	8.24	9.69	**6.71**

Table 31: Euclidian distance matrix for */a:m/-90%*

	1	2	3	4	5	6	7	8	9	10
1	7.03	8.51	7.68	8.24	8.69	**6.99**	7.98	8.38	9.62	8.07
2	6.78	**6.06**	7.79	8.55	8.96	6.42	7.18	7.69	7.69	7.58
3	6.85	8.41	**6.20**	6.56	6.76	7.52	7.09	7.78	8.48	7.49
4	6.94	8.91	6.27	**6.02**	6.31	7.68	6.95	7.44	8.62	7.83
5	8.14	9.90	7.73	7.65	**7.40**	8.88	8.17	8.85	9.88	8.85
6	6.57	6.79	7.65	8.38	9.27	**5.82**	7.22	7.30	8.35	7.27
7	9.54	10.07	9.07	8.68	8.36	10.20	**8.15**	8.76	9.56	9.19
8	6.37	7.65	6.47	6.62	7.31	6.47	6.42	**5.92**	8.52	6.73
9	9.26	**8.83**	9.66	10.21	10.41	9.00	9.21	10.22	8.87	9.91
10	7.06	8.58	6.59	7.52	8.01	7.07	7.26	7.11	9.80	**5.93**

Table 32: Euclidian distance matrix for */a:n/-80%*

	1	2	3	4	5	6	7	8	9	10
1	**4.58**	5.85	5.59	6.03	6.77	5.77	5.07	5.48	5.62	6.33
2	6.80	**6.69**	7.19	7.67	8.12	7.39	6.97	7.41	6.98	7.57
3	6.63	7.59	6.25	6.00	6.23	6.55	6.01	**5.82**	8.25	7.93
4	6.17	7.49	6.77	**5.31**	5.65	6.08	5.76	5.46	8.20	8.52
5	5.99	6.99	6.45	**4.75**	4.87	5.08	5.58	4.89	8.00	8.28
6	4.76	5.88	5.87	5.05	5.85	**4.31**	4.88	4.99	5.70	6.72
7	5.83	6.68	5.61	6.69	7.33	6.88	**5.33**	6.27	6.49	6.53
8	6.66	7.96	7.42	6.82	7.20	7.12	6.72	**6.56**	8.27	8.66
9	5.14	5.10	6.17	7.21	7.99	6.47	5.49	6.63	**4.65**	6.29
10	6.99	7.43	6.36	8.96	9.65	8.92	7.22	8.39	6.09	**5.38**

Table 33: Euclidian distance matrix for */i:m/-80%*

	1	2	3	4	5	6	7	8	9	10
1	**5.45**	6.50	6.14	6.11	6.78	6.28	6.32	6.02	7.17	6.38
2	6.24	**4.93**	5.08	5.58	5.41	4.99	6.57	6.24	5.47	7.02
3	7.19	7.32	7.07	7.48	7.81	7.58	**6.77**	7.43	7.83	6.78
4	7.29	8.79	7.47	**6.91**	7.22	8.18	8.84	7.62	9.73	9.54
5	5.95	6.75	6.06	5.91	**5.59**	6.43	7.30	6.57	8.09	7.21
6	6.05	6.11	5.79	5.74	5.70	**5.57**	7.45	6.48	7.22	7.56
7	6.78	6.79	6.80	7.29	8.29	7.15	**5.33**	6.55	6.53	5.95
8	5.94	5.72	5.47	5.61	**5.36**	5.51	6.83	6.12	6.91	7.03
9	7.58	6.74	7.20	7.76	8.34	7.24	6.81	7.55	**6.42**	7.37
10	6.24	6.24	6.66	6.99	7.52	6.42	5.73	6.79	6.72	**4.21**

Table 34: Euclidian distance matrix for */i:n/-80%*

	1	2	3	4	5	6	7	8	9	10
1	**4.46**	5.90	4.81	5.28	5.14	5.45	5.39	5.09	6.66	4.99
2	6.33	**4.19**	5.93	5.69	5.61	5.10	5.07	6.66	4.84	7.16
3	5.08	5.61	**4.69**	5.39	5.05	6.11	5.21	4.86	6.00	5.40
4	5.95	7.57	5.83	6.16	6.35	7.41	6.68	6.12	7.93	**5.66**
5	6.36	**4.93**	5.63	5.73	5.37	6.17	5.00	6.26	5.32	6.94
6	5.20	5.20	5.21	6.25	5.33	**4.69**	5.43	5.75	5.68	6.33
7	6.21	5.62	5.89	6.16	5.94	6.03	**5.61**	6.67	6.01	7.01
8	4.54	4.91	4.53	4.83	4.68	5.05	4.76	**4.52**	5.66	5.22
9	6.55	5.21	6.09	6.48	6.19	5.76	5.79	6.98	**5.04**	7.38
10	5.70	7.28	5.84	5.49	6.20	7.14	6.56	6.06	7.96	**4.45**

Table 35: Euclidian distance matrix for */u:m/-80%*

	1	2	3	4	5	6	7	8	9	10
1	6.67	7.95	6.57	**6.56**	7.04	7.14	6.96	6.70	6.68	7.34
2	8.79	**7.41**	8.72	9.06	8.27	8.73	8.53	8.50	8.51	9.18
3	6.24	6.75	**5.96**	6.03	6.27	6.56	6.01	5.98	5.97	7.01
4	5.97	7.03	5.60	**5.43**	5.56	6.52	5.75	5.64	6.05	6.73
5	6.70	7.25	6.40	6.57	6.48	7.03	6.64	**6.40**	6.56	7.38
6	6.25	6.83	6.20	6.46	6.55	**5.87**	6.62	6.37	6.00	6.87
7	7.05	8.31	7.11	6.94	7.01	8.09	**6.75**	6.80	7.72	7.38
8	6.95	7.04	7.12	7.35	7.01	7.01	6.88	**6.67**	7.08	7.99
9	7.85	6.76	7.58	8.03	7.66	7.06	7.88	7.88	**6.75**	8.04
10	5.48	6.33	5.64	5.66	5.71	5.60	5.92	6.06	5.66	**4.77**

Table 36: Euclidian distance matrix for */u:n/-80%*