Research Article

Closed set speaker identification using fricative /s/ and /ʃ/ in Kannada speaking individuals

Arjun M. S.[a] and Hema N.[b]

JAIISH(2014)
Vol 33, pp. 72-81

Affiliations

[a]Maharaja's College, University of Mysore
[b]All India Institute of Speech and hearing, Mysore.

Corresponding Author

Hema N.
Lecturer, Department of Speech-Language Sciences, All India Institute of Speech and Hearing, Manasagangothri, Mysore
hema_chari2@yahoo.com.

Key Words

Fricative duration
Fricative amplitude
Centre frequency of frication
Euclidean distance

Abstract

The aim of the present study was to obtain the percentage of speaker verification using fricative sounds in Kannada speaking individuals. The participants chosen for the study were ten Kannada speaking neuro-typical adults, in the age range of 20-30 years, constituted as Group A and further sub grouped as Group B and Group C constituting five males and five females respectively. The material used was eight commonly occurring, meaningful Kannada words containing the voiceless fricatives /s/ and /ʃ/ in the word initial position. Spectrographic analysis using the PRAAT software was done and each token were measured for (1) the frication duration, (2) fricative amplitude ratio, (3) center frequency of frication/turbulence. The results of the present study showed that fricative sound /s/ and /ʃ/ was good for the condition - fricative duration with fricative amplitude (T2) which could be considered as an efficient parameter for speaker verification compared to other conditions.

Introduction

One of the applications of acoustic analysis is forensic speaker identification. During the committing of a crime for example a bomb threat, ransom demand, sexual abuse, hoax emergency call or drug deal, the serious problem in forensic speaker identification is to identify an unknown speaker whose voice has been recorded during such crimes. 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 and verify the identity claim of the speaker. Speaker verification has been used in a variety of criminal cases, including extortion, drug smuggling, murder, political corruption, rape, money-laundering, wagering-gambling investigations, tax evasion, terrorist activities, bomb threats, burglary and organized crime activities. Acoustic analysis is one of the methods used for speaker verification. Forensic acoustic analysis involves tape authentication, tape filtering and enhancement, reconstruction of conversation, gunshot acoustics and analysis of any other questioned acoustic event.

Speech signal conveys several types of information because speech is the most natural and common way used to communicate information by humans. From the speech production point of view, the speech signal conveys linguistic information (example- language and message) and speaker information (example- regional, emotional, and physiological characteristics). Majority of population is aware of the fact that voices of different individuals do not sound alike and person can be recognized solely from his voice (perceptually).

According to Hecker (1971), any decision making processes uses the speaker dependent features of the speech signal. There are three major methods of speaker recognition suggested by Hecker (1971) and Bricker and Pruzansky (1976), firstly by listening, here a person hears a voice and then attempt to match it to a particular individual, i.e., the one whose speech they heard. Second is by visual inspection of spectrograms, it is a three dimensional (time, amplitude and frequency) display of speech sounds. These were used in attempts to identify unknown speakers by matching their speech/voice patterns with those of known speakers (or suspects). The third is a machine methods, 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. In the *Automatic speaker identification (AUSI)*, the computer does all the work and the participation of the examiner is minimal. For the purpose of automatic identification, special algorithms are used which differ based on the phonetic context. 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.

The identification of speakers from spectrographic representations of voice has obvious forensic applications, but it has been a controversial technique (Bolt et al., 1970). The results of Tosi et al. (1972) have led them to suggest that speaker verification from voice prints may be validly used in forensic situations to identify speakers objectively. Although spectrographic representations are only rather crude representation of some of the more prominent acoustic characteristics of speech, there have been few attempts to factor out the contributions of these characteristics to speaker verification. Until such data are fully elaborated, there can be no legitimacy to claim concerning the objectivity of speaker recognition from visual representations of speech. The Federal Bureau of Investigation (FBI) made a survey of 2000 voice identification comparisons, the examiners (Koenig, 1986) under actual forensic conditions observed and determined the error rate of the spectrographic voice identification technique. The survey revealed that decisions were made in 34.8% of the comparisons with a 0.53% false elimination error rate and a 0.31% false identification error rate. Under actual forensic conditions, these error rates are expected to represent the minimum error rates.

Many researchers have done studies related to identification of speakers' gender 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 can be 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.

There are studies designed on identification of speaker gender from isolated production of voiceless fricatives. One such study by Schwartz (1968), designed to investigate the ability of listeners to identify speaker gender from isolated productions of /s/, /ʃ/, /f/ and /θ/. Among the total of 18 participants nine males and nine females recorded the four fricatives in isolation. The stimuli were randomized and presented through auditory mode via loudspeaker to ten listeners for gender identifications. The obtained results indicated that, only from the isolated productions of /s/ and /ʃ/ the listeners could identify the gender of the speakers, but could not from the /f/ and /θ/ 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 gender of a speaker from hearing voiceless fricatives in isolation and gender was better identified on fricative /h/.

Whiteside (1998) conducted an experimental study using brief (30 msec to 100 msec) voiceless fricative segments. Here the aim of the study was to test whether three phonetically naive listeners were able to identify the speaker's gender. Sentences were spoken by members of a group of 3 men and 3 women with a British general Northern accent, all speech segments were extracted from these spoken sentences. With an accuracy of 64.4 % the consonants segments were significantly identified by the listeners. Using spectrographic analysis, phonetic and acoustic differences related to a speaker's gender were investigated from a sample of chosen fricative segments. Analysis showed that on the average the frication of the men's voiceless fricatives was significantly lower in frequency than that of women. Apart from identification of speaker gender by a listener using spectrographic analysis of /s/ and /ʃ/ stimuli, for speaker verification there is a need to study few specific acoustical parameter of fricatives which represent the individual identity.

An Indian study by 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 time in India, an attempt to establish benchmarking was done.

The present study is an extension of the previous studies on speaker verification using only vowels, nasal consonants and fricatives. Where, the results of previous study say fricatives are used only for the gender identification. Here an attempt is made to study the acoustic features of fricatives in speaker verification. The study focus on the third method of speaker identification called voice print analysis or visible speech (Juang & Rabiner, 1993), a semiautomatic process with the help of spectrograms. Kersta (1962) had coined the word "voice print" in a report discussing identification of speaker by visual inspection of spectrograms and concluded this method seemed to offer good possibility. Stevens (1968) compared aural with the visual examination of spectrogram using a set of eight talkers and found that error rate for listening is 6% and for visual is 21%. These scores depended upon the talker, duration of the speech material and phonetic content. These variables would result in varied effect on any spectrographic analysis.

The present study was focused on Kannada fricatives /s/ and /ʃ/ which falls under the category of unstructured consonants of the Kannada script. The mean percentage and standard deviation of frequency of phonemes /s/ and /ʃ/ in Mysuru dialect conversational Kannada is 1.75 (0.21) and 0.42 (0.25) respectively (Vikas & Sreedevi, 2012). These fricatives are nonresonant consonants and during the production of these sounds a narrow constriction is maintained somewhere in the vocal tract. When air passes through the constriction, a condition called turbulence results. Turbulence means that the particle motion in the air stream becomes highly complex, forming small eddies in the region just beyond the constricted segment. The aerodynamic condition of turbulence is associated with the generation of turbulence noise in the acoustic signal. Thus, fricatives are identified by the formation of a narrow constriction in the vocal tract, development of turbulent airflow, and generation of turbulence noise.

Perceptual study using turbulent portions of fricatives and identification of speaker (gender) from voiceless fricative segments has been conducted earlier. Vowels, nasals and fricatives (in decreasing order) are commonly recommended for voice recognition because their spectra contain features that reliably distinguish speakers and they are relatively easy to identify in speech signals. 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). Fricatives are not only the class of sounds involving noise generation. However, compared to stops and affricates, fricatives have relatively longer duration of noise, and it is this lengthy interval of a periodic energy that distinguishes fricatives as a sound class. The duration of fricatives can range from 50ms to 200ms. Voiced fricatives are produced with two sources of energy, the quasiperiodic energy of vocal fold vibration and the aperiodic energy of turbulence noise whereas the unvoiced fricatives have only the latter source of energy. Voiced fricatives have greater amplitude of the fricative interval than their unvoiced counterparts. Fricatives have been of particular interest because the cavities of different speakers are distinctive and not easily modified (except via colds). Sole (2003) found that frication duration greatly affects the shapes and heights of the amplitude of fricatives. In other words, the amplitude of a fricative tends to be higher when a fricative is in an initial position, but lower, in a final position. Relatively extended period of noise (frication) that is the principal acoustic cue to the perception of fricatives as well as the noise in terms of amplitude or frequency concentration generated by the turbulent airstream as it passes through the articulatory constriction. It is the frication onset duration which is quick and less for /s/ and slow and more for /ʃ/. The frication energy concentration is between 4 to 6 kHz for /s/ and for /ʃ/ it is from 2.5 kHz onwards till 4 kHz. Not many studies have examined the usefulness of fricatives based on acoustic parameters (turbulent duration, amplitude and spectral peak) using spectrogram and therefore in this context it is necessary to examine if fricatives provide high percent of identification when two or more (closed set) speech samples are compared. In this context the present study was planned.

Method

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

Material: The material used was eight commonly occurring, meaningful multisyllabic Kannada sentences containing the voiceless fricatives /s/ (anterior alveolar) and /ʃ/ (posterior alveolar palatal) in the word initial and medial position embedded in the sentences. These sentences were developed in such a way that it embedded most of the fricatives in Kannada and were the hypothetical sentences commonly present in any forensic speech sample. The sentences were written on a white background card of 4 by 6 inches in size. The sentences (words containing fricatives are bold and fricatives /s/ and /ʃ/ in each word is separated by hyphen sign) are given below:

1. "ninna magananna **jeeva-*s*-ahita** nooDabeekaadare eraDu ***sh*-arattugaLive**" *(You have two conditions if you want to see your son alive)*
2. "aivattu **lak-*sh*-a** ruupaayena ippattunaalaku ganTeyoLage koDbeeku" *(Firstly you have to give me 50 lakhs within 24 hours)*
3. "ninna ***s*-amaya** iiga ***sh*- uruvaagide**" *(Your time begins now)*
4. "yaarigaadru **vi-*sh*-aya thiLi- *s*-idalli** evanannu ***s*-aahi-*s*- ibiDuttene**" *(Do not disclose this information to anyone, if you do so I will kill him)*
5. "**Pooli-*s*-arige vi-*sh*- aya tiLi-*s*-abaaradu** matte ***s*- uLLu** heeli vishaya **tiLi-*s*-adre** adara pariNaama ***s*-ari iralla**." *(You should not inform police and if you lie and do so, you will face the consequences)*
6. "ii **pari-*s*-thitiyalli** yoochane maaDi nirdhaara togoo." *(In this situation think twice before you take any decision)*
7. "haNa **talupa-*s*-oodakke vyava- *s*-the** maaDu." *(Make arrangements to send the money)*
8. "aaga ninna maga **k-*sh*-eemavaagi** mane ***s*-eeruttaane."** *(Only then your son will reach home safely)*

Procedure: All the participants were provided prior notice that their speech samples will be audio recorded and the recording will be started when they will be ready for the same. The testing was done in a laboratory condition. Speech samples of participants were recorded individually. The sentences were presented visually to the participants. Participants were informed about the nature of the study and were asked to read and familiarize the sentences before the final recordings. They were instructed to utter the sentences in a normal modal voice. Direct (live) recording of maximum of three repetitions of these sentences by the participants was taken for the present study. The distance between the mouth and the voice recorder was kept constant at approximately 10 cm. All the recordings were carried out in one sitting (contemporary speech samples).

The recordings were done using CSL- 4500 (computerized speech lab) software (Kay Pentax, New Jersey). All these were recorded on a computer memory using a desired Bit (analog-digital) converter at a required sampling frequency of 44 KHz. CSL was used only for recording purpose and PRAAT was used for analysis. The words and in turn the target fricatives containing /s/ and /ʃ/ was truncated from the recorded samples from the wide band bar type of spectrograms using PRAAT (Boersma & Weenink, 2009). The fricatives were displayed as waveform and were acoustically zoomed to do spectrographic analysis using the same software. Each token was measured for (1) the fricative duration, (2) Fricative amplitude ratio and (3) center frequency of frication/turbulence.

Analyses: Phase I: PRAAT (Boersma & Weenink, 2009) computer software was used for analysis. The target fricatives were extracted from the samples from the wide band bar spectrogram and were stored in separate folders. The files were opened in .wav format of PRAAT. Trial II/second repetition of each recording was considered as *reference sample* or the reference set and the Trial III/third repetition was considered as *test sample* or test set. The fricatives in the recorded speech sample were displayed as waveform and were acoustically zoomed to do spectrographic analysis for the following three parameters:

1. The fricative duration was measured by manually selecting the steady state of frication of each fricatives /s/ and /ʃ/ from the spectrogram as shown in Figure 1.
2. After manually selecting the steady state of frication, the fricative amplitude was extracted automatically using the option 'Intensity' and 'Intensity listing' as shown in Figure 1, thus T1 (the onset ratio), T2 (50% amplitude ratio) and T3 (the offset ratio) was obtained as shown in Figure 2.
3. The centre frequency of frication was obtained using the option 'Spectrum' and 'View Spectral Slice' and thus the spectrum window was obtained in which the value of centre spectral peak was noted down.

In this manner the fricative duration, the fricative amplitude and the centre frequency of frication for initial position of /s/ and /ʃ/ for all the speech samples of each participant were extracted and entered in excel sheet. Both average and individual values of the trials were calculated for inter and intra speaker analysis. The present study focuses on initial position of /s/ and /ʃ/ to avoid the influence of preceding and following vowels.

Phase II: Calculating Euclidean distance manually for three different types of combinations were done in order to obtain the percentage of correct identification for the fricative sound /s/ and /ʃ/ i.e. fricative duration with fricative amplitude (FD Vs FA), fricative duration with centre frequency of frication (FD Vs CF) and fricative amplitude with centre frequency of frication (FA Vs CF).

The Euclidean distance within and between participants were noted down and the participants having the least Euclidean distance were considered as same. The formula for Euclidean distance is given by: $ED = \sqrt{(x_1 - y_1)^2 - (x_2 - y_2)^2}$ (Thakur & Sahayam, 2013). The Euclidean distance within and between the participants was noted for calculation of percent correct identification. The Euclidean distance is the distance between two points and is a measure of similarity or dissimilarity. The least Euclidean distance between reference sample and test sample was considered as identification. For this purpose the confusion matrix was used. If the Euclidean distance between the test sample and corresponding reference sample is least, then the identification was considered as correct identification. If the Euclidean distance between the test sample and the corresponding reference sample is more than the least, then the speaker is considered to be falsely identified as another speaker. The percent correct identification was calculated using the following formula:

$$\% \text{ correct identification} = \frac{\text{No. of correct identification X 100}}{\text{No. of total identification}} \quad (1)$$

Percentage of correct identification for three groups of different number of known speakers was examined. In other words, the reference samples (Trial II) of speakers were considered as "known" speakers and the test samples (Trial III) of speakers were considered as "unknown" speakers. All the ten speakers were randomly listed as speaker 1 to speaker 10. Three groups of speakers were examined- Group A, Group B and Group C. In group A, all the ten speakers were considered. One 'unknown speaker' was one among the ten 'known speakers'. One 'unknown' speaker was compared with the ten 'known' speakers of Group A. All the ten speakers were further grouped into five male speakers (Group B) and five female speakers (Group C). In group B, the five 'known' male speakers were assigned numbers as speaker 1 to speaker 5. One 'unknown male speaker' was one among the five 'known male speakers'. One 'unknown' male speaker was compared with the five 'known' male speakers of Group B. In group C, the five 'known' female speakers were assigned numbers as speaker 1 to speaker 5. One 'unknown female speaker' was one among the five 'known female speakers'. One 'unknown' female speaker was compared with the five 'known' female speakers of Group C. The unknown speakers (Trial III/third repetition) was labeled as US1, US2, US3, US4, US5, US6, US7, US8, US9 and US10, and the known speakers (Trial II/second repetition) were labeled as KS1, KS2, KS3, KS4, KS5, KS6, KS7, KS8, KS9 and KS10. If US1 was compared with KS1, KS2, KS3, KS4, KS5, KS6, KS7, KS8, KS9 and KS10, this process includes both inter and intra speaker verification because US1 (third repetition) and KS1 (second repetition) belongs to same speaker (intra speaker), US1 (third repetition) and second repetitions of KS2, KS3, KS4, KS5, KS6, KS7, KS8, KS9 and KS10 belong to different speakers (inter speaker). Similar procedure was followed in all three groups (A, B and C). In this study, all the speech samples are contemporary, as all the recordings of the same person were carried out in the same session. Closed set speaker identification tasks were performed, in which the examiner was aware that the 'unknown speaker' was one among the 'known speakers'.

Results and Discussion

Results of the study is discussed under two sections, section A was the speaker verification among all *ten speakers* (Group A) and section B is speaker verification among *five male* (Group B) *and five female speakers* (Group C) by considering average scores of five words with /s/ and three words with /ʃ/. The sub-sections within section A and B are the three conditions of combination among the acoustical parameters fricative duration, centre frequency of frication and fricative amplitude is as follows.

1. Fricative duration with Centre frequency of frication
2. Fricative duration with Fricative amplitude (T1, T2, T3, T2-T1 and T2-T3)
3. Centre frequency of frication with Fricative amplitude (T1, T2, T3, T2-T1 and T2-T3)

Section A: Speaker verification among all ten speakers (Group A)

Condition I - Fricative duration with centre frequency of frication: Correct percent identification score for average of /s/ and /ʃ/ was seen to be 20% and 30% respectively. The reference sample was taken along the row and the test sample was taken along the column. The Euclidean distance of the samples were obtained manually using excel sheet for the test and reference sample of the same speaker initially. Euclidean distance was tabulated with reference sample against test sample to find the correct percentage of identification (confusion matrix). Later these were then compared against all the speakers. The one with the minimum displacement from the reference was identified as the test speaker. The numbers in bold in the Table 1 -8 indicates the correct identification of speaker sample as belonging to the same speaker as the reference sample. The numbers in italics in the Table 1-8 indicates the wrong identification of test sample as belonging to a different reference speaker in condition II of section A and section B.

The above results imply that these acoustic measurements does not support as a good cue for speaker verification even though, the experimental data might have suggested possible perceptual strategy for fricative identification in reference to two points. Firstly, listeners identify a fricative because they hear a noisy, aperiodic component of relatively long duration. They then seem to place

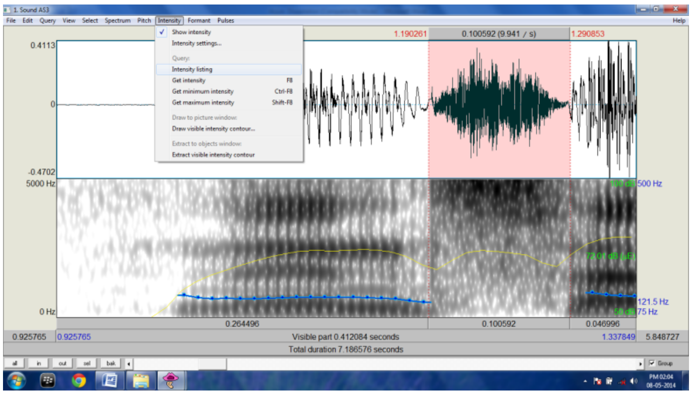

Figure 1: Steady state of frication of a fricative and use the option 'Intensity' and 'Intensity listing' to obtain fricative amplitude.

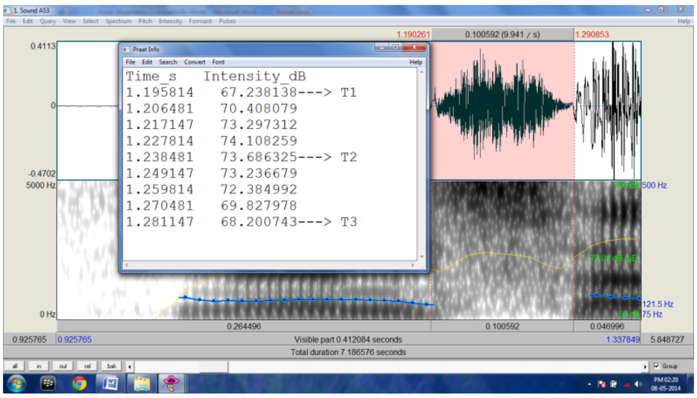

Figure 2: Fricative amplitude (T1, T2 & T3) was automatically obtained.

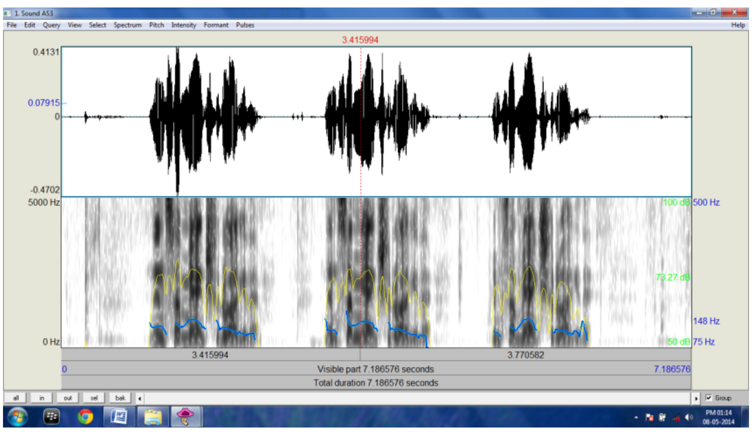

Figure 3: Waveform (above) and Spectrogram (below) of the sentence "ninna ***samaya*** iiga ***sh*-uruvaagide.**

Table 1: Euclidean distance for fricative duration with fricative amplitude (T3) for /s/

	SP1	SP2	SP3	SP4	SP5	SP6	SP7	SP8	SP9	SP10
SP1	**0.009**	3.4	3.8	13.8	3.4	2.6	0.016	1.6	2.4	2.6
SP2	2.6	**0.8**	*1.2*	11.2	6	*0.0034*	2.6	1	0.2	*0.010*
SP3	1.4	2	2.4	12.4	4.8	1.2	1.4	*0.2*	1	1.2
SP4	13.8	10.4	10	**0.0006**	17.2	11.2	13.8	12.2	11.4	11.2
SP5	5.2	8.6	9	19	**1.8**	7.8	5.2	6.8	7.6	7.8
SP6	5.4	2	1.6	8.4	8.8	2.8	5.4	3.8	3	2.8
SP7	0.02	3.4001	3.8001	13.8	3.4	2.6001	**0.0008**	1.6	2.4	2.6
SP8	1.0001	2.4	2.8	12.8	4.4	1.6	1	0.6	1.4	1.6
SP9	2.4	1	1.4	11.4	5.8	0.2	2.4	0.8	**0.0014**	0.2
SP10	1.2002	2.2001	2.6001	12.6	4.6	1.4001	1.2	0.4	1.2	1.4

Table 2: Euclidean distance for fricative duration with fricative amplitude (T2) for /ʃ/

	SP1	SP2	SP3	SP4	SP5	SP6	SP7	SP8	SP9	SP10
SP1	**1.66**	14.66	11.66	18.66	2.33	15.66	3.66	10.33	6.33	4.34
SP2	13.66	0.66	3.66	3.33	13	0.33	11.66	5	9	11
SP3	8	5	2	9	7.33	6	6.1	0.69	3.33	5.34
SP4	16.66	3.66	6.66	**0.33**	16	2.66	14.66	8	12	14
SP5	1.66	11.33	8.33	15.33	**1**	12.33	0.33	7	3	1.06
SP6	13.66	*0.66*	3.66	3.33	13	**0.33**	11.66	5	9	11
SP7	2	11	8	15	1.33	12	**0**	6.67	2.67	0.67
SP8	9.33	3.66	*0.66*	7.66	8.66	4.66	7.33	**0.66**	4.67	6.67
SP9	2.66	10.33	7.33	14.33	2	11.33	0.66	6	**2**	*0.01*
SP10	2	11	8	15	1.33	12	0.01	6.66	2.66	0.69

the fricative into one of two groups, based on relative intensity; posteriorly articulated sibilants of higher intensity, /s, z, ʃ, ʒ/, or anteriorly articulated nonsibilant fricatives of low intensity, /f, ð, θ, v/ (Raphael, Borden & Harris, 2011). In the present study since both the target stimuli belongs to the same group of sibilants of higher intensity, the identification might have become difficult. There could be a difference in identification if the target stimuli were of two different types in terms of place of articulation followed with differences in intensity. But secondly, the difference distinguished according to the place of articulation on the basis of spectral cues, the alveolar fricatives /s/ and /z/ having a first spectral peak at about 4 kHz, and the alveolar-palatal fricatives /textipaS/ and /ʒ/ have a first spectral peak at about 2.5 kHz. These cues have facilitated minimal difference in correct percent speaker verification among /s/ and /textipaS/ fricatives. The same first point of this discussion holds good for the following condition II.

Condition II - Fricative duration with fricative amplitude (T1, T2, T3, T2-T1, T2-T3): Here the results are discussed under five points of fricative amplitude measurement, initially T1 (the onset ratio), T2 (50% amplitude ratio), T3 (the offset ratio), ratio between T2 and T1 and the final is the ratio between T2 and T3. Among the five conditions the highest percent speaker identification was 60% for /s/ as shown in Table 1 and 70% for /ʃ/ as shown in Table 2 for the condition FD vs FA (T3) and FD vs FA (T2) respectively. The results of all the five conditions showing the percent correct identification for average of /s/ and /ʃ/ is as follows:

- *20% and 40% respectively for FD vs FA (T1)*
- *30% and 70% respectively for FD vs FA (T2)*
- *60% and 50% respectively for FD vs FA (T3)*
- *10% and 30% respectively for FD vs FA (T2-T1)*
- *20% and 10% respectively for FD vs FA (T2-T3)*

Condition III - Centre frequency of frication with Fricative amplitude (T1, T2, T3, T2-T1 and T2-T3): As mentioned in the previous condition, here also the results are discussed under five points of fricative amplitude measurement. Results showed the percent correct identification for average of /s/ and /ʃ/ was seen to be 30% and 30% respectively for CF vs FA (T1), CF vs FA (T2) and CF vs FA (T3), 20% and 30% respectively for CF vs FA (T2-T1) and CF vs FA (T2-T3).

From this result it is observed that, the parameter centre frequency of frication with fricative amplitude is not a very sensitive parameter for speaker identification. Spectrograms are not ideal for examination of the detailed spectral features of fricatives, For this purpose, it is preferable to use spectra determined by methods such as Fast Fourier Transformation (FFT) or Linear Predictive Coding/Coefficient (LPC). FFT and LPC spectra reveal that both alveolar and palatal fricatives have numerous minor maxima and minima in their spectra which are referred to as poles and zeros.

Section B: Speaker verification among five male (Group B) and five female (Group C)

Condition I - Fricative duration with centre frequency of frication: The results indicating correct percent identification score for average of /s/ and /ʃ/ was seen to be 20% and 60% respectively among five male speakers (Group B), 20% and 20% respectively among five female speakers (Group C). From this result it is observed that the male speaker had higher percent correct speaker verification for /ʃ/ sound compare to female speaker.

The percent correct identification increases as the number of participants decreased. This result contradicts the Hollein (2002) findings that decrease in error rate with increase in the number of participants. But irrespective of the paradigm or the material considered, 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 power 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.

Condition II - Fricative duration with fricative amplitude (T1, T2, T3, T2-T1, T2-T3): Here the results are discussed under five points of fricative amplitude measurement, initially T1 (the onset ratio), T2 (50% amplitude ratio), T3 (the offset ratio), ratio between T2 and T1 and the final is the ratio between T2 and T3. In Group B, the results showed the percent correct identification for average of /s/ and /ʃ/ was seen to be 40% and 60% respectively among five male speakers for FD vs FA (T1), 80% as shown in Table 3 and 100% as shown in Table 4 for FD vs FA (T2), 80% and 80% for FD vs FA (T3), 20% and 40% for FD vs FA (T2-T1), 20% and 60% for FD vs FA (T2-T3). On comparison between /s/ and /ʃ/, from the above results, it is found that /ʃ/ fricative shows higher percentage of correct identification at all the five amplitude measurement conditions.

Table 3: Euclidean distance for fricative duration with fricative amplitude (T2) for /s/

	SP1	SP2	SP3	SP4	SP5
SP1	2	5.6	8.2	17.2	1
SP2	3.2	**0.4**	3	12	4.2
SP3	5	1.4	**1.2**	10.2	6
SP4	15.8	12.2	9.6	**0.6**	16.8
SP5	*1.8*	5.4	8	17	**0.8**

Table 4: Euclidean distance for fricative duration with fricative amplitude (T2) for /ʃ/

	SP1	**SP2**	**SP3**	**SP4**	**SP5**
SP1	**1.6**	14.6	11.6	18.6	2.3
SP2	13.6	**0.6**	3.6	3.3	13
SP3	8	5	**2**	9	7.3
SP4	16.6	3.6	6.6	**0.3**	16
SP5	1.6	11.3	8.3	15.3	**1**

Table 5: Euclidean distance for fricative duration with fricative amplitude (T2) for /s/

	SP1	SP2	SP3	SP4	SP5
SP1	**3**	2.4	***0***	21.1	3.8
SP2	4.4	**1**	1.4	10.2	2.4
SP3	5.2	*0.2*	2.2	5.7	1.6
SP4	7.2	1.8	4.2	**0.1**	*0.4*
SP5	4.8	0.6	1.8	7.8	2

Table 6: Euclidean distance for fricative duration with fricative amplitude (T2) for /ʃ/

	SP1	SP2	SP3	SP4	SP5
SP1	**0.3**	11.6	5	9	11
SP2	12	**0**	6.6	2.6	0.6
SP3	4.6	7.3	**0.6**	4.6	6.6
SP4	11.3	0.6	6	**2**	*0*
SP5	12	0	6.7	2.6	0.6

In Group C, the results showed the percent correct identification for average of /s/ and /ʃ/ was seen to be 40% and 40% respectively among five female speakers for FD vs FA (T1), 40% as shown in Table 5 and 80% as shown in Table 6 for FD vs FA (T2), 40% and 40% for FD vs FA (T3), 20% and 60% for FD vs FA (T2-T1), 20% and 40% for FD vs FA (T2-T3). On comparison between /s/ and /ʃ/, from the above results, it is found that /ʃ/ fricative shows higher percentage of correct identification at all the five amplitude measurement conditions.

Condition III - Centre frequency of frication with Fricative amplitude (T1, T2, T3, T2-T1 and T2-T3): As mentioned in the preceding conditions, here the results are discussed under five points of fricative amplitude measurement. In Group B, the results showed the percent correct identification for average of /s/ and /ʃ/ was seen to be 40% in Table 7 and 60% in Table 8 respectively among five male speakers for CF vs FA (T1), CF vs FA (T2) and CF vs FA (T3), 20% and 60% respectively for CF vs FA (T2-T1) and CF vs FA (T2-T3). On comparison between /s/ and /ʃ/, from the above results, it is found that /ʃ/ fricative shows higher percentage of correct identification at all the five amplitude measurement conditions.

Table 7: Euclidean distance for centre frequency of frication with fricative amplitude (T2) for /s/

	SP1	SP2	SP3	SP4	SP5
SP1	186	375	*44.9*	561	578
SP2	264	454	*122*	640	657
SP3	*82.9*	272	59	458	475
SP4	452	262.8	593.8	**76.4**	61.7
SP5	451	*262.4*	593.6	78	**59.2**

Table 8: Euclidean distance for centre frequency of frication with fricative amplitude (T2) for /ʃ/

	SP1	SP2	SP3	SP4	SP5
SP1	142	372	*73.2*	327	113
SP2	213	**17**	282	682	468
SP3	*46.3*	184	115	514	300
SP4	597	827	527	**128**	342
SP5	219	449	149	250	**35.6**

In Group C, the results showed the percent correct identification for average of /s/ and /ʃ/ was seen to be 20% and 20% respectively among five female speakers for CF vs FA (T1), CF vs FA (T2) and CF vs FA (T3), 20% and 20% respectively for CF vs FA (T2-T1) and CF vs FA (T2-T3). On comparison between /s/ and /ʃ/, from the above results, it is found that /s/ and /ʃ/ fricative shows similar percentage of correct identification at all the five amplitude measurement conditions. There was no difference in either using /s/ or /ʃ/ for any speaker verification with reference to this particular condition. The speaker verification was at higher error rate in this condition compared to other conditions. Possible reason could be the speaker verification conducted in a laboratory condition and in factual forensic conditions the results may differ. According to Koenig (1986), under actual forensic conditions the error rates observed and determined in the spectrographic voice identification technique are expected to represent the minimum error rates.

With the fundamental knowledge of Pamela and Savithri (2002) investigation of the reliability of voiceprints by extracting acoustic parameters in the speech samples using wideband spectrograms, as a general discussion there might be high intrasubject variability which might cause lower correct percent of speaker verification. However, the results of the present study show the application of acoustical parameters in speaker verification. From this preliminary research on fricatives, it is observed and signifies that the acoustic feature fricative amplitude in combination with fricative duration and centre frequency of frication could be considered as a better parameter for speaker verification compared to only fricative duration and centre frequency of frication combination. The results imply that fricative amplitude was different across participants which emphasized the identity of a person. Similar study with large number of participants should be generalized to other languages too.

Conclusions

The results of this study using fricative sounds in Kannada speaking individuals showed several points of interest. Under section A, the high percent correct identification of average /s/ and /ʃ/ was *60% for the condition FD vs FA (T3) and 70% for the condition FD vs FA (T2) respectively.* Under section B, the high percent correct identification of average /s/ and /ʃ/ among five male speakers was found to be 80% and 100% and among five female speakers was 40 % and 80% respectively for the condition FD vs FA (T2). The high percent correct identification of average /ʃ/ was the highest among both Group B and Group C for the condition FD vs FA (T2) compare to the average /s/. Thus among the two fricatives, /ʃ/ is more sensitive for speaker identification compared to /s/. And among the overall five conditions FD vs FA (T1, T2, T3, T2-T1, T2-T3) may be considered as good measurement for speaker verification using fricatives /s/ and /ʃ/ in Kannada language. The other conditions/acoustic measures may hence not be considered for speaker verification using fricative sounds. The results of the present study are restricted to ten speakers (five male and five female) and eight sentences of Kannada language. Hence the results cannot be generalized to other languages and words. The results of the study are relatively interesting and future research is defensible on other Indian languages.

Acknowledgements

We would like to express our sincere thanks to Dr. S. R. Savithri, Director, All India Institute of Speech and Hearing for permitting us to do this study. Our heartfelt gratitude to the participants

in the study for their cooperation.

References

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

Bolt, R. H., Cooper, F. S., David, E. E., Denes, P. B., Picket, J. M., & Stevens, K. N. (1970). Speaker identification by speech spectrogram. *The Journal of the Acoustical Society of America, 47*, 597-613.

Bricker, P. S. & Pruzansky, S. (1976). Effects of stimulus content and duration on talker identification.*The Journal of the Acoustical Society of America, 40*, 1441-1450.

Deepa, A. & Savithri, S. R. (1972).*Re-Standardization of Kannada Articulation Test.* A Dissertation submitted in part fulfillment of final year M.Sc Speech-Language Pathology, University of Mysore, Mysore.

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

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

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

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

Juang, B. H. & Rabiner, L. R. (1993).*Fundamentals of Speech Recognition.* Englewood Cliffs, NJ, Prentice Hall.

Kersta, l. G. (1962).Voice identification, Nature, 196, 1253-1257. In F. Nolan, 1983, (Ed.), *The Phonetic Bases of Speaker Identification.* Cambridge: Cambridge University Press.

Koenig, B. E. (1986). Spectrographic Voice Identification: A forensic survey (letter to the editor). *The Journal of Acoustical the Society of America, 79*, 2088-2090.

LaRiviere, C. (1974). Speech Science Laboratory, University of Missouri-Kanas City, 5220 rockhill Rd, Kanas City, MO 64110 (USA), *Phonetica, 29*, 246-252.

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

Raphael, J. L., Borden, J. G. & Harris, S. K. (2011).*Speech Science Primer: physiology, acoustics, and perception of speech,* 6^{th}ed.Philadelphia: Williams & Wilkins press, USA.

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

Sole, M. (2003). Aerodynamic characteristics of onset and coda fricatives. *Proceedings of the 15th International Congress of Phonetic Sciences,* Barcelona, 2761-2764.

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

Su, L., Li, K., & Fu, K. (1974). Identification of Speakers by Use of Nasal Co articulation. *The Journal of the Acoustical Society of America, 56*(5), 1876-1882.

Thakur, A. S. & Sahayam, N. (2013). Speech Recognition Using Euclidean Distance. *International Journal of Emerging Technology and Advanced Engineering, 3*(3), 587-590.

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

Vikas, M. D., & Sreedevi, N. (2012). *Frequency of Occurrence of Phonemes in Kannada.* A project funded by AIISH Research Fund.

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