World Journal of Advanced Engineering Technology and Sciences
eISSN: 2582-8266
Cross Ref DOI: 10.30574/wjaets
Journal homepage: https://wjaets.com/

(RESEARCH ARTICLE)

Check for updates

Is tonal language a problem for speaker identification (SPID)?

Lakshmi Prasanna P [1, 2, *] and Vanlalhriati C [2]

[1] *Speech – Language Pathology, and 2MSc (SLP).*
[2] *Helen Keller's Institute of Research and Rehabilitation for the Disabled Children, RK Puram, near Neredmet X roads, Secunderabad-500056, Telangana, India.*

World Journal of Advanced Engineering Technology and Sciences, 2022, 07(02), 163–173

Publication history: Received on 27 October 2022; revised on 04 December 2022; accepted on 06 December 2022

Article DOI: https://doi.org/10.30574/wjaets.2022.7.2.0140

Abstract

The availability of numerous technologies has led to an increase in the usage of bioinformatics in recent years. Siri, Alexa, and other artificial intelligence systems assist us in our daily lives. Voice recognition systems are used to confirm an individual's identity based on particular elements retrieved from his or her voice. In this regard, the current study attempted to assess the proportion of speaker identification in tonal language speaking persons. The study included 20 participants in the age range from 20 to 40 years. All participants were given a few phrases to speak and were recorded. PRAAT software was used to analyze the obtained data. A vector was developed by using the first two formants, which was then utilized to calculate the percentage for speaker identification. From small sample size to bigger sample size, three groups were formed: A-5, B-10, and C-20 speakers. In a lower sample size, results showed a benchmark of 65% for vowel /i:/, which is better for SPID, 60% for /a:/, which is above chance level, and 45% for /u:/, which is below chance level. The authors stated that increasing the sample size has an influence on speaker identification.

Keywords: Tonal language; Speaker identification; Benchmark; PRAAT software; Artificial Intelligence

1. Introduction

The challenge presented by speaker identification is fundamentally distinct from the challenge presented by any form of identification that relies on constant cues. The voice of an individual is far from constant. The majority of lay people are unaware that such variations occur because typically no one uses the same word twice with all attributes being the same. For instance, recognizing person's voice on telephone banking systems, in alerting automated systems of speaker changes and identifying speaker in a discussion. Forensic speaker identification can also be performed by matching unknown voice with other voices present in the data list through various verification process to match the speaker. The emphasis in these systems is on the vocal qualities that make speech rather than on the sound or pronunciation of speech. The dimensions of the vocal tract, mouth, and other speech processing machinery in the human body can all have an effect on vocal qualities. The quality of the voice is altered by noisy environments and the emotional state of the users. A person's voice may also alter owing to health issues such as a cold, which may impair voice recognition. The size of the voice template library is big, which can affect matching performance. Automated voice recognition systems can identify people based on their voices with less than a 1% error rate. The error rate is even lower when speakers say a pre-programmed sentence. The accuracy of these technologies is comparable to that of fingerprint scanners. These systems have a 2% FAR and a 10% FRR if the sensor distance is 20 cm. SR systems are employed when the only accessible biometric is voice, which is affordable and requires minimal expenditure. [26].

Companies utilize biometric-based voice recognition software to prevent voice mimicry. The key to developing this type of software is to record and analyze a person's natural and distinctive aspects of voice and speech in the same way that a fingerprint or iris pattern is recorded. Voiceprint is a secure authentication solution that enables

* Corresponding author: Lakshmi Prasanna P

businesses to do two things: secure call centers from fraudulent calls and protect impersonated people whose identities have been stolen. A brief speech sample is sufficient to identify someone and exclude imposters. [10]

Controlling access to protected resources necessitates the personal identification verification of users. Presenting a special personal item like a key, a badge, or a password is typically how one establishes their identification. But they might get misplaced or taken. In addition, a claim based solely on identity is insufficient if there is a high risk of loss and a harsh penalty for misrepresenting oneself. The asserted identification must therefore be confirmed. This can be done by looking at a person's unique biometric features, such as fingerprints, hand geometry, or retinal patterns, or by looking at specific features that come from their distinctive activity, such as speech or handwriting. Every time, the traits were contrasted with those that had already been saved for the subject whose identification was being asserted. If the decision criterion determines that this comparison is favorable, the asserted identification is confirmed. In terms of practical application, identity verification based on a person's voice stands out among these methods. Since speech is the method of communication that we all utilize most naturally, user approval of the system would be very high. There are three methods of speaker identification listening/perceptual method, visual method and machine. There have been several measures for speaker identification, first and second formant frequencies [30, 2, 25, 15 & 19], fundamental frequency [4], pitch contour [1], Linear Prediction Coefficients [24 & 29] Cepstral Coefficients & Mel Frequency Cepstral Coefficients [12, 3, 28] and Cepstrum [23, 3, 13, 22, 14, 9 & 17] have been used in the past.

By using the visual method, [18] claimed an identification rate of more than 95%. However, in training and experimental tasks, [33] found that identification rates were 78.4% and 37.3%, respectively. An error rate of 21% was observed by [30]**.** Researchers have reported speaker identification utilizing various acoustical metrics in modern, non-modern, field, lab, and concealed situations using semi-automatic approaches. [31] Automatically calculated the voice speech's lowest three formants and pitch period. The outcome showed strong performance. The findings of the study by [11] showed that (a) formants changed as speakers' ages increased and (b) imitators struggled to match formants and pitch. In typical circumstances, two vectors—time energy distribution and voiced unvoiced speech time contrast—produce 100% speaker identification scores [16]. [7] stated that the formant frequencies were in the first factor (factor analysis) in differentiating talking gender using the frequencies of the first three formants, F0, jitter, shimmer, and duration. Two speech samples can be considered as belonging to two different speakers if 67% of the measurements in [27] analysis of spectral and temporal measures in Hindi-speaking normal subjects were different. [32] Attempted benchmarking for a temporal and spectral measure in normal and four disguised speaking settings in direct and telephone recordings as part of an investigation on acoustics and similarities and differences. In direct recordings, formant frequencies were benchmarked at 68%, 50%, and 40%, while in telephone recordings, they were 76%, 68%, and 58%. When taking into account 5 speakers, [19] benchmarking results for the vowels /i:/ 70% and /a:/ 65% which she derived using formants F2 and F1, were above the chance level. [5] Speaker identification scores are affected similarly by the nasal continuants /m/ and /n/. A semi-automatic speaker recognition system's performance is significantly impacted by the number of speakers. In contrast to other vowels, the study discovered that the vowel /a:/ before both the nasals /m/ and /n/ was reliable for speaker identification. [21] The production of fundamental frequency and voicing is determined by the tension in the vocal folds. This work studies the interaction between tone and voicing in Mizo, a lesser-known tone language. Results showed that for Mizo, the first 25% of the F0 contour displays notable coarticulatory effects. Inducing tone-specific F0 features while reducing IF0 and CF0 requires speakers to modify their vocal cords during this time. [22] did a study on to check the vowel space in a disguised voice by comparing it with the normal voice. Vowel space was larger in the original app voice than in the normal and other disguised voices. The author also stated that the formant frequencies were increased in the disguised voice with the effect of the voice changer app. The male and boy-disguised voices also show higher formant frequencies, which is close to the normal recorded (female) voice. Hence, there is no significant difference between the normal voice and the other disguised voices (original app, male and boy).

Biometric voice recognition systems can accurately recognize the unique features of each person's voice apparatus in order to generate a user profile. This is accomplished by capturing and evaluating personal characteristics such as the buccal and cranial cavity, vocal frequency, and other voice parameters. Voice impersonation scams via phone calls are on the rise since they are a low-cost and, sadly, often rather effective technique of fraud due to inadequate security solutions deployed in the targeted companies. Voice recognition using biometric parameters is not a new technology. However, far too few organizations employ it to safeguard themselves, and, more critically, the data of their customers or clients, from fraud. These approaches have a promising future in voice recognition. With your voice registered and entirely safe, you will be able to unlock your phone, open your vehicle, or sign a transaction with full legal validity. Physical (static) traits that can be used for authentication include retinas, irises, face patterns, hand vein patterns, and palm geometry, whereas behavioral features include signature, stride, and typing (dynamic). Some biometric traits, such as voice, have both physical and behavioral elements. [26].

Except for [11], single acoustic parameters were employed for comparisons, according to the review. Since the vocal tract's size and shape are reflected in the formant frequency, high intra-speaker similarities were anticipated. Statistical analysis of the data, however, has not produced a strong benchmarking. Better benchmarking of F2≈F1 may be obtained by using a vector, such as F2≈F1, rather than single metrics. This was taken into consideration when designing the current study. The goal of this study was to compare F2 and F1 for Mizo speaker identification. In the present study, three variables were considered i.e., vowels (3) and the number of known speakers (5, 10, 20). The effect of these two variables on the percentage of correct speaker identification was examined as outlined below.

2. Methodology

2.1. Subjects

Twenty Mizo-speaking*** normal male subjects in the age range of 20 -40 years participated in the study. all the subjects had normal speech and hearing and without any neurological or oro-motor abnormalities. The hearing screening was done for all subjects. Participants without any history of speech, language or hearing problem, normal oral structure and no associated problems are included in the study.

***The Mizo language, or Mizo ṭawng, is a Kuki-Chin-Mizo language belonging to the Tibeto-Burman family of languages, spoken natively by the Mizo people in the Mizoram, Meghalaya, Bangladesh, Assam, Nagaland, Tripura, Manipur states of India and Chin State in Myanmar. Around 843,750 (2011, census) uses this language.

2.2. Material

The long vowels /a:/ /i:/ /u:/ occurring in nine words were selected; these were embedded in three sentences in word-initial and word-medial positions. The data was given in the table 1.

Table 1 The stimulus used in the study

Sl. No	Sentences
1.	A zuang sa:ng ber hi dinti:r tu:r
2.	Tu:nlaia a fel zia enti:r na:n ani
3.	Zi:ngah kawngka:ah a lo lu:t

2.3. Procedure

The subject's consent to participate in the study was obtained. Subjects were instructed to speak the sentence normally. Subjects were informed about the nature of the study and were instructed to speak the sentence in a normal manner. The sentence was recorded live (direct) by using PRAAT 6.0 version software [9]. The sampling rate was selected at 44kHz. The distance between the mouth and the voice recorder was kept constant at approximately 10 cm. Each word was truncated from the sentence and stored in the computer's memory. All the recordings were done at different places according to the subject's convenience and the noise was controlled as much as possible at that place. All the recordings were analyzed.

2.4. Analysis

Before the analysis the keywords were judged by three qualified speech-language pathologists in order to check the accuracy of the production of vowels in words. The best sample out of the three was chosen for further analysis.

2.5. Acoustical Analysis

- The formant frequencies (F1-F2) for long vowels in the words were extracted using the PRAAT software.
- In the steady state of each vowel, F1-F2 at 10 evenly spaced segments were extracted. For instance, figure 1 shows a spectrogram of a word /zuang sa:ng/.

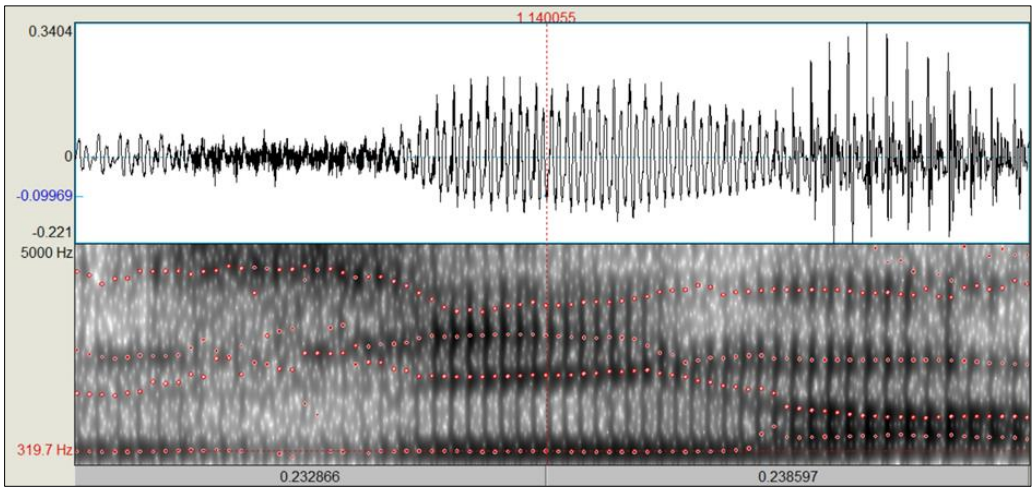

Figure 1 waveform, wide band bar type of spectrogram, and formant tracks of the word /zuang sa:ng/

Thirty F1 and F2 values were obtained for each vowel, and each speaker's 90 values were derived. For each subject, the average of the first fifteen values served as the speaker who was known, while the remaining values were utilized to represent unknown speakers.

Three factors—vowels, recording circumstances, and the quantity of 'known' speakers—were taken into account in the current investigation. These two variables were looked at for their impact on the accuracy of speaker identification.

(1) Vowels: The percentages of /a:/, /i:/, and /u:/ being correctly identified were looked at.

(2) Number of "known" speakers: For three groups with various numbers of "known" speakers, the percentage of correct identification was evaluated. Speakers 1 through 20 were chosen at random from among all twenty speakers. Group A, Group B, and Group C were the three speaker groups that were evaluated. All twenty speakers in Group A were divided into four subgroups of five speakers each. These five "known" speakers were given the numbers 1 through 5, 6 through 10, 11 through 15, and 16 through 20. The five "known" speakers of Group A were contrasted with one "unknown" speaker. All twenty speakers in Group B were divided into two groups of ten speakers each. Speakers 1 through 10 and Speakers 11 through 20 were given to these ten "known" speakers. Ten "known" speakers from group B were compared to one "unknown" speaker. One "unknown" speaker was compared to all twenty known speakers in Group C.

Since all of the recordings were made in one sitting for this investigation, every voice sample was current. The examiner knew the 'unknown' speaker was one of the 'known' ones when they completed closed-set speaker identification tasks. In all scenarios, the remaining values served as unknown speakers while the average of the first fifteen values served as known speakers. For all vowels and recording settings, all "known" speakers were given numbers ranging from 1 to 20, while matching "unknown" speakers were given numbers ranging from US 1 to US 20. For illustration, the same speaker is represented by speaker 01 (known) and speaker US1 (unknown).

The F2≈ F1 was plotted for groups of varying numbers of speakers from known speaker's vs one unknown speaker, with F1 on the horizontal axis and F2 on the vertical axis. The recognized speakers and the mystery speaker were contrasted. The distance between the unknown and known speakers was used to determine if a speaker may be identified positively or negatively. If there was less distance between the unknown speaker and the corresponding known speaker, speaker identification was regarded to be accurate; if there was more distance between the unknown speaker and any other known speaker, speaker identification was deemed to be incorrect or false. The percent correct identification was calculated by using the following formula

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

2.6. Statistical analysis

The mean and SD were measured and paired t-tests were done to compare F2-F1 vector between vowels

3. Results and discussion

3.1. F2≈F1 vector among vowels

The result indicated variation in F2≈F1 among subjects. Subject 10 had the lowest F2≈F1, and subject 6 had the highest F2≈F1 for the vowel /a:/. Subject 9 had the lowest F2≈ F1, and subject 17 had the highest F2≈ F1 for the vowel /i:/. Subject 3 had the lowest F2≈ F1 and, subject 20 had the highest F2≈ F1 for the vowel /u:/. Vowel /i:/ had the highest F2≈ F1 followed by vowel /a:/ and /u:/. The range of F2≈F1 was 478Hz, 559Hz, and 436 Hz. The data was tabulated in the table 2.

Table 2 The mean and SD of 30 observations (3 × 10 observations), F2≈F1 (Hz) for each of the vowels /a:/, /i:/ and /u:/. *SD was calculated as '0' due to the distribution of values in the same range

Speakers	/a:/	/i:/	/u:/
1	762 (0)	1434 (0)	792 (0)
2	671 (0)	1359 (0)	773 (0)
3	742 (0)	1665 (0)	648 (0)
4	847 (0)	1398 (0)	1007 (0)
5	684 (0)	1335 (0)	784 (0)
6	1087 (0)	1615 (0)	791 (0)
7	845 (0)	1520 (0)	935 (0)
8	914 (0)	1499 (0)	1107 (0)
9	761 (0)	1260 (0)	891 (0)
10	609 (0)	1502 (0)	859 (0)
11	635 (0)	1297 (0)	784 (7.02)
12	732 (0)	1287 (0)	705 (0)
13	840 (0)	1560 (0)	817 (0)
14	729 (0)	1312 (0)	762 (0)
15	866 (0)	1414 (0)	818 (0)
16	765 (0)	1344 (0)	812 (0)
17	924 (0)	1819 (0)	958 (0)
18	1050 (0)	1716 (0)	876 (0)
19	864 (0)	1663 (0)	784 (0)
20	1013 (0)	1769 (0)	1084 (0)

3.2. Interspeaker identification

The F2 ≈ F1 was plotted with F1 on the horizontal axis and F2 on the vertical axis for groups of varied numbers of speakers from known speaker's vs one unknown speaker. To identify the speakers, a total of 180 figures (3 vowels, 1 recording condition, 3 groups of speakers, and 20 speakers) were plotted. Here are a few samples of both correct and inaccurate speaker identification for each of these.

If there was less distance between the unknown speaker and the corresponding known speaker, the speaker was deemed to have been correctly identified. If there was a significant gap between the speaker and the corresponding known speaker, they were considered to be unidentified. Figures 2 to 7 display the true/false speaker identification.

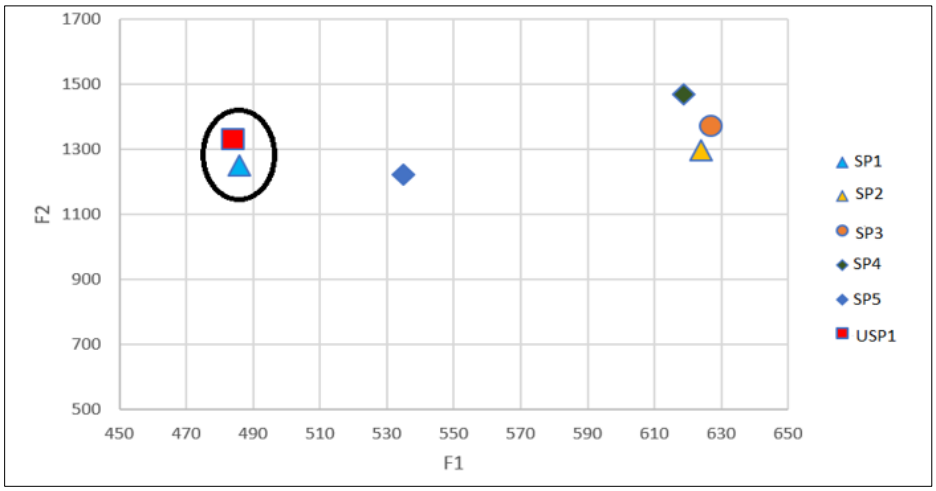

Figure 2 Correct identification of US1 with S1

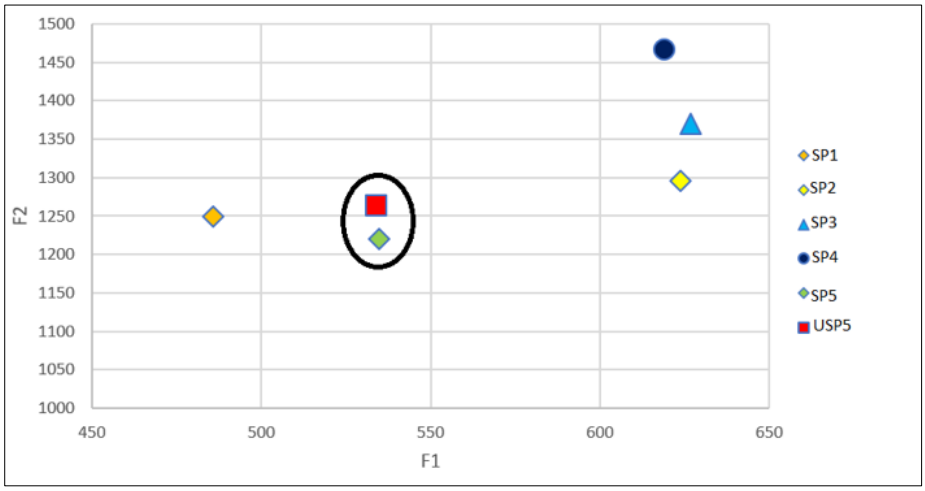

Figure 3 Correct identification US5 with SP5

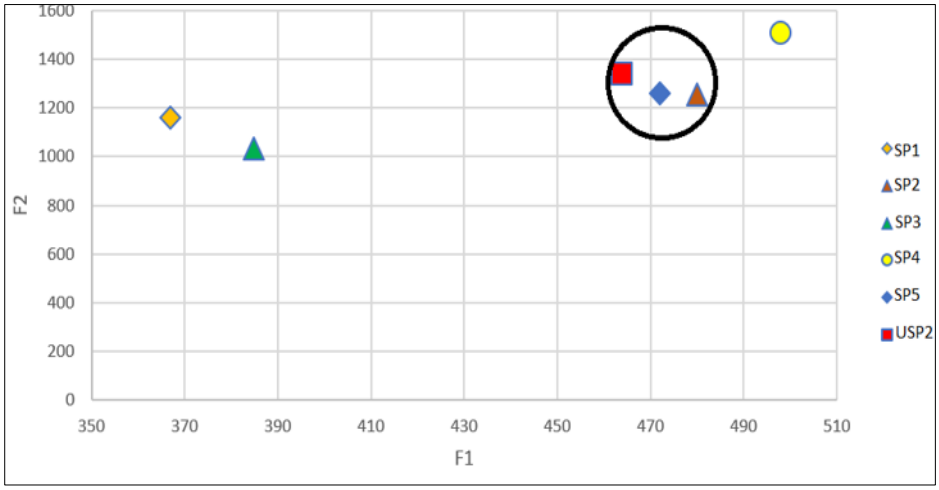

Figure 4 False identification of US2 with SP5

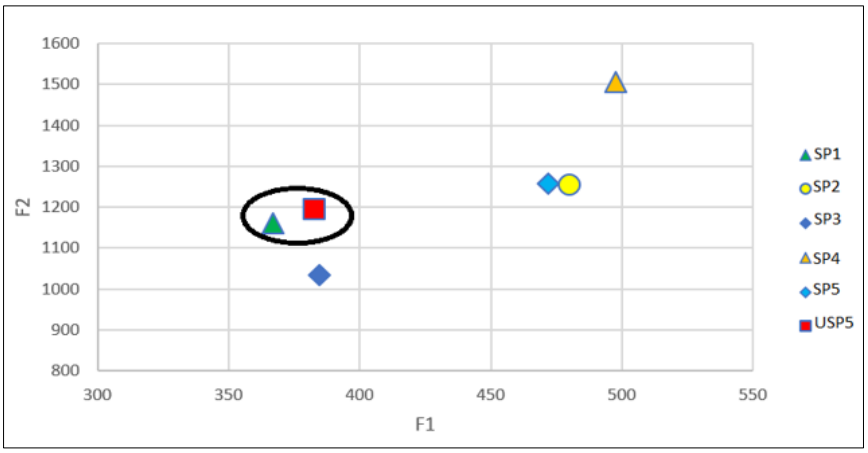

Figure 5 False identification of USP5 with SP1

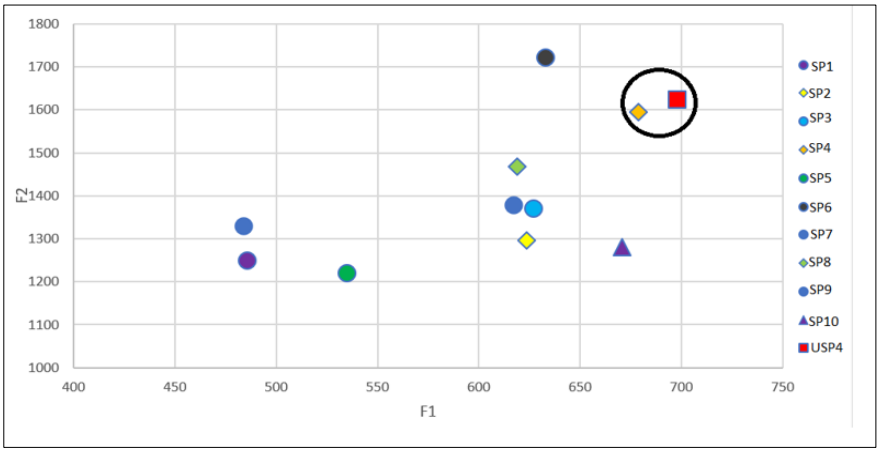

Figure 6 Correct identification of US4 with SP4

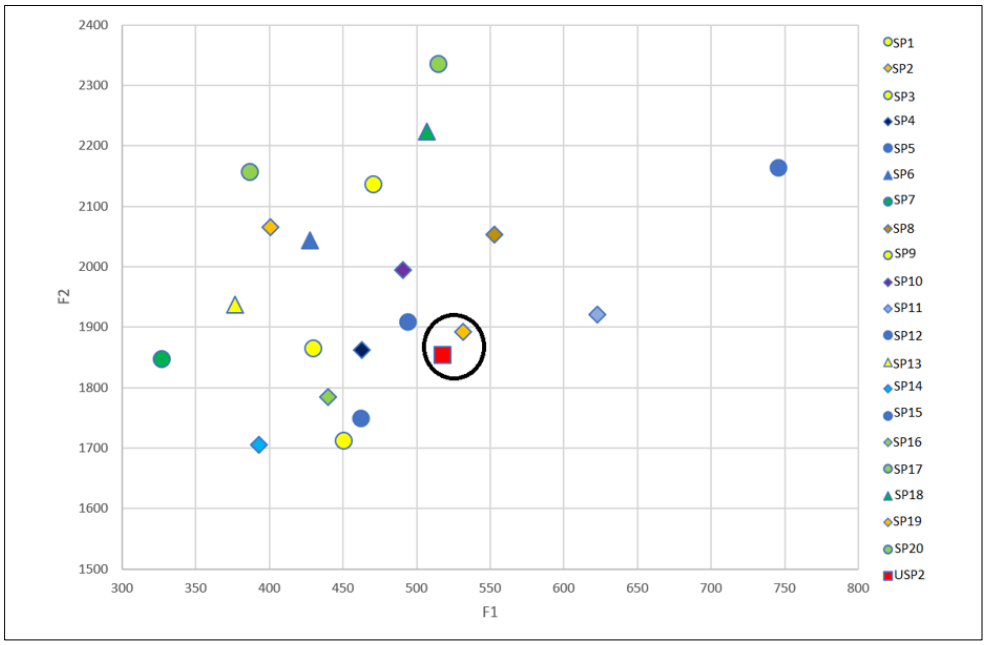

Figure 7 Correct identification of USP2 with SP2

3.3. Comparison of F2≈F1 vector between the groups

- **Group A** showed a benchmark of around 60% for vowel /a:/ whereas 65% for vowel /i:/ this indicates slightly better speaker identification which means it was noticed that above probability. Vowel /u:/ showed 45% which is said to be below probability and considered as poor benchmarking.
- **Group B** showed a benchmark of around 50% for the vowel /a:/ which states probable level and vowel /i:/ showed 40% whereas /u:/ showed 20% which is said to be below probability and considered as poor benchmarking.
- **Group C** showed a below probability and considered as a poor benchmark for speaker identification in vowel /a:/ (20%) and vowel/i:/ (25%) whereas 0% for vowel /u:/.

Overall, the interspeaker identification revealed that benchmarking depended on the no. of speakers and vowels. The per cent correct identification decreased as the number of speakers increased from 5 to 20. The mean per cent correct identification was 57% (5 speakers), 37% (10 speakers) and 15% (20 speakers). There is no particular trend observed across the speakers in relation to the vowels. This indicates that these vowel vectors are not useful for speaker identification in larger groups. The present study authors stated that the benchmarks may be affected by the increase of speakers and various vowels. The whole data was tabulated in the table 3 and depicted in figure 8.

Table 3 Percent correct identification of speakers

Sl.No	Group	% Correct Identification			
		/a:/	/i:/	/u:/	Mean
1	A (5 speakers)	60	65	45	57
2	B (10 speakers)	50	40	20	37
3	C (20 speakers)	20	25	0	15

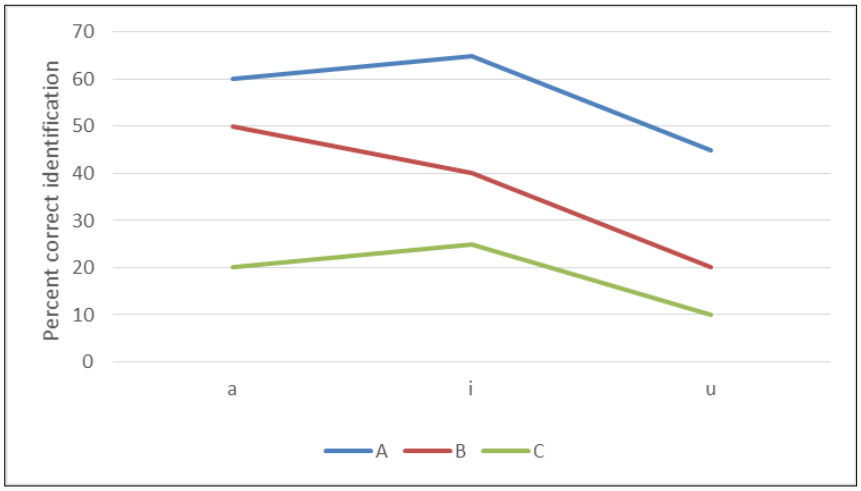

Figure 8 Percent correct identification for vowels /a:/, /i:/ and /u:/

3.4. Comparison of F2≈F1 vector between vowels

Paired sample t-test was done to compare the F2-F1 vectors among vowels where it observed that higher means were observed in vowel /i:/ and the normal pattern was followed. There is a highly significant difference found between pair 1 (a:-i:) and pair 3 (i:-u:) where $p<0.000$. There is no significant difference between vowels a:-u: where $p>0.05$ this could be due to the effect of place and manner of articulation as both are back vowels. The whole data was given in the table 4.

Table 4 Paired Samples Test

		Mean	Std. Deviation	T	Significance
Pair 1	a:	817.00	133.404	-23.644	0.000**
	i:	1488.40	172.021		
Pair 2	a:	817.00	133.404	-1.116	0.278
	u:	849.35	118.265		
Pair 3	i:	1488.40	172.021	16.265	0.000**
	u:	849.35	118.265		

The result indicated several points of interest. First of all, benchmarking for F2≈ F1 for the vowel /a:/ was 60% when five speakers were considered, 40% when ten speakers were considered and 20% when twenty speakers were considered and the benchmarking for vowels was below chance level.

Second, benchmarking for F2≈ F1 for vowel /i:/ was 60% when five speakers, 30% in ten speakers and, 25% in 20 speakers and benchmarking for vowels was below chance level.

Thirdly, benchmarking for F2≈ F1 for vowel /u:/ was 45% when five speakers, 20% in ten speakers and, 0% in 20 speakers and benchmarking for vowels was below chance level. This indicates very poor benchmarking which states that the use of F2≈ F1 of this vowel for forensic speaker identification is not useful.

In the past, [31] created an algorithm and automatically assessed the lowest three formants and pitch period of spoken speech. The algorithm's performance on vowels glides, and semi-vowels was demonstrated by the results. However, considering the poor quality of the benchmarking, the results of the present study do not appear to agree with those of [31].

[27] Used F1≈F2 transition 67% of the measures varied between participants, and 61% of the measures varied within subjects. Under direct and telephone recordings, [32] tried benchmarking for temporal and spectral measurements in normal and four disguised speech circumstances. The benchmarking for formant frequencies in direct recording was 68%, 50%, and 40%, and in telephone recording, it was 76%, 68%, and 58%. She does not, however, provide the benchmarking for each vowel.

[19] Did a study on benchmarks for SPID by using the F1-F2 vector, where results revealed that a high percentage was observed in /i/ in 70% and for vowel /a/ 65%. And benchmarking for other vowels was below the chance level when 5 speakers were considered and benchmarking obtained was above the chance level for ten and twenty speakers for all three vowels.

When five Hindi speakers were taken into account, [13] employed Quefrency to benchmark and produced benchmarks of 83.33% (live vs. live) and 81.67% (mobile vs. mobile). Additionally, when live speech was contrasted with mobile recording, a benchmark of 78.33 was obtained. Compared to Jakhar, the benchmarking in the current study was poor.

[5] Speaker identification scores are affected similarly by the nasal continuants /m/ and /n/. A semi-automatic speaker recognition system's performance is significantly impacted by the number of speakers. In contrast to other vowels, the study discovered that the vowel /a:/ before both the nasals /m/ and /n/ was reliable for speaker identification. Suresh et.al. (2015) 90% identification for /n/ in vowels that follow the nasal sound. Formant frequencies varies with speakers as the current study shows better benchmark for /i:/ and poor benchmarks for /u:/ due to the variations in the tonal language and is also agreeing with [22] which states the formant frequencies showed higher in various disguise condition.

[6] developed a benchmark for speaker identification on children with cleft lip and palate which revealed 75% for/i:/ and 60% for /a:/ among 5 speakers and the other vowels showed below probability. Authors stated that speaker identification is difficult as the speaker size increased and there is no significant difference found between vowels which could be due to the effect of manner of articulation.

Paired sample t-test was done to compare the F2-F1 vectors among vowels where it observed that higher means were observed in vowel /i:/ and the normal pattern was followed. There is a highly significant difference found between pair 1 (a:-i:) and pair 3 (i:-u:) where $p<0.000$. There is no significant difference between vowels a:-u: where $p>0.05$ this could be due to the effect of place and manner of articulation as both are back vowels.

The field of speaker identification has benefited from the findings of this study. In summary, it came to the conclusion that speaker identification in tonal language was not well served by F2≈F1. Except for /i:/, when a percentage of 65% is permissible. Although the results of other studies were compared with those of the current study, they were completely different because the other studies used different methods, including automatic and quefrency, in various speech sounds like nasal continuants and other consonants, where they discovered between 85% and 90% speaker identification. The F2≈F1 vector in tone language, which shows 65% for /i:/, was used in the current investigation. This indicates a SPID benchmark that is suitable for a smaller population group, and the authors of the current study stated that increase in number of speakers leads to poor SPID. Finally, the current study authors stated that in the Tonal language also vowels /a:/ and /i:/ found better benchmark in smaller sample size similarly to other languages.

4. Conclusion

The present study was restricted to less population, only one tonal language, restricted to vowels and gender. The present study results can be used as reference, field of forensics and also can be used in the assessment and management of speech perception aspects of tonal languages. The future studies can be carried out in other tonal languages, between various regions, larger sample size, various speaking aspects, different age groups and gender variations.

Compliance with ethical standards

Acknowledgments

Sincere thanks to the participants and Helen Keller's Institute for allowing the authors to complete the study.

Disclosure of conflict of interest

The authors declared that there are no conflicts of interest.

Source of Funding

This study was done under the part of Research at Helen Keller's Institute, Secunderabad, INDIA.

References

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

[2] Atal, B. S. (1972). Text-Independent Speaker Recognition. The Journal of the Acoustical Society of America, 52(1A), 181-181.

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

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

[5] Arjun (2015). Benchmark for speaker identification using Mel frequency cepstral coefficients on vowels preceding's nasal continuants in Kannada, unpublished dissertation, PGDFSST, University of Mysore.

[6] Akanksha Kumari and Lakshmi Prasanna P (2022). Benchmarks for Speaker Identification in Children with Cleft Lip and Palate, unpublished dissertation, Osmania University, Hyderabad.

[7] Bachorowski, J. A., & Owren, M. J. (1999). Acoustic correlates of talker sex and individual talker identity are present in a short vowel segment produced in running speech. The Journal of the Acoustical Society of America, 106(2), 1054–1063. https://doi.org/10.1121/1.427115

[8] Boersma, P. & Weenink, D. (2019). PRAAT 6.1.16 software. Retrieved from Download Praat 6.1.16 for Windows

[9] Che, C., & Lin, Q., (1995), Speaker recognition using HMM with experiments on the YOHO database, In EUROSPEECH, 625-628

[10] Danny Thakkar, Top Five Biometrics: Face, Fingerprint, Iris, Palm and Voice (www.bayometric.com), blog.

[11] Endres, W., Bambach, W., & Flösser, G. (1971). Voice spectrograms as a function of age, voice disguise, and voice imitation. The Journal of the Acoustical Society of America, 49(6B), 1842-1848.

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

[13] Furui, S. (1981), Cepstral Analysis Technique for Automatic Speaker Verification, IEEE Transactions on Acoustics, Speech and signal Processing, Vol-29, 254-272.

[14] Higgins, A., & Wohlford, R. E. (1986), A new method of text Independent Speaker Recognition, In International Conference on Acoustics, Speech and Signal processing in Tokyo, IEEE ,869-872.

[15] Hollien, H (1990), The acoustics of Crime, The New Science of Forensic Phonetics, Plenum, Nueva York.

[16] Johnson. C, Hollien. H & Hicks. J, (1984). Speaker Identification utilizing selected temporal speech features, journal of phonetics, vol; 12, 319-326.

[17] Jakhar, S. S. (2009). Benchmarks for speaker identification using Cepstral coefficient in Hindi. Unpublished project of Post graduate Diploma in Forensic Speech Science and Technology, University of Mysore, Mysore.

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

[19] Lakshmi, P. (2009). Benchmark for speaker identification using vector F2 ≈ F1 vector. Unpublished project of Post graduate Diploma in Forensic Speech Science and Technology, University of Mysore, Mysore.

[20] Lakshmi Prasanna P (2022). Vowel Space in Disguise Voice. J Biomed Sci, 11(8), 76.

[21] Lalhminghlui, W., & Sarmah, P. (2020). Interaction of Tone and Voicing in Mizo. In INTERSPEECH (pp. 1903-1907).

[22] Li, K. P., & Wrench, E. H. (1983), Text Independent Speaker Recognition with short Utterances, In international Conference on Acoustics, Speech and Signal Processing in Boston, IEEE, 555-558.

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

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

[25] Nolan, F., & Hollien, H. (1985). The phonetic bases of speaker recognition by Francis Nolan. The J Acoust. Soc. Am. 78, 817 (1985); https://doi.org/10.1121/1.392415

[26] Pavel Jirik (2021), Is a Fingerprint More Secure Than a Voiceprint? - PHONEXIA; Blog.

[27] Pamela, S. (2002). Reliability of voice prints. Unpublished dissertation, no.462, University of Mysore.

[28] Rabiner, L., & Juang, B.H. (1993), Fundamentals of Speech Recognition, Prentice Hall PTR.

[29] Soong, F., Rosenberg, A. E., Rabiner, L., & Juang, B.H. (1985), A vector Quantisation Approach to Speaker Recognition, In International Conference on Acoustics, Speech and Signal Processing in Florida, IEEE, 387-390.

[30] 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.

[31] Schafer, R. W., & Rabiner, L. R., (1970). System for automatic formant analysis of voiced sounds, j. Acoust. Soc. Amer., vol;47, 634-48.

[32] Savithri (2008) Speaker identification by native and non-native speakers Proceedings of the International Symposium on Frontiers of Research on Speech and Music.

[33] Young, M., & Campbell, R. (1967). Effects of context on talker identification. The Journal of the Acoustical Society of America, 42 6, 1250-4