



Forensic Linguistic Inquiry into the Validity of F0 as Discriminatory Potential in the System of Forensic Speaker Verification



Susanto^{1,2*}, Wang Zhenhua¹, Wang Yingli³ and Deri Sis Nanda²

¹Martin Centre for Applicable Linguistics, Shanghai Jiao Tong University, China

²Centre for Studies in Linguistics, Bandar Lampung University, Indonesia

³Centre of Criminal Technology, Public Security Bureau of Guangdong Province, China

Submission: January 2, 2017; **Published:** September 26, 2017

***Corresponding author:** Susanto, Head of Centre for Studies in Linguistics, Bandar Lampung University, Indonesia; Post doc Research Fellow in Forensic Linguistics at Martin Centre for Applicable Linguistics, Shanghai Jiao Tong University, China, Tel: +8615821962135 Email: susanto@ubl.ac.id

Abstract

In the provision of linguistic evidence as one of the foci in Forensic Linguistics, Forensic Speaker Verification (FSV) includes an analysis of speech recordings to verify the voice of a criminal. As an inquiry into the validity of the available FSV, we present the analysis on Indonesian FSV system. The system consists of pairing, tagging, acoustic features extraction, and statistical analysis. There is a claim that the system meets the demand for presenting legal evidence in Indonesian court. In the system, one of the acoustic features extracted from the speech data is fundamental frequency (F0). Then, the paper aims at reviewing the method in Indonesian FSV system in terms of fundamental frequency (F0) used as the discriminatory potential. The results show that F0 has not represented an adequate interpretation of the linguistic evidence in our experimental data. It leads us to suggest that more experimental studies are required to scrutinize F0 in the system.

Keywords : Forensic Speaker Verification; F0; Discriminatory potential; Validity

Introduction

In Indonesia, some court cases involve speech recordings of suspects as legal evidence. In the courts, an expert is invited to explain for the verification to find out whether the speech recordings are spoken by the suspects or not. The task is called as Forensic Speaker Verification (FSV). It is one of the areas in the application for Forensic Linguistics as the provision of linguistic evidence [1]. FSV system includes an analysis of speech recordings to verify the voice of a criminal. In the system, fundamental frequency (F0) is one of the acoustic features which are extracted from the speech data [2]. Then, they are analyzed as the discriminatory potential. It is important to note that there is always a question in the context of human speech sound and its forensic relevance as an inquiry into its validity [3]. The critical question brings the requirement for always reviewing an available system in forensic speaker verification or identification. Later on, the review can be used for the improvement of the system. In line with that, the paper aims at reviewing the method in Indonesian FSV system in terms of the extracted acoustic feature of F0, which is used as the discriminatory potential.

Method

The data are derived from Indonesian speech sounds of two (2) telephone conversations with Speakers LR (f;21) and MR (m;23) in the first conversation; Speakers DS (f;22) and RD (m;22) in the second. The data were recorded in Centre for Studies in Linguistics, Bandar Lampung University. The conversation is designed as a simulation for a corruption case. The speech data are categorized as Unknown (Uk), following the scenario used in Indonesian FSV system [4]. For Known (K) category, twenty (20) words spoken by each speaker are recorded to be paired with the same words in Uk sample. Praat [5] is used for the acoustic analysis of K and Uk samples. One-way analysis of variance (ANOVA) and Likelihood Ratio (LR) approach are used to evaluate the findings statistically.

Results and Discussion

In Indonesian FSV system, there are two samples from speech data which are compared for evaluation [4,6]. They are called Known (K) and Unknown (Uk) samples respectively. The K sample is derived from speech data of a suspected person. In the

sample, it is already known who the speaker is. Meanwhile, the Uk sample is derived from speech data of a recorded telephone conversation which is not known yet who is speaking. The main purpose of the comparison is to find out who is really the speaker in the recorded telephone conversation. The evaluation provides some evidence if the speaker is the suspected person or not. In presenting the evidence, there are four main steps to conduct data analysis such as pairing, tagging, acoustic features extraction, and statistical analysis. In the analysis, F0, F1, and F2 are observed to find out the patterns of habitual pitch range, minimum-maximum pitch, first-second formant, and speaking style for pitch and formant. As the evaluation of the current Indonesian FSV system, there is a claim that it “meets the demand for presenting legal evidence in Indonesian court” [6].

To review the method in Indonesian FSV system, we scrutinize

Table 1: Target words for K and Uk samples in telephone conversation 1.

| TELEPHONE CONVERSATION 1 | | | |
|--------------------------|-----------------|--------------------|---|
| Speaker LR (f;21) | Meaning | Speaker MR (m;23) | Meaning |
| <i>Dengan</i> | With | <i>Selamat</i> | Congratulations |
| <i>Dibantu</i> | Helped | <i>Saya</i> | 1 st person singular |
| <i>Menyediakan</i> | Provide | <i>Proyek</i> | Project |
| <i>Jelaskan</i> | Explain | <i>Jenis</i> | Type |
| <i>Kualitas</i> | Quality | <i>Sepuluh</i> | Ten |
| <i>Menggunakan</i> | Use | <i>Harga</i> | Price |
| <i>Menawarkan</i> | Offer | <i>Minta</i> | Ask for |
| <i>Berminat</i> | Want | <i>Tulis</i> | Write |
| <i>Tertarik</i> | Interested | <i>Paket</i> | Package |
| <i>Ulang</i> | Repeat | <i>Iya</i> | Yes |
| <i>Berikan</i> | Give it away | <i>Terimakasih</i> | Thank you |
| <i>Buatkan</i> | Make it | <i>Kita</i> | 3 rd person plural inclusive |
| <i>Benarkah</i> | Is it true | <i>Siang</i> | Noon |
| <i>Kerahasiaan</i> | Confidentiality | <i>Ada</i> | Available |
| <i>Usahakan</i> | Try it | <i>Pembangunan</i> | Development |
| <i>Segera</i> | Soon | <i>Aspal</i> | Asphalt |
| <i>Pastikan</i> | Be sure | <i>Kwalitas</i> | Quality |
| <i>Tentu</i> | Certain | <i>Nota</i> | Notes |
| <i>Total</i> | Total | <i>Jalan</i> | Street |
| <i>Terimakasih</i> | Thank you | <i>Sedang</i> | Medium |

Table 2 : Target words for K and Uk samples in telephone conversation 2.

| TELEPHONE CONVERSATION 2 | | | |
|--------------------------|-------------|-------------------|----------|
| Speaker DS (f;22) | Meaning | Speaker RD (m;22) | Meaning |
| <i>Bangunan</i> | Building | <i>Rancangan</i> | Design |
| <i>Pak</i> | Sir | <i>Bangunan</i> | Building |
| <i>Material</i> | Material | <i>Material</i> | Material |
| <i>Kerjasama</i> | Cooperation | <i>Disain</i> | Design |
| <i>Oh</i> | Oh | <i>Kerja</i> | Work |
| <i>Rencana</i> | Plan | <i>Sama</i> | Same |

| | | | |
|----------------------|---------------------------------|--------------------|---------------------------------|
| <i>Bisa</i> | Can | <i>Kwalitas</i> | Quality |
| <i>Saya</i> | 1 st person singular | <i>Anda</i> | 2 nd person singular |
| <i>Dengan</i> | With | <i>Jadi</i> | So |
| <i>Menggunakan</i> | Use | <i>Rendah</i> | Low |
| <i>Apa</i> | What | <i>Terbaik</i> | Best |
| <i>Seperti</i> | As | <i>Penanggung</i> | Insurers |
| <i>Dijelaskan</i> | Explained | <i>Selanjutnya</i> | Next |
| <i>Menyelesaikan</i> | Complete | <i>Nanti</i> | Later |
| <i>Dikerjakan</i> | Do it | <i>Dan</i> | And |
| <i>Iya</i> | Yes | <i>Seperti</i> | As |
| <i>Aman</i> | Secure | <i>Jawab</i> | Answer |
| <i>Hasil</i> | Results | <i>Apa</i> | What |
| <i>Mengerti</i> | Understand | <i>Bagaimana</i> | How |
| <i>Akan</i> | Will | <i>Mengenai</i> | About |

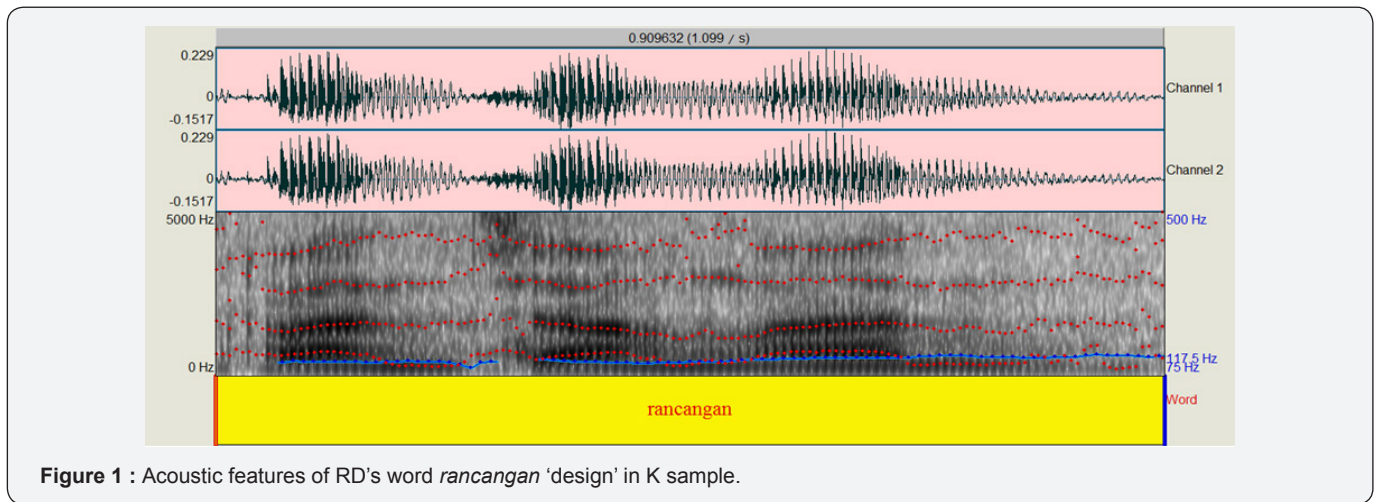


Figure 1 : Acoustic features of RD's word *rancangan* 'design' in K sample.

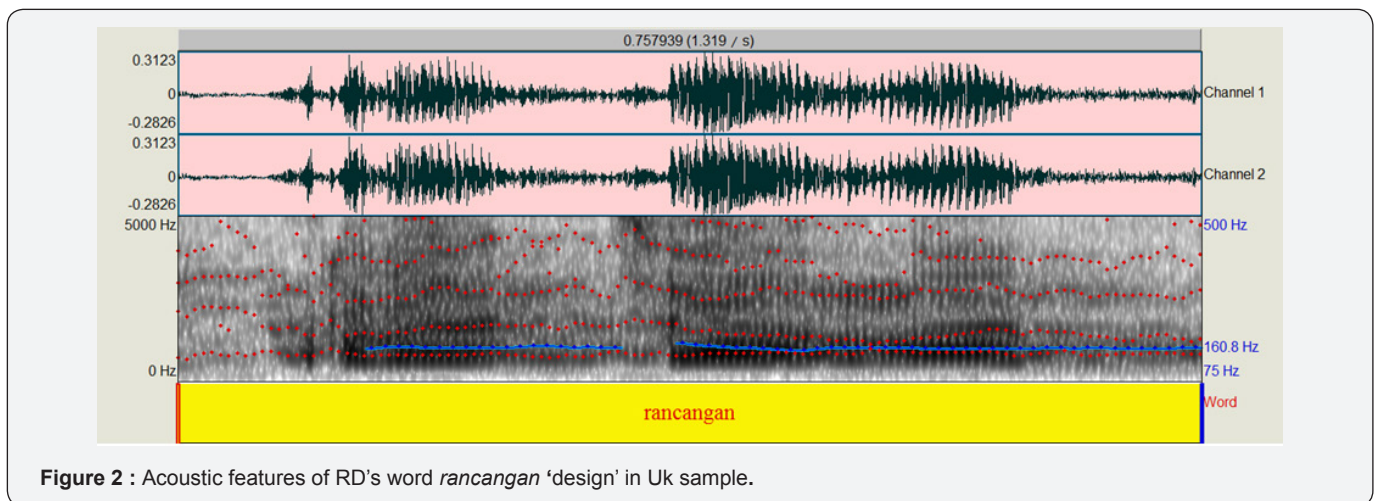


Figure 2 : Acoustic features of RD's word *rancangan* 'design' in Uk sample.

Since the speech sounds are spoken by the same speaker, we presume that pitch values of RD's speech in K and Uk samples will match. However, it is found that in the pitch analysis of its mean and standard deviation (SD) of 20 words spoken by RD in K and Uk samples, only few values match (Figure 3).

In the pitch analysis of minimum and maximum values, it is also found that the maximum values in the Uk samples do not match their K counterparts (Figure 4). Meanwhile, the minimum values in K and Uk samples only match at several points.

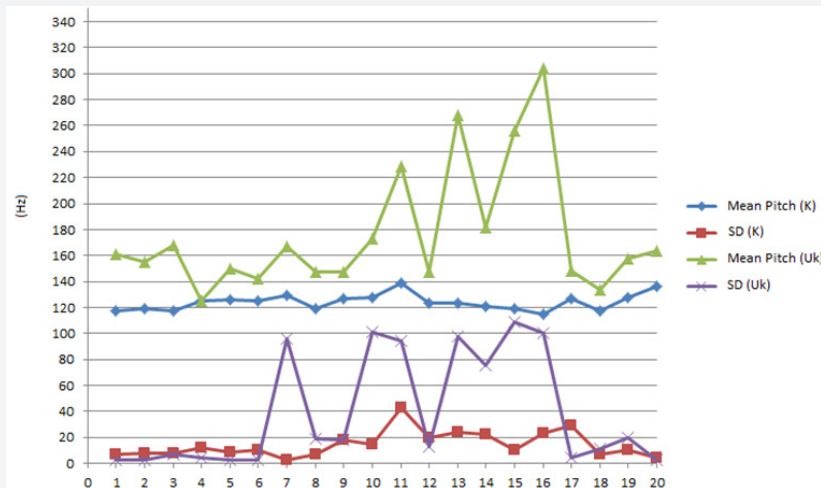


Figure 3 : Mean pitch and its standard deviation (SD) of RD's speech in K and Uk samples.

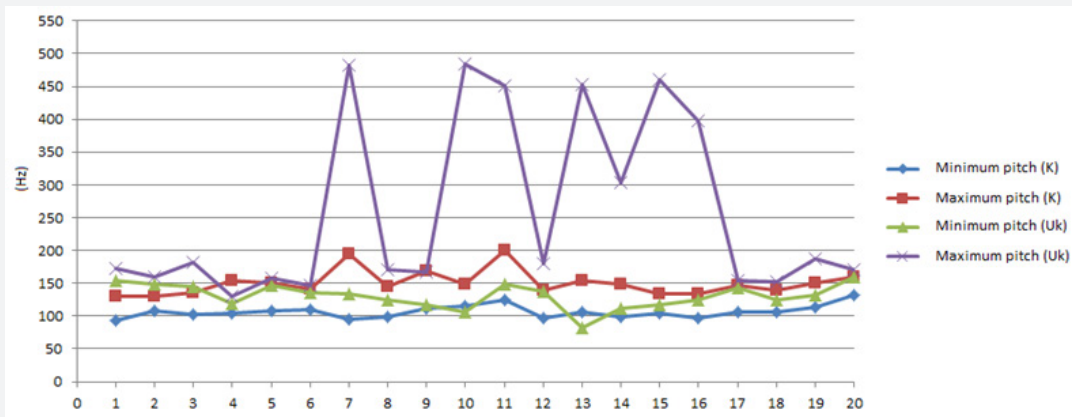


Figure 4 : Maximum and minimum pitch of RD's speech in K and Uk samples.

In addition, in one-way analysis of variance (ANOVA), it is also found that the pitch of each word spoken by RD in K and Uk samples is significantly different ($p < 0.05$). RD's F0s are significantly different in both K and Uk samples (Figure 5).

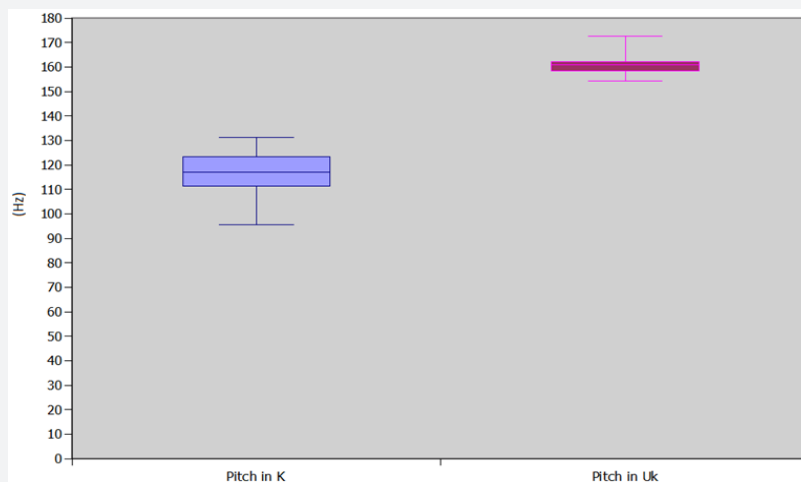


Figure 5 : F0s of RD's speech *rancangan* in both K and Uk samples.

Further, for the evidence evaluation using Likelihood Ratio (LR) approach [2], we analyze the probability of the samples. The result indicates that the pitch in the data can be categorized as 'very strong evidence against' the fact that the K and Uk samples are derived from the same speaker ($LR < 0.0001$).

$$LR = \frac{p(E | H_p)}{p(E | H_d)}$$

$p(E | H_p)$ = the probability of the samples from same speakers
 $p(E | H_d)$ = the probability of the samples from different speakers

From the results in one-way ANOVA and LR approach, it can be inferred that F0 cannot be used as a discriminatory potential in the experimental data. ANOVA says that the pitch in K and

Uk samples is significantly different. And LR also indicates that the sounds are derived from different speakers. In the contrary, they are from the same speaker, i.e. RD. We highlight three main problems that may arise in terms of fundamental frequency (F0) used as the discriminatory potential based on the experimental data following Indonesian FSV system. The first problem is about the default setting in pitch range for analysing connected speech [7]. The F0 reading with the default setting may not show the actual value of the speaker's F0 [Figure 6]. The second problem is about the telephone transmission [8]. The transmission could have effects [9], especially on the vowel quality [10] that may result in the discrepancy in values of the speaker's F0. The third problem is about the lack of theoretical background for the Indonesian FSV system which uses F0 as one of its discriminatory potentials.

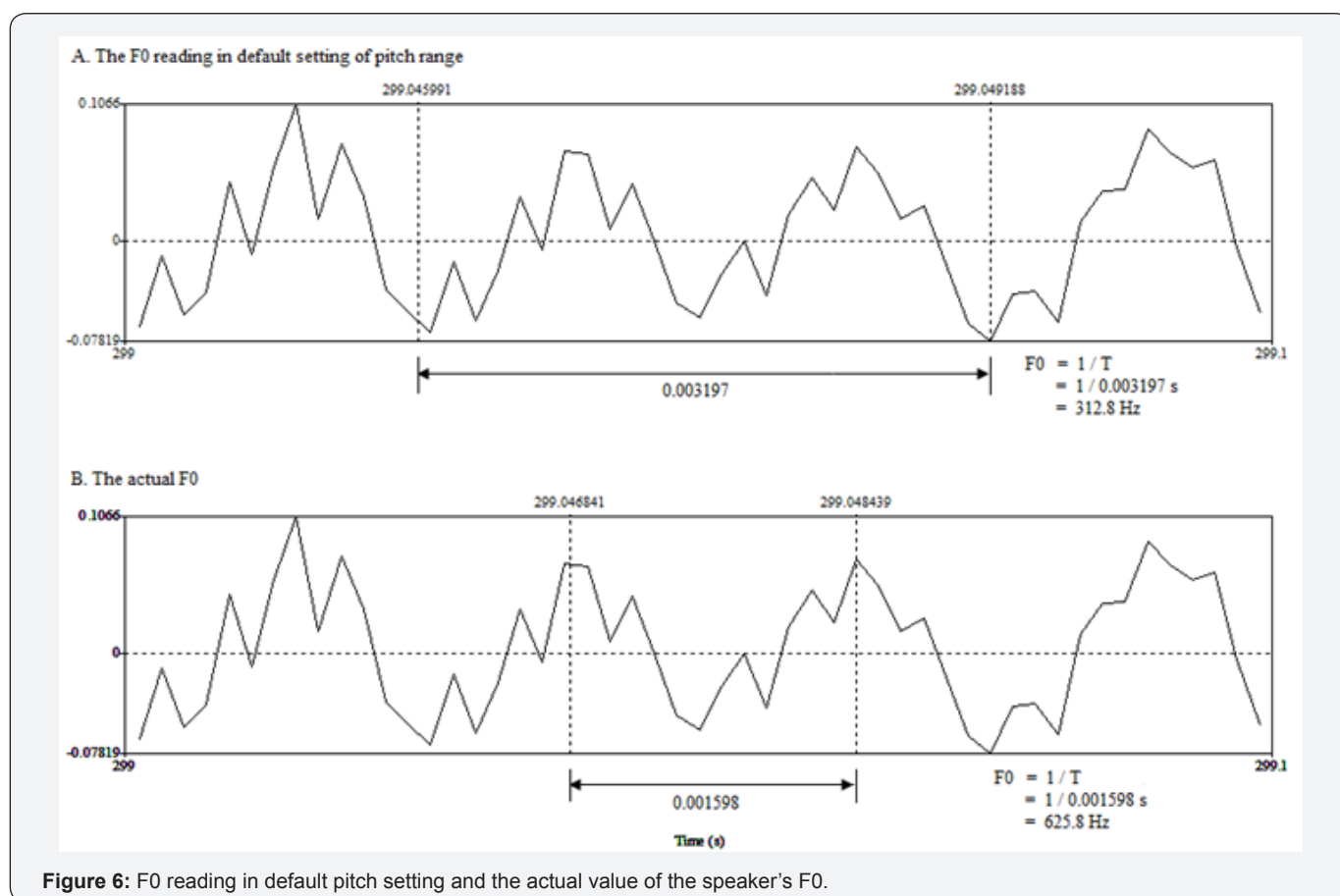


Figure 6: F0 reading in default pitch setting and the actual value of the speaker's F0.

Conclusion

F0 as the physical correlates to a speaker's pitch is analyzed to review the method in Indonesian FSV system. In the experimental data, although the speech data are derived from the same speaker (RD), only few values in pitch analysis of its mean and SD in K and Uk samples match. Maximum and minimum pitch values also show the same result. Furthermore, using one-way ANOVA and LR approach, the study proves that it fails in providing the evidence for F0s derived from the same

speaker. Therefore, it is suggested that more studies should be proposed to look at another strategy if F0 is still used for Indonesian FSV system, e.g. using pitch alignment features [11], adjusting advanced pitch settings and framing sentences by using the intonation system [7], and considering the effect of pitch span on intonational plateau [12]. Highlighting some functional aspects in the conversational structure in spontaneous dialogue [13] is also necessary to consider in getting the required K and Uk samples. Moreover, insights on phonological variation for discriminatory aspects in forensic speaker verification [14] and

other related aspects in forensic phonetics [15, 16] and forensic linguistics [17] are suggested to the system as some of theoretical

backgrounds to provide linguistic evidence in legal settings.

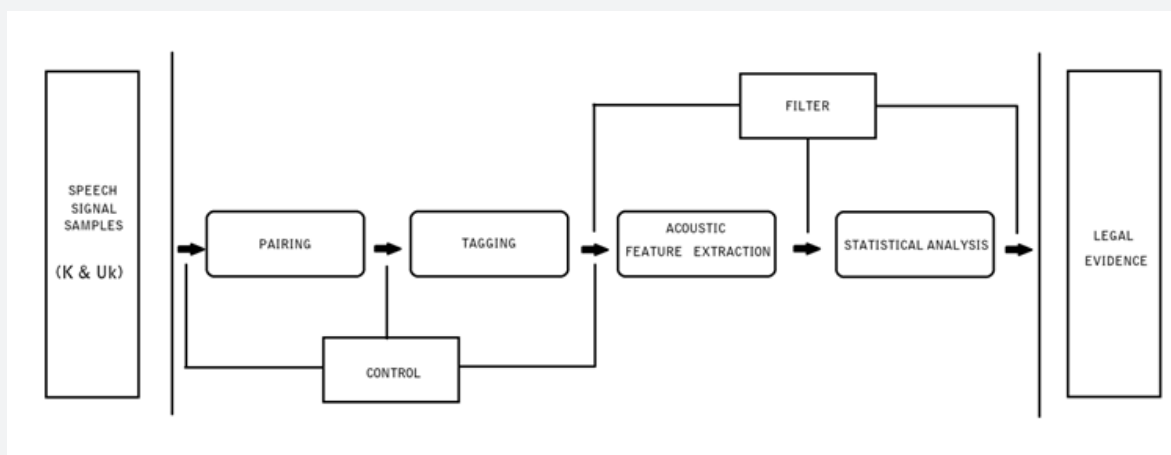


Figure 7: Steps in forensic speaker verification system.

The experimental study on Indonesian FSV system leads us to propose a scenario for forensic speaker verification [Figure 7]. In the system, K and Uk samples are paired for the same words. For tagging, syllables are derived from the paired words. Starting from pairing to the end of tagging, a control is conducted to scrutinize the effects of telephone transmission. Then, it moves forward to the acoustic feature extraction. Starting from the acoustic feature extraction to the end of statistical analysis, a filter is implemented to get high qualified performance. The filter is in terms of what acoustic features will be analyzed, what the theoretical backgrounds are for the analysis, and how the factors of reliability and validity can be achieved. Finally, the result is ready to present as legal evidence.

Acknowledgment

We gratefully acknowledge the supports from the Martin Centre for Applicable Linguistics at Shanghai Jiao Tong University and the Centre for Studies in Linguistics at Bandar Lampung University. This work is also supported by the Chinese National Research Project (16BYY501). We also thank the reviewers and appreciate their comments.

References

- Olsson J (2004) Forensic linguistics: An introduction to language, crime and the law. Continuum, London.
- Rose P (2002) Forensic speaker identification. Taylor & Francis, London.
- Rose P (2005) Forensic speaker recognition at the beginning of the twenty first century: An overview and a demonstration. Australian Journal of Forensic Sciences 37: 49-72.
- Al-Azhar MN (2011) Audio forensic: Theory and analysis. DFAT Puslab-for.
- Boersma P, Weenink D (2015) Praat: Doing phonetics by computer [Computer program]. Retrieved on 2 December 2015 from <http://www.praat.org/>
- Sarwono J, Mandasari, Suprijanto (2010) Forensic speaker identification: An experience in Indonesian court. In Proceedings of 20th International Congress on Acoustics, ICA 2010. 23-27 August, Sydney, Australia.
- Susanto (2014) Intonation in Bahasa Indonesia: A systemic functional study. Unpublished PhD thesis, EFLU.
- McDougall K, Nolan F, Hudson T (2015) Telephone transmission and earwitnesses: performance on voice parades controlled for voice similarity. *Phonetica* 72(4): 257-272.
- Nolan F, McDougall K, Hudson T (2013) Effects of the telephone on perceived voice similarity: implications for voice line-ups. *International Journal of Speech, Language and the Law* 20(2): 229-246.
- Lawrence S, Nolan F, McDougall K (2008) Acoustic and perceptual effects of telephone transmission on vowel quality. *International Journal of Speech, Language and the Law* 15(2): 161-192.
- Nolan F (2002) Intonation in speaker identification: an experiment on pitch alignment features. *Forensic Linguistics* 9(1): 1-21.
- Knight RA, Nolan F (2006) The effect of pitch span on intonational plateaux. *Journal of the International Phonetic Association* 36(1): 21-38.
- Susanto (2015) A systemic functional study on the conversational structure of an Indonesian spontaneous dialogue. *English Review: Journal of English Education* 4(1): 122-129.
- Moosmüller S (1997) Phonological variation in speaker identification. *Forensic Linguistics* 4(1): 29-47.
- Nolan F (1991) Forensic phonetics. *Journal of Linguistics* 27(2): 483-493.
- Nolan F (2009) The phonetic bases of speaker recognition. Cambridge University Press, Cambridge.
- Olsson J (2008) Forensic linguistics. Continuum, London.



This work is licensed under Creative Commons Attribution 4.0 License
DOI: [10.19080/JFSCI.2017.05.555664](https://doi.org/10.19080/JFSCI.2017.05.555664)

**Your next submission with Juniper Publishers
will reach you the below assets**

- Quality Editorial service
- Swift Peer Review
- Reprints availability
- E-prints Service
- Manuscript Podcast for convenient understanding
- Global attainment for your research
- Manuscript accessibility in different formats
(Pdf, E-pub, Full Text, Audio)
- Unceasing customer service

Track the below URL for one-step submission

<https://juniperpublishers.com/online-submission>.