Proceedings of 20th International Congress on Acoustics, ICA 2010

23-27 August 2010, Sydney, Australia

Forensic speaker identification: an experience in Indonesians court

J. Sarwono, M.I. Mandasari, and Suprijanto

Department of Engineering Physics, Insitut Teknologi Bandung, Jl. Ganesa 10, Bandung, INDONESIA
jsarwono@tf.itb.ac.id

PACS: 43.60.DH , 43.70.BK, 43.72.UV

ABSTRACT

Since 2008, intercepted speech communication can be accepted as one of legal evidence in Indonesians court. The first case of using this evidence in the court was a case of corruption handled by the Indonesian Corruption Eradication Commission (KPK). To take the speech evidence to court, such speech evidence should be examined by a Speaker Identification Forensic Examiner to verify the identity of the speaker. Two groups of speech samples were needed. The first group is the intercepted voice belongs to unidentified speaker, while the second is the voice recording of an identified suspect. Examiner serves to compare the two sample groups and confirm the identity of speaker's voice in the first sample. This examination process is strongly influenced by processes that occur while the second sample group was recorded from the suspect. If the defendant cooperated with investigators, then the identification process will go well, and vice versa. This paper shows the development of Forensic Speaker Identification methods which are used in handling corruption cases in Indonesian court. Results of identification using real data as well as a simulated speech signal were also discussed.

BACKGROUND

The development of forensic speaker identification in Indonesia was started from when speech signal samples used as legal evidence in the court of corruption cases. Before taken as legal evidence to the court, the speech signal samples should be first examined by the expert. The process or speech signal samples analysis, or known as Speaker Verification (SV) system, is a system that identify the subject of the Unknown (UK) sample by comparing it with the Known (K) sample. K sample is speech signal recording taken from the suspected person with the known identity, and UK sample is speech signal recording taken from suspected conversation. For example, the UK sample is taken from telephone conversation recording. Both of the samples (K and UK) based on the Voice Comparison Standards that came from American Board of Recorded Evidence, a board of the American College of Forensic Examiners [1]. The legal evidence which is proceed from SV analysis is depend on the method of analysis that used in the SV system.

ANALYSIS METHOD OF SPEAKER VERIFICATION

The method of analysis that used in SV system to provide as legal evidence from speech signal samples in Indonesian Court Justice, consist of four main steps. The four steps are pairing, tagging, acoustic features extraction, and statistical analysis as can be seen in Figure 1.

On the pairing step, words from the K sample were paired to the same words of the UK sample in order to make SV analysis more efficient. Standard number of pairing words that commonly used is ranging from 10 until 20 words, but sometimes higher depending on desirable result. In the tagging step, a word of the speech signal samples were divided into syllables, because the speech signal segment per syllable will be used in the statistical analysis step.

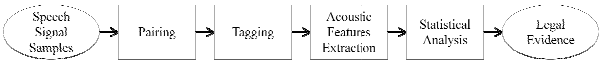

Figure 1. The four steps in the analysis method of SV system that used in Indonesian court.

The SV system use Praat software for the acoustic features extraction step, since Praat is a freeware that can be used as analysis and manipulate tools for speech signal. Pitch extraction is using an autocorrelation method, and formant extraction is using the LPC-Burg method. Until now, the SV system for producing legal evidence in Indonesia Court still use only pitch and formant as the acoustic features. It is however, bandwidth formant can also be used in some special cases of statistical analysis.

Statistical analysis was carried out to make statistic features of acoustic features set from the K and UK samples.

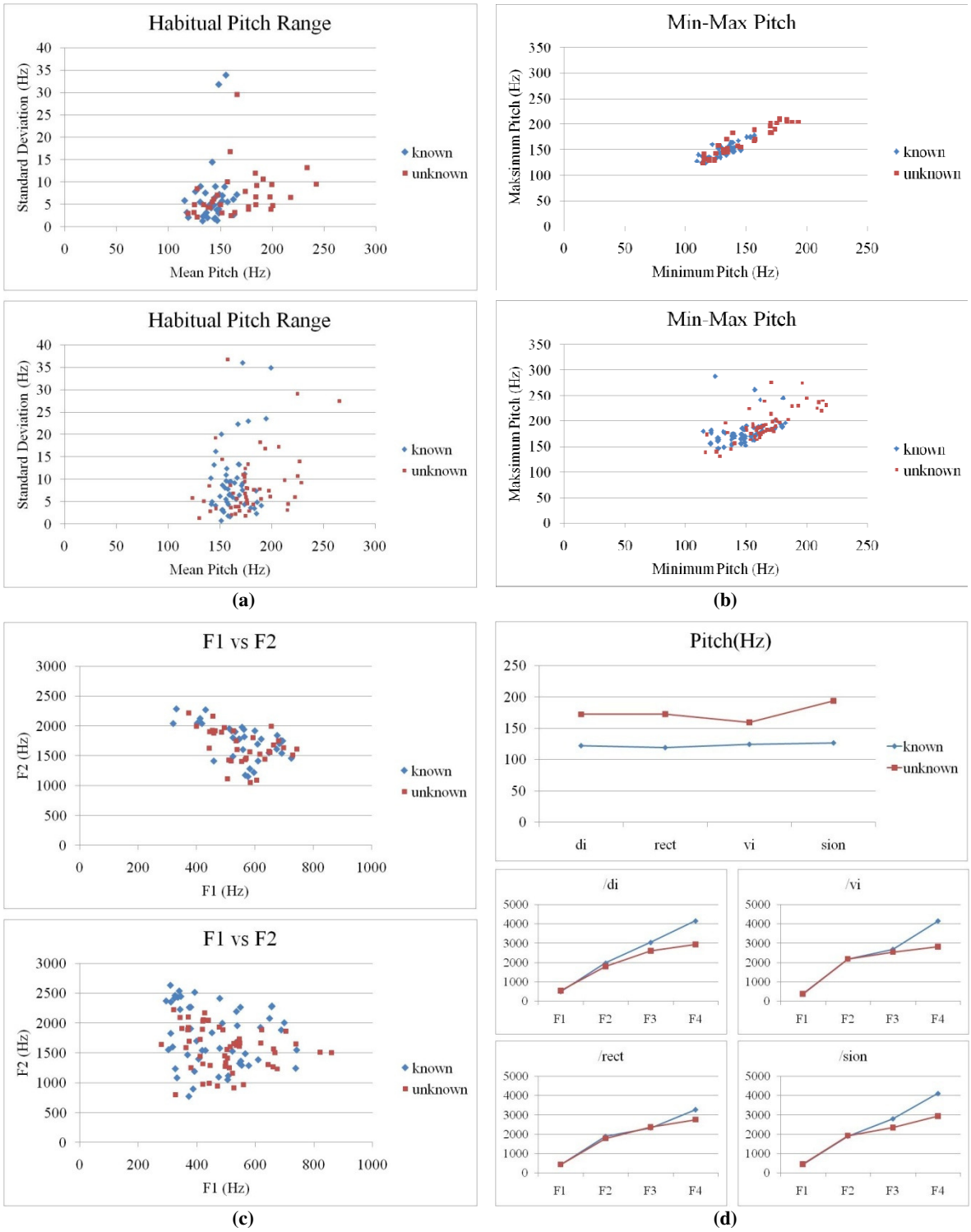

Figure 2. Example of acoustic and statistic features for: (a) habitual pitch range, (b) minimum-maximum pitch, (c) first-second formant, and (d) speaking style for pitch and formant.

Statistics features which are used in the analysis are the followings: the spreading of mean pitch to standard deviation pitch –which known as habitual pitch range as we seen in Figure 2(a), and the spreading of minimum pitch to maximum pitch as we seen in Figure 2(b). For formant acoustic features, the spreading of first formant (F1) to second formant (F2) statistic feature as we seen in Figure 2 (c) was used. In addition of these spreading statistic features, a supra-segmental analysis by seeing the contour of the pitch and formant were carried out to see the speaking style of the speaker from the speech signal samples as seen in Figure 2(d).

EVALUATION OF THE CURRENT SV SYSTEM

Up to now, there were more than 30 SV cases in preparing legal evidence for corruption courts in Indonesia have been done, with more than 95% of the cases give a result that both of the samples (K and UK) are coming from the same speaker using this SV system. So far, the current SV system meets the demand for presenting legal evidence in Indonesian court. It is however, the system performances improvement is continuously carried out. The aim of this development is to make comparison method that can support the result from the current SV system.

In the implementation of this current SV system, there is a difficulty related to how to get the K sample due to the lack of the available standard recording equipments, noise free room and recording technique. The development to overcome this problem is currently being planned. Other problems affecting the current SV system are the emotional and health condition of the suspect during the recording of the K samples in the interrogation room. An active disguising speech by the suspect was also arosen in several cases.

SPEAKER VERIFICATION SYSTEM DEVELOPMENT

There is quite a view of developments for SV system in Indonesia. One of them was the development of acoustic features extraction method and algorithm using Matlab instead of Praat software[2]. For the statistical analysis process, the use Anova (Analysis of Variance) was also considered as mathematical tools to find the significant difference between acoustic features in K and UK samples. The result shown that Anova has not give the best performance as mathematical tools in SV system statistical analysis because the result of Anova is depend on the condition of K and UK samples as the inputs for the system[3]. Another proposal to enhance the urrent SV system was to apply Likelihood Ratio (LR) in presenting the legal evidence to the court [3]. An ongoing research is currently carried out to tackle the problem of active disguishing speech by looking at the possibility of someone shifting their usual pitch and formant frequency range. Indonesian shadow puppet *dalang* (narrator and puppeteer) were used as the source of speech samples.

The development of SV system in Indonesia is not only in acoustic features extraction and statistical analysis method, but also in searching new features both in acoustic and statistic features. Experimental research using Indonesian trained speaker to produce an active disguised speech shown that bandwidth formant is a robust acoustic feature in case disguised speech signal samples [4]. For the statistic features, the experiment found that second, third, and fourth statistical moments (variance, skew, and kurtosis) give no significant difference between K and UK samples for standard deviation pitch acoustic feature. This result means that those statistic features can be potential features for improving the performance of SV system in Indonesia.

CONCLUSION

The current Speaker Verification system has a good performance in preparing legal evidence for corruption court in Indonesia. It is however the system needs to be improved in the area of searching the new acoustic and statistic features, and better extraction and statistical analysis method for giving the best system that can support the current SV system.

REFERENCES

1 S. Cain, “American board of recorded evidence – voice comparison standards” *Expert Pages* (2008).

2 G. Y. Koesoemah, M. I. Mandasari, Suprijanto and J. Sarwono., “Comparative study of speech signal processing algorithm for the extraction of pitch and formant” *Proceedings on International Conference in Instrumentation, Control and Automation* (2009).

3 M.I. Mandasari, “Pengembangan sistem identifikasi pengucap untuk aplikasi forensik di Indonesia (Development of speaker verification system for forensic application in Indonesia)” *Thesis, Engineering Physiscs Department, Institut Teknologi Bandung, Indonesia* (2008).

4 Anugerah, M. I. Mandasari, J. Sarwono and Suprijanto , “Acoustic speech analysis for disguised voice identification” (invited speakers) *Proceedings on International Conference in Instrumentation, Control and Automation* (2009).