Analysing Forensic Speaker Verification by Utilizing Artificial Neural Network

Susanto Susanto[1,*] Deri Sis Nanda[2]

[1] *Universitas Bandar Lampung, Indonesia*
[2] *Universitas Bandar Lampung, Indonesia*
[*]*Corresponding author. Email: susanto@ubl.ac.id*

ABSTRACT
In this paper, we describe the use of Artificial Neural Network (ANN) to compute the acoustic features in analysing forensic speaker verification. In the computation, there are two datasets derived from speech recording of a simulated human trafficking crime, namely Forensic Evidence Data (FED) and Comparative Evidence Data (CED). In both datasets, sound segmentation is performed and then the acoustic features (Formant Frequencies F1, F2, F3, and F4) are extracted. The acoustic feature values are computed with ANN to predict an output with a targeted sound classification /a/, /i/ and /u/. The results are interpreted as forensic evidence against sound data in recorded evidence. With a result rate of more than 80%, this method might be studied more deeply to be developed and applied in evaluating recorded sound evidence for the legal case process.

Keywords: *Acoustic Features, Artificial Neural Networks, Forensic Linguistics, Formant Frequency*.

1. INTRODUCTION

Forensic linguistics can be defined as the application of linguistics in the legal field [1]; [10]; [2]; [3]. Hence, forensic linguistics is the application of linguistic science that can include theories, methods and language analysis for legal purposes, for example, criminal law, civil law, constitutional law, customary law, environmental law, and others. The application of linguistics in the field of law continues to increase. One of them is in the settlement of legal cases, for example in the settlement of cases of defamation, threats, extortion, murder, disputes, plagiarism, corruption and so on [4]; [5]; [7]; [8]). Apart from being used to help resolve legal cases, forensic linguistics is also used in counter-terrorism and intelligence efforts, including identifying and verifying intercepted voice data.

Developments in the settlement of legal cases make the role of forensic linguistics more dynamic so that various studies of forensic linguistics are required not only as a scientific discipline but also as a professional field of linguistics. The studies that can be done from the scope of language analysis include language analysis on legal products, legal case trials, legal documents and legal evidence. In educational and training institutions, linguists introduce and teach forensic linguistics. Research in this field is also developing, marked by the existence of various writings on reports of empirical research results. In addition, linguists also form professional organizations. Among them are the International Association of Forensic and Legal Linguistics (https://www.iafl.org/) and the International Association of Forensic Phonetics and Acoustics (https://www.iafpa.net/). In Indonesia, research centres have been established that specialize in forensic linguistics studies, including the Centre for Studies in Linguistics at Universitas Bandar Lampung in 2015 (https://csl.ubl.ac.id/). Meanwhile, the Indonesian Forensic Linguistics Community (KLFI-Komunitas Linguistik Forensik Indonesia) was also formed in 2015 (https://klfi.weebly.com/) and it successfully held its first conference in Pekanbaru (February 2019).

1.1. Forensic linguistics in the legal process, product and evidence

In general, the field of forensic linguistics can be divided into 3 (three) groups. The first is the study of language in the legal process. This can be exemplified by language studies in police investigations and court proceedings. In the police investigation process, research can be carried out to determine the strategy of police investigators in examining a criminal case [9]; [10]; [11]; [12]). While in the trial process in the court, research can be conducted to find out how judges,

prosecutors, lawyers, witnesses and defendants communicate [4]; [13]; [6].

The second is the study of language in legal products. This study is exemplified in research on the language of legislation and also research on the language of court decisions. Research in this study can be carried out to understand the use of language that is specifically used in legal products [14]. The third is the study of language in legal evidence. It analyses the language used in the documents that are the cause of dispute cases, such as employment contract documents [15]; [16]) or patent documents [18]). In addition, it can also be carried out in language research on identifying voice telephone conversations [19]. Through telephone conversations, messages conveyed may cause legal problems if the message contains things that are prohibited by law such as threats, blackmail, or insults. These study groups in forensic linguistics can be carried out separately or integrated, depending on what research objectives are going to achieve.

1.2. Artificial Neural Network

As ideally in forensic studies, the evidence and its explanations in forensic speaker identification and verification provided by experts contain an objective evaluation of the acoustic features obtained from recorded evidence [20]. Then, the evaluation results are compared with those of the speech recorded by the defendant. Comparative evaluation results are used to support the expert's arguments in his or her testimony in determining the level of sound similarity in the recorded evidence with the defendant's voice.

In this paper, we report the results of a forensic linguistic study related to the application of Artificial Neural Networks (ANN). It is part of the measurement dimension in forensic linguistics [7]). The computing is applied to the values of formant frequency as the acoustic feature which is extracted from the data. ANN is a computational-based adaptive computing network, whose design follows neural [21]. Currently, the use of ANN is growing not only in the field of artificial intelligence system studies but also in various other fields including forensic studies [22].

We propose a model for using ANN in performing the prediction for evaluating the acoustic features in the data. In the model, there are two groups of data. In both groups, the sound segmentation is performed and then its acoustic features are extracted. The values of the acoustic features are computed in ANN to predict the targeted sound classifications as the outputs.

2. METHODOLOGY

In this study, we use the voice recording data of a telephone conversation between two speakers (Sp1=male,24:3; and Sp2=male,24:7). The conversation is in the context of a simulated human trafficking crime. The speech sounds of Sp1 in the conversation are for Forensic Evidence Data (FED). In addition, we use other speech sounds of Sp1 from a separated recording for Comparative Evidence Data (CED). In both FED and CED, 10 words are segmented, namely *aman* 'safe', *anak* 'child', *bayi* 'baby', *bisa* 'be able', *cari* 'search', *itu* 'that', *jual* 'sell', '*masalah'* 'problem', *takut* 'fear' and *umur* 'age'. In total, the syllable sounds of those words contain 21 nucleus elements. Thus, for both FED and CED, there are 42 nucleus elements altogether. The recordings for the data were conducted at the Centre for Studies in Linguistics, Universitas Bandar Lampung, Indonesia.

The ANN model used in this study is backpropagation ANN which is built in the R programming language [23]. Backpropagation ANN is used with three layers: input layer (F1-F4), hidden layer, and output layer (/a/, /i/, and /u/). In this ANN model, FED is used as the training data and CED as the testing data. From FED, specific words were selected and segmented. Then, the extraction of their acoustic features at the syllable nucleus elements was done by using Praat software [24]. The values of the extracted acoustic features were computed with ANN to predict the sound classification as the outputs. Meanwhile, the inputs are the values of those extracted acoustic features, i.e. formant frequencies F1, F2, F3, and F4. Then, in CED, the same sounds were also segmented and its acoustic features as the inputs are tested in ANN. The classification target for ANN is the phoneme sound in the nucleus elements, namely /a/, /i/, and /u/. There are 63 extracted points for each formant frequency in each data. Hence, in total, 504 extracted points were used in the study.

3. RESULTS AND DISCUSSION

3.1. Acoustic Features of Nuclei in Syllable Unit

A syllable is a unit in speech sound consisting of three elements, namely onset, nucleus and coda [25]; [26]; [27]; [28]; [29]. In syllables, the onset is at the beginning, the nucleus is in the middle and the coda is at the end. The onset and coda elements usually consist of consonant sounds, while the nucleus elements usually consist of vowel sounds. In Systemic Functional Grammar [30]; [31], the syllable unit is one of the speech sound units that are below the rhythm unit and above the phoneme unit. For the study, the syllable unit is considered for providing the context of the target phonemes in the nuclei.

In the nuclei of syllable sounds, the extracted acoustic features are the formant frequencies F1, F2, F3 and F4. Formant frequencies are frequencies that result from acoustic resonance in the human vocal tract [32];

[33]. F1-F4 were extracted precisely at the nucleus element of the syllable. This is to get pure values that don't mix with those in the onset and coda elements. To further avoid mixing these values, the extract points are not taken directly starting from the point on the border between the nucleus element and the onset element to the point on the border between the nucleus element and the coda element. Instead, we only extract the values from the central point. See Table 1 and Table 2 for the listed values of F1-F4 at the extracted points of the nuclei in FED and CED respectively.

3.2. Computation with ANN

Figure 1 shows the ANN diagram with four formant frequencies (F1, F2, F3, and F4) at the input layer and three phoneme sounds (/a/, /i/ and /u/) at the output layer. In the system, there is a hidden layer between the input and output layers. Thus, the counting network of the system consists of several layers of units (input, hidden and output) that are interconnected with each other. This looks like the nerve cells in the brain. Its calculation system can adapt to the input information from outside and inside that flows through the network [34]; [17].

In the counting process, the number of units in the hidden layer is created. It started with 1 to 10 units. From the calculation in the ANN model, CED was tested for similarity by referring to FED which had been trained in the ANN. The result rates can reach more than 80% at the 10-unit stage in the hidden layer as illustrated in Figure 2. It can still be improved in several ways.

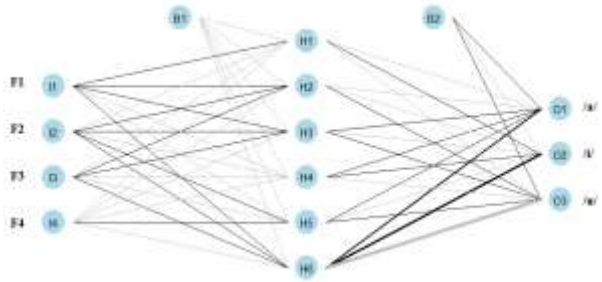

Figure 1 ANN diagram with 4 units (I1-I4) in the input layer for F1-F4, 6 units (H1-H6) in the hidden layer and 3 units (O1-O3) in the output layer for /a/, /i/, and /u/.

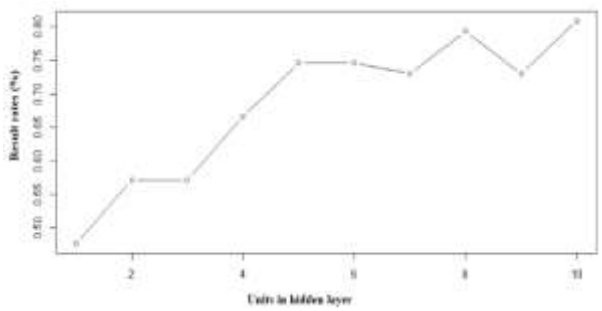

Figure 2 The result rates with up to 10 units in the hidden layer.

One of the ways is to add another variable acoustic feature to increase the number of units in the input layer. Another way is to choose a more varied syllable sound so that the target phoneme sound classification can also include the phonemes /e/ and /o/. In addition, adding the hidden layer could be in the system.

Table 1. F1-F4 at the extracted points of the nuclei in FED

No	F1 (Hz)	F2 (Hz)	F3 (Hz)	F4 (Hz)	Code[a]
1	592	1382	2858	3712	DBF_I-SIL1_K1
2	626	1358	2832	3710	DBF_I-SIL1_K1
3	577	1339	2806	3662	DBF_I-SIL1_K1
4	519	1183	2905	3579	DBF_I-SIL2_K1
5	546	1235	2887	3539	DBF_I-SIL2_K1
6	545	1165	2896	3560	DBF_I-SIL2_K1
7	816	1416	2964	3816	DBF_I-SIL1_K2
8	776	1364	2960	3675	DBF_I-SIL1_K2
9	675	1334	2928	3650	DBF_I-SIL1_K2
10	782	1275	2857	3875	DBF_I-SIL2_K2
…	…	…	…	…	…
63	473	817	3079	4312	DBF_I-SIL2_K10

[a]*Note: DBF_I-SIL1_K1 means the point is derived from the nucleus /i/ in the the frist syllable of Word 1 in FED.*

Table 2. F1-F4 at the extracted points of the nuclei in CED

No.	F1 (Hz)	F2 (Hz)	F3 (Hz)	F4 (Hz)	Code[b]
1	729	1223	2584	3752	DBP_I-SIL1_K1
2	702	1182	2576	3750	DBP_I-SIL1_K1
3	652	1084	2551	3685	DBP_I-SIL1_K1
4	565	1047	2683	3585	DBP_I-SIL2_K1
5	573	1176	2731	3606	DBP_I-SIL2_K1
6	539	1293	2793	3596	DBP_I-SIL2_K1
7	548	893	2576	3888	DBP_I-SIL1_K2
8	486	883	2862	3707	DBP_I-SIL1_K2
9	768	849	2888	3603	DBP_I-SIL1_K2
10	782	1387	2956	4322	DBP_I-SIL2_K2
…	…	…	…	…	…
63	417	742	3133	3250	DBP_I-SIL2_K10

[b]*Note: DBP_I-SIL1_K1 means the point is derived from the nucleus /i/ in the first syllable of Word 1 in CED.*

4. CONCLUSION

Forensic linguistics can be seen as applied linguistics because it uses the application of language analysis in either phonetics, phonology, lexico-grammatical, discourse semantics and other linguistic elements into another science, i.e. legal science for the benefit of the legal process. For example, in phonetics and phonology, language analysis is applied for forensic speaker verification.

In the study, we have discussed the use of ANN to compute the acoustic features in analysing forensic speaker verification. The segmentation of the speech sounds and their acoustic features (F1-F4) was carried out in two datasets FED and CED. Then the ANN was applied to predict an output with a targeted sound classification /a/, /i/ and /u/. With a result rate of more than 80%, this method might be studied more deeply to be developed and applied in evaluating recorded sound evidence for the legal case process.

AUTHORS' CONTRIBUTIONS

The first author conceived of the presented idea developed the model and performed the experiments. The second author helped supervise the experiments and verified the analytical method. All authors discussed the results and contributed to the final manuscript.

ACKNOWLEDGMENTS

This work was supported by Kementerian Riset dan Teknologi/Badan Riset dan Inovasi Nasional, Indonesia [grant number 1424/SP2H/LT/LL2/2021].

REFERENCES

[1] Coulthard, M., & Johnson, A. (2010). *The Routledge handbook of forensic linguistics*. New York, NY: Routledge.

[2] Gibbons, J., & Turell, T. (2008). *Dimensions of forensic linguistics*. Amsterdam: John Benjamins.

[3] Olsson, J. (2004). *Forensic linguistics: An introduction to the language, crime and the law*. London: Continuum.

[4] Shuy, R. W. (1993). *Language crimes: The use and abuse of language evidence in the courtroom*. Cambridge, Mass.: Blackwell Publishers

[5] Solan, L. M., & Tiersma, P. M. (2005). *Speaking of crime: The language of criminal justice*. Chicago: University of Chicago Press.

[6] Susanto, S. (2016, May). Language in courtroom discourse. In *International Conference on Education and Language (ICEL)* (p. 26).

[7] Susanto, S., & Nanda, D. S. (2020). Dimensi analisis bahasa dalam linguistik forensik. *IJFL (International Journal of Forensic Linguistic), 1*(1), 17-22.

[7] Susanto, S. (2020, January 19). *Potensi dan tantangan linguistik forensik di Indonesia*.

[8] Susanto, S., Zhenhua, W., Yingli, W., & Nanda, D. S. (2020, January 13). *Forensic linguistic inquiry into the validity of F0 as discriminatory potential in the system of forensic speaker verification*

[9] Baldwin, J. (1993). Police interview techniques: Establishing truth or proof? *The British Journal of Criminology*, *33*(3), 325-352.

[10] Gibbons, J. (1996). Distortions of the police interview revealed by video tape. *International Journal of Speech, Language and Law*, *3*(2), 289-298.

[11] Gregory, M. (2011). A Comparison of US police interviewers' notes with their subsequent reports. *Journal of Investigative Psychology and Offender Profiling*, *8*(2), 203-215.

[12] Heydon, G. (2012). Helping the police with their enquiries: Enhancing the investigative interview

[13] Solan, L. M. (1993). *The language of judges.* Chicago: The University of Chicago Press.

[14] Wagner, A., & Cacciaguidi, S. (2006). *Legal language and the search for clarity*. Bern: Peter Lang.

[15] Fawzy, S. A., & El-adaway, I. H. (2012). Contract administration guidelines for managing conflicts, claims, and disputes under world bank–funded projects. *Journal of Legal Affairs and Dispute Resolution in Engineering and Construction*, *4*(4), 101-110.

[16] Watts, V. M., & Scrivener, J. C. (1993). Review of Australian building disputes settled by litigation. *Building Research & Information*, *21*(1), 59-63.

[17] Tomandl, D., & Schober, A. (2001). A Modified General Regression Neural Network (MGRNN) with new, efficient training algorithms as a robust 'black box'-tool for data analysis. *Neural Networks*, *14*(8), 1023–34.

[18] McJohn, S. (2017). Top tens in 2016: Patent, trademark, copyright and trade secret cases. *Northwestern Journal of Technology and Intellectual Property*, *15*(2), 77-110.

[19] Künzel, H. J. (2001). Beware of the "telephone effect": The influence of telephone transmission on the measurement of formant frequencies. *Forensic Linguistics*, *8*, 80–99.

[20] Rose, P. (2002). *Forensic speaker identification*. London: Taylor & Francis.

[21] Fausett, L.V. (1994). *Fundamentals of neural networks: Architectures, algorithms, and applications.* New Jersey: Prentice Hall.

[22] Kozan, N.M., Kotsyubynska, Y.Z., & Zelenchuk, G.M. (2017). Using the artificial neural networks for identification unknown person. *IOSR Journal of Dental and Medical Sciences*, *16*(4), 107–113.

[23] R Core Team. (2018). *R: A language and environment for statistical computing*. Software.

[24] Boersma, P., & Weenink, D. (2019). Praat: Doing phonetics by computer [Computer program]. Version 6.0.50.

[25] Duanmu, S. (2008). *Syllable structure: The limits of variation*. Oxford: OUP.

[26] Goldsmith, J. (2011). The syllable. In J. Goldsmith, J. Riggle & A.C.L. Yu (Eds.), *The Handbook of phonological theory.* Chichester: Wiley-Blackwell.

[27] Gordon, M.T. (2006). *Syllable weight: Phonetics, phonology, typology*. New York, NY: Routledge.

[28] Jong, K.D.E. (2009). Temporal constraints and characterising syllable structuring. *Phonetic Interpretation Papers in Laboratory Phonology VI*, 253–268.

[29] Zec, D. (2009). The syllable. In P.D. Lacey (ed.) *The Cambridge handbook of phonology*. Cambridge, UK: Cambridge University Press.

[30] Halliday, M.A.K., & Greaves, W.S. (2008). *Intonation in the grammar of English*. London: Equinox Publishing.

[31] Halliday, M.A.K., & Matthiessen, C.M.I.M. (2004). *An introduction to functional grammar*. London: Edward Arnold.

[32] Johnson, K. (2003). *Acoustics and auditory phonetics.* Malden, MA: Blackwell.

[33] Titze, I. R. (1994). *Principles of voice production*. Englewood Cliffs, NJ: Prentice Hall.

[34] Hastie, T., Tibshirani, R., & Friedman, J. (2009). *Elements of statistical learning*. New York, NY: Springer.