IJFL (International Journal of Forensic Linguistic)
Vol. 3, No. 1 April 2022, Page 69-74

P-ISSN: 2723-1542
E-ISSN: 2723-5254
Available Online at https://www.ejournal.warmadewa.ac.id/index.php/ijfl/index

AN ANALYSIS OF MALE'S VOICE SPECTRUM CHARACTERISTICS THROUGH PRAAT SOFTWARE

Asep Surahman
Universitas Warmadewa
asep@mybalitrips.com

How to cite:

Surahman, A. (2022). An Analysis of Male's Voice Spectrum Characteristics Through PRAAT Software. *IJFL (International Journal of Forensic Linguistic)*, Vol. 3 (1), 69-74. Doi: https://doi.org/10.22225/ijfl.3.1.4776.69-74

Abstract- This study is aimed to analyze the characteristics of male voices recording by using PRAAT software. Technically, this study was conducted by using direct measurement method with the assistance a laptop equipped with PRAAT software. In addition, the data were collected from the sound of men which has recorded by using a recording application which is available on smartphone that has been connected to a laptop. The sample of men voices consist one native speaker and 3 non-native speakers from the range age 22 to 30 years old. The basic frequency values obtained from the voice recordings were compared between the voices of native and non-native speakers, the formant values were obtained and the duration of time required in pronunciation. As result, this study found that the average basic frequency value for men (native speaker) shows a value that is almost the same as the basic frequency value for non-native speakers, it is caused by the shape of the sound-producing organs and conditions when recording sound. The value of the first formant (F1) and the fourth formant (F4) in men (native speaker) very flexible can be lower, higher and even the same to the non-native speaker. Therefore, the value of pitch and formant and the spectrum which has shown on PRAAT do not show the consistency and validity of the data so that in-depth analysis is needed with the help of other supporting media.

Keywords: Digital-forensic, PRAAT, Spectrum.

I. INTRODUCTION

The ability of humans to speak, especially in expressing something by meaningful sound has placed humans in a special stage compared to other creatures on earth. However, the human voice tends to show a significant difference. this is certainly caused by several factors such as age, gender, race, and others. In some situations, for identifying human voices without seeing faces, such as in telephone conversations, adult male or female voices are easier to identify than children's voices (Smith and Patterson., 2005).

Briefly, the development of forensic linguistics science which is also related to audio forensics and digital forensics has made a considerable contribution to several institutions or related parties such as prosecutors and the police to analyze and determine the validity of legal evidence in the form of language material such as speech and others. According to Olsson (2004) defines it as an application of linguistics in the context of crime, court proceeding, or arguments in law. Similarly, Coulthard and Johnson (2010) mention that forensic linguistics ranges from courtroom discourse and legal language to plagiarism.

Moreover, technological developments have contributed to the emergence of technological media in the forensic-linguistic process. Specifically, one of the language materials that commonly analyzed is voices recording. However, the voice recording is often used by someone to immortalize a conversation Directly or by telephone. According to The ITE Law No. 19 of 2016 mentions that voice recordings are one of the pieces of evidence, as described in Article 1. It's just that sound recordings cannot be used as evidence without going through a fairly long analysis process, which is carried out by an expert in the field of audio forensics.

In addition, Digital forensic science is by definition a combination from the disciplines of law and computer science in collect and analyze data from computer systems, networks, wireless communications, and storage devices digital data for later use as evidence in problem-solving in the realm of law. (Binyamin Widi Prasetya, dkk, 2008).

Therefore, in identifying the sound spectrum based on acoustic measurements for a man's or woman's voice by measuring the value of the fundamental (pitch) and formant frequencies. While in this study, the characteristics of the human voice spectrum were analyzed by determining the fundamental and formant frequency values.

Technically, sound is a mechanical compression or longitudinal wave that propagates through a medium and is produced by vibrations causing changes in the air pressure around it. The vibrations that occur have a certain wave pattern. The sound signal is a signal that varies with time, in a short period of time (5-100 milliseconds) its characteristics are stationary (Rabiner and Juang, 1993).

Furthermore, the sound waves are classified into two, namely pure waves (waves generated by a sound wave-producing device (tuning fork)) and complex/complex waves (waves produced by various sound wave-producing devices with different and irregular frequencies.). Specifically, voice is the air that comes out of the lungs through the throat vibrating the vocal cords, then some air through the nose and some air through the mouth (Gabriel, 2001). The sound production process is divided into two steps, namely, the production of audible sounds and control of sounds to produce phonemes.

Phoneme is the smallest sound quantity that has a functional function in language. The lungs pump to inhale and exhale air. The air exhaled by the lungs exits through an area called the glottal area. The vocal cords in this state vibrate to produce various types of sound waves. The air then passes through a passage called the pharynx. From the pharynx, air passes through two passages, namely through the nose and through the oral cavity. The tongue, teeth, lips and nose act as modulators to produce a variety of different sounds. Infants have a vocal tract length of between 6 to 8 cm, whereas in adults, males and females have a vocal tract length of between 15 to 18 cm.

In this study, the main software that is specifically used is PRAAT. PRAAT (Dutch means 'sound') is a freeware developed by Paul Boersma & David Weenink of the University of Amsterdam's Phonetic Sciences Department. PRAAT program is a program containing tutorials on the analysis and reconstruction of acoustic speech signals. This PRAAT software can be downloaded on the internet at the address: http://www.fon.hum.uva.nl/praat/.

According to Septiyansyah (2015) stated that Praat app is able to record sound from microphone or other audio devices, besides that this application can also reads sound from an input file or disk. With PRAAT, the user is able to see into the audio. Technically, PRAAT software can be used on different operating systems which can be seen on the web above. PRAAT program is a very flexible program for analyzing human speech. In addition, this program has a wide range of standard procedures such as spectrographic analysis, articulatory synthesis and neural networks. In this PRAAT program several topics such as:

a. Search manually
b. Creating a sound object
c. Processing signals
d. Waveform label

e. General analysis (wave, intensity, sonogram, pitch, duration)
f. Spectrographic analysis
g. Intensity analysis
h. Pitch analysis

Briefly, pitch is the perception of human hearing of the difference in the frequency of a sound or pitch (F0) is the period of repetition of pulses caused by the opening and closing of the vocal cords. Pitch has a scale that is logarithmic to the frequency of its formation. The unit of pitch is Hz. In normal conversation conditions, the level of habitual pitch ranges from 50-250 Hz for men and 120-500 Hz for women (M. Nuh AL-Azhar, 2012). Furthermore, Pitch or fundamental frequency of sound is determined by the lowest frequency of the vibration of the speaker's vocal cords. Pitch represents the basic tone of the human voice. High and low pitch is influenced by age, gender and also influenced by vocal cords.

In speech and phonetics, formant also means acoustic resonance (Titze, 1994) of the human vocal tract. The frequency spectrum of a sound is often measured as peak amplitude, using a spectrogram or spectrum analyzer, although in vocals spoken with a high fundamental frequency, such as in a female or child voice, the resonant frequency may lie between the broad-scattered harmonics and therefore no visible peak. natural resonant frequency that occurs in the cavity of the sound field, vocals generally have 3 formants F1, F2, and F3 and to be more completely it can be F4.

Moreover, one of the digital forensic techniques is Voice Recognition, namely digital forensic techniques for detecting records Voice. People who have conversations can identity is known through audio forensic examination for speech recognition by comparison method, namely, comparing the voices in the recorded evidence (unknown sample) with sound recorded as Comparison (known sample). If the result of voice recognition indicates that the sound of the unknown sample is identical to known sample voice, then the voice in the conversation in the recording Evidence can be obtained from the owner of the vote Comparison (Septiyansyah, 2015).

Based on the background above, the problem in this study can be formulated, namely: how do the male's voice spectrum characteristics can be identified on PRAAT software? And what are the value of pitch and formant on male's voice recording?

Thus, based on the formulation above, the purposes of this study are to identify and find out the male's voice spectrum characteristics trough PRAAT software. In addition, this study is also aimed to find out the value of pitch and formant for the data sounds which has been recorded through smartphone.

II. METHODS

Technically, the direct measurement method was used in this study with the assistance a laptop equipped with PRAAT software. Data collection on the sample sound spectrum is done by recording sound using a recording application which is available on smartphone that has been connected to a laptop equipped with PRAAT software. The sample of men voices consist one native speaker and 10 non-native speakers. The basic frequency values obtained from the voice recordings were compared between the voices of native and non-native speakers, the formant values were obtained and the duration of time required in pronunciation.

Technically, this research was started by recording vocal sounds from various speakers. From one speech source will be recorded several random vowels that are separated automatically with the help of software. Each recorded sound, which is approximately two seconds long, is stored in a single file. The recording format is mono. The condition of the room during recording is not sterile from noise although it is limited to the noise that is common for an ordinary workspace in an office.

The recording stage is followed by the verification and editing stage. The recording process is carried out to ensure that the recorded sound is the same as the file name. This is intended to avoid data errors for certain phonemes being grouped into other phoneme groups. The verification process is done manually with human assistance.

After editing stage, the research proceeds to the transformation stage to the frequency domain with the help PRAAT. From this set of frequency responses, the dominant frequency for each phoneme will be determined.

Determination of the number of frequencies and formant that will be used as a database of Indonesian word sounds in this research is carried out by a) sorting the amplitude values of the frequency components from the largest to the smallest and b) calculating the summation value repeatedly starting from the top, second, third and so on until a certain percentage level is reached (compared to the total value of all frequency components). However, the PRAAT application and Microsoft excel are used analyze the data appropriately. In addition, PRAAT is used to search information from the comparison between records of native speaker's voice and recorded comparison voice. While, Microsoft Excel to use to measure the formant, pitch of each word spoken original and comparison.

III. RESULT

3.1 The Analysis of Spectrum on PRAAT

The spectrum of the sound is a representation of vibration energy at each frequency where the frequency in the sound spectrum varies. It can be seen that for the three sound spectrums there are differences in the amplitude and time intervals between the spectra (see figures 1 – 3).

This difference is caused by the intensity of the sound and the speaking speed of the speaker. The recorded voice samples have the similarity of every word spoken in the first pronunciation, second pronunciation and third pronunciation.

Figures 1 to 3show the shape of the sound spectrum which has a slight difference in spectrum height and duration even though the spoken word is the same from the native speaker.

Sample data: Native Speaker
Audio and spectrum word:

- *Celah*
- *Permintaan*
- *Nikel*

Figure 3.1
Spectrum word "Celah"

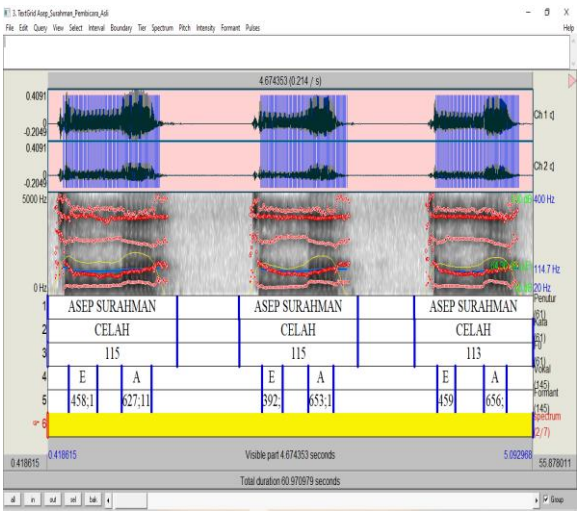


Figure 3.2
Spectrum word "Permintaan"

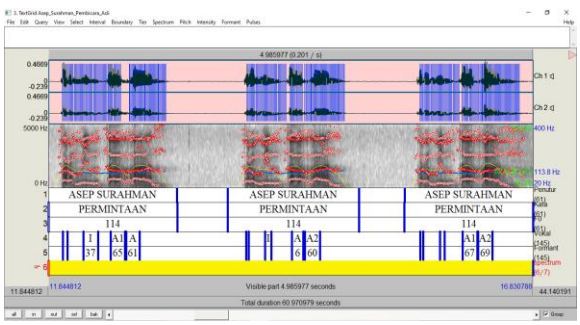
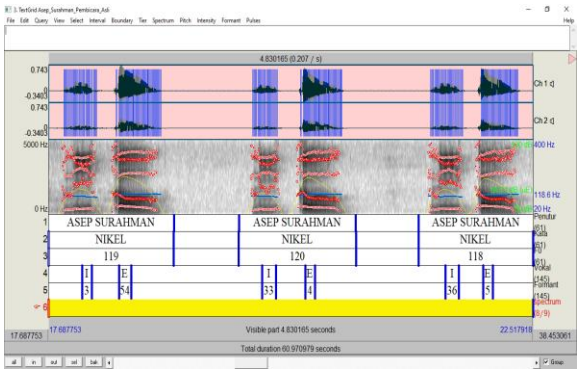


Figure 3.3
Spectrum word "Nikel"



From the figures above, it can be seen that spectrum of the male voice uttering:
Figure 3.1: *"Celah"* (a) the first pronunciation, (b) the second pronunciation and (c) the third pronunciation.
Figure 3.2: *"Permintaan"* (a) the first pronunciation, (b) the second pronunciation and (c) the third pronunciation.
Figure 3.3: *"Nikel"* (a) the first pronunciation, (b) the second pronunciation and (c) the third pronunciation.

Meanwhile, the spectrum from non-native speaker can be seen on the figures above.

Sample data: non-native speaker (male) consisted of: PP1 (comparison audio_1), PP2 (comparison_2), and PP3 (comparison audio_3).
Word list: PP1 (Celah), PP2 (Permintaan), PP3 (Nikel).

Figure 3.4
Spectrum word "celah" (non-native speaker)

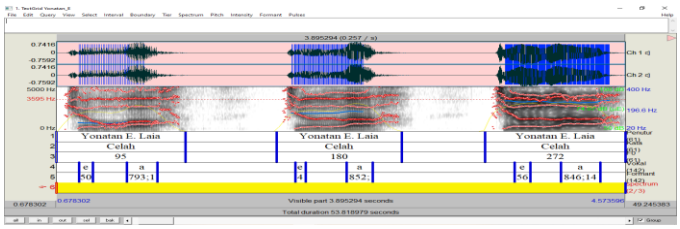


Figure 3.5
Spectrum word "Permintaan"
(non-native speaker)

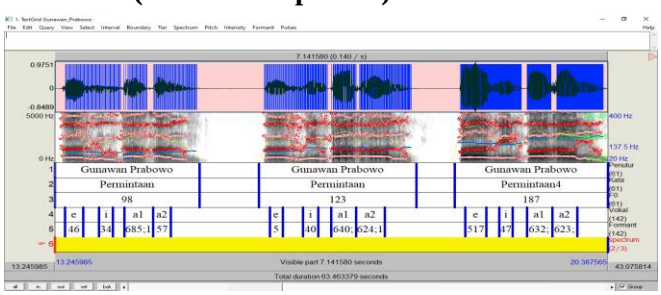
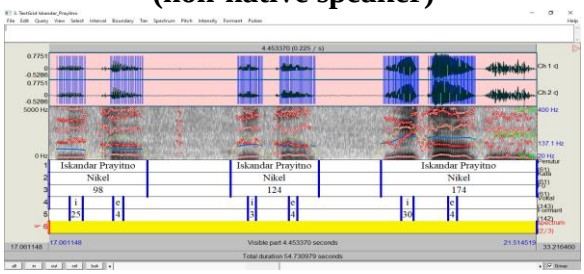


Figure 3.6
Spectrum word "Nikel"
(non-native speaker)



Based on the figures 4 to 6, the audio was taken from non-native speakers (male), their age range is 22 until 30 years old. Thus, it can be seen that the spectrum of each respondent compared to the native speakers are mostly different. In this case, the data result can be depended on the value of pitch and formant.

Table 3.1
Pitch value of speakers

No.	Speaker	Words	Pitch (Hz)		
			p-1	p-2	p-3
1	Native speaker	*Celah*	115	115	113
		Permintaan	114	114	114
		Nikel	119	120	118
2	PP1	Celah	95	180	272
3	PP2	Permintaan	98	123	187
4	PP3	Nikel	98	124	174

It is obvious the value of pitch on each respondent are different which caused by the different intonation, and frequency of their sounds. Technically, it was sounded form the lower tone, normal tone, and high tone.

In this case, the characteristics of male voices' spectrum can be indicated by the spectrogram and the value pitch which are indicated that if the tone are lower than the value of pitch is lower and spectrogram will be smaller, meanwhile if the tone of the voice is high then the value of pitch is higher about 150 Hz – 300 Hz and the spectrogram will be increased.

However, the value of formant of each data samples can be the indicators of the spectrum characteristics. Thus, the result of formant's value on each respondent are:

Figure 3.7
Formant's value (Native Speaker)

Penutur :	Asep Surahman					
No.	Kata	Penutur	F1	F2	F3	F4
1	Celah_1_e	1	458	1069	2672	3838
2	Celah_1_a	1	627	1188	2547	3763
3	Celah_2_e	1	392	1233	2674	3813
4	Celah_2_a	1	653	1184	2556	3756
5	Celah_3_e	1	459	1179	2697	3837
6	Celah_3_a	1	656	1214	2550	3757
13	Permintaan_1_e	1	586	1320	2220	3609
14	Permintaan_1_i	1	371	1548	2650	3584
15	Permintaan_1_a1	1	652	1319	2544	3813
16	Permintaan_1_a2	1	613	1316	2544	3771
17	Permintaan_2_e	1	537	1306	2297	3428
18	Permintaan_2_i	1	375	1869	2585	3537
19	Permintaan_2_a1	1	680	1348	2540	3784
20	Permintaan_2_a2	1	601	1346	2584	3852
21	Permintaan_3_e	1	546	1409	2342	3586
22	Permintaan_3_i	1	360	1640	2600	3531
23	Permintaan_3_a1	1	676	1357	2601	3799
24	Permintaan_3_a2	1	695	1415	2688	3830
25	Nikel_1_i	1	383	2144	2949	3421
26	Nikel_1_e	1	546	1112	2708	3662
27	Nikel_2_i	1	334	1901	2802	3444
28	Nikel_2_e	1	490	1139	2708	3628
29	Nikel_3_i	1	369	2162	2943	3433
30	Nikel_3_e	1	558	1077	2662	3664

Figure 3.8
Formant's value Non-Native speaker

SPEAKER	: NON-NATIVE SPEAKERS					
	Daftar Formant					
No.	Kata	Penutur	F1	F2	F3	F4
1	Celah_1_e	PP1	500	1594	2602	3592
2	Celah_1_a	PP1	793	1078	2787	3683
3	Celah_2_e	PP1	431	1839	2748	3614
4	Celah_2_a	PP1	852	1350	2751	3757
5	Celah_3_e	PP1	569	1447	2771	3679
6	Celah_3_a	PP1	846	1407	2821	3643
13	Permintaan_1_e	PP2	460	1389	2604	3805
14	Permintaan_1_i	PP2	347	1129	2478	3940
15	Permintaan_1_a1	PP2	685	1423	2892	3876
16	Permintaan_1_a2	PP2	579	1283	2642	3719
17	Permintaan_2_e	PP2	511	1327	2550	3849
18	Permintaan_2_i	PP2	401	1980	3082	3847
19	Permintaan_2_a1	PP2	640	1321	2680	3771
20	Permintaan_2_a2	PP2	624	1228	2710	3777
21	Permintaan_3_e	PP2	517	1248	2799	3720
22	Permintaan_3_i	PP2	470	2185	2596	3723
23	Permintaan_3_a1	PP2	632	1209	2671	3821
24	Permintaan_3_a2	PP2	623	1274	2776	3789
25	Nikel_1_i	PP3	253	1805	2961	3664
26	Nikel_1_e	PP3	496	1392	2862	4062
27	Nikel_2_i	PP3	314	1549	3028	3925
28	Nikel_2_e	PP3	453	1403	2925	4229
29	Nikel_3_i	PP3	300	2247	3064	3600
30	Nikel_3_e	PP3	470	1414	3184	4448

Figure 3.7 and 3.8 clearly shows that the value of the formant increases from the first formant (F1) to the fourth formant (F3). F1 and F4 values in native are lower than non-native speakers. However, it does not meet the value consistency due to the very varied level of intonation accuracy and tone pressure. so that the formant value of each respondent sometimes goes up and down. however, the difference in value will appear when the tone is adjusted to its level where the low tone will indicate a low value and the normal tone will be normal around 90-100 Hz, while the high tone will show the pitch and formant values which are also high ranging from 150-300 Hz.

IV. CONCLUSION

As conclusion, the results showed that the average basic frequency value for men (native speaker) shows a value that is almost the same as the basic frequency value for non-native speakers, it is caused by the shape of the sound-producing organs and conditions when recording sound. The value of the first formant (F1) and the fourth formant (F4) in men (native

speaker) very flexible can be lower, higher and even the same to the non-native speaker. Technically, the value of pitch and formant depend on the tone. The words are pronounced with high intonation shown the high value. However, if the tone is lower, then the value will lower also. Thus, the spectrum of men's voices on the PRAAT can be easily distinguished from the comparison voice which is on average high, very high, low or very low intonation. As result, the value of pitch and formant and the spectrum which has shown on PRAAT do not show the consistency and validity of the data so that in-depth analysis is needed with the help of other supporting media.

REFERENCES

Septiyansyah, Hasbi. 2015. Implementasi Metode Forensik Suara Pria Menggunakan Teknik Voice Recognize Untuk Analisis Kemiripan Suara Pada Media Alat Rekam Telepon Selular. Jurusan Sistem Informasi, Universitas Teknologi sepuluh November, Surabaya.

Olsson, J. (2004). Forensic linguistics: An introduction to language, crime and the law. London and New York: Continuum.

Coulthard, M., & Johnson, A. (2010). Handbook of forensic linguistics. London: Routledge.

M. Nuh AL-Azhar. 2012. Digital Forensic : Panduan Praktis Investigasi Komputer. Jakarta: Salemba Infotek.

Rabiner, L., Juang, B.H.1993. *Fundamental of Speech Recognation,* Pretice-Hall Inc., New Jersey.

Gabriel, J.F.2001. *Fisika Lingkungan*. Jakarta:Hipokrates

Smith, R. R., and Patterson, R. D. 2005. The interaction of glottal-pulse rate and vocal-tract length in judgements of speaker size, sex, and age. *J. Acoust. Soc. Am*., 118, 3177-3186.

Titze, I.R., (1994). "Principles of Voice Production", Prentice Hall.