Journal of Critical Reviews

ISSN- 2394-5125 Vol 7, Issue 8, 2020

REMOVING SILENCE IN THE SPEECH OF MALE AND FEMALE ADULTS FOR MALAY WORDS USING FRAMING AND WINDOWING METHOD

Rafizah Mohd Hanifa[1,2], Khalid Isa[2], Shamsul Mohamad[2]

[1]Center for Diploma Studies, Universiti Tun Hussein Onn Malaysia, 86400, Parit Raja, Batu Pahat, Johor.
[2]Faculty of Electrical and Electronic Engineering, Universiti Tun Hussein Onn Malaysia, Batu Pahat, Johor
E-mail: [1]rafizah@uthm.edu.my

Received: 20.05.2020 Revised: 17.06.2020 Accepted: 06.07.2020

Abstract
The most important domain in digital processing is speech signal processing. A variety of noise signals could degrade the original speech signal and make it unclear to listener. This research contributes to the literature by developing a program to remove silence from the original speech signal using framing and windowing method. The research also aims to assess voice quality features and the extent of variability of these features in specified groups of adults which are chosen based on gender. For this preliminary stage, eight subjects have been asked to utter four isolated Malay words and their voices are recorded. The silence signal for each of the word is removed using the developed program. Next, the signals are analyzed based on selected voice quality features to study the differences in voice quality characteristics between men and women. The results revealed that the developed program can be used effectively for removing silence and there exist differences in voice quality characteristics between men and women.

Index Terms— Digital signal processing; Isolated Malay words; Fundamental frequency; Voice analysis, Windowing and framing

DOI: http://dx.doi.org/10.31838/jcr.07.08.273

INTRODUCTION

Voice is the most natural communication tools used by human [1]. Lungs, larynx, pharynx, nose and various parts of the mouth are involved in producing speech [2].

Realizing the importance of analyzing and understanding human uttered speech, many research have been conducted to produce various application on speech. Basically, speech can be classified into voiced and unvoiced [3,4] parts. According to Gaikward et. al, there are four different classes of speech which are based on the type of utterances: isolated, connected, continuous and spontaneous word [5].

The voice features very much depend on the pace or speed, volume, pitch level and voice quality while the articulation rate and speech pauses depend on the speaking style of the speaker [6].

Since the focus of the research is on gender, it is important to understand the gender-specific articulatory-acoustic relations between male and female. Figure 1 shows the mid-sagittal articulator positions for both genders [7].

Vocal folds vibrate when the air is under pressure from the lungs. Pitch is controlled by the tension in the vocal muscles. The range of frequencies for the normal speaking human voice is 70-200 Hz for males and 140-400 Hz for females [8][9]. Thus, men have denser and longer vocal folds. Adult refers to those aged 20-64 years old.

Isolated Malay words have been chosen in this research as Malay language is widely used in Malaysia, Singapore, Indonesia, Brunei and southern Thailand but with different dialects and accents. The reason for selecting these words is to later help those on smart wheelchair to utter simple words to help them move from one location to another.

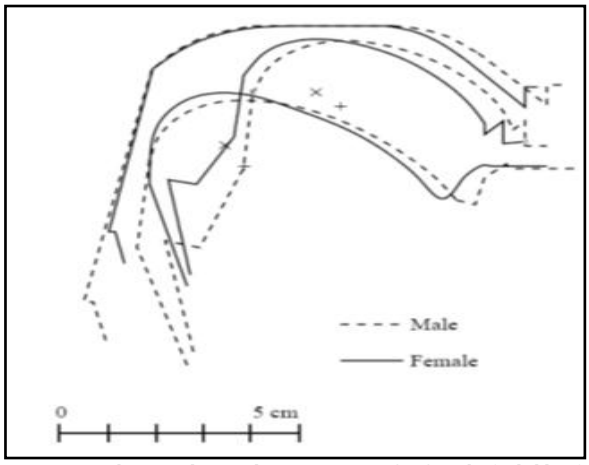


Figure 1. Mid-sagittal articulator positions for female (solid line) and male (dashed line)

We organized our paper in the following sequence: Section II describes our experiment setup and Section III reveals the result based on the analysis conducted in our research. Finally, Section IV draws the conclusion and our future works.

METHODS

Since the main objective of this paper is to remove silence for the isolated Malay words using framing and windowing method, we have therefore designed our program to cater for this situation.

The words uttered by the eight subjects (four females and four males) are recorded and saved in .m4a formats. Figure 2 shows the flow chart for the developed program.

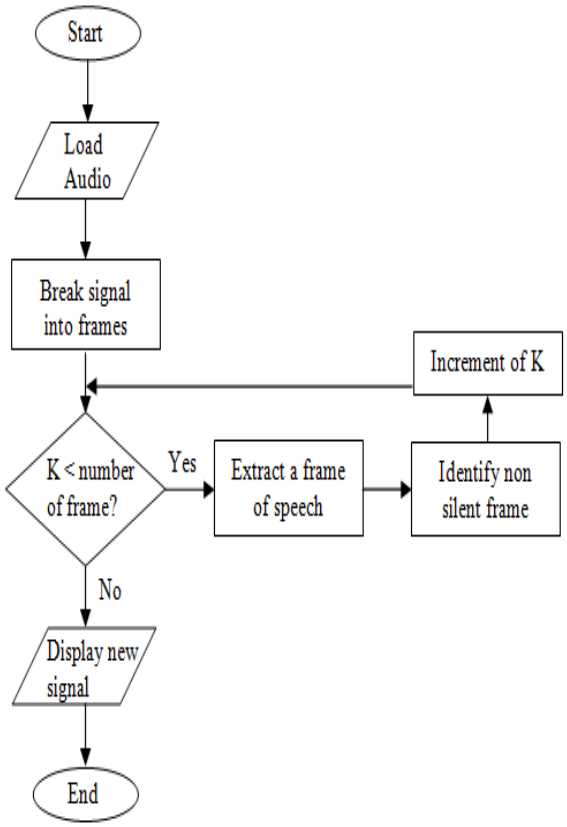


Figure 2. Flow chart for the silence removal of isolated Malay words using framing and windowing method

First, the audio is loaded and the original signal is plotted. To remove silences, the signal is first split into frames of 0.1s (depending on the sample). The speech signal is then divided into intervals within each frame without any overlapping. The timing used for splitting the frames is vital as it determines the frame length and also the number of frames. Through the use of looping, each frame is next extracted and the non-silent frames with maximum amplitude of less than 0.05 are identified. This task is repeated until it reached the end of the signal. Once no more frames can be extracted in the process, the new signal is then plotted without the silence signal. Initially, K is assigned as 1 to start the looping and its value will increase after each loop.

MATLAB R2016b is used for this research as it offers many advantages. In designing the graphical user interface, we used GUIDE in MATLAB since it offers quick layout and easy setup to get fast results in a more appropriate manner. Figure 3 shows the main interface of the program. There are three main buttons: Load File, Play and Reset. Each of these buttons has its own functions. Load button is used to choose the recorded audio as shown in Figure 4. To play the selected audio and plot the original signal and the new signal with silence removal, the Play button is chosen (refer to Figure 5). The Reset button is to clear the selected signal before choosing the other audio for testing purposes.

Besides removing silence from the original signal, the aim of this research is also to assess information on voice quality features of adults based on gender, thus PRAAT is used to help expedite the analysis. PRAAT is a suitable tool to do speech analysis and offers a wide range of standard and non-standard procedures, including spectrographic analysis, articulatory synthesis, and neural networks.

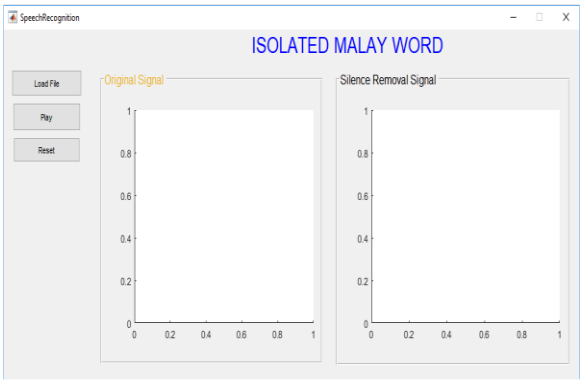


Figure 3. The main interface of the program

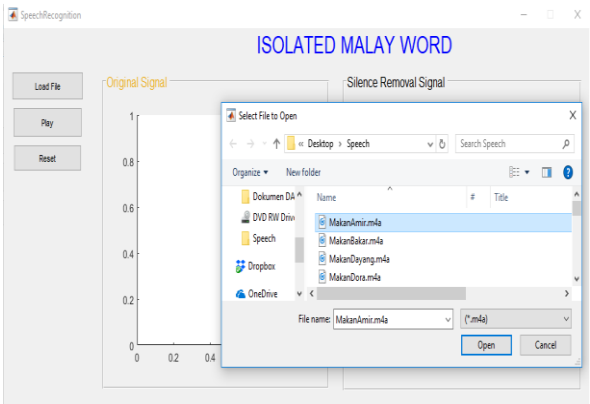


Figure 4. The audio is chosen to be loaded

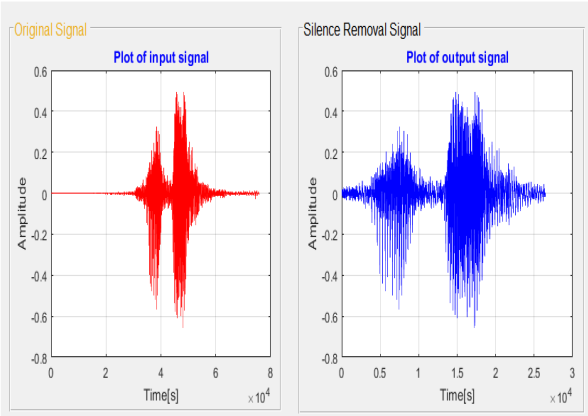


Figure 5. Examples of loaded audio with original and silence removal signal

In this research, eight adults of different gender (4 males and 4 females) have been asked to utter four words: *Makan* (eat), *Mandi* (bathe), *Tidur* (sleep) and *Wuduk* (make ablution), in order to study the influence of their voice quality features. Since PRAAT does not support *.m4a files, the earlier saved files are converted to *.wav files using an online audio converter.

The converted audio is then loaded and read using PRAAT. Figure 6 shows the options in PRAAT that can be manipulated by users. To view the chosen audio, View & Edit button is selected. We then proceed with the analysis of the signal based on the pulses, formants, intensity, pitch and spectrogram. Figure 7 and Figure 8 illustrate our analysis for the selected sound portion

based on the features mentioned earlier for male and female uttering the same word, *Makan*.

Figure 6. Options in PRAAT

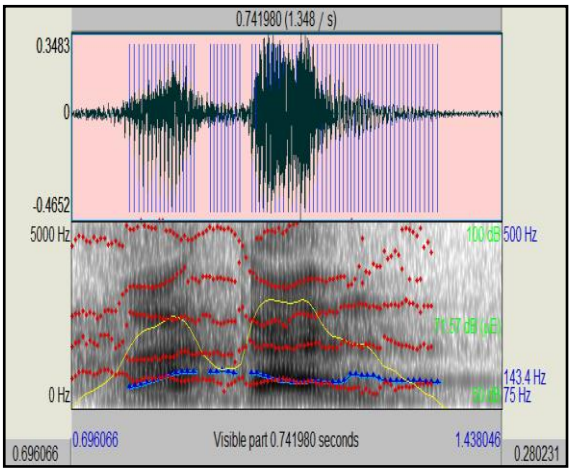


Figure 7. Sample of voice features for male

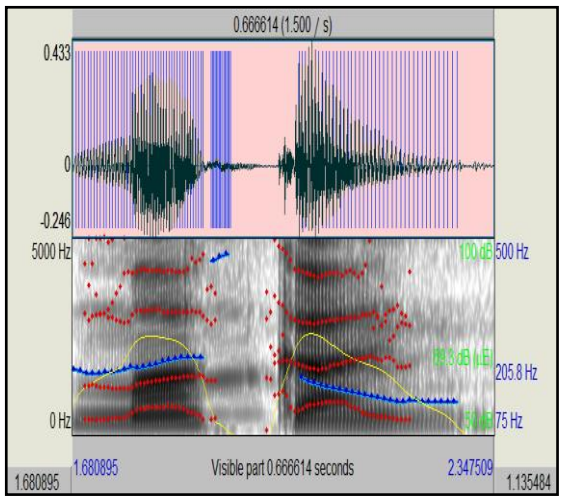


Figure 8. Sample of voice features for female

The pitch signal is shown in bright blue solid line and the value at the cursor position is indicated on the right hand outside of the panel in a dark blue font. The intensity signal is shown in yellow solid line and the value at the cursor position is indicated on the right hand side of the panel in a bright green font. The formants are shown in red dotted lines. Pulses are indicated in the top panel with vertical blue solid lines.

RESULT AND ANALYSIS

The demographic profiles of the eight respondents are shown in Table 1. The reason of choosing this age group is to observe variance in their pitch and intensity.

Table 1. Respondents profiles

Gender	Respondent	Age	No
Male	Speaker 2	20 - 30	1
	Speaker 3	31 - 40	1
	Speaker 4	41 - 50	1
	Speaker 1	51 - 60	1
Female	Speaker 2	20 - 30	1
	Speaker 4	31 - 40	1
	Speaker 1, 3	41 - 50	2
	None	51 - 60	0

In evaluating the effectiveness of the developed system in removing silence from the original signal, we plotted the sample results for one male and one female respondent that uttered the four isolated Malay words as shown in Table 2 and Table 3. The results show that the developed program has successfully eliminated the silence signal. In fact, the program can even produce audio for the original signal and also the new signal without the silent signal to let the user hear the audio differences before and after the elimination of the silence signal.

Table 2. Results for a male respondent who uttered four isolated Malay words

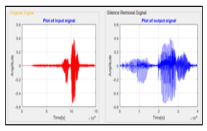
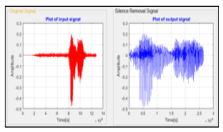
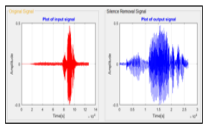
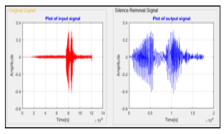
Word	Signal	Word	Signal
Makan		*Mandi*	
Tidur		*Wuduk*	

Table 3. Results for a female respondent uttered four Malay isolated words

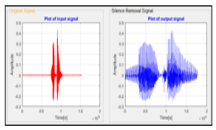
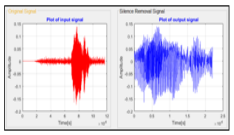
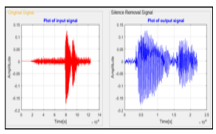
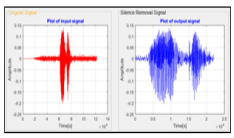
Word	Signal	Word	Signal
Makan		*Mandi*	
Tidur		*Wuduk*	

Table 4 shows the mean of the pitch and intensity for each male speaker who uttered the four words. It can be seen that there are variations in the results which may be attributed to their age difference. However, their pitch still falls within the ranges of 70 – 200Hz.

Table 4. Male mean pitch and intensity for male adults

Gender	Word	Respondent	Mean pitch (Hz)	Mean intensity (dB)
Male	Makan	Speaker 1	143.416	71.568
		Speaker 2	129.301	69.696
		Speaker 3	101.788	60.759
		Speaker 4	142.509	70.524
	Mandi	Speaker 1	143.642	73.306
		Speaker 2	145.177	68.754
		Speaker 3	96.870	62.054
		Speaker 4	177.698	70.854
	Tidur	Speaker 1	148.886	74.094
		Speaker 2	161.738	72.679
		Speaker 3	100.122	61.247
		Speaker 4	184.018	71.760
	Wuduk	Speaker 1	137.285	74.095
		Speaker 2	213.121	70.376
		Speaker 3	106.807	62.288
		Speaker 4	149.796	73.086

The mean pitch and intensity for each female speaker who uttered the four words are shown in Table 5. There are also variances in the results but their pitch is still within the range of 140 – 400Hz. The average pitch and intensity based on the results in Tables 4 and 5 is presented in Table 6. It can be seen from the table that male has low pitch but high intensity while female has high pitch but low intensity.

Table 5. Mean pitch and intensity for female adults

Gender	Word	Respondent	Mean pitch (Hz)	Mean intensity (dB)
Female	Makan	Speaker 1	200.851	69.376
		Speaker 2	192.432	59.268
		Speaker 3	199.255	67.038
		Speaker 4	192.917	67.816
	Mandi	Speaker 1	186.288	65.102
		Speaker 2	116.525	59.459
		Speaker 3	217.756	68.382
		Speaker 4	203.749	68.706
	Tidur	Speaker 1	188.470	64.312
		Speaker 2	198.154	62.730
		Speaker 3	215.103	69.823
		Speaker 4	199.180	71.018
	Wuduk	Speaker 1	191.001	66.151
		Speaker 2	177.370	59.816
		Speaker 3	229.668	70.761
		Speaker 4	223.961	70.771

Table 6. Average pitch and intensity of adults by gender

Gender	Word	Average Pitch (Hz)	Average Intensity (dB)
Male	Makan	129.253	68.137
	Mandi	140.847	68.742
	Tidur	148.691	69.945
	Wuduk	151.752	69.961
Female	Makan	196.364	65.874
	Mandi	181.079	65.412
	Tidur	200.227	66.971
	Wuduk	205.500	66.875

CONCLUSION

We have developed a program for removing silence in the original signal using framing and windowing method based on four words uttered by eight subjects for testing the effectiveness of the program. The idea is to remove the unwanted or silence signal before feeding the speech signal to feature extraction. We also analyzed various parameters for four isolated Malay words in the speech of male and female adults to see the difference in terms of both pitch and intensity.

We found that male has low pitch and high intensity and vice versa, for female. This might be related to the phonation frequency and intensity between the two gender groups. In our future research, we will focus on the extraction of the features for speech signal and represent them using an appropriate data model of the input signal.

ACKNOWLEDGMENT

The authors would like to thank the University Tun Hussein Onn Malaysia for funding this research under Tier-1 grant Vote U926.

REFERENCES

1. J. Keshet, and S. Bengio, *Automatic Speech & Speaker Recognition: Large Margin and Kernel Methods*, John Wiley & Sons, Ltd, Publication, 2009.
2. J. Holmes, and W. Holmes, *Speech Synthesis & Recognition*, 2nd Edition, Taylor & Francis, 2002.
3. M. Wolfel, and J. McDonough, *Distant Speech Recognition*, John Wiley & Sons, Ltd, Publication, 2011.
4. R. Bachu, S. Kopparthi, B. Adapa, and B. Barkana, *Voiced/Unvoiced Decision for Speech Signals Based on Zero-Crossing Rate and Energy*, Advanced Techniques in Computing Sciences and Software Engineering, Springer, Dordrecht, 2010.
5. S.K. Gaikwad, B.W. Gawali, and P. Yannawar, *A Review on Speech Recognition Technique*, International Journal of Computer Applications, Volume 10 – No 3, 2010.
6. M. Christian, *Speaker Classification 1: Fundamental, Features and Methods*. New York: Springer, 2007.
7. A. P Simpson, Gender-specific articulatory-acoustic relations in vowel sequences. *Journal of Phonetics*, 30(3):417-435, 2002.
8. *The Human Voice*. Retrieved from https://www.colorado.edu/physics/phys1240/phys1240_sm09/index_files/L17.pdf
9. Hussain, A., Mkpojiogu, E.O.C., Nawi, M.N.M. *Capturing customer satisfaction and dissatisfaction in software requirements elicitation for features in proposed software systems.* Journal of Engineering and Applied Sciences, 12 (21), pp. 5590-5597. 2017.