ISSN 2347 - 3983
Volume 8. No. 7, July 2020
International Journal of Emerging Trends in Engineering Research
Available Online at http://www.warse.org/IJETER/static/pdf/file/ijeter16872020.pdf
https://doi.org/10.30534/ijeter/2020/16872020

Audio-based Assessment in Determining Language

Aaron Don M. Africa, Ray Vincent Alin B. Lamdagan, Juan Miguel C. Lacanilao
Department of Electronics and Communications Engineering
De La Salle University, Manila
2401 Taft Ave., Malate, Manila 1004,
Philippines, aaron.africa@dlsu.edu.ph

ABSTRACT

The most apparent way humans communicate is verbal with the use of Language. With language being so complex, the average person would not be able to recognize another language besides their native own. But one thing is universal for all languages spoken throughout the world and it is that they produce sound. The average human being can hear sounds with frequencies ranging from 20 Hz to 20,000 Hz and with an intensity above the standard threshold of hearing. Hence, speaking produces sound with frequencies in between the said range. This paper aims to determine the language spoken of a given voice signal. By analyzing the given voice signal, its frequency range can be determined. Languages are spoken differently therefore have different frequency ranges. By using a filter to remove any ambient sounds from the input voice signal, its equivalent spectrogram reading gives an accurate range of its frequencies.

Key words: Language, Audible sound, Frequency, Filter, Magnitude Spectrum Plot, Spectrogram

1. INTRODUCTION

Communication is deeply intertwined with human existence. It allows human beings to express their thoughts and feelings with each other. While there are many forms of communication, it can generally be narrowed down into two forms: verbal and non-verbal. Verbal communication makes use of words either spoken or written to communicate. These words come in many different languages that are native to several different countries. Non-verbal communication however doesn't make use of words. Some examples of non-verbal communication according to [1] would be Kinesics or body movements including facial expressions and eye contact; Vocalists or adjusting the rate, volume, and pitch of voice; personal appearance or how one expresses himself visually; haptics or any form of interaction involving the sense of touch. There are numerous more forms of non-verbal communication since it is not bound to the use of words [2,3,4].

Verbal communication in the form of speaking makes use of sounds that we hear and then interpret its message or meaning. The sound that an average human being can hear produces frequencies that range from 20 Hz to 20 kHz [5,6,7]. Definitely, for a person to hear the sound, that sound must also be loud enough. The sound must also have an intensity higher than the threshold of hearing which is 0 dB or decibels. When human beings communicate by speaking to each other, they produce sound waves with frequencies in between the previously said range. Compared to non-verbal communication, verbal communication would be the more apparent choice for human beings to interact with each other [8,9,10,11]. However, verbal communication is not as simple as it seems. Language is one of the most complex ways of communicating for humans. Each country has its language. And most of that country's inhabitants can only understand or speak that specific language. This makes it hard for the average person to communicate with another person from a different country.

2. BACKGROUND OF THE STUDY

Language is the most apparent form of communication used around the globe today. It is a structured and conventional system that people follow to share information. Because of this, language can be generalized to the system being used by a certain country or community. An example would be English which is one of the most commonly known languages globally. The number of languages being used all across the globe can be estimated to vary between 5000 and 7000 [12,13]. The reason for such a large amount is the number of different dialects present in just one language. Another reason for the numerous amounts of languages is because languages evolve and diversify over time. Some languages were based on other languages. Others were changed or simplified over time. Because language such an important form of communication, some people study other languages to be able to speak and write in them other than their own [14,15,16].

Frequency is the number of times a certain event repeats per unit of time. This parameter is very important in the field of science and engineering for it is used to specify the rate of oscillations produced in signals. This study will be dealing with an audio frequency which is generally sound frequency but bounded only to a frequency range an average human being will be able to hear. The standard hearing range for humans is 20 Hz to 20 kHz [17]. This range can be divided into seven different frequency bands: 20 Hz to 60 Hz range is called Sub-bass which produces a sound so low that it is felt more than it is heard. 60 Hz to 250 Hz range is called Bass which determines how fat or thin the sound is. 250 Hz to 500 Hz range is called Low midrange which contains the low

order harmonics of most musical instruments. 500 Hz to 2 kHz range is called Midrange which is the frequency range for intelligible human speech. 2 kHz to 4kHz range is called Upper midrange which makes human hearing sensitive and can cause listening fatigue. 4 kHz to 6kHz range is called Presence which handles how clear and defined a certain sound is [18,19,20]. Lastly, 6 kHz to 20kHz range is called Brilliance which is the sound of bells and cymbals when they are ringing.

3. STATEMENT OF THE PROBLEM

Language is so complex, presents such a challenge even to know just one alternative or second language next to your native language. The average person cannot different languages if heard only for the first time. Many only hear 1-2 languages for the rest of their life. The group aims to provide a solution to this problem by presenting an effective way to determine the language of a given voice input [21,22]. With the frequency ranges of the popular languages known, it would be easy to determine the language just by comparing the input voice signal's frequency to the known frequency range values [23].

4. SIGNIFICANCE OF THE STUDY

The theory of creating a device that functions is to take a given voice signal and with that be able to determine the language spoken has many potential uses. Because language is used all around the world, the potential prototypes that support this theory's function would be used globally as well [24]. Examples of some of the uses of this device are as follows: One would be in telecommunications, calling hotlines and support when not in your respective country would almost be impossible. But with this device's function, the caller would be redirected to an agent that speaks their language and thus be serviced [25]. The caller would otherwise be fed with a recording speaking their language if no agent is available. Another use would be calibrating user interfaces to best support the language presented by these interfaces in electronic devices. With electronic devices being used largely around the world, this function would be able to help its user especially those who are not used to using these devices particularly the elderly.

5. THEORETICAL CONSIDERATIONS

Spoken language produces sound. For that reason, the sound coming from speaking has an audio frequency under the range of audible sound. The group is interested in whether different languages also have different frequency ranges. Knowing this information will be crucial in designing systems being able to determine a language just by feeding them an input voice signal via signal processing [26].

Started designing this system would have to of course be with the voice input signal. To analyze and determine the frequency of the input accurately, the input must not have any artifacts or unwanted noise. This noise may come from ambient sounds or even just faulty voice input devices. To remove the noise, the voice signal must be manipulated so that the noise frequencies may be eliminated. This can be done by filtering the voice signal before it is examined for its frequency range [27].

6. DESCRIPTION OF THE SYSTEM

The system will first consist of a device that can receive the spoken voice to be analyzed. This will be any microphone or audio input device. To maintain clarity of the voice signal, a filter will be used to rid the signal of any noise or artifacts. The clear voice signal will then be analyzed for its frequency range and then be compared [28,29].

Since the proposal is a theory, the group will be making use of known frequency range values of different languages to be compared to the input voice signal. These frequency ranges vary from country to country. Figures 1 and 2 show the range and table of different languages.

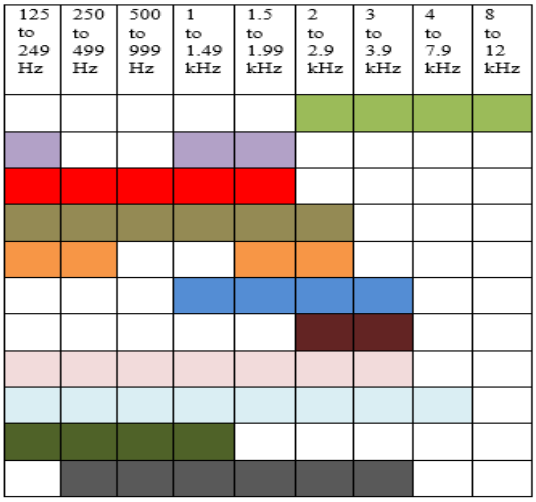

Figure 1: Frequency Range Table of Different Languages

Figure 2: Legend of Frequency Table in Figure 1

7. METHODOLOGY

Figure 3 shows the flowchart of the proposed system.

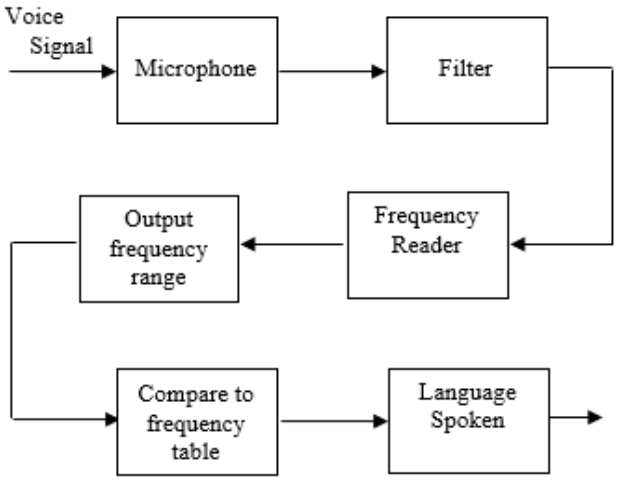

Figure 3: Flowchart of the Proposed System

The system functions by following the workflow presented in Figure 3. A person speaks through a microphone which takes in the input voice signal. This voice signal is then filtered to remove any unwanted sounds like ambient sounds or generally noise. After filtering, the signal is now analyzed and processed to find its frequency range. This can be done through MATLAB functions than will generate frequency/time plots or even spectrograms to accurately determine the frequency range of the input signal [30].

8. DATA AND RESULTS

The input voice signal was taken from the Open Speech Repository in a .wav format. This file was then imported into MATLAB using the audio read () command [31]. Because of the presence of noise in the signal, its frequency range could not accurately be determined. The group made use of a high pass filter to get rid of the noise present in the signal. The reason for this is because the voice signal initially is high frequency dominant. Figure 4 shows the plot of the filtered voice signal.

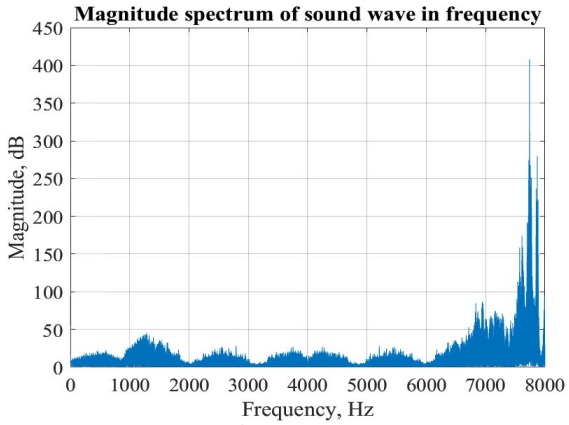

Figure 4: Plot of the Filtered Voice Signal

The plot shows that the voice signal has higher magnitudes at frequency levels 7000 to 8000. Meaning these are the frequencies of sound most present in the input voice signal. Therefore, it can be concluded that the frequency range of this voice signal is from 7000 to 8000.

9. ANALYSIS OF DATA

The sample voice file taken from the Open Speech Repository was first imported into MATLAB before being processed. The group was able to remove artifacts or any unwanted sound in the voice signal. Afterward, the group plotted the signal to find its frequency range. The plot showed large amounts of magnitude in the 7000 Hz to 8000 Hz region. This can be considered as the frequency range of the signal. Looking back to Figure 1 and Figure 2, wherein some of the frequency ranges of different languages were shown, we can see that the frequency range of the signal falls in between two languages being British English and Russian. Because the frequency range of the Russian language starts from such a low frequency of 125 Hz up to 8 kHz, the frequency range of our voice input signal is more relevant to the range of British English which is from 2 kHz up to 12 kHz. Upon checking the source of our voice signal, we confirmed that it was a recording of a British woman speaking English. Therefore, validating the frequency range of our signal to fall under British English. However, because this is just an assumption, there may be chances of error upon processing other voice signals.

10. CONCLUSION

The group was able to successfully follow its presented workflow in achieving the required output. Given a voice signal, the group was able to filter the voice signal using a high pass filter. The filtered voice signal was then plotted to show its Magnitude Spectrum wherein the group was able to find its frequency range. Comparing the frequency range of the voice signal to the previously presented frequency ranges of the different languages, the group was able to conclude that the voice signal comes from a person who speaks British English.

This study successfully showed the potential of analyzing frequencies of sound particularly spoken language. Because audio-assessment of speech communication was shown to be a way of determining a language. The future holds many possible ways of applying this theory. Because technology is rapidly evolving, there will be easier and more effective theories and applications to tackle a certain need. Voice analysis can come a long way even up to a point where it can stand on par with fingerprint recognition.

11. RECOMMENDATIONS

The group recommends the use of a pop filter to go with the microphone in order for popping sounds to be eliminated as this can affect the processing of the voice signal. The spoken voice should also be clear and understandable. It is

recommended that the speech signal is pronounced and annunciated properly to minimize any error that may occur. In using MATLAB, the group recommends the use of a spectrogram as an alternative way to plot the filtered signal to determine its frequency range. This is just another approach and will more or less arrive at the same output of the method tested in this study. The frequency range database of different languages is also recommended to be expanded so it will be able to cater to other languages as well. Increasing the number of possible languages to choose from will also increase the efficiency of the system.

REFERENCES

[1] J. K. Burgoon and D. B. Buller, "Interpersonal deception: III. Effects of deceit on perceived communication and nonverbal behavior dynamics," Journal of Nonverbal Behavior. Vol. 18, No. 2, pp. 155-184, 1994. https://doi.org/10.1007/BF02170076

[2] Y. Shi, H. Liu, Y. Wang, M. Cai and W. Xu, "Theory and Application of Audio-Based Assessment of Cough," Journal of Sensors. 2018.

[3] C. S. Seelamantula and T. V. Sreenivas, "Blocking artifacts in speech/audio: Dynamic auditory model-based characterization and optimal time-frequency smoothing," Signal Processing. Vol. 89, No. 4, pp. 523-531, 2009. https://doi.org/10.1016/j.sigpro.2008.10.014

[4] C. Cheng, A. Rashidi, M. A. Davenport and D. V. Anderson, "Activity analysis of construction equipment using audio signals and support vector machines," Automation in Construction. Vol. 81, pp. 240-253, 2017.

[5] H. Luo, M. Wang, P. K. Wong and J. C.P. Cheng, "Full body pose estimation of construction equipment using computer vision and deep learning techniques," Automation in Construction. Vol. 110, 2020.

[6] N. Bajaj, J. R. Carrión, F. Bellotti, R. Berta and A. De Gloria, "Automatic and tunable algorithm for EEG artifact removal using wavelet decomposition with applications in predictive modeling during auditory tasks," Biomedical Signal Processing and Control. Vol. 55, 2020. https://doi.org/10.1016/j.bspc.2019.101624

[7] A. K. Kumar and G. Saha, "Improved computerized cardiac auscultation by discarding artifact contaminated PCG signal sub-sequence," Biomedical Signal Processing and Control. Vol. 41,pp. 48-62, 2018.

[8] E. Verschueren, B. Somers and T. Francart, "Neural envelope tracking as a measure of speech understanding in cochlear implant users," Hearing Research. Vol. 373, pp. 23-31, 2019.

[9] T. Roy, T. Marwala and S. Chakraverty, "Precise detection of speech endpoints dynamically: A wavelet convolution-based approach," Communications in Nonlinear Science and Numerical Simulation. Vol. 67, pp. 162-175, 2019. https://doi.org/10.1016/j.cnsns.2018.07.008

[10] A. Borowicz, "Using a multichannel Wiener filter to remove eye-blink artifacts from EEG data," Biomedical Signal Processing and Control. Vol. 45, pp. 246-255, 2018. https://doi.org/10.1016/j.bspc.2018.05.012

[11] C. Preibisch, P. Raab, K. Neumann, H. A. Euler, A. W. von Gudenberg, V. Gall, H. Lanfermann and F. Zanella, "Event-related fMRI for the suppression of speech-associated artifacts in stuttering," NeuroImage. Vol. 19, No. 3, pp. 1076-1084, 2003.

[12] Y. Xu, Y. Tong, S. Liu, H. M. Chow, N. Y. AbdulSabur, G. S. Mattay and A. R. Braun, "Denoising the speaking brain: Toward a robust technique for correcting artifact-contaminated fMRI data under severe motion," NeuroImage. Vol. 103, pp. 33-47, 2014.

[13] F. Lieb and H. Stark, "Audio inpainting: Evaluation of time-frequency representations and structured sparsity approaches," Signal Processing. Vol. 153, pp. 291-299, 2018.

[14] S. Masaya, "Audio signal separation through complex tensor factorization: Utilizing modulation frequency and phase information," Signal Processing. Vol. 142,pp. 137-148, 2018.

[15] A. Covic, C. Keitel, E. Porcu, E. Schröger and M. M. Müller, "Audio-visual synchrony and spatial attention enhance processing of dynamic visual stimulation independently and in parallel: A frequency-tagging study," NeuroImage. Vol. 161, pp. 32-42, 2017.

[16] M. Savari, A. W. A. Wahab and N. B. Anuar, "High-performance combination method of electric network frequency and phase for audio forgery detection in battery-powered devices," Forensic Science International. Vol. 266, pp. 427-439, 2016.

[17] K. Wang, H. Cao, C. Duan, J. Huang and F. Li, "Three-dimensional scalar controlled-source audio-frequency magnetotelluric inversion using tipper data," Journal of Applied Geophysics. Vol. 164, pp. 75-86, 2019.

[18] K. H. Lee, H. K. Lee, H. J. Byun, I. J. Cho, J. U. Bu and E. Yoon, "An audio frequency filter application of micromachined thermally-isolated diaphragm structures," Sensors and Actuators A: Physical. Vol. 89, Nos. 1–2, pp. 49-55, 2001.

[19] E. Erçelebi and L. Batakçı, "Audio watermarking scheme based on embedding strategy in low frequency components with a binary image," Digital Signal Processing. Vol. 19, No. 2, pp. 265-277, 2009.

[20] A. Potchinkov, "Digital signal processing methods of global nonparametric frequency domain audio testing," Signal Processing. Vol. 85, No. 6, pp. 1225-1254, 2005. https://doi.org/10.1016/j.sigpro.2004.12.007

[21] J. Monge-Álvarez, C. Hoyos-Barceló, K. Dahal and P. Casaseca-de-la-Higuera, "Audio-cough event detection based on moment theory," Applied Acoustics. Vol. 135, pp. 124-135, 2018.

[22] K. Hirai, M.Nukaga, H. Tabata, M. Enseki, H.Furuya, F. Niimura, K. Yamaguchi and H. Mochizuki, "Objective measurement of nocturnal cough in infants with acute bronchiolitis," Respiratory Investigation. Vol. 57, No. 6, pp. 605-610, 2019. https://doi.org/10.1016/j.resinv.2019.06.005

[23] X. Wang, X. Zhao, Y. He and K. Wang, "Cough sound analysis to assess air quality in commercial weaner barns," Computers and Electronics in Agriculture. Vol. 160, pp. 8-13, 2019.
[24] P. Klco, M. Kollarik and M. Tatar, "Novel computer algorithm for cough monitoring based on octonions," Respiratory Physiology & Neurobiology. Vol. 257, pp. 36-41, 2018. https://doi.org/10.1016/j.resp.2018.03.010
[25] H. Mohammadi, A. Samadani, C. Steele and T. Chau, "Automatic discrimination between cough and non-cough accelerometry signal artefacts," Biomedical Signal Processing and Control. Vol. 52, pp. 394-40, 2019.
[26] N. V. S. P. Kumar, J. K. R. Sastry and K. R. S. Rao, "Mining negative frequent regular itemsets from data streams," International Journal of Emerging Trends in Engineering Research.Vol. 7, No. 8, pp. 85-98, 2019. https://doi.org/10.30534/ijeter/2019/02782019
[27] H. Hwang and J. A. Shin, "Cumulative effects of syntactic experience in a between- and a within-language context: Evidence for implicit learning," Journal of Memory and Language. Vol. 109, 2019.
[28] A. Africa, A.Tabalan andM.Tan, "Speech emotion recognition using support vector machines," International Journal of Emerging Trends in Engineering Research. Vol.8, No. 4, pp. 1212-1216, 2020. https://doi.org/10.30534/ijeter/2020/43842020
[29] J. Schepens, R. van Hout and T. F. Jaeger, "Big data suggest strong constraints of linguistic similarity on adult language learning," Cognition. Vol. 194, 2020.
[30] A. Presbitero, "Foreign language skill, anxiety, cultural intelligence and individual task performance in global virtual teams: A cognitive perspective," Journal of International Management. 2019. https://doi.org/10.1016/j.intman.2019.100729
[31] Matlab. https://www.mathworks.com/products/matlab.html. 2020.