Southeast Europe Journal of Soft Computing
Available online: http://scjournal.ius.edu.ba

ISSN 2233 – 1859

Vol 9, No. 2, September 2020, pp. 1-5

Bosnian Vowels Analysis Using Formant Frequencies

[1*]Melisa Kunovac and [2]Ali Abd Almisreb

[1,2]Faculty of Engineering and Natural Sciences,
International University of Sarajevo International University of Sarajevo,
Hrasnicka Cesta 15, Ilidža 71210 Sarajevo,
Bosnia and Herzegovina
**Corresponding Author*: melisa.kunovac1@gmail.com

Article Info

Article history:

Article received on 1 February 2020

Received in revised form 20 February 2020

Keywords:

Bosnian language; Vowel recognition; Vowels; PRAAT; Formant frequencies; Formants

ABSTRACT: One way for analysis of vowels in any language is an analysis using formant frequency analysis. The Bosnian language has five vowels and those are a, e, i, o, u. The research was conducted in such a way that words with a minimum of two identical vowels per word were selected for each vowel. Several samples were then collected that recorded each of the words, and then those words were analyzed in PRAAT software. The total number of samples was 1050, twenty-one subjects were included, twelve females and nine males. Each of them recorded ten words for each of five vowels, therefore fifty words by each subject. The outcomes are based on related articles and dissertations, recognition, and analysis of vowels. Recognition was based on the statement, reading the literature, that each person has a narrow band of F4 formant values that should identify the person. And the analysis part was done by comparing formant values. Also, the work was based on gender differences for this analysis, as well as some other small observations, for example, the difference between native and other speakers of the Bosnian language.

1. INTRODUCTION

Formant frequency analysis, in some simple way, can be explained as a comparison between formant values of the analyzed word. Formant analysis also can be used for testing any vocal, not just vowels. But the best result was shown by analyzing vowels. The formant values are extracted and then, based on their value, compared to achieve some new perception or conclusion on how they behave pronounced by different people [1][2]. From a recognition situated view formants are striking parts in the range that can serve to recognize, for instance, unique vowel phonemes from one another. In acoustics, formants are characterized as tops in the range envelope of a discourse sound. Furthermore, a formant is the centralization of acoustic energy around a specific recurrence in the discourse wave. There are a few formants, each at an alternate recurrence, around one in each 1000Hz band. Or then again, to put it in another way, formants happen at generally 1000Hz stretches. Each formant relates to reverberation in the vocal parcel. With regards to the number of formants, it very well may be said that there are five of them. (F1, F2, F3, F4, F5) [3]. Each one of them is expressed in Hz. Normally for examination, just an initial three are utilized, however in this paper likewise F4 is utilized and it is a formant that ought to distinguish the speaker. formats frequencies have been used widely to various vowels from different languages such as [4]–[6].

For this paper, the Bosnian language was used. Bosnian language, in the opinion of people whose native language is not Bosnian, is very hard to understand and learn. The language is, by pronunciation, can be said to be similar to the languages of the countries in the region, but the grammar and spelling are much different. The language has seven cases and five vowels. There are three divisions of vowels. These are divisions that are made based on the position of the tongue when pronouncing them (first and second division) and based on the size of the opening of the oral cavity when pronouncing them (third division) [7].

2. BOSNIAN VOWELS

There are three divisions of vowels. These are divisions that are made based on the position of the tongue when pronouncing them (first and second division) and based on the size of the opening of the oral cavity when pronouncing them (third division). The first division is determined based on differences in the horizontal position of the tongue (for moving back and forth), while the second division starts from the differences in the vertical position of the tongue (its movement up and down) [7]. Within the first division, the vowels are grouped into three types:

- Front row vowels (*i, e*) - which are pronounced by moving the tongue forward, towards the front of the oral cavity and leaning on the lower teeth, rising towards the hard palate.
- Middle-row vocal (*a*) - during the pronunciation of which the tongue does not move forward or backward, but descends together with the lower jaw; (This vocal is included in the back row vocals in some sources).
- Back-row vowels (*o, u*) - in the pronunciation of which the tongue is pulled towards the back of the oral cavity.

According to the second division, vowels are also divided into three types:

- High vowels (*i, u*) - in the pronunciation of which the tongue approaches the upper jaw with its front and back parts;
- Middle vowels *(e, o*) - in the pronunciation of which the tongue occupies a middle position and is located somewhere in the middle between the upper and lower jaw;
- Low vowel *(a*) - in the pronunciation of which the tongue descends towards the lower jaw.

According to the third division, the vocals compete with each other in terms of openness, ie closedness. The most open is the vowel a, in whose pronunciation the angle between the upper and lower jaw is the largest, ie the mouth is the most open. The vowels o and e are medium open vowels. Closed vowels are u and i because when they are pronounced, the mouth is the most closed, that is, the distance between the jaws is the narrowest [7].

The classification of vowels can be seen in Figure 1

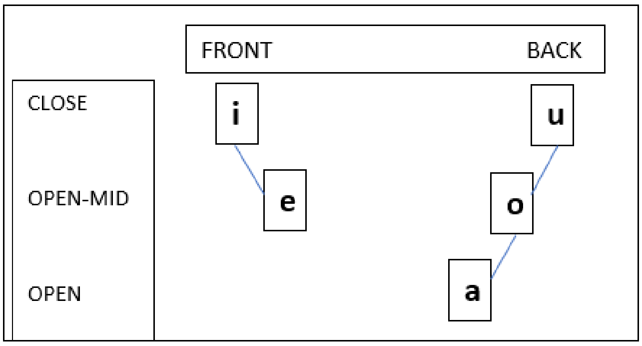

Figure 1 Classification of vowels

3. Dataset

The data analyzed through this paper are audio recordings. When creating the work plan, the goal was to gather ten people, both genders who would record their audio recordings of the pronunciation of the words given to them. However, to achieve better and more accurate results, but also the diversity of voices, ages, and accents of the Bosnian language (BL), the number was increased to twenty-one (21) subjects.). Twelve (12) subjects were female, and nine (9) were male. As mentioned earlier, the BL has five (5) vowels (a, e, i, o, u). Ten words in the BL were selected for each of the vowels, each of which has at least two identical vowels in it. This means that each word contains only one or five vowels and contains at least two vowels, for example, the word banana contains only the vowel a, and contains three times that vowel. Therefore, fifty (50) words for each sample. In total, the number of one thousand five hundred (1050) samples was analyzed. All audio recordings were recorded on a mobile phone. Table 1 shows the words per vowels.

Table 1 Words used per vowel

Words					
Numbering	**Vowel**				
	a	*e*	*i*	*o*	*u*
1.	Banana	Ceker	Ćilim	Kokos	Humus
2.	Flaša	Ekser	Isti	Kolo	Kukuruz
3.	Kanta	Element	Ići	Kopno	Kupus
4.	Krastavac	Fenjer	Pišti	Lopov	Kusur
5.	Krava	Melek	Piti	Odmor	Putuju
6.	Lama	Melem	Prišiti	Oko	Ubrus
7.	Lampa	Negdje	Siliti	Okolo	Unuk
8.	Tabla	Pelet	Šiti	Porok	Usput
9.	Trava	Pješke	Sitni	Potok	Uzduž
10.	Čarapa	Reljef	Štititi	Slovo	Zumbul

4. METHODS

The analysis was focused on the separation of two genders and observing F1, F2, F3, and F4. The average value for each formant and each vowel was calculated, and based on that graphs were plotted. This type of value was used for easier analysis of formants per each vowel. The plan was to show which formants have the higher and lower values and is it possible to recognize the gender-based formant values. The processing and analysis were done using PRAAT software [8]. Figure 2 shows samples of waveforms for each of the five vowels.

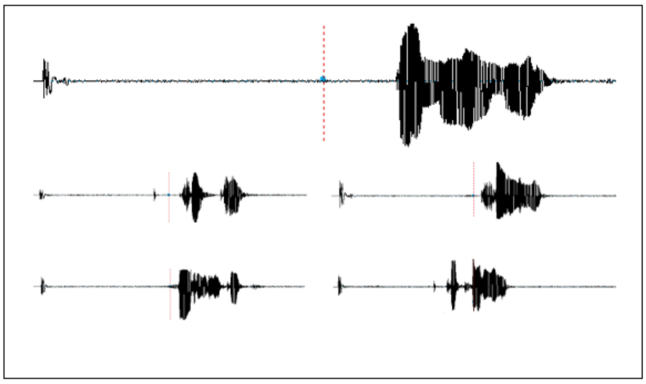

Figure 2 Example of waveform per each vowel respectively

5. RESULTS

From Figure 3 it can be seen that data is slightly unexpected. What does this mean? It means that by doing the research, almost all of the resources claimed that values will be higher for female samples rather than male samples. The unexpected was that for four vowels F1 for males was slightly higher than F1 for females. F3 was higher for females in vowels *e*, *i*, *o*. For vowel, *a*, and vowel *u* (both classified as 'back' vowels) all values from F1 to F4 were higher for males. The average results of female formants, especially F4, might have a small deviation. This is because for 21 recordings (2%) F4 could not be read. After all, the value is too high (over 5000Hz). Therefore, where the results could not be read, zero was counted so it might reduce the final value.

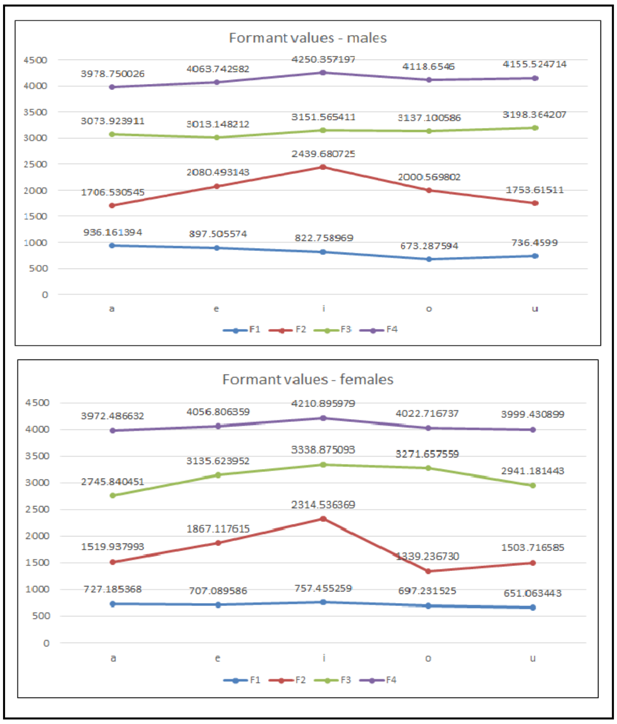

Figure 3: Formant values males/females

The analysis could be better seen in the graphs above. As stated, vowel (*i*) has the highest values for all formants in both genders. Also, the unexpected values for females were lower than for males (F4 especially) since the female voice, in general, has higher frequencies, but it might be due to mentioned software limitations.

When it comes to the difference between native speakers and those not living for a long time in Bosnia and Herzegovina. Things that can be seen from the comparison of these two are 1. Again, the vowel with the highest values of formants frequencies was i. 2. In general, the highest values were for speakers living in Bosnia and Herzegovina region rather than the Scandinavia region. This is the case for both female and male samples. 3. Small deviations were for vowels e and o, but this was expected since the Swedish language and in general, all languages spoken in the Scandinavia region have emphasized pronunciation of these vowels and also a couple of different variants of it. Person 1 is a native Bosnian speaker and Person 2 is a Bosnian speaker living in the Scandinavia region. Furthermore, the next graphs (Figures 4, 5 and 6) are showing these differences.

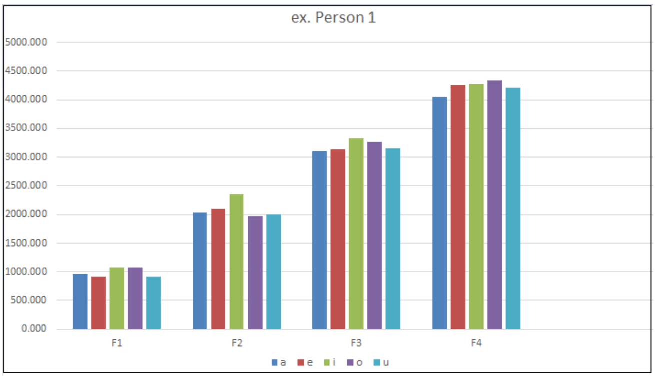

Figure 3 Vowel values per formant – Person 1

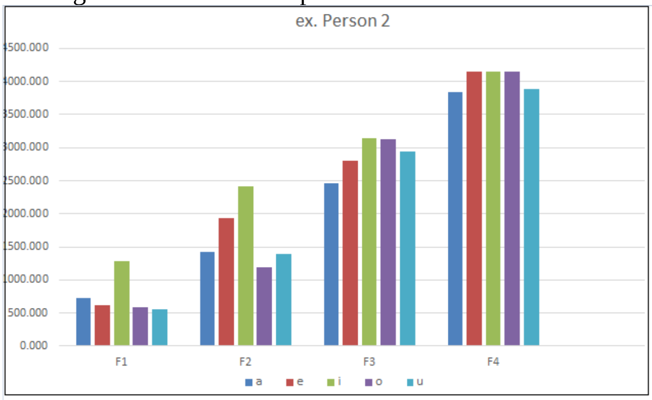

Figure 4 Vowel values per formant – Person 2

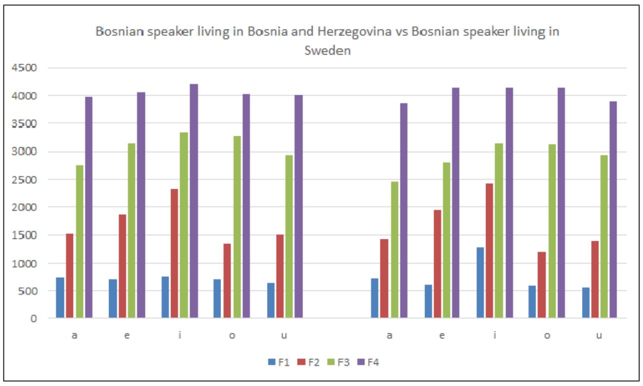

Figure 6 Bosnian speaker living in Bosnia and Herzegovina vs Bosnian speaker living in Sweden

6. CONCLUSION

What could be concluded from the analyzed data is that almost all of the resources claimed that values will be higher for female samples rather than male samples. The unexpected was that for four vowels F1 for males was slightly higher than F1 for females. F3 was higher for females in vowels *e*, *i*, *o*. For vowel, *a*, and vowel *u* (both classified as 'back' vowels) all values from F1 to F4 were higher for males. The average results of female formants, especially F4, might have a small deviation. This is because for some words F4 could not be read because the value is too high (over 5000Hz). So, where the results could not be read, zero was counted so it might reduce the final value. For future work, Research on this topic would be based on further analysis related to these patterns, such as reading a person pronouncing a particular word based on the F4 value alone. Also, the analysis would be performed based on other software that could read the values unavailable in the one used. The data set would expand to as many patterns as possible, and the words that would be used to record the audio would contain more than two of the same vowels in it. More samples from different spoken areas whose mother tongue is Bosnian would be collected and compared with samples from people living in Bosnia and Herzegovina. Also, the wider range of age of people recording voices would be collected. The whole process of recording would be placed in special conditions, so the noise could be reduced to a minimum. After doing all of this, the plan is to do analysis based on deep learning methods, so extracting the information about formant frequency values and then doing formant estimation and formant tracking.

7. REFERENCES

[1] D. J. Broad, "Formants in automatic speech recognition," *Int. J. Man. Mach. Stud.*, vol. 4, no. 4, pp. 411–424, 1972.

[2] H. Cao and V. Dellwo, "The role of the first five formants in three vowels of mandarin for forensic voice analysis," 2019, [Online]. Available: https://www.zora.uzh.ch/id/eprint/177494/1/ICPhS_666.pdf.

[3] S. Wood, "Praat for beginners [Manual]," *Available on-line from http//www. ling. lu. se/persons/Sidney/praate*, 2005.

[4] A. A. Almisreb, A. F. Abidin, and N. M. Tahir, "Acoustic analysis of Arabic phonemes based- /a:/ vowel," 2016, doi: 10.1109/CSPA.2016.7515845.

[5] A. A. Almisreb, N. Md Tahir, A. F. Abidin, and N. Md Din, "Acoustical comparison between /u/ and /u:/ Arabic vowels for non-native speakers," *Indones. J. Electr. Eng. Comput. Sci.*, 2018, doi: 10.11591/ijeecs.v11.i1.pp1-8.

[6] N. Jamil, I. Ramli, and N. Ardi, “Formant characteristics of Malay vowels of Perlis, Kelantan And Terengganu,” *J. Inf. Commun. Technol.*, vol. 18, no. 4, pp. 529–544, 2019.

[7] D. Jahić, S. Halilović, and I. Palić, *Gramatika bosanskoga jezika*. Dom štampe, 2000.

[8] D. Boersma, Paul & Weenink, “Praat,” *doing phonetics by computer [Computer program]*, 2020. http://www.praat.org/ (accessed Sep. 01, 2020).