Warsaw University of Technology

FACULTY OF
POWER AND AERONAUTICAL ENGINEERING

Project

in the subject of study of Signal Processing

Formant-Based Speech Analysis

Usman Ahmad
David Ibrahim

Project supervisor
Maciej Stefańczyk, M.Sc.

Warsaw, 2021

Contents

Introduction 3
Problem Description 3
Methodology 4
LPC Coefficients 4
Autoregressive Model 4
Program 4
Data Visualization 5
Formants Calculation 10
Comparison 11
[0] Female 11
[1] Female 12
[2] Male 13
[5] Female 14
Results 15
Conclusions 21
References 22

Introduction

Signal processing is an engineering subfield that focuses on analyzing, modifying, and synthesizing signals such as sound, images, and scientific measurements. Signal processing techniques are used to improve transmission, storage efficiency and subjective quality and to also emphasize or detect components or features of interest in a measured or stored signal. In the field of Signal Processing, the speech analysis makes one of the most important and widely researched topics. Speech analysis is a massive research aspect as it has many applications in various industries, as it can be used for security, to increase efficiency in the workplace, in the health care, or to create a faster experience of normal day services like banking. The speech analysis which leads to speech recognition is a very delicate and advanced subject, but in its core, it's simply signal processing. The audio waves detected by the computer are signal, and to be able to analyze those signals, they must be processed and filtered to make a useful aspect of the signal to be analyzed and used for any applications.

The basis of this project is speech analysis in the frequency domain. The formant-based speech analysis will be used to determine the similar features in the same word, spoken by a man and a woman. The main conflict in this project is how to increase those similarities, so that the computer can recognize the word as the same whether it is spoken by a man or woman.

What is a formant? A formant is the broad spectral maximum that results from an acoustic resonance of the human vocal tract. Formants are frequency peaks in the spectrum which have a high degree of energy. They are especially prominent in vowels.

Problem Description

This project will consist of five main parts:

1. Generating a spectrogram for each speech signal.
2. Determining and showing formants (local energy maxima).
   The formants are the basis of comparison and the core of this project. To calculate the formants, a few steps will be applied. Using then two different methods to calculate the formants: from the LPC function and the autoregressive model of the signal.
3. Comparing the two voices (of a man and a women) speaking the same word or sentence.
4. Finding out an approach to increase the similarity of both spectrograms.
   The most straightforward approach is to change the frequency of the segments using the formants frequency average of the man and the woman, stretching if the average is higher than the actual frequency and compressing if it is lower. That approach is very straightforward and should in theory work in the project. The real implication of it is yet to be tested and evaluated.
5. Making visualizations of the time-domain signal and its spectrogram, of intermediate and final processing results.

The spectrogram will be the main aspect of comparison. The spectrograms of the same word for the man and the woman will be shown side by side, for a valid comparison. The formants of the signals will be calculated. Using those formants and the spectrograms of the signals, the signals will be refined and analyzed to increase the similarities and making the word spoken by either the man or the woman be detectable as the same word easily.

Methodology

There are two main methods for computing the formants of the signals in this project: the LPC coefficients method and the autoregressive model.

LPC Coefficients

Linear predictive coding (LPC) is a method used mostly in audio signal processing and speech processing for representing the spectral envelope of a digital signal of speech in compressed form, using the information of a linear predictive model.

In this project the way it is implemented is through a few step to calculate the formants. The first step in this process is applying a few filters to the signal to make the information needed more accessible. The first filter is hamming filter, and the second filter is pre-emphasis filter. Then the LPC coefficients are calculated, the number of coefficients is determined with the use of the formula that they have to be 2 + the sampling frequency in kHZ. Because the LPC coefficients are real-valued, the roots occur in complex conjugate pairs, so we retain only the roots with one sign for the imaginary part and then determine the angles corresponding to those roots. The next step is to convert the angular frequencies in rad/sample represented by the angles, just calculated, to hertz and then calculate the bandwidths of the formants. The bandwidths of the formants are calculated by the distance of the prediction polynomial zeros from the unit circle. The next step is to filter the formants using the criterion that formant frequencies should be greater than 90 Hz with bandwidths less than 400 Hz to determine the proper formants. Then lastly, extracting the first three formants as they provide most of the information needed to adjust the frequency of the male and female speech signals.

Autoregressive Model

An autoregressive (AR) model predicts future behavior based on past behavior. It is used for forecasting or predicting when there is some correlation between the values in a time series and the values that precede and succeed them. Only past data is used to model the behavior, that is what gave the name “autoregressive”. The process is basically a linear regression of the data in the current series against one or more past values in the same series.

In this project the way it is implemented is through a few step to calculate the formants. The first step in this process is applying a hamming filter to the signal to make the information needed more accessible. Then the autoregressive model for the signal is computed, the order of the model is determined with the use of the formula that they have to be 2 + the sampling frequency in kHZ. The next step is about getting the transfer function of the voice tract and using the coefficients of the transfer function to compute the frequency response. Then, the peaks in the frequency response are extracted as they represent the locations of the formants. Using those locations the formants are extracted. Then lastly, extracting the first three formants as they provide most of the information needed to adjust the frequency of the male and female speech signals.

Program

The program is simple as it uses everything that was explained previously. The first step is loading the data in a loop to go through the signal of the same number by the male and the female. After loading the speech signals, the spectrograms are then used to visualize the properties of the signals without any changes occurring on it yet. The formants are then computed using the two methods explained in details above: the LPC method and the autoregressive method. Those two methods will be compared over the next part of the project with the manual calculation of formants to determine which is more accurate or which is better overall for this implementation. The next step in the program is to determine the rate of difference between the formants of the female and the male signals. This rate will then be used to increase and decrease the frequency of the male and female signals, respectively. The last step is to see the effect of that alteration of frequency by comparing between the formants of the adjusted signals, using the method deemed best by the comparison conducted, and visualizing their new spectrograms.

Data Visualization

The sample data given were sound signals of the a man and a woman counting from 0 to 9 in the Polish language. The best way to efficiently analyze the data is to represent the spectrograms of the signals, to be able to see the signals and their similarities easily. Below are the spectrograms for the 20 signals with each two of the same number grouped together.

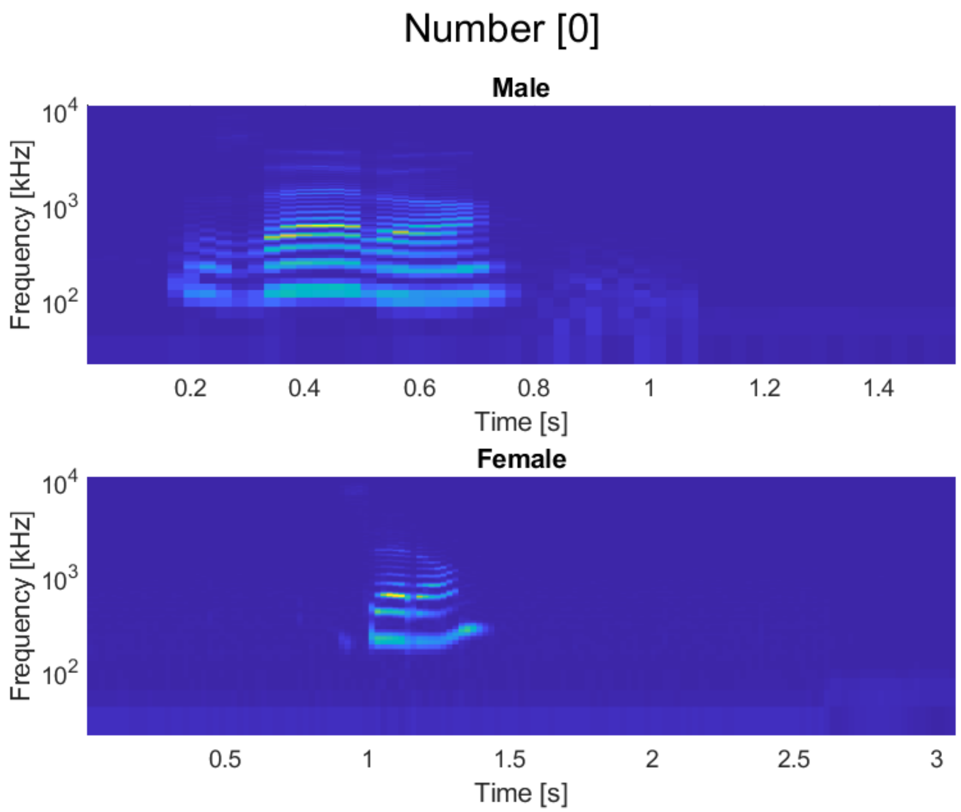

Figure 1 Number [0] (zero)

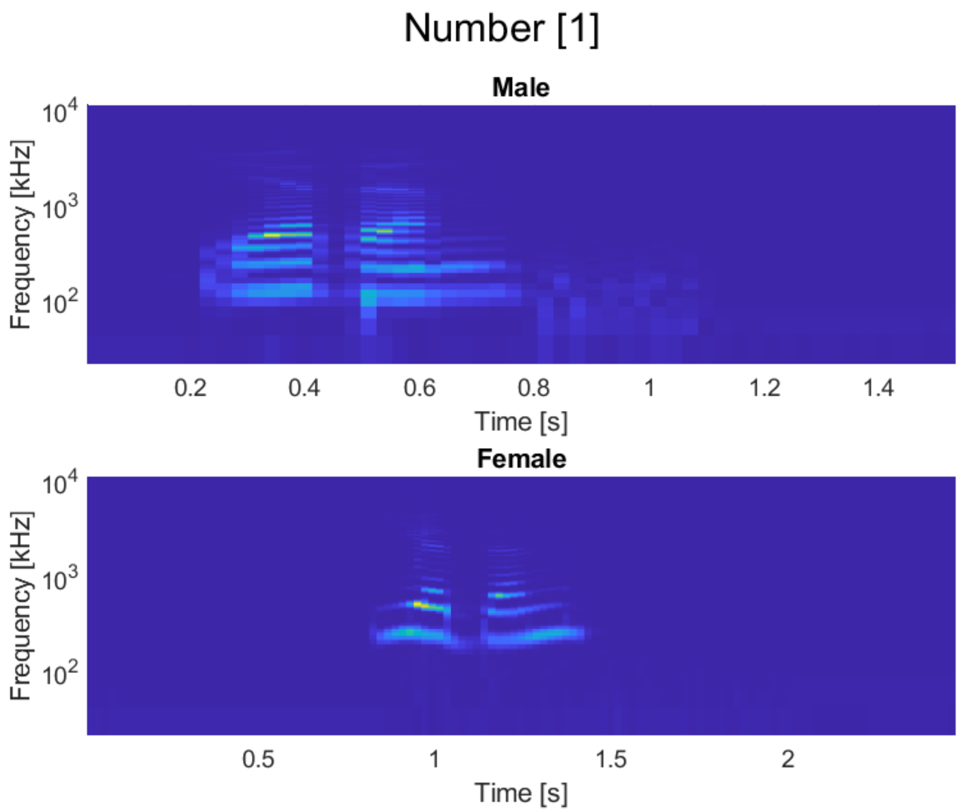

Figure 2 Number [1] (jeden)

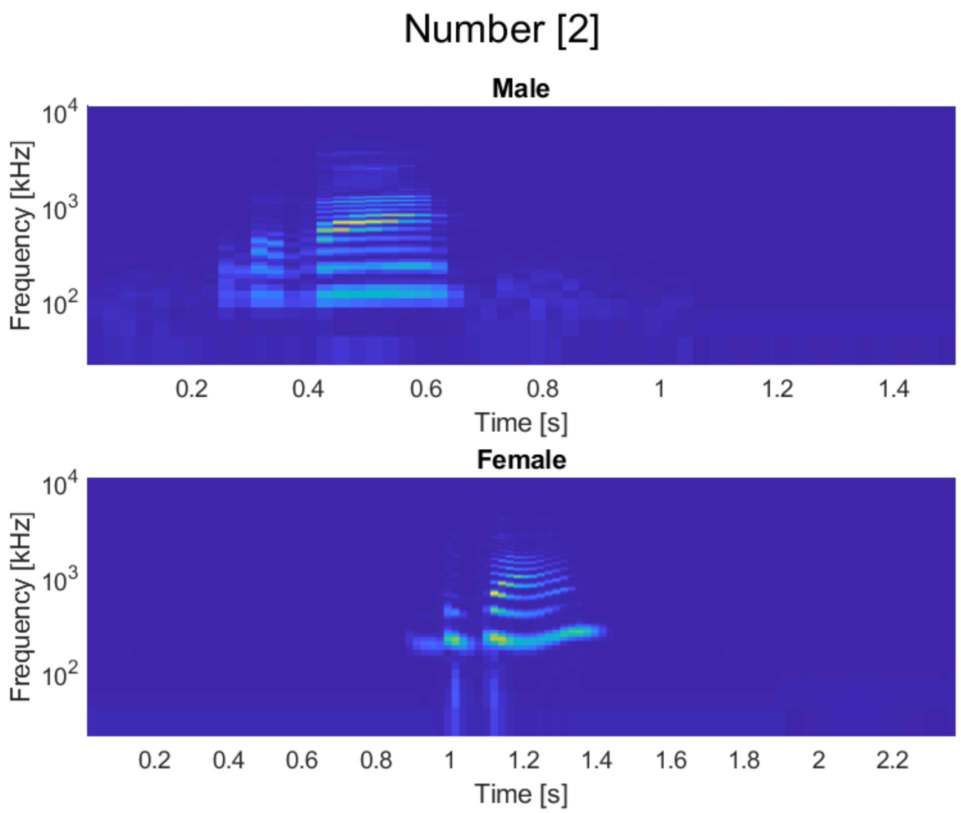

Figure 3 Number [2] (dwa)

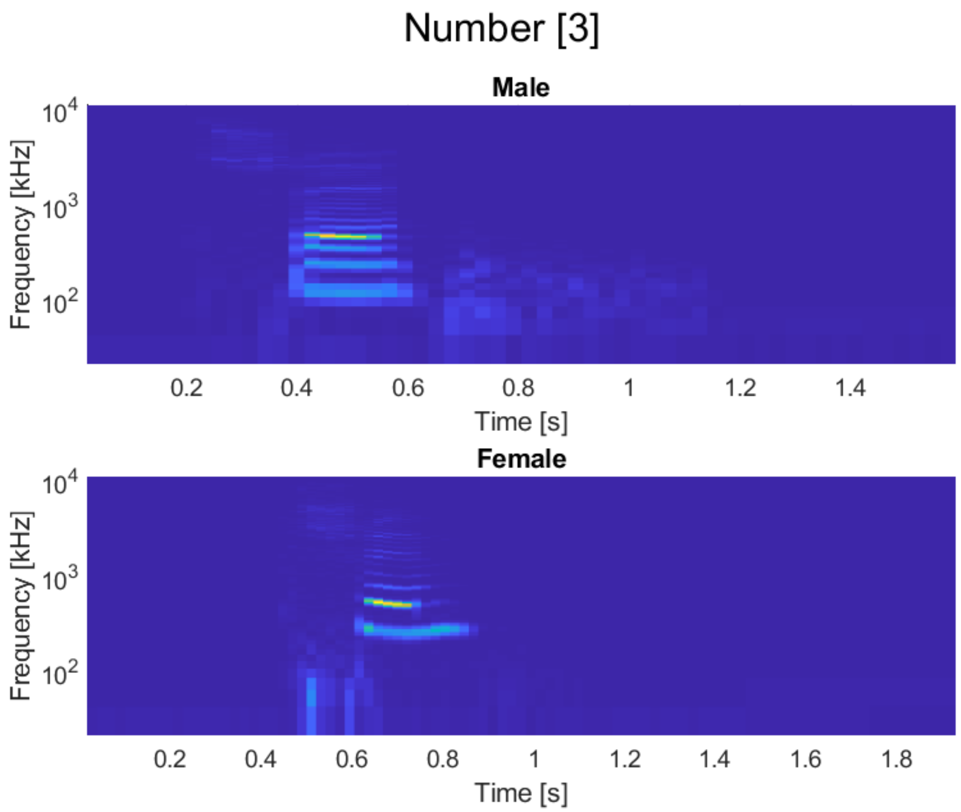

Figure 4 Number [3] (trzy)

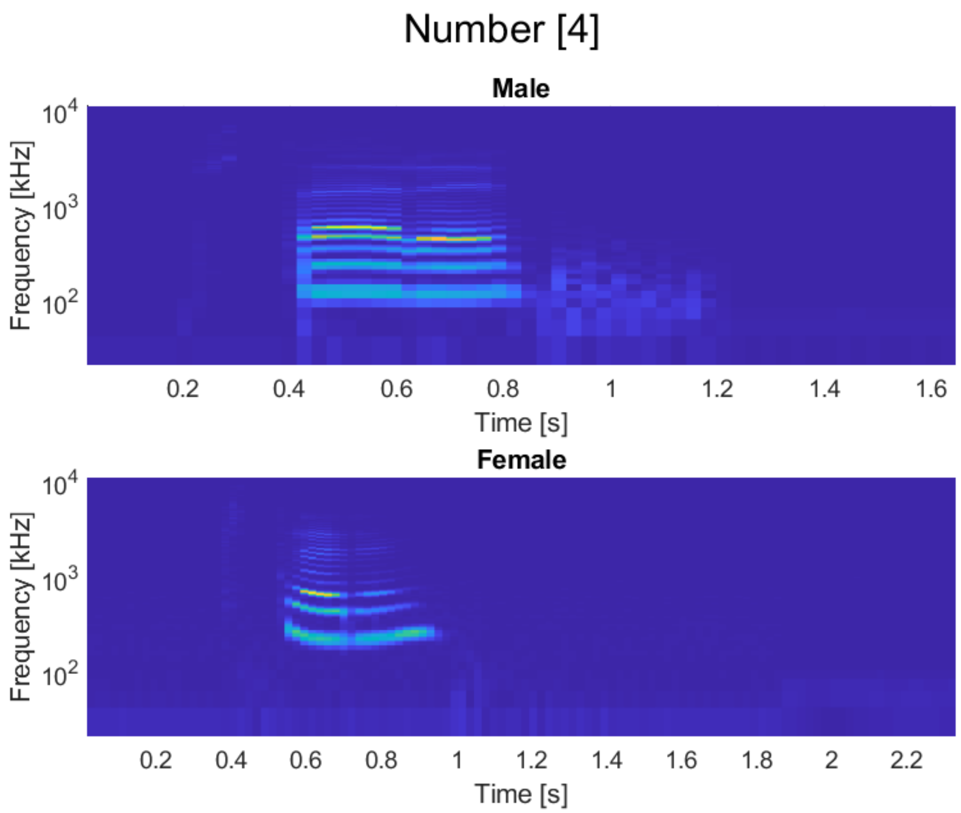

Figure 5 Number [4] (cztery)

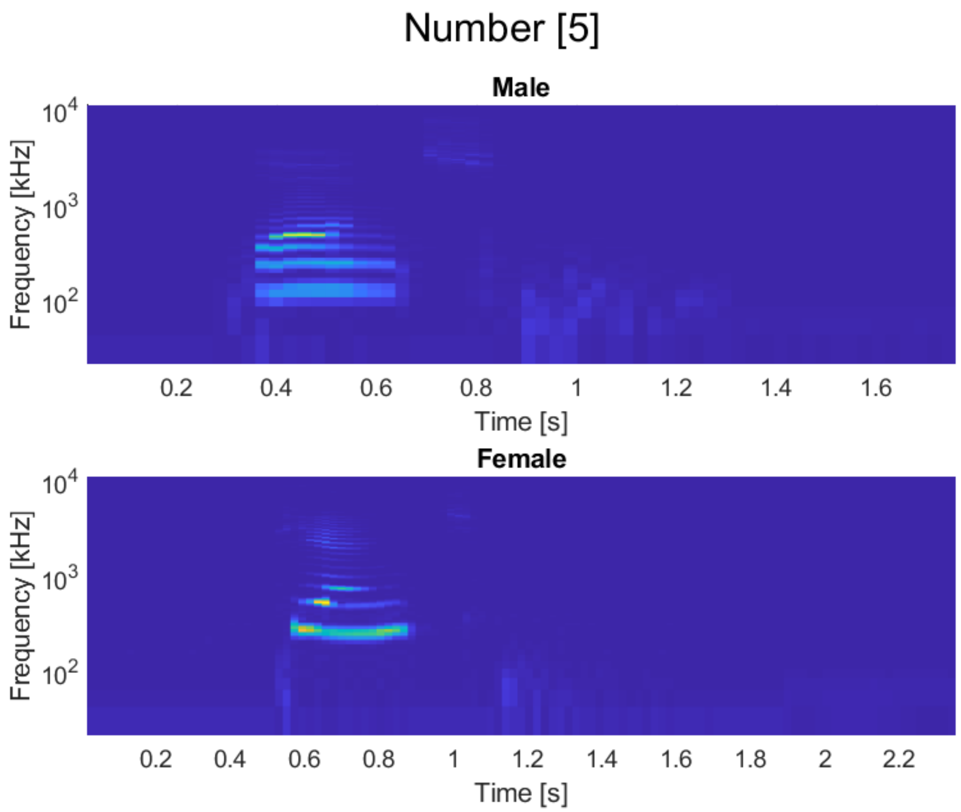

Figure 6 Number [5] (pięć)

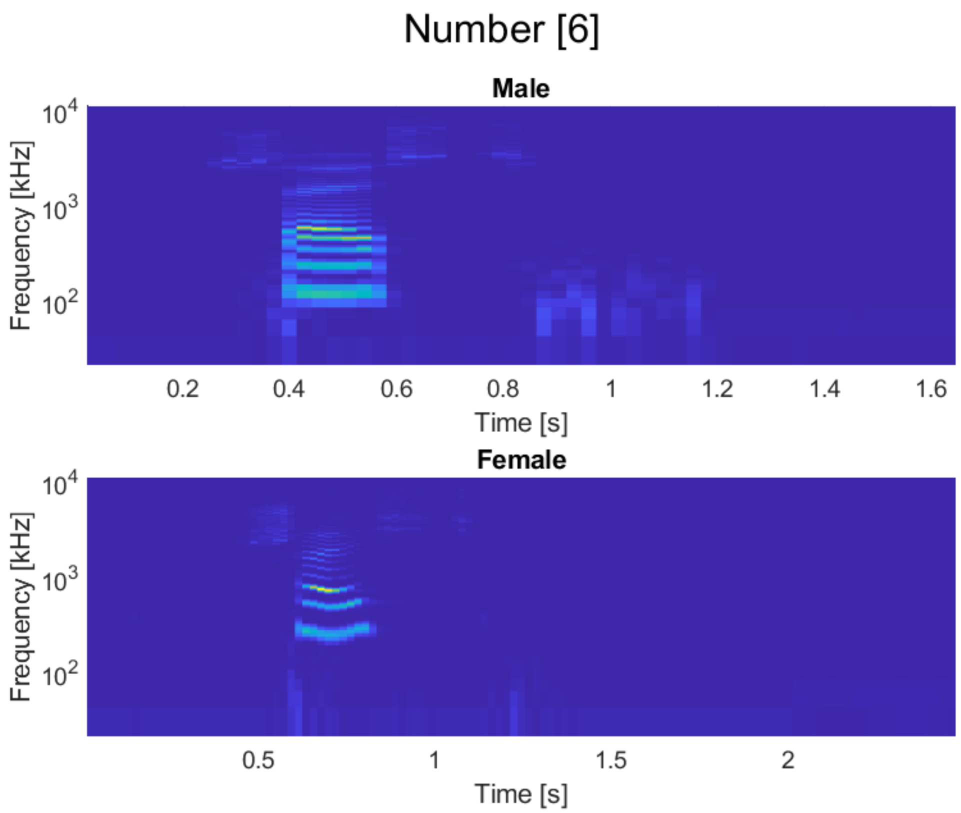

Figure 7 Number [6] (sześć)

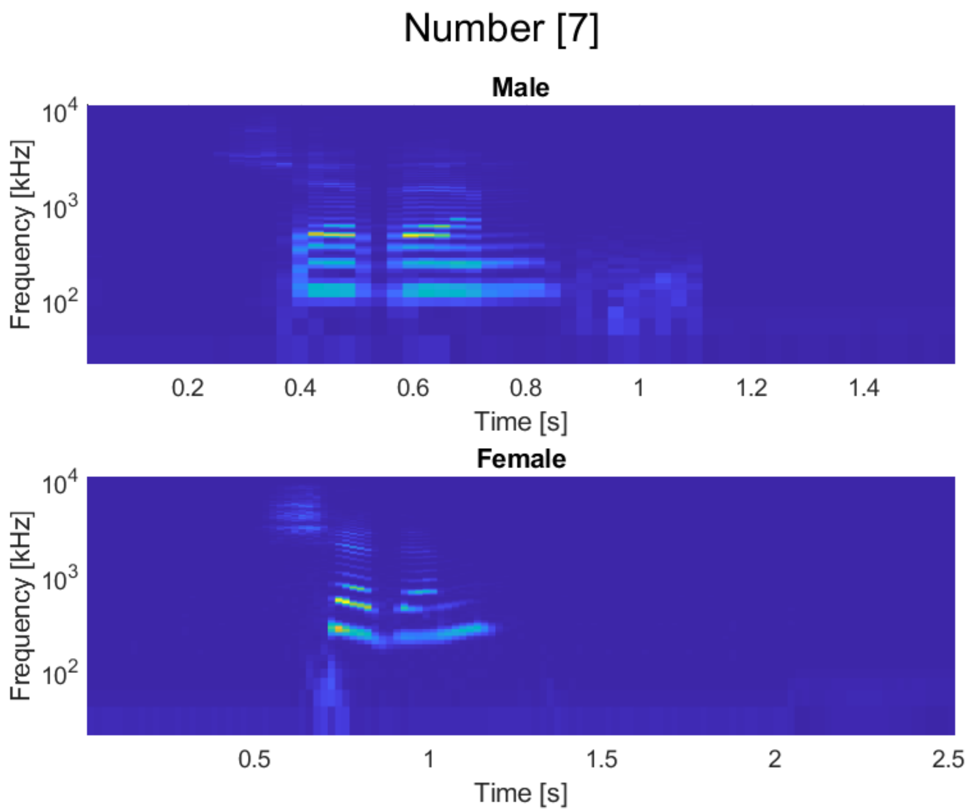

Figure 8 Number [7] (siedem)

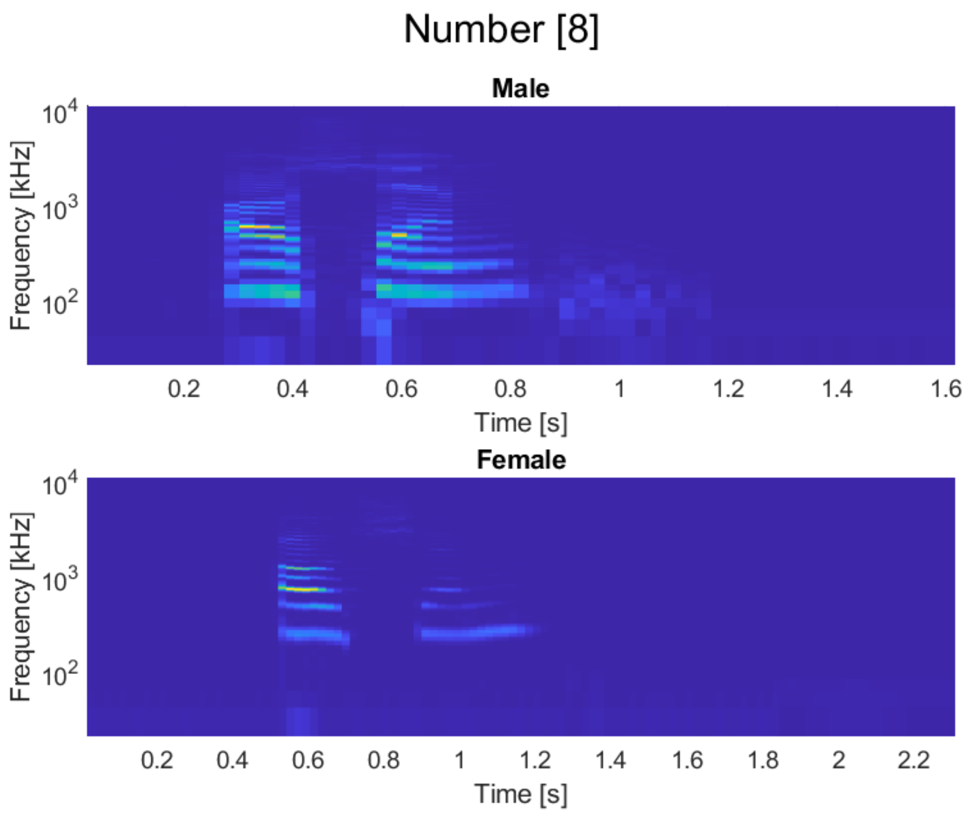

Figure 9 Number [8] (osiem)

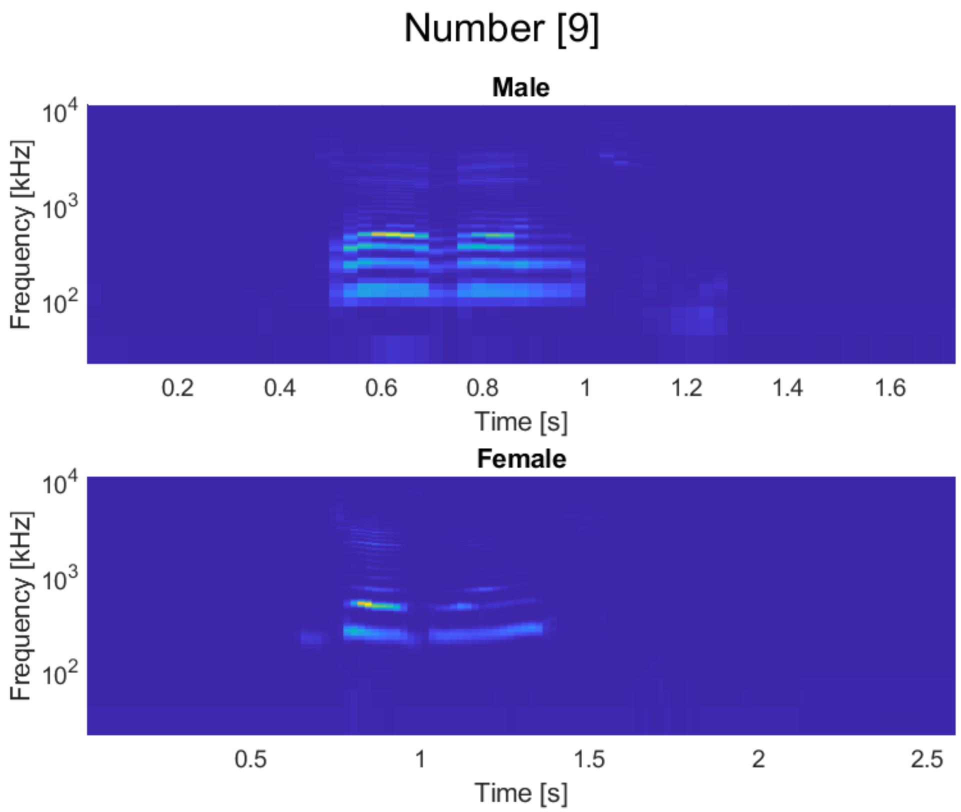

Figure 10 Number [9] (dziewięć)

Formants Calculation

In this section, the two methods for calculating the formants are going to be compared with the manual observation of the formants on the spectrogram. There are 20 different speech signals, so comparing the methods for all of them would be impractical. Instead, a few samples will be taken to establish a general assumption about the best method possible. Below is a table of all the formants values calculated from both the methods.

Table 1 Formants Values From LPC & AR Methods Before Any Adjustments

			F1	F2	F3
0	LPC	male	558.9768711	1226.214503	2421.246255
		female	431.0240333	703.2429245	1776.970856
	AR	male	538.3300781	1162.792969	2433.251953
		female	581.3964844	1744.189453	4349.707031
1	LPC	male	521.8876053	1437.077227	2579.878502
		female	400.9784328	544.3760408	2007.459259
	AR	male	516.796875	1442.724609	2583.984375
		female	473.7304688	2002.587891	2906.982422
2	LPC	male	691.0725463	1202.121295	2464.592894
		female	305.8387963	901.5554243	1541.214773
	AR	male	689.0625	2454.785156	3552.978516
		female	301.4648438	882.8613281	1485.791016
3	LPC	male	445.0288767	1490.902572	2582.620256
		female	492.3534833	1709.711179	2737.933704
	AR	male	452.1972656	1464.257813	2562.451172

		female	495.2636719	1679.589844	2734.716797
4	LPC	male	495.5990776	1516.411565	2503.090027
		female	577.9139614	1877.616814	2826.491271
	AR	male	495.2636719	1507.324219	2497.851563
		female	559.8632813	1873.388672	2799.316406
5	LPC	male	460.6138599	2742.094967	3437.578389
		female	354.26098	566.9373898	2219.561231
	AR	male	452.1972656	2756.25	3380.712891
		female	344.53125	2217.919922	2906.982422
6	LPC	male	447.7216122	1550.722402	2678.027283
		female	669.4357545	1831.905153	2953.268879
	AR	male	473.7304688	1528.857422	2627.050781
		female	667.5292969	1830.322266	2950.048828
7	LPC	male	500.2770134	1481.604682	2593.553915
		female	562.1794803	2029.65905	3087.227747
	AR	male	495.2636719	1464.257813	2583.984375
		female	538.3300781	2024.121094	3079.248047
8	LPC	male	508.1956401	1494.308293	2504.496928
		female	657.0832772	1208.580071	2646.899751
	AR	male	495.2636719	1485.791016	2497.851563
		female	667.5292969	1184.326172	2627.050781
9	LPC	male	437.7832273	1756.529661	2570.653131
		female	473.5147025	2149.433643	2956.454328
	AR	male	430.6640625	1765.722656	2562.451172
		female	473.7304688	2153.320313	2950.048828

Observing the values above, the difference between the formants calculated form both methods for the same signal was computed, and the 4 signals with the highest differences were taken. The ones with highest differences were chosen as there must be one closer to the real values observed manually from the spectrogram. Those 4 signals are [0] female, [1] female, [2] male, and [5] female.

Comparison

[0] Female

The figure below show the signal with its spectrogram, with the formants plotted on it.

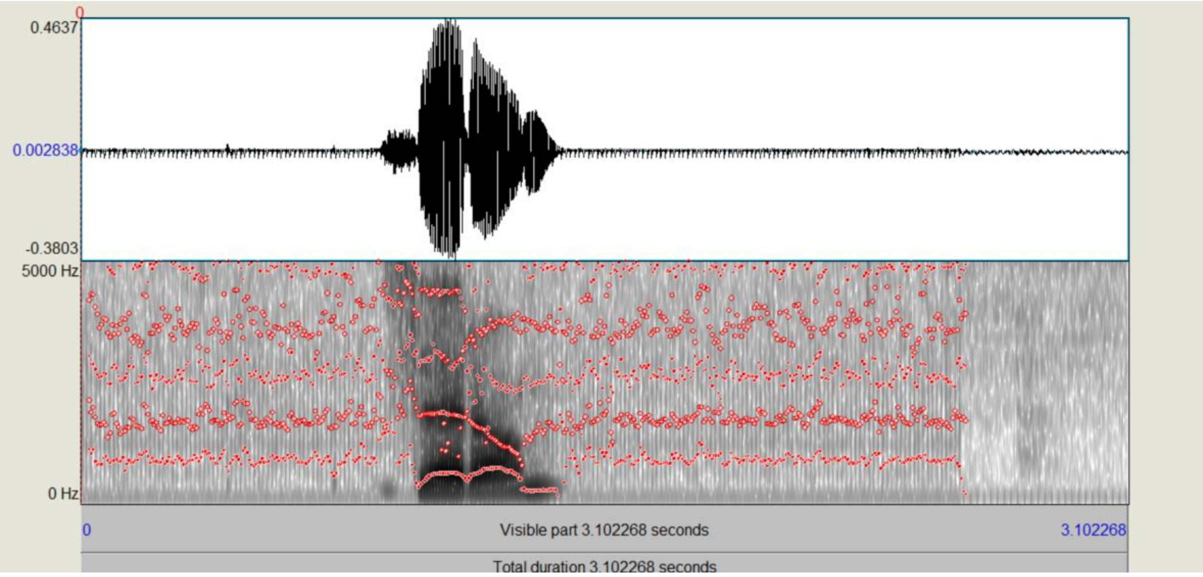

Figure 11 [0] Female Formants

The formants can be easily detected from the figure above, but as it can be seen that there is a lot of noise around the actual formants, so that may cause inaccurate results from the methods implemented to calculate the formants. The formants were estimated to be as follows.

Table 2 [0] Female Formants Manually from Spectrogram

F1	F2	F3	F4
575	1850	2955	4309

By analyzing those values, they can be compared to those obtained by LPC and AR. The 4th formants was added as it will crucial to the comparison below.

Table 3 [0] Female Formants with LPC & AR

0	LPC	female	431.0240333	703.2429245	1776.970856
	AR	female	581.3964844	1744.189453	4349.707031

After seeing all the results, it can be deemed that the values obtained by the AR method are much closer to the values observed on the spectrogram. The F1 was very close; the F2 was not so far for the real value; but when it comes to F3, it can be seen that the AR method mistakenly skipped the values of F3 and computed F4 as F3.

[1] Female

The figure below show the signal with its spectrogram, with the formants plotted on it.

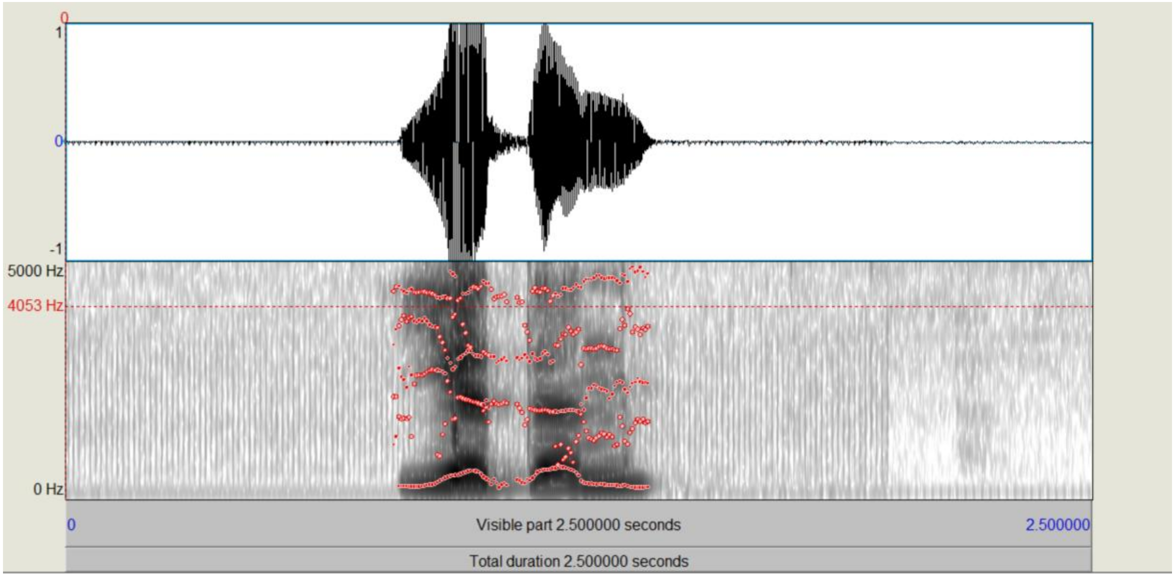

Figure 12 [1] Female Formants

The formants can be easily detected from the figure above, and unlike the previous signal the noise is minimal. The formants were estimated to be as follows.

Table 4 [1] Female Formants Manually from Spectrogram

F1	F2	F3	F4
473	2478	2964	4218

By analyzing those values, they can be compared to those obtained by LPC and AR.

Table 5 [1] Female Formants with LPC & AR

1	LPC	female	400.9784328	544.3760408	2007.459259
	AR	female	473.7304688	2002.587891	2906.982422

After seeing all the results, it can be deemed that the values obtained by the AR method are much closer to the values observed on the spectrogram. The F1 is exact; the F2 was a little bit off from the real value; and F3 is very close to the actual value.

[2] Male

The figure below show the signal with its spectrogram, with the formants plotted on it.

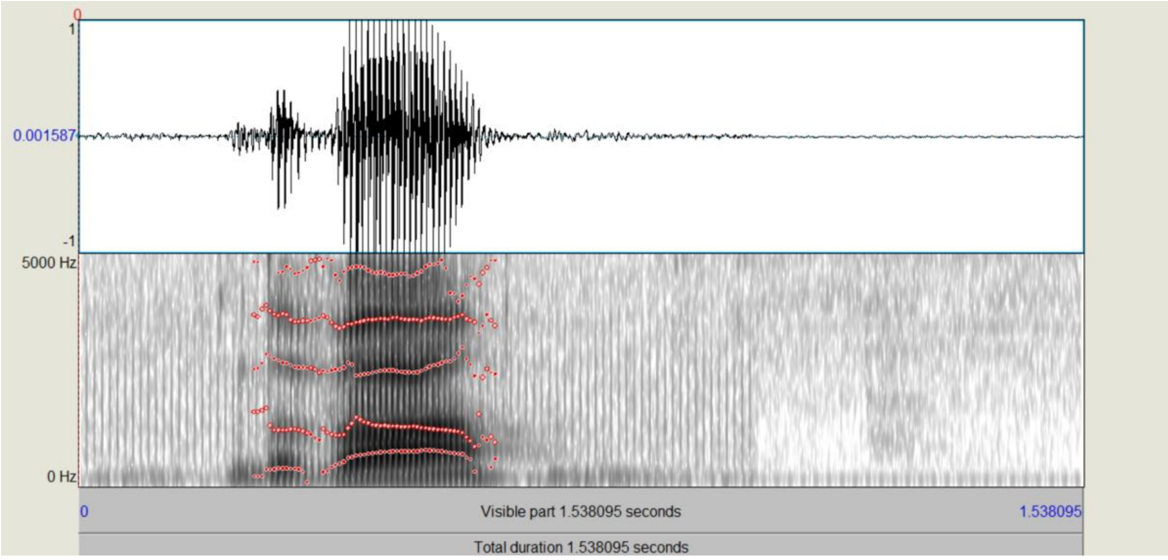

Figure 13 [2] Male Formants

The formants can be easily detected from the figure above, and unlike the previous signal the noise is minimal. The formants were estimated to be as follows.

Table 6 [2] Male Formants Manually from Spectrogram

F1	F2	F3	F4
687	1378	2411	3545

By analyzing those values, they can be compared to those obtained by LPC and AR. The 4th formants was added as it will crucial to the comparison below.

Table 7 [2] Male Formants with LPC & AR

2	LPC	male	691.0725463	1202.121295	2464.592894
	AR	male	689.0625	2454.785156	3552.978516

After seeing all the results, it can be deemed that the values obtained by the LPC method are much closer to the values observed on the spectrogram. The F1 in both methods is very close to the actual value. The F2 in LPC is very far from the actual value, while in AR F2 was mistakenly skipped and F3 and F4 were computed as F2 and F3. The F3 in LPC is very close to the actual value, while the one for AR, as it was explained, is mistakenly the value for F4.

[5] Female

The figure below show the signal with its spectrogram, with the formants plotted on it.

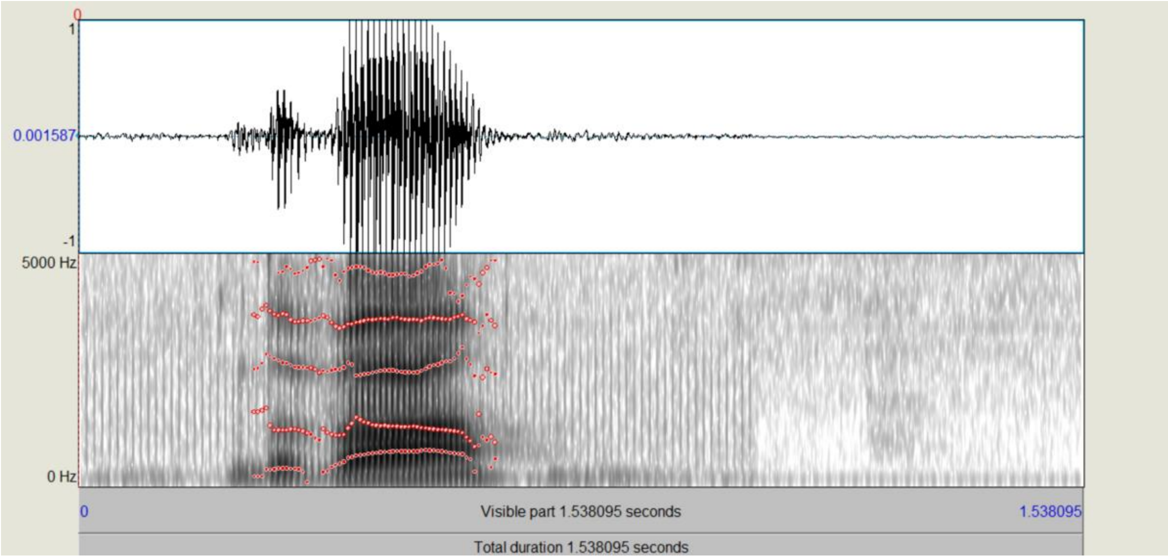

Figure 14 [5] Female Formants

The formants can be easily detected from the figure above, and unlike the previous signal the noise is minimal. The formants were estimated to be as follows.

Table 8 [5] Female Formants Manually from Spectrogram

F1	F2	F3	F4
367	2397	3176	4107

By analyzing those values, they can be compared to those obtained by LPC and AR.

Table 9 [5] Female Formants with LPC & AR

5	LPC	female	354.26098	566.9373898	2219.561231
	AR	female	344.53125	2217.919922	2906.982422

After seeing all the results, it can be deemed that the values obtained by the AR method are much closer to the values observed on the spectrogram. The F1 in both methods is very close to the actual value. The F2 in LPC is most likely the result of some noise, as F3 in LPC is very close to the actual F2. The F2 and F3 in AR are not that far from the actual values for the formants.

After analyzing the comparison results from those 4 signals, it can be seen that the AR model was more accurate in calculating the formants, except for one signal but overall it provided better results.

Results

The next step is to use the values of the formants of the male and female signals to adjust the original signals to increase the similarities between them. Below the new formants values along with the spectrograms of the adjust signals will be displayed.

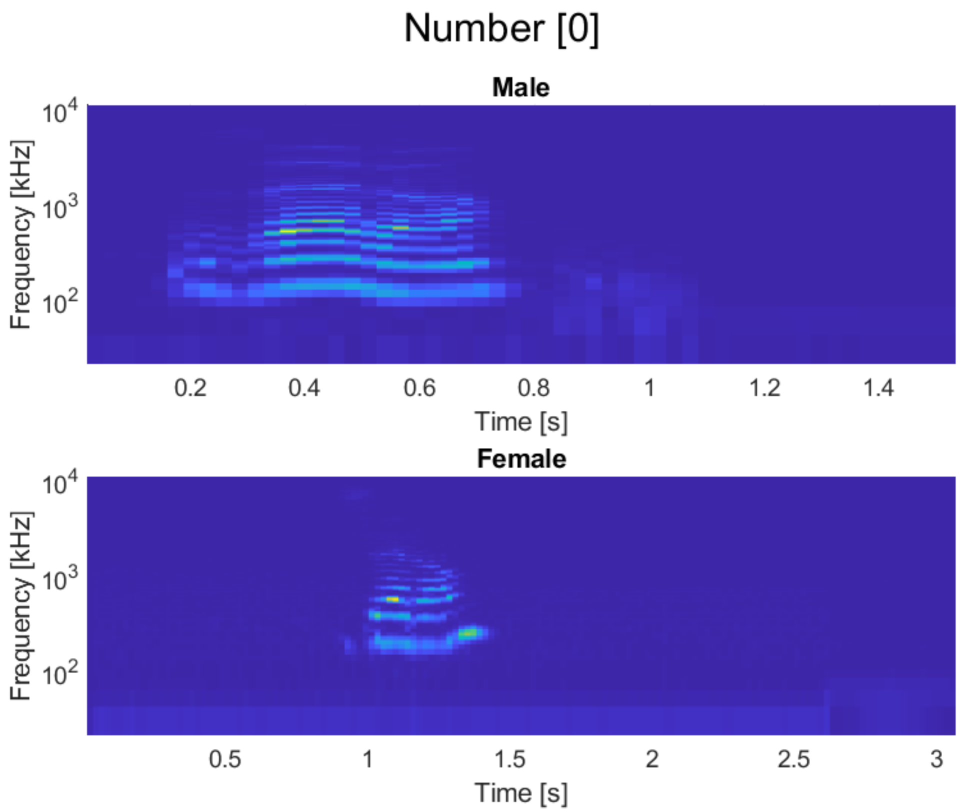

Figure 15 Adjusted Number [0] (zero)

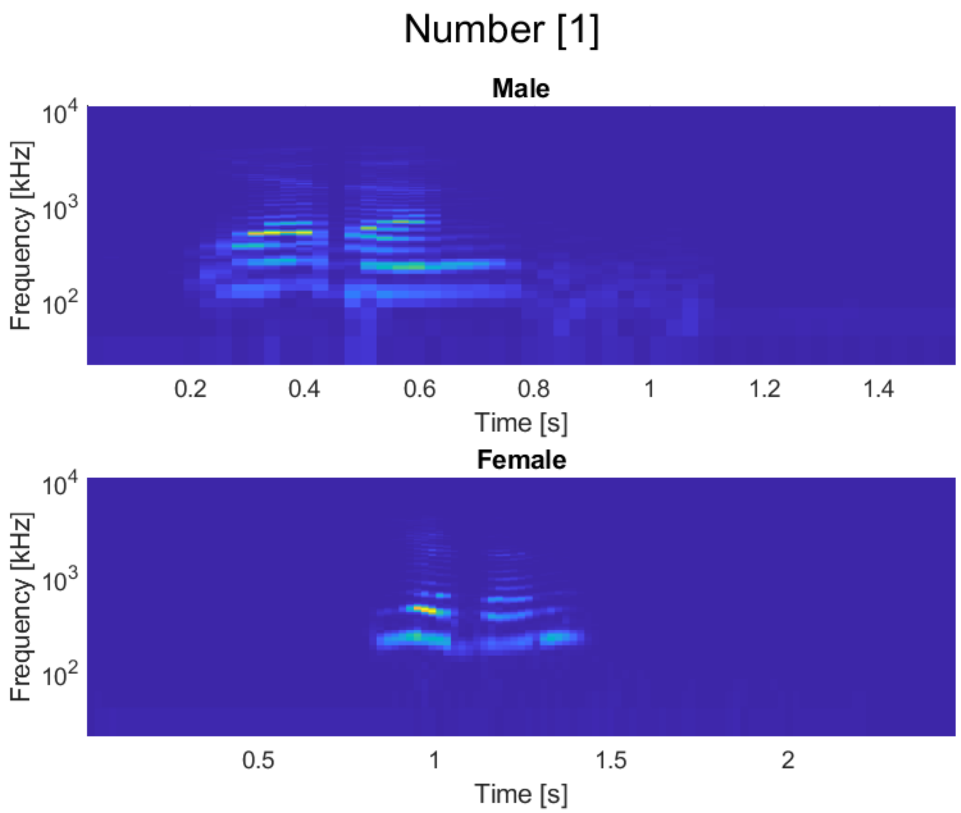

Figure 16 Adjusted Number [1] (jeden)

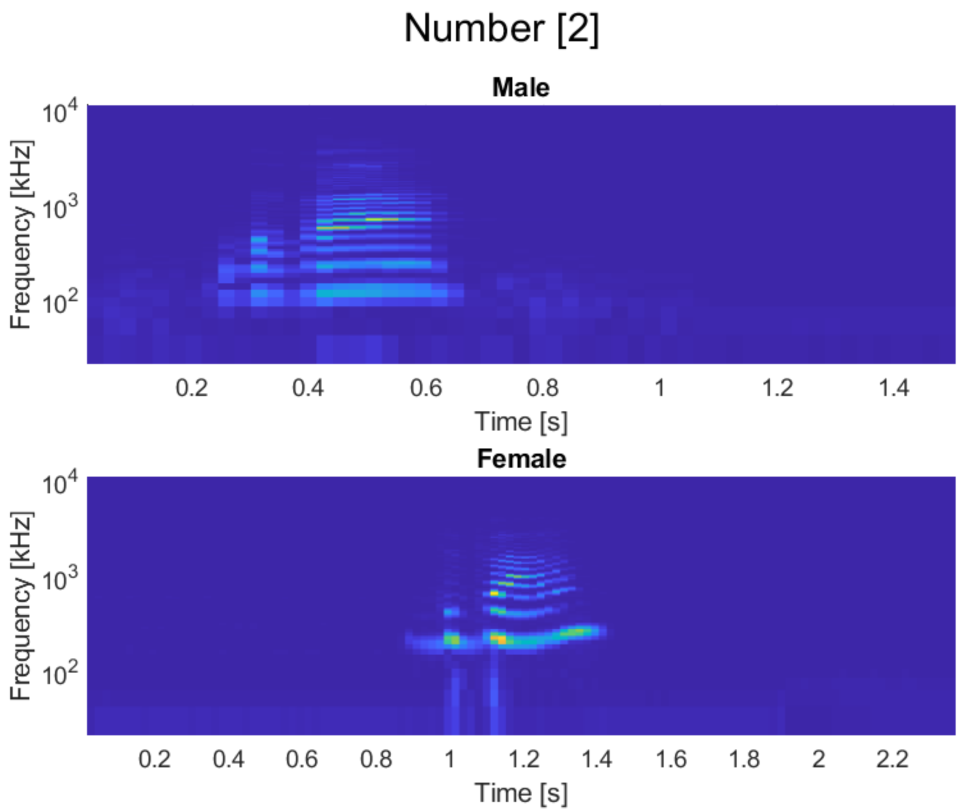

Figure 17 Adjusted Number [2] (dwa)

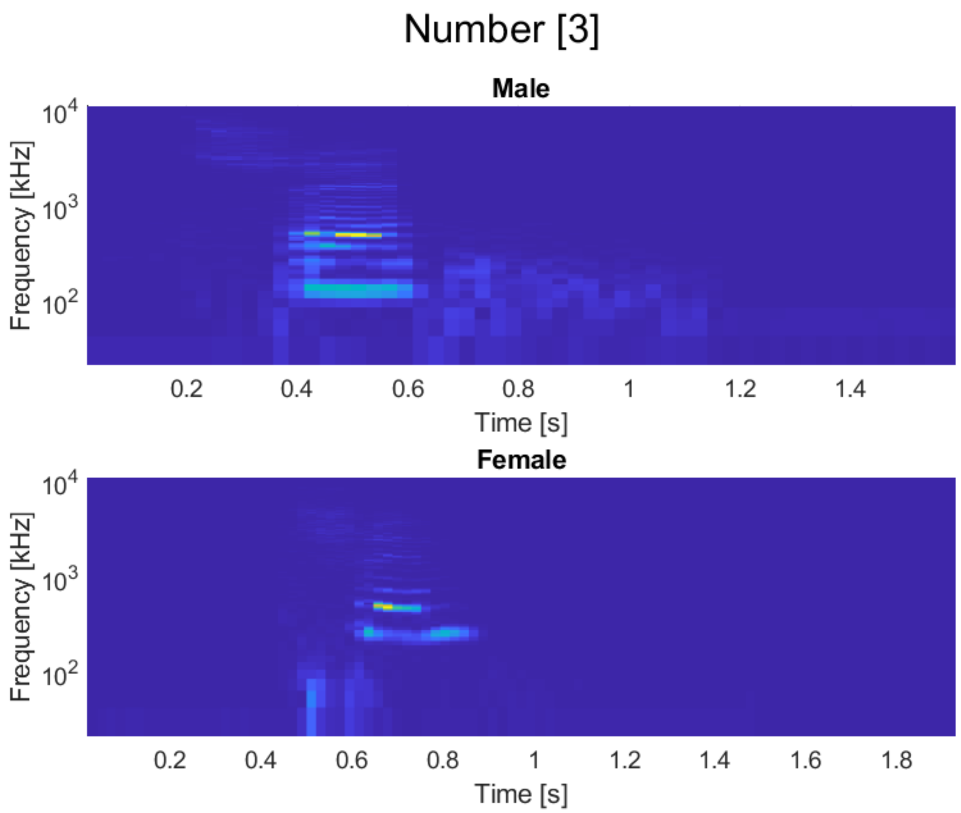

Figure 18 Adjusted Number [3] (trzy)

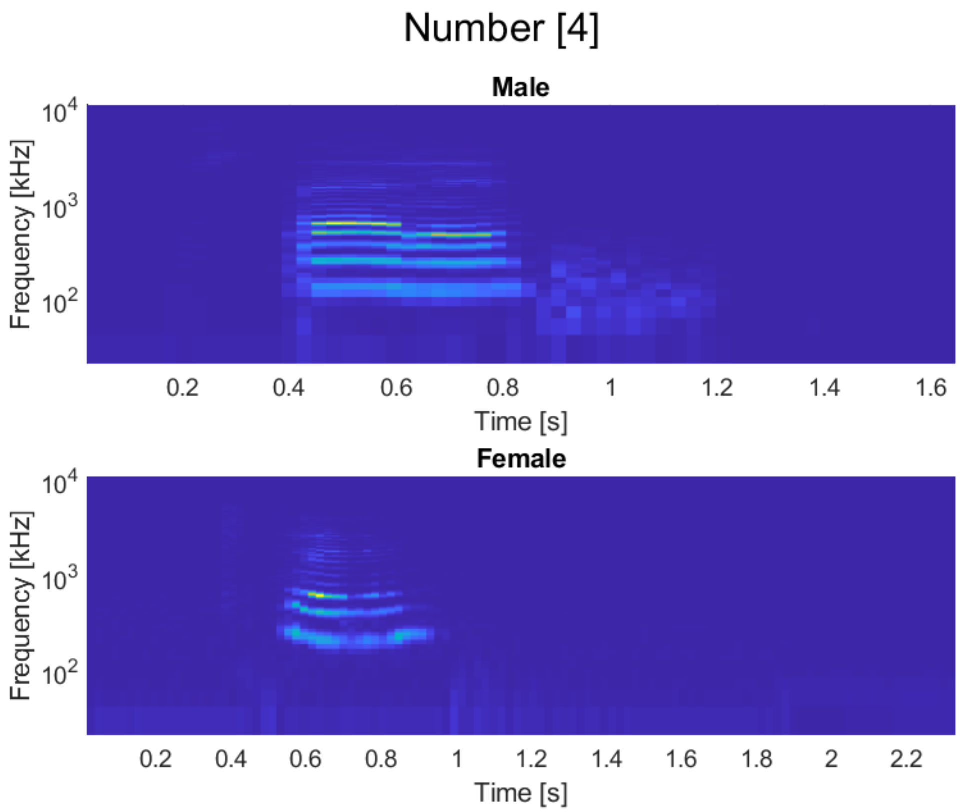

Figure 19 Adjusted Number [4] (cztery)

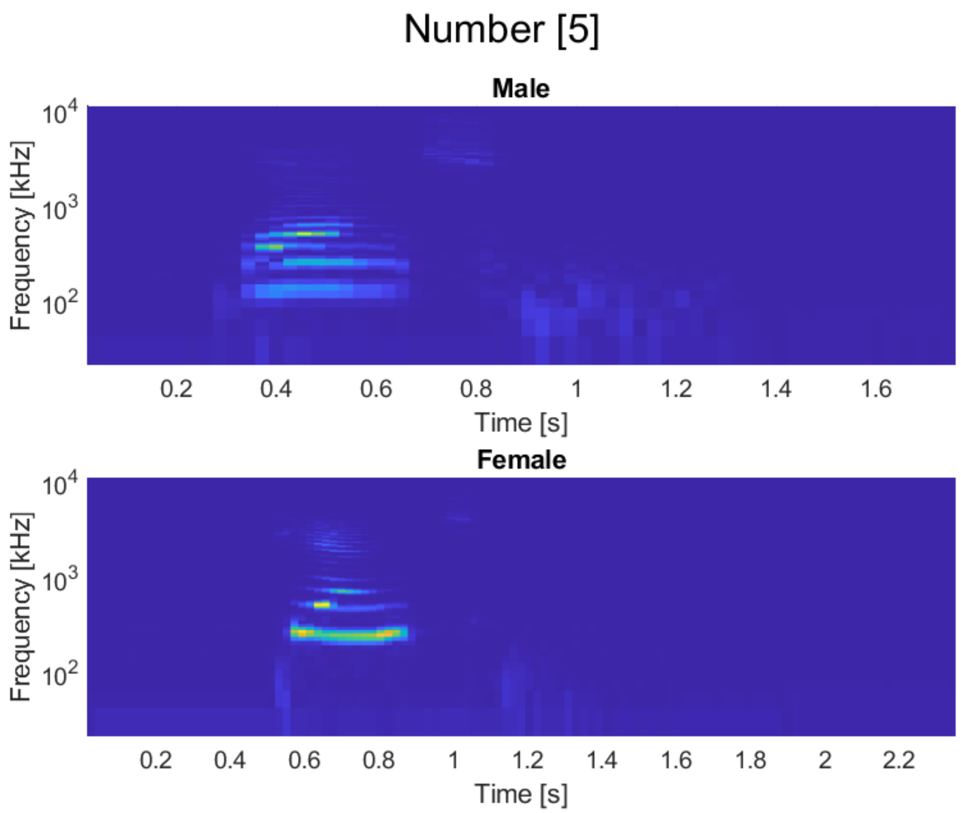

Figure 20 Adjusted Number [5] (pięć)

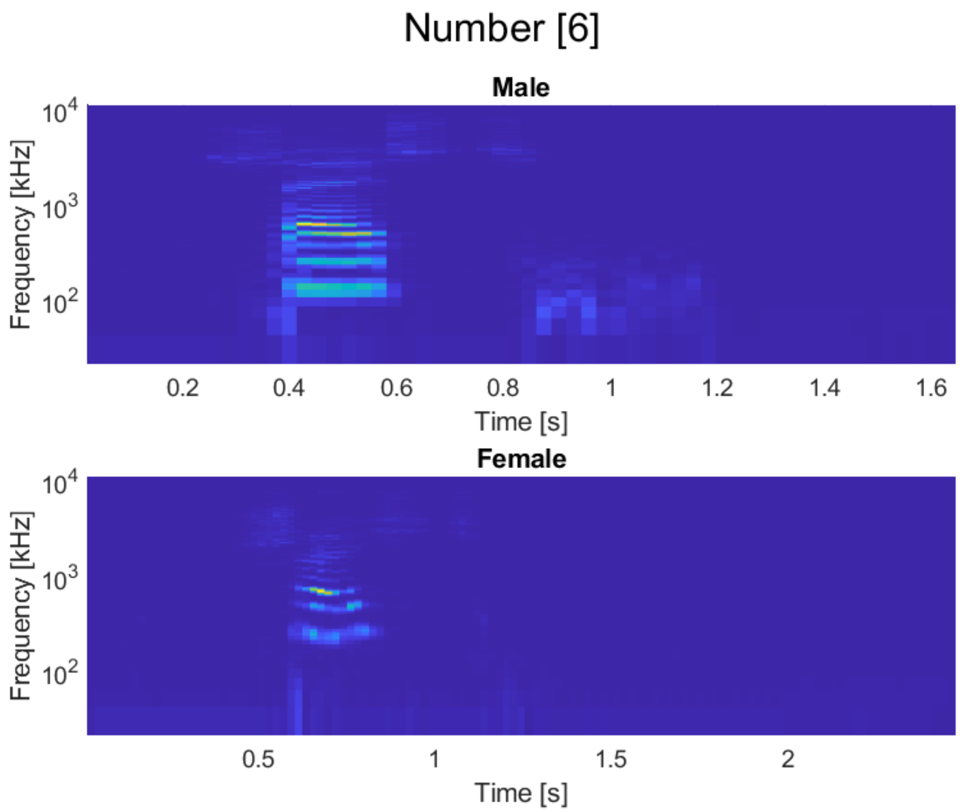

Figure 21 Adjusted Number [6] (sześć)

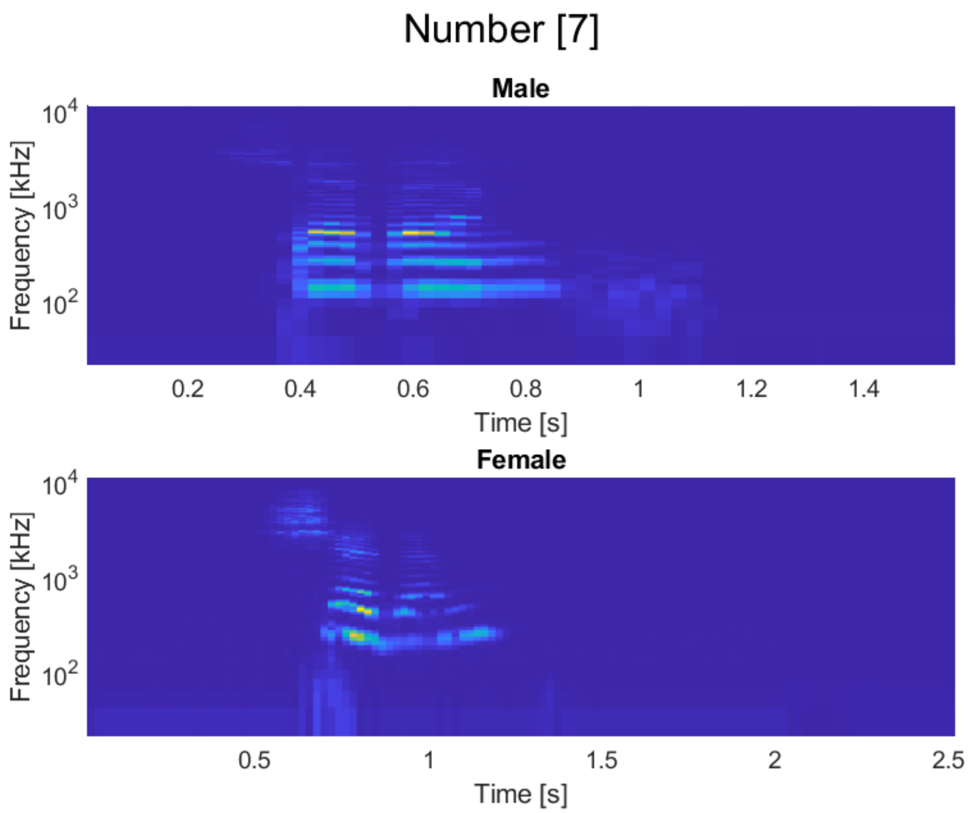

Figure 22 Adjusted Number [7] (siedem)

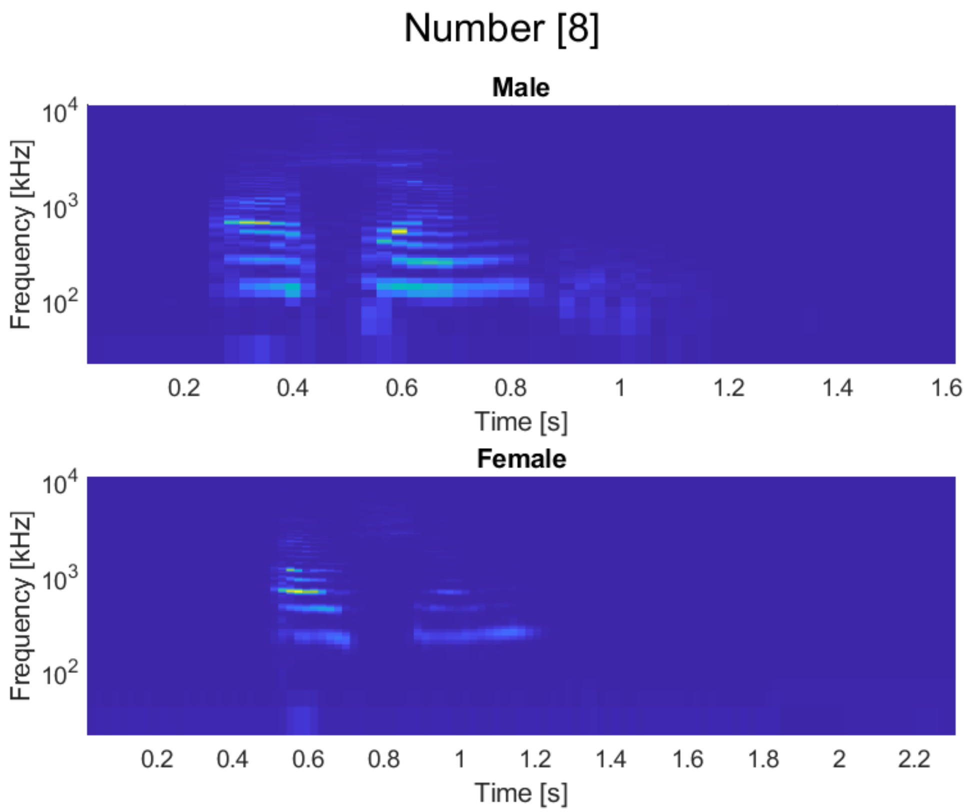

Figure 23 Adjusted Number [8] (osiem)

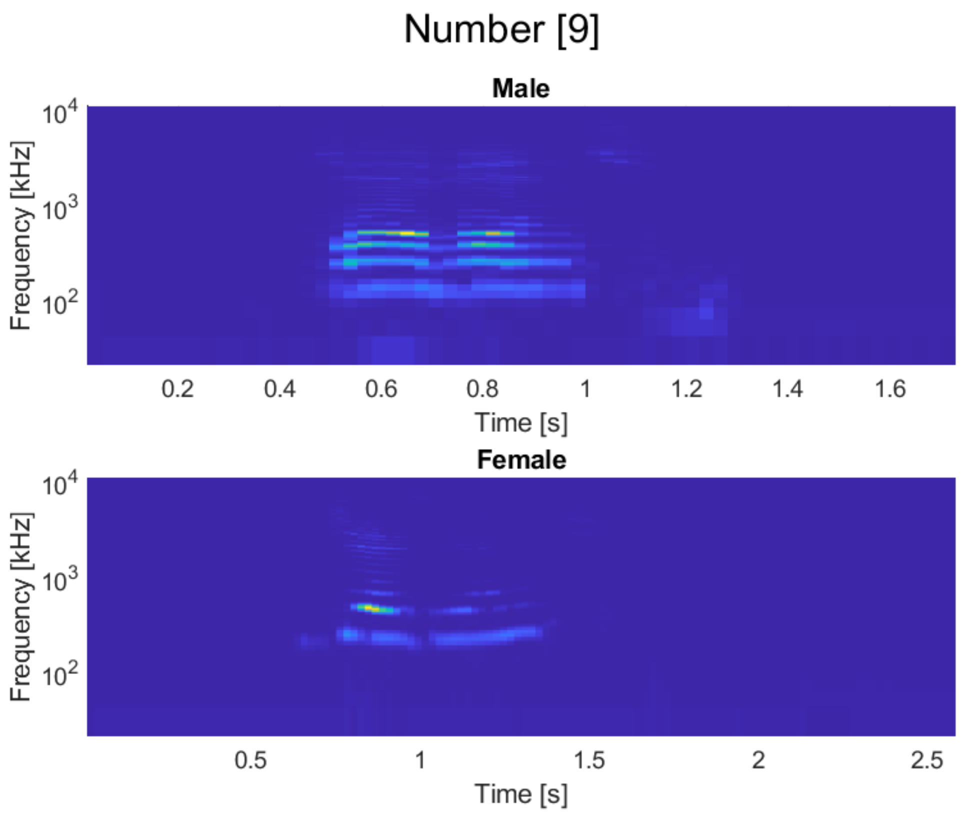

Figure 24 Adjusted Number [9] (dziewięć)

The new formants are as follows.

Table 10 Formants of the Adjusted Signals

			F1	F2	F3
0	AR	male	581.3964844	1335.058594	2691.650391
		female	409.1308594	1571.923828	3962.109375
1	AR	male	559.8632813	1593.457031	2777.783203
		female	430.6640625	1830.322266	2713.183594
2	AR	male	689.0625	2519.384766	3639.111328
		female	279.9316406	882.8613281	1485.791016
3	AR	male	495.2636719	1550.390625	2691.650391
		female	452.1972656	1614.990234	2605.517578
4	AR	male	538.3300781	1679.589844	2691.650391
		female	538.3300781	1744.189453	2648.583984
5	AR	male	473.7304688	2906.982422	3574.511719
		female	344.53125	2131.787109	2756.25
6	AR	male	559.8632813	1636.523438	2756.25
		female	645.9960938	1636.523438	2713.183594
7	AR	male	516.796875	1636.523438	2799.316406
		female	495.2636719	1851.855469	2842.382813
8	AR	male	495.2636719	1636.523438	2713.183594
		female	602.9296875	2540.917969	3875.976563
9	AR	male	473.7304688	1916.455078	2777.783203
		female	452.1972656	2002.587891	2756.25

Conclusions

Upon observing the spectrograms and the formants of the adjusted signals, the following can be concluded. The idea of using the rate of change of the formants to adjust the signals was satisfactory, as the values for the formants are much closer from the male and female signals in the adjusted signals, but it is also clear that were a lot errors in the formants values. Those errors are simply the errors in the formants calculations. Although two different methods were used and compared, the formants calculation is not as good as it should be. This lack of performance in the formants calculation is the main part that could be improved in any future versions of this project. Overall, the project works in the way it was intended, but its performance in terms of formants calculation could be improved as stated above.

References

1. https://en.wikipedia.org/wiki/Formant#:~:text=In%20speech%20science%20and%20phonetics,of%20the%20human%20vocal%20tract.&text=A%20room%20can%20be%20said,from%20its%20walls%20and%20objects.
2. https://www.statisticshowto.com/autoregressive-model/#:~:text=An%20autoregressive%20(AR)%20model%20predicts,that%20precede%20and%20succeed%20them.
3. https://www.clear.rice.edu/elec431/projects96/digitalbb/index.html
4. https://en.wikipedia.org/wiki/Linear_predictive_coding
5. https://www.sciencedirect.com/topics/medicine-and-dentistry/formant
6. https://home.cc.umanitoba.ca/~krussll/phonetics/acoustic/formants.html
7. P. Padmini, D. Gupta, M. Zakariah, Y. A. Alotaibi and K. Bhowmick, "A Simple Speech Production System Based on Formant Estimation of a Tongue Articulatory System Using Human Tongue Orientation," in IEEE Access, vol. 9, pp. 4688-4710, 2021, doi: 10.1109/ACCESS.2020.3048076.
8. https://sail.usc.edu/~lgoldste/General_Phonetics/Week10/Formant_Analysis/index.html