Voice Recognition Research

Francisco Capuzo[1] , Lucas Santos[2] , Maria Reis[3] and Thiago Coutinho[4]
[1234] Lab245 Research Center, Lab245 , Rio de Janeiro, Rio de Janeiro, Brazil

Abstract - *The main objective of the software is to create a database of voice recording files, which has an interface that can store and analyze audio data. Therewith, it is possible to gather and identify many information, for example: the voice tone and the peak of the frequency.*

Our studies were based on data collection and analysis. Therefore, we've built a structure with many information that was really relevant to the word comparing and identification processes.

Our final intention is to expand this database in order to increase the number of users and improve the accuracy of the software.

Keywords - Voice, Frequency, Studies, Samples, Word.

1. Introduction

The voice is a tool of communication that's exclusive to humans. Talking is our main way to express our feelings and exchange information in everyday life.

It has characteristics that may vary according to the gender, person characteristc or age. It's also able to reflect each person's emotional status. The voice carries the identity of each individual, and it works similar to a fingerprint.

Our first idea was to create an app that could simultaneously write in words what people were talking. Considering that the process of voice recognition is very complex, we started converting the voice into amplitude spectrum in a *time domain graphic*. The ordinate corresponds to signal amplitude and abscissa corresponds to the time.

After this, we realized that the most important thing was to transform the amplitude spectrum into the *frequency domain graphic*, in which, we had more information than in the *time domain* one. The axis of the abscissa is built with frequency values. So, it is possible to analyze if the voice is low or high for example.

We have an interface that makes this conversion. So, we can take reliable frequency signature (see **Picture-1**):

The graph construction is another useful application. It shows all information about the voice, such as: where is the largest signature and the smallest signature.

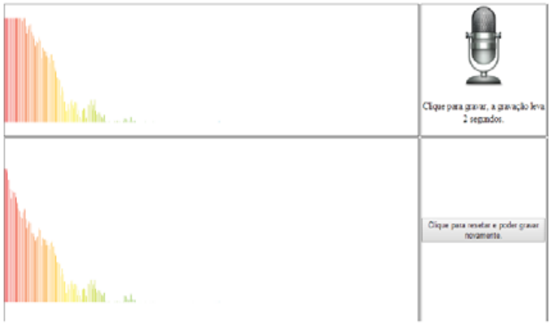
2. Development

The first step in the development process was to create a code that was able to convert the recorded voice into a frequency domain.

In the beginning, to complete the recording process, it was necessary two clicks. One click for starting and another one to finish.

But after some tests, we changed that.

For many reasons, mainly for human error (timing and miss-clicks), it would be easier if the user had a predetermined time (two seconds, in this case) to make the recording, right after click the *Start* button. It can be observed in **picture 2.1**:



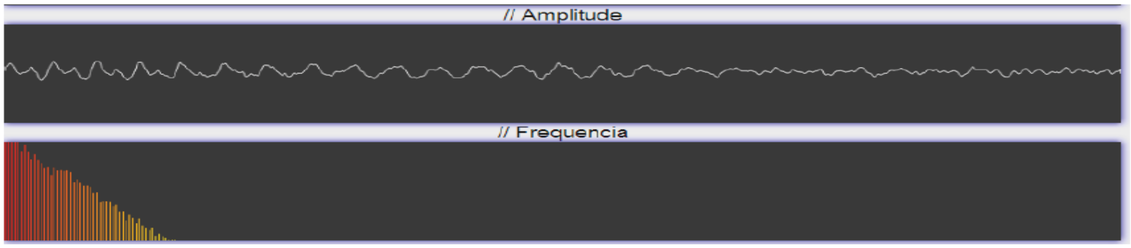
Clique para gravar, a gravação leva 2 segundos.

Picture-2.1-Recoding Interface

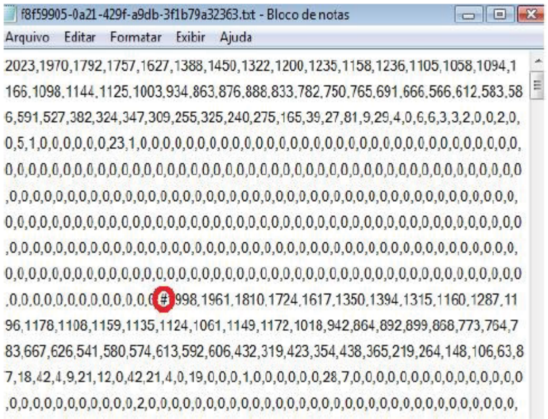
For a new recording, the user must reset the recording process clicking on the *Reset* button.

The collected data is going to be inserted in a database. For the data analysis, the software has an algorithm that transforms the whole recording into numbers, commas, and hashes.

In order to make clearer where each recording starts and ends, it is set a "#" between each recording every time it captures a sample. There is an example that can be observed in **Picture 2.2**:



Picture-1-Amplitude Spectrum/Frequency Domain.



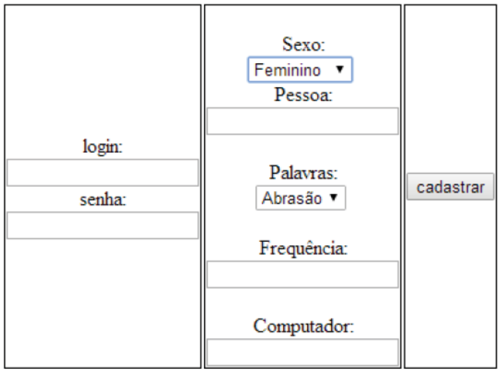
f8f59905-0a21-429f-a9db-3f1b79a32363.txt - Bloco de notas
Arquivo Editar Formatar Exibir Ajuda

```
2023,1970,1792,1757,1627,1388,1450,1322,1200,1235,1158,1236,1105,1058,1094,1
166,1098,1144,1125,1003,934,863,876,888,833,782,750,765,691,666,566,612,583,58
6,591,527,382,324,347,309,255,325,240,275,165,39,27,81,9,29,4,0,6,6,3,3,2,0,0,2,0,
0,5,1,0,0,0,0,0,23,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,
0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0
,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,
0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0
,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,
0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0
,0,0,0,0,0,0,0,0,0,0,0,0,0#998,1961,1810,1724,1617,1350,1394,1315,1160,1287,11
96,1178,1108,1159,1135,1124,1061,1149,1172,1018,942,864,892,899,868,773,764,7
83,667,626,541,580,574,613,592,606,432,319,423,354,438,365,219,264,148,106,63,8
7,18,42,4,9,21,12,0,42,21,4,0,19,0,0,0,1,0,0,0,0,0,28,7,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0
,0,0,0,0,0,0,0,0,0,0,0,2,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,
```

Picture-2.2-Samples Exemples.

It’s important to know that, in these two seconds, the frequency is stored continuously, until the recording process ends. So, the frequency changes according to the sample vector and, there are maximum and minimum of each vector positions. Thereby, we can analyze the entire recording process.

For the next step **(Picture-2.3**), the user must enter his login and password. After that, the frequency is automatically stored in a page interface. In order to finish the process, the user must fill all the fields (it gathers all the data that were previously mentioned: word, gender, and computer) and click on “cadastrar” button:



login:

senha:

Sexo: Feminino ▼

Pessoa:

Palavras: Abrasão ▼

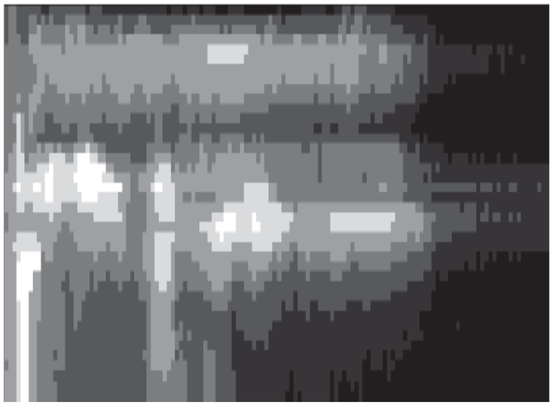
Frequência:

Computador:

Picture-2.3-Fill in the data interface.

After all this data have been stored in a database, it is possible to analyze the vectors of a recording according to the gender of the individual; computer used, or even, take the data samples per person. Once necessary the database can make calculations with these signatures, calculate the meaning of all vectors of the same word for example.

Also, there was an idea to plot these vectors in a shape of a matrix. The vector number is represented by the column, and the frequency value in the vector represents the rows, three hundred forty rows. The most interesting is that each value range has a different color, in which, we can examine the **picture 2.4** better:



Picture-2.4-Voice Graphic.

In this picture, the color white represents the highest frequency values, and the color black the lower ones. If there is a large white color concentration at the beginning, it is possible to conclude that there’s a predominantly lower voice. On the opposite side, when the situation occurs at the end, there is a high predominance voice.

3. Conclusion

Until the end of the development of this article, we went through several stages, and at each step, we have improved. From the initial moment in which we modified the frequency domain graphics until the final step in which we had to plot a graphic. Even so, the initial purpose of the software has always been to store frequency signatures in a reliable database. This would be the start of other great applications that could be used all over the world, such as:

- Build a word recognizer that would help people who want to speed up a job, or a physically handicapped or injured person, who cannot type or who otherwise has some sort of difficulty writing a text would be able to accomplish this task, whereas the reverse procedure would help a person who is visually impaired to hear a text.
- Considering that the database it is able to do any math operation, you can create a unique identity for each word and also use it as a key to unlock cell phones and computers for example.
- As the standard deviation of the words is analyzed, if a peculiar variation related to some samples sets is found, this fact may be related to some emotional state, becoming a useful tool for psychologists, pedagogues, with

the intention of recognizing the Emotional state of a patient.

- Also, it could be useful for Energy Distribution Companies. The software could monitor the frequency that is used and its lag degree, thereby finds, quickly, some technical failure.

We believe that is a study with a very big relevance that can provide structure for many possible other discoveries. Thereat, the purpose is to expand this tool and the number of words to be recorded. As a result, there will be a larger database of samples giving greater reliability to the application that can be useful for a greater number of people.

4. References

[1] IEEE ASSP MAGAZINE (Volume: 3, Issue: 1, Jan 1986)
[2] Lawrence Rabiner and Biing-Hwang Juang, Fundamental of Speech Recognition”, Prentice-Hall, Englewood Cliffs, N.J., 1993.
[3] R. M. Gray, ``Vector Quantization," IEEE ASSP Magazine, pp. 4-29, April 1984.