Available online at www.sciencedirect.com

ScienceDirect

Procedia Technology 9 (2013) 1112 – 1122

Procedia Technology

CENTERIS 2013 - Conference on ENTERprise Information Systems / HCIST 2013 - International Conference on Health and Social Care Information Systems and Technologies

Vocal Acoustic Analysis - Jitter, Shimmer and HNR Parameters

João Paulo Teixeira*, Carla Oliveira, Carla Lopes

Polytechnic Institute of Bragança, Bragança, Portugal

Abstract

A new procedure for automatic diagnosis of pathologies of the larynx is presented. The new procedure has the advantage over other traditional techniques of being non-invasive, inexpensive and objective. The algorithms for determination of jitter and shimmer parameters by their Jitta, Jitt, RAP, ppq5 in case of jitter and Shim, SHDB, apq3 and apq5 in case of shimmer are presented. The algorithm developed and implemented for determining the HNR (Harmonic to Noise Ratio) are also presented. The developed tools allow the diagnosis that indicates whether or not the voice is pathologic.

Keywords: Fundamental frequency; speech jitter; speech shimmer; Harmonic to Noise Ratio; larynx pathologies.

1. Introduction

The present work has as a backdrop the study of the speech signal in Signal Processing, and has as main objective to make the automatic diagnosis of laryngeal pathologies. From the speech signal it is possible to extract a set of parameters of the voice. Thus, it is possible to detect pathologies of the vocal cords in individuals by comparing the data of patients with certain pathology with the data of persons considered with wealthy voice.

The disorder's voice can be noticeable by the analysis of several acoustic signal parameters.

In the medical field various techniques have been used to assess the patient's voice quality. One of them

* Corresponding author. Tel.: +351 273303129; fax: +351 273313051.
E-mail address: joaopt@ipb.pt.

doi:10.1016/j.protcy.2013.12.124

consists in the auditory perceptual analysis. However, these may lead to different results depending on the experience of the practitioner involved. This is a subjective assessment technique which leads to the lack of consensus among professionals. Therefore it became necessary to search for an objective assessment, in which the voices were analyzed by devices which are capable of measuring several acoustic parameters, as Almeida [1].

The most common signs that may indicate changes in the larynx relate hoarseness, breathiness and roughness. The transient hoarseness may result from abuse of the voice or the casual flu. But when the hoarseness persists and becomes a characteristic voice, is indicative of pathology of the larynx. Hoarseness can also be an early symptom of cancer of the larynx, Teixeira, et al. [2]. The most common pathologies affecting voice are vocal nodules, the laryngitis, the paralysis, polyps, cysts and Reinke's Edema. Other pathologies of the larynx that may lead to dysphonic speech are ulcers of contact, as Lopes [3].

The parameters obtained by the acoustic analysis have the advantage of describing the voice objectively. With the existence of normative databases characterizing voice quality or using intelligent tools combining the various parameters, it is possible to distinguish between normal and pathological voice or even identify or suggest the pathology. These tools allow the monitoring of clinical standpoint and/or employment and reduce the degree of subjectivity of perceptual analysis, as Teixeira, et al. [2].

Currently, acoustic parameters commonly used in applications of acoustic analysis as well as the most referenced in the literature, are the fundamental frequency, jitter, shimmer and HNR. The fundamental frequency (F0), measured in Hertz, is defined as the number of times a sound wave produced by the vocal cords repeats during a given time period. It is also the number of cycles of opening/closure of the glottis. There is a typical range of values of this frequency for the different genders and ages. But these values are not stationary since F0 is also used to convey prosody. Besides, it also vary with sex and age, thought to depend on factors such as the state of mind of the person, the time of day that fit the lifestyle and professional use of voice, as Teixeira, et al. [2].

Measurements of F0 disturbance jitter and shimmer, has proven to be useful in describing the vocal characteristics. Jitter is defined as the parameter of frequency variation from cycle to cycle, and shimmer relates to the amplitude variation of the sound wave, as Zwetsch et al. [4]. In figure 1 it can be seen the representation of these parameters.

These parameters can be analyzed under a steady voice producing a vowel continuously.

The jitter is affected mainly by the lack of control of vibration of the cords; the voices of patients with pathologies often have a higher percentage of jitter. Most researchers considered as typical value variation between 0.5 and 1.0% for the sustained phonation in young adults.

The shimmer changes with the reduction of glottal resistance and mass lesions on the vocal cords and is correlated with the presence of noise emission and breathiness. It is considered pathological voice for values less than 3% for adults and around 0.4 and 1% for children, as Guimarães [5].

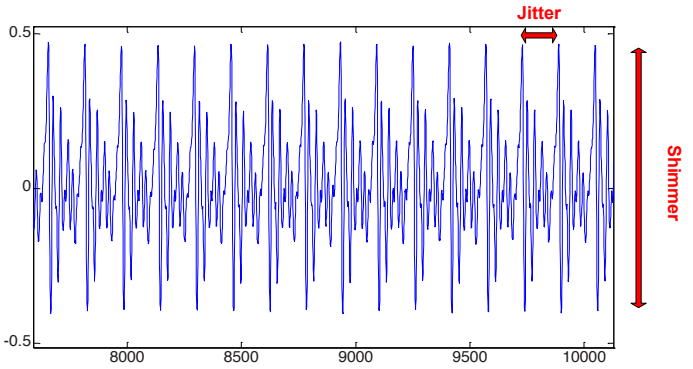

Figure 1: Representation of Jitter and Shimmer perturbation measures in speech signal.

The HNR is an assessment of the ratio between periodic components and non periodic component comprising a segment of voiced speech, as Murphy and Akande [6]. The first component arises from the vibration of the vocal cords and the second follows from the glottal noise, expressed in dB. The evaluation between the two components reflects the efficiency of speech, i.e., the greater the flow of air expelled from the lungs into energy of vibration of vocal cords. In these cases the HNR will be greater. A voice sound is thus characterized by a high HNR, which is associated with sonorant and harmonic voice. A low HNR denotes an asthenic voice and dysphonia. That is, with a value of less than 7 dB in HNR is considered pathological, as Boersma [7].

Some authors (Guimarães, [5]) presented the values of table 1 for the threshold of voice pathology.

Table 1. Threshold values for pathologic voice.

Parameter	Threshold value
Jitt (%)	1.04
Jitta (μs)	83.2
Rap (%)	0.68
ppq5 (%)	-
Shim (%)	3.81
ShdB (dB)	0.35
Apq3 (%)	-
Apq5 (%)	-

Some systems that use this set of parameters to diagnose pathologic voices do not agree in the measured parameters as reported by Bielamowicz et al. [8]. Therefore the improvement in the algorithms used to determine these parameters still been required. Brockmann-Bauser [9] developed some techniques to improve the algorithms. Other authors such Vasilakis and Stylianou [10] determine the jitter parameter in the frequency domain. The author Brockmann-Bauser [9] reported also that several factor may influence the parameters values, such the sound-pressure-level (SPL) or even the way the voice were used during the day. The age also influences the threshold for pathologic voices, as reported by Wertzner et al. [11] for children.

2. Methodology

2.1. Signal Record

The signal that is intended to be analyzed corresponds to a continuous and sustained pronunciation of one vowel. For this work the subjects reproduced the vowel /a/.

This study involved several subjects with ages between 20 and 23 years. All the subjects are students. After collecting the corresponding acoustic signals only one male and one female signal were selected for analysis. It should be noted that the selection of the subjects took into account the fact that they have no signs or symptoms of voice disorders.

Initially, the record consisted in a 3-4 seconds of sustained sound of the vowel /a/ for each speaker, with a minimum duration of 2 seconds. The record was performed using the Praat program and digitally recorded in the .wav format. The signal record was performed inside a laboratory with minimal acoustic conditions. In this room, each speaker sat comfortably and with a microphone (Sony ECM - MS907) 10cm away from the mouth. The sampling frequency used for recording these signals was 22.05 kHz, with 16 bit resolution and mono. It should be noted that the laboratory did not have the ideal characteristics, however, took up all the necessary precautions so that the signals were collected in an environment as good as possible.

2.2. Determination of jitter

To determine this parameter, that reflects the variation of the successive periods, the algorithm started to implement a function that detects the timing of the fundamental period. The output vector of the function contains the peaks levels corresponding to the beginning of the glottal pulse signal, this means, this function returns a vector of the same size but only with the peaks.

This function removes the linear trends of the signal and then uses a moving average with length corresponding to about 10 ms (a length similar to one glottal period). Then the peak is searched as the maximum of the acoustic signal under a window of 15 samples before and 15 samples after the index of the maximum of the moving average.

Analyzing the results of the algorithm the peaks are correctly extracted except when the maximum is a negative peak, as can be seen in figure 2.

Therefore, as one can ascertain the function of this timing does not detect the true maximum absolute peak, because the positive peak are detect when it should detect the negative peak because it presents a higher magnitude compared to the positive peak. This situation was corrected using the module of the input signal. The situation became corrected as shown in figure 3 for negative peaks and figure 4 for positive peaks. A problem may arise when from one period to the following the maximum peak changes from the negative peak to the positive or vice versa.

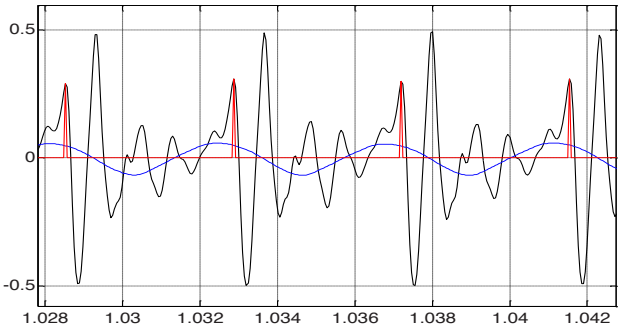

Figure 2: Representation of the signal peak corresponds to the glottal pulse in a women voice.

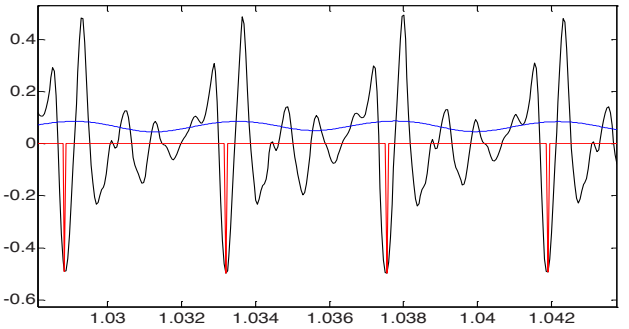

Figure 3: Visualization of the absolute maximum peaks after using the module function for negative peaks.

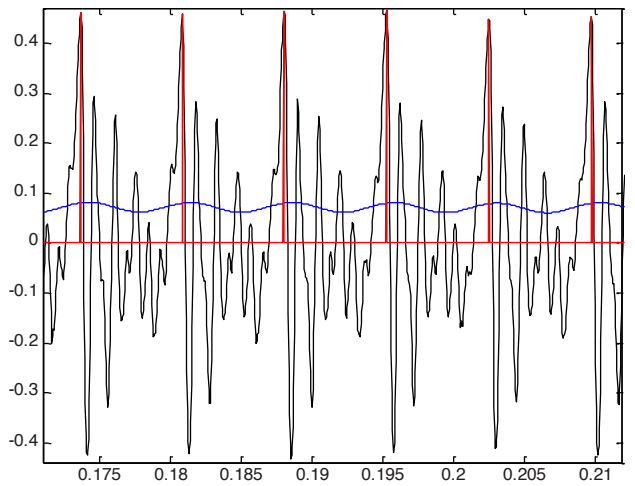

Figure 4: Visualization of the absolute maximum peaks after using the module function for positive peaks.

After the determination of the onset time of the glottal pulses the jitter can be determined for its several measures shapes given by the formulas shown below (Boersma [7]; Teixeira et al. [2]).

Jitter (local, absolute): Represents the average absolute difference between two consecutive periods and is known as jitta. The threshold value to detect pathologies in adults is 83.2 μs as reported by Guimarães [5].

$$jitta = \frac{1}{N-1}\sum_{i=1}^{N-1}\left|T_i - T_{i-1}\right| \quad (1)$$

Jitter (local): Represents the average absolute difference between two consecutive periods, divided by the average period. It is known as jitt and has 1.04% as the threshold limit for detecting pathologies.

$$jitt = \frac{jitta}{\frac{1}{N}\sum_{i=1}^{N} T_i} \times 100 \quad (2)$$

Where T_i is the duration in seconds of each period and N is the number of periods.

Jitter (rap): Represents the average for the disturbance, i.e., the average absolute difference of one period and the average of the period with its two neighbors, divided by the average period. The threshold value to detect pathologies is 0.68%.

$$rap = \frac{\frac{1}{N-1}\sum_{i=1}^{N-1}\left|T_i - \left(\frac{1}{3}\sum_{n=i-1}^{i+1} T_n\right)\right|}{\frac{1}{N}\sum_{i=1}^{N} T_i} \times 100 \quad (3)$$

Jitter (ppq5): Represents the ratio of disturbance within five periods, i.e., the average absolute difference between a period and the average containing its four nearest neighbor periods, i.e. two previous and two subsequent periods, divided by average period.

$$ppq5 = \frac{\frac{1}{N-1}\sum_{i=2}^{N-2}\left|T_i - \left(\frac{1}{5}\sum_{n=i-2}^{i+2} T_n\right)\right|}{\frac{1}{N}\sum_{i=1}^{N} T_i} \times 100 \quad (4)$$

Despite the same formulas used by different algorithm author's, the usage of the formulas can vary in the algorithms implementation considering different length of the used signal segment or even using several small segments and averaging its parameters for the whole signal. Besides, there are differences in the determination of the onset time of the glottal pulses.

2.3. Determination of shimmer

The methods used for determine the Shimmer are identical to jitter, the main difference is that the jitter considers periods and shimmer takes into account the maximum peak amplitude of the signal.

To determine the Shimmer parameters the methods used for the jitter was followed. The algorithm began by determining the onset time of the glottal pulses of the signal and the respective magnitude of the signal at that sample. Then the algorithm was applied to determine the values of each parameter of Shimmer similarly as for the jitter. The shimmer parameters are given by following expressions (Boersma [7]; Teixeira et al. [2]).

Shimmer (local): Represents the average absolute difference between the amplitudes of two consecutive periods, divided by the average amplitude. It's called a shim and this parameter was 3.81% as the limit for detecting pathologies.

$$Shim = \frac{\frac{1}{N-1}\sum_{i=1}^{N-1}\left|A_i - A_{i+1}\right|}{\frac{1}{N}\sum_{i=1}^{N}A_i} \times 100 \tag{5}$$

Shimmer (local, dB): Represents the average absolute difference of the base 10 logarithm of the difference between two consecutive periods and it is call ShdB. The limit to detect pathologies is 0.350 dB.

$$ShdB = \frac{1}{N-1}\sum_{i=1}^{N-1}\left|20*\log\left(\frac{A_{i+1}}{A_i}\right)\right| \tag{6}$$

Shimmer (apq3): represents the quotient of amplitude disturbance within three periods, in other words, the average absolute difference between the amplitude of a period and the mean amplitudes of its two neighbors, divided by the average amplitude.

$$apq3 = \frac{\frac{1}{N-1}\sum_{i=1}^{N-1}\left|A_i - \left(\frac{1}{3}\sum_{n=i-1}^{i+1}A_n\right)\right|}{\frac{1}{N}\sum_{i=1}^{N}A_i} \times 100 \tag{7}$$

Shimmer (apq5): Represents the ratio of perturbation amplitude of five periods, in other words, the average absolute difference between the amplitude of a period and the mean amplitudes of it and its four nearest neighbors, divided by the average amplitude.

$$apq5 = \frac{\frac{1}{N-1}\sum_{i=2}^{N-2}\left|A_i - \left(\frac{1}{5}\sum_{n=i-2}^{i+2}A_n\right)\right|}{\frac{1}{N}\sum_{i=1}^{N}A_i} \times 100 \tag{8}$$

2.4. Determination of HNR

The implementation of the harmonic to noise ratio was based on the mathematical fundaments presented by Boersma [7]. It started by the detection of the autocorrelation function of the voice signal, as the example displayed in figure 5.

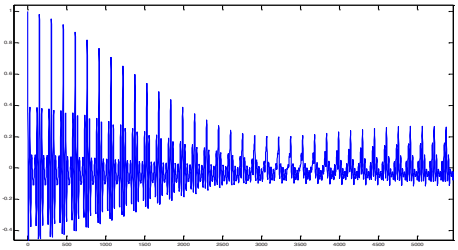

Figure 5: Visualization of 1 local maximum of the autocorrelation result.

The first local is the peak identified in Fig. 5 and corresponds to the peak after the index 1. The $AC_V(T)$ of eq. 9 is the peak at the index position corresponding to the period of the signal. Therefore the expected values for F0 define a position for that peak between two indices. Taking into account the set value for the fundamental frequency (for women of 200 to 300Hz in the case of Man between 80 and 200 Hz and in the case of child voice it varies between 400 and 500 Hz) for the first index (fs/F0max) and the second index (fs/F0min). After determining the indices the local maximum is found within the first and second index, finding their respective amplitude.

Then applying the following formula it is found the value of HNR, Boersma [7].

$$HNR = 10 * \log_{10} \frac{AC_V(T)}{AC_V(0) - AC_V(T)} \quad (9)$$

Despite the usage of the same mathematical formula the algorithms are different also because of the length of used segments or even because of the usage of several segments.

3. Presentation and Discussion of Results

In this section we present the results of jitter, shimmer and HNR for two signals under study, one male and the other female subject. In tables 2, 3, 4, 5, 6 and 7 the set of parameters are presented for the two signals. The second column shows the values obtained using the Praat software (Boersma and Weenink [12]), and the last column reports the values obtained with the algorithm developed in this work.

Table 2: Values for the jitter with a female signal

Female Signal	Praat Alg.	Devel. Alg.
Jitta (μs)	17	29
Jitt (%)	0,39	0,66
RAP (%)	0,23	0,43
ppq5 (%)	0,25	0,46

Table 3: Values for jitter with a male signal

Male Signal	Praat Alg.	Devel. Alg.
Jitta (μs)	18	30
Jitt (%)	0,26	0,43
RAP (%)	0,15	0,28

ppq5 (%)	0,15	0,28

Table 4: Values for the shimmer with a female signal

Female Signal	Praat Alg.	Devel. Alg.
Shim (%)	2,28	2,43
ShdB(dB)	0,20	0,45
apq3 (%)	1,30	2,70
apq5 (%)	1,37	0,72

Table 5 Values for the shimmer with a male signal

Male Signal	Praat Alg.	Devel. Alg.
Shim (%)	1,72	2,01
ShdB(dB)	0,15	0,10
apq3 (%)	1,00	1,37
apq5 (%)	1,07	0,79

Table 6: Values for the HNR with a female signal

Female Signal	Praat Alg.	Devel. Alg.
HNR(dB)	21,7	15,3

Table 7: Values for the HNR with a male signal

Male Signal	Praat Alg.	Devel. Alg.
HNR (dB)	23,7	17,3

The used voices did not include any pathologic voice. An analysis of jitter parameters shows that the parameters are under the threshold values for pathologic values. Comparing the output of the developed algorithm and Praat it can be considered similar and with correct diagnose. The values for RAP and ppq5 are similar in each algorithm.

The Shimmer parameters gives a healthy voice for the Shim (<3.81%), but for ShdB the developed algorithm gave a pathologic voice (female) recommending attention to the other parameters. Comparing the output of the algorithm and Praat the differences are higher than for Jitter parameters.

For HNR both cases (male and female) for the developed algorithm and for Praat the values are higher than 7 dB meaning healthy voices, although this threshold cannot be taken as scientific limit.

Since almost all values are within the range of values considered healthy voices the small differences among the developed algorithms and Praat software cannot be considered significant. The ShdB for the female voice is contradictory to the other parameters.

Analyzing the differences between Praat and the developed algorithm the following aspects should be considered. The number of peaks under analysis for both algorithms do not match, meaning that it is not analyzed the same number of peaks, hence the values differ, and therefore the length of the compared signal are not exactly the same, being this one of the reasons that can explain the small differences.

Regarding the results obtained for HNR the difference of output is due to the fact that the Praat make an average of every 80 ms, i.e., every 80 ms determines a value HNR providing in its final the average of all intervals. In the algorithm developed here the HNR value is provided taking into account the first local maximum, considering only one value using the entire signal.

The values for F0 are another factor to take into account because they are not standardized as their alteration can affect the value of HNR.

Finally, in order to verify the accuracy of the developed algorithm and Praat algorithm a synthesized /a/ vowel with exactly the same period and amplitude along periods, i.e. no jitter and no shimmer, was produced using the formant synthesizer (Teixeira and Fernandes [13]). The signal was submitted to the Praat and the developed algorithm. Tables 8 and 9 present the output for Jitter and Shimmer, respectively. The total number of cycles analyzed was 194 with the developed algorithm and 199 with Praat. It can be seen that the values presented in tables 8 and 9 are almost insignificant for both algorithms, but lower with the developed algorithm.

Table 8: Values for the Jitter with a synthesized signal

Synth. Signal	Praat Alg.	Devel. Alg.
Jitta (μs)	0.003	0.000
Jitt (%)	0,00003	0,00000
RAP (%)	0,00002	0,00000
ppq5 (%)	0,00002	0,00000

Table 9: Values for the Shimmer with a synthesized signal

Synth. Signal	Praat Alg.	Devel. Alg.
Shim (%)	0,0008	0,0003
ShdB(dB)	0,00007	0,00002
apq3 (%)	0,0003	0,0000
apq5 (%)	0,0001	0,0000

4. Conclusion

In this paper the algorithms and its implementation to determine parameters associated with jitter, shimmer and HNR in its various measures such as the jitt, jitta, RAP and ppq5 for jitter and the shim, SHDB, and Apq3 Apq5 for the shimmer was presented.

Comparing the output for the parameters for two healthy voices (one male and other female) using the Praat and the developed implementation it can be considered at the same level, and both produced a diagnose of healthy voices. Despite, a small difference exists between the algorithms that can be partially explained by the different length of signal used.

Regarding the algorithms developed to determine the jitter and shimmer, it is worth noting that the absolute maximum peak detection is extremely important for the accuracy of the output. Once these parameters are measured in relatively small values, any very small error of the index can affect the measurement and results.

Finally, the implemented algorithm can be considered accurate for determine the above mentioned parameters. For the future, several pathologic voices should be used to test the algorithm within a real situation.

References

[1] Almeida, N. Sistema Inteligente para Diagnostico da Patologias na Laringe Utilizando Maquinas de Vetor de Suporte. Msc., Universidade Federal Rio Grande do Norte – Natal – Brasil, 2010.

[2] Teixeira, J. P.; Ferreira, D.; Carneiro, S.. Análise acústica vocal - determinação do Jitter e Shimmer para diagnóstico de patalogias da fala. In 6º Congresso Luso-Moçambicano de Engenharia. Maputo, Moçambique, 2011.

[3] Lopes, J.. Ambiente da análise robusta dos principais parametros da voz. Msc. University of Porto, 2008.

[4] Zwetsch, I., Fagundes, R., Russomano, T., Scolari, D.. Digital signal processing in the differential diagnosis of beningn larynx diseases, Porto Alegre, 2006.

[5] Guimarães, Isabel. A Ciência e a Arte da Voz Humana. Escola Superior de Saúde de Alcoitão, 2007.

[6] Murphy, P. and Akande, O. Cepstrum-Based Estimation of the Harmonics-tonoise Ratio for Synthesized and Human Voice Signals. In Nonlinear Analyses and Algorithms for Speech Processing. Barcelona, LNAI 3817, Springer, 2005.

[7] Boersma, P. Accurate short-term analysis of the fundamental frequency and the harmonic-to-noise ratio of a sample sound. IFA Proceedings 1993; 17, 97-110.

[8] Bielamowicz, S.; Kreiman, J.; Gerratt, B.; Dauer, M.; Berke, G. Comparison of Voice Analysis Systems for Perturbation Measurement. Journal of Speech and Hearing Research, 1996; 39, 126-134.

[9] Brockmann-Bauser, M. Improving jitter and shimmer measurements in normal voices. Phd Thesis of Newcastle University 2011.

[10] Vasilakis M.; Stylianou, Y. Spectral jitter modeling and estimation. Biomedical Signal Processing and Control 2009; 129.

[11] Wertzner, H.; Schreiber, S.; Amaro, L. Analysis of fundamental frequency, jitter, shimmer and vocal intensity in children with phonological disorders. Rev Bras Otorrinolaringologia 2005; 71, 5, 582-88.

[12] Boersma, Paul and Weenink, David. Praat: doing phonetics by computer. Phonetic Sciences, University of Amsterdam. http://www.fon.hum.uva.nl/praat/

[13] Teixeira, J. P; Fernandes, A. Didactic Speech Synthesizer – Acoustic Module – Formants Model. Proceedings of BioSignals, 2013. Barcelona.