SEARCH OF AN ADEQUATE METHOD FOR QUANTITATIVE MEASUREMENT OF SPEECH SOUNDS DISTORTIONS DUE TO AN INADEQUATE DENTAL MOUNTING

G.E. DRĂGĂNESCU, C. SINESCU**, D. DODENCIU***, Z. FLORIȚA***

*Department of Mechanics and Vibrations, Polytechnic University of Timişoara, 1, Mihai Viteazu Blvd, 300222 Timişoara, Romania
**Department of Dental Materials and Dental Prostheses Technology, "Victor Babeş" University of Medicine and Pharmaceutics of Timişoara, 2, E. Murgu St., 300041 Timişoara, Romania
***Department of Biophysics and Medical Informatics, "Victor Babeş" University of Medicine and Pharmaceutics of Timişoara, 2, E. Murgu St., 300041 Timişoara, Romania

Abstract. In this paper we search the most adequate quantitative method based on signal analysis in order to reflect the dependence of the speech quality on the correct positioning of the dental mounting. We search also quantitative parameters, which reflect correct positioning of dental mountings.

Key words: signal analysis, speech quality, Fourier transform, autocorrelation function, power spectrum, prosthodontic products, quantitative analysis.

INTRODUCTION

It is well known that the quality and the correct positioning of the dental mountings have influence on the quality of speech production. This is due to a complex relation between the speech quality and the type and the quality of the prosthetic materials, the geometrical shape and the accuracy of building the prosthodontic products.

On the other hand, the speech sounds are complex oscillations of the air pressure around atmospheric pressure and contain an enormous quantity of information. For this reason, it is difficult to reflect in a simply manner the sound distortions which affect the sound quality.

This complex character of the speech sounds appears also after the analysis of these sounds. This fact makes difficult a quantitative description of the speech sound alterations.

Received March 2004;
in final form March 2005.

ROMANIAN J. BIOPHYS., Vol. 13, Nos. 1–4, P. 83–97, BUCHAREST, 2003

It results that the hearing sense is an analysis device more sensitive than the known manmade analysis devices.

There exist a series of physical models of speech sounds production [13], but due to the complexity of speech mechanisms these models are not able to reflect this distortion. Of major difficulty is the modeling of the contribution in the speech production of the volitional control and the learning mechanisms.

Due to the complexity of speech production phenomena the correlation between the prosthetic product quality and form and the speech quality is difficult to be done. For these reasons it results that high quality dental mountings are obtained due to the science and the professional *art* of the dentistry specialist.

On the other part, there are important investigations [5, 6, 9, 10] in order to analyze the correlation between the quality of speech production and the prosthetic mount position, generally based on the Fourier spectrum of the speech signals and on the short time Fourier spectrum (i.e. time frequency representation of the speech signals or sonograms). In these papers it was found that it is efficient to make these studies on a single phoneme like /s/, and it was also established that the analysis methods express the speech alterations. Speech distortions are also analyzed in [2, 11].

There exist also time-to-time transforms of the signals, as for example the Hilbert transform [12] and the *cepstrum* transform [4], useful in the analysis of the speech sounds distortions. These transforms are useful in order to detect a series of anomalies in the content of signals. Recently there were introduced two-dimensional representations (transforms) of the signals. Two-dimensional representations of the signals were defined long time ago, but these methods were effectively used only in our days, when the computing speed of the computers permitted to calculate these representations of the signals in real time or in an acceptable time interval, due to the fact that the calculation of these transforms implies a great number of elementary operations. On the other hand the physical significance of some of these transforms is not simple or unclear indeed. In some case these transforms are useful in order to illustrate small modifications of the signals, which are not illustrated by standard Fourier analysis. For example, the ambiguity function (Wigner-Ville transform), the wavelet analysis, Gabor transform and so on can be used [3, 4].

The final goal of the studies of the correlation between the prosthetic products and prosthetic position and the speech sound alteration is of practical nature: to obtain a rigorous, computer assisted method to establish the optimal parameters of the prosthetic mounts and the best speech quality. For this reason the aim of this paper is to search an optimal analysis method of the speech distortions and a series of quantitative parameters of speech sounds which sensitively express the minimal distortion of speech sounds.

In this preliminary paper we used a single patient, due to the fact that it has been established [8] that the characteristics investigated here reflect in an accurate manner repeatable personal characteristics of the human voice (i.e. spectral bands of the voice, fractal Hurst exponent), which can be used in the person recognition techniques for forensic needs.

MATERIALS AND METHOD

One of the most sensitive sounds to the speech alterations, which reflect the incorrect prosthetic mounts, is the phoneme /s/. For this reason there was used the first /s/ phoneme, from the word *Sisyphe*, recorded from a patient with a prosthetic mount corresponding to different mounting positions. The recordings correspond to different θ angular positions, the reference direction corresponding to the normal position, the positive angles corresponding to the labial inclination and the negative, to the opposite inclinations.

The sound recordings were done in a special room phonically isolated. We used one subject with dental problems in the front part of the dental arch (Class IV Kennedy), which received different kinds of fixed dental prostheses. We have made five fixed dental prostheses for each subject, one in the axial implantation of the teeth, one with 30º inclined in oral position, one with 60º inclined in oral position, one with 30º inclined in vestibular position and one with 60º inclined in vestibular position.

The first /s/ phoneme, which was extracted from the acoustic records, contains the *Sisyphe* word, using the *Cool Edit 2000* software for acoustic manipulations.

For the experiments we used a professional dynamic microphone SHURE SM58, a professional sound interface AUDIGY 2 and a personal computer with Cool Edit 2000 sound analysis software. The word was recorded approximately at the same sound level of 40 dB(A) for all situations. In order to ensure the same level of the signal, it was used the *Normalize* function from *Cool Edit 2000* software.

From all signals it was extracted the first /s/ phoneme. This phoneme is produced in the front alveolar region of the oral cavity.

We denote by $x(t)$ the sound pressure of the phoneme signal, recorded with a professional microphone placed in a fixed position ahead the mouth of the patient.

The auto-correlation function of the $x(t)$ signal is defined [1, 7, 14] as:

$$C_{xx}(t) = \int_{-\infty}^{\infty} x(\tau)x(\tau + t)d\tau \, . \tag{1}$$

An important property of the correlation functions is the fact that by correlation the effect of noise is reduced. The Fourier transform of the autocorrelation, obtained practically by FFT method, gives the power spectrum $S(\nu)$ of the signal. It results that in the power spectrum the effect of noise is reduced. For this reason we decided to investigate the speech alterations with the aid of power spectra.

There was used the sampling frequency of 44.1 kHz with 16 bits resolution.

The correlation functions and the power spectra were obtained with the aid of *SYSTAT* software.

The power spectra were established also from the signal derived with respect to time. In this case the power spectra are $S_{deriv}(\nu) = (2\pi\nu)^2 S(\nu)$, i.e. the spectrum is amplified (magnified) at the higher frequencies, due to the multiplication of S with ν^2.

RESULTS

Form the spectra it results that a series of changes appears in the configuration of the amplitude spectra with the mounting angle (θ). These changes can be expressed as a set of quantitative parameters. From these parameters we intend to find a number of parameters which have a sooth and reproducible variation with the θ mounting angle. Among these parameters it will be of major practical interest the parameters which will depend by a one to one mathematical manner, or an injective mathematical manner, on the θ mounting angle, or obey an extreme due value at $\theta = 0^{\circ}$ (in the normal position).

Due to the fact that the spectra of the signals contain a lot of irregularities as the power spectra do, in the paper the power spectra $S(\nu_k)$ were used. The power spectra for different θ mounting angles are presented in Figures 1–5.

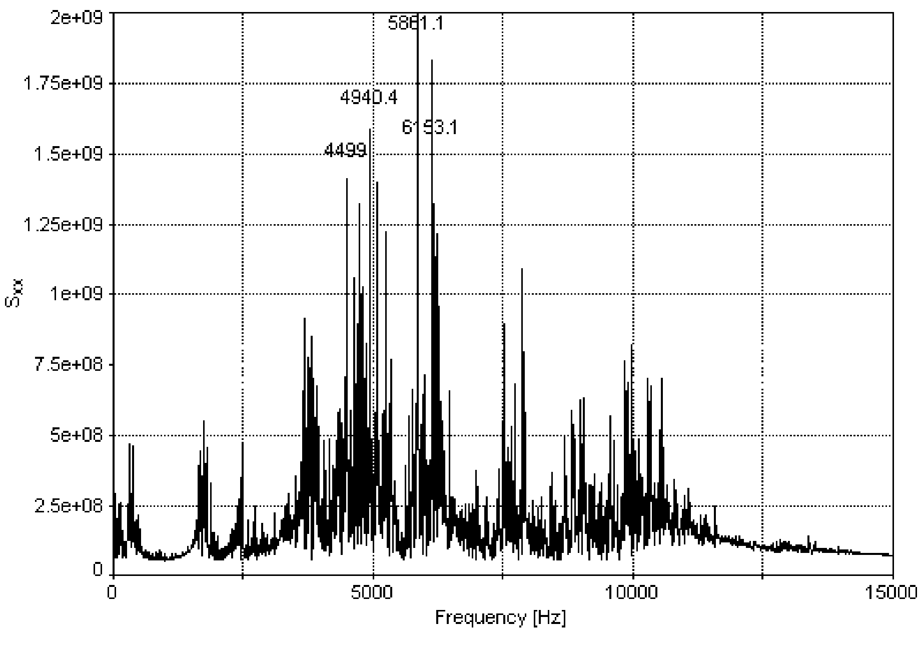

Fig. 1. – The power spectrum for $\theta = -60^{\circ}$.

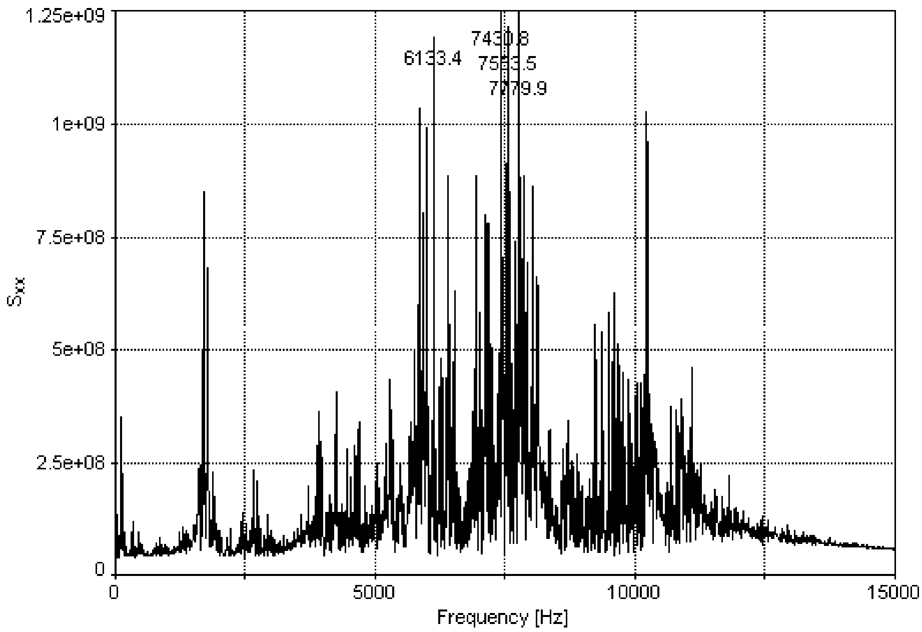

Fig. 2. – The power spectrum for $\theta = -30^{\circ}$.

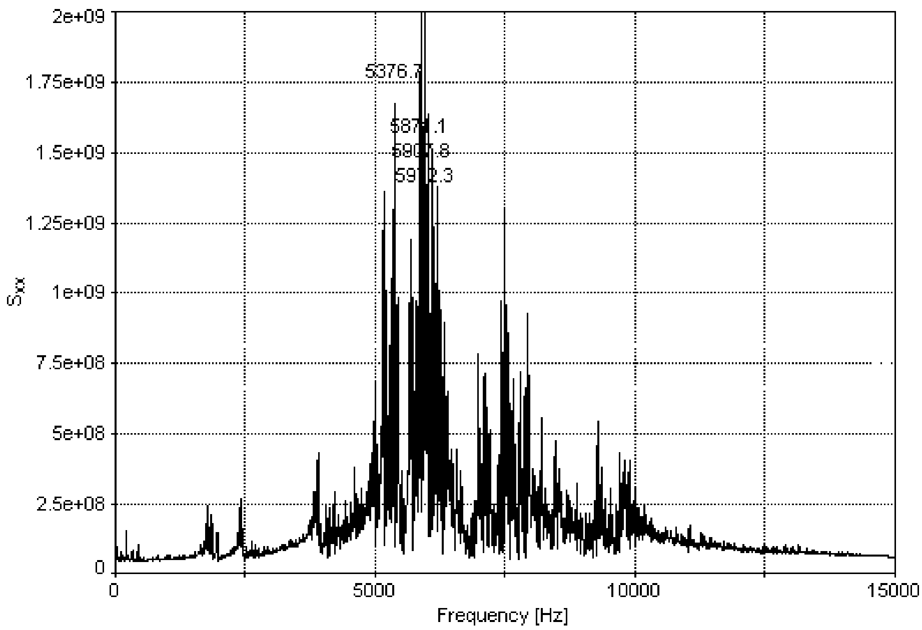

Fig. 3. – The power spectrum for $\theta = 0^{\circ}$.

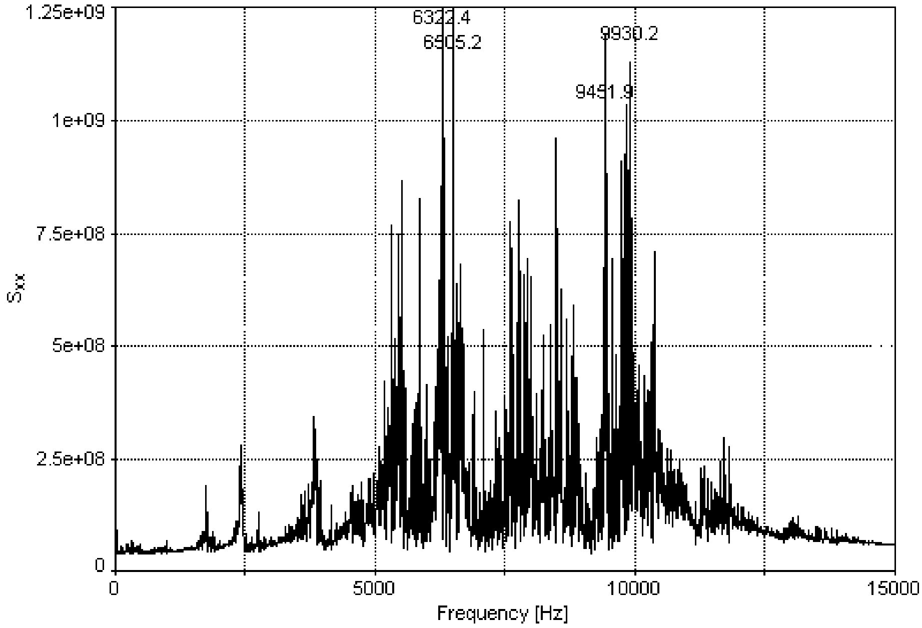

Fig. 4. – The power spectrum for θ = 30º.

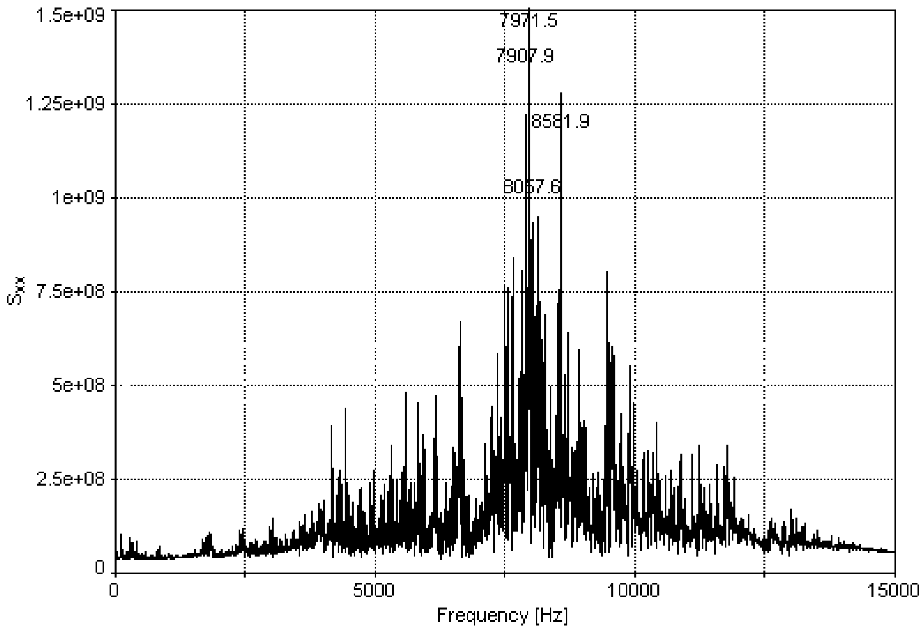

Fig. 5. – The power spectrum for θ = 60º.

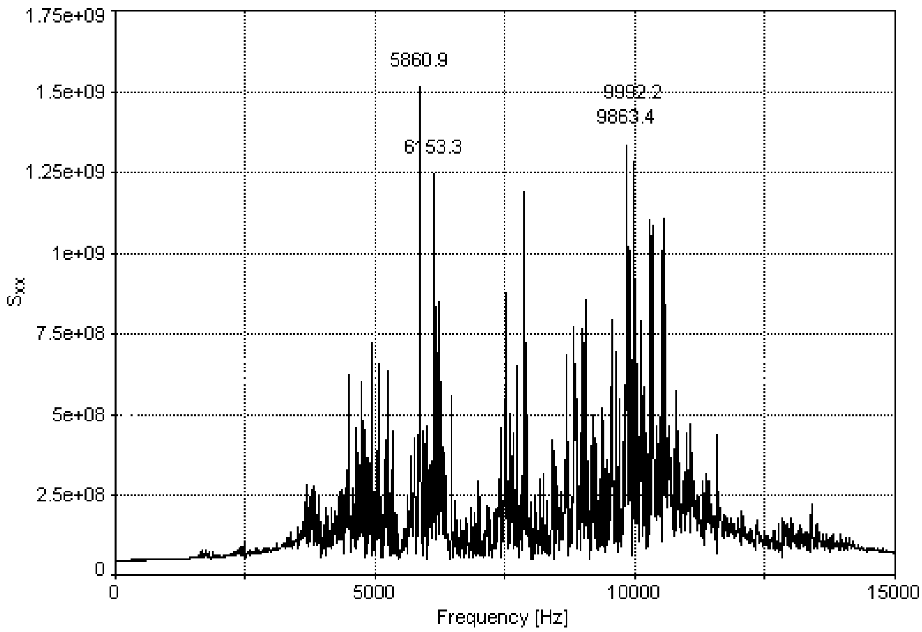

Fig. 6. – The power spectrum from the derived signal for $\theta = -60°$.

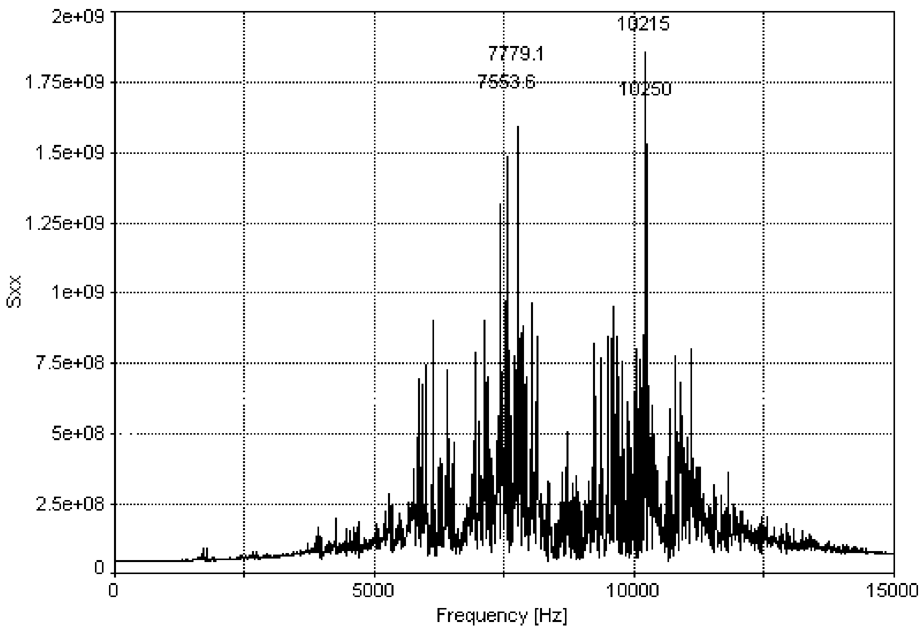

Fig. 7. – The power spectrum from the derived signal for $\theta = -30°$.

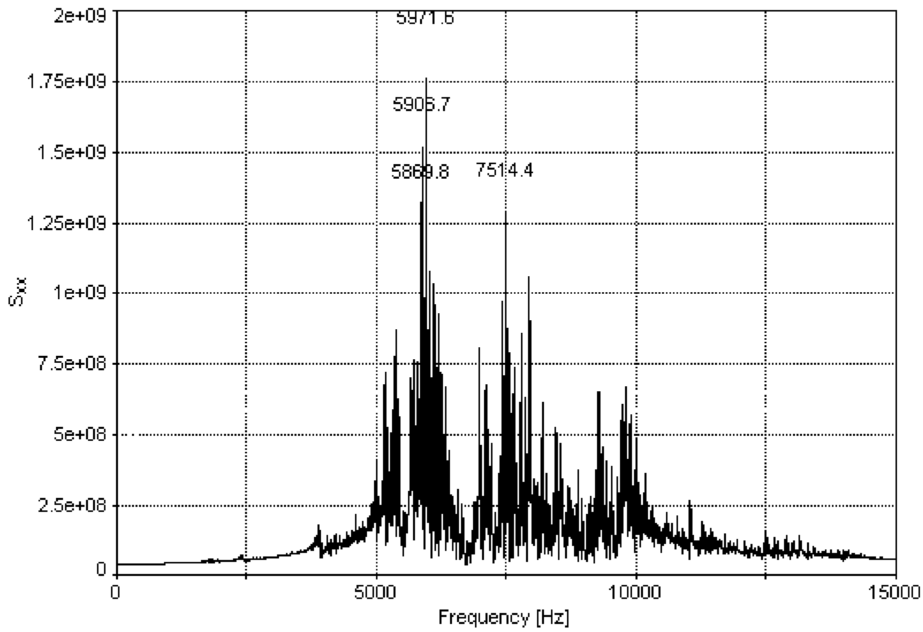

Fig. 8. – The power spectrum from the derived signal for θ = 0º.

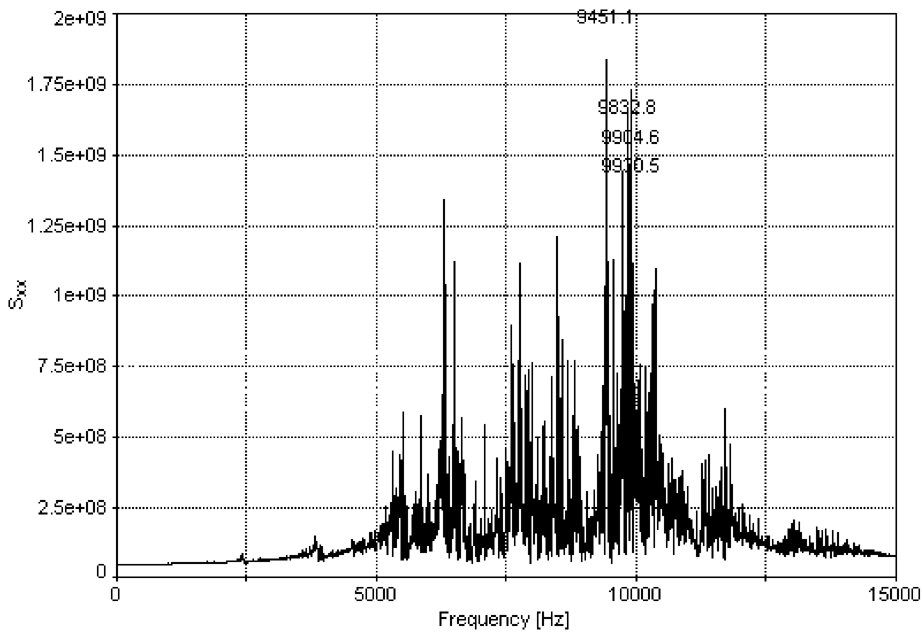

Fig. 9. – The power spectrum from the derived signal for θ = 30º.

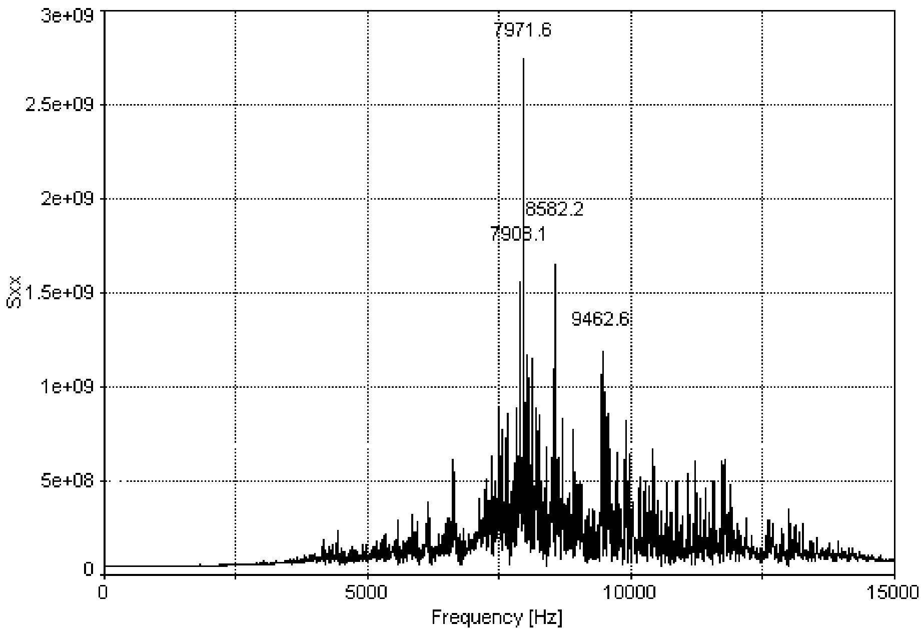

Fig. 10. – The power spectrum from the derived signal for θ = 60º.

From Figure 3 it results that for the normal position ($\theta = 0$º) the spectrum of the phoneme has minimal dispersion, the most significant components of the spectrum being localized in the domain 5–7 kHz. For negative angles $\theta < 0$º, the dispersion of the spectrum, and the spectral components for frequencies greater than 5 kHz increase at $\theta = –30$º, after which decrease the at $\theta = –60$º, and the most significant components of the spectrum move to higher frequencies in the domain 5–7 kHz.

The spectral components, localized in the domain 0–5 kHz, increase for θ near –30º after which decrease for θ near –60º. These spectral components decrease also for positive values of θ, with the increasing of θ.

From the figures it results that in the spectra from the signals derived with respect to time the high frequency part of the spectra are magnified and diminished at low frequencies. We restricted the presentation of these spectra only to the cases in Figs. 6–7. A similar dependence on θ of the spectra can be observed in the case of the signal derived as for the signals not derived, dependency discussed above. The use of the signal derived is useful for the study of the spectral components in the domain 10–15 kHz.

It results that in the spectra obtained from the signal derived the modifications induced by the mounting angle θ are more visible at high frequencies.

For practical reasons, discussed in the Introduction section, it is important to introduce a set of quantitative parameters, which describe the alterations of the speech discussed above in Figures 1–10, with the mounting angle θ.

Based on the above discussion on the changes in the spectra, we will propose a set of quantitative parameters as follows.

In order to obtain comparable measurements the amplitude spectra were normalized to the sum.

$$\frac{S(\nu_k)}{\sum_k S(\nu_k)} \rightarrow Y(\nu_k) \tag{7}$$

We will introduce three quantitative parameters in order to express the modifications appeared in spectral bands, delimited by ν_{min} and ν_{max}. After the examination of the modifications of the spectra with θ it resulted that it was useful to introduce the spectral bands defined in Table 1. The limits of the bands were established after the examination of the spectral modifications, so that in each band a group of peaks modify with θ.

Table 1

The spectral bands of interest

The band	B1	B2	B3	B4	B5	B6
ν_{min} [kHz]	0.0	2.5	5.0	7.5	10.0	12.5
ν_{max}[kHz]	2.5	5.0	7.5	10.0	12.5	15.0

These parameters will be the sum of amplitudes of the spectral components, i.e. the power of the band:

$$A = \sum_{\upsilon_k} Y(\nu_k)\,, \tag{8}$$

where $\nu_{min} \le \nu_k \le \nu_{max}$.

The power of a band:

$$P = \sum_{\upsilon_k} (Y(\nu_k))^2\,, \tag{9}$$

where $\nu_{min} \le \nu_k \le \nu_{max}$;

the average frequency of the band:

$$F = \sum_{\upsilon_k} \nu_k Y(\nu_k)\,, \tag{10}$$

where $\nu_{min} \le \nu_k \le \nu_{max}$;

and the dispersion from the average frequency.

$$D = \sqrt{\sum_{\upsilon_k} (\nu_k - F)^2 Y(\nu_k)} \tag{11}$$

where $\nu_{\min} \le \nu_k \le \nu_{\max}$.

Due to the fact that A and P parameters are of similar nature, both will manifest the same type of variation and it is sufficient to use only A in order to do our quantitative analysis.

Table 2

Dependence of A on θ for the signal $x(t)$

θ (º)	A					
	B1	B2	B3	B4	B5	B6
–60	0.08925	0.20034	0.20336	0.16573	0.13782	0.06948
–30	0.08229	0.09915	0.24298	0.22997	0.16149	0.07516
0	0.05771	0.11662	0.35788	0.19211	0.09666	0.06120
30	0.06206	0.09916	0.23474	0.26761	0.16783	0.07878
60	0.05900	0.11106	0.17678	0.32541	0.15174	0.08813

Table 3

Dependence of F on θ for the signal $x(t)$

θ (º)	F					
	B1	B2	B3	B4	B5	B6
–60	116.1	814.7	1233.6	1454.6	1519.2	950.0
–30	112.3	385.6	1553.1	1960.3	1781.9	1026.2
0	80.6	469.4	2169.0	1646.5	1071.5	837.0
30	88.9	387.5	1449.9	2358.9	1851.6	1076.7
60	78.3	435.5	1126.4	2789.9	1690.1	1202.1

Table 4

Dependence of D on θ for the signal $x(t)$

θ (º)	D					
	B1	B2	B3	B4	B5	B6
–60	421.32	1484.40	2198.64	2997.86	3537.83	3359.33
–30	408.64	1125.26	2411.09	3169.66	3729.15	3467.58
0	362.52	1237.37	2359.09	3053.65	3121.43	3181.43
30	384.91	1129.54	2311.12	3364.11	3772.92	3539.58
60	352.69	1183.59	2227.01	3324.36	3690.91	3698.63

Table 5

Dependence of *A* on θ for the signal *x(t)* derived

θ (º)	A					
	B1	B2	B3	B4	B5	B6
–60	0.04915	0.12209	0.16195	0.23369	0.20762	0.08671
–30	0.04651	0.07659	0.21077	0.24583	0.22000	0.08931
0	0.04622	0.09871	0.30053	0.2358	0.12157	0.07340
30	0.04679	0.07494	0.17574	0.30874	0.20960	0.09161
60	0.04745	0.07678	0.15847	0.33279	0.18350	0.10941

Table 6

Dependence of *F* on θ for the signal *x(t)* derived

θ (º)	F					
	B1	B2	B3	B4	B5	B6
–60	65.89	491.34	996.782	2078.75	2276.18	1184.11
–30	60.84	298.63	1362.19	2124.69	2421.57	1216.98
0	61.47	395.17	1843.62	2039.48	1343.35	1002.10
30	61.15	292.35	1098.05	2749.71	2311.90	1251.84
60	61.48	298.86	1021.48	2872.31	2048.36	1489.26

Table 7

Dependence of *D* on θ for the signal *x(t)* derived

θ (º)	D					
	B1	B2	B3	B4	B5	B6
–60	325.44	1257.48	2093.39	3316.44	3970.26	3678.24
–30	310.63	1015.59	2365.08	3256.49	4040.46	3714.84
0	314.38	1155.54	2380.87	3232.84	3393.87	3432.86
30	312.27	1007.17	2176.99	3448.50	4005.08	3763.19
60	311.91	1014.86	2178.81	3348.71	3915.86	4016.67

The dependence of the parameters *A, F, D* on mounting angle θ is given in Tables 2–4 for the signal not derived, and in Tables 5–7 for the signal derived with respect to time.

From the Tables 2–4 it results that the parameters *A, F, D* present a variation, similar to the above-discussed variation of spectra. The parameters present a typical value for θ = 0º. For example *A*, corresponding to the *B6* band, has a variation presented in Figure 11.

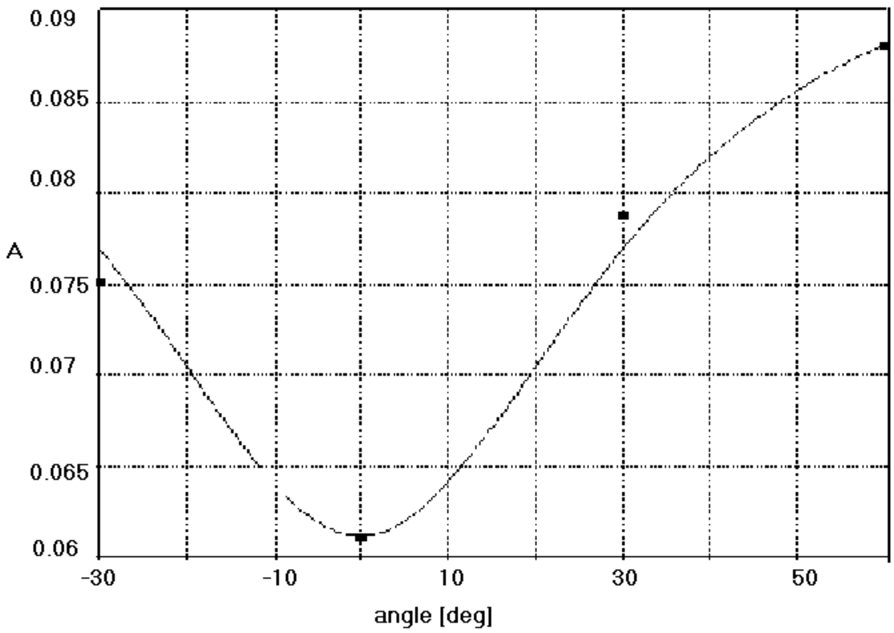

Fig. 11. – Dependence of A on θ, for the $B6$ band, from Table 2.

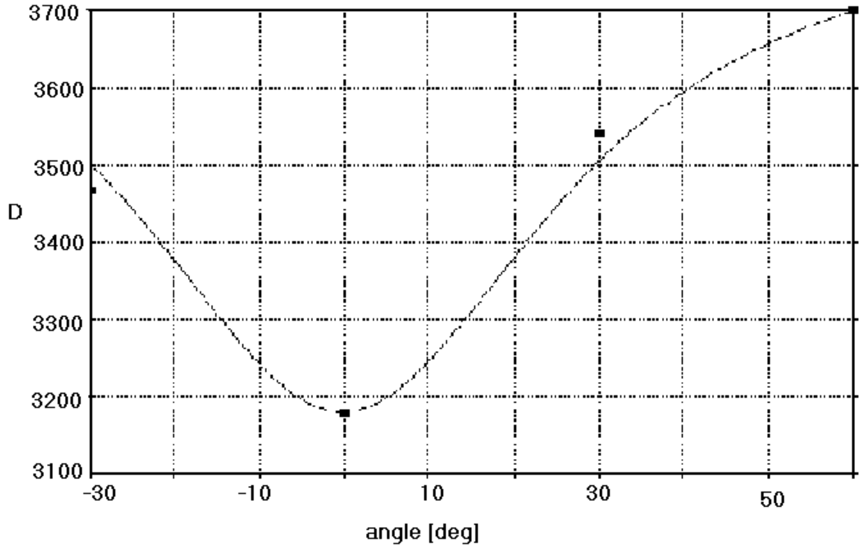

Fig. 12. – Dependence of D on θ, for the $B6$ band, from Table 2.

This kind of variation was fit by the function:

$$A = \frac{a + c\theta^2}{1 + b\theta^2}, \tag{12}$$

where: a = 0.0612, b = 0.000899, c = 8.67 10^{-5}, for a correlation quotient r = 0.9824695.

A similar type of variation, i.e. (12), presents D for $B1$, with a = 3181.43, B = 0.001098, and c = 4.2068, for r = 0.981538.

The fitting curves for all parameters and all frequency bands can be established.

From Tables 2–4 it results that the parameters *A, F, D* take a *local* minimal value, for –30º < θ< 30º, for the bands *B1, B4, B5, B6*. It also results that it is not of interest to study the variation of the *A, F, D* parameters for angles outside of this domain. For a next more detailed study, results that the analysis can be restricted in the –30º < θ< 30º domain, for a number a series of intermediate positions, so that the dependency of the parameters *A, F, D* on θ can be fit.

Similar results can be established for the parameters obtained from the power spectra of the signal derived with respect to time. From Figures 11–12 it results that the parameters *A* and *F* have the same kind of variation.

From Tables 5–7, it results that the parameters *A, F, D* take a *local* minimal value, for –30º < θ< 30º, for the bands *B4, B5,* and *B6*.

In a previous paper it was established that the speech sounds have reproducible spectral components [8] and can be used for speaker identification, including the forensic purposes. For this reason the statistical investigation of the values of the parameters was omitted in this preliminary study, our study being restricted to a single patient.

CONCLUSIONS

For a due frequency band of the /s/ phoneme investigated here, it was established that the quantitative parameters: the power of the spectral band (*A*), the average frequency of the band (*F*), and the dispersion from the average frequency (*D*) have typical values in the case of unaltered speech sound production, i.e. for good quality prosthetic mounting.

These parameters can be used as quantitative indicators of the correct position of the tooth (take a *local* minimal value) only in the *B1, B4, B5, B6* bands (i.e. 0–2.5 kHz, 7.5–10 kHz, 10–12.5 kHz and 12.5–15 kHz), as can be seen in the Tables 2–4.

It also results that the same parameters – *A, F, D* – in the *B1, B4, B5, B6* bands can be used as quantitative indicators of the correct position of the tooth, for the signal derived with respect to time (Tables 5–7).

The parameters above defined can also be used in the speech identification [8].

It must be noticed that it is possible to define other parameters which can be used as quantitative indicators of the correct position of the tooth, as the fractal Hurst exponent, entropy of the signal, introduced and used in speaker identification techniques. In a next paper we intend to illustrate that it is possible to use the Hurst exponent, entropy of the signal as indicators of the correct position of the tooth.

Acknowledgements. The authors are indebted to acknowledge Professors D. Bratu and M. Românu, for encouraging the research in this field and for their helpful opinions.

REFERENCES

1. AKAY, M., *Biomedical signal processing*, Academic Press, N. Y., 1994.
2. BLESSER, B., Speech perception under conditions of spectral transformation: I. Phonetic characteristics, *J. Speech Hear. Res.*, 1972, **15**, 5–41.
3. COHEN, L., *Time frequency analysis*, Prentice Hall Englewood Cliffs, NJ, 1995.
4. ERLEBACHER, G. *et al*., *Wavelets: theory and applications*., Oxford Univ. Press, NY, 1996.
5. JINDRA, P., M. EBER, J. PEŠÁK, The spectral analysis of sylabes in patient using dentures, *Biomed. Papers*, 2002, **146**, 91–94.
6. ÖĞÜTCEN-TOLLER, M., Sound analysis of temporomandibular joint internal derangements with phonographic recordings, *Journ. of Prosth. Dent.*, 2003, **89**, 311–318.
7. PAPOULIS, A., *The Fourier integral and its applications*, J. Wiley and Sons, New York, 1965.
8. PETRY, A., D.A.C. BARONE, Speaker identification using dynamical features, *Chaos, solitons and fractals,* 2002, **13**, 221–231.
9. RITCHIE, G. M., Y. T. ARIFFIN, Sonographic analysis of speech sounds with varying positions of the upper anterior teeth, *J. Dent.*, 1982, **10**, 17–27.
10. RUNTE, C., M. LAWERINO, D. DIERKESEN, F. BOLLMAN, E. SEIFERT, The influence of maxillary central incisor position in complete dentures on /s/ sound production, *Journ. of Prosth. Dent*., 2001, **85**, 405–415.
11. SHANNON, R.V., F.-G. ZENG, J. WYGONSKI, Speech recognition with altered spectral distribution of envelope cues, *J. Acoust. Soc. Am.,* 1989, **104,** 1467–2477.
12. THRANE, N., *Hilbert transform*, Technical Review, Nr. 3, Bruel & Kaer, Naerum, 1984.
13. TITZE, I.R., *Principles of voice production*, Prentice Hall, New Jersey, 1994.
14. TOMPKINS W. S., *Biomedical digital signal processing*, Prentice Hall, New Jersey, 1993.