Journal of ELECTRICAL ENGINEERING, VOL. 63, NO. 5, 2012, 296–302

AN EXPERIMENT WITH SPECTRAL ANALYSIS OF EMOTIONAL SPEECH AFFECTED BY ORTHODONTIC APPLIANCES

Jiří Přibil * —**Anna Přibilová** ** — **Daniela Ďuračková** **

The contribution describes the effect of the fixed and removable orthodontic appliances on spectral properties of emotional speech. Spectral changes were analyzed and evaluated by spectrograms and mean Welch's periodograms. This alternative approach to the standard listening test enables to obtain objective comparison based on statistical analysis by ANOVA and hypothesis tests. Obtained results of analysis performed on short sentences of a female speaker in four emotional states (joyous, sad, angry, and neutral) show that, first of all, the removable orthodontic appliance affects the spectrograms of produced speech.

K e y w o r d s: emotional speech, spectrogram, periodogram, statistical analysis

1 INTRODUCTION

At present, orthodontic appliances of various kinds are widely used by children as well as adults. Many of them cause problems with articulation and speech intelligibility. It is valid mostly for lingual appliances [1, 2]. Similar effect had been studied also for dental prostheses [3–7]. However, since articulation together with phonation and respiration is affected by physiological changes accompanying emotional arousal of a speaker [8], it might be expected that different emotions manifest differently in speech uttered while wearing the orthodontic appliances. Our present work is aimed at investigation of influence of the upper removable plate and the lower conventional fixed orthodontic brackets and their combination on emotional speech representing anger, joy, sadness, and a neutral emotional state for comparison.

2 SUBJECT AND METHOD

The effect of the fixed and removable orthodontic appliances on spectral properties of emotional speech should be evaluated by a listening test as a subjective comparison criterion. However, a problem with repeating of this test in a short time interval together with often non-representative results motivated us to find another method for evaluation. Our alternative evaluation approach was visual comparison and matching of displayed spectrograms. This method works in time/frequency domain and we can compare speech corresponding to a short sentence or an isolated word by this approach. Disadvantage of subjectivity of this method can be eliminated by spectrogram classification with the help of statistical parameters analysis.

2.1 Short Description of Orthodontic Appliances

An orthodontic appliance is a mechanism for application of a pressure to the teeth and their supporting tissues to produce changes in the relationship of the teeth and/or the related osseous structures [9]. There are two large categories of these appliances: fixed and removable. The appliances can be active or passive – some of them actively move the teeth, while others, such as retainers, are designed to keep the teeth where they are [10]. The orthodontic appliances are custom-designed and built for the individual patient. Typical types of the fixed orthodontic appliances are braces (from stainless steel, alloys of nickel, titanium, copper, and cobalt), in which small metal brackets are bonded to the centre of the teeth. A metal wire is then run horizontally through the brackets to connect them — see Fig. 1a). Removable orthodontic appliances are devices that can easily be taken off by the patient. The removable appliance consists of an active element or elements (*ie* metal wires and/or screws), which exert orthodontic forces on the teeth, and retentive elements (*ie* clasps), which help to retain the appliance in the mouth. A plastic plate holds these two sets of elements together — see Fig. 1b).

2.2 Spectrogram calculation, displaying, and comparison

In general, the analyzed speech signals (isolated words or short sentences) used in a comparison experiment can have different time duration. Therefore, time domain normalization must be carried out before the spectrogram calculation. For this time duration normalization the linear or non-linear time scale mapping function [13], or dynamic time warping (DTW) algorithm can be used [14] — see the block diagram of the spectrogram calculation in Fig. 2.

* Institute of Measurement Science, Slovak Academy of Sciences, Bratislava, Slovakia, Jiri.Pribil@savba.sk; ** Institute of Electronics and Photonics, Faculty of Electrical Engineering and Information Technology, Slovak University of Technology, Bratislava, Slovakia, {Anna.Pribilova; Daniela.Durackova}@stuba.sk

DOI: 10.2478/v10187-012-0043-5, ISSN 1335-3632

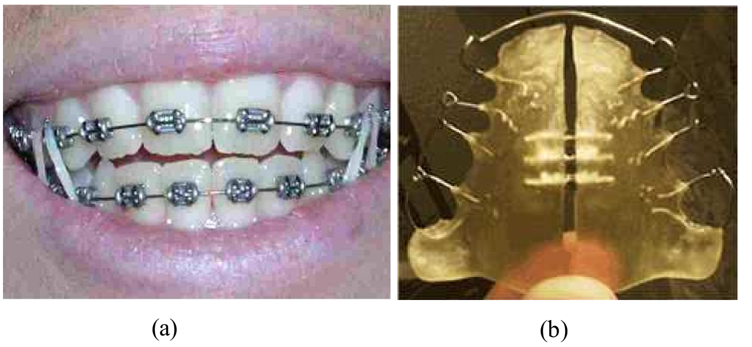

Fig. 1. Photo of upper and lower fixed brackets (a) and an upper removable plate (b)

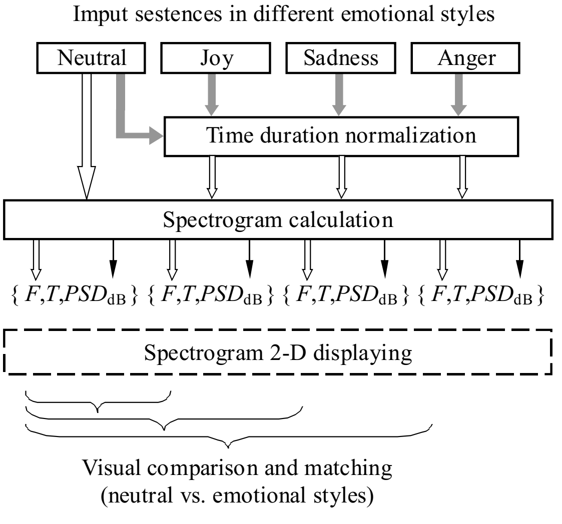

Fig. 2. Block diagram of spectrogram calculation and comparison of sentences in different emotional speaking styles

However, results of visual comparison of the whole spectrograms (representing speech signals of short sentences or isolated words) depend much on a person that makes this matching. For objective comparison and matching the statistical approach based on analysis of variance (ANOVA) [15] and hypothesis tests can be applied [16]. To obtain serious matching results it is necessary to select the region of interest (ROI) of the input signal. From the chosen ROI area the mean periodogram calculated by the Welch method [17] can be determined. The periodogram for an input signal of a sample sequence $[x_1, \dots, x_n]$ weighted by a window $[w_1, \dots, w_n]$ is defined as

$$S\left(e^{j\omega}\right) = \frac{\frac{1}{n}\left|\sum_{k=1}^{n} w_k x_k e^{-j\omega k}\right|^2}{\frac{1}{n}\sum_{k=1}^{n}|w_k|^2} . \tag{1}$$

This expression represents an estimate of the power spectral density (PSD) of the input speech signal. In our case, periodogram uses an $N_{\rm FFT}$-point FFT to compute the power spectral density as $S\left(e^{j\omega}\right)/f_s$ where f_s is a sampling frequency. The resulting Welch's periodogram in [dB] can be used for subsequent comparison. From these mean periodograms the first three format frequencies F_1, F_2, and F_3 can be also determined as the first three local maxima where its gradient changes from positive to negative — see Fig. 3. For exact numerical comparison (objective matching method) it is possible to calculate the RMS spectral distance $D_{\rm RMS}$ between different periodograms corresponding to the basic sentence in a neutral style and other three sentences in emotional styles. This method enabling objective comparison of the same ROI area ($\Delta T_N = \Delta T_J = \Delta T_S = \Delta T_A$) of four spectrograms after speech signal time normalization based on the neutral style is illustrated by the block diagram in Fig. 4.

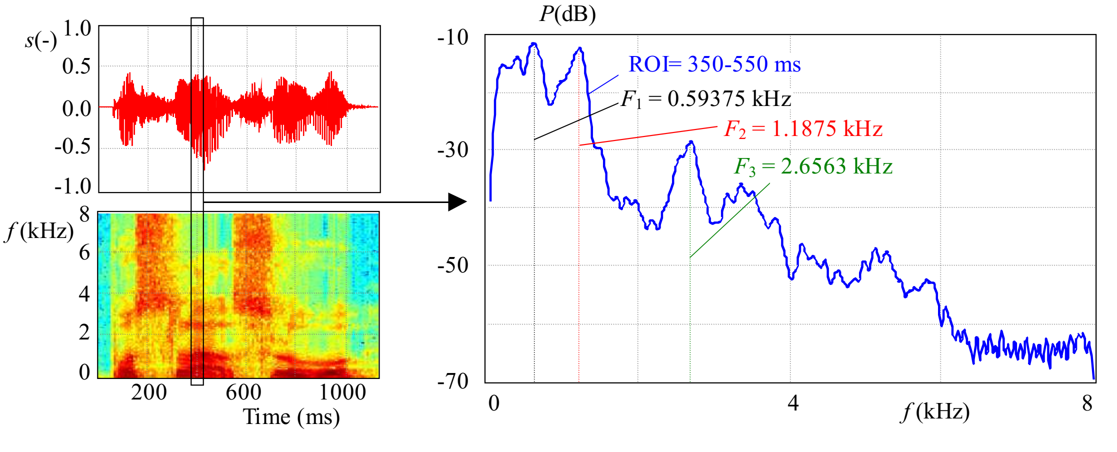

Fig. 3. Example of the mean Welch's periodogram calculation: selected ROI corresponding to long vowel "a" with determined F_1, F_2, and F_3 formant frequencies

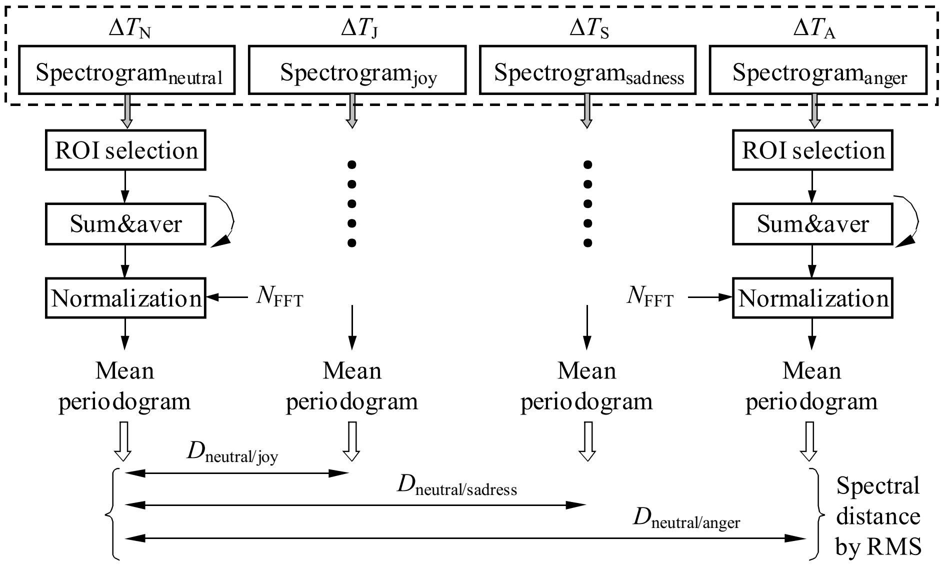

Fig. 4. Block diagram of the averaged Welch's periodograms calculation and the spectral distance determination from the selected ROI areas between neutral and emotional styles

3 MATERIAL, EXPERIMENTS, AND RESULTS

Speech material for analysis was recorded with the help of the Behringer professional Podcastudio USB with the dynamic cardioid microphone Ultravoice XM8500 and the mixing console Xenyx 502 connected to a personal computer through the UCA200 high-performance audio interface. The collected speech database consists of 72 short sentences (with duration from 0.5 to 1.5 seconds) spoken in four emotional states — neutral, joyous, sad, and angry in Slovak and Czech languages. For every emotional state, the sentences were uttered under three types of conditions:

1. without orthodontic appliances (NO OA),
2. with the lower fixed orthodontic brackets (LF OB),
3. with the upper removable plate and the lower fixed orthodontic brackets (UP LB).

It means that there exist always 12 variants for every sentence included in the database, uttered by a female speaker with F0 $\approx$ 200 Hz, originally recorded at 32 kHz, and resampled to 16 kHz. The parameters settings for spectrogram was chosen in correspondence with the speaker's mean F0 as: window length $L_W = 180$, window overlapping $L_O = 40$ (in [samples]), and $N_{\rm FFT}$ =1024 for the used sampling frequency $f_s = 16$ kHz.

Full comparison between all sentences represents a 2-D task which is not easy to be solved. In our experiment, we have found an easy way: at first, comparison between the neutral and emotional styles was carried out for the whole sentences. Then, comparison between sentences obtained with different configuration of orthodontic appliances selectively for every emotion was realized. This method reduces comparison task to a 1-D matching process and enables to do detailed analysis of interesting regions. As follows from visual comparison, the orthodontic appliances bring about the most significant spectral changes for voiced speech (see Fig. 5). For this purpose the second database consisting of the selected basic vowels "a", "e", "i", "o", "u" and the voiced consonants "m", "n" was created.

Evaluation of spectral analysis of emotional speech affected by orthodontic appliances was carried out in four steps:

1. visual comparison of calculated spectrograms (the whole sentences from the main speech database),
2. statistical comparison of the whole spectrograms by ANOVA and hypothesis tests,
3. visual comparison of the calculated Welch's periodograms (for selected ROI from the database of the vowels and the voiced consonants),
4. numerical matching of results from the calculated RMS spectral distances between corresponding periodograms.

The spectrograms corresponding to speech uttered in different emotional styles with different configurations of orthodontic appliances were evaluated by one-way ANOVA in the second step of our experiment. Then series of t tests and visualization of differences between group means were performed. Obtained results in the form of graphs are presented in Fig. 6, the numerical results of performed Ansari-Bradley hypothesis tests for 5 % significance level are summarized in Tab. 1. Mean periodograms of the analysed sound database with different configurations of orthodontic appliances with speech uttered in a neutral style are shown in Fig. 7, calculated corresponding $D_{\rm RMS}$ values are summarized in Tab. 2, Tab. 3 consists of F_1, F_2, and F_3 formant frequencies. Results of detailed analysis of the sound "e" are presented

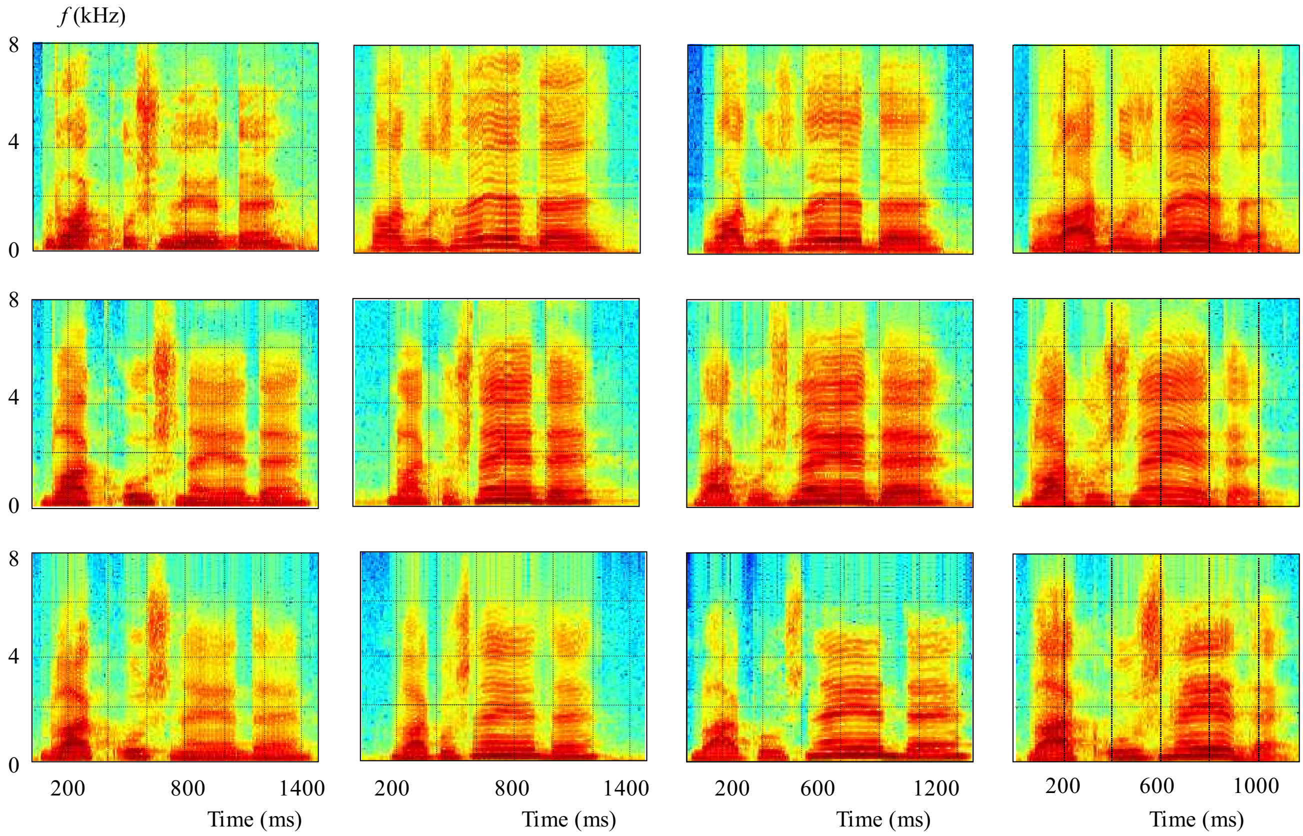

Fig. 5. Documentary set of pictures for visual comparison of spectrograms of the sentence "Vlak už nejede" (No more train leaves today — female speaker, F0 ≈ 200 Hz, $f_s = 16$ kHz) uttered in neutral (left column), joyous (left-middle column), sad (right-middle column), and angry emotional style (right column): NO OA (upper line), LF OB (middle line), and UP LB (bottom line) — speech signal time normalization was applied always in the frame of the same emotional style

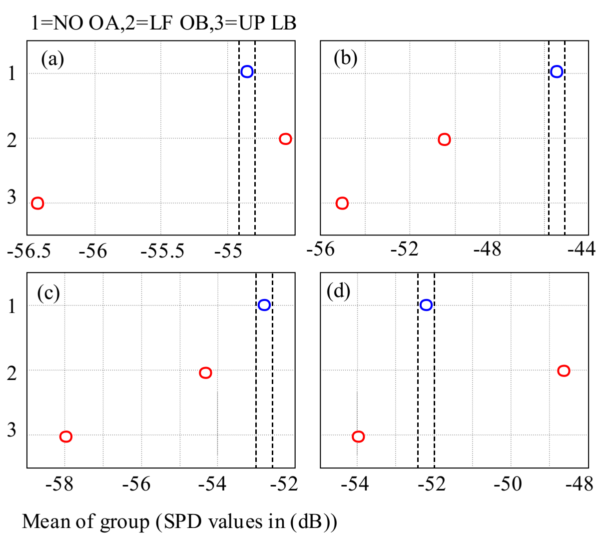

Fig. 6. Results of multiple comparison of groups applied to ANOVA results of the sets of spectrograms with different configurations of orthodontic appliances in emotional styles: neutral (a), joyous (b), sad (c), and angry (d)

by the graphs of mean periodograms in Figs. 8 and 9. Partial results of spectral distances of mean periodograms of the sound "e" are stored in Tab. 4, and spectral distances between different emotional styles are in Tabs. 5 and 6.

Table 1. Results of Ansari-Bradley hypothesis tests of the sets of spectrograms with different configurations of orthodontic appliances in neutral and emotional styles

h/p*	Neutral		Joyous	
	LF OB	UP LB	LF OB	UP LB
NO OA	$1/1.35\cdot10^{-8}$	$1/3.74\cdot10^{-19}$	$1/2.52\cdot10^{-25}$	$1/6.42\cdot10^{-79}$
LF OB	0/1	$1/2.23\cdot10^{-22}$	0/1	$1/2.23\cdot10^{-22}$
h/p*	Sad		Angry	
	LF OB	UP LB	LF OB	UP LB
NO OA	$1/1.07\cdot10^{-16}$	$1/8.56\cdot10^{-68}$	$1/3.74\cdot10^{-28}$	$1/6.38\cdot10^{-22}$
LF OB	0/1	$1/3.35\cdot10^{-44}$	0/1	$1/4.54\cdot10^{-61}$

* hypothesis/probability values for 5% significance level

Table 2. Summary results of spectral distances of mean periodograms of sounds with different configurations of orthodontic appliances uttered in a neutral style

Sound type	D_{RMS} (dB) NO OA to LF OB	D_{RMS} (dB) NO OA to UP LB	D_{RMS} (dB) LF OB to UP LB
"a"	5.844	5.272	5.228
"e"	9.346	6.599	8.469
"i"	4.375	3.482	3.900
"o"	4.257	4.001	2.759
"u"	5.343	3.009	4.719
"m"	4.324	5.250	3.235
"n"	3.843	5.133	3.059

Table 3. Mean formant frequencies in (Hz) of sounds with different configurations of orthodontic appliances uttered in a neutral style

Sound	NO OA			LF OB			UP LB		
type	F_1	F_2	F_3	F_1	F_2	F_3	F_1	F_2	F_3
"a"	790	1346	2773	769	1285	2846	791	1215	2732
"e"	551	1840	2800	544	1941	2799	522	1823	2754
"i"	475	1239	2713	485	1254	2845	483	1152	2879
"o"	525	982	2705	536	989	2741	541	996	2736
"u"	412	951	2691	383	945	2637	375	1058	2738
"m"	393	1617	2852	379	1598	2796	387	1609	2804
"n"	379	1082	2467	357	1134	2402	363	1457	2443

Table 4. Partial results of spectral distances of mean periodograms of the sound "e" with different configurations of orthodontic appliances uttered in different emotional styles

Style	D_{RMS} (dB) NO OA to LF OB	D_{RMS} (dB) NO OA to UP LB	D_{RMS} (dB) LF OB to UP LB
neutral	9.345	6.599	8.469
joyous	14.076	7.185	9.254
sad	6.706	4.327	7.024
angry	9.454	7.632	11.012

Table 5. Detailed results of spectral distances of mean periodograms of the sound "e" uttered in different emotional styles with different configurations of orthodontic appliances

Appliance	D_{RMS} *(dB) J to S	D_{RMS} *(dB) J to A	D_{RMS} *(dB) S to A
NO OA	12.283	7.683	7.139
LF OB	7.992	2.674	9.249
UP LB	6.808	3.569	5.267

* calculated between: joyous (J), sad (S), and angry (A) styles

Table 6. Detailed results of spectral distances of mean periodograms of the sound "e" uttered in different emotional styles with different configurations of orthodontic appliances

Appliance	D_{RMS} *(dB) N to J	D_{RMS} *(dB) N to S	D_{RMS} *(dB) N to A
NO OA	13.128	4.032	7.342
LF OB	8.555	2.196	9.831
UP LB	8.105	3.471	7.059

* calculated between: neutral (N), sad (S), and angry (A) styles

4 DISCUSSION AND CONCLUSION

The statistical approach based on ANOVA analysis with multiple comparison of groups was applied to evaluation of the whole spectrograms (see example in Fig. 5). For objective statistical comparison and matching of calculated spectrograms, the hypothesis probability Ansari-Bradley test (variance) was applied. If is the test of the hypothesis that two independent samples come from the same distribution against the alternative that they come from distributions having the same median and shape but different variances. This test also returns the probability of observing the given result, or one more extreme by chance if the null hypothesis is true. Small values of this probability cast doubt on the validity of the null hypothesis. From comparison of the whole spectrograms next follows, that use of the orthodontic appliance brings about the most significant spectral changes for voiced speech. Therefore the extended analysis of sounds based on Welch's periodograms was subsequently performed.

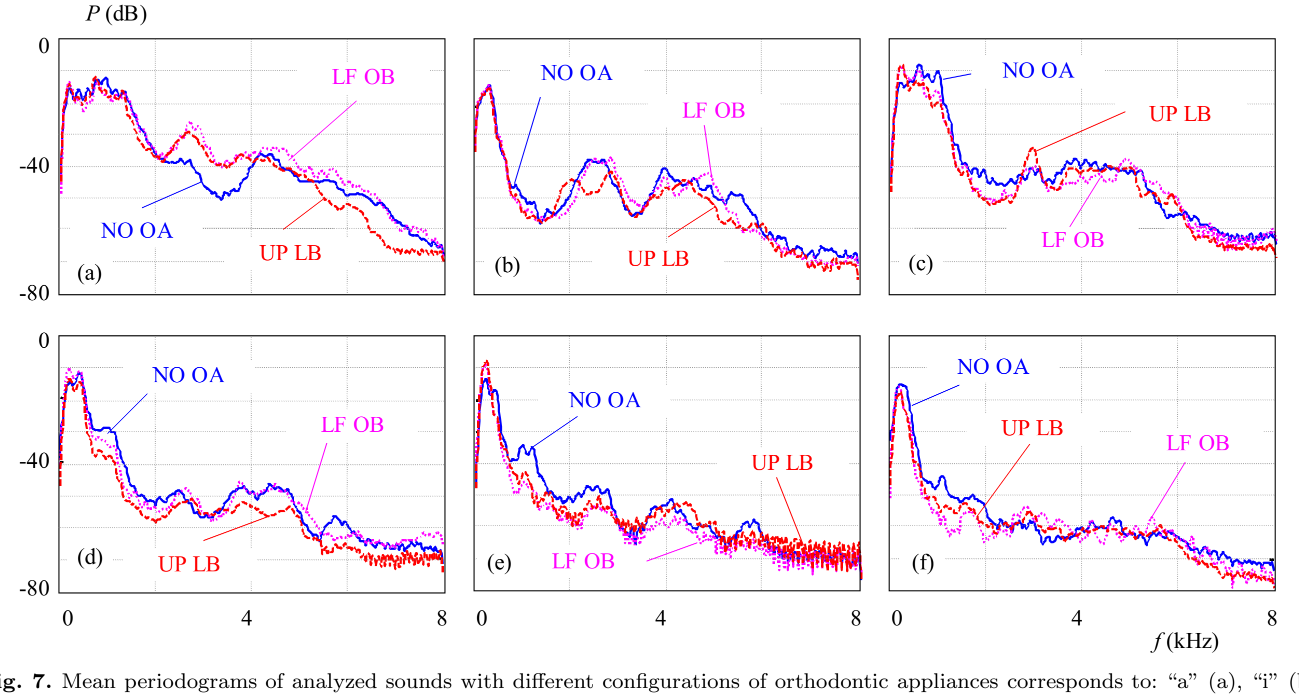

Fig. 7. Mean periodograms of analyzed sounds with different configurations of orthodontic appliances corresponds to: "a" (a), "i" (b), "o" (c), "u" (d), "m" (e), and "n" (f) — uttered in a neutral style

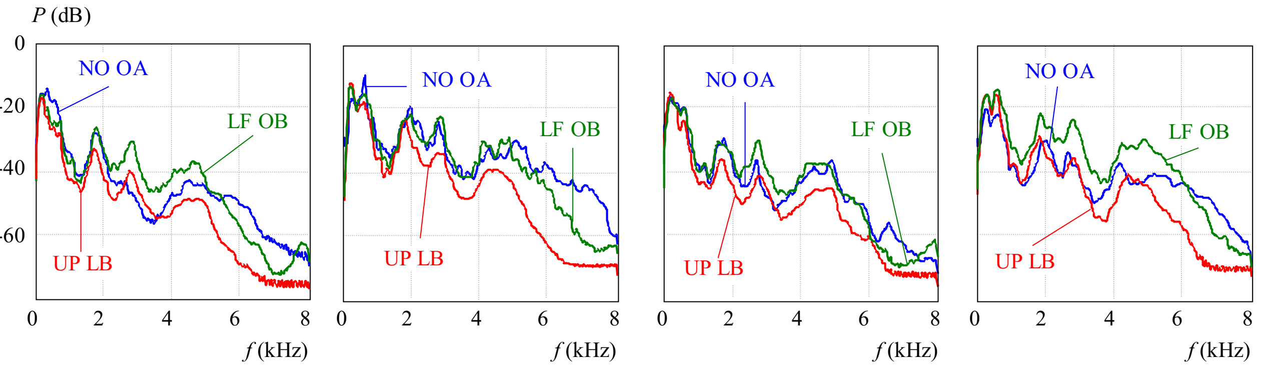

Fig. 8. Mean periodograms of the sound "e" with different configurations of orthodontic appliances in emotional styles: neutral (a), joyous (b), sad (c), and angry (d)

Statistical results of ANOVA analysis together with the results of hypothesis tests in Tab. 1 confirms the effect of the orthodontic appliances on speech uttered in all four emotional styles. According to the obtained results of realized analysis of selected voiced sounds uttered in the neutral style (shown in Tab. 2) based on the mean periodograms calculated by the Welch method, the vowel "e" was chosen for the next detailed analysis. Calculation of the spectral distances of the sound "e" picked from the utterances with different types of the orthodontic appliances and uttered with different emotional colouration was done. As follows from the partial results in Tab. 4 and the detailed results in Tabs. 5 and 6 the most significant changes are observed for less expressive types of speech — neutral and sad emotions. Application of the lower fixed orthodontic brackets alone is accompanied by fluctuations and local rise of the magnitude spectrum particularly in the area of higher frequencies (higher than 5 kHz). The

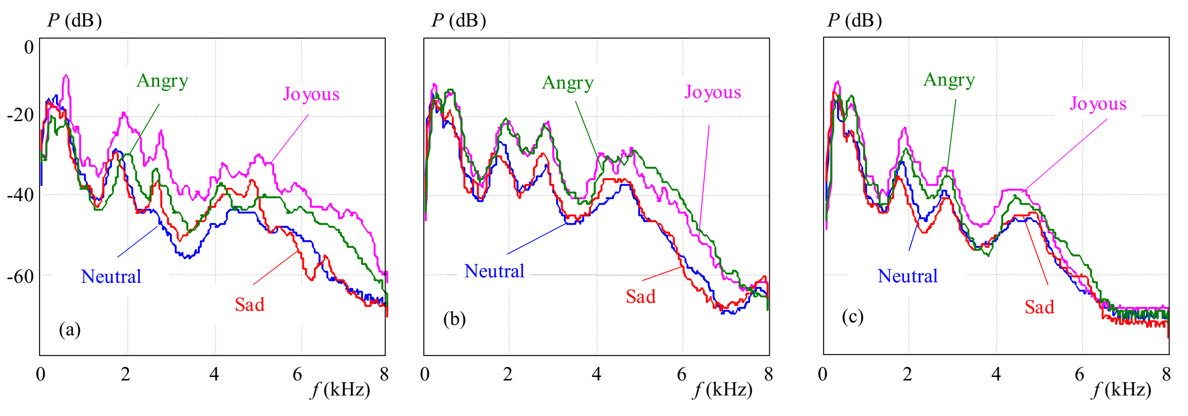

Fig. 9. Mean periodograms of the sound "e" uttered in different emotional styles with different configurations of orthodontic appliances: NO OA (a), LF OB (b), and UP LB (c)

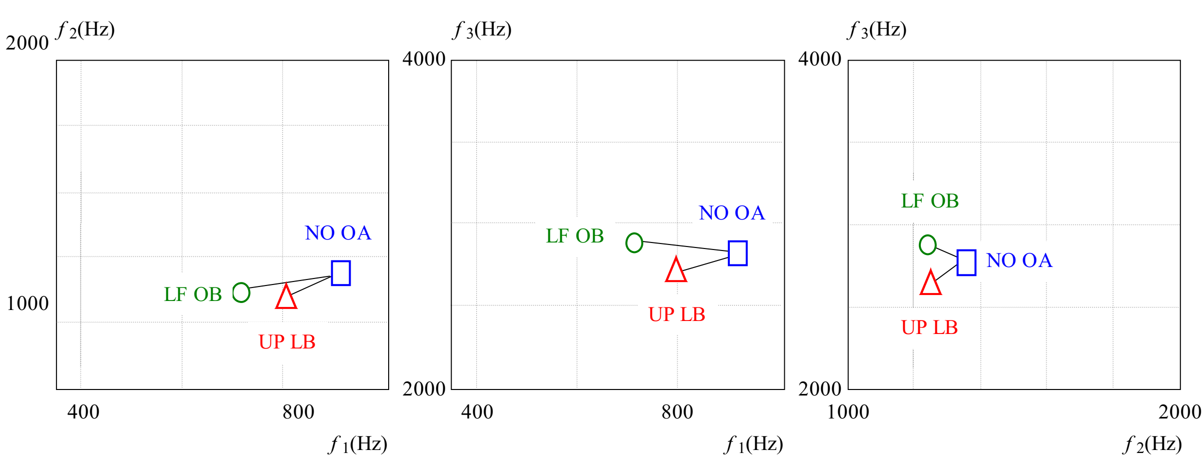

Fig. 10. Diagrams of F_1/F_2 (a), F_1/F_3 (b), F_2/F_3 (c) mutual frequencies of the sound "e" with different configurations of orthodontic appliances uttered in a neutral style

overall orthodontic appliances effect (the upper removable plate and the lower fixed orthodontic brackets together) is manifested by suppression of high frequencies, even higher than 4 kHz — see Figs. 8 and 9. These spectral changes appear in all of the analyzed sounds, in the case of the consonants "m" and "n" the differences (spectral distances) were lower due to smaller absolute amplitudes of the speech signal of the vowels. On the other hand, the orthodontic appliances have small influence to the first three formant positions — see results in Tab. 3; it can be also seen in diagrams of F_1/F_2, F_1/F_3, and F_2/F_3 mutual frequencies of the sound "e" in Fig. 10.

Finally, a question arises how a speaker wearing the orthodontic appliances can adjust his / her articulation after some training under abnormal conditions. It is a theme for the next study — similar to the influence of "foreign objects" in the mouth, investigated as the well-known bite-block experiments [18, 19].

Acknowledgment

The work has been supported by the Grant Agency of the Slovak Academy of Sciences (VEGA 2/0090/11) and the Ministry of Education of the Slovak Republic (VEGA 1/0987/12).

References

[1] HOHOFF, A.—SEIFERT, E.—FILLION, D.—STAMM, T.—HEINECKE, A.—EHMER, U. : Speech Performance in Lingual Orthodontic Patients Measured by Sonagraphy and Auditive Analysis, American Journal of Orthodontics and Dentofacial Orthopedics **123** (2003), 146-152.

[2] HOHOFF, A.,—STAMM, T.—GODER, G.—SAUERLAND, C.—EHMER, U.—SEIFERT, E. : Comparison of 3 Bonded Lingual Appliances by Auditive Analysis and Subjective Assessment, American Journal of Orthodontics and Dentofacial Orthopedics **124** (2003), 737-745.

[3] SEIFERT, E.—RUNTE, C.—RIEBANDT, M.—LAMPRECHT -DINNESEN, A.—BOLLMANN, F. : Can Dental Prostheses Influence Vocal Parameters?, The Journal of Prosthetic Dentistry **81** (1999), 579-585.

[4] RUNTE, C.—LAWERINO, C.—DIRKSEN, D.—BOLLMAN, F.—LAMPRECHT-DINNESEN, A.—SEIFERT, E. : The Influence of Maxillary Central Incisor Position in Complete Dentures on /s/ Sound Production, The Journal of Prosthetic Dentistry **85** (2001), 485-495.

[5] DRĂGĂNESCU, G. E.—SINESCU, C.—DODENCIU, D. : Quantitative Measurement of Speech Sound Distortions Due to Inadequate Dental Mounting, Computer Methods in Biomechanics and Biomedical Engineering No. 9 (2006), 25-34.

[6] SINESCU, C.—DRĂGĂNESCU, G. E.—DODENCIU, D.—BERETEU, L.—NEGRUŢIU, M.—ROMÂNU, M. : Quantitative Parameters Which Describe Speech Sound Distortions Due to Inadequate Dental Mounting, Physics A **387** (2008), 1205-1217.

[7] KONG, H. J.—HANSEN, C. A. : Customizing Palatal Contours of a Denture to Improve Speech Intelligibility, The Journal of Prosthetic Dentistry **99** (2008), 243-248.

[8] SCHERER, K. R. : Vocal Communication of Emotion: A Review of Research Paradigms, Speech Communication **40** (2003), 227-256.

[9] HILTON, L. : Orthodontic Appliances Information on Healthline, Gale Encyclopedia of Nursing and Allied Health, The Gale Group Inc., Gale, Detroit, 2002.

[10] BEARN, D. R. : Bonded Orthodontic Retainers: A Review, American Journal of Orthodontics and Dentofacial Orthopedics **108** (1995), 207-213.

[11] SLIFKA, J.—ANDERSON, T. R. : Speaker Modification with LPC Pole Analysis, The IEEE International Conference on Acoustics, Speech, and Signal Processing, Detroit, 1995, pp. 644-647.

[12] HANSON, H. M.—STEVENS, L. N.—KUO, H.-K. J.—CHEN, M. Y., SLIFKA, J. : Towards Models of Phonation, Journal of Phonetics No. 29 (2001), 451-480.

[13] PŘIBIL, J.—PŘIBILOVÁ, A. : Emotional Style Conversion in the TTS System with Cepstral Description, In: A. Esposito, M. Faundez-Zanuy, E. Keller, M. Marinaro (eds.) Verbal and Nonverbal Communication Behaviours. LNCS (LNAI), vol. 4775, Springer, Heidelberg, 2007, pp. 65-73.

[14] BENESTY, J.—SONDHI, M. M.—HUANG, Y. (eds.) : Springer Handbook of Speech Processing, Springer-Verlag, Berlin Heidelberg, 2008.

[15] RENCHER, A. C.—SCHAALJE, G. B. : Linear Models in Statistics, Second Edition, John Wiley & Sons, 2008.

[16] MIZUSHIMA, T. : Multisample Tests for Scale Based on Kernel Density Estimation, Statistics & Probability Letters **49** (2000), 81-91.

[17] OPPENHEIM, A. V.—SCHAFER, R. W.—BUCK, J. R. : Discrete-Time Signal Processing, Second Edition, Prentice Hall, 1999.

[18] FOWLER, C. A.—TURVEY, M. T. : Immediate Compensation in Bite-Block Speech, Phonetica **37,** (1980), 306–326.

[19] LANE, H. *et al* : Effects of Bite Blocks and Hearing Status on Vowel Production, Journal of Acoustical Society of America **118** No. 3 (Sep r2005), 1636-1646.

Received 9 January 2012

Jiří Přibil (Ing, PhD), born in 1962 in Prague, Czechoslovakia. He received his MS degree in computer engineering in 1991 and his PhD degree in applied electronics in 1998 from the Czech Technical University in Prague. At present he is a scientific worker at the Department of Imaging Methods Institute of Measurement Science, Slovak Academy of Sciences in Bratislava. His research interests are signal processors and their applications, signal analysis and synthesis, and text-to-speech systems.

Anna Přibilová (Ing, PhD) received her MSc and PhD degrees from the Faculty of Electrical Engineering and Information Technology, Slovak University of Technology (FEEIT SUT) in 1985 and 2002, respectively. For six years she had been with Chirana Research Centre for Medical Equipment as a research assistant. Since 1992 she has been working as a university teacher at the Radioelectronics Department and since 2011 at the Institute of Electronics and Photonics of the FEEIT SUT in Bratislava. The main field of her research and teaching activities is audio and speech signal processing.

Daniela Ďuračková (Prof, Ing, PhD) received her MSc and PhD from Faculty of Electrical Engineering and Information Technology , Slovak University of Technology in 1974 and 1981 respectively. Since 1991 she has been an associate Professor and since 2005 Professor at the Microelectronics Department of the Faculty of Electrical Engineering and Information Technology, Slovak University of Technology in Bratislava. The main field of her research and teaching activities has moved from semi-conductor devices towards the design of analog and digital ASICs and neural network implementation on-chip.