# Comparison of Formant Features of Male and Female Emotional Speech in Czech and Slovak

J. Pribil[1,2], A. Pribilova[3], J. Matousek[1]
[1]*Department of Cybernetics, Faculty of Applied Sciences, University of West Bohemia, Univerzitní 8, CZ-306 14 Plzeň, Czech Republic*
[2]*Institute of Measurement Science, Slovak Academy of Sciences, Dúbravská cesta 9, SK-841 04 Bratislava, Slovakia*
[3]*Institute of Electronics and Photonics, FEI, Slovak University of Technology, Ilkovičova 3, SK-812 19 Bratislava, Slovakia*
*jiri.pribil@savba.sk*

***Abstract*—The paper describes analysis and comparison of formant features comprising the first three formant positions together with their 3 -dB bandwidths and the formant tilts. These features were determined from the smoothed spectral envelopes or directly calculated from the complex roots of the LPC polynomial. Subsequently, statistical analysis and comparison of the formant features from emotional speech representing joy, sadness, anger, and a neutral state was performed. In this experiment we use the speech material in the form of sentences uttered by male and female professional speakers in Czech and Slovak languages. For detailed analysis, the derived speech database consisting of manually selected sounds corresponding to the stationary parts of five vowels and two nasals was created. The determined formant positions and their value ranges are in correspondence with the general knowledge for male and female voices. Obtained statistical results and values of parameter ratios will be used for emotional speech conversion or they can also be applied for extension of the text-to-speech system enabling expressive speech production.**

***Index Terms*—Speech processing, spectral analysis, speech analysis, emotion recognition.**

## I. Introduction

Identification of emotions in speech depends on the chosen set of features extracted from the speech signal. These features are systematically divided into segmental and supra-segmental ones [1]. Short-term segmental features derived from speech frames with short duration are usually in relation with the speech spectrum. These include traditional features like linear predictive coefficients, line spectral frequencies, mel-frequency cepstral coefficients (MFCC), linear prediction cepstral coefficients [2], or unconventional ones like perceptual linear predictive coefficients, log frequency power coefficients, etc. [3]. Supra-segmental features comprise statistical values of parameters describing prosody by duration, fundamental frequency, and energy. This category comprises also a separate group of features constituting voice quality parameters: jitter, shimmer, glottal-to-noise excitation ratio, normalized amplitude quotient, spectral tilt, and spectral balance [2].

Emotion detection from speech is usually carried out with combination of various features among which almost invariably the MFCC are used [4]. Improvement in speech emotion recognition can be attained by using various sets of new features, e.g. harmony features [5] or semantic labels [6]. However, within the standard features the formants convey information about the speaker's vocal tract which differs not only for different speakers but it changes its shape for the same speaker in different emotional state.

This paper describes analysis and comparison of the formant features (FF) of male and female Czech and Slovak acted speech in four emotional states: joy, sadness, anger, and a neutral state. Czech and Slovak languages (belonging to the Slavonic languages) are similar but different, therefore we can use a common speech corpus to obtain spectral parameters, but on the phonetic and prosodic level the synthetic speech must be processed separately. Motivation of our work was to find out the parameters for extension of the text-to-speech (TTS) system enabling expressive speech production [7]. The obtained parameter ratios between male and female voices as well as ratios between emotional and neutral speech will be used in emotional speech transformation (conversion) method.

## II. Methods Used for Spectral Envelope Smoothing

The FF consist of the basic frequency parameters, as the first three formant positions ($F_1$, $F_2$, and $F_3$) together with their bandwidths, and the complementary parameters (usually the formant tilts) that can be calculated by several techniques [8]. In practice two basic approaches to FF determination are mostly used: the first one calculates them from the complex roots of the LPC polynomial; the second one consists in finding of the local maxima of the smoothed spectral envelope where its gradient changes from positive to negative.

The formant positions and their bandwidths are

Manuscript received May 23, 2012; accepted April 3, 2013.

The work has been supported by the Technology Agency of the Czech Republic (TA01030476), the Ministry of Education of the Slovak Republic (VEGA 1/0987/12), and Grant Agency of the Slovak Academy of Sciences (VEGA 2/0090/11).

prevailingly determined from the smoothed envelope of the voiced parts of the speech signal. To obtain the smoothed spectral envelope, the mean periodogram of the chosen region of interest (ROI) areas – voiced parts of the speech signal – can be computed by the Welch method. The periodogram for an input signal of a sample sequence $[x_1, \ldots, x_n]$ weighted by a window $[w_1, \ldots, w_n]$ is defined as

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

In our case, the periodogram uses an $N_{\mathrm{FFT}}$-point FFT to compute the power spectral density (PSD) of the input speech signal as $S(e^{j\omega})/f_s$ where $f_s$ is a sampling frequency.

The smooth spectral envelope of the speech signal can also be determined during cepstral analysis [9]. Cepstral analysis of the speech signal is performed in the following way: first, the complex spectrum using FFT algorithm is calculated from the input samples (after segmentation and weighting by a Hamming window). In the next step, the log power spectrum is computed. Application of the inverse FFT algorithm gives the symmetric real cepstrum. By limitation to the first $N_0+1$ coefficients, the truncated cepstrum represents approximation of a log spectrum envelope

$$S\left(e^{j\omega}\right)=c_0+2\sum_{n=1}^{N_0} c_n\cos\left(n\cdot\omega\right), \quad (2)$$

where the first cepstral coefficient $c_0$ corresponds to the signal energy.

Formant frequencies and bandwidths can be determined simply using an autoregressive (AR) model well known in speech processing as a linear predictive coding (LPC) model being an all-pole model of a vocal tract. The autocorrelation method uses the Levinson-Durbin recursion to compute the parameters $\{a_k\}$ describing the speech spectral envelope $S\left(e^{j\omega}\right)$

$$S\left(e^{j\omega}\right)=\frac{G}{1+\sum_{n=1}^{N_A} a_k\exp^{-j\omega n}}, \quad (3)$$

where $N_A$ is the order of the AR model.

## III. Calculation and Evaluation of Formant Features

We apply two methods for basic formant features determination:

1) Indirect – the formant positions as the first three local maxima of the smoothed spectral envelope where its gradient changes from positive to negative, corresponding bandwidths are obtained as the frequency intervals between the points of the 3 -dB decrease of the magnitude spectrum from the formant amplitudes;

*2)* Immediate – the estimation of the formant frequencies and their bandwidths directly from the complex roots of the LPC polynomial *A(z)*.

Using the Newton-Raphson or the Bairstow algorithm [10] we obtain the complex root pairs $z_n=|z_n|\ e^{j\pm\theta_n}$ corresponding to the poles of the LPC transfer function. The formant frequency $F_n$ and the 3 -dB bandwidth $B_n$ in [Hz] can be determined by

$$F_n=\frac{f_s}{2\pi}\theta_n,\ B_n=-\frac{f_s}{\pi}\ln|z_n|, \quad (4)$$

where $\theta_n$ is the angle in [rad] of the complex root.

Indirect determination of the basic FF is realized by combination of all three mentioned approaches for the spectral envelope calculation and smoothing. In the case of the LPC envelope calculated by (3) the higher order is applied; in the case of direct calculation from the roots of the LPC polynomial by (4), the lower order is applied. Correctness of the basic FF values obtained by all three indirect methods as well as by direct calculation from roots is assessed by two criteria:

1) The resulting values of 3 -dB bandwidths must be less than 500 Hz [11];

2) The found values of the first three formant positions must fall within the corresponding frequency interval depending on the voice type (male/female).

Resulting values fulfilling these conditions are used finally for next processing. The whole algorithm of the used method of the basic and complementary formant features determination is described by the block diagram in Fig. 1.

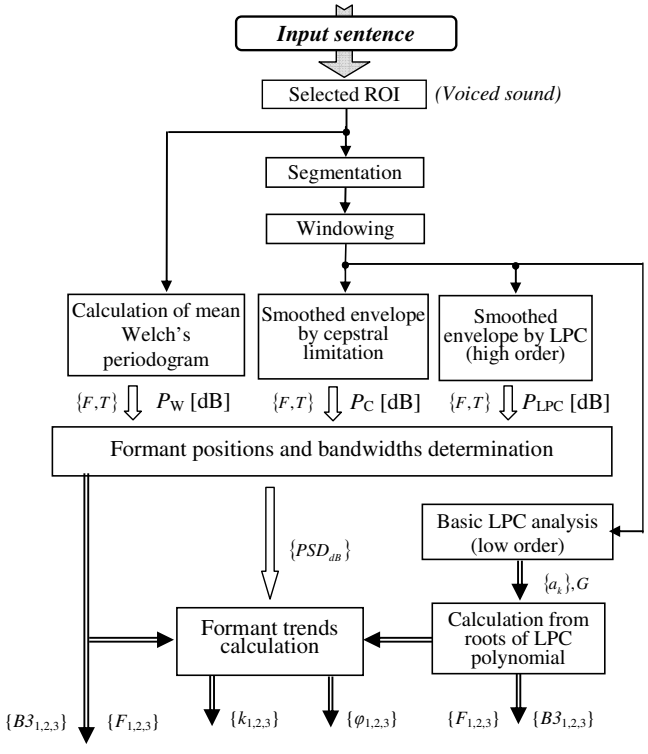


Fig. 1. Block diagram of the formant features determination.

The complementary FF can be defined as formant tilts – angles between spectrum peaks in the place of determined first three format positions (see documentary Fig. 2). The general bisector formula in the parametric form using the direction *k* is defined as

$$y - y_1 = k(x - x_1)\ ,\ k = \frac{y_2 - y_1}{x_2 - x_1},\quad k = \mathrm{tg}(\varphi)\ , \qquad (5)$$

where $y_{1,2}$ represent values of PSD in [dB] of determined formants, and $x_{1,2}$ are the positions of the formants on the frequency axis in [Hz]. When $k < 0$, the formants have declining trend, when $k > 0$, the formants have ascending trend. The resulting angle $\varphi$ in degrees is defined as $\varphi = (\mathrm{Arctg}(k)/\pi)\cdot 180$. Obtained basic and complementary FF values are separately processed depending on a voice type (male / female), subsequently sorted by emotional styles, and stored in separate stacks.

The whole evaluation and comparison process of the FF values consists of six steps:

1) Calculation of basic statistical values of formant frequencies and their 3 -dB bandwidth (minimum, maximum, mean values, and standard deviation);
2) Calculation and building of the histograms for $F_{1,2,3}$ frequencies;
3) Building of bar diagrams of $F_{1,2,3}$, and $B3_{1,2,3}$ values for visual comparison;
4) Building of diagrams of formant tilts (bisectors with directions), and $F_1/F_2$, $F_1/F_3$ and $F_2/F_3$ mutual frequency positions;
5) Calculation of ratios of the basic FF mean values for emotional and neutral states;
6) Numerical matching of formant tilts (directions and angles between first three spectral maxima of a smoothed envelope).

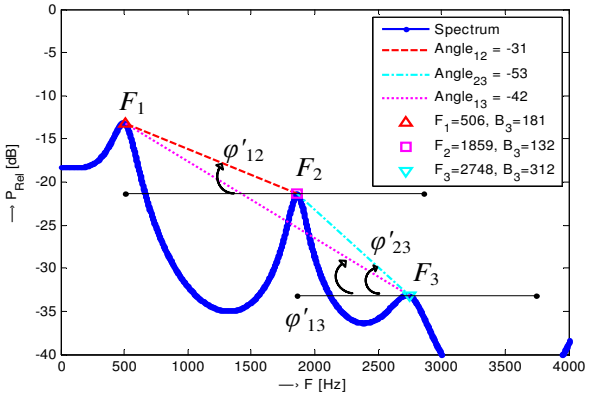


Fig. 2. Example of formant tilts determination – the LPC spectrum with $N_A$=32, stationary part of the vowel "e", female voice, F0≈190 Hz, $f_s$ = 16 kHz; the complementary angles are calculated as $\varphi' = \varphi - 180$.

## IV. Material, Experiments and Results

In our experiment the used speech corpus extracted from the Czech and Slovak stories performed by professional actors contains sentences with different contents expressed in four emotional states: "neutral state", "joy", "sadness", and "anger" uttered by several speakers (134 sentences spoken by male voices and 132 sentences spoken by female voices, 8+8 speakers altogether). The processed speech material consists of sentences with duration from 0.5 to 5.5 seconds, resampled at 16 kHz. The frame length for spectral analysis depends on the mean pitch period of the processed signal. For spectral analysis we had chosen 24-ms frames for the male voices, and 20-ms frames for the female voices. Calculation of the FF values was supplemented with determination of the fundamental frequency F0 by autocorrelation analysis method with experimentally chosen pitch ranges as follows: 55 ÷ 250 Hz for the male voices and 105 ÷ 350 Hz for the female ones. Then, the F0 values were compared and corrected by the results obtained with the help of the PRAAT program [12] with similar internal settings of F0 values. For voicing frame classification, the value of the detected pitch period $L$ was used. If the value $L \neq 0$, the processed speech frame is determined as voiced, in the case of $L = 0$ the frame is marked as unvoiced.

From the main speech signal database of spoken sentences, the next one consisting of manually selected ROIs corresponding to the stationary parts of the vowels "a", "e", "i", "o", "u", and consonants "m" and "n" was consequently created for detailed analysis. Number of analyzed voiced frames was in total:

a) Male: neutral - 5103, joy - 4927, sadness - 4642, anger - 4391.
b) Female: neutral - 5223, joy - 4541, sadness - 4203, anger - 4349.

Partial results of analysis of all voiced frames of the main speech corpus are presented in the form of the box-plot graphs of basic statistical parameters of the $F_{1,2,3}$ values for male and female voice determined from the neutral and emotional speech (Fig. 3).

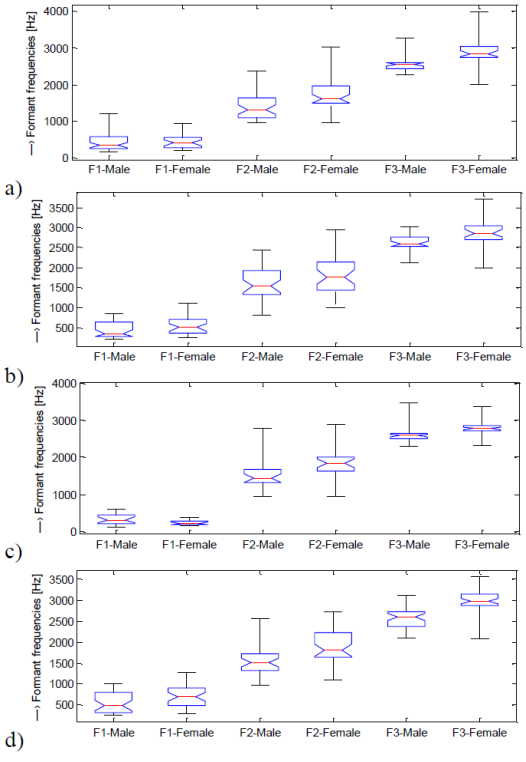


Fig. 3. Basic statistical parameters of the first three format frequencies for male and female voices: speech in "neutral" (a), "sad" (b), "joyful" (c), and "angry" emotional styles (d).

The common bar graphs of mean values of the first three formant 3 -dB bandwidths for both voices are presented in Fig. 4. Summary histograms of the first three formant frequencies for the male and female speech in different emotional styles are shown in Fig. 5 (male voice) and Fig. 6 (female voice). Two diagrams of bisectors with directions

given by formant tilts from the male and the female voices are shown in Fig. 7. Comparison between obtained mutual formant mean frequencies $F_1/F_2$, $F_1/F_3$, and $F_2/F_3$ for different emotional states of male and female voices is presented in Fig. 8. Diagrams of detailed analysis of formant mutual frequency positions for different emotional states of voiced sounds corresponding to vowels "a", "e", "o", and consonant "n" selected from the speech material of male and female voices are shown in Fig. 9 and Fig. 10.

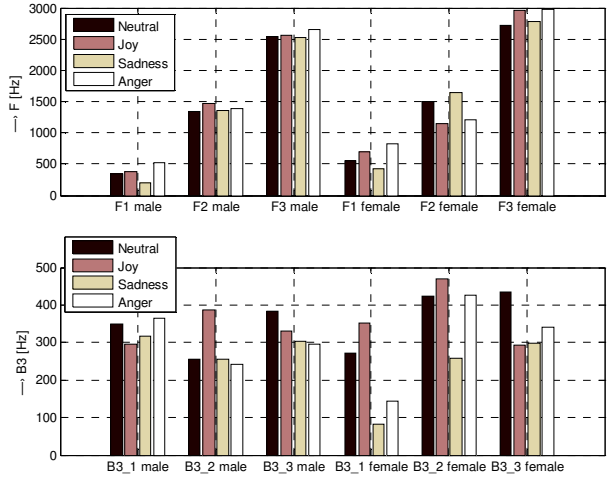


Fig. 4. Common bar graphs of mean values of the first three formant frequencies (upper), and their 3 -dB bandwidths (lower) for different emotional states of male and female voices.

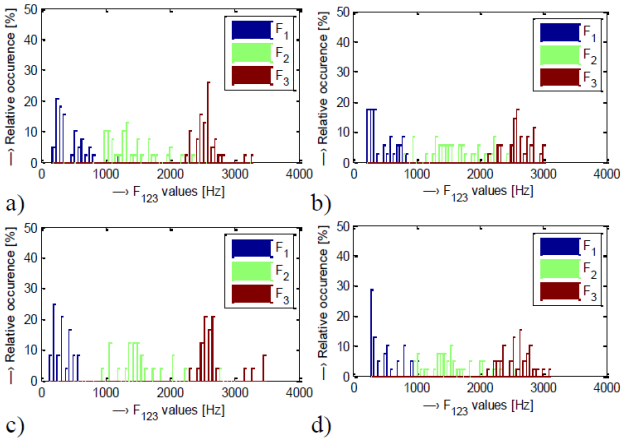


Fig. 5. Histograms of $F_{1,2,3}$ values for different emotional states: neutral (a), joy (b), sadness (c), and anger (d) – male voices.

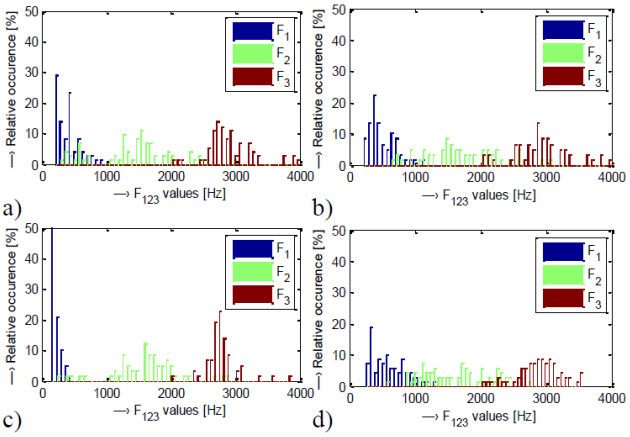


Fig. 6. Histograms of $F_{1,2,3}$ values for different emotional states: neutral (a), joy (b), sadness (c), and anger (d) – female voices.

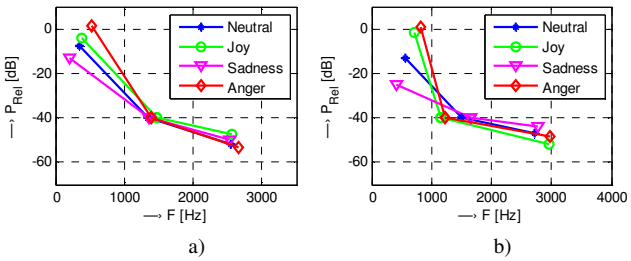


Fig. 7. Summary diagrams of bisectors with directions given by formant tilts of male (a); and female (b) voices.

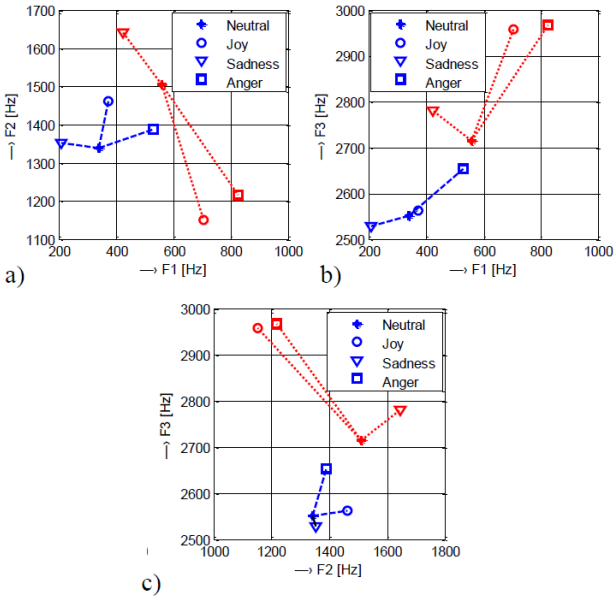


Fig. 8. Summary diagrams of mean $F_1/F_2$ (a), $F_1/F_3$ (b), $F_2/F_3$ (c) mutual frequency positions for different emotional states of male voices (blue dashed lines and marks) and female ones (red dotted lines and marks).

The mean emotional-to-neutral formant position ratios between different emotional states and a neutral state for male and female voices are presented in Table I; summary female-to-male ratios of formant positions are shown in Table II. Detailed results of the mean $F_{1,2,3}$ frequencies and their 3 -dB bandwidths of the selected voiced sounds in neutral speaking style are shown together with numbers of analyzed voiced frames $N_F$ in Table III and Table IV for male and female voices. Summary results of detailed analysis of formant tilts (complementary angles in [deg]) of voiced sounds in neutral speaking style for male and female voices are presented in Table V.

TABLE I. MEAN EMOTIONAL-TO-NEUTRAL FORMANT POSITION RATIOS.

| Formant ratio | $F_{1male}$ | $F_{2male}$ | $F_{3male}$ | $F_{1female}$ | $F_{2female}$ | $F_{3female}$ |
|---|---|---|---|---|---|---|
| Joyous: neutral | 0.712 | 1.025 | 1.038 | 0.898 | 1.082 | 1.049 |
| Sadness: neutral | 1.043 | 0.813 | 0.899 | 1.353 | 0.948 | 0.938 |
| Angry: neutral | 1.123 | 0.795 | 0.762 | 1.282 | 0.885 | 0.887 |

TABLE II. FEMALE-TO-MALE RATIOS OF FORMANT POSITIONS.

| Female:male ratio | Neutral | Joy | Sadness | Anger |
|---|---|---|---|---|
| $F_1$ | 1.238 | 1.261 | 1.297 | 1.142 |
| $F_2$ | 1.102 | 1.022 | 1.038 | 1.113 |
| $F_3$ | 1.108 | 1.126 | 1.043 | 1.117 |

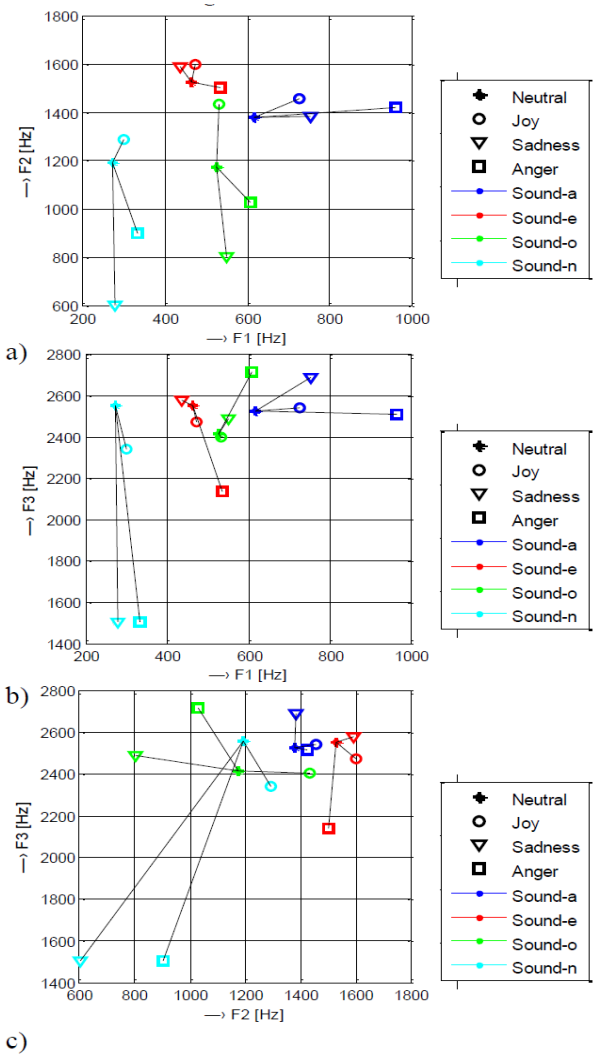


Fig. 9. Diagrams of formant mutual frequency positions for different emotional states of sounds corresponding to vowels "a", "e", "o", and consonant "n": $F_1/F_2$ (a), $F_1/F_3$ (b), $F_2/F_3$ (c) – male voice.

TABLE III. DETAILED RESULTS OF THE MEAN $F_{1,2,3}$ FREQUENCIES AND THEIR BANDWIDTHS – NEUTRAL SPEAKING STYLE, MALE VOICE.

| Sound | $N_F$ [-] | $F_1 / B3_1$ [Hz] | $F_2 / B3_2$ [Hz] | $F_3 / B3_3$ [Hz] |
|---|---|---|---|---|
| "a" | 758 | 631 / 261 | 1363 / 193 | 2529 / 366 |
| "e" | 720 | 451 / 140 | 1590 / 240 | 2490 / 402 |
| "i" | 684 | 297 / 94 | 1879 / 412 | 2615 / 276 |
| "o" | 608 | 514 / 80 | 1178 / 320 | 2594 / 486 |
| "u" | 735 | 393 / 191 | 1113 / 173 | 2531 / 327 |
| "m" | 784 | 209 / 143 | 1186 / 309 | 2414 / 126 |
| "n" | 934 | 271 / 99 | 1191 / 427 | 2557 / 345 |

TABLE IV. DETAILED RESULTS OF THE MEAN $F_{1,2,3}$ FREQUENCIES AND THEIR BANDWIDTHS – NEUTRAL SPEAKING STYLE, FEMALE VOICE.

| Sound | $N_F$ [-] | $F_1 / B3_1$ [Hz] | $F_2 / B3_2$ [Hz] | $F_3 / B3_3$ [Hz] |
|---|---|---|---|---|
| "a" | 782 | 806 / 321 | 1545 / 220 | 2725 / 292 |
| "e" | 802 | 621 / 152 | 1754 / 402 | 2744 / 473 |
| "i" | 576 | 311 / 80 | 1922 / 397 | 2814 / 361 |
| "o" | 618 | 451 / 263 | 1671 / 466 | 2606 / 356 |
| "u" | 684 | 438 / 120 | 1428 / 167 | 2943 / 292 |
| "m" | 696 | 314 / 156 | 1257 / 433 | 2613 / 370 |
| "n" | 945 | 357 / 130 | 1272 / 373 | 2690 / 313 |

TABLE V. DETAILED ANALYSIS OF FORMANT TILTS; COMPLEMENTARY ANGLES IN [DEG] FOR MALE AND FEMALE VOICES.

| Sound | Male voice | | | Female voice | | |
|---|---|---|---|---|---|---|
| | $\varphi'_{12}$ | $\varphi'_{13}$ | $\varphi'_{23}$ *) | $\varphi'_{12}$ | $\varphi'_{13}$ | $\varphi'_{23}$ *) |
| "a" | -42 | -27 | -49 | -31 | -22 | -36 |
| "e" | -45 | -51 | -34 | -32 | -36 | -28 |
| "i" | -43 | -59 | 33 | -46 | -52 | 9 |
| "o" | -51 | -56 | -48 | -44 | -57 | -34 |
| "u" | -53 | -64 | -44 | -48 | -59 | -37 |
| "m" | -44 | -65 | -3 | -52 | -70 | -10 |
| "n" | -48 | -70 | -2 | -47 | -72 | -12 |

*Note:* *) *negative angle values* $\Rightarrow$ *the formants have declining trend, otherwise the formants have ascending trend*

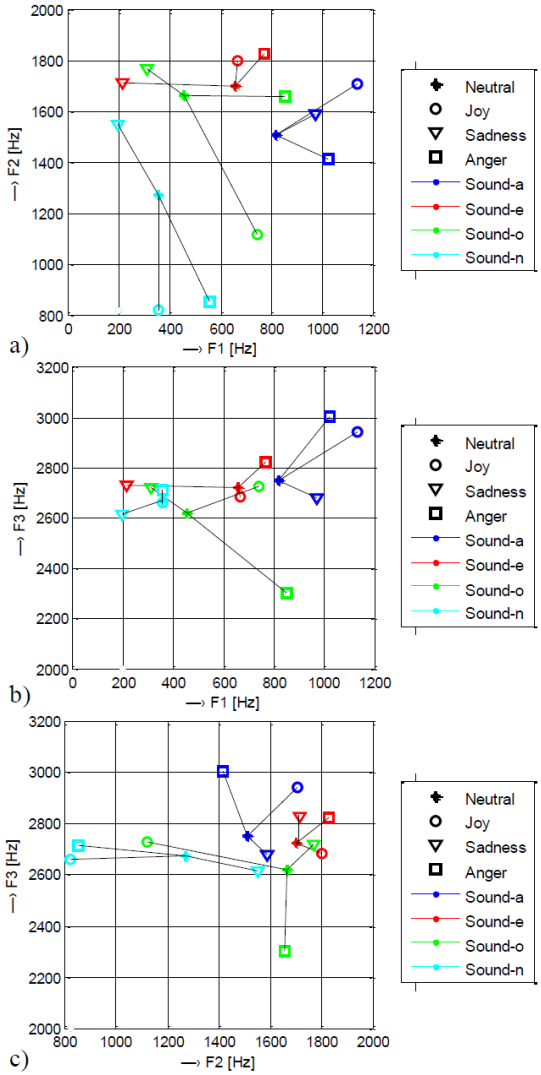


Fig. 10. Diagrams of formant mutual frequency positions for different emotional states of sounds corresponding to vowels "a", "e", "o", and consonant "n": $F_1/F_2$ (a), $F_1/F_3$ (b), $F_2/F_3$ (c) – female voice.

## V. DISCUSSION AND

Our experiment was aimed at analysis and comparison of the formant features in emotional and neutral speech (voiced parts of recorded speech signal). For speech synthesis the source-filter model with cepstral description of the vocal tract transfer function [9] was applied. The parameters used in the original realization of the cepstral speech synthesizer had been obtained by evaluation of a speech signal in the

database of phones uttered by a male speaker in a neutral speech. Subsequently we decided to carry out analysis of the first three formant positions ($F_1$, $F_2$, and $F_3$) got from speech signals expressing different emotional states and compare results for male and female voices. Results of the first three formant position ratios (see Table I) together with summary of female-to-male ratios (see Table II) will be used for emotional speech transformation and production in the multi-voice TTS system.

Values of the basic formant features obtained from the female voices have higher standard deviation (compare box-plot graphs of basic statistical parameters in Fig. 3), and in correspondence with our expectancy formant frequencies are approximately about 15 % higher than that of the male voices. In addition, obtained results correlate with conclusions in [13] that during pleasant emotions the first formant is falling and resonances are raised. For unpleasant emotions the first formant is rising, and the second and the third formants are falling. We can also conclude that the first formant and the higher formants of emotional speech shift in opposite directions. For pleasant emotions the first formant shifts to the left, and the higher formants shift to the right. For unpleasant emotions the opposite situation occurs: the first formant shifts to the right, and the higher formants shift to the left. Contrary to it, the values of the formant 3 -dB bandwidths have no correlation with the type of the speaking style or the type of the voice (see common bar graphs in Fig. 4). On the other hand, the comparison of the formant tilts shows good differentiation between neutral and emotional styles (Fig. 7) for both voices.

Results of detailed analysis of basic five vowels indicate differences between $F_{1,2,3}$ positions for neutral and emotional styles which are visible well in the graph of formant mutual frequency positions (Fig. 8 and Fig. 9). However, in the case of the consonants "m" and "n" the differences of the $F_{1,2,3}$ values were lower due to smaller absolute amplitudes of the speech signal of the vowels and they cannot be correctly compared visually. These obtained results are in good correspondence with the general knowledge of [14], [15], that vowel formant areas of the male voice lies in the ranges $F_1 \approx 250 \div 700$ Hz, $F_2 \approx 700 \div 2000$ Hz, $F_3 \approx 2000 \div 3200$ Hz, and the female voice vowel formant areas are higher, lying about $F_1 \approx 300 \div 840$ Hz, $F_2 \approx 840 \div 2400$ Hz, $F_3 \approx 2400 \div 3840$ Hz.

Numerical matching of the mean $F_{1,2,3}$ values of all voiced sounds in a neutral style also documents sufficient differentiation, again the mean $B3_{1,2,3}$ values don't carry this information – see Table III and Table IV. The complementary angles between PSD at frequencies $F_1$ and $F_2$ ($\varphi'_{12}$) and the complementary angles between PSD at frequencies $F_1$ and $F_3$ ($\varphi'_{13}$) have always negative values. The complementary angles between PSD at frequencies $F_2$ and $F_3$ ($\varphi'_{23}$) can have also positive values or values near zero (see Table V).

## VI. CONCLUSIONS

Knowledge about the effect of emotional states on speech signals is very important not only for emotion recognition but for standard speech recognition when influence of various factors (including the speaker's emotional state) is taken into consideration as well [16]. Considering the fact that our current database contains only speech with acted emotional styles, the analysis of FF properties using also speech material representing real emotions should be recorded. Last but not least, we would like to use broader comparison with other databases in different languages (e.g. the German speech database Emo-DB [17], or international COST 2102 Italian Database of Emotional Speech [18]).

## REFERENCES


[1] M. Chetouani, A. Mahdhaoui, F. Ringeval, "Time-Scale Feature Extractions for Emotional Speech Characterization", *Cognitive Computation*, vol. 1. pp. 194–201, 2009. [Online]. Available: http://dx.doi.org/10.1007/s12559-009-9016-9

[2] I. Luengo, E. Navas, I. Hernáez, "Feature Analysis and Evaluation for Automatic Emotion Identification in Speech", *IEEE Transactions on Multimedia*, vol. 12, pp. 490–501, 2010. [Online]. Available: http://dx.doi.org/10.1109/TMM.2010.2051872

[3] T. L. Pao, Y. T. Chen, J. H. Yeh, W. Y. Liao, "Combining Acoustic Features for Improved Emotion Recognition in Mandarin Speech", *Affective Computing and Intelligent Interaction,* vol. 3784. pp. 279–285, 2005. [Online]. Available: http://dx.doi.org/10.1007/11573548_36

[4] O. A. Schipor, S. G. Pentiuc, M. D. Schipor, "The Utilization of Feedback and Emotion Recognition in Computer Based Speech Therapy System", *Elektronika ir Elektrotechnika (Electronics and Electrical Engineering)*, no. 3, pp. 101–104, 2011.

[5] B. Yang, M. Lugger, "Emotion Recognition from Speech Signals Using New Harmony Features", *Signal Processing*, vol. 90, pp. 1415–1423, 2010. [Online]. Available: http://dx.doi.org/10.1016/j.sigpro.2009.09.009

[6] C. H. Wu, W. B. Liang, "Emotion Recognition of Affective Speech Based on Multiple Classifiers Using Acoustic-Prosodic Information and Semantic Labels", *IEEE Transactions on Affective Computing*, vol. 1, no. 2, pp. 10–21, 2011.

[7] M. Grůber, Z. Hanzlíček, "Czech Expressive Speech Synthesis in Limited Domain Comparison of Unit Selection and HMM-Based Approaches", in *Proc. of the TSD 2012*, Springer-Verlag Berlin Heidelberg, 2012, vol. 7499, pp. 656–664.

[8] S. G. Koolagudi, R. S. Krothapalli, "Two Stage Emotion Recognition Based on Speaking Rate", *International Journal of Speech Technology*, vol. 14, pp. 35–48, 2011. [Online]. Available: http://dx.doi.org/10.1007/s10772-010-9085-x

[9] R. Vích, J. Přibil, Z. Smékal, "New Cepstral Zero-Pole Vocal Tract Models for TTS Synthesis", in *Proc. of the IEEE Region 8 EUROCON 2001*, 2001, vol. 2, pp. 458–462.

[10] N. H. Shah, *Numerical Methods with C++ Programming*, Prentice-Hall Of India Learning Private Limited, New Delhi, 2009, p. 251.

[11] H. G. Ilk, O, Eroğul. B. Satar, Y. Özkaptan, "Effects of Tonsillectomy on Speech Spectrum", *Journal of Voice*, vol. 16. pp. 580–586, 2002. [Online]. Available: http://dx.doi.org/10.1016/S0892-1997(02)00133-9

[12] P. Boersma, D. Weenink, *Praat: Doing Phonetics by Computer (Version 5.2.20).* [Online]. Available: http://www.praat.org/

[13] K. R. Scherer, "Vocal Communication of Emotion: A Review of Research Paradigms", *Speech Communication,* vol. 40, pp. 227–256, 2003. [Online]. Available: http://dx.doi.org/10.1016/S0167-6393(02)00084-5

[14] G. Fant, *Acoustical Analysis of Speech* (*Encyclopedia of Acoustics)*, John Wiley & Sons, 1997, pp. 1589–1598.

[15] G. Fant, "Speech Acoustics and Phonetics", Kluwer Academic Publishers, Dordrecht, 2004.

[16] G. Čeidaitė, L.Telksnys, "Analysis of Factors Influencing Accuracy of Speech Recognition", *Elektronika ir Elektrotechnika (Electronics and Electrical Engineering)*, no. 9, pp. 69–72, 2010.

[17] F. Burkhardt, A. Paeschke, M. Rolfes, W. Sendlmeier, B. A. Weiss, "Database of German Emotional Speech", in *Proc. of the INTERSPEECH 2005*. ISCA, Lisbon, Portugal, 2005, pp.1517–1520.

[18] H. Atassi, M. T. Riviello, Z. Smékal, A. Hussain, A. Esposito, "Emotional Vocal Expressions Recognition Using the COST 2102 Italian Database of Emotional Speech", *Development of Multimodal Interfaces: Active Listening and Synchrony*, Springer-Verlag Berlin Heidelberg, 2010, vol. 5967, pp. 255–267.