Parameters of Fricatives and Affricates in Russian Emotional Speech

Valery A. Petrushin [1)] and Veronika Makarova [2)]

[1)] Accenture Technology Labs
Accenture, Chicago, USA
valery.a.petrushin@accenture.com

[2)] Department of Languages and Linguistics
University of Saskatchewan, Saskatoon, Canada
v.makarova@usask.ca

Abstract

The paper investigates the effect of emotive states on the characteristics of Russian fricatives and affricates. The experimental data come from RUSLANA, a database containing neutral utterances along with the ones that portray surprise, happiness, anger, sadness and fear. The paper focuses on the role of duration, energy and dynamic ranges in the expression of emotions at the segmental level.

1. Introduction

Due to the immense industrial demand in electronic devices, toys and robots that can adequately interact with humans, recognition and synthesis of human emotions is one of the fastest growing areas in speech processing [1]. Emotion expression is an important features of human evolution [2], it carries a significant part of intended meaning in communication [3], constitutes an important component of motivation [4], and contributes to the differentiation of genders and cultures [5,6]. Although emotions can be expressed via non-linguistic means, such as facial expressions, postures, gestures, art, etc., the emotional component of speech is extremely important in all the world's languages [6].

All the structural layers of language, such as lexis, syntax, discoursal units, can be involved in the expression of emotions [7]. This paper considers only the acoustic cues of the expression of emotive states, which are often associated with the role of prosody in speech [8]. However, it is also known that certain segmental features may also contribute to the expression of emotions, in particular, duration and spectral characteristics of segments [9, 10].

In our earlier studies, we examined the effect of emotive states on the production of Russian vowels [11]. This paper concentrates on the segmental cues of emotive states in Russian fricatives and affricates.

2. Materials and methods

The data in the study is taken from RUSLANA (Russian Language Affective) database described in [12]. The database includes utterances from 61 subjects (12 male and 49 female). Each subject recorded 10 sentences of different syntactical types and intonational patterns portraying the above mentioned six emotional states. In our research we used 600 utterances from 10 subjects (5 male and 5 female), which count 60 utterances per subject or 100 utterances per emotional state. The subjects were selected based on their high ranks in ability to express emotions.

2.1. Analyzed features

One list of parameters has been extracted to account for all the segmental features in the database: vowels and consonants. The full list of parameters is reported in [11].

The following parameters have been analyzed for the fricative and affricate phonemes in the study: phoneme duration (Dur), average energy (E), power spectra (Fq) for the following bandwidths: 0 - 500 Hz (Fq500), 501 - 1000 Hz (Fq1000), 1001 - 1500 Hz (Fq1500), 1501 - 2000 (Fq2000) Hz, 2001 - 2500 Hz (Fq2500), 2501 - 3000 Hz (Fq 3000), 3001 - 3500 Hz (Fq3500), 3501 - 4000 Hz (Fq4000), 4001 - 5000 Hz (Fq5000), 5001 - 6000 Hz (Fq6000), 6001 - 7000 Hz (Fq7000), 7001 - 8000 Hz (Fq8000), 8001 - 10000 Hz (Fq10000), 10001 - 12000 Hz (Fq12000), 12001 - 14000 Hz (Fq14000), 14001 - 16000 Hz (Fq16000).
The list of analyzed phonemes in Sampa notation includes 13 fricatives: [f, f', v, v', s, s', z, z', S, S', Z, x, x'], and two affricates: [ts, tS].

2.2. Statistical analysis

Parameters were analyzed for descriptive statistics, ANOVA with subsequent post-hocs were conducted to determine the existence of statistically significant differences in parameter values across emotional states. The total of 2292 segments were analyzed.

3. Results

3.1. Duration

Duration of fricative and affricates does not show statistically significant differences across emotions, with the exception of phonemes [x, x', z']. For [x, x'] phonemes, 'happy' has the shortest, and 'sad' - the longest durations. Post-hoc tests do not reveal statistically significant differences between any individual pairs of emotions. The phoneme [z'] has a different distribution of duration across emotions, whereby the shortest durations are associated with 'neutral' and 'sad' emotions, and the longest - with 'angry' and 'afraid', See Table 1 for the parameter values.

Table 1. Average duration (ms) of [x, x', z'].

	Sig (p)	angry	sad	afraid	happy	neutral	surprised
x (120)	0.019	80.14	91.92	89.87	78.01	88.76	89.75
x' (59)	0.038	132.02	151.64	123.9	112.3	110.9	112.9
z' (232)	0.039	86.74	74.85	87.68	77.44	68.57	78.69

Table 2. Average energy values across emotional states.

E	N	Sig (p)	angry	sad	afraid	happy	neutral	surprised
ts	118	0.000054	0.019	0.011	0.018	0.022	0.008	0.014
tS	237	0	0.043	0.017	0.032	0.035	0.015	0.022
x	120	0	0.02	0.007	0.012	0.016	0.006	0.011
x'	59	0.038	0.019	0.01	0.017	0.018	0.01	0.014
s	586	0	0.026	0.014	0.02478	0.02472	0.011	0.018
s'	183	0.00001	0.032	0.018	0.026	0.031	0.014	0.022
z	115	X	0.036	0.043	0.023	0.03	0.016	0.023
z'	232	0.00001	0.024	0.01	0.018	0.021	0.011	0.015
S	60	0.0019	0.061	0.034	0.058	0.06	0.025	0.042
S'	119	0	0.05	0.019	0.042	0.039	0.014	0.03
f	244	0.00007	0.014	0.006	0.011	0.013	0.006	0.009
f'	60	0.002	0.011	0.004	0.007	0.006	0.003	0.005
v	229	0.0019	0.014	0.008	0.011	0.018	0.007	0.013
v'	158	X	0.026	0.031	0.056	0.063	0.01	0.015
Z	119	0	0.04	0.022	0.03	0.039	0.014	0.026
Total/Average	2639		0.029	0.016933	0.025719	0.029048	0.011333	0.0186

Table3. Results of pair-wise post-hoc comparisons of parametric differences across emotional states.

	ang sad	ang afr	ang hap	ang neu	ang sur	sad afr	sad hap	sad neu	sad sur	afr ha	afr neu	afr sur	hap neu	hap sur	neu sur	SUM param
E	10	3	0	12	6	3	11	0	0	1	7	1	12	3	1	70
Fq500	10	3	0	11	2	2	10	0	1	2	6	0	11	2	2	62
Fq1000	8	3	0	12	0	1	9	0	1	2	3	0	12	1	1	53
Fq1500	12	2	0	13	3	5	10	0	2	0	4	0	11	1	3	66
Fq2000	8	1	0	12	1	2	9	0	2	0	4	0	10	0	3	52
Fq2500	10	1	0	11	2	4	7	0	2	0	5	0	9	1	3	55
Fq3000	11	1	1	12	3	5	7	0	1	0	8	0	11	0	6	66
Fq3500	11	1	0	11	3	4	9	0	3	0	9	0	10	0	6	67
Fq4000	9	1	0	10	3	2	6	0	1	0	6	0	9	0	6	53
Fq5000	9	0	0	11	1	3	4	0	0	0	8	0	8	1	7	52
Fq6000	9	0	0	13	2	3	5	0	1	0	8	0	8	1	4	54
Fq7000	11	0	0	13	3	5	6	0	2	0	8	0	10	0	6	64
Fq8000	12	0	1	13	4	5	8	0	2	0	8	0	11	1	7	72
Fq10000	11	1	0	13	3	4	8	0	3	0	7	0	12	1	7	70
Fq12000	13	4	0	15	5	7	11	0	3	3	11	0	15	3	8	98
Fq14000	13	2	0	14	5	7	12	0	4	2	12	0	15	2	10	98
Fq16000	12	1	0	15	4	8	12	0	4	0	11	0	14	1	6	88
SUM emotion	179	24	2	211	50	70	144	0	32	10	125	1	188	18	86	

3.2. Energy

Energy is a parameter of high significance for the expression of emotions in all the analyzed phonemes except [v, v', z].
The parameter values are represented in Table 2. The highest energy values are found in 'happy', 'angry' states, medium - in 'afraid, surprised', and the lowest - in 'neutral' and 'sad' states.

3.3. Power spectra

All the analyzed dynamic ranges of fricatives and affricates are highly significant for the expression of emotions. Figures 1 -15 represent the distribution of energy over the frequency ranges for all the analyzed consonants. As the figure demonstrates, all the consonants have higher spectral energies for 'happy' and 'angry' states, and the lowest for 'neutral' and 'sad', whereby 'sad' has somewhat higher energy than 'neutral' at frequencies over 3000-3500Hz. Most consonants also demonstrate a tilt in the distribution of energy towards higher frequencies with the 'angry' and 'happy' emotions, as compared to 'neutral' and 'sad'. This tendency is accompanied by a rightward shift in the position of the dynamic peak for the 'angry and 'happy' emotions as compared to 'neutral' in fricatives and affricates with energies concentrated in medium (2500-5000Hz) and high (5000-12000) ranges: [s, s', z, z', S, Z, tS, x']. This tendency is not observed for [ts], where most energy is concentrated in the narrow high range between 6000 and 12000Hz, although the energy values for this phoneme are also greater for 'angry' and 'happy' emotional states.

Fricatives without clear spectral concentrations of energy in the dynamic ranges [f, f', v, v'] display more energy at higher frequencies for the 'angry' state than for neutral and 'sad' ones. The only exceptions with no significant differences across emotive states were the following: phoneme [x']: Fq500, 1000, 2100, 2500, 3000, 3500; phoneme [z]: Fq 5000; phoneme [S] : Fq 2500, 3000, 3500 ; phoneme [f'] : Fq 500.

The absence of significant differences in these cases is most likely explained by relatively small samples as compared to other phonemes.

3.4. Post-hoc comparisons of power spectra features

The comparison of all the pairs of emotional states across parameters shows that the area of power spectrum most salient for the expression of differences across emotional states is from 8000 to 16000 Hz (see Table 3)
The results allow to group pairs of emotions into four categories (in the decreasing order of differences):

1) Pairs with maximum parametric distinctions: angry/neutral, happy/neutral, angry/sad, happy/sad, afraid/neutral.
2) Pairs with strong parametric distinctions: neutral/surprised, sad/afraid, angry/surprised.
3) Pairs with moderate parametric distinctions: sad/surprised, angry/afraid, happy/surprised, afraid/happy.
4) Pairs with little or no parametric distinctions: angry/happy, afraid/surprised, sad/neutral.

4. Discussion

A recent promising development in speech technology is concatenative synthesis based on emotive natural speech corpora [13]. Databases of emotional speech like RUSLANA [12] are essential for synthesizing high intelligibility speech with realistic portrayal of human emotion. Synthesis and recognition of emotional speech can be more successful, if researchers gain more information regarding the acoustic parameters responsible for the expression of emotion. The advantage of RUSLANA is in the large number of speakers of both genders (61 speakers), which allows the extraction of speaker-independent parameters. While most current studies of the segmental characteristics of emotion concentrate on vowel quality [14], our study confirms that consonants (fricatives and affricates) also contain information about emotive states.

The results of this study confirm some earlier findings that demonstrate the importance of mean intensity and increase/decrease in high frequency energy spectrum for the expression of emotion and the increase in medium energy and high spectrum energy with emotions of happiness and anger [15, 16].

Our study sheds some additional light on the reasons why some pairs of emotions get misconstrued more than others. For example, anger can me misperceived as happiness, and sadness is often confused with neutrality [15]. Some similarities in the segmental clues of these emotions, such as the ones observed in our study in energy and spectra, must be contributing to misconstrued perception.

Linguistic (potentially language-specific) vs. paralinguistic (potentially universal) features in the expression of emotion has been a widely debated issue [17]. The data in our study come from simulated emotions. We believe that since these emotions can be simulated and are therefore under the control of the speaker, their acoustic cues should be treated as ‘linguistic’ ones, although it does not mean that they necessarily overlap with the expression of ‘real’ emotions.

5. Conclusions

This study demonstrated that duration of fricatives and affricates is of little salience for the expression of emotions in Russian, whereas the total energy and dynamic ranges are of extreme importance. The study also demonstrates that the emotive states fall into natural classes depending on the proximity/distance of the parametric values.

The results have to be treated with caution, since the database is not well balanced for the position of analyzed segments in regards of the syllable and word boundaries, as well as because the emotions in the database are ‘simulated’, and the results may not be fully applicable to the expression of emotions in natural settings.

Further directions of the study are extracting and analyzing the segmental features for all the consonants as well as addressing the prosodic characteristics. The practical outcome is application to synthesis and automatic recognition of emotional speech for Russian.

6. References

[1] Cowie, R., Douglas-Cowie, E. Tsapatsoulis, G., Votsis, G., Kollias, S., Fellenz, W., and Taylor, J. 2001. Emotion recognition in human-computer interaction, *IEEE Signal Processing Magazine*, 18 (1): 32-80.

[2] Darwin,C.1965 (repr.1872). The expression of emotions in man and animals. Chicago: U of Chicago Press.

[3] Iida, A., Campbell, N., Higuchi, F., Yasumura, M. 2003. A corpus-based speech synthesis system with emotion, *Speech Communication,* 40: 161-187.

[4] Lazarus, R.S. 1991. Emotion and Adaptation. New York: Oxford University Press.

[5] O'Kearney, R., Dadds, M. 2004. Developmental and gender differences in the language for emotions across the adolescent years. Cognition and Emotion, 18 (7): 913-938

[6] Wierzbicka, A. 1998. Russian Emotional Expression. *Ethos,* 26(4): 456-483.

[7] Min Lee, C. and Narayanan, S.S. Toward detecting emotions in spoken dialogs. *IEEE Transactions on Speech and Audio Processing,* vol 13, N 2, 2005, pp. 293-303.

[8] Mozziconacci, S. 2001. Emotion and attitude conveyed in speech by means of prosody. 2nd Workshop on Attitude, Personality and Emotions in User-Adapted Interaction, Sonthofen, Germany, July 2001, 1-10.

[9] Kienast, M. & W. F. Sendlmeier. 2000. Acoustical analysis of spectral and temporal changes in emotional speech. *Proc. ISCA Workshop on Speech and Emotion: A Conceptual Framework for Research*, Belfast, 5-7 September, pp. 92-97.

[10] Tickle, A. English and Japanese speakers' emotion vocalisation and recognition: A comparison highlighting vowel quality. *Proc. ISCA Workshop on Speech and Emotion: A Conceptual Framework for Research*, Belfast, 5-7 September, 2000, pp. 104-109

[11] Makarova, V., Petrushin, V. Vowel quality in Emotional Speech. SPECOM 2005. In. Proc. 10th Intl Conf. Speech and Computer, 17-19 October, 2005, Patras, Greece, 449-454.

[12] Makarova, V., Petrushin, V. RUSLANA: a Database of Russian Emotional Utterances. In Proc. ICSLP 2002, September 16-20, 2002, Denver, CO, 2041-2044.

[13] Iida, A., Campbell, N., Higuchi, F., Yasumura, M. 2002. A corpus-based speech synthesis system with emotion. Speech Communication, 40: 161-187.

[14] Airas, M., Alku, P. 2004. Emotion in short vowel segments: Effects of the glottal flow as reflected by the normalized amplitude quotient. *Lecture Notes in Artificial Intelligence*, v 3068, *Affective Dialogue Systems*:13-24.

[15] Pereira, C., Watson, C. 1998. Some acoustic characteristics of emotion. Proceedings, 5th ICSLP: 927-929.

[16] Pittam, J., Scherer, K. 1993. Vocal expression and communication of emotion. In: The handbook of emotion, ed. M. Lewis & J. Haviland, NY: Guildford: 185-197.

[17] Fernandez, R. 2004. A computational model for the automatic recognition of affect in speech. PhD thesis. MIT.

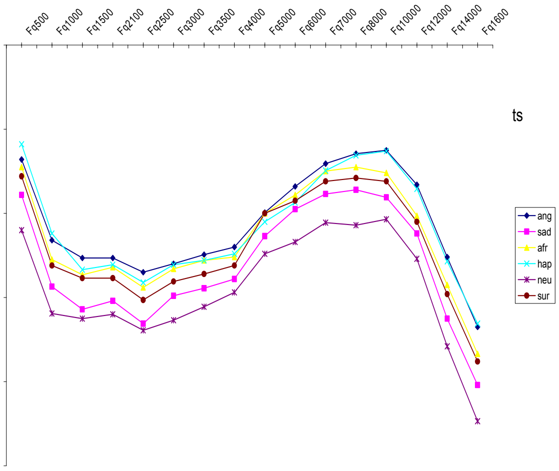

Figure 1. Power spectrum for phoneme [ts].

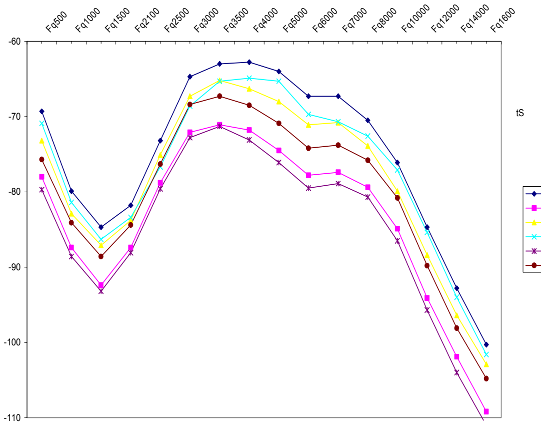

Figure 2. Power spectrum for phoneme [tS].

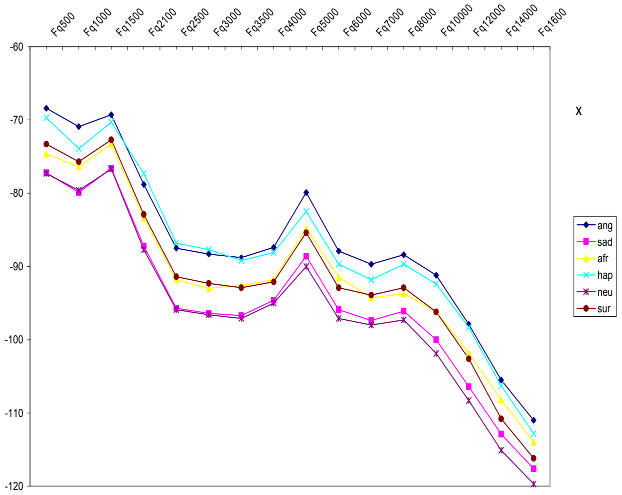

Figure 3. Power spectrum for phoneme [x].

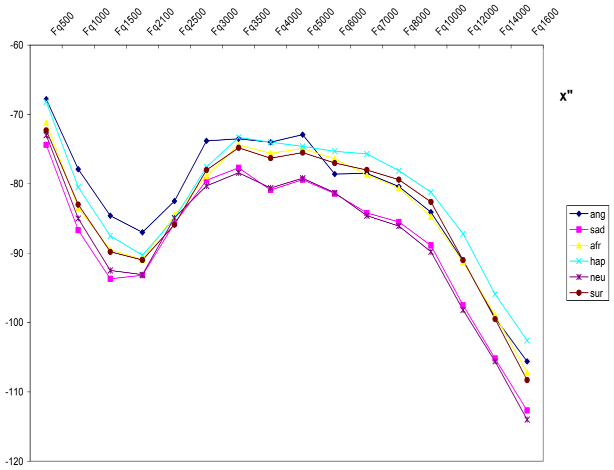

Figure 4. Power spectrum for phoneme [x'].

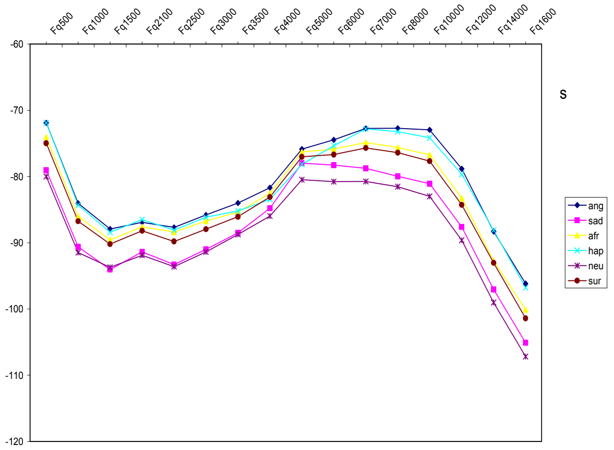

Figure 5. Power spectrum for phoneme [s].

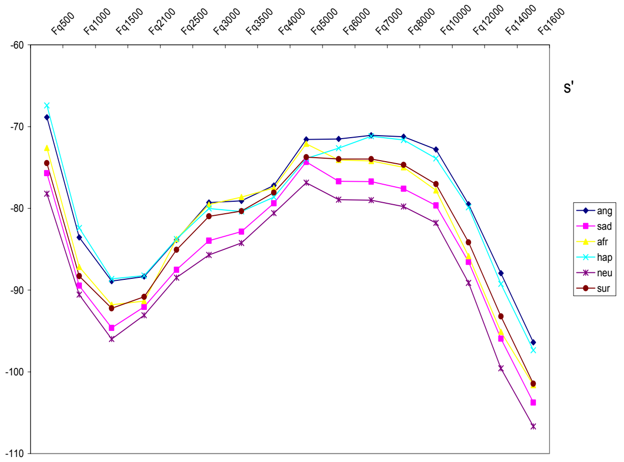

Figure 6. Power spectrum for phoneme [s'].

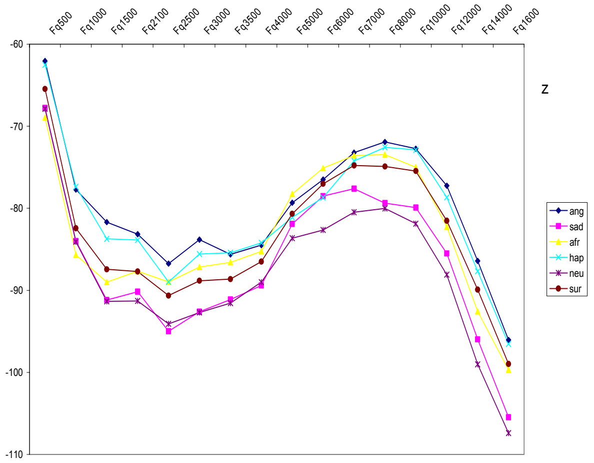

Figure 7. Power spectrum for phoneme [z].

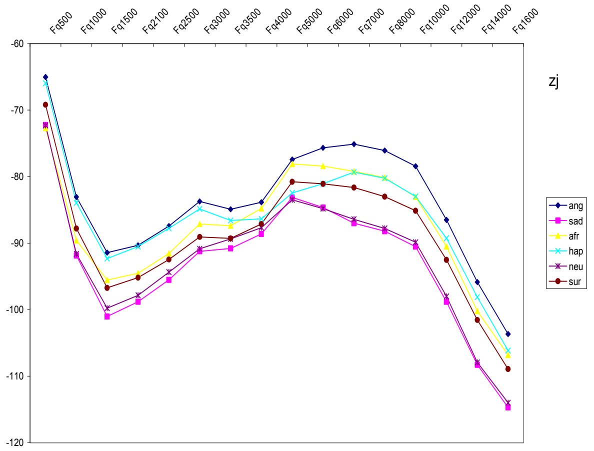

Figure 8. Power spectrum for phoneme [z'].

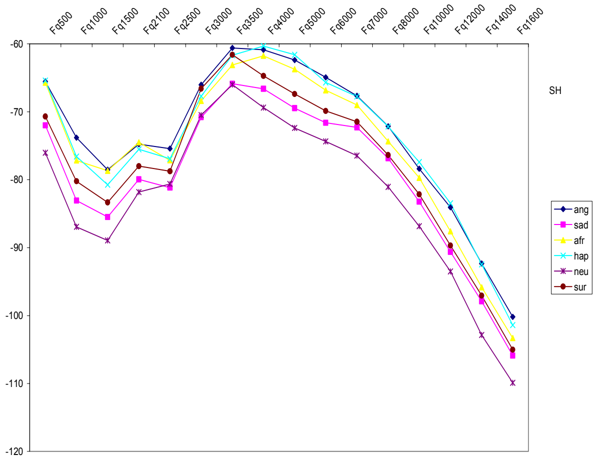

Figure 9. Power spectrum for phoneme [S].

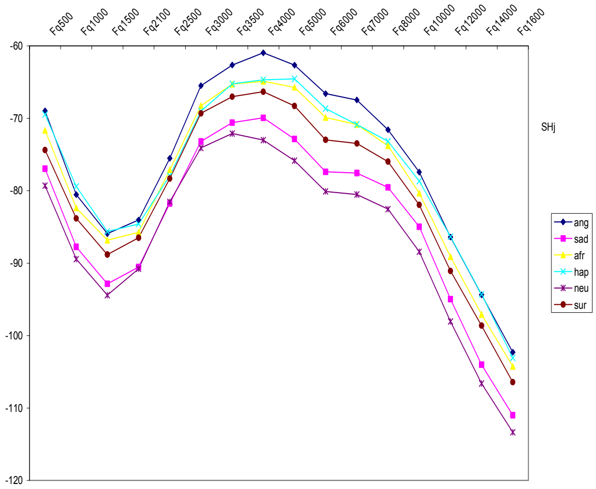

Figure 10. Power spectrum for phoneme [S'].

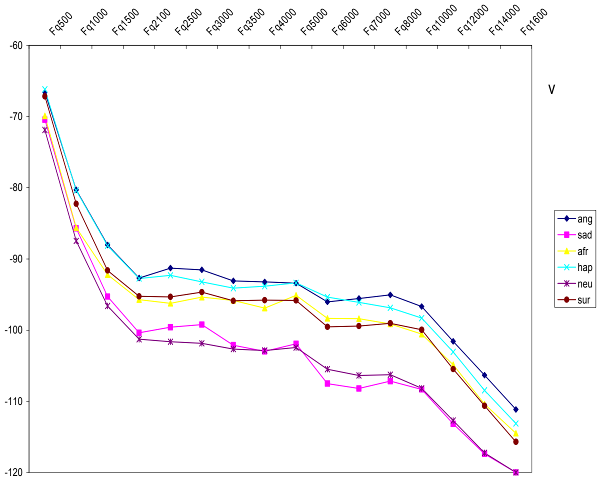

Figure 13. Power spectrum for phoneme [v].

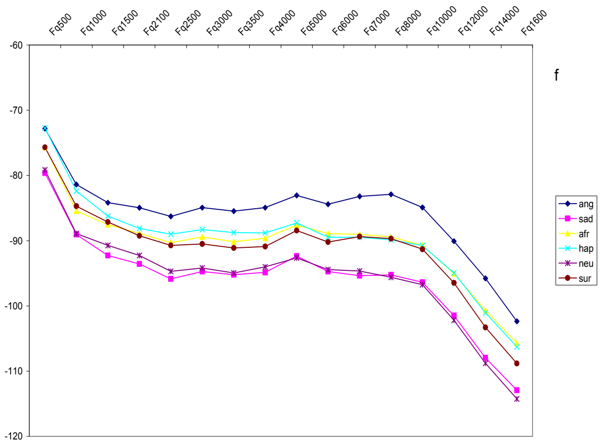

Figure 11. Power spectrum for phoneme [f].

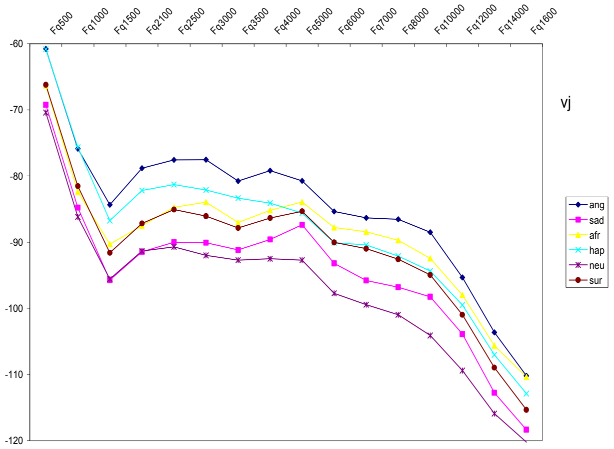

Figure 14. Power spectrum for phoneme [v'].

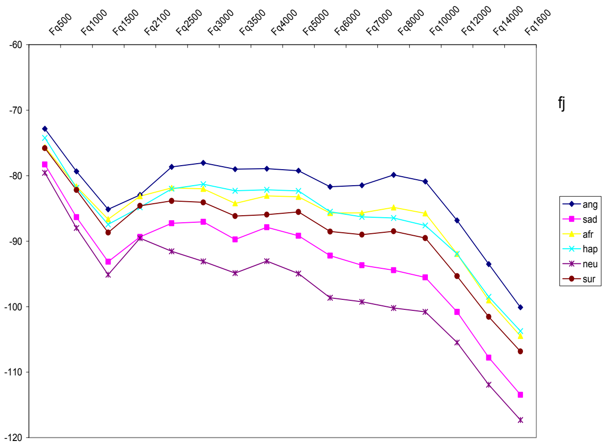

Figure 12. Power spectrum for phoneme [f '].

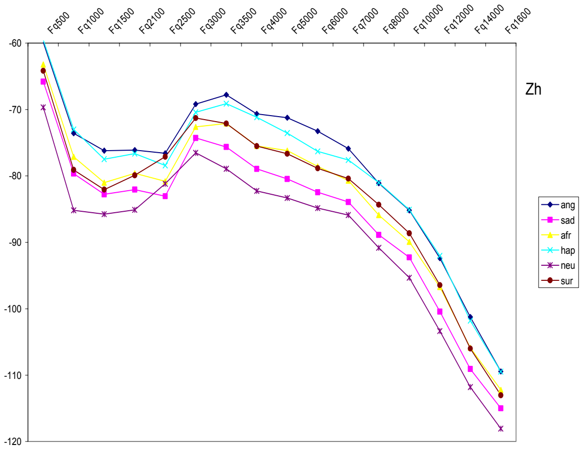

Figure 15. Power spectrum for phoneme [Z].