KAIRI TAMURI, MEELIS MIHKLA (Tallinn)

EMOTIONS AND SPEECH TEMPORAL STRUCTURE*

Abstract. The focus of the article is on whether emotions could be traced in the temporal structure of Estonian speech. There are two research questions, namely, (a) Do emotions affect speech rate? and (b) What detectable traces, if any, might emotions generate in word prosody? To answer question (a), the articulation rate of emotional utterances was measured and the results were compared with those on neutral speech. The difference revealed was statistically significant. To answer question (b), the relations between emotions and the temporal characteristics of words with a vowel-centered structure were investigated. Sound durations were measured, various duration ratios were computed and various combinations of the characteristics where subjected to statistical analysis. The results revealed a certain difference between the temporal characteristics of Q2 and Q3 feet, and a loss of the difference between the second and third quantity degrees in sad speech.

Keywords: Estonian, emotional speech, speech rate, word prosody, quantity degrees.

Introduction

The Estonian Emotional Speech Corpus (EESC) is currently used to study the acoustic characteristics of three emotions — anger, sadness and joy — and of neutral speech. The aim is to ascertain the definitive and distinctive parameters enabling recognition of emotions in Estonian speech as well as their identifiable synthesis. The results of the acoustic analysis of Estonian emotional speech, concerning pauses (Tamuri 2010), as well as formants and articulation precision (Tamuri 2012), also confirm that emotions do affect those parameters.

Studies of emotional speech acoustics have shown, for example, that variation in speech rate may, inter alia, be eloquent of the speaker's state of emotion. Thus, a very slow rate may indicate that the speaker is sad or depressed. It should be kept in mind, however, that speech rate is a rather subjective characteristic, which depends not only on the state of emotion but also on the speaker's gender, age, speech style, language, cultural space, communicative situation etc. (Laver 1994). Speech rate is measured in speech segments per unit time (e.g. speech sounds per second), either with pauses included or excluded (Braun, Oba 2007). In the present study a speaker's rate of articulation is measured on speech material where pauses have been excluded.

Estonian is a word-central language. It is in word prosody that quantity degrees are manifested, which certainly belong to the pivotal phenomena of Estonian phonetics. Notably, in Estonian the phonologically significant three-way opposition

between the Q1, Q2, and Q3 quantity degrees is manifested in the foot (Lehiste 1997). On the acoustic level, the distinctive features of those three quantity degrees include the duration ratio of the rhyme of the stressed syllable and the nucleus of the unstressed syllable,[1] and the F0 contour (Ross, Lehiste 2001), which together form a mutually complementary system. The temporal parameter is the duration ratio (V1 : V2) of the stressed and unstressed syllables in the word; this ratio is hitherto the most stable parameter distinguishing between the quantity degrees. The numerous different experiments from more than a half century have established the following general V1 : V2 ratios for the Estonian quantity degrees: 2 : 3 for the first degree (Q1), 3 : 2 for the second (Q2), and 2 : 1 for the third quantity degree (Q3) (Kalvik, Mihkla, Kiissel, Hein 2010; Lippus, Pajusalu, Allik 2007; Eek, Meister 1997; Krull 1993; Liiv 1961; Lehiste 1960). As, according to previous research (Kalvik, Mihkla 2010), the duration ratio of the stressed and unstressed syllables covers three fourths of data variation this ratio was placed in the focus of the present study as well. Besides the parameters mentioned earlier it has also been suggested that quantity degrees could be determined from the durational ratios of adjacent speech sounds (Eek, Meister 2003), using perception and weighting of durational differences. Up to now, emotional speech has never been studied from the point of view of temporal characteristics of word prosody.

The present study investigates the possible connections between the temporal characteristics of uttered words with a CV[V]CV structure and emotions, considering the words stressedness, phrase position and part of speech. Statistical methods (logistic regression and CART) are used to study combinations of different parameters, possibly enabling to detect small, covert, yet essential connections between the input parameters and emotions (Sagisaka 2003).

1. Hypotheses and research material

Hypothesis One states that emotions affect Estonian speech rate. Estonian being a word-central language hypothesis Two concerns word prosody: emotions affect word temporal structure, while the influence can be detected in temporal parameters and in the duration ratio V1 : V2 of the stressed and unstressed syllables, which is the main distinctive feature of quantity degrees. In addition it is investigated in precisely which temporal parameters of word prosody the possible rate specifics of emotional speech is manifested.

The acoustic base of the study consists of the Estonian Emotional Speech Corpus[2] of the Institute of the Estonian Language. The corpus has been generated on the principle that emotions can sufficiently well be identified from natural, non-acted speech and that natural speech synthesis should be based on non-acted speech (Iida, Campbell, Higuchi, Yasumura 2003). The corpus contains read sentences of anger, joy and sadness, and neutral speech. Those basic emotions also cover the following emotions: anger = displeasure, irony, dislike, contempt, schadenfreude, rage; joy = gratitude, happiness, pleasure, enthusiasm; sadness = lonelyness, disconsolateness, uneasiness, hopelessness. Neutral means 'without particular emotions'. The corpus items (text paragraphs) have been selected so that their content is likely to excite a state of emotion in the reader. Therefore the reader has not been prompted as to with what emotion the paragraph should be read. The corpus contains paragraphs of journalistic texts read by a female voice, which have been segmented into sentences, words and speech sounds. The emotional colouring (anger, joy, sadness) or neutrality of the corpus sentences has been found by using perception tests, see (Altrov, Pajupuu 2012).

---

[1] The quantity degree of a foot is defined as follows: $\sigma_{stressed}$(nucleus+[coda]) / $\sigma_{unstressed}$(nucleus).

[2] See http://peeter.eki.ee:5000/.

To study the speech rate, at least three-word emotional (joyous, sad, angry) or neutral sentences were used, the emotionality or neutrality of which had been confirmed by more than 50% of perception test listeners. To study speech prosody, words of all three quantity degrees with a CV[V]CV structure were picked from the corpus (see Table 1).

Material for speech rate and speech prosody investigation

Emotion	No. of sentences	CV[V]CV words		
		Q1	Q2	Q3
joy	55	45	28	20
sadness	84	62	42	31
anger	77	83	40	24
neutral	98	111	35	37
TOTAL	**314**	**301**	**145**	**112**

2. Method

To find out whether emotions actually affect speech rate, the articulation rate of emotional vs. neutral speech was measured. EESC contains journalistic texts, where all emotional sentences are different, therefore speech sounds per second was considered the most adequate unit of measurement. As from the phonological point of view, long speech sounds represent sequences of two short phonemes (Eek 2008) long sounds were counted as two sounds for speech rate calculations. Emotion results were compared pairwise and with neutral speech. In addition it was investigated whether variation of emotionality is accompanied by speech rate variation within an utterance. For that purpose separate speech rate measurements were conducted on phrase-final word and non-phrase-final words.[3]

To find out whether emotions cause changes in prosody, correlations between the temporal parameters of the vowelcentered word structure CV[V]CV and emotions were modelled, considering the stressedness, phrase position and part of speech of the word. Spech sound durations were measured and the duration ratios of V1/V2, V1/C1 and V2/C2 were computed. Logistic regression and the CART method were used to prove the significance of the effect caused by emotion characteristics in word temporal structure. The corpus has been tagged in the Praat environment. The measurements were analysed and modelled using the SYSTAT12 package.

Klaus R. Scherer (1986) has elaborated a model that predicts the effect of emotion on vocal expression. Scherer's c o m p o n e n t p r o c e s s m o d e l (CPM) considers the psychological and physiological factors associated with emotional vocal expression and demonstrates that there exist certain emotionspecific acoustic patterns. Scherer describes emotion as a series of a d a p t i v e c h a n g e s, which are mutually related. Having received an emotional impulse the nervous system affects the speaker's breathing as well as the muscular tension of the speech organs, which all causes changes in the acoustics of the speech signal. The CPM also predicts what difference emotions will probably cause in the speech rate as compared to normal speech. For a comparison of the results of the present study with those of the CPM and some earlier studies see Ch. 4.

---

[3] As phrase-final lengthening also has an inevitable effect on word temporal structure the final word of a phrase were considered separately from the rest.

3. Results

3.1. Speech rate

The results reveal that emotions do affect the overall speech rate and that of the non-phrase-final words, while the differences across emotions as well as between emotional and neutral speech are statistically significant (see Tables 2 and 3). The rate of pronunciation of the final word of the phrase has no differentiating power, neither between emotions nor between emotional and neutral speech.

Table 2 shows that the overall speech rate is the highest in anger utterances and the lowest in sadness utterances: anger (17.5 sound/s) > joy (17.1 sound/s) > neutral (16.9 sound/s) > sadness (16.6 sound/s). Table 3 indicates that for emotion pairs differences in overall speech rate are statistically significant. As for neutral speech its overall rate only differed significantly from anger utterances, whereas no significant difference was observed between the rates of neutral vs. joy or neutral vs. sadness.

To find out whether speech rate may, depending on changing emotions or emotionality, also vary within a single utterance, separate rate measurements were conducted on phrase-final word and non-phrase-final words. As is demonstrated in Table 2 speech rate differences are the most salient in non-phrase-final words. Here, too, the rate is the highest in anger utterances and the lowest in sadness utterances: anger (18.44 sound/s) > joy (17.62 sound/s) > neutral (17.54 sound/s) > sadness (17.04 sound/s). The differences in the pronunciation rate of non-phrase-final words were statistically significant in emotion pairs as well as between emotional and neutral speech (see Table 3). Here the olny exception is joy, in which case the articulation rate of non-phrase-final words is not significantly different from that of neutral speech.

The average articulation rates of phrase-final word do not differ much (see Table 2), let alone significantly (see Table 3), across emotions.

Table 2

Average pronunciation rates in sounds per second

Emotion	Overall	Non-phrase-final words	Phrase-final word
joy	17.1	17.6	14.0
sadness	16.6	17.0	14.4
anger	17.5	18.4	14.4
neutral	16.9	17.5	14.1

Table 3

Results of ANOVA pairwise analysis of speech rate differences (statistically significant differences, p < 0.05, are highlighted by gray background)

Emotions	p-value for speech rate differences		
	overall	non-phrase-final words	phrase-final word
joy vs. anger	0.010	0.043	0.452
joy vs. sadness	0.039	0.036	0.317
joy vs. neutral	0.557	0.810	0.985
sadness vs. anger	0.001	0.001	0.837
sadness vs. neutral	0.107	0.033	0.237
anger vs. neutral	0.031	0.008	0.370

A closer look at the variation of emotional and neutral speech rates reveals that in non-phrase- inal words the rate of neutral speech varies considerably less than that of emotional speech (see Figure 1).

In phrase-final word speech rate variation does not differ much across emotions (see Figure 2), which means that phrase-final lengthening is realized similarly for all emotions.
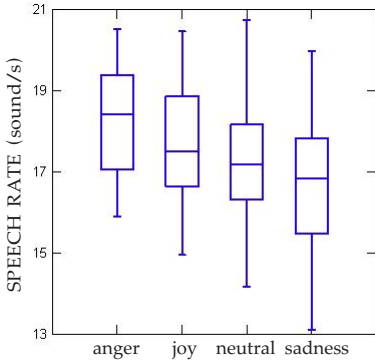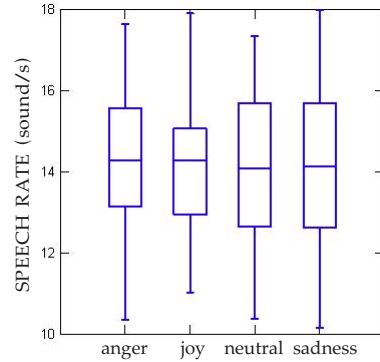

Figure 1. **Speech rate of non-phrase-final words.**

Figure 2. **Speech rate of phrase-final word.**

3.2. Word prosody

Table 4 contains the mean values of the temporal parameters of nearly six hundred words with a vowel-centered structure C1V1[:]C2V2 as distributed across emotions and quantity degrees. This table also contains the most relevant parameters of the duration model: V1, C2, V2 and V1 : V2. Experimental attempts of modelling emotional speech have shown that like in neutral speech (Kalvik, Mihkla 2010), the duration ratios of adjacent speech sounds, V1 : C1 and V2 : C2, are considerably less important than the classical ratio of V1 : V2. Table 4 uses boldface to highlight those parameters whose averages displayed significant statistical differences across emotions (p < 0.05).

Table 4

Temporal parameters of emotional and neutral speech in words with a CV[:]CV structure across three quantity degrees (Q1, Q2, Q3)

Emotion	Q1				Q2				Q3			
	V1	C2	V2	V1 : V2	**V1**	C2	V2	V1 : V2	V1	C2	**V2**	V1 : V2
joy	66	53	90	0.82	**126**	47	78	1.73	141	53	**67**	2.38
sadness	63	53	81	0.83	**132**	51	70	2.11	154	63	**80**	2.20
anger	69	48	91	0.82	**124**	49	71	1.80	153	61	**64**	2.76
neutral	65	52	87	0.80	**145**	54	78	1.94	150	58	**63**	2.64

For the first degree, Q1, the temporal parameters were quite similar, with no significant differences between emotions or emotional vs. neutral speech. More salient differences between emotional and neutral speech could be observed in the Q2 foot. In emotional speech, the duration of the vowel of the stressed syllable, V1, was significantly shorter than in neutral speech (see Figure 3).

However, there were no considerable differences in the durations of the consonant, C2, and vowel, V2, of the unstressed syllable.

Although there is a noticeable difference between the average duration ratios (V1 : V2) of the vowels of the stressed and unstressed syllables in the words with a Q2
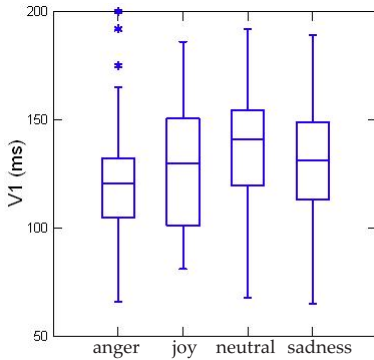
Figure 3. **Distributions of the durations (in ms) of the stressed syllable vowel V1 in words with a Q2 foot for different emotions and neutral speech.**

foot when uttered with different emotions, those differences between the mean values are not significant statistically (p > 0.05). Evidently this is due to the behaviour of the unstressed syllable vowel V2, which differs from that of V1. The highest value of V1/V2 is observed in Q2 words of sad speech. This ratio (2.11) is rather more like the third quantity degree Q3, which in sad speech equals 2.20. Thus, in case the vowel-centered words of the given material are pronounced sadly, their second and third quantity degrees converge, so that the three-way opposition is replaced by a dual one. In the words with a Q3 foot statistically significant differences can be observed between the mean values of the unstressed syllable vowels, V2, if articulated in emotional speech (see Figure 4). For joy, anger and neutral speech, the durations of the unstressed syllable vowel V2 are relatively similar (67, 64 and 63 ms, respectively). A notable V2 lengthening is observed in the Q3 foot in the case of sad speech.
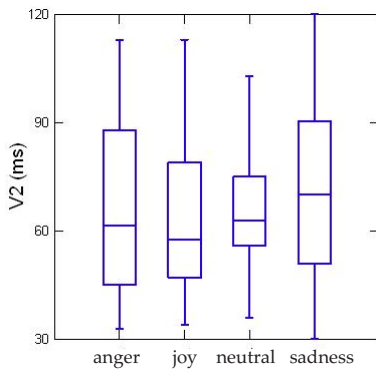

Figure 4. **Distributions of the durations (in ms) of the unstressed syllable vowel V2 in words with a Q3 foot for different emotions and neutral speech.**

According to the speech rate study previously discussed, sad speech was the slowest. What exactly is the role of V2 stretchingin the speech rate decrease of sad speech is pending future studies focused on words with a consonant-centered structure. An analogous situation occurs in the case of anger, where the speech rate is the highest and, in a Q2 foot in words of a vowel-centered structure, the unstressed syllable vowel is the shortest. Again, the significance of that local hange cannot be judged from the vowel-centered structure.

4. Discussion

According to the prediction of CPM, anger makes the speech rate rise. This is confirmed by our results, where the speech rate in anger utterances is indeed higher than in the joy, neutral and sadness ones (see Table 2), while the differences are statistically significant (see Table 3). In the case of anger, rise in the speech rate

has also been observed in some earlier studies, e.g. Murray, Arnott 2008; Iida, Campbell, Higuchi, Yasumura 2003; Juslin, Laukka 2003; Banse, Scherer 1996.

For sadness, the CPM predicts a fall in the speech rate. According to our results sadness also brings a lower speech rate than anger, joy or neutral (see Table 2), but the differences are statistically significant only for sadness *vs*. anger and sadness *vs*. joy (see Table 3). Again, a lower speech rate has been observed in sad speech by many other researchers of emotional speech acoustics, e.g. Murray, Arnott 2008; Yildirim, Bulut, Lee, Kazemzadeh, Deng, Lee, Narayanan, Busso 2004; Banse, Scherer 1996.

For happiness, the CPM prediction reads that the speech rate should be lower, whereas elation should raise it higher. The present study, however, does not treat happiness, joy and elation separately and the category of 'joy' covers happiness and elation as well (see Ch. 1). According to our results the joy utterances are produced more rapidly than the neutral and sad ones (see Table 2), but the diffrence is statistically significant only for joy *vs*. sadness (see Table 3). Again, a higher speech rate has also been observed in joy utterances by several earlier researches, e.g. Murray, Arnott 2008; Iida, Campbell, Higuchi, Yasumura 2003; Banse, Scherer 1996.

As Estonian is unique for its three-way opposition of quantity degrees (Q1, Q2, Q3) it is complicated to achieve a word prosodic comparison with other languages and with universal language models. The classical duration ratios (Eek, Meister 1997), turned out to hold for emotional speech as well, except that in sad speech the duration ratio V1 : V2 of the stressed and unstressed syllables was almost similar for the second and third quantity degrees, due to which the three-way opposition gave way to a dual one. There was no statistically significant correlation between emotions and V1 : V2, which is the main distinctive feature of Estonian phonetic quantity degrees. The general durational model of the Estonian speech sounds (Mihkla 2007), does not include emotion as an argument feature. In the words with a vowel-centered structure emotion was only significant for the stressed syllable vowel V1 in Q2 foot words and for the unstressed syllable vowel V2 in Q3 foot words. Future experiments should deal with the possible effect of emotions on words with a consonant-centered structure CVC[C]V and on monosyllables, with a view of an overall durational model of emotional speech.

The results of the present study once again confirm that speech rate tends to follow a universal pattern, notably, it is higher in utterances of anger and joy, whereas sadness makes it fall. Of the three emotions analysed, anger is the farthest from neutral speech; the difference is statistically significant.

5. Conclusion

The present analysis of read Estonian emotional utterances has proved that emotions do affect the speech rate. According to our measurements, the overall speech rate was higher than neutral in utterances of joy and anger, whereas sadness made it drop lower than neutral. However, the difference was statistically significant only between anger and neutrality. As for emotion pairs the speech rates were always different, and the difference was statistically significant.

Separate analyses were carried out on phrase-final word and non-phrase-final words. The differences were more salient in the latter, showing the highest rate for angry utterances and the lowest for sad ones. Almost all speech rate differences in emotion pairs as well as between emotions and neutral speech proved statistically significant, except for joy vs. neutral, which difference was not significant in non-phrase-final words. In phrase-final word the articulation rates did not substantially differ. Consequently the influence of emotions on speech rate is confined to non-phrase-final words.

The working hypothesis of the possible influence of emotions on the temporal characteristics of the words with a vowel-centered structure CV[V]CV and on the duration ratio of the stressed and unstressed syllables was but partly confirmed. Emotions did not have a significant influence on the main distinctive feature, V1:V2, of Estonian quantity degrees, however, the average durations of V1 in words with a Q2 foot differed across emotions, as well as the duration of the vowel V2 of the unstressed syllable in words with a Q3 foot.

In sad speech, the parameters of the second and third quantity degrees converged, so that the typical three-way quantitativ opposition was reduced to a dual one. But this was proved just for words with a vowel-centered structure. To reach a final conclusion about the influence of emotions on Estonian word prosody and the emotion-induced local changes in the speech rate our research should also cover the correlation of emotions with the temporal parameters of words with a consonantcentered CVC[C]V structure and of monosyllabic words.

Addresses

Kairi Tamuri
Institute of the Estonian Language
E-mail: kairi.tamuri@eki.ee

Meelis Mihkla
Institute of the Estonian Language
meelis.mihkla@eki.ee

R E F E R E N C E S

A l t r o v, R., P a j u p u u, H. 2012, Estonian Emotional Speech Corpus: Content and options. – Variation and Change in Spoken and Written Discourse. Perspectives from Corpus Linguistis, Amsterdam [forthcoming]

B a n s e, R., S c h e r e r, K. R. 1996, Acoustic Profiles in Vocal Emotion Expression. — Journal of Personality and Social Psychology 70, 614—636.

B r a u n, A., O b a, R. 2007, Speaking Tempo in Emotional Speech — a Cross-Cultural Study Using Dubbed Speech. — ParaLing'07, 77—82.

E e k, A. 2008, Eesti keele foneetika I, [Tallinn].

E e k, A., M e i s t e r, E. 1997, Simple Perception Experiments on Estonian Word Prosody: Foot Structure vs. Segmental Quantity. — Estonian Prosody: Papers from a Symposium, Proceedings of the International Symposium on Estonian Prosody, Tallinn, 71—99.

E e k, A., M e i s t e r, E. 2003, Foneetilisi katseid kvantiteedi alalt. — KK, 815—837, 902—916.

I i d a, A., C a m p b e l l, N., H i g u c h i, F., Y a s u m u r a, M. 2003, A Corpus-Based Speech Synthesis System with Emotikon. — Speech Communication 40, 161—187.

J u s l i n, P. N., L a u k k a, P. 2003, Communication of Emotions in Vocal Expression and Music Performance. Different Channels, Same Code? — Psychological Bulletin 129, 770—814.

K a l v i k, M-L., M i h k l a, M. 2010, Modelling the Temporal Structure of Estonian Speech. — Human Language Technologies — The Baltic Perspective. Proceedings of the Fourth International Conference Baltic HLT 2010, Amsterdam, 53—60.

K a l v i k, M-L., M i h k l a, M., K i i s s e l, I., H e i n, I. 2010, Estonian: Some Findings for Modelling Speech Rhythmicity and Perception of Speech Rate. — Text, Speech and Dialogue, Berlin—Heidelberg, 314—321.

K r u l l, D. 1993, Word-Prosodic fFeatures in Estonian Conversational Speech. Some Preliminary Results. — PERILUS (Phonetic Experimental Research, Institute of Linguistics, University of Stockholm) XVII, Stockholm, 45—54.

L a v e r, J. 1994, Principles of Phonetics, Cambridge.

L e h i s t e, I. 1960, Segmental and Syllabic Quantity in Estonian. — American Studies in Uralic Linguistics 1, Bloomington, 21—28.

— — 1997, Search for Phonetic Correlates in Estonian Prosody. — Estonian Prosody: Papers from a Symposium, Proceedings of the International Symposium on Estonian Prosody, Tallinn, 11—35.

L i i v, G. 1961, Eesti keele kolme vältusastme kestus ja meloodiatüübid. — KK, 412—424, 480—490.

L i p p u s, P., P a j u s a l u, K., A l l i k, J. 2007, The Tonal Component in Perception of the Estonian Quantity. — Proceedings of the 16th International Congress of Phonetic Sciences, Saarbrücken,1049—1052.

M i h k l a, M. 2007, Modelling Speech Temporal Structure for Estonian Text-to-Speech Synthesis: Feature Selection. — Trames 11, 284—298.

M u r r a y, I. R., A r n o t t, J. L. 2008, Applying an Analysis of Acted Vocal Emotions to Improve the Simulation of Synthetic Speech. — Computer Speech and Language 22, 107—129.

R o s s, J., L e h i s t e, I. 2001, The temporal Structure of Estonian Runic Songs, Berlin—New York.

S a g i s a k a, Y. 2003, Modeling and Perception of Temporal Characteristics in Speech. — Proceedings of 15th International Congress of Phonetic Sciences, Barcelona, 1—6.

S c h e r e r, K. R. 1986, Vocal Affect Expression. A Review and a Model for Future Research. — Psychological Bulletin 99, 143—165.

— — 2003, Vocal Communication of Emotion. A Review of Research Paradigms. — Speech Communication 40, 227—256.

T a m u r i, K. 2010, Kas pausid kannavad emotsiooni? — Eesti Rakenduslingvistika Ühingu Aastaraamat 6, 297—306.

— — 2012, Kas formandid peegeldavad emotsioone? — Eesti Rakenduslingvistika Ühingu Aastaraamat 8, 231—243.

Y i l d i r i m, S., B u l u t, M., L e e, C. M., K a z e m z a d e h, A., D e n g, Z., L e e, S., N a r a y a n a n, S., B u s s o, C. 2004, An Acoustic Study of Emotions Expressed in Speech. — Proceedings of InterSpeech 2004, Jeju Island, 2193—2196.

КАЙРИ ТАМУРИ, МЭЭЛИС МИХКЛА (Таллинн)

ЭМОЦИИ И ВРЕМЕННА́Я СТРУКТУРА РЕЧИ

Авторы статьи пытаются выяснить, влияют ли эмоции на временну́ю структуру эстонской речи. С этой целью они ставят перед собой два вопроса: а) воздействуют ли эмоции на темп речи и б) проявляются ли они в просодии слова. Акустической базой исследования послужил корпус эстонской эмоциональной речи Института эстонского языка. Для ответа на первый вопрос сначала измерялась артикуляционная скорость эмоциональных изречений (число звуков в секунду), а полученные результаты сравнивались с нейтральной речью. Выяснилось, что темпы эмоциональной и нейтральной речи различаются между собой и различия эти статистически существенны. Затем изучались соотношения временны́х признаков у слов, имеющих структуру CV[V]CV (преобладание гласных), с учетом ударности слова, его позиции в предложении и морфологической категории. Измерялись длительности звуков, вычислялись соотношения этих длительностей и в ходе статистического анализа комбинировались различные признаки. Результаты показали некоторое различие во временны́х признаках степеней долготы Q2 и Q3, а также слияние второй и третьей долготы в грустной речи.