CROSS-LANGUAGE PERCEPTION OF EMOTIONAL CHILDREN'S SPEECH IN GERMAN AND RUSSIAN

Karina Evgrafova[a], Pavel Skrelin[a], Daria Shatalova[a]
[a]Deparment of Phonetics, Saint-Petersburg State University
evgrafova@phonetics.pu.ru; skrelin@phonetics.pu.ru; daria.s.shat@gmail.com

ABSTRACT

The paper concerns universal and language-specific aspects of emotion perception in children's speech. Three experiments were carried out to investigate differences and similarities in the assessment of emotions by German and Russian adult listeners. The corpora of German and Russian emotional children's speech were employed in the first and second experiments. In the third experiment German and Russian 'delexicalised' utterances were used. They were selected from the both corpora and added white noise to. Thus the semantic content was removed while the prosodic features stayed intact. The experiment was aimed at analyzing recognition strategies when listeners rely only on prosody while segmental level information is not present.

The experiments revealed similar and different patterns of assessing emotions in children's speech in German and Russian.

The study contributes to better understanding of cross-lingual human emotion perception and the role of verbal, segmental and suprasegmental components in emotion recognition.

Keywords: emotional speech, children's speech, speech perception, cross-language studies.

1. INTRODUCTION

The expression of emotion in speech has been addressed in a substantial number of research. [3-6]. The existing studies focus on various facets of the problem which include the role of intonation patterns, voice quality, rhythm, and major acoustic parameters in conveying emotions. The results of the studies to a great extent depend on a type of the data which can be natural, elicited, fully acted or synthetic emotional speech. The analysis of the natural expression of emotions, however, is very rare as collecting of authentic emotions can be rather problematic. [8-10] Besides, the usage of the prosodic means expressing emotions can vary across languages. [1-2], [8], [9]

The existing corpora and databases of emotional speech are exploited for investigating the recognition of emotion categories and issues related to emotion perception. On the one hand, these topics are significant linguistically, on the other hand, they are of great importance for applications in the areas of human-robot communication and machine learning. [7], [9-10]

As to research on cross-language differences of emotional speech perception, there is a relative paucity of studies. In general, it is shown that emotions can be recognized with relative accuracy even in unknown languages. However, the level of accuracy is higher in native languages. Considerable differences between native and non-native speakers when judging emotions for a language are observed on both valence and arousal dimensions. Strongly positive and strongly negative emotions in a language can be confused by non-native speakers due to the absence of lexical prompt and the similarity of prosodic features associated with them. Thus, the data suggest that confusion patterns of emotion perception are not symmetrical across languages. [8]

The aim of our study was to detect universal and language-specific patterns of perceiving emotions in the speech of German and Russian children by adult native speakers of the two languages. Particularly, we were interested in the way listeners identified emotions in case when the semantic content was not present. The hypothesis was that lexixal and segmental level components would be strongly required for successful identification of emotions.

The human emotion expression is based both on universal mechanisms and cultural conventions. The main motivation for analyzing children's speech was the consideration that children’s emotional expression is spontaneous as their behaviour is practically not determined by social conventions.

That is why the emotional children's speech is suitable for investigating direct correlation between acoustic characteristics of speech and emotional verbal reactions.

2. MATERIAL

The study was based on the speech material of two corpora: the pre-existing FAU Aibo Emotion Corpus and the Corpus of Russian Children's Emotional Speech which was specially recorded for the objectives of the study.

The audio data and emotion label files of FAU Aibo Emotion Corpus were kindly made available for the purpose of our study by the developers. It is a corpus of German spontaneous speech with recordings of children at the age of 10 to 13 years communicating with a pet robot [10]. The general framework for the corpus is child–robot communication and the elicitation of emotion-related speaker states. The robot is Sony's (doglike) robot Aibo.

The Russian corpus was collected strictly according to same scenario and conditions. The model of a robot dog was different though. The corpora vary also with respect to the size and number of speakers: 9 hours and 51 speakers (German) and 5 hours, 15 speakers (Russian). Despite these differences, the both corpora can be considered to be parallel as they contain the expressions of an identical set of emotions elicited in very similar conditions.

3. METHOD

The speech material of the above-mentioned corpora was employed in three types of perception experiments.

3.1. Perception experiment 1

The aim of the first experiment was to obtain the evaluations of German emotion utterances from the Russian listeners in order to compare them with the ones which had been previously made by Germans. We predicted the mismatch of the evaluations among the German and Russian listeners which resulting from the differences in prosodic systems of the languages.

Thirty Russian adults aged 25-35 were asked to listen to 111 files picked out from the Aibo Emotion Corpus. The files contained the samples of all the types of emotional states labeled in the Aibo corpus. All the listeners were native speakers of Russian with no knowledge of German. The stimuli were short utterances that had been pronounced by German children in situations which evoked emotional verbal reactions. The listeners had to make a decision which emotion was expressed in each phrase. They were asked to select from the list of 11 types of emotional states. The same list of emotions had been used by German annotators of the Aibo corpus. It included the following types of emotional states:

- neutral,
- emphatic.
- bored
- surprised
- touchy
- hesistant
- motherese
- joyful
- reprimanding
- angry
- other. [10]

The listeners were free to listen to each stimulus as often they like to make their choice.

3.2. Experiment 2.

The second experiment was designed to check if there would be recognition confusion among the native speakers if both semantic and prosodic components are provided. For this purpose the same group of Russian listeners was involved to obtain the evaluations of emotional utterances in their native language.

The stimuli were 40 utterances from the Russian corpus containing the samples of all the types of emotional states which had been taken into account in the previous experiment.

3.3. Perception experiment 3

In the third experiment the data from the two corpora were exploited. We selected 20 German and 20 Russian utterances and added white noise to the signal in order to make them "delexicalised". Thus the semantic content was removed while the prosodic features stayed intact.

The experiment was aimed at analyzing the recognition strategies if the listeners rely only on prosodic features while the lexical meaning and any segmental level information is not present. It should be also noted that the utterances which were selected for the third experiment (both German and Russian) were the ones which had been evaluated unanimously in terms of emotion category by all the listeners.

In the sections below the results of the experiments and their discussion are presented.

4. RESULTS

4.1 Germans and Russians perceiving German speech: mismatch in the interpretation of emotions

The comparison of the emotion evaluations by the Russian listeners and German labelers showed that there are only 5 frequently recognized emotions in the Aibo corpus. Besides, their sets vary across the German and Russian listeners: ***neutral, emphatic, bored, surprised*** *and* ***touchy*** and ***surprised, angry, joyful, scared*** *and* ***sad*** respectively. They arw arranged according to their frequency. The rare types of emotions were not analyzed in the study. The table 1 shows the frequency of the evaluations of the Aibo corpus by the Russian listeners.

Table 1: Frequency of state evaluations: Russian perceiving German speech.

Emotional states	%
surprise	31.5
anger	25.5
joy	24
sadness	9.5
fear	9.5

One can see that the category ***surprised*** was very often chosen by the Russian listeners. (This category has positive valence in the both corpora). In a number of cases an utterance that had been labeled as ***neutral*** by the German annotators was consistently evaluated as ***surprised*** by the Russian listeners.

It should be also mentioned that the Russian listeners were more specific in judging emotions. They used the categories ***sad*** and ***scared*** while the German listeners had evaluated the same utterances as ***neutral*** and ***emphatic*** respectively or ***other***.

The table 2 shows the confusion matrix of the evaluations done by the Russian listeners and allows comparing the intended emotions and interpreted emotions in the German utterances. The intended emotions are considered to be the ones labeled by the native annotators.

Table 2: Russian perceiving German speech. The confusion matrix: intended vs. interpreted emotions (in percentage).

German	Joy	Anger	Sadn.	Surpr.	Fear
Joy	**41**	4	6	34	15
Anger	24	**46**	8	20	2
Sadness	13	8	**46**	33	0
Surpr.	16	25	9	**50**	0
Fear	0	2	3	18	**77**

One can see that in most cases the Russian evaluations were similar to the German ones. However, the correct recognition rate is 40-50%, except for the ***fear*** category which is relatively high and comes up to 77%. This can indicate that the emotions having the high degree of both valence and arousal are expressed by common means in terms of acoustics across the languages.

4.2 Russians perceiving Russian speech: recognition patterns

The experiment based on the Russian speech showed the recognition patterns of Russians perceiving the Russian emotional utterances. The set of most frequently perceived emotions remained consistent while the frequency rate turned to be different. The table 3 shows the percentage of emotion categories recognition. The most positive emotions had the highest rate and the most negative the lowest.

Table 3: Frequency of state evaluations: Russian perceiving Russian speech.

Emotional states	%
joy	35.5
surprise	22.5
sadness	21.5

fear	13
anger	7,5

The table 3 shows the confusion matrix which compares the evaluations done by the listeners (interpreted emotions) and assessments done by the corpus developers which were based on experiment protocols and video recordings (intended emotions).

Table 3: Russian perceiving Russian speech. The confusion matrix: intended vs. interpreted emotions (in percentage).

Russian	Joy	Anger	Sadn.	Surpr.	Fear
Joy	**62**	0	4	30	4
Anger	19	**48**	23	4	6
Sadness	0	1	**79**	12	8
Surpr.	7	3	0	**90**	0
Fear	0	10	18	8	**64**

As it shown in the matrix, the most correctly recognized emotions were ***surprise*** (90%) and sadness (79%) and the least recognized one was ***fear*** (48%). On the whole, the recognition rate in the native speech was much higher in comparison with the one of non-native speech.

4.3 Russians perceiving Russian and German "delexecalized" utterances

The third experiment yielded the following results. The correct recognition was significantly hampered. The categories ***surprise*** and ***sadness*** turned out to be most recognizable: 71% and 64% respectively which is comparable with the recognition rate in "normal" speech. However, the listeners reported having strong difficulties in evaluating emotions in "delexecalized" utterances and admitted their decisions being random.

Table 2: Russian perceiving Russian and German "delexecalized" utterances. The confusion matrix (in percentage).

	Joy	Anger	Sadn.	Surpr.	Fear
Joy	**17**	20	20	***31***	12
Anger	18	**29**	3	***32***	18
Sadness	0	0	**64**	9	27
Surpr.	19	0	0	**71**	10
Fear	0	0	37	25	**38**

5. DISCUSSION AND CONCLUSIONS

The analysis of the cross-language experimental results showed that strategies of the emotion recognition in children's speech in German and Russian are not symmetrical in terms emotion categories perceived and the correct recognition rate. The experiments on the identification of emotions in native speech (both for German and Russian) yielded very high correct recognition rate which ranges from 50% to 90% (the perception is based on analyzing semantic and prosodic components). The recognition of emotion categories in non-native speech was normally below 50% for the Russian listeners (semantic component is not present). To find out if it is true we intend to conduct one more perception experiment.

The experiment with "delexecalized" utterances tested our hypothesis that not only lexical and prosodic components matter, but also segmental characteristics such as the set of phonemes, number and type of syllables. In our experiment all these characteristics were masked with white noise which resulted in poor and random emotion recognition. The study contributes to better understanding of cross-lingual human emotion perception and the role of verbal, segmental and suprasegmental components in emotion recognition.

6. REFERENCES

[1] Albas, D.C., McCluskey, K.W., Albas, C.A. 1976. Perception of the emotional content: A comparison of two Canadian groups. *J. of Cross-Cultural Psychology* 7(4), 481-490.

[2] Abelin, A. & Allwood, J. 2000. Cross-linguistic Interpretation of Emotional Prosody. In R. Cowie, E. Douglas-Cowie & M. Schröder (Eds.) *Proceedings of the ISCA Workshop on Speech and Emotion.* Belfast, Ireland.http://www.qub.ac.uk/en/isca/proceeding

[3] Banzieger, T., Sherer, K.R. 2005. The role of intonation in emotional expressions. *Speech Communication* 46 (3-4), 252-267.

[4] Chen, A. J. 2005. Universal and language-specific perception of paralinguistic intonational meaning. PhD thesis. Utrecht: LOT.

[5] van Bezooijen, R., Ottoo, S.A., Heenan, T.A. 1983. Recognition of vocal expression of emotions: A three-nation study to identify universal characteristics. *J. of Cross-Cultural Psychology* 14 (4), 387-406.
[6] Frick, R. W. 1985. Communicating emotion: The role of prosodic features // I V. 97. N 3. P. 412—429.
[7] Hagen, A., Pellom, B., and Cole, R. Highly accurate children's speech recognition for interactive reading utors using subword units. *Speech Communication* 49(12):861–873, 2007.
[8] Pfitzinger, H., Amir, N. 2011. Cross-langiage perception of Hebrew and German authentic emotional speech. *Proceedings of IChS XVII.* 1586-1589.
[9] Steidl, S., Schuller, B., Batliner, A. and Seppi, D. The hinterland of emotions: Facing the open-microphone challenge. In Proc. of ACII, pages 690–697, 2009.
[10] Steidl, S, 2009. Automatic classification of emotional-related user states in spontaneous children's speech. Erlangen,