

UNIVERSAL AND LANGUAGE-SPECIFIC PATTERNS OF PERCEIVING EMOTIONS IN CHILDREN'S SPEECH: A CROSS-LANGUAGE GERMAN-RUSSIAN STUDY

Karina Evgrafova¹, Pavel Skrelin¹, Tatjana Chukajeva¹, Elmar Neuth²

Saint Petersburg University¹, Friedrich-Alexander University Neurnberg-Erlangen²
evgrafova@phonetics.pu.ru

ABSTRACT

The paper describes research on universal and language-specific patterns of perceiving emotions in children's speech in German and Russian by the native speakers of the two languages. The children's speech presents a reliable material for analysing emotions as children's emotional expression is spontaneous and their verbal behaviour is least determined by social conventions. Two parallel corpora containing the expressions of an identical set of emotions (FAU AIBO Emotion Corpus and Corpus of Russian Children's Emotional Speech) were used. The emotions were elicited in very similar conditions. The data from four types of cross-language experiments were obtained: Germans evaluating German speech, Germans evaluating Russian speech, Russians evaluating Russian speech, Russians evaluating German speech. The confusion matrices of emotion recognition in all types of experiments were compared. The universal and language specific patterns of perceiving emotions were detected and described in the paper

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

1. INTRODUCTION

Emotional speech has been focused in many studies [3-6] which address a wide range of problems. They include a role of intonation patterns, voice quality, rhythm and major acoustic parameters in expressing and perceiving emotions. The existing studies employed different types of emotional speech 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. [10-12] The existing corpora and databases of emotional speech are exploited in speech emotion recognition and emotional speech synthesis systems. 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. [9], [11-12]

We are mostly interested on cross-language differences of emotional speech perception.

It is found out that prosodic patterns employed for expressing different emotions can vary across languages. [1-2], [9], [11] It may result in different emotion recognition strategies in cross-language perception. Data suggest that confusion patterns of emotion perception are not symmetrical across languages. [10]

Although emotions can be identified with relative accuracy even in unknown languages, the level of accuracy is higher in native languages. Native and non-native speakers judging emotions for a language demonstrate different recognition patterns on both valence and arousal dimensions. The correct emotion recognition may be hampered for non-natives speakers due to the absence of lexical prompt and the dissimilarity of prosodic features associated with different emotions in different languages.

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 lexical and segmental level components would be strongly required for successful identification of emotions. However, there should be some universal ways for expressing emotions across the two languages which allow high accuracy in perceiving certain emotions. We were interested in finding out which types of emotions are normally confused and which ones are identified correctly even when the lexical prompts are absent.

The given paper sums up the results of the research which consisted of six experimental parts. Some results of the previous work have been reported earlier. [5, 6]

2. MATERIAL

The human emotional verbal behavior is based both on universal psychophysiological mechanisms and cultural conventions. Our main motivation for analyzing children's speech was the consideration that children's emotional behaviour is spontaneous as it is less determined by social conventions compared to that of adults'.

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

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 the 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 six types of cross-language perception experiments.

Subjects

30 Russians native speakers (15 male + 15 female), aged 25-35, no knowledge of German, no parental experience.

30 German native speakers (15 male + 15 female) adults, aged 25-35, no knowledge of Russian, no parental experience.

3 set of stimuli

45 German utterances

45 Russian utterances

40 modified (20 German + 20 Russian)

The selected utterances (both German and Russian) were the ones which presented no ambiguity in terms of emotion category.

Procedure

Task - listen to an audio file and identify the emotion expressed (marking the choice in the emotion circle in fig. 1).

Description of emotions

Joyful – the child enjoys the play or finds something funny

Surprised – the child is (positively) surprised

Motherese – the child addresses the toy in the way parents address their babies when they are well-behaving

Reprimanding – the child is reproachful, “wags the finger”

Emphatic – the child speaks in an accentuated way, but shows no specific emotion

Neutral – the child shows no emotion

Touchy – the child is slightly irritated

Angry – the child is annoyed or clearly angry

Other – the emotion is different from any on the list, suggest your variant.

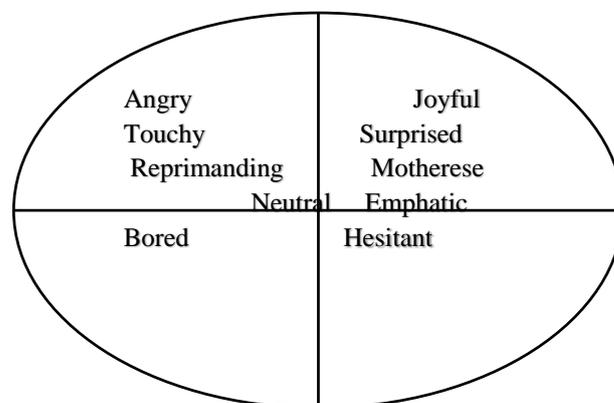


Fig. 1 The emotion circle.

3.1. Perception experiment 1

The aim of the first experiment was to obtain the evaluations of German emotional utterances from the Russian listeners.

The stimuli were short utterances (one or two words long) 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 having only prosodic and segmental prompts. They were asked to

select from the emotion circle (fig.1). The listeners were free to listen to each stimulus as often they liked.

3.2. Perception 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 (meaning of the words + intonation clues) were 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 45 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 using PRAAT procedures. Thus we removed the semantic content 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.

3.4. Perception experiments 4-6

The group of German listeners was involved into the same types of perception experiments. They also evaluated three sets of stimuli: 45 Russian utterances, 45 German utterances and 40 mixed utterances. The conditions of all the experiments were identical. In the sections below the results of the experiments and their discussion are presented.

4. RESULTS

4.1 Russian listeners' evaluations of three sets of stimuli: recognition patterns

The comparison of the evaluations done by the Russian listeners and German listeners showed that there are only 5 frequently recognized emotions in the both corpora. They are *joy*, *anger*, *neutral*, *surprise*, *emphatic*. The rarely occurring types of emotions were not analyzed in the study.

The table 1 shows the confusion matrix of the evaluations of the three sets of stimuli (**German**, **Russian**, **modified**) done by the Russian listeners. The left column shows intended emotions (assessments done by the corpus developers which were based on experiment protocols and video

recordings). The row above shows interpreted emotions (obtained during the perception experiment).

Table 1: Russian listeners' evaluations. The confusion matrix: intended vs. interpreted emotions (in percentage).

German	Joy	Anger	Neutral	Surpr.	Emph
Joy	43	4	6	32	15
Anger	21	48	8	23	0
Neutral	8	23	36	33	0
Surprise	7	20	0	73	0
Emphatic	26	4	0	50	20
Russian	Joy	Anger	Neutral	Surpr.	Emph
Joy	74	0	0	19	7
Anger	7	62	0	4	27
Neutral	0	0	83	7	10
Surprise	7	2	0	91	0
Emphatic	0	13	8	12	67
Modified	Joy	Anger	Neutral	Surpr.	Emph
Joy	21	18	17	24	20
Anger	16	12	15	44	13
Neutral	12	4	37	20	27
Surprise	20	10	1	59	10
Emphatic	0	7	13	27	53

4.2 German listeners' evaluations of three sets of stimuli: recognition patterns

Table 2: German listeners' evaluations. The confusion matrix: intended vs. interpreted emotions (in percentage).

German	Joy	Anger	Neutral	Surpr.	Emph
Joy	82	0	0	8	10
Anger	0	92	0	0	8
Neutral	2	0	98	0	0
Surprise	15	0	0	80	5
Emphatic	26	5	0	5	64
Russian	Joy	Anger	Neutral	Surpr.	Emph
Joy	41	7	0	22	30
Anger	24	37	0	9	30
Neutral	18	17	45	6	14
Surprise	16	8	0	49	27
Emphatic	32	13	0	7	48
Modified	Joy	Anger	Neutral	Surpr.	Emph
Joy	23	5	0	32	40
Anger	26	29	0	8	37
Neutral	24	8	30	9	29
Surprise	18	11	0	34	37
Emphatic	17	19	8	28	28

5. DISCUSSION AND CONCLUSIONS

One can observe the mismatch of the evaluations among the German and Russian listeners which must be due to the differences in prosodic systems of the languages. The category *surprised* was very often chosen by the Russian listeners. In a number of cases an utterance evaluated as *neutral* by the German listeners 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 evaluated the same utterances as *neutral* and *emphatic* respectively or *other*.

On the whole, the correct cognition rate in the native speech was much higher in the comparison with that of non-native speech. As the matrices show, there were cases when one type of emotion was perceived as another. This observation is true for *joy* and *anger* which were often recognized as *surprise*.

The most correctly recognized emotions in all types of material was *surprise* (up to 80% in non-native speech). The least recognized by German listeners was *anger* (48% in non-native speech) while Russian listeners had difficulties with recognizing the category *emphatic* (20% in non-native speech).

The correct recognition of modified signals was significantly hampered. However, the category *surprise* turned out to be most recognizable: up to 57% which is comparable with the recognition rate in “non-native” speech but lower than correct recognition in “normal native” speech. However, the listeners reported having strong difficulties in evaluating emotions in “delexicalized” utterances and admitted their decisions being random.

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 of emotion categories perceived and the correct recognition rate. The identification of emotions in native speech (both for German and Russian) showed very high correct recognition rate which ranges from 50% to 98% (the perception is based on analyzing semantic and prosodic components). The recognition of emotion categories in non-native speech was normally below 50% (semantic component was not present).

The experiment with modified 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.

7. 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] Karina Evgrafova, Pavel Skrelin, Daria Shatalova Cross-Language Perception of Emotional Children's Speech in German and Russian // Proceedings of the 18th International Congress of Phonetic Sciences. Glasgow, UK, 2015.
- [6] Karina Evgrafova, Pavel Skrelin, Elmar Noeth Cross-language Research on Expression and Perception of Emotional Children's Speech in German and Russian // 23. Linguistik- und Literaturtage in St. Petersburg, Russland: in St. Petersburg, Russland: und darüber hinaus. Abstracts. 2015.
- [7] 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.
- [8] Frick, R. W. 1985. Communicating emotion: The role of prosodic features // I V. 97. N 3. P. 412—429.
- [9] Hagen, A., Pellom, B., and Cole, R. Highly accurate children's speech recognition for interactive reading tutors using subword units. *Speech Communication* 49(12):861–873, 2007.
- [10] Pfitzinger, H., Amir, N. 2011. Cross-language perception of Hebrew and German authentic emotional speech. *Proceedings of IChS XVII*. 1586-1589.
- [11] 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.
- [12] Steidl, S, 2009. Automatic classification of emotional-related user states in spontaneous children's speech. Erlangen,