

ASPECTS OF CULTURAL COMMUNICATION IN RECOGNIZING EMOTIONS

Rene Altrov

Institute of the Estonian Language, Tallinn

Abstract. The study addresses cultural influence in the recognition of moderately expressed emotions in a second language (L2) and foreign speech. The web-based listening test consisted of context-free sentences drawn from the Estonian Emotional Speech Corpus. The task was to recognize the emotion (joy, anger, sadness) or neutrality of each sentence without seeing the speaker. Three adult groups participated: (1) 36 Estonians, with Estonian as mother tongue; (2) 16 highly educated Russians living in Estonia, with Russian as their mother tongue and Estonian as a second language; (3) 16 highly educated Russians living in Russia, with Russian mother tongue and no knowledge of Estonian. The results showed a significant difference between Estonians and Russians living in Estonia in their recognition of joy and neutrality; Russians living in Russia differed significantly from Estonians and Russians living in Estonia on all emotion scores. This confirms that cultural norms are mastered through interaction: to recognize vocal emotions expressed in another language it is necessary to live in the culture and communicate in its language.

Keywords: emotional speech, cross-cultural comparison, L1, L2, Estonians, Russians

DOI: 10.3176/tr.2013.2.04

1. Introduction

Emotions play a central role in interpersonal relations. Cowie and colleagues (2011:11) have described an experiment for HUMAINE¹, confirming, among other things, that hardly any act of communication takes place without the involvement of emotions. The experiment used video recordings of people dealing with challenges in a novel outdoor environment. The company's psychologist selected about 5 hrs of material that she regarded as representative of the types of experience found in the material. After that, four testers watched the videotapes

¹ The HUMAINE Association is a professional, worldwide association for researchers in emotion-oriented/affective computing.

and indicated which of the three categories given best described their impression of the person being recorded: 79% of the recorded material was perceived as moderately emotional, 14% as strongly emotional and only 7% as unemotional.

It is well known that in order to understand messages in verbal communication it is important to pay attention not only to what is said but also to how it is said. Accurate recognition of the emotional intentions of others contributes to the efficiency of communication by enabling better prediction and interpretation of the behaviour and emotional states of others, as well as better formation and maintenance of social bonds. This makes the production, perception, interpretation and response to emotional signals an important element of our interpersonal lives (see also Mitchell 2007, Soto and Levenson 2009, Van Rijn et al. 2005).

There is more than one way to communicate an emotional message. In parallel with the words actually spoken, emotions are communicated by body movements, facial expressions, changes in voice quality, and so on. If one can see the speaking person, comprehension may be relatively easy, yet there are various situations where emotion has to be guessed from vocal cues only (a telephone conversation, listening to the radio, overhearing distant talk, or human-machine interaction); hence, the importance of the ability to recognize emotions from the voice. Moreover, it is important to be able to interpret vocal emotions even if the emotional meaning of the words differs from that conveyed by the tone of voice (Cheang and Pell 2008).

2. Cultural influence on emotions

Recent studies have confirmed the general view of culture as an important factor in determining how we express and understand emotions and what emotional signals are socially acceptable. Therefore, a shared culture can facilitate emotional communication, while cultural differences can make understanding another person's emotional state more difficult (Soto and Levenson 2009).

In normal conversation, verbal, vocal and facial emotional cues are transmitted and received simultaneously. But if we cannot see the speaker well enough or at all, the emotions have to be decided just by vocal channels. The results of recent cross-cultural studies focused on emotion recognition from nothing but voice have demonstrated the influence of culture on emotion recognition, as well as some universal tendencies.

In their study on the recognition of basic emotions, Bryant and Barrett (2008) have shown that people are capable of the vocal recognition of emotions from quite disparate cultures, which suggests that despite cultural variations it is possible for individuals to communicate across wide cultural boundaries. Their experiment involved 28 students of the University of California, whose mother tongue was English, and 23 adults from two Shuar villages (Amazonian Ecuador), whose first and primary language was Shuar. Both groups were asked to listen to ten English utterances produced by two native English speakers using five basic

emotions – anger, happiness, sadness, fear and disgust. For the native English participants the utterances were content-filtered by removing the words, but retaining the global prosodic information. A comparison of the results showed that both groups were capable of recognizing the basic emotions without content information, scoring better than chance probability, while the only emotion scoring no higher than chance probability was disgust as recognized by the Shuar group.

Thompson and Balkwill (2006) also supported the view that emotional prosody is decoded by a combination of universal and culture-specific cues. They had 20 English speakers judge the emotive intent of utterances spoken by male and female speakers of English, German, Chinese, Japanese and Tagalog. The content was neutral, but the utterances were spoken in a way that communicated each of four intended emotions: joy, sadness, anger and fear. Recognition accuracy was better than chance probability for all emotions in all languages. Across languages, sadness and anger were more accurately recognized than joy and fear. The results showed that there was an in-group advantage for decoding emotional prosody, with the highest recognition rates being for English utterances and lowest for Japanese and Chinese ones. The conclusion was that the capability of listeners to decode the vocal expression of emotions in unfamiliar languages demonstrates the universality of some prosodic cues for emotions. The evidence of an in-group advantage, however, is indicative of cultural factors influencing the production and/or perception of emotional prosody.

Pell et al. (2009) have also corroborated the influence of universal and culture-specific factors on the recognition of vocal emotions. In their study, 61 adult monolingual speakers of Argentine Spanish were asked to determine five basic emotions (anger, disgust, fear, sadness, joy) and neutrality in acted pseudo-utterances (i.e. utterances without semantic content) in Spanish and in three foreign languages – English, German and Arabic. The Spanish participants' recognition scores were better than chance probability for all the languages involved, although they performed significantly better in their native language ('in-group advantage'). Similar to Thompson and Balkwill (2006), their results also indicated that independent of language, vocal emotion recognition tended to be highest for anger (73%) and sadness (66%), and the lowest for disgust (42%). The only emotion where there was a clear difference between native and foreign recognition was joy, which scored 89% when native Spanish was used compared to a mere 32% in the case of the English test material.

Scherer et al. (2011), who generalized studies on the determination of the emotion of vocal expression using speech materials from different languages, stated that "all the studies in this area have found better-than-chance accuracy of cross-cultural recognition". The best recognition scores belonged to anger and sadness. The accuracy percentage was somewhat lower for non-Westerners trying to recognize Western emotions.

Although a better-than-chance probability of cross-cultural emotion recognition suggests the existence of universal aspects of emotional expression, it does not exclude cultural influences on emotion recognition, as has also been shown by the

above-referenced studies. In cross-cultural interaction it makes a big difference whether an emotion is recognized at a level no better than chance guessing (e.g. scoring 30% in a choice of four, where chance probability is 25%) or at a level significantly higher than chance (e.g. 70% when there is a 25% chance probability). Where the latter is the case, confusion with other emotions is less likely and thus interaction with a person from the other culture may proceed more adequately. It is also important to identify what other emotion this or that emotion is likely to be mixed up with. If confusion occurs among members of the same emotion family (like anger and rage) it causes fewer problems than confusion among members of different emotion families, such as joy and anger (see also Scherer et al. 2011:422–423).

Altrov and Pajupuu (2010) have examined cross-cultural emotion recognition from an interaction point of view. Their focus was on the interpretation of emotions expressed in the native speech of foreigners from geographically close areas and with a relatively similar historical and cultural background to that of the listeners. In their study, 32 adult native Latvians were asked to listen to non-acted context-free emotional expressions uttered in native Estonian and to decide for each expression whether it conveyed sadness, joy, anger or a neutral attitude. Although the Latvians recognized Estonian emotions with better than chance probability, their scores differed significantly from those of the Estonians. The Latvians' scores of lower than 50% for joy, anger and neutrality suggest a high probability of confusion occurring in the interaction. Sadness was the only Estonian emotion recognized relatively well (72.5%) by the Latvian test subjects.

Kamaruddin et al. (2012) used automatic emotion recognition to demonstrate the influence of culture on the expression and perception of emotions. Using three culturally homogeneous data sets – Berlin (German), NTU-American (American English language), NTU-Asian (Malay language) – they tested the recognition of sadness, happiness, anger and neutrality separately for each data set (emotions were recognized in the same data set as that in which the recognizer was trained) as well as between the data sets (the recognizer trained on one data set was tried on the emotions of another data set; for instance, the Berlin data set was used for training, but tested with the NTU-American or the NTU-Asian and vice versa – a so-called inter-cultural experiment). The results showed high scores for the intra-cultural experiment: the average was 65.1% for the Berlin data set, 70.4% for the NTU-American and 64.5% for the NTU-Asian one, whereas the inter-cultural experiment showed a very poor overall performance accuracy, which was almost as low as chance guessing (25% for four choices). For those trained on the NTU-American and tested on the Berlin data set, for example, *happiness* was recognized in a mere 8.1% of cases. The best inter-cultural scores were collected from those trained on the NTU-American and tested on the NTU-Asian data set, which gives evidence of the relatively greater similarity of how emotion is expressed in those languages. The results indicate that cultural influence on emotion recognition is an issue also to be considered in speech technology.

Most studies of cultural differences in the recognition of vocal expressions of emotion have compared members of different ethnic groups living in different countries. Many of the studies have used acted speech. In the present study the focus lies on how emotions are recognized by people who speak different languages, but live in the same country; more precisely, it focuses on how vocal expressions of emotions are recognized not in one's first language (L1), but in one's second (L2), when the emotions are not acted but expressed only moderately.

The focus group consisted of Russians living in Estonia.

Two research questions were posed:

- 1) Do Russians living in Estonia, whose second language (L2) is Estonian recognize emotions expressed in native Estonian?
- 2) Do Russians living in Estonia recognize emotions in the same way as Estonians or as Russians living in Russia?

3. Background information on the communication contacts between Estonians and Russians living in Estonia

According to the 2011 census the Estonian population is about 1.3 million, of which 68.5% (886,859) speak Estonian (the official national language) as their mother tongue and 29.6% (383,062) speak Russian as their first language².

Due to the official language requirements in public service, Estonian language skills are tested on three levels. Command of the official language at advanced level (C1) is obligatory in demanding positions requiring higher education. Every year about 1,000 people who are mainly Russian-speaking take the C1-level exams³.

Despite the professional language requirements, studies have established that when it comes to Russian–Estonian interaction only 33% of Russian speakers mainly use Estonian, while 33% prefer Russian and 34% use both languages. It is estimated that about a third of the Estonian minority population speaks Estonian at work, while in the private sphere (friends or family circle) only one in every six or seven does. In the public space, on the street and in services, Russians are slightly more active in speaking Estonian than in the private sphere, as in those spheres one in every four or five Russians is ready to start an interaction in Estonian or switch over to Estonian (Vihalemm 2011).

Thus, there is a considerable number of Russian speakers who know Estonian but do not use it in every occasion.

² Population and housing census 2011: http://www.stat.ee/64628?parent_id=32784

³ Examination statistics: <http://www.innove.ee/et/eesti-keele-tasemeeksamid/tasemeeksamite-statistika-ja-analysid>

4. Specifics of emotional expression in Estonian and Russian

According to many studies (Juslin and Laukka 2003, Scherer et al. 2011), the acoustic parameters mostly used to describe vocal emotions are pitch, intensity and speech rate. Studies of the prosody of (mainly acted) basic emotions have shown that expressions of sadness tend to be produced with a relatively low pitch and slow speaking rate, whereas expressions of anger and happiness tend to be produced with a moderate or high mean pitch and fast speaking rate. In addition, anger and happiness usually display high pitch variation, whereas sadness often exhibits less pitch variation (see the survey by Pell, Paulmann et al. 2009). This does not mean, however, that there are no cross-cultural differences in the expression of emotion (Anolli et al. 2008, Scherer et al. 2011).

For Estonian, the acoustics of three emotions – joy, anger and sadness – and of neutral speech have been studied statistically, using the material of the Estonian Emotional Speech Corpus⁴. The respective Russian emotions and neutrality have been studied using the RUSLANA⁵ database of Russian emotional speech.

Certain differences can be observed in the prosodic expression of joy, anger, sadness and neutrality in L1 speakers of Russian as compared to L1 speakers of Estonian (see Tamuri 2012, Tamuri and Mihkla 2012, Makarova and Petrushin 2003, 2012). The articulation rate of Estonian emotional and neutral sentences was measured in speech sounds per second (Tamuri and Mihkla 2012). Ranking the average speech rates from the most rapid to the slowest yields the following sequence: anger > joy > neutral > sadness. The rate of emotional speech in Russian does not much depend on the emotion expressed (Makarova and Petrushin 2003). Intensity makes a significant difference between emotions and neutral speech in Estonian except in the case of the pair, joy and sadness (Tamuri 2012). The mean intensity of neutrality and the three emotions (from strongest to weakest) are ranked as follows: neutral > anger > joy > sadness. The intensity ranking of the respective Russian emotions running from strongest to weakest is anger > happiness > sadness > neutral (Makarova and Petrushin 2012). For emotions in Estonian, the mean pitch is highest for joy and lowest for anger: joy > neutral = sadness > anger (Pajupuu 2012). The respective ranking of emotions in Russian from highest to lowest is happiness > anger > sadness > neutral (Makarova and Petrushin 2012). The main difference between Estonian and Russian vocal emotions is observed in the expression of anger and neutrality. Estonian neutral speech is intense, and the pitch is high; for anger the pitch is lowered. In Russian the situation is the other way around; notably, angry speech is

⁴ Estonian Emotional Speech Corpus contains 1,234 sentences isolated from read passages, with emotion class – joy, anger, sadness or neutral – defined by means of listening tests. All corpus sentences are different. <http://peeter.eki.ee:5000/>

⁵ The Russian Language Affective speech database (RUSLANA) contains 3,660 sentences recorded by 61 speakers. They were asked to read ten sentences neutrally (unemotionally) and in five emotional states (surprise, happiness, anger, sadness and fear). http://universal.elra.info/product_info.php?cPath=37_39&products_id=2294

intense and has a high pitch, while neutral speech has a weak intensity and the lowest pitch. Joy is also expressed differently in Estonian and Russian: although the mean pitch for joy is similarly high in both languages, the intensity is high in Russian, but low in Estonian. The Estonian joy is more like quiet happiness – high, but not particularly intense (Pajupuu 2012, Tamuri 2012).

The different traditions of expressing emotions may easily affect the recognition of Estonian emotions by Russians.

5. Material and method

5.1. Material

The research material comes from the Estonian Emotional Speech Corpus (see Altrov and Pajupuu 2012). The speech in the corpus has been obtained by extracted from longer text passages – it is assumed that every text evokes a certain mood in the reader, which is vocally expressed. The corpus does not contain acted emotions. The read passages were segmented into sentences, which were separated from context and presented to evaluators to decide whether the utterances sounded joyous, angry, sad or neutral. The listeners were briefed that each of the three emotions listed (joy, anger, sadness) also comprised several other closely related emotions: joy included gratitude, happiness, pleasure and exhilaration; anger included resentment, irony, reluctance, contempt, malice and rage; sadness covered loneliness, disconsolation, concern and hopelessness; while, neutral speech was to be understood as normal speech, without special emotions. For 73.5% of corpus sentences the same emotion or neutrality was suggested by over 50% of the listeners, and each corpus sentence carries its own rate for the identification of emotion. In addition, the same sentences were presented to evaluators in text form, and they were given the task of determining the emotion from the semantic content without hearing the sentence. A comparison of the listening and reading results enabled the corpus sentences to be divided into two groups: (1) sentences where the semantic content and vocal expression convey a similar emotion (defined as such both from listening and reading); (2) sentences where the tone of voice changes the emotion conveyed by the semantic content (i.e. different emotions were identified from listening and reading, plus those cases where no emotion was detected from the semantic content and the emotional tone has to be conveyed only by the voice). The sentences in the second group represent a situation where the cue for the emotional content is not in what is said, but in how it is said.

For the present study the corpus was searched for 10 joyful, 10 sad, 10 angry and 5 neutral sentences with an emotion identification rate of no less than 65% (i.e. more than 2.5 times better than chance) and which belonged to the second group, that is, where the tone of voice not the semantic content played the decisive role in emotion identification (see Table 1). The sentences were arranged into a web-based listening test. The test sentences were sequenced so that no two consecutive sentences would make a logical whole.

Table 1. Emotional corpus sentences picked for the listening test

Corpus sentences selected for listening (translated into English)	Recognition rate from listening (corpus data)	Recognition rate from reading (corpus data)
1. Although Ott knew nothing of my existence...	joy 86.2%	?
2. As I see it, the parliament acts as a “rubber stamp”.	anger 75.8%	?
3. Musician and artists seem to enjoy interacting with their fellow citizens unshaved.	anger 77.4%	?
4. Our family never discussed any problems.	sadness 82.8%	?
5. I saw how the district chief of Nõmme passed out at the Raba race yesterday.	neutral 71.0%	sadness 55.2%
6. Once, I wanted hot semolina porridge at three o'clock in the morning.	joy 69.7%	?
7. I live in Tallinn now, together with my invalid daughter, and we feel much more comfortable here than in Narva.	neutral 65.5%	joy 69.2%
8. However, the Estonian national team deserves praise.	joy 87.1%	?
9. At yesterday's concert Padar smoked like a chimney.	neutral 66.7%	anger 55.6%
10. I can't move my cowshed to the vicinity of Tallinn, can I?	sadness 74.2%	?
11. His spirit will have a most prolonged influence on us.	sadness 93.1%	?
12. Only state officials can expect a 10% rise.	anger 75.9%	?
13. She even called my granny, urging me to come back to school.	joy 81.2%	?
14. There could be a medical professional or two available by the wood trail, couldn't there?	anger 90.3%	?
15. What I may and what I may not.	sadness 80.6%	?
16. The more painful the truth will be later.	neutral 75.0%	?
17. A waiter said in the newspaper that his constantly bad mood was due to the numerous Finnish clientele.	neutral 70.0%	?
18. In this case, it is a beautiful suffering.	joy 75.0%	?
19. I can't see why people would want to look so much uglier than they really are.	anger 96.7%	?
20. Can't we really manage anything without training any more?	anger 67.7%	neutral 53.3%
21. At that moment he was the only person there for me 24/7.	sadness 94.8%	?
22. So I quitted work without notice.	sadness 90.6%	?
23. At three in the morning!	joy 75.0%	?
24. If a man works at a restaurant, he is usually believed to have something wrong with him.	anger 66.7%	?
25. It would be so genuine!	joy 74.2%	?
26. Whatever I do, he is never happy.	sadness 78.1%	anger 64.3%
27. And dizzy like sharks.	anger 65.6%	?
28. I miss Enn even in the daytime when I work.	joy 67.7%	sadness 86.2%
29. This is an enormous hole.	sadness 75.9%	?
30. He practically doesn't touch alcohol.	joy 77.4%	neutral 61.1%
31. What those four years have done ...	sadness 77.4%	?
32. Like a dead body.	anger 79.3%	?
33. How to cope with such a situation?	anger 83.9%	sadness 62.1%
34. But just a little bit.	joy 78.1%	?
35. A difficult passage.	sadness 86.2%	neutral 53.8%

Note. The question mark (?) indicates those cases where the evaluators disagreed over the emotion or neutrality of the corpus sentence (i.e. none of the four choices – joy, sadness, anger, neutral – had achieved more than 50% agreement of the evaluators).

5.2. Participants

The listening test was taken by three groups of adult subjects: (1) 36 ethnic Estonians (average age $M = 41.3$ years, $SD = 9.2$) who spoke Estonian as their mother tongue; (2) 16 ethnic Russians living in Estonia (average age $M = 35.8$ years, $SD = 8.1$) whose first language (L1) was Russian and who spoke Estonian (L2) at an advanced level (C1); (3) 16 ethnic Russians living in Russia (average age $M = 33.1$ years, $SD = 5.0$) whose mother tongue was Russian and who had no knowledge of Estonian. Both L1 Russian groups had received higher education. The subjects were chosen from those over 30 years old. One of the reasons for this was a recommendation made in an earlier study to the effect that people who are too young might not be the best subjects, because it can be assumed that recognition of vocal emotions is, to some extent at least, a culture-specific skill that can only be acquired over time; in other words, the successful decoding of stereotypical vocal cues for emotions is based on the long-term experience of being a member of a culture (Toivanen et al. 2004). Similarly, it has been confirmed that people aged over 30 are more capable of recognizing emotions in the voice than younger adults aged 20 to 28. Moreover, younger people classify considerably more of the emotions of sentences as neutral (see Altrov and Pajupuu 2010).

5.3. Procedure

The test was carried out using the web-based testing environment of the Estonian Emotional Speech Corpus. The participants, who needed computer access and either speakers or headphones, were provided with links so that they could do the listening test. On entering the testing environment the participants were asked to fill in the following personal data: sex, age, education, ethnicity, mother tongue and main language of education. Instructions were available when the test was opened; the text of these was in Estonian for all subjects living in Estonia, and in Russian for the Russians living in Russia. The subjects were asked to listen to separate context-free Estonian sentences and decide, without seeing the text, upon the emotion of the sentence. The choice was between joy, anger, sadness and a neutral attitude. It was explained that in normal speech, these emotions are seldom encountered in their full form and this is why joy should be interpreted as an emotion that also covers gratitude, happiness, pleasure and exhilaration; anger also includes resentment, irony, reluctance, contempt, malice and rage; and sadness covers loneliness, disconsolation, concern and hopelessness, while neutral refers to normal speech without special emotions. The subjects could listen to each sentence as many times as they wished.

6. Results

Pearson's χ^2 test was used to reveal whether the three groups of participants – Estonians living in Estonia, Russians living in Estonia and Russians living in

Russia – significantly differed from each other in terms of emotion recognition, see Tables 2, 3 and 4.

Table 2. Recognition of Estonian emotions: Comparison of Estonians and Russians living in Estonia

Target emotion	Groups	Response emotions				p
		Joy	Anger	Sadness	Neutral	
Joy	Estonians	234	13	26	76	.005
	Russians living in Estonia	83	19	17	41	
Anger	Estonians	12	195	61	80	.501
	Russians living in Estonia	4	101	23	32	
Sadness	Estonians	4	40	280	23	.056
	Russians living in Estonia	3	26	112	19	
Neutral	Estonians	10	18	21	126	.037
	Russians living in Estonia	11	11	14	44	

Note. Pearson's χ^2 results: significant difference between group recognitions if $p < .05$.

Table 3. Recognition of Estonian emotions: Comparison of Estonians living in Estonia and Russians living in Russia

Target emotion	Groups	Response emotions				p
		Joy	Anger	Sadness	Neutral	
Joy	Estonians	234	13	26	76	.001
	Russians living in Russia	57	22	12	69	
Anger	Estonians	12	195	61	80	.001
	Russians living in Russia	16	68	21	55	
Sadness	Estonians	4	40	280	23	.001
	Russians living in Russia	10	11	117	22	
Neutral	Estonians	10	18	21	126	.001
	Russians living in Russia	15	24	6	35	

Note. Pearson's χ^2 results: significant difference between group recognitions if $p < .05$.

Table 4. Recognition of Estonian emotions: Comparison of Russians living in Estonia and in Russia

Target emotion	Groups	Response emotions				p
		Joy	Anger	Sadness	Neutral	
Joy	Russians living in Estonia	83	19	17	41	.005
	Russians living in Russia	57	22	12	69	
Anger	Russians living in Estonia	4	101	23	32	.001
	Russians living in Russia	16	68	21	55	
Sadness	Russians living in Estonia	3	26	112	19	.017
	Russians living in Russia	10	11	117	22	
Neutral	Russians living in Estonia	11	11	14	44	.022
	Russians living in Russia	15	24	6	35	

Note. Pearson's χ^2 results: significant difference between group recognitions if $p < .05$.

According to the results, the Estonians and Russians living in Estonia did not significantly differ in their recognition of anger or sadness ($p < .501$ and $p < .056$, respectively), but there was a significant difference in the recognition of joy and neutrality. The Estonians living in Estonia and the Russians living in Russia differed considerably in the recognition of all three Estonian emotions and Estonian neutrality, and so did the Russians living in Estonia and the Russians living in Russia.

The confusion matrix (Table 5) demonstrates how the subjects' judgements were distributed between correct and incorrect emotions.

Table 5. Confusion matrix: Recognition of Estonian emotions by Estonians living in Estonia, Russians living in Estonia and Russians living in Russia (% of target recognition)

Target emotion	Response emotions				Number of responses
	Joy	Anger	Sadness	Neutral	
<i>Estonians</i>					
Joy	67.0	3.7	7.5	21.8	349
Anger	3.5	56.0	17.5	23.0	348
Sadness	1.2	11.5	80.7	6.6	347
Neutral	5.7	10.3	12.0	72.0	175
<i>Russians living in Estonia</i>					
Joy	51.9	11.9	10.6	25.6	160
Anger	2.5	63.1	14.4	20.0	160
Sadness	1.9	16.2	70.0	11.9	160
Neutral	13.8	13.8	17.5	55.0	80
<i>Russians living in Russia</i>					
Joy	35.6	13.8	7.5	43.1	160
Anger	10.0	42.5	13.1	34.4	160
Sadness	6.3	6.9	73.1	13.7	160
Neutral	18.8	30.0	7.5	43.7	80

The data demonstrates that each emotion was recognized at least twice as often as chance probability (i.e. over 50%) by Estonians and by Russians living in Estonia, i.e. the target emotion was their most frequent answer. Sadness was the only Estonian emotion recognized twice as often as chance probability by the Russians living in Russia.

For Estonians no wrong choice scored better than chance probability. The Russians living in Estonia confused joy with neutrality, while Russians living in Russia confused both joy and anger with neutrality, and neutrality was confused with anger. Thus, we can conclude that Russians find it difficult to distinguish Estonian joy from neutrality, no matter whether they live in Russia or Estonia; in addition, the Russians living in Russia may mistake Estonian neutrality for anger, as shown in Figs. 1, 2, 3 and 4.

Target emotion – JOY

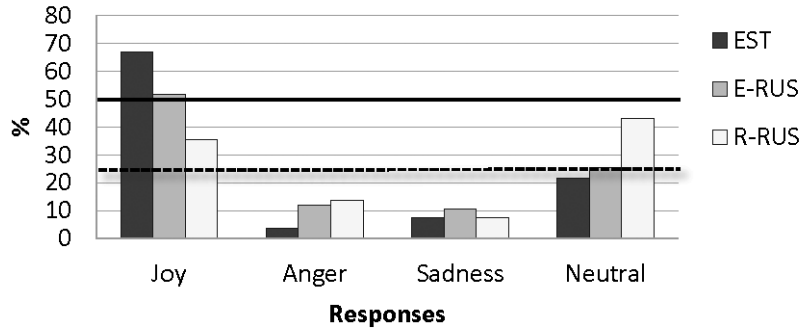


Fig. 1. Recognition of Estonian joy. Dashed line: *chance probability*; solid line: *twice the chance probability*.

Target emotion – ANGER

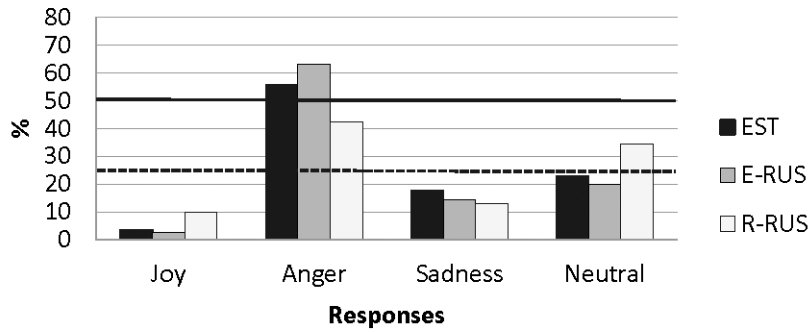


Fig. 2. Recognition of Estonian anger. Dashed line: *chance probability*; solid line: *twice the chance probability*.

Target emotion – SADNESS

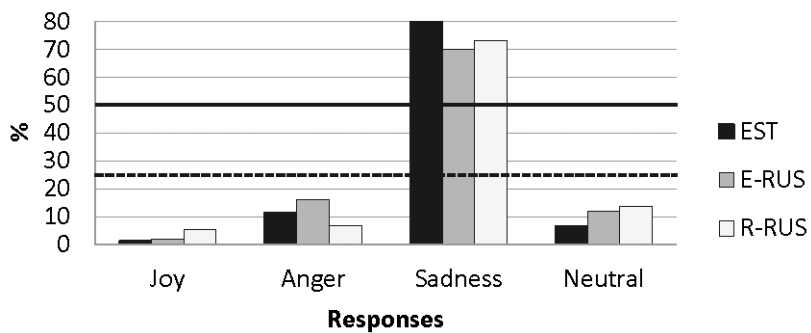


Fig. 3. Recognition of Estonian sadness. Dashed line: *chance probability*; solid line: *twice the chance probability*.

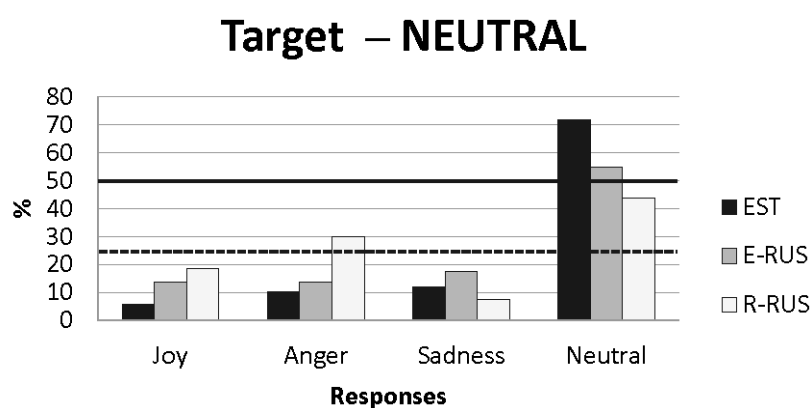


Fig. 4. Recognition of Estonian neutrality. Dashed line: *chance probability*; solid line: *twice the chance probability*.

7. Discussion

The research question was whether Russians who live in Estonia and whose second language (L2) is Estonian recognize emotions expressed in Estonian and whether Russians living in Estonia recognize emotions in the same way as Estonians or as Russians who live in Russia.

A comparison of the results of three test groups – Estonians living in Estonia, Russians living in Estonia and Russians living in Russia – demonstrated the influence of cross-cultural contacts on emotion recognition. The Russians living in Estonia recognized all Estonian emotions with an accuracy of over 50%, while anger and sadness were recognized with a similar accuracy both by Russians living in Estonia and Estonians, but there were differences when it came to joy and neutrality, where Russians had a lower accuracy because they mistook joy for neutrality.

The Estonians living in Estonia and the Russians living in Russia differed significantly in their recognition of all three emotions and neutrality, and so did the Russians living in Estonia compared to those living in Russia. Thus, the Russians living in Estonia recognized Estonian emotions in a way that was rather more like the local Estonians than the Russians living in Russia.

The Russians living in Russia gained scores of over 50% only for Estonian sadness. They mistook Estonian joy for neutrality and could not distinguish the difference between Estonian anger and neutrality.

Cross-cultural surveys have shown that negative emotions, such as anger and sadness, are better recognized across cultures (e.g. Thompson and Belkwill 2006, Pell et al. 2009, Scherer et al. 2011). The fact that the Russians living in Russia recognized Estonian sadness relatively well is in accord with these studies. At the same time, they could not recognize Estonian anger, which they confused with neutrality. The reason for this could be the opposite way of expressing anger and neutrality in Estonian and Russian: Russian neutral speech is quiet and has a low

pitch and Russian anger is expressed with loud speech at a high pitch, whereas in Estonian the intensity and pitch of neutral speech are higher than those of angry speech (Makarova and Petrushin 2012, Pajupuu 2012, Tamuri 2012). The fact that the Russians living in Estonia recognized Estonian anger almost as well as the local Estonians proves that interaction in a common cultural environment has helped them to learn the Estonian way of expressing that emotion.

The intensity of Estonian neutral speech caused confusion in the recognition of Estonian joy by both Russian groups. Although both Russian joy and Estonian joy have a high pitch, Russian joy is intense like Estonian neutral speech (Makarova & Petrushin 2012, Pajupuu 2012, Tamuri 2012). Although the Russians living in Estonia recognized Estonian joy with an accuracy that is two times better than chance, they still confused it with neutrality. The Russians living in Russia, however, tended to interpret Estonian joy as a neutral attitude rather than joy. This is another difference that demonstrates the importance of living and interacting in a culture for emotion recognition since, despite scoring lower on joy than the local Estonians, the Russians living in Estonia were still capable of recognizing the Estonian emotion, unlike the Russians living in Russia who mostly failed in the task. That is to say, the Russians living in Estonia have acquired the Estonian pattern of expressing joy.

Consequently, our results comparing the performance of Russians living in Estonia with those in Russia in the recognition of Estonian emotions suggests that understanding emotions is dependent on cultural factors and social interactions. That is, the social norms of a culture are learnt during practical interaction.

According to research, the Russians living in Estonia mainly interact in Estonian in the working environment and at school (Vihalemm 2011). This suggests that the high level (C1) of Estonian language proficiency of the test group is sufficient to enable the acquisition of the Estonian pattern of emotion perception if they work or study in Estonian environment. In other words, high language proficiency and close interaction with Estonians guarantees that Russians who live in Estonia will receive not only the verbal Estonian messages, but also an understanding of their emotional content.

Acknowledgements

The study was supported by the project SF0050023s09 “Modelling inter-modular phenomena in Estonian”.

Address:

Rene Altrov
Institute of the Estonian Language
Roosikrantsi 6
10119 Tallinn, Estonia

Tel.: +372 522 7203

E-mail: rene.altrov@eki.ee

References

- Altrov, Rene and Hille Pajupuu (2010) "Estonian emotional speech corpus: culture and age in selecting corpus testers". In *Human language technologies – the Baltic perspective: proceedings of the fourth international conference Baltic HLT 2010*, 25–32. Inguna Skadina, Andrejs Vasiljevs, eds. Amsterdam: IOS Press.
- Altrov, Rene and Hille Pajupuu (2012) "Estonian emotional speech corpus: theoretical base and implementation". In *4th international workshop on corpora for research on emotion sentiment & social signals (ES3)*, 50–53. Laurence Devillers, Björn Schuller, Anton Batliner, Paolo Rosso, Ellen Douglas-Cowie, Roddy Cowie, and Catherine Pelachaud, eds. Istanbul.
- Anolli, Luigi, Lei Wang, Fabrizia Mantovani, and Alessandro De Toni (2008) "The voice of emotion in Chinese and Italian young adults". *Journal of Cross-Cultural Psychology* 39, 5, 565–598.
- Bryant, Gregory A. and H. Clark Barrett (2008) "Vocal emotion recognition across disparate cultures". *Journal of Cognition & Culture* 8, 1–2, 135–148.
- Cheang, Henry S. and Marc D. Pell (2008) "The sound of sarcasm". *Speech Communication* 50, 5, 366–381.
- Cowie, Roddy, Naomi Sussman, and Aaron Ben-Ze'ev (2008) "Emotions: concepts and definitions". In *Emotion-oriented systems: the HUMAINE handbook*, 9–31. Roddy Cowie, Catherine Pelachaud, and Paolo Petta, eds. Berlin and Heidelberg: Springer.
- Justlin, Patrik N. and Petri Laukka (2003) "Communication of emotions in vocal expression and music performance: different channels, same code?". *Psychological Bulletin* 129, 5, 770–814.
- Kamaruddin, Norhaslinda, Abdul Wahab, and Chai Quek (2012) "Cultural dependency analysis for understanding speech emotikon". *Expert Systems with Applications* 39, 5, 5115–5133.
- Makarova, Veronika and Valery A. Petrushin (2003) "Phonetics of emotion in Russian speech". In *Proc. of the XVth international conference of phonetic sciences*. D. R. M. J. Sole and J. Romero, eds. Barcelona, Spain.
- Makarova, Veronika and Valery A. Petrushin (2012) "Phonetics: tracing emotions in Russian vowels". In *Russian language studies in North America: new perspectives from theoretical and applied linguistics*, 3–42. Veronika Makarova, ed. (Anthem Series on Russian, East European and Eurasian Studies.) London and New York: Anthem Press.
- Mitchell, Rachel L. C. (2007) "Age-related decline in the ability to decode emotional prosody: primary or secondary phenomenon?". *Cognition & Emotion* 21, 7, 1435–1454.
- Pajupuu, Hille (2012) "Emotsioonid – kõnetehnoloogia olevik ja tulevik". [Emotions – the present and future of speech technology.] *Keel ja Kirjandus* (Tallinn) 8-9, 629–643.
- Pell, Marc D, Laura Monetta, Silke Paulmann, and Sonja A. Kotz (2009) "Recognizing emotions in a foreign language". *Journal of Nonverbal Behavior* 33, 2, 107–120.
- Pell, Marc D., Silke Paulmann, Chinar Dara, Areej Alasseria, and Sonja A. Kotz (2009) "Factors in the recognition of vocally expressed emotions: a comparison of four languages". *Journal of Phonetics* 37, 4, 417–435.
- Scherer, Klaus R., Elizabeth Clark-Polner, and Marcello Mortillaro (2011) "In the eye of the beholder? Universality and cultural specificity in the expression and perception of emotikon". *International Journal of Psychology* 46, 6, 401–435.
- Soto, Jose Angel and Robert W. Levenson (2009) "Emotion recognition across cultures: the influence of ethnicity on empathic accuracy and physiological linkage". *Emotion* 9, 6, 874–884.
- Tamuri, Kairi (2012) "Intensity of Estonian emotional speech". In *Human language technologies – the Baltic perspective: proceedings of the fifth international conference Baltic HLT 2012*, 238–246. Arvi Tavast, Kadri Muischnek and Mare Koit, eds. Amsterdam: IOS Press.
- Tamuri, Kairi and Meelis Mihkla (2012) "Emotions and speech temporal structure". *Linguistica Uralica* 3, 209–217.
- Thompson, William Forde and L.-L. Balkwill (2006) "Decoding speech prosody in five languages". *Semiotica* 158, 1-4, 407–424.

- Toivanen, Juhani, Eero Väyrynen, and Tapio Seppänen (2004) "Automatic discrimination of emotion from spoken Finnish". *Language & Speech* 47, 4, 383–412.
- van Rijn, Sophie, André Aleman, Eric van Diessen, Celine Berckmoes, Guy Vingerhoets, and René S. Kahn (2005) "What is said or how it is said makes a difference: role of the right fronto-parietal operculum in emotional prosody as revealed by repetitive TMS". *The European Journal of Neuroscience* 21, 11, 3195–3200.
- Vihalemm, Triin (2011) "Keelepraktikad, kollektiivne identiteet ja mälu". [Linguistic practices, collective identity and memory.] In *Eesti ühiskonna integratsiooni monitoring 2011*, 113–156. [The monitoring of the integration of the Estonian society.] Triin Vihalemm, ed. Available online at <http://www.kul.ee/webeditor/files/integratsioon/Intmon_2011_pt_4.pdf>. Accessed on 16.02.2013.