Lund University, Dept. of Linguistics and Phonetics
Working Papers 53 (2008), 179–187

Perception of emotions in speech

Joost van de Weijer and Sigrún Gunnarsdóttir

1 Introduction

Emotions in speech may be real or pretended. The first type occurs when a speaker is truly happy, sad or angry, and this emotional state is reflected in his or her speech. The second type occurs when the emotional load of an utterance is not the same as the speaker's emotional state. This is the case, for instance, when an actor pretends to be sad or happy, or when a parent pretends to be angry with a child.

Most of the research on emotions in speech has focused on the latter type of emotions in speech (Scherer 2003). The material used in these studies consists of utterances produced by an actor or a professional speaker who adopts different kinds of emotions.

In the present study, we focus on the former type of emotions in speech. We investigate whether listeners are able to distinguish between speech produced by speakers in different emotional states. In addition, we look at cultural differences in the expression and perception of emotions, and at gender effects, e.g., are there differences between male and female speakers and male and female listeners?

The distinction of different types of emotions is a difficult matter (Cowie & Cornelius 2003). A common classification is the following: angry, disgusted, fearful, happy, sad and surprised (Ovesdotter Alm & Sproat 2005). However, this categorization is not unproblematic since the boundaries between the categories are fuzzy, and there may be overlap, for instance between surprised and happy, or angry and sad. In order to avoid these problems we focus on two basic categories that we call positive and negative emotions. We also include a third category of neutral emotions for comparison. These emotions were elicited from a group of Swedish participants by presenting them with pictures that represented a positive scene (e.g., a baby animal), a neutral scene (e.g., an everyday object) or a negative scene (e.g., the victim of an accident). The participants were asked

questions about the scene to which they had to respond using a complete sentence. Their responses were subsequently presented to listeners who rated them on a seven-point scale as to how negative or positive they thought the utterance sounded.

We compared the ratings from Swedish and Icelandic listeners in order to establish whether there were any cultural differences in the perception of emotions in speech. It has been a much debated question whether emotions are universal or not. Results of research in perception of facial expression has provided evidence that emotions tend to be universal, but, at the same time, that participants tend to be better at judging people from their own culture than people from a different culture (Elfenbein & Ambady 2003).

Apart from cultural differences, the perception of emotions may also be influenced by gender differences. This issue is as yet unsettled (Wester et al. 2002). The expression of emotions may be more apparent in either female or male voices. Additionally, either female or male listeners may be more attentive to the acoustic cues that determine the speaker's emotional state. Finally, it is possible that speaker gender interacts with listener gender, i.e., female listeners are more responsive to female speakers, whereas male listeners are more responsive to male speakers.

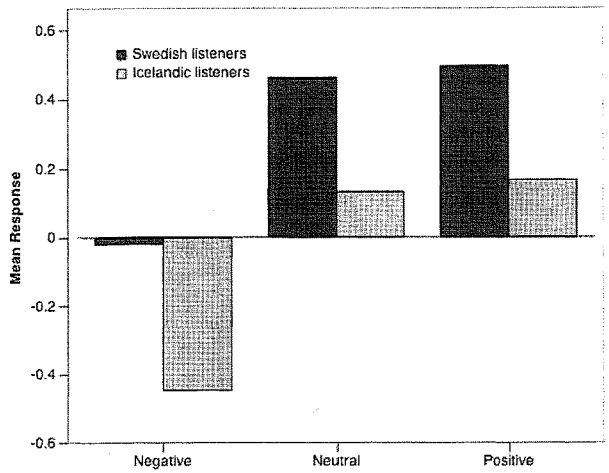

Figure 1: Mean responses of Swedish and Icelandic listeners.

2 Method

2.1 Experimental materials

For the elicitation of the auditory stimuli, subjects were asked to answer simple, neutral questions related to pictures shown on a computer screen. There were 60 pictures all taken from the *International Affective Picture System* (*IAPS*), a set of 480 pictures that have been rated for the emotional response (pleasure, dominance, arousal) that they elicit. The pleasure ratings in the *IAPS* vary from 1 (negative) to 9 (positive), which served as the basis for selecting 60 pictures of which 20 were negative (mean rating of 2.16), 20 were neutral (mean rating of 5.38) and 20 were positive (mean rating of 7.43). The negative pictures were of people with severe injuries, dangerous animals, starving children. Positive pictures were of small animals, babies, happy faces. Neutral pictures showed objects, people with neutral faces.

The 60 pictures were shown in random order to a group of 12 Swedish subjects. The subjects were six women and six men with an average age of 21.8 years. The subjects were instructed to answer a question about the picture with a complete sentence. Naturally, there was nothing in the instructions about the way in which the subjects were expected to answer. The questions were displayed next to or below the pictures on the computer screen. All questions were semantically neutral, so as not to disclose the nature of the picture in the perception test. Examples of questions were: 'Is it

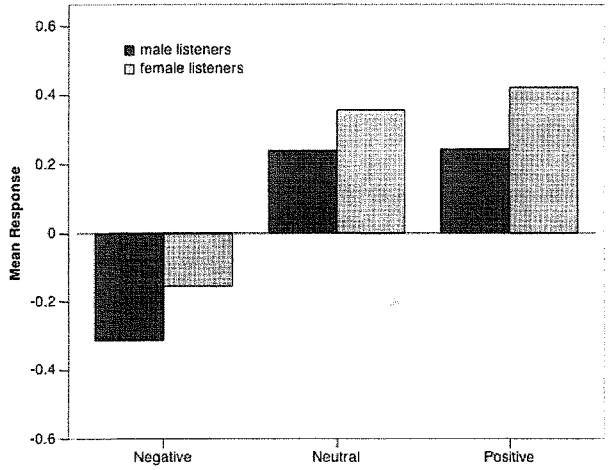

Figure 2: Effect of listener gender.

a man or a woman on the picture?', 'Is the situation inside or outside?'. The responses were recorded directly on the hard disk of the computer. All subjects were informed beforehand that some of the pictures would be strong, and that they could opt to disrupt at any time during the recording. None of the subjects, however, chose to do so.

Subsequently, 10 responses from each subject were selected for the perception experiment. In principle, we selected the first 10 responses, unless we considered one of these unsuitable, e.g., because the response was not a complete sentence. The resulting test items were 120 utterances, of which 39 were negative, 35 were neutral and 46 were positive.

2.2 Perception test

The test items were played in random order to a group of 16 Swedish listeners (eight men, eight women, average age 24.0 years) and a comparable group of 16 Icelandic listeners (eight men, eight women, average age 24.4 years). The listeners were instructed to listen to the utterances and to rate them on a seven-point scale from –3 to +3 as to how positive or how negative they thought the utterance sounded. The Icelandic subjects knew only little or no Swedish at all. The stimuli were played over headphones at a comfortable listening level. The listeners were able to adjust the volume if they wanted to.

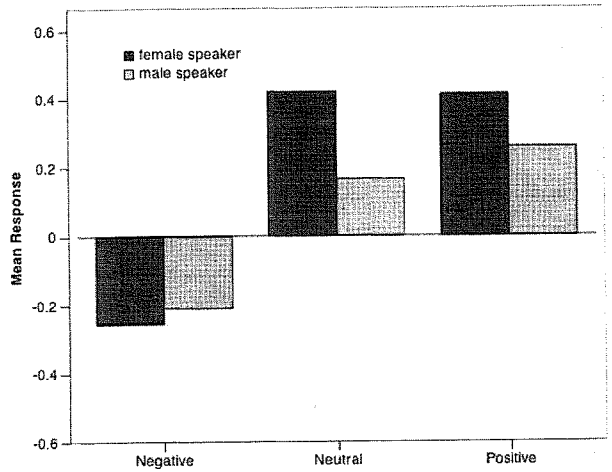

Figure 3: Effect of speaker gender.

3 Results

The results were subjected to an ANOVA (*F1*/*F2* analysis) with the following factors: condition (negative, neutral, positive), language (Swedish, Icelandic), speaker gender, listener gender. The results of the analysis were the following.

Two effects were significant ($p < 0.05$) in both the *F1* and the *F2* analysis: Condition ($F1[2,56] = 6.417$; $F2[2,114] = 6.210$), language ($F1[1,28] = 9.098$; $F2[1,114] = 67.402$). These two effects are illustrated in Figure 1 which shows the average ratings of the Swedish and the Icelandic listeners. The figure shows that overall ratings for negative items were lower than for neutral and positive utterances. This was confirmed by a Tukey *post hoc* comparison: There was no significant difference between positive and neutral utterances, but the ratings for the negative utterances were significantly lower. Furthermore, the Icelandic listeners tended to give overall lower ratings than the Swedish listeners.

Furthermore, the following trends (i.e., the effect was significant in one of the analyses only) were observed. Male listeners tended to give lower ratings than female listeners (see Figure 2). This effect was significant in the *F2* analysis ($F2[1,154] = 17.905$), but not in the *F1* analysis. Furthermore,

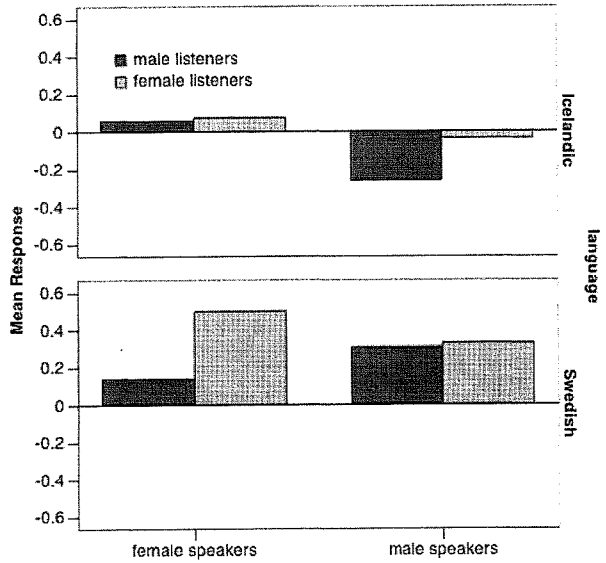

Figure 4: Three-way interaction of speaker gender, listener gender and language.

compared to the male speakers, the female speakers received on average lower scores on negative items, and higher scores on the neutral and the positive items (see Figure 3). In other words, ratings for the male speakers were less extreme than those for the female speakers. This interaction was significant in the *F1* analysis ($F1[2,56] = 9.697$), but not in the *F2* analysis. There was an interaction between listener gender and speaker gender, but the shape of this interaction was not the same for the two languages (see Figure 4). Roughly speaking, the male Swedish listeners gave higher ratings to the male speakers, whereas the female listeners gave higher ratings to the female speakers. In the Icelandic group, on the other hand, all listeners gave higher ratings to the female speakers than the male speakers. This three-way interaction of listener gender, speaker gender and language was significant in the *F2* analysis ($F[1,114] = 13.911$) but not in the *F1* analysis. A final significant effect was a three-way interaction between language, listener gender and condition. This effect was significant in the *F2* analysis only ($F2[2,114] = 3.301$). Inspection of the average ratings (Figure 5) indicated that the Swedish female listeners gave higher ratings than the male listeners to the negative items, whereas the difference between male and female Icelandic listeners was negligible.

In sum, it appeared that the listeners made a difference between the three types of utterances. We made an attempt to establish which factors

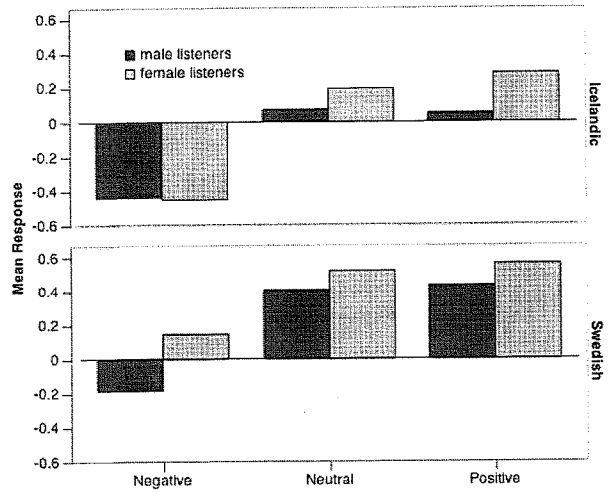

Figure 5: Three-way interaction of listener gender, language and condition.

distinguished the three conditions, and performed some basic measurements, known to correlate with emotional expression, on the materials. These measures included F0 characteristics, duration, length, speaking rate, and whether or not the utterance ended in a terminal rise (this final feature was established by a phonetically trained listener). The results, separated for male and female speakers, are displayed in Table 1.

There was only one significant difference in the measurements. The number of words per utterance for the female speakers ($F[2,57] = 4.626$, $p < 0.05$), indicating that the neutral utterances produced by the female speakers were significantly shorter than the negative or the positive utterances. All the other differences were non-significant.

4 Discussion

The results of the present study suggest that listeners are able to distinguish between utterances that are spoken in different emotional states, irrespective of the utterance's semantic content. We found significantly lower ratings for negative utterances compared to neutral and positive utterances. We did not, however, find any difference between ratings for positive and neutral utterances. The probable cause is that, using picture material, positive emotions are more difficult to elicit than negative emotions. Reactions to scenes such as severely injured victims or starving children tend to be much stronger than those to happy faces or baby animals.

The same pattern of results was obtained for the Swedish and the Icelandic group, which suggests that the perception of emotions, in any case these two

Table 1. *Post hoc* measurements on the experimental materials.

	male			female		
	negative	neutral	positive	negative	neutral	positive
F0 mean (Hz)	123	117	115	216	217	220
F0 std (Hz)	42	38	35	60	64	60
F0 range (Hz)	66	48	38	136	128	126
utterance length (words)	4.79	4.94	4.17	4.05	3.33	4.14
utterance length (syllables)	6.11	7.12	6.58	5.40	5.06	5.55
articulation rate (syllables/second)	3.92	4.02	4.40	3.56	3.81	3.72
duration (seconds)	1.85	2.00	1.59	1.53	1.35	1.60
terminal rise (%)	68.4	64.7	70.8	90.0	77.8	86.4

basic types, is universal. Naturally, many more replications of the perception test would be necessary to further confirm this conclusion. The Icelandic and Swedish societies are comparatively close, and so are the two languages. It would be interesting to see whether the results are the same with listeners who have a more remote native language, and come from a different society.

Even though the Icelandic listeners responded in a similar way as the Swedish listeners, their overall ratings were significantly lower than those of the Swedish listeners. Utterances that sounded negative to the Icelandic listeners sounded neutral (i.e., average rating of approximately zero) to the Swedish listeners. Utterances that sounded positive to the Swedish listeners (i.e., average rating above zero) sounded neutral to the Icelandic listeners. This finding suggests that, in addition to universal aspects, there may be language-specific or cultural-specific aspects to the perception of emotion.

In addition to cultural effects, we also found indications (statistically not fully reliable) of gender effects. Male listeners gave on average lower ratings than female listeners, and female speakers received more extreme ratings than male speakers. In other words, the male listeners were more negative in their ratings than the female listeners, and the female speakers expressed the emotions more clearly than the male speakers. These gender effects, however, should be interpreted cautiously since they were not consistent across the languages and the conditions.

The fact that we did not find the same ratings for utterances in the three conditions suggests that there must have been differences in the realizations of the utterances. We did not, however, find any reliable acoustic correlates to the expression of emotions. The only significant difference in the *post hoc* measurements (the number of words per utterance for the female speakers being significantly smaller for neutral utterances than for positive or negative utterances) was not consistent with the results of the perception test. More fine-grained measurements are thus necessary to establish what caused the listeners to distinguish between the different emotions.

References

Cowie, R. & R. Cornelius. 2003. 'Describing the emotional states that are expressed in speech'. *Speech Communication*, 40, 5-32.

Elfenbein, H. and Ambady, N. 2003. 'Universals and cultural differences in recognizing emotions'. *Current Directions in Psychological Science* 12, 159-164.

Ovesdotter Alm, C. & R. Sproat. 2005. 'Perceptions of emotions in expressive storytelling. *Proceedings of InterSpeech 2005*, 533-36. Lisbon, Portugal.

Scherer, K. 2003. 'Vocal communication of emotion: a review of research paradigms'. *Speech Communication*, 40, 227-256.

Wester, S., D. Vogel, P. Pressly & M. Heesacker. 2002. 'Sex differences in emotion: a critical review of the literature and implications for counselling psychology'. *The Counseling Psychologist* 30, 630-652.