

## Gender differences in the ability to discriminate emotional content from speech

Juhani Toivanen<sup>1</sup>, Eero Väyrynen<sup>2</sup> and Tapio Seppänen<sup>2</sup>

<sup>1</sup>MediaTeam, University of Oulu and Academy of Finland

<sup>2</sup>MediaTeam, University of Oulu

### Abstract

*In this paper, an experiment is reported which was carried out to investigate gender differences in the ability to infer emotional content from speech. Fourteen professional actors (eight men, six women) produced simulated emotional speech data representing the most important basic emotions (three emotions in addition to neutral). Each emotion was simulated when reading aloud a semantically neutral text. Fifty-one listeners (27 males, 24 females) were asked to listen to the speech samples and choose (among the four options) the most appropriate emotional label describing the simulated emotional state. The female listeners were consistently better at discriminating the emotional state from speech than the male subjects. The results suggest that females are emotionally more sensitive than males, as far as emotion recognition from voice is concerned.*

### Introduction

Phoneticians, speech scientists and engineers are taking increasing interest in the role of the expression of emotion in speech communication. In addition to so-called basic emotions, other global speaker-states are investigated, for example, irritation and trouble in communication (Batliner et al. 2003). A major approach in basic (phonetic) research has been to investigate the vocal parameters of specific emotions, and these parameters are now understood relatively well. Nowadays, the role of the vocal expression of emotion is gaining increasing importance in the computer speech community, for example, in the applied context of the automatic discrimination/classification of emotional content from speech (ten Bosch 2003). It can be argued that, after a long exploratory stage, the study of the vocal expression of emotion is reaching a level of maturity where the main focus is on important applications, particularly those involving human-computer interaction.

In the study of the vocal communication of emotion, an important taking-off point is the

base-line data, i.e. the human emotion discrimination performance level. There is now a relatively large literature on the human discrimination of emotions from speech: reviewing over 30 studies of the subject conducted up to the 1980's, Scherer (1989) concludes that an average accuracy percentage of about 60 % can be obtained in experiments where listeners are to infer emotional content from vocal cues only (without any help from lexis etc.). In a recent large-scale cross-cultural study (Scherer et al. 2001), an accuracy level rate of 66 % was found, across emotions (neutral, anger, fear, joy, sadness and surprise) and cultural contexts (Europe, Asia and the US). In a western cultural context, vocal recognition of six emotions (neutral, anger, fear, joy, sadness and disgust) was 62 %.

Typically, in investigations of the human discrimination of emotions, a standard speech sample (an utterance or a short passage) is used: the same lexical content is produced (often by actors) with different simulated emotions and test subjects are asked to choose the most appropriate emotional label for each sample (among the intended emotion categories). The emotions investigated in these studies usually represent "basic emotions": it is argued that certain emotions – at least fear, anger, happiness, sadness, surprise and disgust – are the most important or basic emotions (because they are seen to represent survival-related patterns of responses to events in the environment).

Although the vocal expression of emotions has been investigated rather intensively, at least as far as simulated data is concerned (and empirical evidence has cumulated indicating how well basic emotions can be discriminated by human listeners in different cultures), there has been little research on inter-subject differences (within a culture) in emotion discrimination ability. Usually, the emotion recognition performance level of a group of test subjects is reported as a single numerical value, without making any intra-group distinctions. Thus there is very little reported empirical evidence concerning possible differences between female

and male subjects, for example, in their ability to infer emotional content from vocal cues only.

This paper concentrates on the inter-gender differences in emotion recognition ability in a simulated emotional speech data context. The research question is: within a speech community, are female listeners better than male listeners at distinguishing between different emotions in speech? And if they are better, are they consistently so, that is, are they better than male listeners also at discriminating emotions from speech produced by male speakers? To our knowledge, these are questions no-one has systematically addressed in the literature on the vocal communication of emotion.

## Speech data

For the purposes of the present study, simulated emotional speech data was collected. Fourteen professional actors (eight men, six women) produced the speech data. The speakers were aged between 26 and 50 (average age was 39); all were speakers of the same northern variety of Finnish. The speakers read out a phonetically rich Finnish passage of some 120 words simulating three basic emotions, in addition to neutral: sadness, anger and happiness/joy. The text was emotionally completely neutral, representing matter-of-fact newspaper prose. The recordings were made in an anechoic chamber using a high quality condenser microphone and a DAT recorder to obtain a 48 kHz, 16-bit recording. The data was stored in a PC as wav format files. Each monologue was divided into five consecutive segments of equal duration for discrimination experiment purposes: thus there were a total of 280 emotional speech samples with an average duration of 13 seconds (five samples for four emotions by fourteen speakers).

## Human discrimination experiments

A performance test for human emotion discrimination was performed in the form of listening tests. The listeners were students in a junior high school, aged between 14 and 15. Fifty-one subjects (27 males, 24 females) participated as volunteers. All were speakers of the same northern variety of Finnish (the actors were speakers of the same variety of Finnish). The listening tests took place in a classroom where the subjects heard the speech data (280

speech samples) from two high-quality computer speakers. The emotional labels to choose between were limited to the intended emotions, not containing any distracters. The subjects heard the samples in random order in eight consecutive sessions within a period of two months (each session was arranged at the beginning of a lesson).

## Results

Tables 1-9 show the results of the experiment: the emotion discrimination performance of the subjects is first presented *in toto* (female and male subjects listening to female and male speakers), and then the results are broken down into sub-categories (females listening to all speakers, females listening to females only, etc.). Each table is a confusion matrix, where the column on the left indicates the intended emotions and the rows indicate the recognized emotions. The underlined percentages indicate the average discrimination accuracy for each specific emotion. The average emotion recognition performance level in each setting is given as the "TOTAL" percentage.

Table 1. Emotion discrimination from voice: females and males listening to females and males.

TOTAL	Neutral	Sad	Angry	Happy
<b>76.9 %</b>				
Neutral	<u>78.4 %</u>	16.9 %	2.6 %	2.1 %
Sad	12.9 %	<u>85.3 %</u>	1.0 %	0.8 %
Angry	14.9 %	2.9 %	<u>76.9 %</u>	5.3 %
Happy	24.3 %	5.4 %	3.3 %	<u>67.0 %</u>

Table 2. Emotion discrimination from voice: females and males listening to males.

TOTAL	Neutral	Sad	Angry	Happy
<b>76.1 %</b>				
Neutral	<u>77.6 %</u>	17.2 %	2.9 %	2.2 %
Sad	14.7 %	<u>83.9 %</u>	1.1 %	0.4 %
Angry	14.6 %	1.2 %	<u>78.2 %</u>	6.0 %
Happy	26.2 %	5.3 %	3.6 %	<u>64.9 %</u>

Table 3. Emotion discrimination from voice: females and males listening to females.

<b>TOTAL</b>	Neutral	Sad	Angry	Happy
<b>77.9 %</b>				
Neutral	<u>79.5 %</u>	16.3 %	2.2 %	2.0 %
Sad	10.5 %	<u>87.1 %</u>	0.9 %	1.4 %
Angry	15.3 %	5.3 %	<u>75.2 %</u>	4.2 %
Happy	21.7 %	5.5 %	3.0 %	<u>69.8 %</u>

Table 4. Emotion discrimination from voice: males listening to females and males.

<b>TOTAL</b>	Neutral	Sad	Angry	Happy
<b>74.4 %</b>				
Neutral	<u>78.4 %</u>	15.9 %	3.3 %	2.4 %
Sad	14.4 %	<u>83.0 %</u>	1.6 %	1.0 %
Angry	16.8 %	3.8 %	<u>73.3 %</u>	6.1 %
Happy	26.8 %	6.0 %	4.3 %	<u>62.8 %</u>

Table 5. Emotion discrimination from voice: females listening to females and males.

<b>TOTAL</b>	Neutral	Sad	Angry	Happy
<b>79.7 %</b>				
Neutral	<u>78.4 %</u>	18.0 %	1.8 %	1.8 %
Sad	11.1 %	<u>87.9 %</u>	0.3 %	0.7 %
Angry	12.7 %	1.9 %	<u>81.1 %</u>	4.3 %
Happy	21.4 %	4.6 %	2.2 %	<u>71.7 %</u>

Table 6. Emotion discrimination from voice: males listening to males.

<b>TOTAL</b>	Neutral	Sad	Angry	Happy
<b>73.9 %</b>				
Neutral	<u>77.2 %</u>	16.6 %	3.7 %	2.5 %
Sad	15.9 %	<u>81.8 %</u>	1.8 %	0.5 %
Angry	16.4 %	1.8 %	<u>74.9 %</u>	6.9 %
Happy	28.3 %	5.6 %	4.7 %	<u>61.4 %</u>

Table 7. Emotion discrimination from voice: females listening to males.

<b>TOTAL</b>	Neutral	Sad	Angry	Happy
<b>78.8 %</b>				
Neutral	<u>78.2 %</u>	18.0 %	2.0 %	1.9 %
Sad	13.2 %	<u>86.3 %</u>	0.2 %	0.2 %
Angry	12.5 %	0.5 %	<u>82.0 %</u>	5.1 %
Happy	23.9 %	4.9 %	2.4 %	<u>68.8 %</u>

Table 8. Emotion discrimination from voice: males listening to females.

<b>TOTAL</b>	Neutral	Sad	Angry	Happy
<b>75.1 %</b>				
Neutral	<u>80.1 %</u>	14.9 %	2.7 %	2.3 %
Sad	12.5 %	<u>84.6 %</u>	1.3 %	1.6 %
Angry	17.3 %	6.5 %	<u>71.2 %</u>	4.9 %
Happy	24.9 %	6.6 %	3.9 %	<u>64.7 %</u>

Table 9. Emotion discrimination from voice: females listening to females.

TOTAL	Neutral	Sad	Angry	Happy
<b>81.1 %</b>				
Neutral	<u>78.8 %</u>	18.0 %	1.6 %	1.7 %
Sad	8.2 %	<u>90.1 %</u>	0.5 %	1.3 %
Angry	12.9 %	3.8 %	<u>80.0 %</u>	3.3 %
Happy	18.2 %	4.2 %	2.0 %	<u>75.6 %</u>

The average human emotion discrimination ability was approximately 77 %, which can be regarded as a good result in the light of earlier research. What is more interesting from the viewpoint of this paper is the systematic advantage of the female listeners.

## Discussion and conclusion

Looking at the results, it can be seen that the female subjects were better than the males in the emotion discrimination task in each setting: they were better (79 %) than the male listeners (74 %) at inferring emotional content also from the speech data produced by the male speakers. The male-male listening setting (74 %) in fact produced the lowest emotion discrimination performance level among the nine settings. The best results were, unsurprisingly, obtained in a setting involving females listening to female speakers (81 %). However, it cannot be argued that the male listeners were really poor with female speech data (75 %) bearing in mind the general accuracy results reported in the literature. All in all, the female speakers produced vocal portrayals of emotions which were easier to interpret (by both sexes) than those produced by the male speakers (78 % vs. 76 %). The emotional state which was best recognized in the whole data was sadness in the female-female setting (90 %); the most difficult emotion was happiness in the male-male setting (61 %). It would be interesting to speculate about the possible reasons for this: maybe males are not interested in finding happiness in fellow males while females are generally empathetic towards other females in distress?

That the female listeners/speakers were better with the vocal communication of emotion

than the male listeners/speakers is not surprising. A relevant concept in this context may, in fact, be *empathy*. Psychological research (see e.g. Tannen 1991) has shown that female superiority in empathizing is manifested in interaction by the following trends, for example: females' speech involves much more direct talk about feelings and affective states than "guy talk", females are usually more co-operative and reciprocal in conversation than males, and females are much quicker to respond empathically/emotionally to the distress of other people. It has been shown that, from birth, females look longer at faces, and particularly at people's eyes, while males are more prone to look at inanimate objects (Connellan et al. 2001).

The results of this study support the consensus view that, emotionally, females are more sensitive than males; this time concrete evidence is presented for the vocal (prosodic, non-lexical) communication of emotion. To draw more far-reaching conclusions, however, we need more speakers to produce the speech data, so that we can exclude the possible effect of speaker-specific idiosyncrasies on the results of the listening tests.

## References

- Batliner A., Fischer K., Huber R., Spilker J. and Nöth E. (2003) How to find trouble in communication. *Speech Communication* 40, 117-143.
- Connellan J., Baron-Cohen S. Wheelwright S., Ba'tki A. and Ahluwalia J. (2001) Sex differences in human neonatal social perception. *Infant behavior and Development* 23, 113-118.
- Scherer K. R. (1989) Vocal correlates of emotion. In Wagner H. and Manstead A. (eds.) *Handbook of Psychophysiology: Emotion and Social Behavior*, 165-197. London:Wiley.
- Scherer K.R., Banse R. and Walbott H.G. (2001) Emotion inferences from vocal expression correlate across languages and cultures. *Journal of Cross-Cultural Psychology* 32, 76-92.
- Tannen D. (1991) *You just don't understand: Women and men in conversation*. London:Virago.
- ten Bosch L. (2003) Emotions, speech and the ASR framework. *Speech Communication* 40, 213-225.