

Cross-Cultural Multimodal Interpretation of Emotional Expressions – An Experimental Study of Spanish and Swedish

Åsa Abelin

Department of Linguistics
University of Göteborg, Sweden
abelin@ling.gu.se

Abstract

This study presents an experiment in cross-cultural multimodal interpretation of emotional expressions. Earlier studies on multimodal communication have shown an interaction between the visual and auditive modalities. Other, cross-cultural, studies indicate that facial expression of emotions is more universal than prosody is. Cross-cultural interpretation of emotions could then be more successful multimodally than only vocally.

The specific questions asked in the present study are whether Swedish listeners can interpret Spanish emotional prosody, and whether simultaneously presented faces, expressing the same emotions, improve the interpretation. Audio recordings of Spanish emotional expressions were presented to Swedish listeners, in two experimental settings. In the first setting the listeners only attended to prosody, in the second one they also saw a face, expressing different emotions. The results indicate that cross-cultural interpretation of emotional prosody is improved by visual stimuli.

1. Introduction

The aim of this study is to investigate how speakers of Spanish and Swedish interpret multimodally expressed emotions of each other's language. The reason for choosing these two languages is that they are frequently claimed to express emotions differently, both prosodically and non-verbally, and could thus disprove claims of universality. More specifically the following questions are asked.

- Can Swedish speakers accurately interpret Spanish emotional prosody?
- Is there an influence between the auditive and visual channel in cross-cultural interpretation of emotional expressions?

Several studies have been made on the expression and interpretation of emotional prosody: for a review of different studies on emotional prosody, see e.g. Scherer (2003) where different research paradigms, methods and results are discussed.

The present investigation concerns the interaction between non-verbal and vocal expression of emotions. One question involved in this area of research is if emotional expressions – non-verbal or prosodic, are universal. Many researchers, beginning with Darwin, (1872/1965) have shown that some facial expressions are probably universal. In many cross linguistic studies of emotional prosody it has been shown that

emotional expressions are quite well interpreted, especially for certain emotions, for example anger, while other emotion words, for example joy, are less well interpreted (cf. Abelin and Allwood, 2000, Scherer, Banse and Wallbott, 2001). One possible explanation for this is that the expressions of some emotions vary more than that of other emotions, inter-individually as well as intra-individually; another possibility is that some emotions are expressed more in the gestural dimension. There could be evolutionary reasons for this (Darwin, 1872/1965). These studies also show that speakers are generally better at interpreting the prosody of speakers of their native language.

In the field of multimodal communication, there have been some studies of emotions, see e.g. overview in Scherer (2003:236), showing that judges are almost as accurate in inferring different emotions from vocal as from facial expression. The same study by Scherer shows that the emotions are more accurately identified in western as compared to non-western cultures. Massaro (2000), working under the assumption that multiple sources of information are used to perceive a persons emotion, as well as in speech perception, made experiments with an animated talking head expressing four emotions in auditory, visual, bimodal consistent and bimodal inconsistent conditions. Overall performance was more accurate with two sources of consistent information than with either source of information alone. In another study de Gelder and Vroomen (2000), asked participants to identify an emotion, given a photograph and/or an auditory spoken sentence. They found that identification judgments were influenced by both sources of information, even when they were instructed to base their judgment on just one of the sources.

Matsumoto et al (2002) review many studies on the cultural influence on the perception of emotion, and mean that there is universality as well as culture-specificity in the perception of emotion. What has been particularly studied is facial perception, where almost all studies show good interpretations of the six basic emotions. Universality in combination with display rules, i.e. culture specific rules for how much certain feelings are shown when there are other people present, is generally accepted in psychology. There are also studies that show that people adjust their interpretations of facial expressions to their *expectations* on intensity of expression in other cultures (Matsumoto & Ekman, 1989).

Normal human communication is multimodal and there has been a vast amount of studies in multimodal communication that shows interaction between the senses in perception (e.g.

Massaro, 2000, 2002). This research has shown that speech perception is greatly influenced by visual perception. There is reason to believe that emotional expressions are also interpreted in a holistic way.

Assuming that there is interaction between the senses, and that facial expression of emotion is more universal than prosody is, then cross-cultural interpretation of emotions should be more successful multimodally than only vocally.

Hypotheses:

- Swedish listeners can interpret Spanish emotional prosody to the same extent as Spanish listeners can interpret Swedish emotional prosody, (which is to a lesser extent than native speakers).
- The cross-cultural interpretation of emotional expressions is improved by multimodal stimuli.

2. Method

In the present article a method of elicitation is used where speakers are enacting emotional speech from the stimuli of drawings of facial expressions, originally used in therapy. The emotional expressions were elicited and recorded. Thereafter the speech material was presented to listeners for interpretation of the emotions expressed. The interpretations can be described as belonging to the attributional stage in Scherer's (2003) Brunswikian lens model of vocal communication of emotion. The listeners first listened to voices alone. Then they listened to voices in combination with looking at a facial expression, but were told to judge the voice. The results were compared to other studies.

2.1. Elicitation of speech material

Recordings were made of a male speaker of Spanish expressing eight different emotions. The method of elicitation was the following: the speaker was presented with the stimuli of schematic drawings of faces expressing emotions.

The faces were the following:

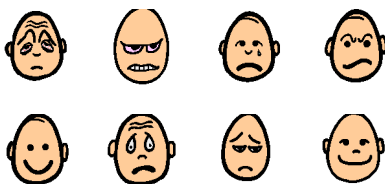


Figure 1: *The eight face stimuli used in the experiment*

The speaker was instructed to try to experience emotionally what the face was expressing and then express this emotion with the voice, while uttering the name "Amanda". The expression was recorded into the software PRAAT. After each recording of an emotion the speaker named the emotion he had just expressed.

In evoking vocal expression directly from facial stimuli, the possibility to get a greater control over the prosodic emotional expressions of speakers of different languages was assumed; this could be less probable if the speakers were to express emotions with the stimuli of emotion words.

The emotions expressed by the Spanish speaker were, according to himself, the following: 1. sad and tired, 2. angry, 3. sad 4. sceptical, 5. delighted, 6. afraid, 7. depressed, 8. very happy (cf. Figure 1).

2.2. Elicitation of listener's responses

The listener group consisted of 15 Swedish native speakers. They were first presented with the speech material over a computer's loud speakers, and named the different emotions they heard, one by one. The speech material was presented once. Later on the listeners were presented with the speech material at the same time as they saw the faces presented on a computer screen, and they named the different emotions as they heard/saw each expression¹. The faces were presented with the emotional expression that was produced for this particular face.

2.3 Comparison with earlier experiments

The results of the experiment were compared with results from a prosodic study (Abelin & Allwood, 2000) of a Swedish speaker who was interpreted by Spanish and Swedish speakers. In this experiment we did not use multimodal stimuli. Table 1 shows the different experimental conditions, which are compared.

Table 1: *Speakers and listeners in the experiments. (The empty slots represent ongoing experiments not yet analyzed.)*

speaker	Sw voice	Sp voice	Sp voice + face	Only face
listener				
Swedish	32	15	15	16
Spanish	25			2

3. Results

The Swedish interpretations of the Spanish speakers prosodic expressions of the emotions are shown below in Figure 2.

There are clear differences in how well the Swedish listeners interpreted the different emotions of the Spanish speaker. The two expressions of sadness were interpreted much more accurately than the other emotions – but only to 53% accuracy. The other emotions were interpreted quite poorly.

¹Utilizing the Psyscope software.

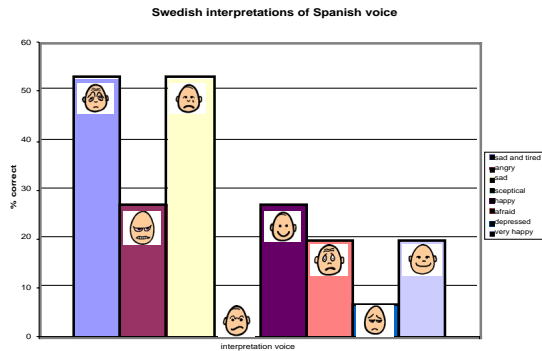


Figure 2: Swedish interpretations of Spanish prosody produced from facial stimuli. To the right are the classifications of the faces made by the speaker.

The multimodal results are presented below in Figure 3.

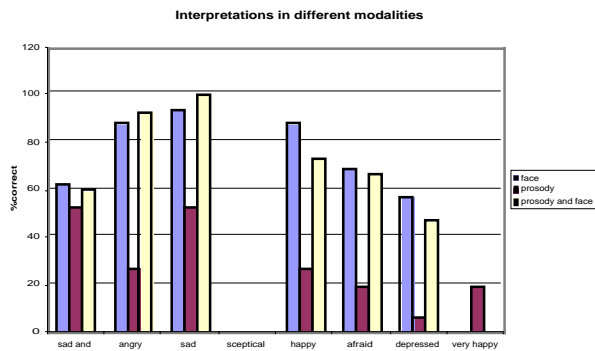


Figure 3: Swedish interpretations of a) the eight faces, b) emotional prosody expressed by the Spanish speaker, and c) simultaneous emotional prosody and facial expression. By correct interpretation is meant an interpretation that is the same as the intended expression as verbalized by the Spanish speaker¹.

Figure 3 shows five findings:

- 1) prosody alone is more difficult to interpret than multimodal stimuli. Prosody with simultaneous facial expression produces a much better interpretation for all but one emotion (very happy).
- 2) some emotions were easier to interpret than others.

¹ There is however reason to believe that the Spanish speaker named emotional expression number 4 less adequately; "sceptical" was not interpreted correctly by anyone, in any modality, but there was a consensus among the listeners to interpret voice and voice+ face as "questioning". The Swedish interpretation of the face only was "angry" to 50%.

- 3) only the face is often easier to interpret than voice + face; adding the voice makes the interpretation less accurate for four of the emotions.
- 4) there were some emotions that were more difficult to interpret in all modalities, while others were easier in all modalities, cf. sad, happy, afraid, depressed.
- 5) the emotion that is recognized relatively well in vocal expression is sadness.

The results for the interpretations of only prosody, in Figure 3, are now compared with the results of Abelin & Allwood (2000) where Spanish speakers interpreted Swedish emotional prosody. The comparison is made for four of the emotions: angry, sad, happy, afraid, in Figure 4.

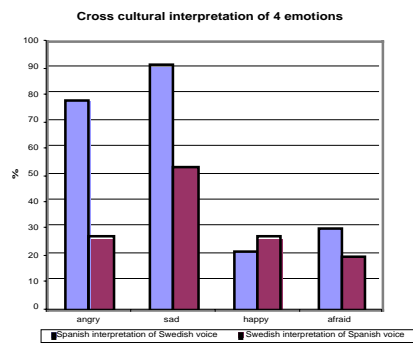


Figure 4: The diagram compares cross-cultural interpretations of four emotions expressed by Spanish and Swedish speakers.

Figure 4 compares cross-linguistic interpretations of four basic emotions. It shows that:

- 1) sadness is interpreted best, for both languages,
- 2) anger is well interpreted by Spanish speakers listening to Swedish, but poorly interpreted by Swedish speakers listening to Spanish
- 3) happiness and fear was interpreted quite poorly cross-linguistically by both groups
- 4) the Spanish group was generally better at interpreting the Swedish speaker than vice versa.

4. Discussion

The results show that perception of emotional expressions are more successful when the stimuli are multimodal, in this cross-linguistic setting. The cross-cultural prosodic comparison with the earlier experiment of Abelin & Allwood shows that the Spanish listeners were better at interpreting Swedish than vice versa. This is presently studied further with Spanish and Finnish listener groups as well as with Swedish and Finnish speaker groups. The present study also shows that certain emotions (happy and afraid) are more difficult to interpret from prosody, by both language groups.

There are a number of problems involved in the study of cross-cultural interpretation of linguistic phenomena such as the expression of emotions. There are translation problems due to different categorizations of the emotional spectrum, different display rules for different emotions, listeners differing knowledge of different display rules, word finding problems etc. This study has tried to handle the translation problem by evoking prosodic expressions directly from facial stimuli. Expectations on different display rules are avoided with listeners not knowing the native language of the speaker.

5. Conclusions

- Swedish listeners could not, in the comparison of the two studies, interpret Spanish emotional prosody to the same extent as Spanish listeners could interpret Swedish emotional prosody. There were also differences between the emotions.
- Emotions were more appropriately interpreted cross-culturally when both the visual and auditory signals were present, than when only the auditory signal was present. Visual stimuli alone gave the best interpretations.

There is reason to believe that the facial expression of emotion is more universal than prosodic expression. The prosodic production of emotional expressions could be universal, at least for certain emotions, but the emotional prosody is never heard in isolation, but always in combination with the speech prosody of each particular language. Another possibility, which will be studied further, is if certain emotions are more dependent on prosodic information and other emotions more dependent on facial expression. This study is presently expanded with more speakers and listener groups.

6. References

- [1] Abelin, Å., Allwood, J., 2000. Cross linguistic interpretation of emotional prosody. *ISCA workshop on Speech and Emotion*. Newcastle, Northern Ireland, 110–113.
- [2] Cornelius, R. R., 2000. Theoretical approaches to emotion. *Proceedings of the ISCA Workshop on Speech and Emotion*. Newcastle, Northern Ireland, 3–8.
- [3] Darwin, C., 1872/1965. *The expression of the emotions in man and animals*. Chicago: University of Chicago Press.
- [4] De Gelder, B. & Vroomen, J., 2000. The perception of emotions by ear and eye. *Cognition and Emotion*, 14, 289–311.
- [5] Massaro, D. W., 2000. Multimodal emotion perception: Analogous to speech processes. *Proceedings of the ISCA Workshop on Speech and Emotion*, Newcastle, Northern Ireland, 114–121.
- [6] Massaro, D. W., 2002. Multimodal Speech Perception, in B. Granström, D. House and I. Karlsson, Eds, *Multimodality in Language and Speech Systems*, Dordrecht: Kluwer Academic Publishers.
- [7] Matsumoto, D., Franklin, B., Choi, J.-W., Rogers, D. Tatani, H., 2002. Cultural influences on the Expression and Perception of Emotion” in W.B. Gudykunst and B. Moody, Eds. *Handbook of International and Intercultural Communication*, Sage Publications.
- [8] Matsumoto, D., Ekman P., 1989. American-Japanese differences in intensity ratings of facial expressions of emotion. *Motivation and Emotion*, 13, 143-157.
- [9] Scherer, K. 2003. Vocal communication of emotion: a review of research paradigms. *Speech Communication* 40 (1). 227-256.
- [10] Scherer, K. R., Banse, R., and Wallbott, H. G., 2001. Emotion inferences from vocal expression correlate across languages and cultures, *Journal of Cross-Cultural Psychology*, 32 (1), 76–92.