Change of Perception of Emotions in Multimodal and Crosslinguistic Setttings

Åsa Abelin

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

Abstract

This paper adresses two questions. The first question is if perception of emotional prosody changes when visual stimuli are present. It has already been shown by others that the perception of vowels and consonants is augmented or changed when visual information is present.The question is then if perception of such categories as emotions, mainly expressed by prosody, are also affected by multimodal stimuli. The second question concerns the cross-linguistic perception of multimodal stimuli. Earlier cross-cultural studies indicate that facial expression of emotions is more universal than prosody is. Cross-linguistic interpretation of emotions could then be more successful multimodally than only vocally. The specific questions asked in the present study are to what degree Spanish and Swedish listeners can interpret Spanish emotional prosody, and whether simultaneously presented faces, expressing the same emotions, change or improve the interpretation. Audio recordings of Spanish emotional expressions were presented to Spanish and 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 intra-linguistic as well as cross-linguistic perception of emotional prosody is improved by visual stimuli. Cross-linguistic interpretation of prosody is more poorly accomplished than inter-linguistic, but seems to be greatly augmented by multimodal stimuli. F0 analysis of the stimuli is made and discussed in relation to the results of he perceptions of the stimuli of the different language groups. The results are also analysed in terms of gender of listeners.

1. Introduction

The aim of this study is to investigate how speakers of Spanish and Swedish interpret vocally and multimodally expressed emotions of Spanish. More specifically the following questions are asked:

- Can Swedish as well as Spanish speakers accurately interpret Spanish emotional prosody?
- Is there an influence between the auditive and visual channel in cross-linguistic 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 [1] 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 is if emotional expressions – non-verbal or prosodic, are universal. Many researchers, beginning with Darwin, [2] 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 [3, 4]. One possible explanation for this is that the expressions of some emotions vary more than that of other emotions; another possibility is that some emotions are expressed more in the gestural dimension. There could be evolutionary reasons for this [2]. 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 [1] showing that judges are almost as accurate in inferring different emotions from vocal as from facial expression. Massaro [5, 6], 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 [7], 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 [8, 9] 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.

Normal human communication is multimodal (cf. Rosenblum [10]) and there has been a vast amount of studies in multimodal communication that shows interaction between the senses in perception [5]. This research has shown that speech perception is greatly influenced by visual perception.

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

Hypotheses:

- Swedish listeners can interpret Spanish emotional prosody to a lesser extent than Spanish listeners can interpret Spanish emotional prosody.
- The cross-linguistic 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 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.

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:

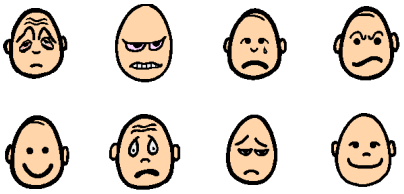

Figure 1: The eight face stimuli used in the experiment

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, and this label was used as the correct answer in the following listening experiments. In this way the facial stimuli can be used for speakers of several languages, and cannot be said to be culture specific.

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 where 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 and 10 Spanish 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. They 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 expression[1]. The faces were presented with the emotional expression that was produced for this particular face. They were told that the emotions expressed were partly different from in the first experiment.

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, in comparison with the Spanish listener. 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.

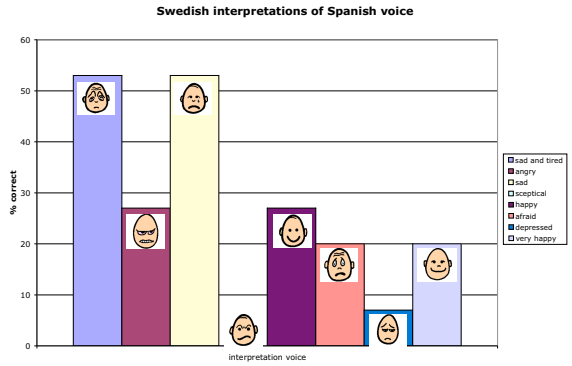

Figure 2: Swedish interpretations of Spanish prosody produced from facial stimuli. To the right are the classifications of the faces made by the speaker.

The results of Figure 2 are presented in comparison with results of only facial and with multimodal stimuli in Figure 3.

[1] Utilizing the Psyscope software.

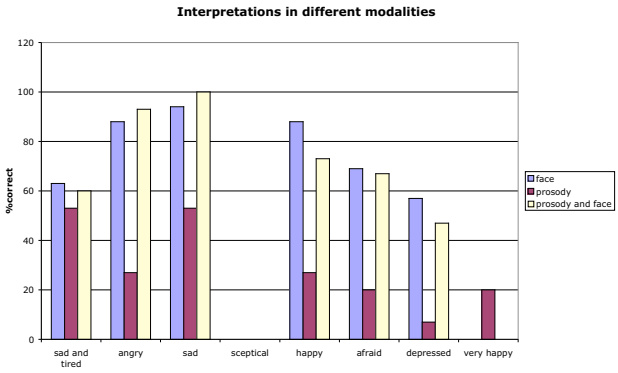

Figure 3: Swedish interpretations of a) the eight faces, b) emotional prosody expressed by the Spanish speaker, and c) simultaneous emotional prosody and facial expression. By correct interpretation is meant an interpretation that is the same as the intended expression as verbalized by the Spanish speaker[1].

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.
3) the face alone is often easier to interpret than voice + face; adding the voice makes the interpretation less actcurate 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.

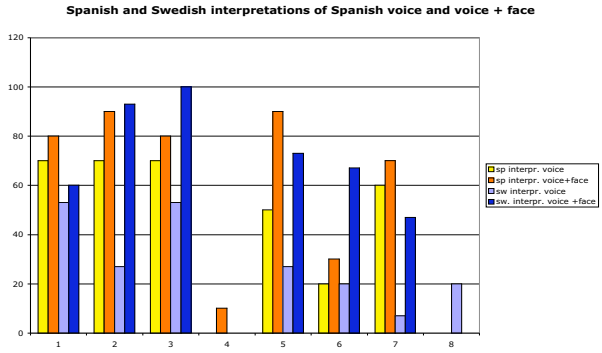

Figure 4: Percent correct interpretations of Spanish voice and voice+face of the eight different emotions. Spanish and Swedish listeners.

Finally we can add to Figure 3 the Spanish interpretations of Spanish vocal and multimodal stimuli, see figure 4. The following conclusions can then be added.

6) Interpretation of emotional expressions is always more successful when the stimuli are multimodal, also for the Spanish listeners
7) The Swedish listeners, who were poor at interpreting the Spanish voice, seem to be greatly aided by the face; the difference between the oral and multimodal interpretations were larger for the Swedish than for the Spanish listeners
8) For multimodal stimuli none of the language groups was generally better at interpreting the stimuli

4. F0 analysis

An F0 analysis of the stimuli shows that the speakers F0 is generally very high. Almost all of the listeners also interpreted him as a woman. The emotion with the highest maximum F0 is "fear", thereafter "sceptical" and the two variants of "happy". The emotions with the lowest minimum F0 is "angry". The ones with the lowest maximum F0 are "sad", "depressed" and "angry". Generally high maximum F0 is accompanied by high minimum F0 and vice versa. The exception is "sceptical". The emotion with the highest F0 is "afraid", which is in accordance with many other studies, see e.g. Laukka [11].

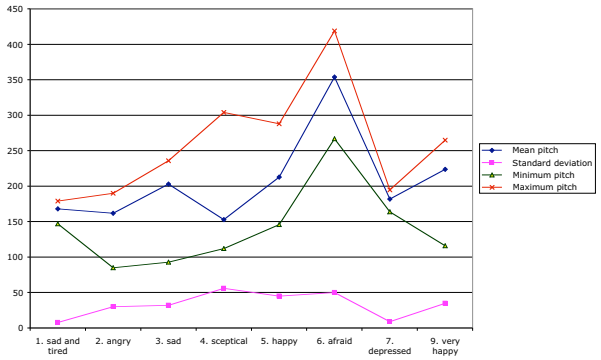

Figure 5: F0 mean, variation, maxima and minima for the eight different emotions.

The F0 values can be related to the interpretations in the following way: the least successfully interpreted emotion (by both language groups) – sceptical – is the one with the largest variation. Sad and tired which is the emotion with the least variation is interpreted best from voice, by both language groups.

5. Gender comparison

The interpretations of Swedish men and women were compared to see if either group was generally better at interpreting the intended emotions. There were 8 men and 7 women in the study. No such tendencies were found. The Spanish group consisted of mainly men and therefore no gender comparison was made.

1 There is however reason to believe that the Spanish speaker did not produce this emotion in accordance with the face; "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%.

6. Discussion

The results show that perception of emotional expressions are more successful when the stimuli are multimodal, in this cross-linguistic setting, even though the facial stimuli are very schematic. The present study 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.

7. Conclusions

- Emotions are more appropriately interpreted intra- and cross-linguistically when both the visual and auditory signals are present, than when only the auditory signal is present. Visual stimuli alone give the best interpretations.
- Some emotions are more easily interpreted, both intra- and cross-linguistically, prosodically and multimodally, e.g. sadness.

There is a possiblitity that the facial expression of emotion is more universal than prosodic expression. The prosodic production of emotional expressions per se 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 also be studied further, is if certain emotions are more dependent on prosodic information and other emotions more dependent on facial expression.

Since the speaker had a high pitched voice which made many, but not all, speakers interpret him as a woman, experiments will be made were the listeners are told that the speaker is either a man or a woman, in order to see if this contextual information will affect the interpretations.

8. References

[1] Scherer, K. 2003. Vocal communication of emotion: a review of research paradigms. *Speech Communication* 40 (1). 227-256.

[2] Darwin, C., 1872/1965. *The expression of the emotions in man and animals*. Chicago: University of Chicago Press.

[3] Abelin, Å..; Allwood, J., 2000. Cross linguistic interpretation of emotional prosody. *ISCA workshop on Speech and Emotion. Newcastle*, Northern Ireland, 110–113.

[4] 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

[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] De Gelder, B. & Vroomen, J., 2000. The perception of emotions by ear and eye. *Cognition and Emotion*, 14, 289–311.

[8] 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.

[9] Matsumoto, D., Ekman P., 1989. American-Japanese differences in intensity ratings of facial expressions of emotion. *Motivation and Emotion, 13,* 143-157.

[10] Rosenblum, L. D., 2005. Primacy of Multimodal Speech Perception in D. B. Pisoni and R. E. Remez, Eds. *The Handbook of Speech Perception*, Blackwell publishing

[11] Laukka, P., 2004. *Vocal Expression of Emotion – Discrete-emotions and Dimensional Accounts*. Uppsala. Acta Universitatis Upsaliensis