A Cross-Cultural Study on the Effectiveness of Visual and Vocal Channels in Transmitting Dynamic Emotional Information

Maria Teresa Riviello[1,2] and Anna Esposito[1,2]

[1] Seconda Università di Napoli, Department of Psychology, via Vivaldi, 43, Caserta, Italy

[2] IIASS, Vietri sul Mare, Salerno, Italy
    e-mails: mariateresa.riviello@unina2.it; iiass.annaesp@tin.it

Abstract: The present work aims to examine emotion recognition within and across cultures. It reports on a set of perceptual experiments designed to explore the human ability to identify emotional expressions dynamically presented through visual and auditory channels. Two cross-modal databases of dynamic verbal and non-verbal emotional stimuli based on video-clips extracted from American and Italian movies, respectively, were defined and exploited for the experiments. In the first study, American, French, and Italian subjects were involved in a comparative analysis of subjective perceptions of six emotional states dynamically portrayed by visual and vocal cues, exploiting the database of American emotional stimuli. In the second study, American and Italian subjects were tested on their ability to recognize six emotional states through the visual and auditory channel, exploiting the database of Italian emotional stimuli. The aim is to investigate if there exists a difference in the efficacy of the visual and auditory channels to infer emotional information and if cultural context, in particular the language, may influence this difference. This hypothesis is investigated including as participants in each of the two studies, one group of native speakers of the language, belonging to the same cultural context of the video-clips used as stimuli (i.e., American subjects for the American stimuli in the former study, and Italian subjects for the Italian stimuli in the latter one). Results showed that emotional information is affected by the communication mode and that language plays a role.

Keywords: emotion recognition; vocal expressions; facial expressions; cultural differences

1   Introduction

In natural social interactions, the communication of emotions is an example of a multi-modal transfer of information: both the visual and auditory channels are involved simultaneously in the transmission and perception of emotions. Understanding the relationship between verbal and non-verbal communication modes, and progress towards their modelling, are crucial for implementing a

friendly human/computer interaction which exploits synthetic agents and sophisticated human-like interfaces and will simplify user access for future telecommunication services. The task has received much attention in the last few years in the context of Human Computer Interaction (HCI), which involves research aimed at investigating the perceptual and cognitive role of visual and auditory channels in conveying emotional information, in order to identify automatic methods and procedures capable of recognize human emotional states exploiting the multimodal nature of emotional expressions [1, 4].

A considerable part of the research on emotions and the related perceptual cues to infer them has focused separately on three expressive domains: facial expressions, voice and body movements. On these research lines, some studies maintain that facial expressions are more informative than gestures and vocal expressions [8, 16] whereas others underline the faithfulness of vocal expressions in portraying emotional states since physiological processes, such as respiration and muscle tension, are naturally influenced by emotional responses [2-3, 23]. In this debate, the data reported in literature seem to favour facial expressions [5, 15].

Nevertheless, most of the studies investigating the perceptual and automatic recognition of emotional facial expressions have exploited static images (see the FACS [5], the Japanese Female Facial Expression (JAFFE) database [18], and the ORL Database of Faces [22]). However, spontaneous emotional facial expressions are dynamic stimuli evolving along time through a change of the facial musculature, as speech and therefore a more appropriate comparison is needed.

The present work aims to investigate the effectiveness of visual and auditory channels in conveying emotional information exploiting dynamicity also in facial expressions. For this purpose, two cross-modal databases comprising dynamic verbal and non-verbal emotional stimuli based on video-clips extracted from American and Italian movies were created and a series of perceptual experiments were defined. The collected stimuli allowed us to characterize emotional cues of some basic emotions dynamically conveyed by the visual and auditory channels and investigate on the preferential channels exploited by humans to perceive emotional states [10-13].

In a cross cultural perspective, it would be worth investigating if the ability to recognize emotional expressions as a function of the channel is also affected by the cultural context and in particular, by the subject's native language. Psychologists have long debated whether emotions are universal versus whether they vary by culture. These issues have been extensively summarized elsewhere and we do not reiterate them [6, 9, 17, 19-21, 26]. Recent theoretical models have attempted to account for both universality and cultural variation by specifying which particular aspects of emotions show similarities and differences across cultural boundaries [9, 14, 19, 24-25].

Along this line, our approach is based on the assumption that culture and language-specific paralinguistic and visual patterns may influence the decoding

process of emotions and that the familiarity of the language and the expositions of the subjects to the cultural environment affect the recognition of the emotional states in particular when they are vocally expressed [12, 24].

Towards this aim, this work investigates the amount of emotional content that participants from different cultures[1] can infer from audio and mute video emotional data extracted from two different cultural contexts and played in two different languages: American English (as a globally spread language) and Italian (as a country-specific language).

In the former study, perceptual experiments devoted to assessing the subjective perception of emotional states, exploiting American emotional audio and mute video stimuli, were conducted on separate groups of American, French and Italian subjects; whereas in the latter study, separate groups of American and Italian subjects were tested, using the same procedure, on their ability to recognize emotional states, exploiting the Italian database of emotional stimuli [11].

2   The Cross-Modal Emotional Databases

The collected data are based on extracts from American and Italian movies whose protagonists were carefully chosen among actors and actresses who are well regarded by critics and considered capable of giving some very real and careful interpretations. The use of audio and video stimuli extracted from movies provided a set of realistic emotional expressions [10-13]. Differently from the other existing emotion databases proposed in literature, in this case the actors/actresses had not been asked to produce an emotional expression, but rather they were acting according to a movie script and their performance had been judged as appropriate to the required emotional context by the movie director (supposed to be an expert). Moreover, even though the emotions expressed in such video-clips were simulations under studio conditions (and may not have reproduced a genuine emotion but a stylized version of it) they were able to catch up and engage the emotional feeling of the spectators (the addressers) and therefore provided more confidence on the value of their perceptual emotional content.

Each database consisted of audio and video stimuli representing 6 emotional states: *happiness*, *sarcasm/irony*, *fear*, *anger*, *surprise,* and *sadness* (except for sarcasm/irony, the remaining emotions are considered by many theories as basic and therefore universally shared [2, 8, 16, 23]). For each database and for each of the above emotional states, 10 stimuli were identified, 5 expressed by an actor and

---

[1]   The term "culture" is here used as a theoretical concept (in a very general sense), to identify the geographical location, language, history, lifestyle, customs and traditions of a country.

5 by an actress, for a total of 60 American and 60 Italian video-clips, each acted by a different actor and actress to avoid bias in their ability to portray emotional states. The stimuli were short in duration (the average length was 3s, SD = ± 1s) to avoid the overlapping of emotional states that could confuse the subject's perception. Care was taken in selecting video clips where the protagonist's face and the upper part of the body were clearly visible. In addition, the semantic meaning of the produced utterances did not clearly express the portrayed emotional state and its intensity level was moderate. For example, the stimuli of sadness, where the actress/actor was clearly crying, or stimuli of happiness, where the protagonist was strongly laughing, were not included in the data. This was an attempt to allow the participants to observe less obvious emotional cues generally employed in a very natural and not extreme emotional interaction.

The emotional labels assigned to the stimuli were first given by two experts and then by three naïve judges independently. The expert judges made a decision on the stimuli carefully exploiting emotional information within facial and vocal expressions such as a frame by frame analysis of changes in facial muscles, and F0 contour, rising and falling of intonation contour, etc., as reported by several authors in literature [5, 27-28] and the contextual situation the protagonist was interpreting. The naïve judges made their decision after watching the stimuli several times. There were no opinion exchanges between the experts and naïve judges and the final agreement on the labelling between the two groups was 100%.

The collected stimuli extracted from movie scenes contain environmental noise and therefore are also useful for testing realistic computer applications. The database is available in the context of the COST Action 2102 (cost2102.cs.stir.ac.uk/) and can be acquired by mailing the second author (iiass.annaesp@tin.it, anna.esposito@unina2.it) of this paper.

Both for the American and Italian database, the audio and mute video alone were extracted from each complete audio-video stimulus (video-clip) coming up with a total of 120 American and 120 Italian stimuli: 60 mute videos and 60 audio stimuli for each database. The stimuli were randomly presented in the first study to separate groups of American, French and Italian participants and in the second one to separate groups of American and Italian participants.

3   American, French and Italian Participants Tested on American Audio and Mute Video Stimuli

3.1   Participants and Testing Procedure

A total of 180 subjects (60 Americans, 60 French and 60 Italians) participated in the perceptual experiments exploiting the American database of emotional stimuli.

For each nationality, 30 subjects (equally distributed for gender) were involved in the evaluation of the American audio and 30 in the evaluation of the American mute video stimuli. The participants' age was similar among countries, ranging from 18 to 35 years. The knowledge of English by Italian and French subjects was comparable, since all of them used it as a second language.

The subjects were randomly assigned to the task and were required to carefully listen to or watch the stimuli via headphones in a quiet room. They were instructed to pay attention to each presentation and decide which of the 6 emotional states were expressed. Responses were recorded on a matrix paper form (60x8), where rows listed the stimuli numbers and columns the 6 selected emotional states plus "*others*" indicating any other emotion not listed and the option of "*no emotion",* which was suggested when according to the subject's feeling the protagonist did not show emotions. For each emotional stimulus the percentage of correct emotion recognition was computed.

3.2   Results

The data obtained from each group of subjects (60 American, 60 French and 60 Italian) evaluating the American emotional stimuli were first analyzed separately.

This analysis provided the percentages of label agreement (recognition accuracy) produced by the representatives of each country, split in subgroups of 30 and tested separately on the American audio and mute video stimuli. For each emotional state, the percentages were computed considering the number of correct answers provided by participants over the total number of expected correct answers. Since the American subjects share both the language and the cultural background of the selected emotional expressions with the encoders of the stimulus material, their performance can be considered as a reference for an optimal identification of the emotional states under examination. Figures 1, 2, and 3 display the percentage of recognition accuracy for the American, French, and Italian subjects, respectively.
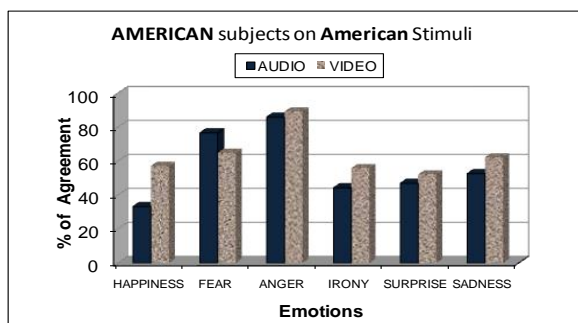

Figure 1
Percentage of agreement obtained under the two experimental conditions (only audio – black bars – and only mute video – gray bars) by the American subjects tested on American stimuli
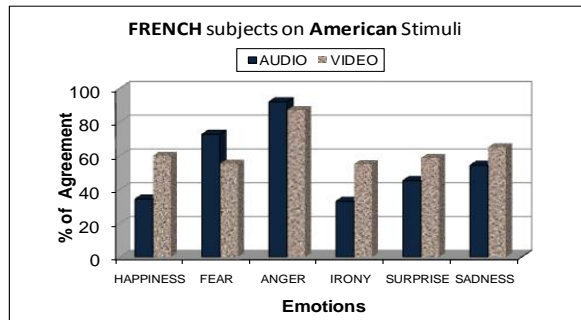
Figure 2
Percentage of agreement obtained under the two experimental conditions (only audio – black bars –
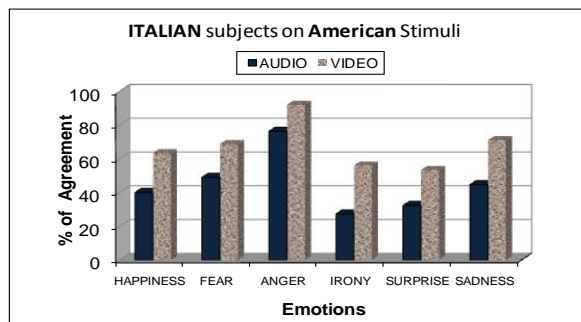and only mute video – gray bars) by the French subjects tested on American stimuli

Figure 3
Percentage of agreement obtained under the two experimental conditions (only audio – black bars –
and only mute video – gray bars) by the Italian subjects tested on American stimuli

An ANOVA analysis was performed on the data obtained by the American, French and Italian subjects evaluating the American audio and video emotional stimuli separately. In the analysis, the *Perceptual Mode* was considered as a between subject variable and the *Emotions and Actor's Gender* as within subject variables. Significance was established for α=.05.

The ANOVA shows that for American (F (1, 8)= 1.696, ρ= .22) and French (F (1, 8)= 3.427, ρ= .1013) subjects, the American mute video and audio convey the same amount of emotional information. This is not the case for Italian subjects (F (1, 8)= 33.74, ρ=.0004), where the perceptual mode plays a significant role. It is worth noting that for Americans, the identification of stimuli is not affected by the specific emotion portrayed (F (5,40)= 1.401, ρ= .24), while this is not true for French (F (5,40)= 2.520, ρ= .04) and Italian (F (5,40)= 4.050, ρ= .0046) subjects, who showed a preference in inferring emotional information of fear (especially from vocal cues for French, and from visual cues for Italians), happiness and sadness (both French and Italian subjects obtained a high percentage of correct recognition on mute video stimuli), and anger (which is very well recognized both from the audio and mute video).

In addition, the gender of the protagonist does not affect the recognition accuracy of the American (F (1, 8)= 3.254, ρ= .10), French (F (1, 8)= 6.443, ρ= .03), and Italian (F (1, 8)= 4.370, ρ= .07) subjects. However, an interaction between the category of emotion and the actor's gender was found for American (F (5, 40)= 9.721, ρ= .0001), French (F (5, 40)= 7.156 ρ= .0001) and Italian (F (5, 40)= 8.532, ρ= .0001) subjects. This interaction is significant for all the emotional categories under examination except for sadness and anger. Anger is also the emotional category with the highest percentage of recognition accuracy independently of the nationality and the perceptual mode.

In order to be able to assess the effectiveness of communication modes (auditory and visual) in conveying emotional information, Figures 4 and 5 report the percentage of correct agreement, on the American audio and video stimuli, obtained by the American, French and Italian subjects.
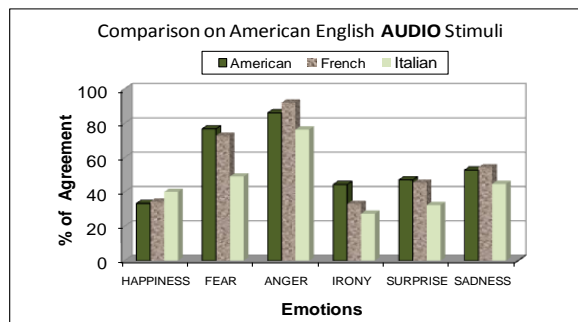

Figure 4

Percentage of agreement obtained under the audio alone experimental condition by the American (black bars), French (grey bars) and Italian (green bars) subjects tested on American audio stimuli
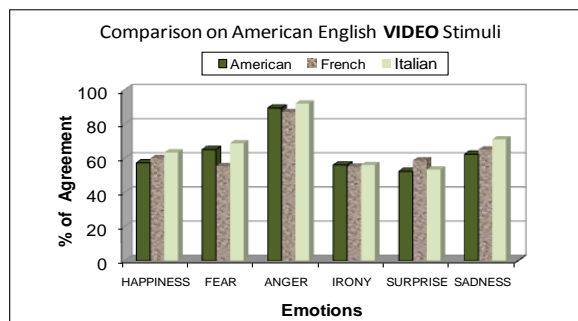

Figure 5

Percentage of agreement obtained under the mute video experimental condition by the American (black bars), French (grey bars) and Italian (green bars) subjects tested on American video stimuli

An ANOVA analysis was performed considering the subjects' *Nationality* as a between subject variable and the *Emotions and Actor's Gender* as within subject variables. Significance was established for α= .05. The analyses show that, when American, French and Italian subjects are tested on American Audio stimuli, nationality plays a significant role (F (2, 12)= 4.288, ρ=.04). A post hoc test revealed that Italian subjects differ significantly both from French and American subjects for the audio, whereas no significant differences were found for the American mute video stimuli (F(2,12)=.324 , ρ=.7294) and hence, it could be hypothesized that the visual channel shares universal emotional features across cultures.

4   American and Italian Subjects Tested on Italian Audio and Video Stimuli

4.1   Participants

A total of 120 subjects (60 Americans, and 60 Italians) were involved in the evaluation of the Italian emotional stimuli. For each nationality, 30 subjects (15 females and 15 males, aged from 18 to 35 years) were tested separately on the audio and mute video stimuli. The American subjects however, differently from the participants to the first study, do not speak Italian as a second language. The testing procedure was exactly the same as reported in section 3.1. Moreover, in such case, the percentage of correct emotion recognition was computed for each emotional stimulus.

4.2   Results

The data obtained on the Italian stimuli from each group (60 American and 60 Italian subjects) were first analyzed separately.

This analysis provided the percentages of recognition accuracy produced by the American and Italian subjects, split into subgroups of 30 and tested separately on Italian audio and mute video stimuli. For each emotional state, the percentages were computed considering the number of correct answers provided by the participants over the total number of expected correct answers. In this case, the performance of the Italian subjects was as a reference for an optimal identification of the emotional states under examination, since Italians are native speakers of the language and share the cultural background of the selected audio-video emotional scenes. Figures 6, 7, display the percentage of the recognition accuracy for the American and Italian subjects, respectively.
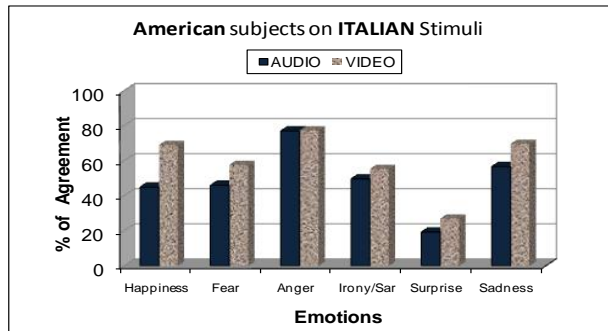
Figure 6
Percentage of agreement obtained under the two experimental conditions (only audio – black bars –
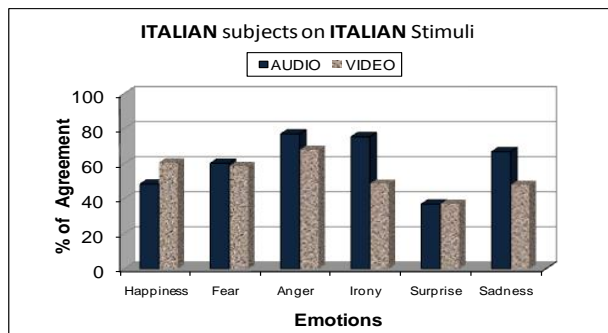and only mute video – gray bars) by the American subjects tested on Italian stimuli

Figure 7
Percentage of agreement obtained under the two experimental conditions (only audio –black bars – and
only mute video –gray bars) by the Italian subjects tested on Italian stimuli

An ANOVA analysis was performed on the frequency of correct answers obtained by the American and Italian subjects, when tested on the Italian audio and mute video stimuli separately. The *Perceptual mode* was considered as a between subjects variable and the *Emotions and Actor's Gender* as within subjects variables. Significance was established for $\alpha=.05$.

The ANOVA shows that for American subjects Italian mute video and audio alone convey the same amount of emotional information (F $(1, 8)= 4.114$, $\rho= .077$), whereas, for Italian subjects there exists a significant difference between the amount of emotional information inferred through the audio and the video channels (F $(1, 8)= 5.517$, $\rho= .468$).

The recognition of a given emotional state significantly depends (both the American (F $(5, 40)= 6.493$, $\rho= .00002$), and Italian (F $(5, 40)= 6.023$, $\rho= .0003$) subjects) on the portrayed emotion. However, for American subjects this is true independently from the perceptual mode (F $(5, 40)= 1.643$, $\rho= .17$), whereas for the Italian subjects an interaction was found between the perceptual mode and the identification of the emotional states (F $(5, 40)= 3.051$, $\rho= .02$).

The actor's gender does not affect the recognition accuracy of emotions for either the American (F (1, 8)= 3.373, ρ= .10), or Italian (F (1, 8)= .211, ρ= .6581) group. No interaction between the emotion category and the actor's gender was found for the Italian subjects (F (5, 40)= 7.533, ρ= .0001). On the other hand, an interaction was found for the American subjects (F (5, 40)= 7.533, ρ= .0001) and post-hoc tests revealed that it is significant for happiness (F (1, 8)= 12.056 ρ=.008), anger (F (1, 8)=19.892, ρ= .002) and surprise (F (1, 8)= 10.86, ρ= .01). In particular, for the American subjects, emotional states portrayed by actresses (F (5, 40)= 1.584, ρ= .187) are better recognized than those portrayed by actors (F (5, 40)= 11.887, ρ= .0001).

It is worth highlighting that in both studies and for both modes (audio and mute video), anger is the emotion that gets the highest percentage of recognition accuracy from all the groups involved, suggesting a perceptually privileged position of anger among the emotions under examination, probably for the phylogenetic value of its clear survival function [7].

With the aim of assessing the effectiveness of the auditory and visual communication modes, the following figures display the percentage of correct agreement obtained by the American and Italian subjects, tested on the Italian audio (Figure 8) and mute video (Figure 9) stimuli.

An ANOVA analysis was performed considering the *Nationality* as a between subject variable and the *Emotions and Actor's Gender* as within subject variables. Significance was established for α= .05. The analyses show that, when American, and Italian subjects are tested on Italian Audio stimuli, nationality plays a significant role (F (1, 8)= 20.987, ρ=.002), whereas no significant differences were found for Italian mute video stimuli (F (1, 8)= 1.055, ρ=.33), supporting the hypothesis that the visual channel shares universal features for a cross-cultural identification of emotional states.
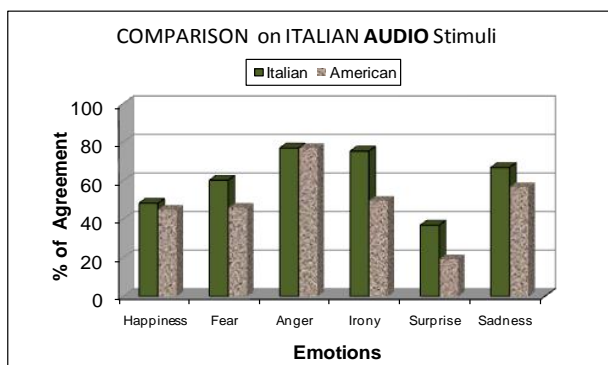

Figure 8

Percentage of agreement obtained under the audio alone experimental condition by the American (black bars), and Italian (grey bars) subjects tested on Italian audio stimuli
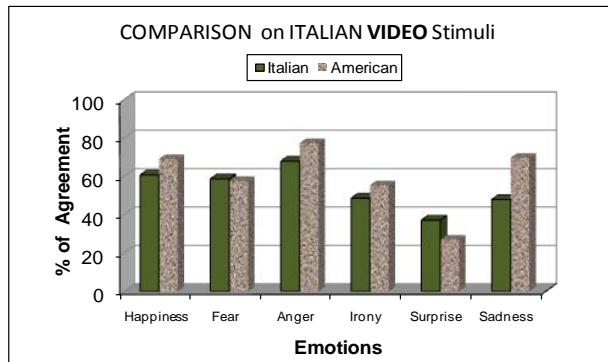
Figure 9

Percentage of agreement obtained under the mute video experimental condition by the American (black bars), and Italian (grey bars) subjects tested on Italian video stimuli

Conclusions

The present work reports results from cross-cultural perceptual experiments aimed at exploring the human ability to recognize emotional expressions dynamically portrayed through the visual and auditory channel, investigating if one channel is more effective than the other for inferring emotional information and if this effectiveness is affected by the cultural context and in particular, by the language. The perceptual experiments were based on dynamic visual and vocal cues exploiting two databases of audio and mute video emotional stimuli based on video-clips extracted from American and Italian movies, respectively. The databases allowed a cross-modal analysis of audio and video recordings with the aims of defining and identifying distinctive, multi-modal and cultural specific emotional features from multi-modal signals, as well as for new methodologies and mathematical models for the automatic implementation of naturally human-like communication interfaces.

In the first study, American, French, and Italian subjects were involved in the evaluation of emotional data, exploiting a database of American emotional audio and mute video stimuli. The three groups were compared on their ability to identify dynamic facial and vocal emotional expressions encoded by exploiting the auditory and visual communication mode separately. The results suggest that the ability to recognize emotional expressions as a function of the communication mode is affected by the cultural context and in particular by the language. In fact, American and French subjects are able to perform equally well both on the visual and vocal cues, whereas Italian subjects rely more on visual information. The data obtained in the second study, where American and Italian subjects were tested on their ability to recognize emotions exploiting a database of Italian emotional audio and mute video stimuli, seems to further support the above hypothesis. In this case, the American and Italian subjects' ability in inferring emotional information from visual cues is comparable, whereas the Americans perform differently from

Italian subjects when tested on their ability to identify emotional vocal expressions.

We hypothesize that speakers of different languages may exhibit a different sensitivity to vocal emotional information. It is possible that at the base of the vocal emotional encoding there is a more language-specific process strictly related to the prosodic feature of the subject's native language.

Cultural specificity seems not to affect the recognition of emotional visual information: The visual channel shares visual emotional features across cultures, supporting the data presented in literature [7-8].

More data are needed to support the above hypotheses by extending the proposed perceptual experiments to the members of other Western and non-Western countries.

Acknowledgement

References

[1]   V. Aubergé, M. Cathiard: Can We Hear the Prosody of Smile? *Speech Communication*, 2003, 40, pp. 87-97

[2]   R. Banse, K. Scherer: Acoustic Profiles in Vocal Emotion Expression. *Journal of Personality & Social Psychology,* 1996, 70(3), pp. 614-636

[3]   J. T. Cacioppo, G. G. Berntson, J. T. Larsen, K. M. Poehlmann, T. A. Ito: The Psychophysiology of Emotion. In J. M. Lewis, M. Haviland-Jones (Eds.), *Handbook of Emotions*, 2nd edition, New York: Guilford Press, 2000, pp. 173-191

[4]   B. De Gelder, P. Bertelson: Multisensory Integration, Levels of Processing and Ecological Validity. *Trends in Cognitive Science 7 (10)*, 2003, pp. 460-467

[5]   P. Ekman, W. V. Friesen, J. C. Hager: *The Facial Action Coding System* Second edition. Salt Lake City: Research Nexus eBook.London: Weidenfeld& Nicolson, 2002

[6]   P. Ekman: Strong Evidence for Universals in Facial Expressions: A Reply to Russell's Mistaken Critique. Psychological Bulletin, 1994, 115, pp. 268-287

[7]   P. Ekman: An Argument for Basic Emotions. Cognition and Emotion, 6, 1992, pp. 169-200

[8]   P. Ekman: The Argument and Evidence about Universals in Facial Expressions of Emotion. In H. Wagner, H., Manstead, A. (eds.) *Handbook of social psychophysiology*, Chichester: Wiley, 1989, pp. 143-164

[9]   P. Ekman: Universals and Cultural Differences in Facial Expressions of Emotion. In J. Cole (Ed.), Nebraska Symposium on Motivation, Lincoln: University of Nebraska Press. 1971, Vol. 19,1972, pp. 207-282

[10]   A. Esposito: The Perceptual and Cognitive Role of Visual and Auditory Channels in Conveying Emotional Information. Cognitive Computation Journal, 2009, 1 (2), pp. 268-278

[11]   A. Esposito, M. T. Riviello, G. Di Maio: The COST 2102 Italian Audio and Video Emotional Database. In B. Apolloni, et al. (Eds) Frontiers in Artificial Intelligence and Applications, Vol. 204, ISBN 978-1-60750-072-8 (print) ISBN 978-1-60750-515-0, 2009, pp. 51-61, http://www.booksonline.iospress.nl/Content/View.aspx?piid=14188

[12]   A. Esposito, M. T. Riviello, N. Bourbakis: Cultural Specific Effects on the Recognition of Basic Emotions: A Study on Italian Subjects. In: Holzinger, A., Miesenberger, K. (eds.) USAB 2009. LNCS, Vol. 5889, 2009, pp. 135-148

[13]   A. Esposito: The Amount of Information on Emotional States Conveyed by the Verbal and Nonverbal Channels: Some Perceptual Data. In Y. Stilianou et al. (Eds): *Progress in Nonlinear Speech Processing*, Lecture Notes in Computer Science, 4392,Springer-Verlag, 2007, pp. 245-264

[14]   A. P. Fiske, S., Kitayama, H. R., Markus, R. E., Nisbett: The Cultural Matrix of Social Psychology. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* Boston: McGraw-Hill ,4th ed., Vol. 2, 1998, pp. 915-981

[15]   C. E Izard, B. P. Ackerman: Motivational, Organizational, and Regulatory Functions of Discrete Emotions. In J. M. Lewis, M. Haviland-Jones (Eds.), *Handbook of Emotions*, 2nd edition, New York: Guilford Press, 2000, pp. 253-264

[16]   C. E. Izard: Innate and Universal Facial Expressions: Evidence from Developmental and Cross-Cultural Research. *Psychological Bulletin*, 1994, 115, pp. 288-299

[17]   C. E. Izard: *The Face of Emotion.* New York: Appleton-Century-Crofts. 1971

[18]   M. Kamachi, M. Lyons, J. Gyoba: Japanese Female Facial Expression Database, Psychology Department in Kyushu University, http://www.kasrl.org/jaffe.html

[19]   B. Mesquita, N. H., Frijda, K. R., Scherer: Culture and Emotion. In J. W. Berry, P. R. Dasen, & T. S. Saraswathi (Eds.), Handbook of cross-cultural

psychology: Vol. 2, Basic processes and human development, Boston: Allyn& Bacon. 1997, pp. 255-297

[20]   B. Mesquita, N. H., Frijda: Cultural Variations in Emotions: A Review. Psychological Bulletin, 1992, 112, 197-204

[21]   J. A. Russell: Is There Universal Recognition of Emotion from Facial Expression? A review of the cross-cultural studies. Psychological Bulletin, 1994, 115, 102-141

[22]   F. Samaria, A. Harter: The ORL Database of Faces. Cambridge University Press, Cambridge,
http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html

[23]   K. R. Scherer: Vocal Communication of Emotion: A Review of Research Paradigms. *Speech Communication* 2003, 40, pp. 227-256

[24]   K. R. Scherer, R., Banse, H. G.,Wallbott: Emotion Inferences from Vocal Expression Correlate across Languages and Cultures. Journal of Cross-Cultural Psychology, 2001, 32, 76-92

[25]   K. R. Scherer: The Role of Culture in Emotion-Antecedent Appraisal. Journal of Personality and Social Psychology, 1997, 73, 902-922

[26]   K. R. Scherer, H. G., Wallbott: Evidence for Universality and Cultural Variation of Differential Emotion Response Patterning. Journal of Personality and Social Psychology, 1994, 66, 310-328

[27]   K. R. Scherer, J. S. Oshinsky: Cue Utilization in Emotion Attribution from Auditory Stimuli. Motivation and Emotion, 1977, 1, pp. 331-346

[28]   D. Ververidis, C. Kotropoulos: Emotional Speech Recognition: Resources, Features and Methods. Elsevier Speech Communication 48(9), 2006, pp. 1162-1181