

Chapter 2

The Proposed Research Work

Abstract This chapter describes in details the materials, procedures, participants, testing, and data analyses exploited for the cross-modal and cross-cultural experiments on the perception of emotions narrated in this book. Two databases consisting of realistic, dynamic, and mutual related visual and vocal emotional information are described. The databases were constructed to allow an actual comparison between the effectiveness of visual and vocal cues in conveying emotional expressions. In order to investigate 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 language, the stimuli were extracted from American (as a global spread language) and Italian (as a country specific language) live recording movies. The experiments involved participants from five different Western cultures and languages to explore possible variations in inferring emotions across similar cultures, highlighting the role of the familiarity with the cultural context and language. The data collected were analyzed computing different repeated ANOVA measurements with the goal to reveal the effect of the communication channel, and both stimuli and participants “cultural context and language” on the recognition of the emotional expressions, as well as, to allow a cross-cultural comparison of the results.

Keywords Database of emotional stimuli • Cross-cultural comparisons • Emotion recognition • Cultural effect on emotions perception • Language familiarity • Cultural specificity • Data analysis • Cultural context • Auditory emotional information • Visual emotional information

This work focuses on the cross-modal and cross-cultural analysis of emotional data in the attempt to clarify mechanisms underlying the human perception of emotional expressions, as well as identifying close cross-cultural differences among such perceptual processes (Esposito 2007, 2009).

For this purpose, perceptual experiments exploiting two multimodal databases of realistic, dynamic and mutually related vocal and visual emotional stimuli were set up. The collected stimuli allowed exploring the amount of emotional information

conveyed dynamically by the visual and auditory channels, and therefore identifying preferential channels exploited by humans in decoding emotional states as well as enlightening which are the emotional cues universally shared.

In a cross-cultural perspective, this work investigates 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. As already mentioned in Sect. 1.1.1, psychologists have long debated whether emotions are universal versus whether they vary across cultures. Our approach is based on the assumption that culture and language-specific paralinguistic patterns may influence the decoding process of emotional speech. In addition, the familiarity to the language and the subject's expositions to cultural norms and social rules may affect the recognition of emotional states in particular when they are vocally expressed.

To this aims, the emotional stimuli used were extracted from two different cultural context and played in two different languages: American English (as a global spread language) (Riviello et al. 2011) and Italian (as a country specific language). In addition, the participants involved in the experiments belong to 5 different, even though close, Western cultures, e.g. Italian, American, French, Hungarian and Lithuanian. In this context, two groups of participants were native speakers of the language and belong to the same cultural context of the administered stimuli. Their performance can be considered as a reference for an optimal identification of the emotional states under examination (Riviello and Esposito 2012).

2.1 Materials: The Cross-Modal Emotional Databases

The collected stimuli are based on extracts from American and Italian live recording movies (Esposito et al. 2009; Esposito and Riviello 2011), 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. Differently from the other existing emotional databases proposed in literature, in this case, the actors/actresses had not been asked to produce an emotional expression for building an associated database, but rather they were acting according to a movie script, their performance was related and considered appropriate to a defined context as the movie director (supposed to be an expert) had judged. In addition, 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. The stimuli were also noisy, as emotions experienced in real environments.

Each database consists of audio and video stimuli representing six emotional states: happiness, sarcasm/irony, fear, anger, surprise, and sadness. Except for

sarcasm/irony, the remaining emotions are considered by many theories as primary ones and therefore universally shared (see Sect. 1.1).

For each database and for each of the emotional states under examination, 10 stimuli were identified, 5 expressed by an actor and 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 actor's bias in the ability to portray emotional states.

The stimuli were short (the average length was 3 s, $SD = \pm 1$ s) 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 database. This was an attempt to allow the participants to observe less obvious emotional cues generally employed in a very natural and ecological setting, rather than in extreme emotional interactions.

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, the rising and falling of F0 intonation contour, 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 labeling between the two groups was 100 %.

The collected stimuli extracted from movie scenes contain environmental noise and therefore are useful for testing realistic computer applications.

Both for the American and Italian data, the audio and mute video were extracted from each complete audio-video stimulus (video-clip) coming up with a total of 180 American and 180 Italian stimuli: 60 mute videos, 60 audio and 60 audio-video stimuli for each database.

2.2 Participants and Testing Procedure

The perceptual experiments involved 180 Italian, 180 American, 180 French, 180 Hungarian, and 180 Lithuanian participants. The participants' age was similar among countries, ranging from 18 to 35 years (26 ± 4.8).

Excluding the Americans, all participants have comparable knowledge of the English, since all of them used it as second language. The participants were volunteers principally recruited among university students.

For each group of 180, hence for each nationality, 90 participants were involved in the evaluation of the American database and 90 of the Italian database of stimuli. For each group of 90, 30 subjects evaluated the audio, 30 the mute video and 30

audio-video stimuli. Gender was equally balanced among the groups composed each by 15 males and 15 females.

The subjects were randomly assigned to the task and were required to carefully listen to and/or watch the stimuli via computer, wearing headphones, in a quiet room. They were instructed to pay attention to each presentation and decide which of the six emotional states were expressed.

Responses were recorded on a matrix paper form (60×8), where rows listed the stimuli numbers and columns the 6 selected emotional states (happiness, fear, anger, irony, surprise and sadness) plus the option for “others” indicating any other emotion not listed and the option for “no emotion”, which was suggested when according to the subject’s feeling the protagonist did not show emotions. The above-mentioned paper form was created in Italian and then translated in American, French, Hungarian and Lithuanian with the help of native speaker of the languages.

2.3 Data Analyses

The data obtained from participants of each nationality were first analyzed separately. For each database and for each set of stimuli (audio, mute video and audio-video) the frequencies of correct answers, intended here as the subjects’ agreement on the label assigned to each stimulus, related to each emotion under consideration, were computed.

To assess the role of the perceptual mode affecting the identification of the emotional stimuli, repeated ANOVA measurements were performed on the frequencies of correct answers obtained by participants from each nationality, separately tested on the American and the Italian stimuli. To set up the analyses the *Perceptual mode* (audio, video, audio-video) was considered as a between subjects variable, while *Emotions* (happiness, fear, anger, irony, surprise and sadness) and *Actors’ gender* (male, female) as within subjects variables. Significance was established for $\alpha = 0.05$.

Further analyses were performed separately on the data gathered from participants of each nationality to assess the role of the language and the cultural context characterizing the emotional expressions exploited as stimuli.

To this aim, for each perceptual mode (audio, video and audio-video) subjects’ performance was assessed and compared on the set of American and Italian stimuli. In the repeated ANOVA measurements, the *Cultural context and the Language of the Stimuli* (American and Italian) was considered as between subjects variable, while *Emotions* and *Actors’ gender* as within subjects variables. Significance was fixed for $\alpha = 0.05$.

To allow a cross-cultural comparison for establishing the effects of the cultural context and the language, further analyses were conducted. Six separated repeated ANOVA measurements for each perceptual mode were performed on the data obtained by each group of 30 American, Italian, French, Hungarian and Lithuanian subjects separately tested on the American and Italian audio, mute video and

audio-video stimuli. In this case the ANOVAs' set up considered the subjects' *Nationality* as a between subjects variable and the *Emotions* and *Actor's Gender* as within subjects variables. Even in this case significance was established for $\alpha = 0.05$. All the statistical analyses were run using SPSS Statistic 17.0 software (SPSS 2008).

References

- Esposito, A. (2007). The amount of information on emotional states conveyed by the verbal and nonverbal channels: Some perceptual data. In Y. Stylianou, M. Faundez-Zanuy, & A. Esposito (Eds.), *Progress in Nonlinear Speech Processing* (Vol. 4391, pp. 249–268), LNCS. Berlin Heidelberg: Springer. ISBN 978-3-540-71503-0.
- Esposito, A. (2009). The perceptual and cognitive role of visual and auditory channels in conveying emotional information. *Cognitive Computation Journal*, 1, 268–278.
- Esposito, A., & Riviello, M. T. (2011). The cross-modal and cross-cultural processing of affective information. In B. Apolloni, et al. (Eds.), *Frontiers in Artificial Intelligence and Applications* (Vol. 226, pp. 301–310). IOS press. ISBN: 978-1-60750-691-1 (print), ISBN 978-1-60750-692-8 (online).
- Esposito, A., Riviello, M. T., & Di Maio, G. (2009). The COST 2102 Italian audio and video emotional database. In B. Apolloni, et al. (Eds.), *Frontiers in Artificial Intelligence and Applications* (Vol. 204, pp. 51–61). ISBN 978-1-60750-072-8.
- Riviello, M. T., & Esposito, A. (2012). A cross-cultural study on the effectiveness of visual and vocal channels in transmitting dynamic emotional information. *Acta Polytechnica Hungarica, Journal of Applied Sciences*, 9(1), 157–170. ISSN 1785-8860.
- Riviello, M. T., Chetouani, M., Cohen, D., & Esposito, A. (2011). On the perception of emotional “voices”: A cross-cultural comparison among American, French and Italian subjects. In A. Esposito, et al. (Eds.), *Analysis of Verbal and Nonverbal Communication and Enactment: The Processing Issue* (Vol. 6800, 368–377), LNCS. Springer. ISBN 978-3-642-25774-2.
- SPSS Inc. Released. (2008). SPSS Statistics for Windows, Version 17.0. Chicago: SPSS Inc.

On the Perception of Dynamic Emotional Expressions: A
Cross-cultural Comparison

Riviello, M.T.; Esposito, A.

2016, VIII, 45 p. 17 illus. in color., Softcover

ISBN: 978-94-024-0885-0