On the Effectiveness of Visual and Vocal Channels in Transmitting Dynamic Emotional Information

Anna Esposito & Maria Teresa Riviello

Seconda Università di Napoli, Department of Psychology, Italy
and
International Institute for Advanced Scientific Studies (IIASS), Italy

3rd SPLab Workshop 2013, Brno

Goals

- Multimodal approach for exploring the ability in decoding emotional dynamic expressions portrayed through visual and vocal channel (singularly or in combination):
  - is one channel more effective than the other in decoding emotional information ?
  - is this effectiveness affected by the cultural context and in particular by the language?
  - Should we look for universal invariants or account for personal traits, cultural specificities and contextual instances?
  - Does multimodality increase our (and machines) ability to code and decode emotional feelings?
  - What role plays language specificity?

A test on static facial expressions

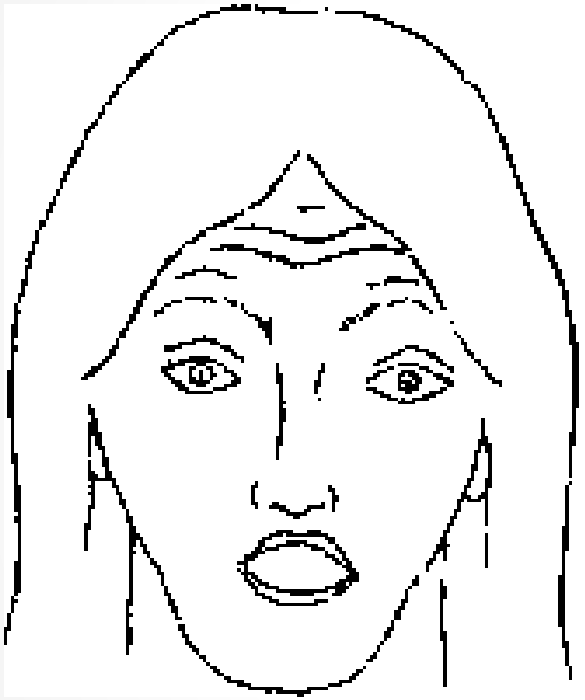


Figure A: The facial expression associated with surprise (Collier 1985)

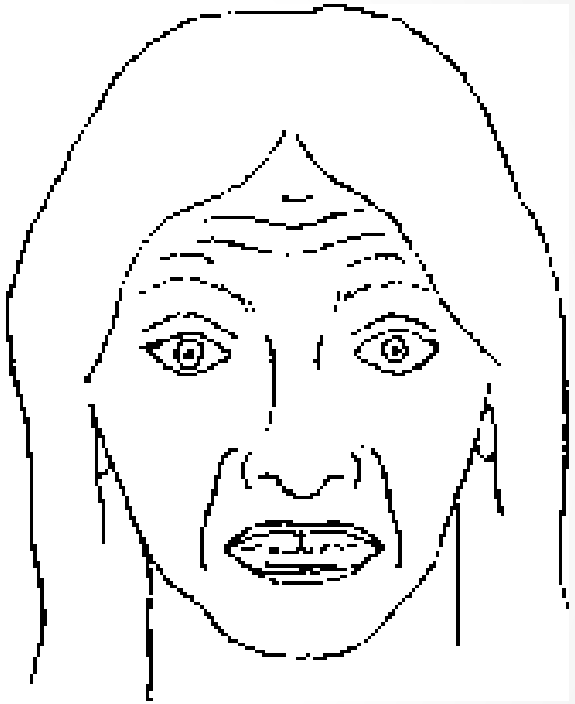


Figure B: The facial expression during fear (Collier 1985)

Materials

Two databases of emotional expressions of *happiness*, *sarcasm/irony*, *fear*, *anger*, *surprise,* and *sadness* extracted from American and Italian movies (Esposito, 2007, 2009):

- 10 stimuli for each emotion: 5 expressed by an actor and 5 by an actress, for a total of 60 American and 60 Italian video-clips
- the audio and the video alone were extracted from each video-clip →180 Italian and 180 American English emotional stimuli
- for each database: 60 stimuli only audio, 60 mute video, and 60 audio-video

Stimuli Specification

- In each video-clip, protagonist's face and the upper part of the body are clearly visible
- The semantic meaning of the utterances produced by actor/actresses is not clearly expressing the portrayed emotional state and its intensity is moderate
- The emotional labels assigned to the stimuli were first given by two experts and then by three naïve judges independently

Participants

1st Study: 90 Americans, 90 French and 90 Italians
2nd Study: 90 Americans and 90 Italians

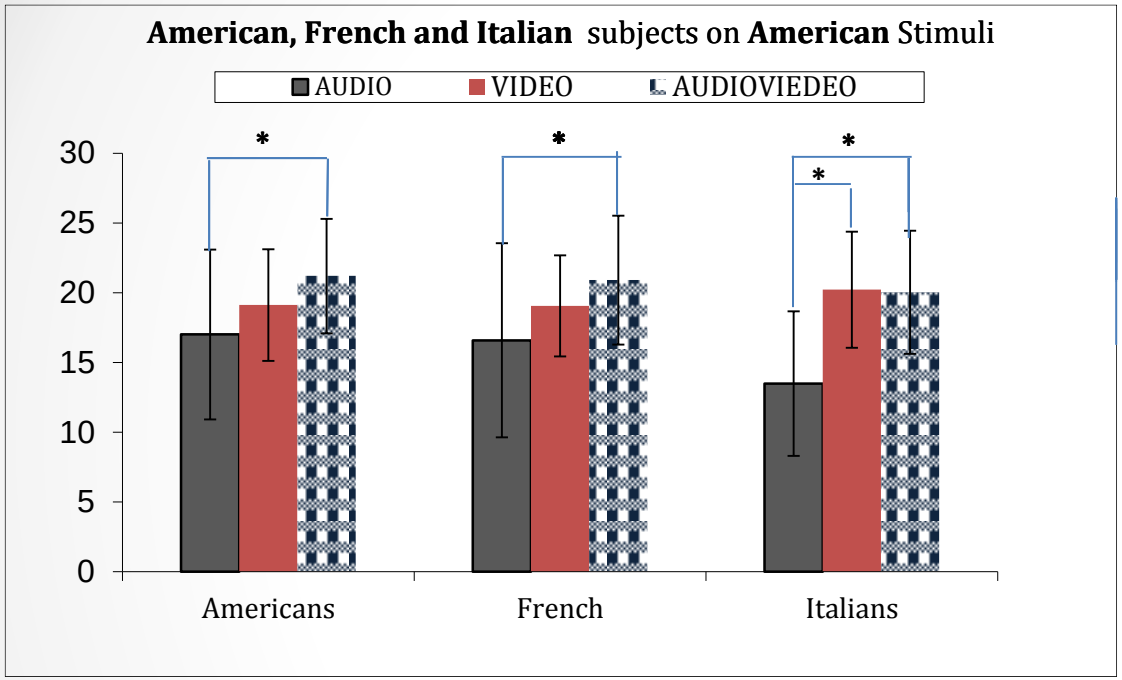
- age: from 18 to 35
- volunteer university students
- Italian and French subjects used English as a second language
- American and French no knowledge of Italian language

- for each group of 90, 30 subjects (balanced by gender) assessed the American\Italian audio, 30 the American\Italian mute video, and 30 American\Italian audio-video stimuli

- subjects were required to carefully listen to and/or watch the stimuli via headphones in a quiet room, and indicate for each presentation which of the 6 emotional states was expressed in it, plus an option for “no emotion” and “other emotion”

Stimuli Examples

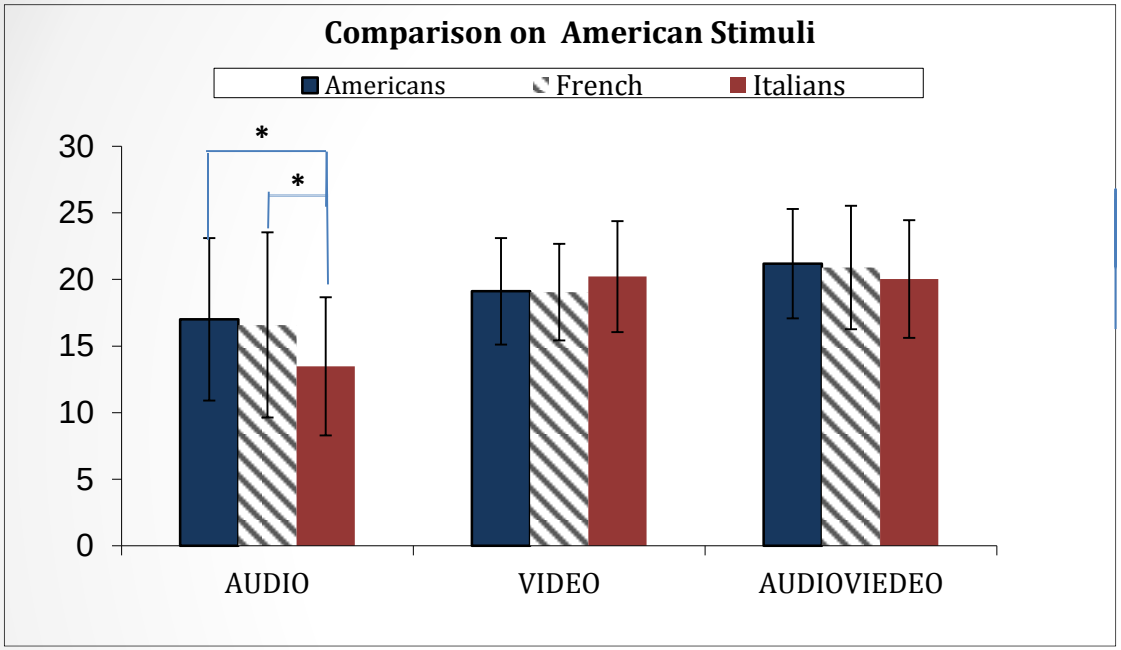
American stimuli: Results on channels



ANOVA: *Communication modes* (Audio, mute video and Audio-video) : between subjects variable; *Emotions*, and *Actors 'gender* (male and female): within subjects variables. Significance was established for α=.05

- **Americans:** **Audio\Audio-video:** F (1,8)= 9.031, p=.017*
- **French:** **Audio\Audio-video**: F (1,8)= 8.493, p=.019*
- **Italians :** **Audio\ Audio-video**: F (1,8)= 53.075, p=.0001*
  Audio\ Mute Video: F (1,8)= 33.722, p =.0001*

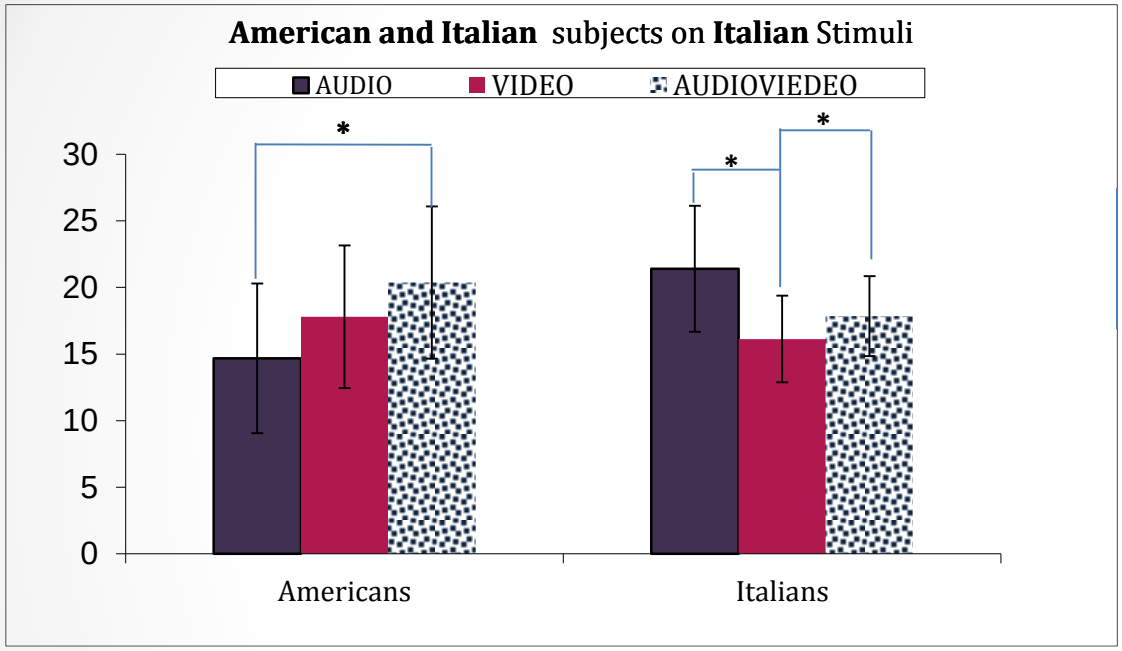
American Stimuli : Results on Subjects' Nationality



ANOVA: Subjects' *Nationality* (American, French and Italian): between subjects variable; *Emotions*, and *Actors 'gender* (male and female): within subjects variables. Significance was established for α=.05

- **American Audio stimuli: Nationality plays a significant role** ($F(2, 12) = 4.288$, ρ=.04*)
- Duncan post-hoc test revealed that the **Italian subjects differ significantly both from French and American subjects** for α=.05

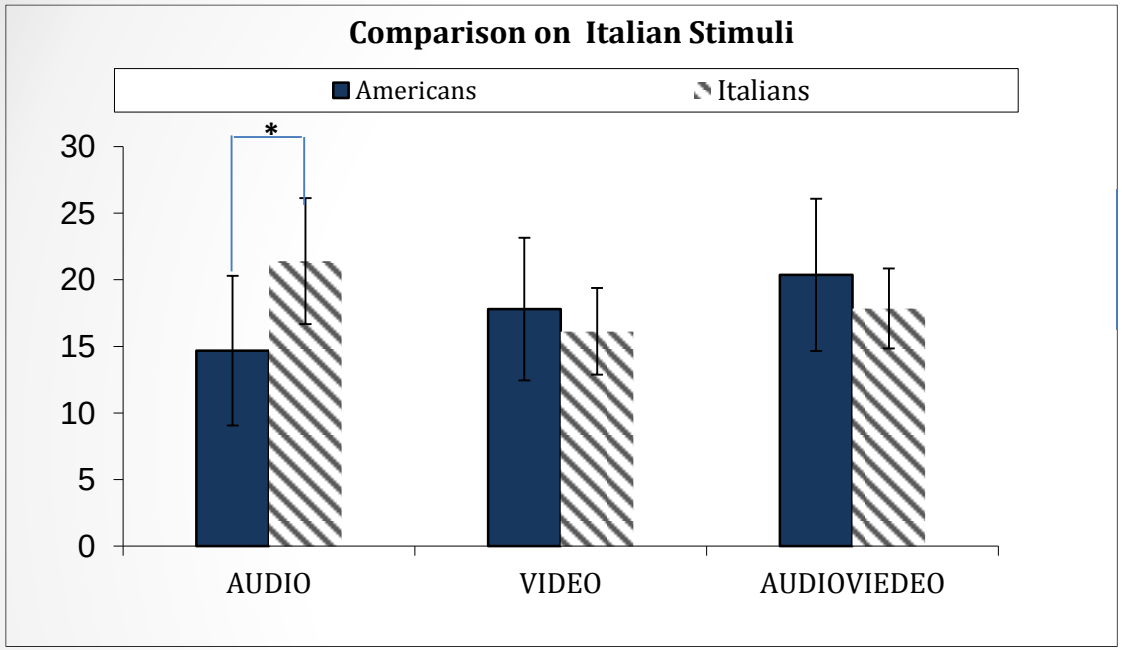
Italian stimuli:results on channels



ANOVA: *Communication modes* (Audio, mute Video and Audio-video) : between subjects variable; *Emotions*, and *Actors 'gender* (male and female): within subjects variables. Significance was established for α=.05

- **Americans: Audio\Audio-video:** F (1,8)= 21.674, p=.002*
- **Italians : Mute Video \ Audio-video:** F (1,8)= 77.58 p=.024*
  Audio\ Mute Video: F (1,8)= 8.436, p=.020*

Italian Stimuli: Results on subjects' country



ANOVA: Subjects' *Nationality* (American and Italian): between subjects variable; *Emotions*, and *Actors 'gender* (male and female): within subjects variables. Significance was established for α=.05

- **Italian Audio stimuli: Nationality plays a significant role:** (F (1, 8)= 20.987, p=.002*)

(Esposito & Riviello 2011. Riviello et al 2012, and unpublished data)

Stimoli Americani			
Soggetti	Audio/Video	Audio/Audiovideo	Video/ Audiovideo
Italiani	**F(1,8)= 33.72, p=.001 ***	**F (1,8)= 53.07, p=.001***	F (1,8)= .025, p=.87
Americani	F (1,8)= 1.7, p=.229	**F (1,8)= 9.03, p=.017***	F (1,8)= 2.323, p= .166
Francesi	F (1,8)= 3.37, p=.104	**F (1,8)= 8.49, p=.019***	F (1,8)= 1.426, p=.267
Ungheresi	**F (1,8)= 6.79, p=.031***	**F (1,8)= 29.27, p=.001***	F (1,8)= 2.749, p=.136
Lituani	**F (1,8)= 5.37, p=.049***	**F (1,8)= 12.97, p=.007***	F (1,8)= 0.421, p=.535
Stimoli Italiani			
Soggetti	Audio/Video	Audio/Audiovideo	Video/ Audiovideo
Italiani	**F (1,8)= 8.43, p=.020***	F (1,8)= .678, p=.434	**F (1,8)= 77.58, p=.024***
Americani	F (1,8)= 4.114, p=.077	**F (1,8)= 21.674, p=.002***	F (1,8)= 2.303, p= .16
Francesi	**F (1,8)= 8.004, p=.022***	**F (1,8)= 28.793, p=.001***	F (1,8)= 0.771, p=.406
Ungheresi	**F (1,8)= 10.15, p=.013***	**F (1,8)= 23.761, p=.001***	F (1,8)= 0.996, p=.347
Lituani	**F (1,8)= 9.835, p=.014***	**F (1,8)= 27.525, p=.001***	F (1,8)= 0.851, p=.383

Conclusions

- The bimodal (audio-video) presentation of the emotional information does not significantly improve subjects recognition accuracy

- The decoding process of emotional states is affected by the communication modality (channel) and language

  - for American stimuli: Americans and French identify emotional information from visual as well as from vocal cues, Italians rely more on visual information
  - for Italian stimuli: American rely more on visual information, Italian on vocal cues
  - ➢ speakers of different languages may exhibit a different sensitivity to vocal emotional information that could be attributed to the language supra-segmental cues specific to the language

- Culture specificity does NOT affect the decoding process of VISUAL emotional information: Visual channel allow a cross-cultural identification of emotions.

We are still comparing WESTERN cultures!


Seconda Università degli Studi di Napoli
SUN


Be
CogSys


IIASS

Thank You