Audio-visual interaction in emotion perception for communication

De Boer, M. J.; Başkent, D.; Cornelissen, F. W.

Published in:
Proceedings - ETRA 2018

DOI:
10.1145/3204493.3207422

Audio-visual interaction in emotion perception for communication

Doctoral Symposium, Extended Abstract

M.J. de Boer
Department of Otorhinolaryngology,
University Medical Center Groningen
Groningen, The Netherlands
minke.de.boer@rug.nl

D. Başkent
Department of Otorhinolaryngology,
University Medical Center Groningen
Groningen, The Netherlands
d.baskent@umcg.nl

F.W. Cornelissen
Experimental Ophthalmology,
University Medical Center Groningen
Groningen, The Netherlands
f.w.cornelissen@umcg.nl

ABSTRACT

Information from multiple modalities contributes to recognizing emotions. While it is known interactions occur between modalities, it is unclear what characterizes these. These interactions, and changes in these interactions due to sensory impairments, are the main subject of this PhD project.

This extended abstract for the Doctoral Symposium of ETRA 2018 describes the project; its background, what I hope to achieve, and some preliminary results.

CCS CONCEPTS

• **Applied computing** → *Psychology*;

KEYWORDS

Emotion perception, Communication, Audio-visual integration

ACM Reference Format:

M.J. de Boer, D. Başkent, and F.W. Cornelissen. 2018. Audio-visual interaction in emotion perception for communication: Doctoral Symposium, Extended Abstract. In *ETRA '18: 2018 Symposium on Eye Tracking Research and Applications, June 14–17, 2018, Warsaw, Poland.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3204493.3207422

1 BACKGROUND AND AIMS

In order to achieve good verbal communication, it is essential to be able to perceive both the content of the words and the communicative intent of the sender. The latter is conveyed through the expression of emotions. While a range of cues can carry emotion, for the communication of emotion cues in the visual and auditory modality are likely to be most important. While it is known that the two modalities interact, for communication in general [Ma et al. 2009; Schwartz et al. 2004] and emotion-in-communication perception (ECP) in particular [Collignon et al. 2008; Ekman and Friesen 1977], the underlying mechanisms remain unknown. Even more so, there is no knowledge on what effects visual impairments (VI) and hearing impairments (HI) have on emotional intent perception. Such sensory impairments are common among the elderly population [Roets-Merken et al. 2014]. Consequently, difficulties in communication are one of the major problems these people face.

Most research into emotion perception has relied on static stimuli, whereas natural communication is characterized by its dynamic nature. Hence, it is important to study sensory interactions using dynamic stimuli. Additional important unknowns reside in the relevance of the face compared to the entire body - which may become more important for people with reduced visual acuity - and the role of viewing strategies in the extraction of emotional cues from face and body, which may also change depending on the level of impairment of the senses. Based on our present knowledge on multisensory perception, interactive effects are not expected to be entirely linear and predictable. It can be expected that VI and HI will interact and amplify each other, and thus be more pronounced than expected on the basis of the domain-specific impairments, as emotion perception relies on both modalities. Yet, the interactions can be expected to change dramatically depending on the presence of sensory impairments.

This PhD project aims to clarify: 1) the interactive effects of visual and auditory components in dynamic ECP, 2) the effects of VI and HI on dynamic ECP, and 3) the role of viewing strategies in dynamic ECP. Our ultimate goal is to create rehabilitation methods to improve ECP in persons with VI and/or HI.

2 APPROACH AND METHODS

These objectives will be tested in a series of experiments. The main paradigm is an ECP test using dynamic audiovisual emotional stimuli [Bänziger and Scherer 2010] and a forced-choice task (clicking on the perceived emotion). ECP is tested in two unimodal (audio-only and video-only) conditions and one bimodal (audiovisual) condition. Audiovisual integration is assessed by comparing performance between the unimodal and bimodal conditions. Eye tracking will be used to gain a deeper insight into decision-making strategies in general, and evaluate strategy changes between conditions.

In addition to this main paradigm, simulations of VI and HI will be used to assess the effects of these impairments in a homogenous group, namely healthy observers. Several visual defects will be simulated: e.g. low-pass filtering the spatial frequency content to simulate decreased visual acuity, low contrast to simulate decreased contrast sensitivity, or adding a gaze-coupled mask to simulate e.g. a central scotoma as occurring in age-related macular degeneration. By coupling the mask location to eye gaze, this mask will be similar to what actual VI persons experience. On top of this, to simulate age-related HI, a realistic HI simulator [Irino et al. 2013] will be

ETRA '18, June 14–17, 2018, Warsaw, Poland

ACM ISBN 978-1-4503-5706-7/18/06.
https://doi.org/10.1145/3204493.3207422

implemented to simulate different types and grades of hearing loss. Additionally, background noise may be added or the audio signal reverberated to simulate speech degradations that happen in real life.

3 PRELIMINARY RESULTS

The first experiment, designed to validate our stimulus set and investigate audiovisual interactions in healthy observers, has recently been carried out.

The data show that the stimulus set works well, as accuracy scores are similar to [Bänziger and Scherer 2010] and performance is much higher than chance level, but not at ceiling (see also Figure 1). The results additionally indicate that auditory and visual information are integrated, but not treated equally; visual information appears to contribute more than auditory information, as indicated by a significantly lower performance in the audio-only condition.

Furthermore, preliminary analysis of gaze data suggests different viewing behavior for the video-only than for the audiovisual condition. Analysis of the fixation duration showed that participants tended to look longer towards the mouth region in the video-only condition and longer at the eyes in the audiovisual condition. This would indicate that participants try to use lip-reading as a substitution for the auditory information in the video-only condition, despite the linguistic content of the stimuli being uninformative. It therefore appears that observers do not always prioritize the eyes for emotion recognition, but use gaze functionally, as an information gathering method. Though this result contradicts the findings from [Vo et al. 2012], their conclusion supports our findings.

To confirm our findings and conclusions, more detailed analysis, taking into account the dynamic nature of the stimuli, are therefore needed.

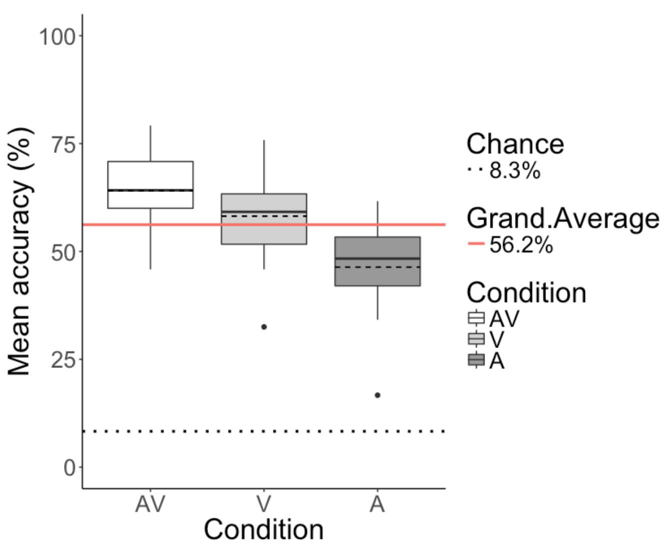

Figure 1: Accuracy scores for each condition, averaged over emotions. Dashed lines within each box indicate mean performance, while the solid line is the median.

4 PLANS FOR FUTURE WORK

In the near future, I will study the effects of simulated VI and HI on ECP in a group of healthy observers. This experiment will give an indication of what kinds of VI and HI lead to performance decrease and will guide the selection of a patient population. This patient population will then undergo the same ECP task as healthy observers.

In order to compare performance of patients, who will most likely be elderly persons, to performance of healthy observers, a group of healthy elderly persons will also be tested. This approach allows for studying the effects of both healthy aging, as well as the effects sensory impairments, on ECP.

Lastly, if time permits, I will include neuroimaging to the experimental paradigm, to study the neural mechanisms of audiovisual integration for ECP.

REFERENCES

Tanja Bänziger and Klaus R. Scherer. 2010. Introducing the geneva multimodal emotion portrayal (gemep) corpus. *Blueprint for affective computing: A sourcebook* (2010), 271–294.

Olivier Collignon, Simon Girard, Frederic Gosselin, Sylvain Roy, Dave Saint-Amour, Maryse Lassonde, and Franco Lepore. 2008. Audio-visual integration of emotion expression. *Brain research* 1242 (2008), 126–135.

Paul Ekman and Wallace V. Friesen. 1977. Facial action coding system. (1977).

Toshio Irino, Tomofumi Fukawatase, Makoto Sakaguchi, Ryuichi Nisimura, Hideki Kawahara, and Roy D. Patterson. 2013. *Accurate Estimation of Compression in Simultaneous Masking Enables the Simulation of Hearing Impairment for Normal-Hearing Listeners.* Springer, 73–80.

Wei Ji Ma, Xiang Zhou, Lars A. Ross, John J. Foxe, and Lucas C. Parra. 2009. Lip-reading aids word recognition most in moderate noise: a Bayesian explanation using high-dimensional feature space. *PLoS One* 4, 3 (2009), e4638.

Lieve M. Roets-Merken, Sytse U. Zuidema, Myrra JFJ Vernooij-Dassen, and Gertrudis IJM Kempen. 2014. Screening for hearing, visual and dual sensory impairment in older adults using behavioural cues: A validation study. *International journal of nursing studies* 51, 11 (2014), 1434–1440.

Jean-Luc Schwartz, Frédéric Berthommier, and Christophe Savariaux. 2004. Seeing to hear better: evidence for early audio-visual interactions in speech identification. *Cognition* 93, 2 (2004), B69–B78.

Melissa L-H Vo, Tim J. Smith, Parag K. Mital, and John M. Henderson. 2012. Do the eyes really have it? Dynamic allocation of attention when viewing moving faces. *Journal of vision* 12, 13 (2012), 3–3.