Expressive Communication: Static vs. Dynamic Gestures of Feeling of Thinking

Anne Vanpé, Véronique Aubergé

"Expressive communication : static vs dynamic gestures of Feeling of Thinking"

Anne Vanpé[1], Véronique Aubergé[1,2]

[1]Gipsa-lab, Speech and Cognition Department, CNRS, Grenoble Universités
[2]LIG lab, CNRS, Grenoble Universités
{anne.vanpe, veronique.auberge}@gipsa-lab.inpg.fr

Keywords: non-verbal prosody, gesture perception, dynamic/static, expressive communication, emotion.

The interaction between two humans (or between human and computer by projection) is cadenced by the turn-taking/speech acting. Nevertheless the communication flow of each interaction participant is never stopped. The out of turn-taking time is continuously full of expressions that concern the processes of listening/ understanding/ reacting ("feedback" in the backchannel expression, (4)). Such information is very varied (3,4) and is expressed voluntary or involuntary: mental states (concentration, "Feeling of Knowing" (4)), intentions, attitudes (politeness, satisfaction, agreement…), emotions (disappointment, irritation…) and moods (stress, relaxing). We called "Feeling of Thinking" such expressions of mental and affective states.

The perceptual study presented here aimed to ascertain the relevance of non-verbal audio-visual stimuli chosen from a large spontaneous expressive French database (Sound Teacher/Ewiz (1)): two (among 17) selected female subjects (introvert S and extravert T) were tricked in a wizard of Oz that made them react with strong contrasted affects and cognition states in a human-computer interaction. The out of turn-taking audio-visual signals for the two subjects and their auto-annotation in terms of affects and other feelings were edited in order to precisely segment and then extract stimuli. Selected stimuli are supposed to be minimal icons representative of the naïve auto-annotation labels given by the subjects. Even if many non speech sounds seem to be informative, only silent visual stimuli were chosen for this experiment. 10 auto-annotation labels were retained for this experiment for subject T: "hesitant", "stressed", “ill at ease/worried", "anxious /oppressed", "at ease/more relaxed", "quiet/fine", "a bit lost/perplexed", "disappointed", "astonished", “concentrating”), and 9 for subject S ("not concentrating and feeling like laughing", "deriding my results", "listening with attention", ""holding over me" by the software", "stressed", "feeling like laughing and answering by chance", "concentrating and answering by chance", "concentrating" and "disappointed"). For each selected minimal gestural icon (long from 0.5 to 4 seconds) was extracted a static picture, supposed to be typical.

This experiment aimed to compare how efficient are the dynamic vs. static icons to convey the information referenced by the labels. Since main studies, in such a topic, are related to emotions, and since it was shown that the face is very informative and that the upper part and the lower part of the face do not carry the same information, according to the multiple works around the Ekman theory, our dynamic and static stimuli were presented in three balanced conditions: upper part of the face ("upper"), lower part of the face ("lower") and whole face ("whole").

Two identical perceptual tests were implemented: the first one with the dynamic stimuli, the second one with static stimuli (pictures) extracted from each dynamic stimulus. Each stimulus was presented once to in each condition for every session of perceptual test. Sixteen judges were presented the two tests, that consisted of closed choices among the self-annotation labels. Although presentation time was not limited for static stimuli, dynamic stimuli could be replayed during 8 seconds.

The main result (cf. figures 1 and 2) is that there is no additivity between the upper and lower part of the face. No part of the face really contains sufficient information, whatever is the label, and more specifically for the mental states expressions (even if, for example, "concentrating" and "feeling like laughing and answering by chance" have more information in the upper part of the face and "stressed" in the lower part). Moreover this sharing between lower and upper part of the face can change, depending on if the stimulus is dynamic or presented on an extracted static picture.

More globally the profit from static to dynamic seems to deeply depend on the nature of the information: for some stimuli, an under chance identification becomes a clear chance score; whereas for some others, the dynamic seems to be a disturbance (recall that the dynamic presentation is the ecological one, and lets the judge free to use static processing when he observes the natural visual signal).

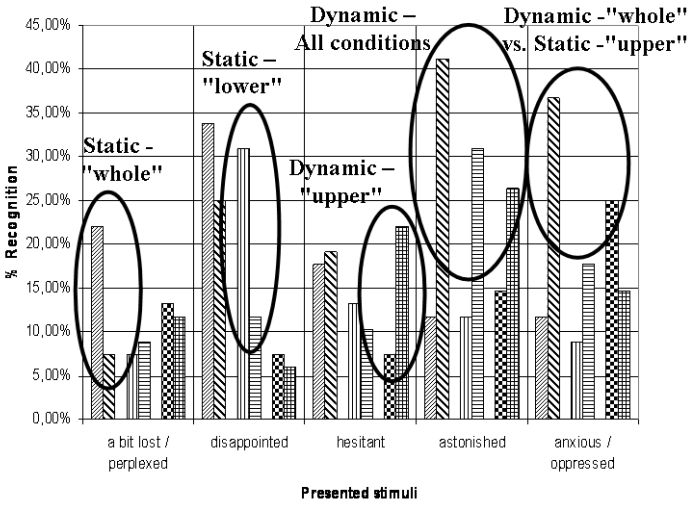
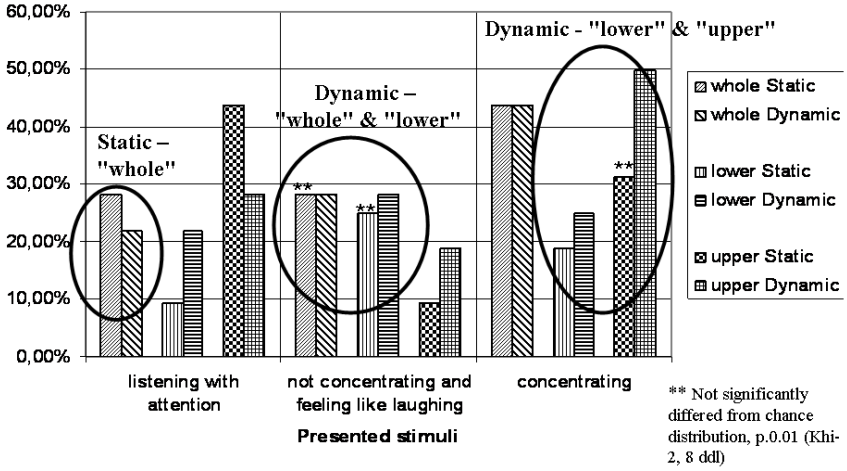


Figure 1: identification from static vs. dynamic stimuli – Subject T



** Not significantly differed from chance distribution, p.0.01 (Khi-2, 8 ddl)

Figure 2: identification from static vs. dynamic stimuli – Subject S

(1) Aubergé, V., Audibert, N., Rilliard, A.: De E-Wiz à C-Clone. Recueil, modélisation et synthèse d'expressions authentiques. In: Revue d'Intelligence Artificielle - "Interactions émotionnelles", 20(4-5), pp.499—528 (2006)

(2) Ekman, P.: Strong evidence for universals in facial expressions: A reply to Russell's mistaken critique. *Psychological Bulletin*, 115, pp. 268—287 (1994)

(3) Loyau, F., Aubergé, V.: Expressions outside the talk turn: ethograms of the Feeling of Thinking. *5th LREC*, pp.47—50 (2006)

(4) Schröder, M., Heylen, D., Poggi, I.: Perception of non-verbal emotional listener feedback. *Speech Prosody*, CD-Rom proceedings, SPS1-4-72 (2006)

(5) Swerts, M., Krahmer, E.: Audiovisual prosody and feeling of knowing. In: Journal of Memory and Language. 53:1, pp.81—94 (2005)