# Perception of social affects from non lexical sounds

*Gilles de Biasi, Véronique Aubergé* [1,2]*, Lionel Granjon*[1]*, Anne Vanpé* [1]

[1] GIPSA-Lab, CNRS, Grenoble, France
[2] LIG, CNRS, Grenoble, France

## 1. Problematics and literacy overview

In the face to face interaction, the communication is continuous, and expressed in a complex multimodality. The subject communicates when taking the talk, when listening the talk, when solving a task (e.g. in HMI). The signals exchanged by the subjects are deeply social, integrating information about the personality, the societal role, the culture and the language.

This paper focused on the perception of audio-visual sounds appearing outside or during the turn talk, that carry directly (by their forms) or indirectly (by their organization) numerous affective and pragmatic functions. Such non lexical sounds can be phonetically labeled (Vanpé, 2011) on a continuum from non phonological/not prosodically controlled sounds (mouth noises) to phonological/prosodically controlled sounds (interjections). The mouth noises, bursts, grunts, fillers, interjections are more and more studied both for their affective functions (Scherer, 1994; Schröder, 2003; Campbell, 2004) and as pragmatic cues for organizing turn talks (Ameka, 1992; Wichmann, 2002: Ward, 2006; Poggi, 2008).

In this paper, some various phonetic kinds of non lexical sounds have been selected from a large spontaneous corpus (Sound Teacher/E-Wiz, (Aubergé et al, 2003)). The French subjects involved in E-Wiz, were convinced to enter a language learning task. They were emotionally induced in a wizard of Oz protocol, by fake results given regularly: phase 1-good results; phase 2-very good results, phase 3-bad results, phase 4-very bad results, alarm about a possible deconstruction of subject's phonetic competences. Subjects are either reading, thinking or producing speech. Speech can consist of spoken answers (isolated monosyllabic words) or free comments. The subjects must reply with command words (*i.e.* keyboard and mouse free), and are given plenty of room for free comments. The machine interaction, which merely consisted of screen instructions and information, and audio stimuli, let no room for dialog. Self-labeling, attributed by the subjects themselves with naive vernacular language (Aubergé et al., 2003) revealed complex cognitive and affective states. These naive labeling describe affects, attitudes, intentions, mental states, cognitive processes, moods and emotions, that is a large panel of functions that was called Feeling of Thinking – FoT (Loyau and Aubergé, 2006).

The phonetic labeling of the non lexical sounds has been performed without any knowledge about self-labeling and induction context.

## 2. Experimental protocol

We selected, for 6 subjects of E-Wiz, some FoT labels, varied between emotions to mental states, in varying the complexity of the labels (simple to blended, or superimposed), the number of labels per subjects varies from 4 to 7 depending of this complexity. For each label, have been selected some sounds varying between prosodically controlled mouth noises to interjections. For each of the 6 subjects have been selected 3 phonetically variant stimuli of the same self-labeling label.

The principle of the experiment is a forced closed choice with the self-labeling labels, and a scale of confidence degree.

15 French native judges held the perceptive experiment balanced in Audio only/ Visual only/ Audio-Visual or Visual only/ Audio only/ Audio-Visual. For each modality, are successively presented, in balanced order, the 6 subjects, the stimuli for each subject being presented in balanced order too.

The stimuli are of course extremely short, but the judges can see, listen or both only one.

## 3. Results

The first main results is that the degree of confidence varies a lot between judges, more than between stimuli and modality, and is in average quite low. Detailed analysis will be presented further.

A confusion matrix is established for each subject in each modality (i.e. 18 matrixes). A Khi2 test is first held to select the significant answers, then is applied (with R) the prop.test function, the graphs for each subject are given hereunder for p<0.05, Audio only, Visual only and AudioVisual being represented on the same graph.

Globally for all the subjects, it has to be noted that many stimuli that are not identified in their own label, are not distributed by hazard, but highly reported on one other label. That could be related to the method to label the FoT: the self-labeling by the subjects concerns a whole sequence inside Sound Teacher task (it can be a reply, reading instructions etc).

For comparable labels, the subjects are differently identified by For comparable labels, the subjects are differently identified by privileging one modality, even if globally Visual seems more relevant than Audio. For comparable labels, the subjects are differently identified by  In Audio only, the labels equivalent to

hesitation/doubt systematically attract the fillers "euh" (typical pragmatic marker of hesitation in French) that were produced  with other labels, whatever the prosodic contours of the "euh" (quite lexical access), whereas in Visual only the label is often well identified (see graphs for subject 1, 2, 6). But what it more interesting is that for these hesitation/doubt labels some non phonetic mouth noises (articulator slackening, respiration releasing) attracted with the some high score than "euh", and could consequently considered as filler traces. These kind of mouth noises and the "euh" that were produced with the label doubt were well identified in Audio (subject 2).

The subject 6 has the best scores, either because his stimuli are relevant or because his self-labeling is precise enough. The subject 5 has good scores in Visual and AudioVisual. The label anxious (subject 1) is particularly well identified in Visual and AudioVisual, whereas Audio is confused with surprise. It must be noted that in the corpus, the subject labeled to be surprises just before to label anxious, perhaps it is a blended affect that were not noticed by the subject. For subject 2, the label unhappy-disappointed is well identified in Audio, but confused in Visual with stress. The label confident is confused with  doubt, that is paradoxical. The label afraid is confused in all modality with doubt: the stimuli are mouth noises similar to those produced in hesitation, perhaps are the filler traces.

## 4. References

Aubergé V., Audibert N., Rilliard A., (2003) "Why and how to control the authentic emotional speech corpora?", Proceedings of Eurospeech, Genève, 321-325.

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

Poggi, I., (2008) "The language of interjections", Multimodal Signals: Cognitive and Algorithmic Issues, COST 2102 School, Vietri, Italy, 170-186,

Scherer, K.R., "Affect bursts" (1994). In S.H.M. van Goozen, N. E. van de Poll & J.A. Sergeant (Eds.), Emotions, Hillsdale (NJ, USA), Lawrence Erlbaum, 161-193,.

Schröder M., Heylen D., Poggi I. (2006) : "Perception of non-verbal emotional listener feedback", *Proc of Speech Prosody 2006*

Signorello, R., Aubergé, V., Vanpé, A., Grandjon, L., Audibert, (2010) N., "A la recherche d'indices de culture et/ou de langue dans les micro-événements audio-visuels de l'interaction face à face", Proceedings of WACA 2010, Lille, France, 69-76.

Swerts, M., Krahmer, E., (2005) "Audiovisual prosody and feeling of knowing", Journal of Memory and Language, 53(1), 81-94

Vanpé A. (2011) Expressions et micro-expressions spontanées de la face et de la voix en Interaction Homme-Machine : esquisse d'un modèle du *Feeling of Thinking,* PhD thesis, Grenoble University.

Ward, N.  (2006) "Non-lexical conversational sounds in American English", Pragmatics & Cognition, 14(1), 129-182 (54).

## 5. Graphs

The labels are naively auto-annotated by the French subjects, often in familiar language, quite written oral. Each table is the result for one subject. A translation of the labels is given for each graph. Each graph represents the significant recognition: Audio in black arrows, Visual in blue arrow, AV in red arrow
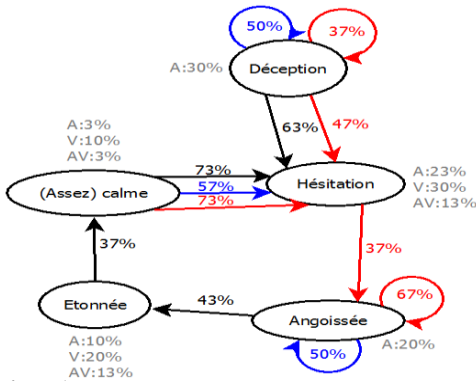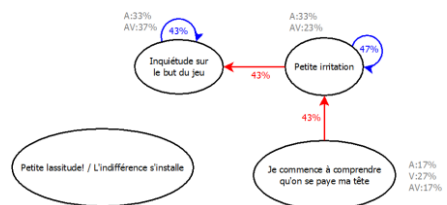


**Subject 1:**
- *(Assez) calme* = quiet calm   - *Hésitation* = hesitation
- *Déception* = disappointment  - *Etonnée* = surprise
- *Angoissée* = anxious
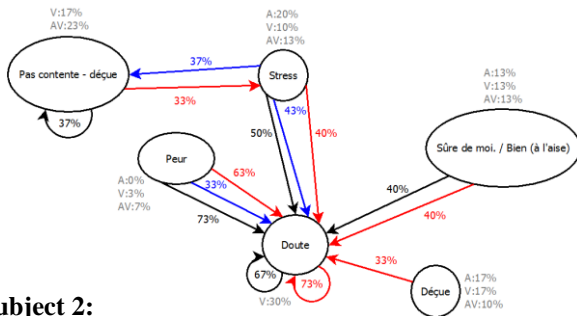


**Subject 2:**
- *Stress* = stress        - *Peur* = afraid
- *Pas contente, déçue* = unhappy, disappointed
- *Déçue* = disappointed        - *Doute* = doubt
- *Sûre de moi / Bien (à l'aise)* = Confident / well (comfortable)



**Subject 3:**
- *Inquiétude sur le but du jeu* = worry about game's goal
- *Petite irritation* = little irritation
- *Petite lassitude / L'indifférence s'installe* = little tireness / Indifference is coming
- *Je commence à comprendre qu'on se paye ma tête* = I start to understand that it is a joke
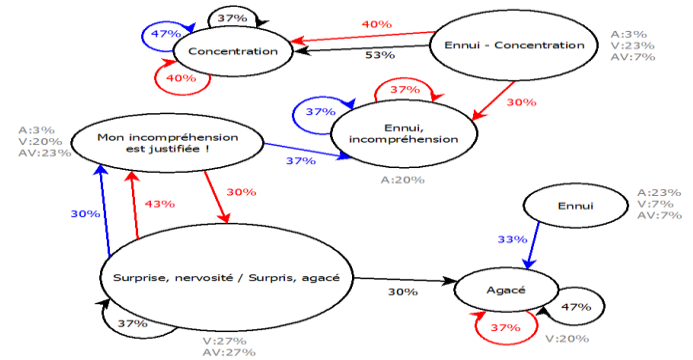
**Subject 6:**
- *Soulagé* = relieved   - *Léger doute* = little doubt
- *Essaie de rester concentré* = try to keep concentrated
- *Déception mais m'en amuse* = disappointed but it gives me fun
- *Léger doute* = little doubt
- *Ne suis plus sûr de moi, léger agacement* = Not anymore confident, little annoyance
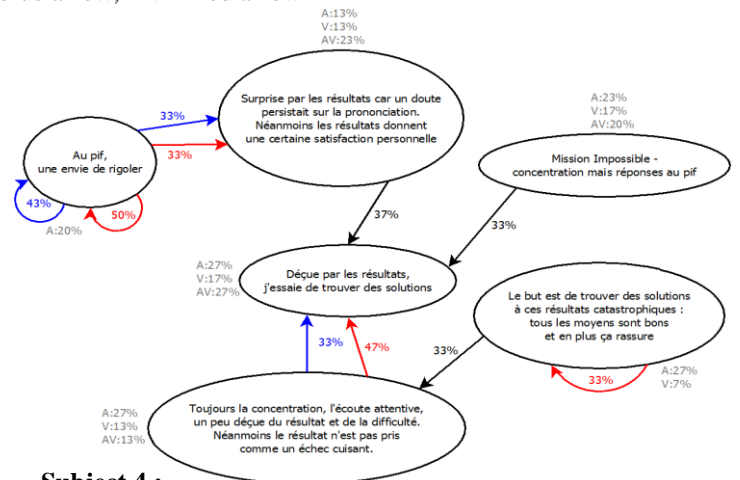


**Subject 4 :**
- *Au pif, envie de rigoler* = by hazard, tempt to laugh
- *Surprise par les résultats car un doute persistait sur la prononciation. Néanmoins les résultats donnent une certaine satisfaction personnelle* = surprise by results because a doubt about pronounciation /that is reply/. However results give me quite satisfaction
- *Mission impossible – concentration mais réponses au pif* = Impossible to reply – concentration but reply by hazard
- *Déçue par les résultats, j'essaie de trouver des solutions* = disappointed by results, I try to get solutions
- *Toujours la concentration, l'écoute attentive, un peu déçue du résultat et de la difficulté. Néanmoins le résultat n'est pas pris comme un échec cuisant* = still concentration, attentive listening, quite disappointed by result and difficulty. However, the result is not un strong failure for me
- *Le but est de trouver des solutions à ces résultats catastrophiques : tous les moyens sont bons et en plus ca rassure* = The aim is to fond solutions to these terrible results, all ways are required, and it reassures.



**Subject 5:**
- *Surprise, nervorsité / surpris, agacé* = surprise, nervosity / surprised, annoyed
- *Concentration* = concentration  - *Ennui* = boredom
- *Mon incomprehension est justifiée* = my misunderstanding is justified
- *Ennui, incompréhension* = boredom, misunderstanding
- *Ennui, concentration* = boredom, conc