Perception de la communication expressive : Icônes Gestuelles statiques vs. dynamiques du *Feeling of Thinking*

VANPE Anne

GIPSA Lab - Institut de la Communication Parlée - UMR 5216 CNRS/INPG/UJF/Stendhal
Université Stendhal, Domaine Universitaire BP25, 38040 Grenoble Cedex 9, FRANCE
Tél. : +33 (0)4 76 82 41 9700 - Fax : +33 (0)4 76 82 43 35
Courriel : anne.vanpe@gipsa-lab.inpg.fr

ABSTRACT

Most studies concerning expressive communication concentrate on (visual/vocal/auditory) expressions of the speaker while he is talking. But information about what a speaker is doing while he is not talking is also important. We first tried to build an empirical methodology of ethograms for information about the (non)talker's mental or affective states, that we called 'Feeling of Thinking'. Then we confronted some of the identified Gestural Icons with the perceptual validation of their relevance, in an association task with subject's self-annotation labels. We tested : (1) the static form of the Icons and their dynamic one; (2) three presentation conditions: whole face, upper part of the face only and lower part of the face only. The Icons were globally well identified, and can consequently be considered as relevant. Moreover, our results showed the importance of the dynamism for the 'Feeling of Thinking' perception and called additivity of the upper and lower parts of the face in terms of affective information into question.

Keywords: ethogram, Feeling of Thinking, 'out of turn-taking', gestural prosody, static vs dynamic perception, expressive speech, perceptual relevance.

.

1. INTRODUCTION

Les modèles sur les interactions verbales et leurs applications pour les Agents Conversationnels Animés (ACA), accordent une importance grandissante aux paramètres contextuels, au locuteur et à sa personnalité, ses émotions, ainsi qu'à ses interlocuteurs. En effet, le rôle communicatif des expressions du « back-channel » est de plus en plus mis en avant, et en particulier les indices de « feed-back » ([5], [7]) fournis par l'interlocuteur, informant sur ses états mentaux et affectifs (concentration, recherche en cours d'information connue par le sujet - « feeling of knowing », doute, accord, désaccord, inquiétude, satisfaction etc).

2. MÉTHODOLOGIE EMPLOYÉE POUR L'ÉTUDE DU *FEELING OF THINKING*

Nous travaillons avec une méthodologie encore empirique empruntée à l'éthologie, consistant à réaliser des éthogrammes, c'est-à-dire des grilles décrivant des comportements, à partir d'un corpus vidéo de sujets humains, piégés émotionnellement.

Notre matériel d'observation est en effet le corpus SoundTeacher [1], qui consiste en des vidéos d'interaction homme/machine émotionnellement induites, grâce à une tâche prétexte d'apprentissage des sons des langues du monde et l'utilisation du paradigme du magicien d'Oz. Il nous permet de travailler sur des expressions émotionnelles authentiques mais contrôlées.

L'analyse multi-modale (voix, parole, langage, expressions faciales, gestualité, signaux physiologiques) de ce corpus d'expressions d'états mentaux et émotionnels authentiques mais contrôlés, nous a permis d'identifier chez les sujets des expressions non seulement de leurs états affectifs (émotions, attitudes, intentions) mais également très largement de leurs états mentaux. Nous avons nommé *Feeling of Thinking* ces informations données par le sujet sur le sentiment qu'il a du déroulement de ses traitements cognitifs, par extension au phénomène du *Feeling of Knowing [8]*, expressions révélant spécifiquement les processus mnésiques du sujet.

Selon notre méthodologie, nous avons dans un premier temps cherché à identifier des Icônes Gestuelles (IG) primitives (c'est-à-dire irréductibles, nécessaires et suffisantes) qui véhiculent de l'information affective et mentale, en nous basant sur la variation des formes et en veillant à ne pas choisir les IG sur une interprétation a priori. Dans un deuxième temps, nous avons procédé à l'étiquetage de notre corpus en fonction de ces IG.

3. PROCEDURE DE VALIDATION PERCEPTIVE DE NOS ICONES GESTUELLES

Nous avons ensuite confronté certaines des IG que nous avons identifiées à la validation perceptive de leur pertinence dans une tâche d'association entre IG et labels issus de l'auto-annotation du corpus par les sujets eux-mêmes.

Nous avons d'abord retenu deux sujets dont les profils psychologiques sont éloignés (l'une est très stressée par le scénario, l'autre réagit aux inductions en se détachant de l'expérience par le rire). Nous avons ensuite sélectionné un sous-ensemble caractéristique d'IG statiques pertinentes méthodologiquement et dont les labels d'auto-annotation associés (items employés par le sujet lui-même) sont représentatifs de l'évolution des états mentaux et affectifs du sujet au cours de la tâche. Nous avons

également extrait les IG dynamiques (vidéos) correspondantes aux IG sélectionnées, en isolant le mouvement complet, figé sur une seule image sur l'IG statique. Deux tests identiques ont donc été montés, le premier avec les stimuli statiques, le second avec les stimuli dynamiques correspondants. Cela nous permet de comparer la perception des IG selon leur nature, statique ou dynamique.

D'autre part, les labels d'auto-annotation utilisés étant évidemment différents d'un sujet à l'autre, chaque sujet a donné lieu à une partie bien distincte lors des tests.

Nous avons ainsi retenu 10 labels pour le sujet T (« hésitante », « stressée », « mal à l'aise/inquiète », « angoissée/oppressée », « rassurée/plus détendue », « calme/va bien », « un peu perdue/perplexe », « déçue », « étonnée », « concentrée »), et 9 pour le sujet S (« pas concentrée et envie de rigoler », « "rit jaune" de ses résultats », « écoute attentivement », « "emprise" du logiciel, dans le sens où elle suit les consignes du mieux qu'elle peut », « stressée », « envie de rigoler et répond au hasard », « concentrée et répond au hasard », « concentrée » et « déçue »).

Trois conditions de présentation en terme d'informations fournies ont également été testées : visage entier (« entier »), seulement le haut du visage (condition « haut ») et seulement le bas (condition « bas ») (cf. un exemple Figure 1), car nous replacerons par la suite nos travaux chez le sujet parlant, et toutes les études en Eye Tracker montrent la prédominance alternée de ces deux zones au niveau de la focalisation visuelle en interaction verbale. De plus, nous avons pu mesurer la pertinence de la proposition d'Ekman qui, sur des formes statiques, donne une spécificité à ces deux grandes zones [3].

Nous cherchons ainsi à connaître la nature des formes relevées : le support est-il toujours la forme statique, comme Ekman l'a longtemps suggéré [3] ? Ou existe-il des informations qui ne sont perceptibles que dynamiquement, ou par la rythmicité du geste ?

Chaque stimulus a été présenté une fois par condition, d'où un nombre de 120 stimuli pour le sujet T (10 labels * 4 sti. par label * 3 conditions), et 48 pour S (9 labels * 2 sti. par label * 3 conditions). Nous avons veillé à ce que leur ordre de présentation soit aléatoire mais avec la condition « entier » en fin de test, de manière à éviter un biais en inter-sujets du fait de l'habituation.

Les tests consistèrent en un choix unique et fermé (cases à cocher) parmi les différents labels d'auto-annotations (cf. l'interface Figure 1). Alors que le temps d'observation des stimuli statiques ne fut pas limité, les vidéos des stimuli dynamiques pouvaient être rejouées pendant une durée de huit secondes. Ils ont été passés par 16 juges.

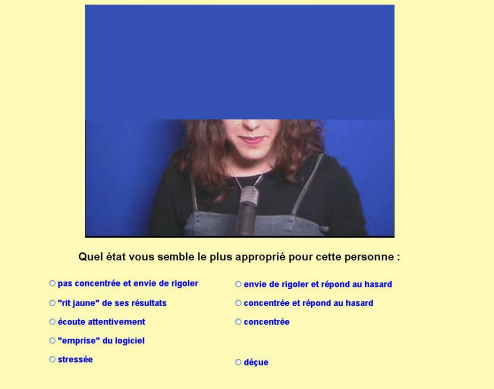

Figure 1: Exemple d'interface, avec un stimulus du sujet S en condition « bas »

4. RESULTATS

4.1. IG dynamiques

Sujet T

En condition « entier », les labels « Concentrée », « Etonnée », « Déçue », « Angoissée/Oppressée » et « Rassurée/plus détendue » ont été reconnus au dessus du seuil de deux fois le niveau du hasard. Des reports et confusions entre labels nous ont permis de dégager quatre méta-classes : « rassurée/plus détendue »/ « calme/va bien » (65,4 %) ; « hésitante »/ « stressée »/ « mal à l'aise/inquiète »/ « un peu perdue/perplexe »/ « déçue »/ « étonnée » (51,7%) ; et deux méta-classes de labels isolés : « angoissée/oppressée » (36,8%) et « concentrée » (30,9%) (voir graphe Figure 2).

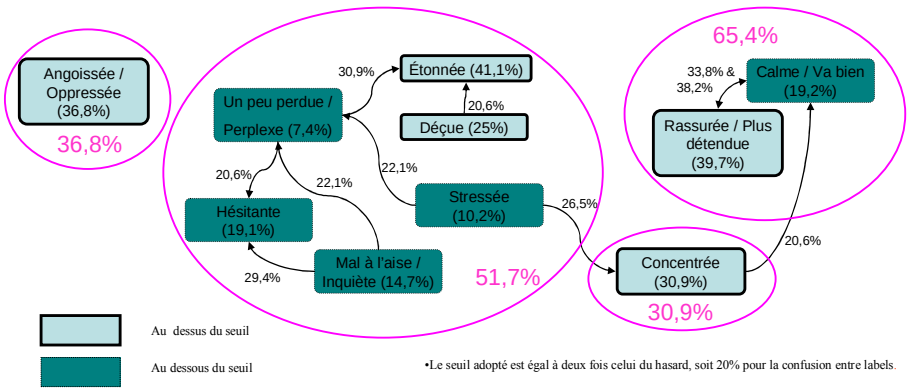

Figure 2 : Graphe de résultats pour le sujet T en condition entier

La distribution des réponses pour les labels « Déçue » en condition « haut » et « angoissée/oppressée » en condition « bas » n'est pas significativement différente de celle du hasard (Khi-2, $p<0.05$, 9 ddl). D'autre part, « Calme/ Va bien », à la fois en condition « haut » et « bas », et « Hésitante » en condition « haut », sont curieusement mieux reconnus qu'en condition « entier ». « Concentrée » est quant à lui le mieux reconnu (de 32,4% « bas », à 41,2% « haut ») et tend à attirer les autres réponses.

Sujet S.

En ce qui concerne le sujet S, en condition entier, les labels « pas concentrée et envie de rigoler », « envie de rigoler et répond au hasard » et « concentrée » ont été reconnus au dessus du seuil de deux fois le niveau du hasard. « Ecoute attentivement » est quant à lui mieux reconnu en condition « haut » qu'en condition « entier ». La distribution des réponses pour les labels « Concentrée et répond au hasard » en condition « haut », et « Ecoute attentivement » en condition « bas » n'est pas significativement différente de celle du hasard (Khi-2, $p<0.05$, 7 ddl). Comme pour le sujet T, le label « Concentrée », a tendance à attirer vers lui les autres réponses, et ce dans toutes les conditions. Nous avons également dégagé une méta-classe : « envie de rigoler et répond au hasard »/ « pas concentrée et envie de rigoler »/ « "rit jaune" de ses résultats », isolée des autres labels (en particulier en condition « entier ») et dans laquelle des confusions entre label existent (taux d'identification alors de 51% en condition « haut » à 76% en condition « entier » - voir graphe Figure 3).

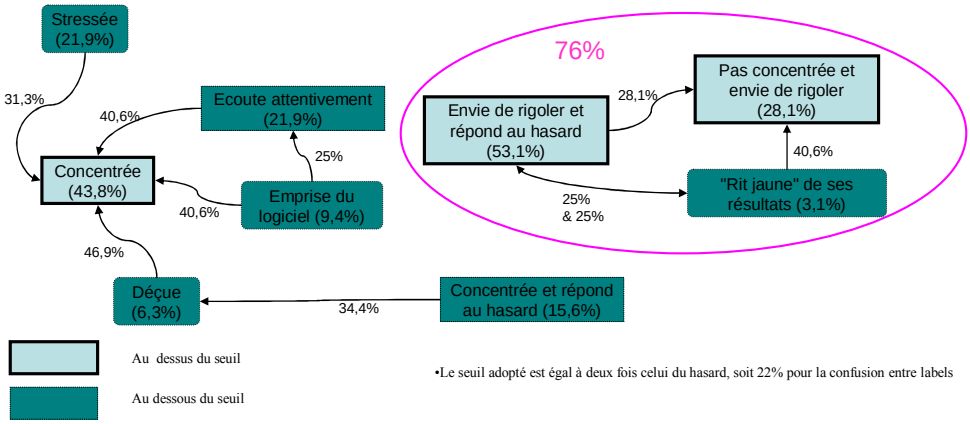

Figure 3 : Graphe de résultats pour le sujet S en condition entier

4.2. Comparaison statique/dynamique

Les résultats obtenus pour les IG dynamiques ont été différents de ceux du test sur les IG statiques [9].

Sujet T

Les quatre méta-classes identifiées lors du test statique (« hésitante »/ « stressée »/ « mal à l'aise/inquiète »/ « angoissée/oppressée » ; « un peu perdue/perplexe »/ « déçue »/ « étonnée » ; « rassurée/plus détendue »/ « calme/va bien » ; et « concentrée », n'ont pas été retrouvées de manière identique pour les mêmes stimuli dynamiques. En effet, si les deux dernières ont été retrouvées telles qu'elles, « angoissée/oppressée » s'est isolé en dynamique et les autres labels n'ont formé plus qu'un seul groupe (cf. Figure 2).

Les labels « étonnée » dans toutes les conditions, « angoissée/oppressée » en condition « entier », et « hésitante » en condition « haut », ont été mieux reconnus en dynamique, alors que « un peu perdue/perplexe » en condition « entier », « angoissée/oppressée » en condition « haut », et « déçue » en condition « bas » l'ont mieux été en statique.

Sujet S

En dynamique et en condition « entier », tous les labels ont eu une distribution de leurs réponses différentes de celle du hasard (vs. en statique, il s'agissait du cas de « pas concentrée et envie de rigoler »). De plus, alors que « pas concentrée et envie de rigoler » en conditions « entier » et « bas », et « concentrée » en conditions « haut » et « bas » ont été mieux reconnus en dynamique, « écoute attentivement » l'a mieux été en statique pour la condition « entier ».

Par ailleurs, la méta-classe identifiée en dynamique (cf. Figure 3) est restée la même que celle relevée en statique.

5. CONCLUSION

Tout d'abord, nos IG ont été globalement correctement identifiées et leur pertinence validée, cela à la fois en statique et en dynamique. Alors que certains labels ont presque toujours été correctement identifiés, nous avons également dégagé des méta-classes de labels à partir des analyses de reports et confusions entre les labels : 4 pour le sujet T, toutefois différentes selon la nature (statique ou dynamique) des stimuli ; une seule dans tous les cas pour S.

De plus, cette évaluation indique que si une répartition entre les informations fournies par le haut et le bas du visage existe, ce qu'il reste à démontrer, alors elle n'est pas additive en termes de reconnaissance de label. Le *FACS* d'Ekman ne serait donc pas écologique, l'expression d'un état émotionnel ne pouvant être réduite à une somme d'*AUs,* IG Primitives selon notre terminologie. Nous pouvons toutefois préciser que la condition (haut/bas/entier) ainsi que la nature statique ou dynamique des stimuli sont des paramètres qui entrent en jeu dans la reconnaissance des IG. Lors du passage du statique au dynamique (cf. [4]) :

- Les informations liées à « concentrée », « envie de rigoler et répond au hasard » se concentrent dans le haut du visage, comme celles d'« hésitante », ou celles d'« écoute attentivement » qui y restent.

- Les informations liées à « stressée » passent du haut au bas du visage, comme celles de « calme/va bien » ou « pas concentrée et envie de rigoler », et celles de « rassurée/plus détendue » qui y restent.

- Les informations liées à « déçue » et « angoissée/ oppressée », auparavant respectivement dans le bas et le haut du visage, se retrouvent alors dans le visage dans son ensemble, comme pour « étonnée ».

Il sera donc important à l'avenir d'approfondir les traitements de ces résultats préliminaires et de continuer à étudier la dynamique du mouvement, car cette dernière semble pertinente dans certains cas pour distinguer certains labels. De plus, il apparaît que c'est parfois le fait même que le sujet fasse le mouvement qui est important.

Par ailleurs, il nous reste à vérifier pour le *Feeling of Thinking,* lors d'une tâche d'interaction, si la rythmicité d'un geste (comme sa régularité et sa fréquence) est un indice fort de l'état affectif dans lequel le sujet étudié se trouve, comme le montre Carlier & Graff [2] chez les joueurs de tennis de haut niveau.

6. PERSPECTIVES

Il serait d'abord intéressant de continuer à établir la typologie des signaux, en gardant l'élargissement du statique au dynamique amorcé ici. Puis mesurer la coordination des expressions dans les différentes modalités et établir des lois d'organisation temporelle pourrait nous éclairer quant au rôle et à l'organisation de la rythmicité dans la communication émotionnelle.

Nous pourrions ensuite comparer les performances en reconnaissance d'icônes naturelles vs synthétiques, en testant perceptivement nos IG simulés sur un ACA.

Puis il s'agirait d'établir un modèle sur la relation existant entre les expressions multimodales et les états mentaux et affectifs du sujet, avant de le vérifier, l'invalider, l'augmenter ou le spécifier sur des données diversifiées d'interactions.

Enfin, l'aboutissement de ce travail serait la simulation de ce modèle avec le contrôle augmenté de l'agent virtuel expressif GRETA, à travers une collaboration avec C. Pelachaud et le LINC [6]. Cette simulation serait évaluée par la réalisation de tests d'usabilité, qui nous permettraient de mettre à l'épreuve notre modèle et de tester la pertinence des simulations en termes écologiques (voir exemple Figure 4).

Figure 4 : Sujet T, corpus E-Wiz vs. Greta (Pelachaud – LINC)

BIBLIOGRAPHIE

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

[2] G. Carlier and C. Graff. Unpredictability as a counter strategy: An analysis of elite matches. *Journal of Sport Sciences*, 2006.

[3] P. Ekman. L'Expression des Emotions. In *Les Emotions,* B. Rimé and K. R. Scherer ed., Neuchâtel, Paris, Delachaux-Niestlé, pages 183-201, 1989.

[4] F. Loyau, V. Aubergé. Expressions outside the talk turn: ethograms of the Feeling of Thinking, *5th LREC,* pages 47-50, 2006.

[5] P. Peters, C. Pelachaud, E. Bevacqua, M. Mancini, I. Poggi. A model of attention and interest using gaze behavior. *IVA'05 International Working Conference on Intelligent Virtual Agents*, pages 229-240, 2005.

[6] I. Poggi, C. Pelachaud, F. de Rosis, V. Caroglio and B. de Carolis. GRETA. A Believable Embodied Conversational Agent. *Multimodal Intelligent Information Presentation*, O. Stock and M. Zancarano eds, Kluwer, pages 3-25, 2005.

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

[8] M. Swerts and E. Khramer. Audiovisual prosody and feeling of knowing, *Journal of Memory and Language*, 53:1, pages 81-94, 2005.

[9] A. Vanpé, V. Aubergé. Pertinence perceptive d'Icônes Gestuelles du « Feeling of Thinking », *WACA 2006*, Toulouse, pages 55-59, 2006