

Perception d'expressions multimodales du *Feeling of Thinking* (états mentaux et affectifs, intentions, attitudes) en interaction

Anne Vanpé

GIPSA-lab, UMR 5216 CNRS/Université de Grenoble
Université Stendhal, domaine universitaire BP 25 38040 Grenoble cedex 9, France
Tél. : ++33 (0)4 76 82 41 97 - Fax : ++33 (0)4 43 35
Courriel : anne.vanpe@gipsa-lab.inpg.fr

ABSTRACT

Human-Machine Interaction, as interaction between two humans, can be considered as a dynamic process where the human is continuously communicating, even when he is “expressively” listening (informative *backchannel* and *feedback*). The present study analyses the audio-visual non speech expressions for two subjects in spontaneous HMI corpora, following an ethology-based methodology. First results reveal a large panel of values expressed outside of turns (e.g. mental states, intentions, attitudes, emotions) that we have globally called *Feeling of Thinking*. We have shown the role of static vs. dynamic processing of visual information and we are now attempting to investigate some specific non speech “vocal events”. Their temporal distribution seems to be particularly relevant for the perception of *Feeling of Thinking* expressions.

1. INTRODUCTION

En Interaction Personne-Machine (IHM), et par extension en interaction entre deux humains, l'agent humain exprime ses intentions, affects sociaux, et ses états mentaux et émotionnels. Bien que les études des expressions et les études en *affective computing* (informatique affective) concernent principalement le locuteur lorsqu'il est en train de parler, l'information que ce dernier envoie en dehors de son tour de parole est également significative. Ainsi, l'inter-actant produit en continu des expressions variées exprimant en particulier l'évolution de ses processus cognitifs, comme son état de compréhension ou de résolution mentale d'une tâche, ou ses réactions affectives aux propos du locuteur, humain ou virtuel (*Mind Markers* [Pog07] –marqueurs de la pensée-, indices de *feedback* dans l'expression du *backchannel* [Pet05] [Sch06] –le *backchannel* étant une forme de participation de l'auditeur à la conversation-). Ces informations couvrent donc un spectre informationnel large [Loy07] et peuvent exprimer, à partir de contrôles volontaires ou involontaires, des états mentaux (concentration, *Feeling of Knowing* [Swe05] –« sentiment de savoir »-), intentions, attitudes (politesse, accord), et des états affectifs (déception, irritation, stress, calme). De plus, elles peuvent aussi bien être exprimées par des mouvements du corps ou de la face que par des événements vocaux comme les interjections, ou autres bruits de bouche [Loy07].

2. UNE MÉTHODOLOGIE INSPIRÉE DE L'ETHOLOGIE POUR ÉTUDIER LE « FEELING OF THINKING »

Ce présent travail porte sur les expressions authentiques audio-visuelles, hors tour de parole, de deux locutrices (choisies parmi 17 sujets) issus du corpus IHM E-Wiz SoundTeacher [Aub06]. Les sujets ont été piégés dans le même scénario d'induction émotionnelle, d'une durée moyenne de 40 minutes par sujet, contrôlée par un paradigme de type magicien d'Oz. Pourtant, les deux sujets sélectionnés pour cette étude présentent des réactions extrêmes et différentes en terme d'états mentaux et affectifs de par leur profil psychologique plutôt opposé sur un axe extra-introverti.

Leurs expressions ont révélé de nombreux événements non-verbaux significatifs, y compris des événements vocaux. Nous cherchons d'une part à savoir dans quelle mesure ces derniers fournissent de l'information sur les états mentaux ou affectifs du (non-)locuteur, d'autre part si ces informations sont portées par des formes perçues par des processus statiques [Ekm94], processus dynamiques, ou plus globalement par l'organisation rythmique de mouvements idiosyncratiques et/ou d'événements vocaux. L'analyse multimodale (voix, parole, langage, gestualité expressions faciales) a permis d'identifier chez les sujets des expressions 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 aux expressions du *Feeling of Knowing* [Swe05], révélant les processus mnésiques du sujet.

L'étiquetage de ces expressions est fondamental : dans une approche déductive, la théorie contraint l'étiquetage sur la nature, les valeurs et les processus des « états » induits par la tâche (SoundTeacher) ; alors que dans une approche inductive, il va être fondé sur les données. La complexité inter- et intra-sujet, ainsi que la diversité des expressions produites nous ont conduits à adopter une méthodologie inductive. Ainsi, les 17 sujets du corpus ont chacun annoté leurs propres états mentaux et affectifs lors de l'expérience, à la suite de l'enregistrement et en utilisant leur propre langage naïf. Ces auto-annotations « naïves » nous servent d'éléments d'échange lors des évaluations perceptives de certaines de nos données par des juges naïfs externes. En parallèle, deux chercheurs

n'ayant aucune information sur les auto-annotations des sujets, ni sur les inductions émotionnelles, ont étiqueté le corpus en événements visuels et vocaux aussi minimaux que possible, en veillant à ne pas les interpréter et sans hypothèse concernant les théories des gestes et des émotions et la pragmatique. En effet, même si de plus en plus d'études entrent dans les détails de mouvements subtils de *Feeling of Thinking*, en particulier au niveau facial [Ken04][Pog05], certains événements pertinents ne semblent pas encore avoir été identifiés comme tels dans les modèles théoriques ou empiriques existants. Cette observation justifie notre protocole relativement inductif et « en aveugle ».

Le but de l'étiquetage est de trouver des « icônes » gestuelles et vocales minimales mais communicativement pertinente pour l'expression du *Feeling of Thinking*. Pour cela, nous avons utilisé une méthodologie empirique empruntée à l'éthologie et avons ainsi établi des éthogrammes (inventaire des comportements d'un animal). L'éditeur d'annotation ANVIL¹ a été utilisé pour étiqueter les signaux audio-visuels du corpus, c'est à dire segmenter les mouvements en icônes gestuelles supposées minimales et fondés sur la variation des formes, sans interprétation. Chaque « icône » a ainsi été décrite précisément selon soit : (1) des critères de direction, localisation, répétition, vitesse, symétrie, stabilité, intensité et/ou amplitude (événements visuels) ; (2) des critères phonétiques et acoustiques (événements vocaux). Par la suite, la pertinence des événements étiquetés a été vérifiée par des analyses et des tests perceptifs.

3. PERCEPTION STATIQUE / DYNAMIQUE DES ICÔNES GESTUELLES

Dans un précédent travail, certaines de nos icônes gestuelles ont ainsi été testé pour être validées [Van06]. Deux tests perceptifs similaires ont été implémentés pour étudier dans quel mesure l'information visuelle exprimée est portée statiquement ou dynamiquement : le premier avec des formes statiques des stimuli (photos), le deuxième avec les mêmes icônes gestuelles sous leur forme dynamique (vidéos). De plus, afin d'évaluer l'hypothèse d'Ekman sur l'importance des informations statiques selon les différentes parties du visage [Ekm94], trois conditions de présentation ont été testées : haut du visage seul (« haut »), bas du visage seul (« bas ») et visage entier (« entier »), la face étant divisée selon une ligne horizontale passant par les ailes du nez. Les tests perceptifs ont consisté pour les juges en une tâche d'association entre les stimuli et un choix fermé parmi les labels d'auto-annotation du sujet lui-même. Alors que le temps d'exposition n'était pas limité pour les stimuli statiques, les stimuli dynamiques pouvaient être rejoués pendant 8 secondes.

Un des résultats les plus intéressants, en référence à des

¹Documentation sur ANVIL : Michael Kipp (2004), "Gesture Generation by Imitation - From Human Behavior to Computer Character Animation", Boca Raton, Florida: Dissertation.com.

modèles combinatoires des expressions faciales (cf. le FACS –Facial Action Coding System- d'Ekman), est qu'il ne semble pas, d'après nos résultats pour l'expression de ces états affectifs (affect étant pris dans son acception la plus large, de l'humeur à l'état mental) que les informations véhiculées par le haut et le bas du visage soient additives ou complémentaires en rapport des informations véhiculées par la face en entier. De plus, aucune des deux parties du visage ne contient une information nulle, et la dichotomie positive vs. négative de la valence des affects respectivement dans le bas vs. le haut du visage n'est pas non plus validée.

Toutefois, certains labels portent plus d'informations dans le haut ou le bas du visage, et cette répartition peut changer en fonction de la nature dynamique ou statique du stimulus présenté [Van08]. Ainsi, quand nous passons du statique au dynamique :

- 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 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 ».

Plus globalement, le gain d'information lors du passage du statique au dynamique semble profondément dépendre de la nature de l'information : l'identification de certains labels semble perturbée par le dynamique. Un point méthodologique peut être rappelé ici : un stimulus statique présenté, par nécessité écologique, pendant une durée de 3 secondes n'est pas perçu par les humains comme une image, mais plutôt comme une vidéo montrant une posture d'une durée de 3 secondes. Parallèlement, le paramètre du « temps d'exposition », la « durée d'immobilité » d'un visage qui ne bouge pas ou peu, peut être le seul paramètre informatif « dynamique » pertinent.

Par ailleurs, comme le fait de percevoir une image statique n'est pas écologique, une vidéo sans modalité auditive n'est pas complète écologiquement, d'autant plus que les événements vocaux semblent être très importants dans l'expression du *Feeling of Thinking*.

4. PREMIÈRE ANALYSE DES ÉVÉNEMENTS VOCaux

Dans une étude préliminaire, Loyau [Loy07] a montré que des juges étaient capables de replacer un rire dans un énoncé en termes d'organisation (par exemple choisir si un rire a lieu au début, au milieu ou à la fin de l'énoncé). Ainsi, un rire pourrait soit indiquer le contexte, soit être intrinsèquement contextuel. Serait-ce le cas pour les interjections et les autres événements vocaux ?

Dans notre corpus, nous avons étiqueté un grand nombre d'événements vocaux, très différents de par leurs paramètres articulatoires et acoustiques. Comme pour les événements gestuels/faciaux, nous cherchons à savoir quels paramètres peuvent être communicativement pertinents pour les expressions du *Feeling of Thinking*. C'est pourquoi, dans un premier temps, nous n'avons négligé aucun paramètre lors de cette première analyse.

4.1. Critères pour l'annotation

Pour étiqueter les événements vocaux de notre corpus, nous avons d'abord distingué les interjections des autres événements vocaux. En effet, les interjections peuvent être identifiées selon un critère de pré-lexicalisation : elles peuvent être décrites en utilisant les phonèmes du français et ont souvent été orthographiées. Nous pouvons même observer que certaines interjections peuvent être réalisées avec des variations phonologiques : par exemple « pfff » peut être articulatoirement produit soit [pf:], soit en utilisant un trill bilabial. Les interjections s'opposent ainsi aux autres événements vocaux (bursts –plosions-, grunts –grognements pré-verbaux, raclements de gorge, clicks, gémissements, bruits de respiration...), que nous grouperons sous la terminologie « bruits de bouche ». Bien que nous ayons distingué deux catégories, nous considérons un continuum allant des signaux spontanés aux pré-lexèmes bien définis, *i.e.* des bursts, puis gémissements, aux interjections et onomatopées.

Pour cette étude, les interjections ont été classées selon un critère phonétique/articulatoire : phonème vocalique (noté « V », *e.g.* [ø:] « euh »), consonantique (« C », *e.g.* [m:] « mmh »), ou combinaisons de phonèmes vocaliques et consonantiques (étiquetés « CV » -*e.g.* [bø:] « beuh »-, « VC » -*e.g.* [ø:m:] « euh mmh »). Une interjection de plus de deux phonèmes a été étiquetée « Comb » (pour combinaison), *e.g.* [bEm:] « ben mmh », [ula] « ouh là ».

De l'autre côté, nous avons utilisé deux distinctions basées sur des critères articulatoires triviaux pour classer les bruits de bouche : (1) s'ils sont produits par le sujet pendant une inspiration vs. une expiration ; (2) s'ils sont

produits avec un flux d'air continu (« continu », *e.g.* gémissement, grande inspiration) vs. si le flux d'air a été bloqué au moins une fois (« bloqué », y compris coup de glotte, *e.g.* click, plosive bilabiale). Ce dernier critère nous a intéressés car il implique une tension suivie d'un relâchement chez le sujet.

Au niveau temporel, nous avons classé les événements vocaux selon leur occurrence par rapport au tour de parole du sujet, et selon la phase du scénario lors de laquelle il apparaît, chacune associée à une induction particulière : phase 1 mise en confiance / familiarisation ; phase 2 feedback positif au sujet ; phase 3 feedback négatif ; phase 4 déstabilisation / tentative de générer peur et stress chez le sujet. Concernant le tour de parole, chaque événement a été noté : « 1 » lorsqu'il apparaissait pendant la production de parole, « 0 » en dehors, « Avant » ou « Après », moins d'une seconde avant ou après, et « ~ » lorsqu'il apparaissait pendant la production de parole, mais en dehors du flux de parole.

4.2. Effets inter-sujets et analyse globale

Les 2 sujets ont utilisé la modalité vocale différemment (Table 1), avec une plus grande proportion de bruits de bouche chez le sujet T (65,7%) que le sujet S (56,7%).

Pour le sujet T, nous avons observé une répartition entre les interjections et les bruits de bouche à peu près similaire selon les phases, avec une apparition globale surtout en phase 3, puis en phases 4, 2 et enfin 1. A l'inverse, pour le sujet S, les bruits de bouche étaient mieux répartis dans le temps (apparaissant tout au long du scénario), que les interjections, en majorité dans la phase 4, à un taux moindre dans la phase 2, et enfin à des taux faibles et similaires dans les phases 1 et 3. Quel que soit le sujet, le nombre d'interjections et bruits de bouche produits n'a pas été constant entre les phases.

4.3. Nature des interjections

Les répartitions temporelles des interjections selon leur classe ont montré de grandes différences quand elles étaient comparées entre elles (Tables 2 et 3).

Table 1 : Comparaison des pourcentages d'interjections et bruits de bouche pendant les différentes phases du scénario pour les sujets T et S (nombre entre parenthèses).

Phase	Sujet S			Sujet T		
	Interjections	Bruits de bouche	Sous-total S	Interjections	Bruits de bouche	Sous-total T
1	18% (17)	21% (25)	19% (42)	12% (10)	18% (30)	16% (40)
2	33% (31)	28% (34)	30% (65)	25% (21)	21% (35)	23% (56)
3	18% (17)	17% (21)	18% (38)	36% (31)	33% (53)	34% (84)
4	30% (28)	34% (42)	33% (70)	27% (23)	28% (45)	27% (68)
Total	100% (93)	100% (122)	100% (215)	100% (85)	100% (163)	100% (248)

Table 2 : Sujet S – nombre d’interjections selon leur nature pendant les différentes phases du scénario.

Sujet S	C		Total C	CV		Total CV	Comb		Total Comb	VC		Total VC	V	Total V	Total
Phase	fff	mmh		ben	mmh euh		ben mmh	et ben euh		euh fff	euh mmh		euh		
1		1	1	3		3			0		1	1	12	12	17
2	1	2	3	1	1	2			0		7	7	19	19	31
3			0	2		2	1		1	2	1	3	11	11	17
4		1	1			0		1	1	1	3	4	22	22	28
Total	1	4	5	6	1	7	1	1	2	3	12	15	64	64	93

Table 3 : Sujet T – nombre d’interjections selon leur nature pendant les différentes phases du scénario.

Sujet T	C		Total C	CV		Total CV	V		Total V	Variantes de V		Total Variantes de V	VC		Total VC	VCV	Total VCV	Total
Phase	fff	mmh		beu	mmh euh		[a]	euh		euh chuchoté	euh soupiré		euh fff	euh mmh		ouh là		
1		1	1			0		9	9			0			0		0	10
2	2		2		1	1		14	14	1	2	3		1	1		0	21
3	5	1	6	1		1		13	13	6	2	8		1	1	2	2	31
4			0			0	1	11	12		5	5	1	4	5	1	1	23
Total	7	2	9	1	1	2	1	47	48	7	9	16	1	6	7	3	3	85

4.4. Nature des bruits de bouche

Pour les deux sujets, les interjections ont été principalement vocaliques (68,5% pour le sujet S et 75% pour le sujet T). Quant à la répartition des interjections consonantiques et plus complexes, elle a différé selon les sujets et les phases du scénario.

Il est notable que pour les 2 sujets l’interjection la plus fréquente a été le phonème vocalique [ø:] (« euh »), apparaissant seul, combiné à d’autres phonèmes, ou encore avec des variations ([ø:] chuchoté ou soupiré) : pour le sujet S 69% des interjections ont été [ø:] et 87% [ø:] ou des combinaisons contenant [ø:] ; pour le sujet T, 55% des interjections ont été [ø:], et 85% [ø:], ses variantes, ou des combinaisons contenant [ø:]. Pour le sujet T, la répartition des variantes de [ø:] selon la phase du scénario a été comparable à la répartition globale des interjections, alors que le nombre de [ø:] simple a été presque constant entre les phases 2, 3 et 4.

D’autre part, les interjections C et CV ont rarement été trouvées lors des phases où sont produites le plus d’interjections (phases 2 et 4 pour le sujet S, phases 3 et 4 pour le sujet T).

Comme pour les interjections, la répartition des bruits de bouche en fonction de leur nature a été différente d’un sujet à l’autre (Table 4). Plus particulièrement, la répartition des bruits de bouche selon leur position temporelle par rapport à la production de parole nous a semblés pertinente. Les analyses du corpus ont montré pour les deux sujets une plus grande quantité de bruits de bouche inspirés qu’expirés (87 et 95 inspirations vs. 38 et 63 expirations, respectivement pour les sujets S et T. Par ailleurs la répartition des bruits de bouche produits avec un flux d’air continu vs. bloqué a différé selon le sujet : alors que nous avons relevé plus de bruits de bouche « bloqués » à la fois en inspiration et en expiration pour le sujet T, les bruits de bouche inspirés ont été principalement « continus » pour le sujet S, alors que ses bruits de bouche expirés ont été plutôt « bloqués ».

À la suite de ces premières analyses, nous focalisons maintenant nos recherches sur les relations existantes entre les types d’inductions, le phénomène de tour de parole et la production d’événements vocaux. Nous nous intéresserons également particulièrement à la prosodie de ces événements, leur durée étant alors considérée comme un point clé.

Table 4 : Comparaison inter-sujets pour le nombre de bruits de bouche selon leur nature et leur position temporelle relative au tour de parole (dans chaque cellule [sujet S / sujet T])

Tour de parole	Inspiration		Expiration		raclement de gorge	déglutit	Total
	continu	bloqué	continu	bloqué			
0	6 / 7	9 / 14	1 / 13	14 / 19	1 / 1	0 / 4	31 / 58
1	0 / 0	0 / 0	1 / 0	0 / 12	0 / 0	0 / 0	1 / 12
~	30 / 4	4 / 8	3 / 0	5 / 3	0 / 0	0 / 0	42 / 15
Avant	7 / 21	7 / 32	2 / 0	0 / 8	0 / 1	0 / 2	16 / 64
Après	23 / 7	1 / 2	4 / 2	8 / 6	1 / 3	0 / 0	37 / 20
Total	66 / 39	21 / 56	11 / 15	27 / 48	2 / 5	0 / 6	127 / 169

5. CONCLUSION ET PERSPECTIVES

Notre approche empirique et relativement éthologique nous a permis d'établir certains facteurs jouant sur la perception du comportement de deux locutrices, l'une introvertie et l'autre extravertie, à partir du corpus expressif IHM spontané SoundTeacher. Les types d'états pour lesquels nous avons étudié les expressions se sont montrés complexes, des états mentaux aux émotions, comme les auto-annotations naïves des sujets eux-mêmes nous l'indiquaient avant même leur validation par des auditeurs naïfs.

Nous avons en particulier montré que l'information visuelle fournie par le haut et le bas du visage n'était pas additive, quel que soit la nature ou la valence de l'état exprimé. D'autre part, l'amplitude des mouvements faciaux ne peut être considéré directement comme un paramètre et doit être relié au locuteur : par exemple un locuteur plutôt introverti va produire des mouvements de plus faible amplitude, mais pas moins informatifs que des mouvements amples produits par un locuteur extraverti. Comme nous l'avions supposé, certains événements visuels ont été identifiés selon un processus de perception statique, avec néanmoins une durée de présentation des stimuli probablement porteuse de sens, alors que d'autres gestes ont été décodés par leur dynamique. Ces deux phénomènes gagneraient à être étudiés plus précisément. Nous avons également l'intention d'étudier certains gestes spécifiques et événements vocaux, intrinsèquement non porteurs d'informations affectives, mais qui pourraient le devenir à travers l'évolution de leur rythmicité (leur régularité et leur fréquence) au cours de l'interaction. Ces événements concerneraient plus le comportement global du locuteur que l'observation d'événements à des moments précis, comme l'ont montré Carlier et Graff [Car06] au cours de leurs travaux sur le comportement de joueurs de tennis de haut-niveau.

Par ailleurs, il est notable que les occurrences d'événements vocaux non-verbaux ont été différentes selon le sujet, à la fois dans la manière de les utiliser que dans le choix de leur nature (e.g. nous n'avons relevé aucun déglutissement particulièrement perceptible pour le sujet introverti). De plus, les résultats apportent des informations pertinentes sur le comportement du sujet, relatives aux phases d'induction et à l'organisation des tours de parole. Ces résultats prometteurs nous poussent donc à étendre l'étude aux autres sujets du corpus, afin d'augmenter la quantité de données traitée en même temps que la variété des comportements observés, fonction des profils de sujets. En parallèle, des travaux sur les rires, antérieurs [Loy07] et actuels, montrent que ces derniers sont acoustiquement variables et peuvent être perceptivement reliés à leur paramètres contextuels (comportement du locuteur, indices pragmatiques du timing de l'interaction, états mentaux et affectifs).

Plus globalement, nous faisons l'hypothèse qu'étudier les événements vocaux non-verbaux dans leur contexte gestuel et observer en détails le contexte acoustique des

icônes gestuelles, nous donneront une large échelle d'interprétation en relation avec les paramètres du contexte, applicable à la fois en situation d'IHM et d'interaction communicative entre deux humains.

RÉFÉRENCES

- [Aub06] Aubergé V., Audibert N., Rilliard A. (2006), "De E-Wiz à C-Clone", *Recueil, modélisation et synthèse d'expressions authentiques. Revue d'Intelligence Artificielle - "Interactions émotionnelles"*, Vol. 20 (4-5), pp. 499-528,.
- [Car06] Carlier P., Graff C. (2006), "Unpredictability as a counter strategy: An analysis of elite matches", *Journal of Sport Sciences*, à paraître.
- [Ekm94] Ekman P. (1994), "Strong evidence for universals in facial expressions: A reply to Russell's mistaken critique", *Psychological Bulletin*, Vol. 115, pp. 268-287.
- [Ken04] Kendon A. (2004), *Gesture: Visible Action as Utterance*. Cambridge University Press.
- [Loy07] Loyau F. (2007), *Expressions des états mentaux et émotionnels de l'humain en interaction : ébauches du Feeling of Thinking*, Thèse de Doctorat en Sciences Cognitives. Institut National Polytechnique de Grenoble.
- [Pet05] Peters P., Pelachaud C., Bevacqua E., Mancini M., Poggi I. (2005), "A model of attention and interest using gaze behavior", *IVA'05 International Working Conference on Intelligent Virtual Agents*, pp. 229-240.
- [Pog05] Poggi I., Pelachaud C., de Rosis F., Caroglio V., de Carolis B. (2005), "GRETA. A Believable Embodied Conversational Agent", *Multimodal Intelligent Information Presentation*, O. Stock & M. Zancaranò [Eds], Kluwer, pp. 3-25.
- [Pog07] Poggi I. (2007), *Mind, hands, face and body. A goal and belief view of multimodal communication*, Weidler, Berlin.
- [Sch06] Schröder M., Heylen D., Poggi I. (2006), "Perception of non-verbal emotional listener feedback", *Speech Prosody*, CD-Rom proceedings, SPS1-4-72.
- [Swe05] Swerts M., Krahmer E. (2005), "Audiovisual prosody and feeling of knowing", *Journal of Memory and Language*, Vol. 53(1), pp. 81-94.
- [Van08] Vanpé A., Aubergé V. (2008), "Static vs. dynamic Gestural Icons of Feeling of Thinking", *2nd International Workshop on Corpora for Research on Emotion and Affect, in LREC'08*, Marrakech, Maroc, CD-ROM proceedings.