HAL
open science

Étude de la perception des émotions dans les micro-événements non verbaux chez l'être humain

Gilles de Biasi

To cite this version:

Gilles de Biasi. Étude de la perception des émotions dans les micro-événements non verbaux chez l'être humain. Linguistique. 2011. dumas-00631033

HAL Id: dumas-00631033

https://dumas.ccsd.cnrs.fr/dumas-00631033v1

Submitted on 11 Oct 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Étude de la perception des émotions dans les micro-événements non verbaux chez l'être humain

Nom : DE BIASI
Prénom : Gilles

UFR Sciences du Langage

Mémoire de master 2 recherche - 30 crédits – Mention Sciences du Langage

Spécialité : Modélisation et Traitements Automatiques en Industries de la Langue : Parole, Écrit, Apprentissage

Parcours : Traitement Automatique de la Langue Écrite et de la Parole

Sous la direction de Véronique Aubergé

Année universitaire 2010-2011

Étude de la perception des émotions dans les micro-événements non verbaux chez l'être humain

Nom : DE BIASI
Prénom : Gilles

UFR Sciences du Langage

Mémoire de master 2 recherche - 30 crédits – Mention Sciences du Langage

Spécialité : Modélisation et Traitements Automatiques en Industries de la Langue : Parole, Écrit, Apprentissage

Parcours : Traitement Automatique de la Langue Écrite et de la Parole

Sous la direction de Véronique Aubergé

Année universitaire 2010-2011

Remerciements

A ma directrice de recherche, Véronique Aubergé, pour avoir encadré ce travail,

Au personnel du Gipsa-Lab, qui m'a accueilli au sein du laboratoire, mention spéciale pour Anne, Rosario et Lionel pour leur aide précieuse.

A mes professeurs de master, qui ont su partager leur savoir,

A mes camarades de classe, avec qui j'ai eu plaisir à travailler durant ces deux années de Master,

Aux juges qui ont participé à mon expérience, sans qui ce travail n'aurait pas abouti,

A ma famille et mes amis, pour leur soutien,

A tous ceux qui prendront le temps de lire le compte-rendu de mon travail.

Sommaire

PARTIE 1 ETAT DE L'ART **7**

CHAPITRE 1 – LES EMOTIONS 8

1.1 Les émotions dans la philosophie *9*

1.2 Les émotions vues par la science *11*

1.3 Conclusion sur les émotions *15*

CHAPITRE 2 – LE LANGAGE 17

2.1 La Communication *17*

2.2 Langage verbal *21*

2.3 Langage non verbal *22*

2.4 Les Agents Conversationnels Animés (ACA) *24*

2.5 Conclusion sur le langage *25*

PARTIE 2 METHODOLOGIE EXPERIMENTALE **27**

CHAPITRE 3 – CADRE THEORIQUE ET ENJEUX DE L'ETUDE 28

3.1 Cadre théorique *28*

3.2 Enjeux de l'étude *30*

CHAPITRE 4 – DU CORPUS A L'EXPERIENCE 32

4.1 Corpus Sound Teacher, plateforme E-wiz *32*

4.2 Sélection des stimuli de chaque sujet *33*

4.3 Les juges *34*

CHAPITRE 5 – DEROULEMENT DU TEST 35

5.1 Interface d'expérimentation *35*

5.2 Procédure *35*

5.3 Limites de l'expérience *36*

PARTIE 3 RESULTATS DE L'EXPERIENCE **38**

CHAPITRE 6 – RESULTATS ET PERSPECTIVES 39

6.1 Traitement statistique *39*

6.2 Analyse des résultats *40*

6.3 Conclusion sur les résultats *50*

6.4 Perspectives *51*

Introduction

Aujourd'hui, rendre l'interaction homme machine écologique est un enjeu technologique majeur pour le domaine des industries de la langue. Nous ne possédons actuellement pas de modèle mathématique scientifique (calculable) des émotions, c'est pourquoi nous réalisons notre étude sur l'humain. Nous nous intéressons particulièrement à l'information contenue dans les expressions présentes en dehors du tour de parole, dans les micro-événements audibles et/ou visibles qui font la richesse de l'interaction humaine. Notre travail, qui n'est qu'un petit pas devant l'ampleur de la tâche à accomplir, vise comme finalité de rendre possible et efficace l'implémentation des émotions, des attitudes et des états mentaux aux Agents Conversationnels Animés.

Ce travail s'articule en trois parties : nous ferons d'abord un tour d'horizon des notions d'émotion et de langage à travers un état de l'art qui pose le cadre théorique de notre étude. Ensuite, nous exposerons la démarche utilisée pour monter notre expérience et la mener à bien. Enfin, nous proposerons une analyse des résultats et discuterons des suites à donner à notre étude.

Partie 1

Etat de l’Art

Chapitre 1 – Les émotions

Le Trésor de la Langue Française informatisé (TLFi) donne cette définition du substantif émotion : « Sens vieilli : mouvement assez vif ; sens plus récent : conduite réactive, réflexe, involontaire, vécue simultanément au niveau du corps d'une manière plus ou moins violente et affectivement sur le mode du plaisir ou de la douleur ».

Toute la difficulté tient dans le fait de donner une définition à une entité psychologique et il en va de même pour le concept d'émotion. La quête de la définition de l'émotion pose un problème car les énoncés ne se rapportent très souvent qu'à un aspect de l'émotion. En effet, le concept d'émotion est répertorié de manière différente s'il est utilisé en référence à l'aspect stimulus, à l'expérience subjective, à une phase d'un processus ou à une réponse.

Un autre problème qui nuit aux progrès vers une meilleure précision dans la définition de l'émotion concerne le langage par lequel on l'exprime. En effet, le langage de tous les jours et le langage scientifique ne visent pas les mêmes objectifs. La langue de tous les jours sert à communiquer, elle permet de répondre aux besoins physiologiques et psychologiques de l'être humain en lui donnant la possibilité d'échanger des informations sur le monde qui l'entoure et sur lui-même. La langue scientifique a quant à elle une visée plus philosophique. Elle permet une normalisation du langage en répondant aux besoins d'un cadre précis qui amène à la compréhension universelle d'une idée. De plus, actuellement, les avancées scientifiques dans ce domaine n'offrent pas de meilleure terminologie, le mot émotion englobant un tas de notions pas toujours faciles à démêler, d'où la complexité à définir un modèle représentatif de l'émotion.

Précisons qu'il peut être intéressant de ne pas avoir de définition trop rigoureuse du terme « émotion », compte tenu du stade de développement dans ce domaine. Donner une définition précise aurait pour conséquence de créer des frontières entre les phénomènes. On courrait le risque d'écarter de l'étude des aspects qui pourraient ensuite se révéler fondamentaux pour la compréhension de l'ensemble du processus.

Dans cette partie, nous nous attarderons sur les théories et modèles émotionnels développés en philosophie et dans les sciences au cours de l'Histoire de l'être humain.

1.1 Les émotions dans la philosophie

De tout temps l'homme a cherché à comprendre, à expliquer, à définir ce qu'est l'émotion. Depuis l'antiquité, on trouve différentes écoles de pensées sur l'émotion. « Il est traditionnel en philosophie, de diviser les phénomènes psychiques selon trois catégories : représentation, volonté et affectivité, cette dernière englobant les émotions, les passions et les sentiments. Les phénomènes affectifs sont ainsi mis à part, sans confusion possible avec l'activité intellectuelle (fondée sur la représentation) ni avec la vie morale (fondée sur la volonté). » (Culture Générale, Hatier, p.8)

Une première tentative de théorisation des émotions est attribuée à Héraclite (550 – 480 av. J.-C.) qui considérait l'émotion comme une augmentation de la température corporelle. Il s'agit de l'un des plus anciens essais de caractérisation physiologique des émotions.

Quelques années plus tard Anaxagore (500 – 428 av. J.-C.) pensait que toute perception sensorielle était accompagnée d'une douleur inconsciente qui passait à l'état conscient lorsque celle-ci devenait suffisamment intense. Cette conception se rapproche déjà de la théorie cognitive de l'émotion, dont nous parlons plus loin.

Socrate (470 – 399 av. J.-C.) et Platon (428 – 347 av. J.-C.) considéraient l'émotion comme une menace pour la raison et un danger pour la philosophie et les philosophes. Ils voyaient dans l'émotion un rôle subordonné, l'émotion serait primitive, bestiale, non intelligente et dangereuse et devrait être contrôlée par la raison.

Aristote (384 – 322 av. J.C.) a passé du temps à lister et analyser les émotions. Il les considérait comme centrales et essentielles pour la « bonne » vie. Les émotions sont des caractéristiques qui permettent de construire des arguments persuasifs (Aristote, l'Ethique). Il attribuait aux émotions une fonction informative. Pour lui l'homme ne subit pas ses émotions. Il construit un univers émotionnel qui lui est propre à partir de ses valeurs, ses goûts, ses croyances et sa morale.

René Descartes (1596 – 1650) définit les « passions », entre autres, comme des émotions de l'âme. Pour lui, l'émotion est un type de passion. Il développa la notion de dualité corps/esprit qu'il considérait comme deux entités distinctes. Son argument est que le raisonnement s'exécute de façon distincte du corps et que les émotions sont liées à l'activité de l'esprit. Les émotions seraient des passions perturbées et influencées par la raison. Contrairement aux théories précédentes, Descartes ne rejeta pas totalement une

certaine physiologie des émotions : l'âme est influencée par le corps et par ses changements d'état. Dans son traité Les Passions de l'âme, il identifia six émotions primaires : la joie, l'admiration (équivalent de la surprise de nos jours), l'amour, la haine, le désir et la tristesse, sans parti pris théologique ou posture moralisatrice (« Les passions sont toutes bonnes de leur nature et nous n'avons rien à éviter que leurs mauvais usages ou leur excès » (Descartes, Les Passions de l'âme)) contrairement à ses prédécesseurs. Toutes les autres émotions seraient ainsi composées de ces six de base. On retrouvera cette approche dans les travaux d'Ekman.

Baruch Spinoza (1632 - 1677) voyait les émotions comme une forme de pensée qui ne comprend pas le monde et nous rend malheureux et frustrés. Il ne croyait pas que les émotions demeurent dans un corps séparé en conflit avec l'âme, car ce sont deux aspects d'une seule réalité. La vision de Spinoza implique l'idée que l'Homme de passion a un comportement changeant : « Les hommes peuvent différer en nature en tant qu'ils sont dominés par des affections qui sont des passions ; et dans la même mesure, le même homme est changeant et inconstant » (Spinoza, L'Ethique, 1677). Il propose un modèle de classification en trois passions : le conatus (désir de persévérer dans son être), la joie et la tristesse. Les combiner entre elles et les lier avec des objets extérieurs permettrait de former toutes les autres passions.

David Hume (1711 - 1776), philosophe fervent partisan des Lumières, donna aux passions une importance supérieure à la raison. Il reprend le terme d'émotion au sens de Descartes et le définit comme un type de sensation qu'il appelle impression et qui serait physiquement stimulée par le mouvement de l'esprit animal dans le sang, c'est-à-dire qui nait quand l'esprit fait face à une stimulation de son appétit, en bien ou en mal.

Tout au long de l'histoire des émotions/passions dans la philosophie on retrouve des points prédominants. Des oppositions, raison contre passion en premier lieu, où l'émotion entrave la connaissance de la vérité en nous condamnant à vivre dans l'illusion, action contre passion dans un second temps, où l'émotion est une dépendance aux événements et aux objets du monde. Vient ensuite le lien entre passion et désir, l'émotion a une forte emprise sur l'esprit et entraine des mouvements difficilement contrôlables. On retrouve également la notion de maîtrise des passions, puisque ces dernières sont dangereuses pour l'homme. Enfin, la valorisation des passions, avec l'idée que priver un homme de ses émotions (dont les propriétés sont bénéfiques), c'est lui enlever une partie essentielle de son être.

1.2 Les émotions vues par la science

Après avoir parcouru le monde des émotions dans la philosophie, attardons-nous à cette même notion dans le domaine scientifique.

1.2.1 L'approche darwinienne et néo-darwinienne

Darwin (1809-1882) est un naturaliste anglais à qui l'on doit la théorie de l'évolution. Il publie en 1872 « l'expression des émotions chez l'homme et les animaux », ouvrage dans lequel il explore l'évolution de la psychologie humaine parallèlement au comportement des animaux. D'après lui les humains et les animaux montrent leurs émotions aux autres, un argument qui lui permet de consolider la théorie d'une évolution partagée entre les espèces.

Selon cette approche, les émotions éprouvées s'adapteraient à l'environnement pour préserver la survie de l'espèce. Les humains et les non humains (génétiquement proches des humains), devraient ressentir les mêmes émotions à cause d'une origine biologique commune. Darwin apporta plusieurs contributions au domaine de l'émotion et de l'expression des émotions. Il fut l'un des premiers à considérer les émotions comme des entités discrètes et séparées. Darwin ne décrivait pas chaque émotion en tant que faisant partie d'une famille, mais il soulignait les caractéristiques qui distinguent une émotion d'une autre. Il tenait pour acquis que c'est la morphologie de l'expression faciale qui transmet les informations qui caractérisent la manifestation d'une émotion. Il prouva aussi qu'une émotion est universelle, au sens où la même émotion peut être éprouvée par des peuples de cultures différentes, et pour soutenir sa thèse il s'appuya sur l'universalité des expressions faciales.

Ekman, né en 1934, est un psychologue américain qui étudie les émotions et les relations de ces dernières aux expressions faciales. Il reprochera à Darwin de ne pas avoir fourni une méthodologie précise pour mesurer les mouvements faciaux et de n'avoir pas non plus observé et défini les frontières de chaque famille d'émotions. Ekman a démontré en 1972 l'existence de six émotions primaires universelles (basic emotions) qui sont la joie, la surprise, la peur, la colère, le dégoût et la tristesse. Il est important de souligner qu'Ekman ne prétend pas que les émotions se réduisent aux « big six » même si les travaux technologiques s'en inspirant se réduisent à ces six catégories simplistes. Ekman dira également que l'expression faciale est le pivot de la communication entre les hommes. En effet, savoir lire sur le visage facilite nos relations sociales ; de même, une interprétation

erronée d'une mimique faciale peut nous faire adopter un comportement mal adapté à la situation. Ekman reconnait le rôle des influences sociales sur l'adaptation des émotions.

Figure 1: Les Big Six d'Ekman, joie, surprise, peur, colère, dégoût et tristesse (Sander et Scherer, 2009, p.88).

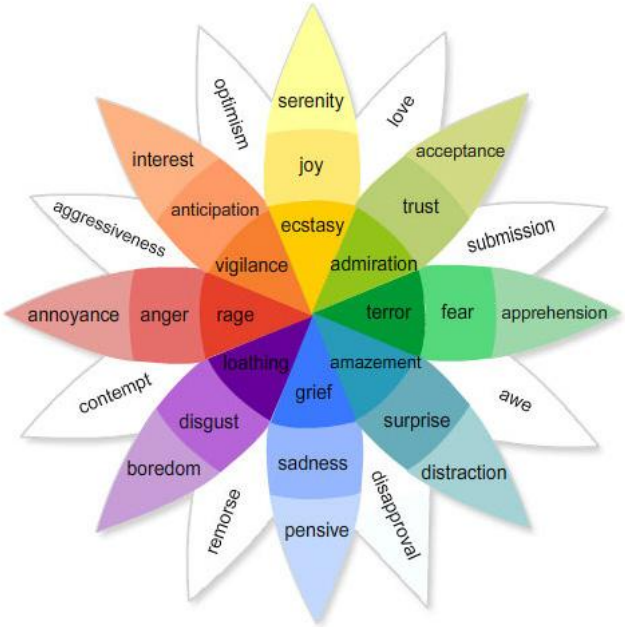

Figure 2 : Une autre vision des émotions de base : Le cercle des émotions de Plutchik (Neil, 2009).

1.2.2 L'approche physiologique

William James (1842 – 1910), est un psychologue et philosophe américain, d'abord professeur adjoint de physiologie (1876) ; il affirme en 1884 qu'une émotion n'est pas une

cause mais un effet. Il prétend qu'on ressent une émotion parce qu'on a d'abord des signes physiologiques mesurables (larmes, accélération des battements du cœur, changement de la pression sanguine…) qui nous amènent à tel ou tel état émotionnel, même si ces phénomènes se produisent de façon involontaire.

La théorie James-Lange prétend qu'une émotion est une réponse aux modifications physiologiques. « Nous nous sentons tristes parce que nous pleurons, en colère parce que nous frappons quelqu'un, et effrayé parce que nous tremblons » (Lange, 1884, La théorie de l'émotion, L'Harmattan, 2006). Cela nous amène à penser que nous contrôlerions nos émotions en agissant différemment, par exemple si je ne pleure pas, je ne suis pas triste.

Les recherches scientifiques en neurobiologie et en neuroscience prétendent que l'être humain est le seul à pouvoir penser et agir sur ses émotions, s'éloignant ainsi du concept d'évolution partagée établi par Darwin. A l'heure actuelle les scientifiques décortiquent les émotions et font appel à des techniques de pointe pour explorer les connections avec le cerveau humain. On peut considérer que James William est celui qui fait le joint entre la partie philosophique/historique et les recherches actuelles sur les émotions.

1.2.3 L'approche cognitive

Cognitif signifie qui est lié au processus d'acquisition de connaissance. Dans certaines visions philosophiques la cognition a toujours été vue comme une adversaire des émotions, du fait que les émotions empêcheraient le bon fonctionnement de la pensée rationnelle. L'approche cognitive des émotions s'oppose à l'approche physiologique dans le sens où ici, on considère que c'est l'émotion la cause, et la réponse physiologique du corps la conséquence.

Magda Arnold (1960) et Richard Lazarus (1966) furent parmi les premiers théoriciens en psychologie à parler de la théorie de l'évaluation cognitive. Ils proposèrent le rôle central de l'appraisal (processus d'évaluation) dans la génération et dans la différentiation des émotions. Les émotions se fonderaient donc sur une évaluation subjective de l'importance d'un événement externe pour le bien-être de la personne concernée. C'est seulement après évaluation d'une situation que le cerveau réagit et que l'émotion émerge. Arnold considère le processus d'appraisal comme inconscient, involontaire, automatique, non réflexif et non-intellectuel. Plus tard, Lazarus (1966) apporta sa contribution en théorisant un processus de reappraisal (réévaluation) qui serait

capable de modifier la première impression ainsi que l'émotion déclenchée. Un événement entraine une évaluation cognitive consciente, qui conduit à une l'émotion.

Pour Klaus Scherer, les émotions représenteraient le carrefour entre un environnement versatile et l'individu. Elles permettraient à ce dernier de s'adapter à l'environnement dans lequel il se trouve et possèderaient une fonction adaptative au sens darwinien du terme. Le modèle cognitif de Scherer s'intéresse aussi à l'expression vocale des émotions (production et perception) en essayant de mesurer les signaux véhiculés afin de les relier aux émotions qui sont exprimées.

Antonio Damasio (né en 1944 à Lisbonne) bâtit sa théorie sur l'interaction entre le fonctionnement des émotions et la raison. Descartes théorisait une conception dualiste entre corps et âme tandis que pour Damasio l'émotion et la raison sont deux parties de la même essence psychophysiologique. Il fait une distinction entre émotions primaires et secondaires et entre émotions fortes et faibles. Pour lui la perception des émotions passe par la perception des changements d'états du corps. La perception de l'émotion serait la conséquence d'une action volontaire ou involontaire. Il a démontré qu'émotions et prises de décision sont liées : « l'esprit et le corps ont des racines communes » (Damasio, Spinoza avait raison : Joie et tristesse, le cerveau des émotions, Odile Jacob, Paris, 2003).

1.2.4 Le constructivisme social

L'approche constructiviste est représentée par James R. Averill. Il donne une importance fondamentale aux interactions sociales et culturelles dans la transmission des émotions. Averill définit les émotions comme des syndromes. Il pense à une série de réponses qui covarient de façon systématique. Les émotions sont décrites comme des constructions sociales (et non biologiquement déterminées), fondées et dépendantes de l'interprétation individuelle du contexte social dans lequel elles se manifestent. Elles fonctionnent à partir des règles sociales apprises par l'individu qui joue un rôle défini comme une série de réactions que l'individu doit avoir dans une société et dans des situations données.

Les syndromes sont alors des manifestations du rôle joué par l'individu et ils représentent la réalisation d'un rôle social momentané. Ils peuvent être complètement compris seulement s'il y a une analyse au niveau social de la part de l'individu. Ce dernier doit forcement connaître le système culturel d'origine d'une émotion pour pouvoir la comprendre. Les émotions ont également un objet qui dépend de l'évaluation individuelle

d'une situation : nous sommes en colère envers quelqu'un, nous avons peur de quelque chose, etc.

Pour un constructiviste l'émotion est une construction sociale et non une réponse adaptative à l'environnement. Le sens d'une émotion est donné par l'apprentissage d'une ou plusieurs lois sociales qui sont apprises ou comprises après une analyse de la société. L'approche constructiviste s'oppose donc au point de vue évolutionniste sur l'existence d'émotions de base primitives et universelles. Ce sont les normes et les attentes sociales qui définissent quelle émotion doit être éprouvée, à quel moment, dans quel contexte et par qui.

1.2.5 L'approche linguistique

Anna Wierzbicka (1992) donna une théorie linguistique de l'émotion. Selon elle les émotions de base, dans le sens courant du terme, seraient des artefacts culturels de notre langage. Elle assigne au langage un rôle important dans la conceptualisation des émotions et elle suggère à tous les chercheurs qui s'occupent des émotions de prendre en compte la façon dont le langage obstrue l'accès direct aux émotions. Elle ne veut pas affirmer qu'il n'existe pas d'émotions universelles et non plus qu'il est impossible de pénétrer et de comprendre les émotions de ceux dont la culture est différente de la nôtre. Au contraire, elle fait une étude des émotions à partir d'une perspective vraiment universelle et donc indépendante du langage et de la culture. Elle précise toutefois que certaines émotions ne peuvent pas être décrites et caractérisées, de façon nette et entière, par les mots de chaque langue.

Wierzbicka croit que la compréhension des émotions progresserait mieux si elle s'appuyait sur ce qu'elle appelle les primitives sémantiques universelles. Elle met en évidence le fait que la plupart des études sur les émotions, conduites par les psychologues, ont été faites suivant une dépendance culturelle. Le vocabulaire des émotions a une signification qui dépend de la culture à laquelle il appartient. Selon Wierzbicka, l'analyse des émotions devrait dépendre d'une explication faite par des petits concepts qui soient à la fois universels et sémantiquement primitifs.

1.3 Conclusion sur les émotions

Actuellement, étudier les émotions devrait amener les scientifiques à travailler de façon pluridisciplinaire, chaque discipline complétant et faisant progresser d'autres

domaines de recherche, que ce soit dans les secteurs de la biologie, de la neurobiologie, de la linguistique, que l'on s'intéresse aux relations entre émotions et cognition ou au verbal/non verbal.

Beaucoup de théories, peu de certitudes. C'est ce qu'il convient de retenir de ce tour d'horizon des émotions. Sartre disait : « A présent nous pouvons concevoir ce qu'est une émotion. C'est une transformation du monde. […] Le passage à l'émotion est une modification totale de « l'être-dans-le-monde » selon les lois très particulières de la magie. » (Sartre, Esquisse d'une théorie des émotions, Paris, Hermann, 1939, p.43, 66). Sans a priori ni parti-pris, notre étude se basera non pas sur une théorie en particulier, mais sur l'observation d'objets physiques repérés dans un corpus émotif audio-visuel.

Chapitre 2 – Le langage

Moyen de communication développé par l'être humain, le langage se définit traditionnellement comme un ensemble de signes doté d'une sémantique et d'une syntaxe. Nous allons ici nous attarder sur la communication, le langage verbal et non verbal et les Agents Conversationnels Animés (ACA).

Le TLFi définit le langage comme suit : « 1 - Le langage comme faculté et comme système ; les hommes expriment leur pensée et communiquent à l'aide d'un système de signes conventionnels formé de sons différents, identifiables et signifiants ; système de signes vocaux et/ou graphiques, langages naturels (langues parlées), langages artificiels (système de symboles, langage de programmation en informatique, SMS, etc.…) ; 2 – Le langage comme moyen d'expression ; d'un point de vue de la forme ou du contenu (langue des signes, grammaire, etc.…) ».

On voit bien ici que la notion de langage est vaste et complexe et comprend tout ce qui est du domaine de la communication ou de l'expression.

2.1 La Communication

La communication est un concept fondamental pour l'homme. Communiquer, c'est :

-établir une relation avec un interlocuteur,

-dire quelque chose à quelqu'un (transmettre une information),

-parler, échanger,

-ressentir,

-émettre un message,

-donner, recevoir, demander une information, obtenir une réponse.

Il existe différentes formes de communication : la communication verbale (les différentes langues du monde), la communication non verbale (gestuelle, mimique, expression, bruits de bouches, etc.…), le langage des gestes (langue des signes notamment), les ondes électriques (radio, télégraphes, signaux lumineux, etc.…). Le point commun de ces formes de communication est qu'elles sont toutes un ensemble de

techniques, de moyens pour transmettre une information d'un émetteur à un récepteur. On voit bien ici que la communication ne s'envisage qu'en interaction, c'est à dire en présence de deux « individus » au minimum.

2.1.1 Communiquer d'un point de vue historique/philosophique

On est tenté de croire que communiquer est propre à l'homme puisqu'il a le monopole du langage, mais il n'en est rien. On trouve des systèmes de communication sophistiqués dans le monde animal aussi. On pourra citer l'exemple de la marmotte, qui siffle pour alerter ses congénères d'un danger, et dont le sifflement changera selon que l'intrus est un randonneur ou un aigle. Même le monde végétal dispose d'un système de communication. « Tout parle dans l'univers, il n'est rien qui n'ait son langage » (Jean de La Fontaine, extrait d'Epilogue).

Depuis les débuts de son existence l'homme a développé la faculté d'émettre des sons qui se sont peu à peu transformés en mots, en langage, nos cordes vocales ainsi que notre maitrise du conduit vocal se sont développées pour émettre des sons de plus en plus élaborés, tout un code s'est mis en place, donnant à l'homme un système plus complexe et unique à notre sens.

L'arrivée de l'écriture a permis de laisser une trace écrite de la parole, avec un impact dans les relations sociales, économiques, politiques, ou encore scientifiques.

Aujourd'hui, grâce à la technologie scientifique, nous disposons de moyens technologiques avancés pour communiquer : la téléphonie, la télévision, internet ; des outils qui nous permettent d'accéder à une somme considérable d'informations, de transmettre des messages, des informations, d'être en relation même en étant très éloignés, et tout cela toujours plus vite. Cependant l'être humain est-il bien capable de rester vigilant quant à la qualité et aux risques induits par cette communication qui envahit peu à peu son quotidien ? (explosion des réseaux sociaux sur internet, applications développées pour les Smartphones afin d'être en « connexion avec le monde » en permanence, etc.…).

2.1.2 Communiquer d'un point de vue scientifique

C'est par la mise au point de différents modèles que le processus scientifique tente d'expliquer le concept de communication.

Le modèle d'Harold Dwight Lasswell (1902,1978), politologue et psychiatre, voit la communication comme un processus d'influence et de persuasion. Il envisage la

communication comme un processus dynamique composé d'une suite d'étapes significatives avec leur problématique et leur spécificité. Néanmoins, la rétroaction ainsi que les contextes psychologiques et sociologiques sont ignorés.

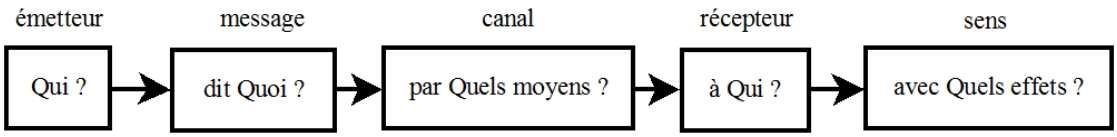

Figure 3 : Le modèle de Lasswell (Lasswell, Structures et fonctions de la communication dans la société, 1948)

Le modèle de Shannon (1916, 2001) et Weaver (1894, 1978), mathématiciens, réduit la communication à la simple transmission d'un message. Il s'agit d'un schéma simpliste qui ignore la pluralité des récepteurs et laisse de côté les éléments psychologiques et sociologiques. De même, ils considèrent le récepteur comme étant passif, privant la communication de la notion de rétroaction.

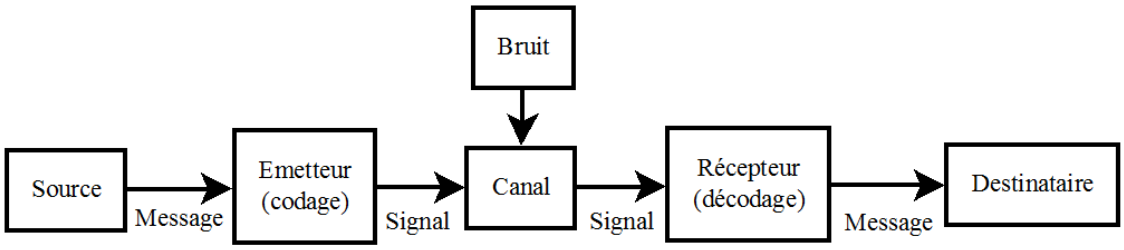

Figure 4 : Le modèle de Shannon et Weaver (Shannon et Weaver, Théorie mathématique de la communication, 1949).

On voit bien apparaitre la faiblesse de ces deux modèles dans le fait que la communication ne peut pas se résumer à un processus linéaire destiné uniquement au transfert d'informations. En outre, ils ignorent totalement la notion d'interaction qui nous est si chère en privant le récepteur de toute possibilité de rétroaction (même si on trouvera dans les versions postérieures des schémas de ces deux modèles une flèche supplémentaire introduisant la rétroaction et permettant de fermer la boucle de la communication).

Le modèle de John Riley et Matilda Riley va proposer une correction à ces modèles en introduisant justement une boucle de rétroaction entre l'émetteur et le récepteur, le feedback, mettant l'accent sur le phénomène d'interaction présent dans la communication. Cette dernière n'est donc plus vue comme un processus linéaire mais bien comme une boucle. Le contexte d'énonciation et l'idée d'appartenance à un groupe entrent également en jeu, mettant l'accent sur l'influence de la représentation du monde vue par un groupe

donné. Le contexte social gagne donc une place importante dans la représentation de la communication.

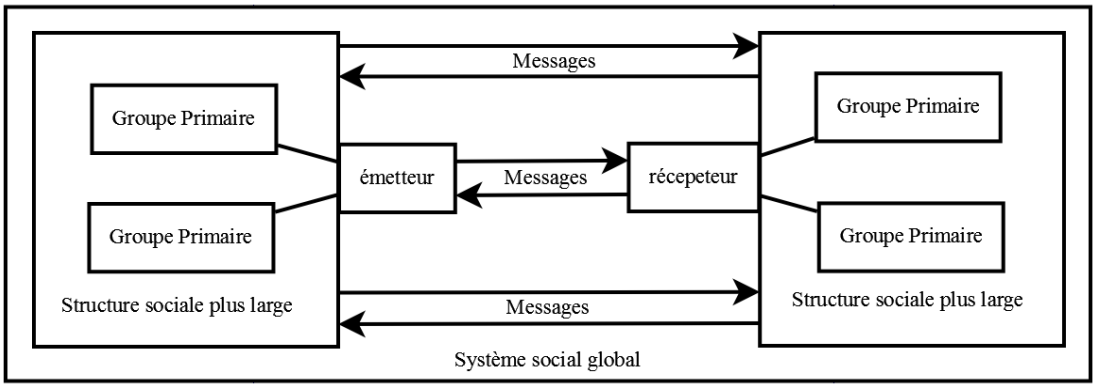

Figure 5 : Riley et Riley, Communication de masse et systèmes sociaux, 1965

Le modèle de Jakobson (1896, 1982) se compose de six facteurs associés à six fonctions du langage :

-Le destinateur, celui qui envoie un message au destinataire, fonction expressive du langage (expression des sentiments du locuteur),

-Le destinataire, celui qui reçoit le message, fonction conative du langage (relative au récepteur),

-Le message, fonction poétique du langage (se rapporte à la forme du message),

-Le contexte, c'est-à-dire l'ensemble des conditions extérieures ayant une influence sur la compréhension du message, fonction référentielle du langage (le message renvoie au monde extérieur),

-Le code, qui est le symbolisme utilisé pour la transmission du message, fonction métalinguistique du langage (le code devient objet du message),

-Le contact, symbolise la liaison physique, psychologique et sociologique entre émetteur et récepteur, fonction phatique du langage (mise en place et maintien de la communication).

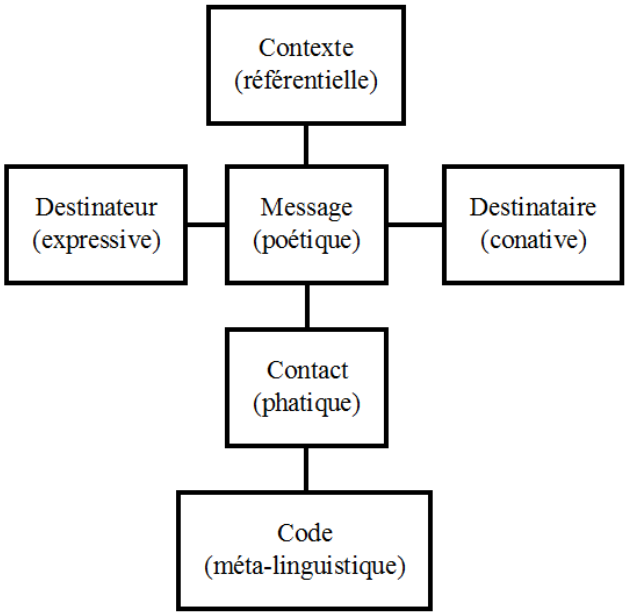

Figure 6 : Schéma de Jakobson (Jakobson, Closing statements : Linguistics and Poetics, in Style in language, 1963)

Avec ces quelques approches non exhaustives de la communication, nous avons posé l'enjeu de l'interaction comme crucial pour bien communiquer. C'est sur cet enjeu que repose un point essentiel de notre étude : l'humanisation des Agents Conversationnels Animés afin de les placer pleinement dans l'échange et d'aboutir à une véritable interaction homme machine. D'énormes progrès ont déjà été réalisés dans le domaine du langage verbal. Le langage non verbal est quant à lui un vecteur de rétroaction sur lequel il reste de nombreux travaux à effectuer afin d'humaniser les ACA ou du moins de s'en rapprocher le plus possible.

2.2 Langage verbal

Que serait la communication sans le langage verbal ? Différent dans chaque langue, le verbal permet de nommer tout objet du monde matériel et un nombre très important de concepts sans forme physique. Il englobe donc les mots et leur sens sémantique. La culture vient jouer un rôle colossal dans le langage, ainsi deux langues différentes n'auront pas forcément les mêmes nuances et le langage en lui-même permettra de faire ressortir de nombreuses facettes des mœurs et comportements d'individus ressortissant d'une culture donnée.

Le langage verbal est propre à l'être humain et il est le médium privilégié lors du processus de communication, cependant il n'est pas le seul langage existant.

2.3 Langage non verbal

La linguistique s'arrête le plus souvent à l'étude du verbal et laisse de côté toute la partie non verbale de la communication. Pourtant, le non verbal est porteur de sens, en particulier dans le domaine des émotions. En effet, un soupir, un mouvement dans le visage, des joues qui rougissent nous donnent chacun une information sur l'état émotionnel de notre interlocuteur. On mesure donc ici l'intérêt du non verbal. A la fois audible et visible, le non verbal comprend différents objets à détailler. Il permet de transmettre un nombre important d'informations complémentaires à ce que l'on dit de manière verbale à notre interlocuteur. De plus, au même titre que le langage verbal, le langage non verbal varie d'une culture à l'autre. Dans notre travail, le non verbal est un phénomène « qui ne véhicule pas une fonction de communication construite par le langage » (Loyau, 2007, p.119). Le non verbal ne se contrôle pas si on n'y pense pas ; par exemple suer quand on a peur, rougir quand on est intimidé ou gêné sont des réactions de l'ordre du réflexe automatique.

Le non verbal dans la relation à l'autre, peut être l'indicateur de notre état émotionnel, peut nous aider ou nous trahir, peut faire que la relation soit empathique ou au contraire éloigner l'autre. Parler du non verbal nous amène à une multitude de théories en psychologie, en sociologie, en management, en sciences du langage selon lesquelles pour mieux communiquer il faut apprendre à gérer ses émotions, à les comprendre, à les décortiquer et à les maitriser.

2.3.1 Le non verbal visible

La partie visible du non verbal inclut les expressions faciales, les gestes, les mimiques, le regard. On peut également inclure dans le non verbal la façon de s'habiller, de se tenir, de se présenter. Déjà dans la seconde moitié du XIXème siècle, le neurologiste Duchenne de Boulogne réalisait une série d'expériences sur l'expression faciale de l'émotion. Il remarqua notamment que les sourires exprimant la joie se réalisent par la contraction d'un muscle situé autour des yeux (orbicularis oculi), ce qui n'est pas le cas des sourires mécaniques produits volontairement. Ces travaux seront confirmés et complétés par Ekman en 1980.

La gestuelle a été étudiée par plusieurs scientifiques dont Jacques Cosnier et David McNeill. Pour le premier, « l'activité mimo-gestuelle » est le processus de co-pilotage des gestes sur la parole : « comment donner sens à cette phrase « c'est celui-ci qui me plaît le

plus » si un geste de pointage n'est pas présent pour le contextualiser ? » (Cosnier, 1996, p.116). McNeill, de son côté, s'est occupé de la fonction communicative des gestes et du caractère multimodal de la parole (Vanpé, 2006). Il distingue deux types de gestes communicatifs : les emblèmes (gestes qui ont du sens en eux-mêmes, déclenchés volontairement par le locuteur et indépendants de la parole (McNeill, 1992, cité par Vanpé, 2006, et par Signorello, 2010)) et les gestes co-verbaux (gestes non autonomes de la parole, non interprétables hors-contexte et se manifestant spontanément pendant la communication verbale).

Pour le visage, on s'intéresse au regard et aux expressions faciales. Le regard est un mode très important pour la régulation de la communication (Cosnier, 1996). Ce processus de régulation se caractérise par l'émission d'un locuteur (activité phatique (voir explication du schéma de Jakobson)), et l'émission d'un allocutaire (activité régulatrice, c'est-à-dire rétroaction/feedback) (Cosnier, 1996).

Du côté des expressions faciales, on trouve Adam Kendon (2002) qui s'intéresse aux mouvements de la tête utilisés par le locuteur dans la prise du tour de parole, et à ceux de l'interlocuteur pour exprimer son feedback. Les expressions faciales sont primordiales dans le développement des Agents Conversationnels Animés, l'objectif étant de rendre les expressions et mouvements des visages de ces faces virtuelles le plus réaliste possible.

Dans les travaux récents et en rapport avec notre étude, on citera la thèse d'Anne Vanpé (2011) et son étude de la perception des icones gestuelles d'un sujet en trois conditions (bas du visage, haut du visage, visage entier), et deux modalités (statique et dynamique).

2.3.2 Le non verbal audible

Nous communiquons quand nous parlons en envoyant des signaux explicites pendant le tour de parole et quand nous ne parlons pas en envoyant des signaux non explicites, chargés de sens, hors du tour de parole.

C'est ici que ce situent les objets les plus intéressants pour notre étude. On trouve dans le non verbal audible tout ce qui est de l'ordre du bruit produit par l'être humain, à savoir l'intonation de la voix, la respiration, la qualité de voix, sa fréquence et les événements vocaux allant du non-lexical (affect bursts, vocal bursts, grunts) au pré-lexical (interjections, vocal hesitation, fillers) (Vanpé et Aubergé, 2010).

Concernant les interjections, on retiendra le travail de Wundt (1900, cité par Scherer, 1994) qui fut l'un des premiers à les distinguer et à spécifier leur rôle dans le langage. Il séparait les interjections en deux classes : primaires (les sons de la nature qui ont pour seule fonction d'interrompre la continuité de la parole) et secondaires (composées des sons assimilés à la langue, elles dépendent donc de la culture et de la langue de référence).

Dans les travaux récents et en rapport avec notre étude, on citera le mémoire de Rosario Signorello (2010) et son étude de la perception de la culture d'un sujet (française ou italienne) au moyen de stimuli non verbaux audio-visuels (semblables à ceux que nous utiliserons).

Les travaux de Signorello (2010) et Vanpé (2006-2011), utilisent des stimuli extraits du corpus Sound Teacher (présenté dans la partie II), dont l'annotation a été réalisée par Anne Vanpé.

2.4 Les Agents Conversationnels Animés (ACA)

L'histoire des ACA remonte à la création des premiers chatterbots, et notamment à celle d'ELIZA. Créée par Joe Weizenbaum en 1966, ELIZA est un logiciel qui tente de dialoguer avec un être humain en donnant l'apparence d'une véritable conversation. Au départ simple logiciel de questions-réponses, la technologie a permis d'aboutir aux ACA et de donner à ces derniers un statut de plus en plus important dans le monde d'aujourd'hui. On peut les définir comme suit : « Les ACA sont un type d'interface visant à transférer les propriétés de l'interaction homme-homme, avec toute leur richesse, à l'interaction homme-machine. Il s'agit de personnages virtuels multimodaux capables de communiquer avec l'utilisateur par de multiples modalités : la voix, les expressions faciales, la direction du regard, les gestes des mains, les postures corporelles, etc. » (Buisine et al., 2006, p. 622).

Figure 7 : GRETA, un ACA en animation 3D (Pelachaud, 2010)

Nous avons vu dans la partie sur la communication l'importance du feedback pour la réussite de l'interaction chez l'être humain, c'est pourquoi les chercheurs du domaine tentent de faire ressembler ces ACA le plus possible à l'homme en voulant leur donner toutes les capacités émotionnelles propres à l'homme. Simuler les émotions et tous les autres traits caractéristiques de l'homme en situation d'interaction est un champ d'application en plein essor. L'implémentation efficace du non verbal dans les ACA est donc un enjeu de taille pour l'évolution de ces derniers.

Sansonnet (2005) attribue trois rôles principaux aux ACA : assistant (accueille l'utilisateur et aide ce dernier à comprendre et à utiliser des applications ou des services informatiques), partenaire (par exemple partenaire ou adversaire de jeu dans un environnement virtuel) et tuteur (pour les apprenants en Environnements Interactifs d'Apprentissage Humain). Ces trois rôles sont employés massivement par tous les secteurs (cinéma, jeux-vidéo, assistance téléphonique, apprentissage, etc…).

2.5 Conclusion sur le langage

Cette présentation du langage humain pose les bases de notre travail. Notre étude se place dans une perspective à long terme, car nous sommes convaincus que l'implémentation efficace de toutes les composantes du langage non verbal humain dans les ACA sera un pas en avant considérable dans l'utilisation des ACA.

Notre étude sur l'humain et les résultats qu'elle fournira font donc partie d'un projet au long cours qui permettra d'appliquer et d'implémenter les connaissances extraites

dans les ACA. Ce mémoire n’est que le prélude d’un projet aux nombreuses perspectives enthousiasmantes.

Partie 2

Méthodologie expérimentale

Chapitre 3 – Cadre théorique et enjeux de l'étude

3.1 Cadre théorique

3.1.1 Différence entre émotion et attitude

L'hypothèse principale à la base de notre travail est qu'il existe différents niveaux cognitifs d'informations affectives, intégrés à la parole.

Nous considérons ainsi que les émotions résultent d'un contrôle involontaire qui s'exprime par et dans la voix. Le processus d'expression des émotions est inné, et dispose d'un contrôle en intensité volontaire allant de l'inhibition à l'exagération.

Les attitudes (affects intentionnels) sont contrôlées de manière volontaire et s'expriment directement, tout particulièrement par la prosodie. Les concepts et morphologies prosodiques sont acquis (installés entre 7 et 11 ans selon Clément (1999 cité par Vanpé, 2006)), en fonction de la langue et de la culture. Ce type d'affects occupe majoritairement le canal de la parole (Campbell 2004, Aubergé et al., 1997, Aubergé, 2002, Shochi et al., 2005). Il permet au locuteur d'informer son interlocuteur sur ses intentions (doute dirigé contre soi ou contre l'autre, politesse, confiance en soi ou en ses propos, autorité, etc.).

3.1.2 Evolution du Feeling of Knowing au Feeling of Thinking

Le Feeling of Thinking (FoT) comprend tous les gestes, mouvements du visage, du buste, et les événements vocaux produits par un sujet donné hors du tour de parole en situation de communication expressive (en IHM dans le cadre de notre étude). Tout état mental et affectif manifesté de façon visuelle et/ou sonore fait parti du FoT (Loyau, 2007). Néologisme introduit par Véronique Aubergé et ses collaborateurs à l'ICP (Institut de la Communication Parlée) de Grenoble, il est le fondement de plusieurs travaux de recherche (Loyau et Aubergé, 2006, Loyau, 2007, Vanpé et Aubergé, 2010, Vanpé, 2011).

Quant au terme FoT, il tire son origine du concept de Feeling of Knowing (FoK), c'est-à-dire le « jugement que font les sujets vis-à-vis de leur propre capacité à se rappeler des informations non accessibles au moment où ce jugement est fait » (Loyau, 2007, p.119). La notion de Fok est apparue pour la première fois dans un article de Hart, comme le rappellent Reder et Ritter : « earlier researchers (Hart, 1965, 1967) defined FoK as only the state of believing that currently unrecallable information will be available (in some form) later because the knowledge is in memory. » (Reder et Ritter, 1992, p.435).

Pour la psychologie cognitive, en particulier dans le domaine de la métacognition, le FoK est inclus dans le processus automatique déclenché lorsqu'on pose une question à un sujet, ce qui entraine un mécanisme de sélection de stratégie de réponse (Nhouyvanisvong et Reder, 1998). D'après Reder, le Fok aurait alors deux fonctions distinctes :

-guider la sélection d'une stratégie avec pour facteur la familiarité (rapid FoK)

-réguler le temps de recherche de la réponse une fois le processus de récupération enclenché, avant d'en fournir une ou pas (post-retrieval failure).

3.1.3 Lien entre la prosodie expressive audiovisuelle et le FoK

Le lien entre prosodie audiovisuelle et FoK a été mis en évidence par Swerts et Krahmer en 2005. Ils se sont intéressés aux indices audio-visuels de métacognition en situation de questions/réponses pour des tâches de perception et de production. Les différents événements sur lesquels Swerts et son équipe se sont focalisés sont :

-les mouvements des sourcils, le nombre de regards vers le haut, le bas, les côtés et les sourires pour le visuel.

-le délai de réponse, le nombre de mots contenus dans une réponse, l'intonation haute en fin d'énoncé (type interrogation) et les filled pauses (type « hum », « euh », etc…) pour l'audio.

Lors d'une première expérimentation, les chercheurs ont ainsi pu démontrer qu'un certain nombre de ces événements audio-visuels sont des indices du FoK en établissant une correspondance étroite entre les taux de FoK et la justesse de la réponse donnée, ainsi qu'une proportionnalité inverse entre FoK, temps de réponse et nombre d'indices produits. Cela signifie que plus le sujet a un FoK élevé, moins son temps de réaction avant de répondre est long et moins il produit d'indices. Pour les non-réponses (de type « je ne sais pas »), ces résultats s'inversent. C'est pour cela que les types de production identiques en terme d'indices pourraient concerner des opérations mentales similaires, plus particulièrement dans le processus de recherche de la réponse du sujet dans sa mémoire.

Une deuxième expérimentation, sur la perception du FoK d'autrui (FoAK, feeling of another's knowing) dans trois modalités (audio seul, visuel seul, audio-visuel) a démontré que les observateurs humains peuvent discriminer une réponse au FoK fort d'une

autre au FoK faible dans les trois modalités et plus particulièrement en condition audio-visuel. Ainsi, les indices audio-visuels des productions semblent être pertinents pour l'interprétation par le juge de l'indice de confiance du locuteur envers sa réponse (Loyau, 2007).

3.1.4 Le FoT, une généralisation du FoK

Nous pensons que la gestualité du corps et du visage est une modalité primordiale des indices idiosyncrasiques et des signaux des états mentaux et affectifs du sujet (Loyau, 2007). En effet, tout locuteur continue à exprimer ses propres états même en dehors du tour de parole. Le facteur le plus important dans cette gestualité est l'organisation temporelle des signaux : une même personne peut produire un même geste à des intervalles plus ou moins rapprochés et avec une amplitude plus ou moins grande, ce qui peut être un indice indirect d'un état spécifique de cette personne. Le FoK ne se limite alors pas à décrire les expressions chez un sujet relevant d'un processus mnésique (récupération d'informations dans la mémoire déclenchée par une question) mais fait également partie d'une phénoménologie plus générique qui relève des processus affectifs et mentaux du sujet lors de toute tâche communicative (même en condition homme-machine), pendant et hors du temps de parole du sujet. Loyau, 2007, Vanpé et Aubergé, 2010, Vanpé, 2011, parlent de « Sentiment de Penser » ou FoT pour désigner les expressions des états affectifs et mentaux.

3.2 Enjeux de l'étude

Nous nous intéressons ici à la perception du non-verbal par l'être humain. Le langage non verbal inclut entre autres : les expressions faciales, le regard, les mouvements et la gestualité des mains, des bras et des jambes, la posture du corps, la tonalité de la voix, sa qualité, sa fréquence, l'intonation, et les événements vocaux. Toute la partie vocale du non verbal est à prendre en compte pour répondre à la question : où commence le langage ?

Notre étude utilise des objets isolés issu d'un corpus annoté par Anne Vanpé, et se place directement dans la continuité de sa thèse (Vanpé, 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").

Notre objectif est d'évaluer la perception d'émotions par des juges français sur des signaux courts. Les stimuli seront présentés selon trois modalités aux juges : Audio seul (A), Vidéo seule (V), et Audio-Visuelle (AV). Dans le but d'éviter un phénomène d'apprentissage

par les juges sur les sujets observés, ou encore un phénomène de report de A sur V ou de V sur A, l'ordre de présentation sera A-V-AV pour la moitié des juges, V-A-AV pour l'autre moitié.

Chapitre 4 – Du corpus à l'expérience

4.1 Corpus Sound Teacher, plateforme E-wiz

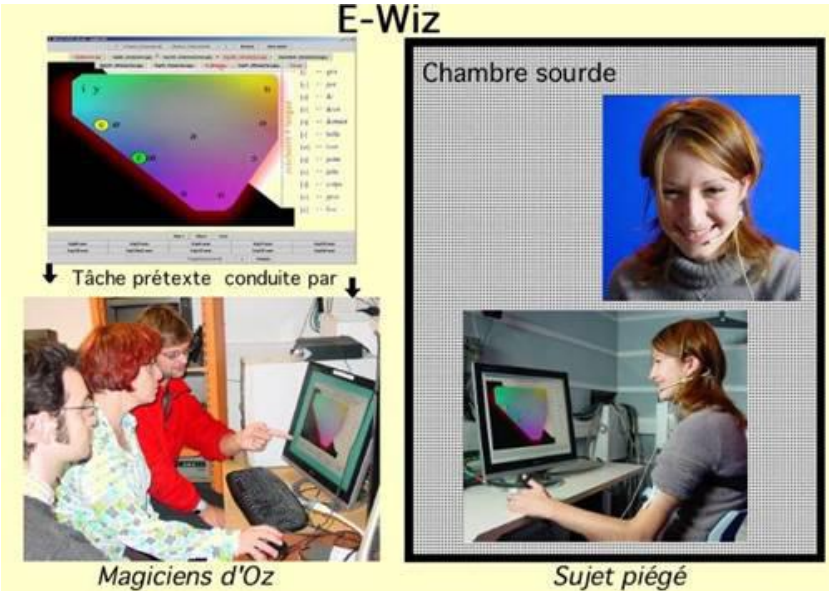

Figure 8 : La plate-forme E-Wiz et le paradigme du Magicien d'OZ. (Vanpé, 2006, p. 48)

Les sujets utilisés proviennent du corpus Sound Teacher, recueilli par Véronique Aubergé, Nicolas Audibert et Albert Rilliard à l'ICP. Corpus enregistré en situation naturelle in vitro, il est composé de vidéos de 17 sujets francophones en situation d'interaction homme-machine.

Pour notre étude nous avons retenus 6 sujets (3 femmes : T, S, M et 3 hommes : J, N, R).

Ce corpus fournit des expressions émotionnelles authentiques mais contrôlées, grâce à l'utilisation du paradigme du « Magicien d'Oz ». Les sujets pensent communiquer avec un ordinateur, alors que le déroulement de l'application est géré à distance par un humain. Le scénario se déroule en quatre phases :

-La phase d'entraînement familiarise le sujet à l'application.

-La deuxième phase implique le sujet dans des tâches très simples sur lesquelles on lui renvoie des appréciations positives, induisant chez l'ensemble des sujets des émotions positives.

-La troisième phase, plus complexe, renvoie au sujet des jugements négatifs, ce qui induit des émotions négatives.

-La quatrième phase, répétition de la tâche initiale, retourne au sujet des résultats très mauvais, dans le but de le déstabiliser, ce qui induit des émotions négatives.

Au cours de ces différentes phases, les sujets se trouvent alternativement en phase de lecture, de réflexion et de production de parole. Chaque sujet pense contrôler le logiciel à l'aide de commandes vocales destinées à une machine. Il peut donc parler :

-pour donner une réponse aux stimuli envoyés par le manipulateur : cinq mots monosyllabiques (sable, rouge, vert, jaune, brique).

-pour dire « page suivante », et ainsi passer à l'étape suivante.

-pour faire des phrases de commentaires, propres à chacun des sujets.

Suite à l'expérience, chaque sujet se voit confier sa vidéo et on lui demande d'annoter ses émotions et/ou états mentaux sur différentes plages de la vidéo. Ces annotations sont cruciales dans notre expérience, puisqu'elles constituent le matériel de réponse fourni aux juges.

4.2 Sélection des stimuli de chaque sujet

Nous avons isolé un ensemble d'annotations pour chaque sujet. Les annotations retenues sont celles pour lesquelles le choix de stimuli correspondant à l'annotation était assez vaste pour permettre une sélection. Nous avons à chaque fois retenu 2 stimuli par annotation. Nous avons donc 4 annotations pour le sujet J (8 stimuli), 5 pour les sujets T et N (10 stimuli), 6 pour les sujets S et M (12 stimuli), et 7 pour le sujet R (14 stimuli). Nous disposons d'un total de 198 stimuli (66 x 3 modalités).

Chacun des sujets piégés dans le corpus Sound Teacher est disponible sous forme de fichier vidéo. Une grille d'annotation réalisée sous le logiciel Anvil par Anne Vanpé (2006 - 2011) est associée à chacun des six fichiers vidéo retenus pour notre expérience. Nous avons repéré dans ces grilles les auto-annotations des sujets qui apparaissent plusieurs fois et pendant lesquels il se passe un nombre important d'événements vocaux. Après avoir dressé une liste de ces stimuli, nous avons découpé les fragments qui nous intéressaient avec le logiciel Virtual Dub. Nous avons ensuite regardé et écouté

individuellement chacun des stimuli extraits afin de déterminer lesquels possédaient le plus d'intérêt pour notre expérience.

Après ce premier visionnage, nous avons dressé un tableau récapitulatif des stimuli en fonction des annotations, en essayant de regrouper les annotations des différents sujets dans des « classes » plus globales[1]. Une deuxième session d'écoute/visionnage a ensuite été réalisée en se basant sur le tableau récapitulatif afin de ne garder que des paires dans un maximum de cases, ce qui nous a amené à conserver 65 fichiers (un des fichiers ayant été dupliqué car esseulé, nous arrivons bien au total de 66 annoncé plus haut)[2].

Les vidéos des sujets n'étant pas toutes encodées de la même façon, par souci d'uniformité, et pour éviter d'éventuels problèmes de compatibilité lors du codage de notre expérience, nous avons également utilisé Virtual Dub pour ré-encoder une partie des vidéos.

Ces événements vocaux ainsi isolés seront soumis à des juges français de manière aléatoire. L'ordre dans lequel les sujets sont présentés est tiré au sort, ainsi que l'ordre de présentation des stimuli de chacun de ces sujets. En effet, chaque sujet est associé à une liste propre de stimuli. On mélange les 6 sujets aléatoirement, puis pour chaque sujet, la liste de stimuli est mélangée à son tour.

4.3 Les juges

Au nombre de 15, les juges ont été sélectionnés selon les critères suivants :

-pas de limite d'âge

-langue maternelle française

-pas de problème d'audition

-pas de discrimination dans le choix du sexe

Nous avons donc retenu 9 femmes et 6 hommes.

[1] Annexe 1
[2] Annexe 2

Chapitre 5 – Déroulement du test

5.1 Interface d'expérimentation

L'interface du test a été conçue avec le logiciel Runtime Revolution Studio[3]. L'application développée nécessite une initialisation en cliquant sur le bouton Init (voir figure 9 ci-dessous). Le manipulateur charge alors en mémoire la liste de stimuli aléatoires à utiliser pour la session en cours, puis il entre les informations concernant le juge sous la forme : numéro d'identification (correspond au numéro d'identification de la liste de stimuli), Sexe (M ou F), nationalité (FR).

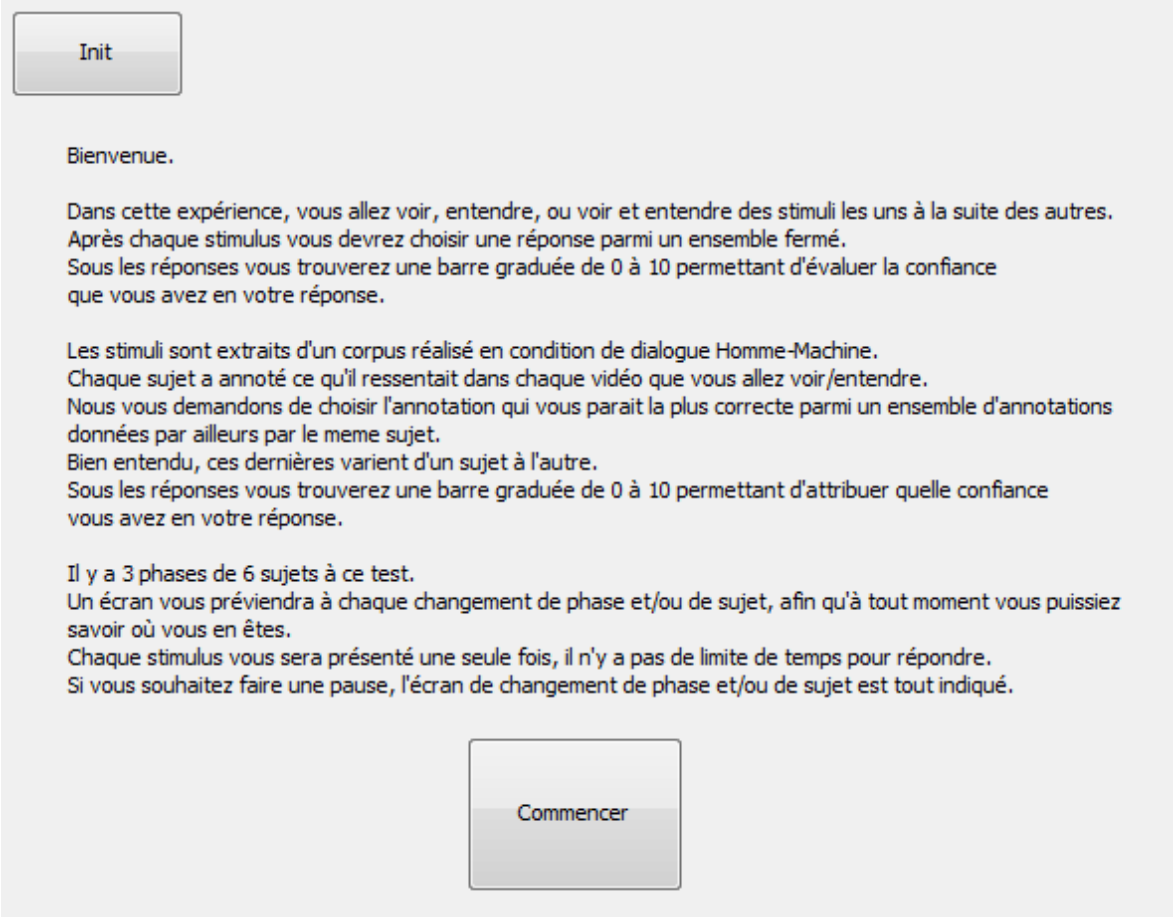

Figure 9 : écran d'accueil de l'expérience.

5.2 Procédure

Une page de consignes (voir figure 9) permet d'abord à l'utilisateur de se familiariser avec l'expérience. Il peut ensuite soit poser des questions au manipulateur, soit commencer l'expérience. Le test se déroule en 3 phases présentant chacune les 6 sujets. Un

[3] Annexe 3

écran prévient l'utilisateur à chaque changement de phase pour qu'il sache s'il va entendre, voir, ou bien les deux en même temps, et à chaque changement de sujet. De cette façon, le juge sait toujours où il en est dans le déroulement de l'expérience. Chaque écran de réponse possède un bouton suivant. Un clic sur ce bouton permet l'enregistrement de la réponse de l'utilisateur dans un fichier texte, et le lancement du stimulus suivant jusqu'à épuisement de la liste. L'écran de réponse contient de 4 à 7 réponses (selon le nombre d'annotations retenu pour le sujet) sous la forme de boutons radios, ainsi qu'une barre de mesure graduée de 0 à 10, permettant à l'utilisateur de quantifier la confiance qu'il a en sa réponse.

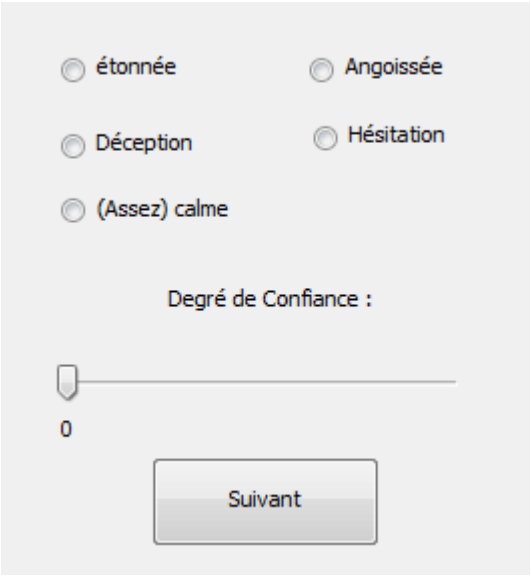

Figure 10 : écran réponse du sujet T.

Il n'y avait aucune limite de temps pour répondre, mais les juges ont généralement répondu assez rapidement, compte tenu du fait que chaque stimulus est très court et ne peut être visionné/entendu qu'une seule fois. Au total, l'expérience a duré de 45 minutes à 1 heure 30 minutes selon les juges.

5.3 Limites de l'expérience

Les auto-annotations fournies par les sujets sont une monnaie d'échange entre eux et les juges. Elles ont pour avantage de faire appel au ressenti naïf des sujets quant à leurs émotions et leurs états mentaux. C'est ce même ressenti naïf qui va permettre aux juges d'interpréter le sens de l'auto-annotation.

Les auto-annotations possèdent donc un inconvénient majeur. En effet, quand le sujet M annote avec le mot doute, on est en droit de se demander si le sujet doute de lui ou

de son interlocuteur. De plus, la manière même de rédiger l'auto-annotation peut donner un indice sur la personnalité du sujet. A titre d'exemple, les juges n'ont pas eu la même réaction lors de la première apparition de l'écran de réponse du sujet T (qui contient des annotations très courtes) et l'apparition de l'écran de réponse du sujet S (dont les annotations sont particulièrement longues et détaillées).

L'utilisation des auto-annotations comme support de réponse est donc à double tranchant, elle est une façon de ne pas risquer de dénaturer les informations qu'elles contiennent en les formalisant, mais également un risque pouvant entrainer un biais dans l'interprétation par les juges de ce qu'ont voulu exprimer les sujets.

Partie 3

Résultats de l’expérience

Chapitre 6 – Résultats et perspectives

Nous présentons ici le prétraitement statistique effectué sur les résultats avant de les analyser.

6.1 Traitement statistique

Avant de déterminer si les résultats obtenus étaient significativement différents d'une répartition au hasard, nous avons d'abord extrait des matrices de confusions des fiches-réponses des 15 juges sous excel[4]. Nous en avons obtenus 18, soit 3 par sujet (une pour la condition audio seul, une pour la vidéo seule et une pour l'audiovisuelle). A partir de ces dernières, nous avons effectué des tests de Khi-deux afin de déterminer si la répartition des réponses données à un même couple de stimuli par les 15 juges était significative. Ensuite, nous avons déterminé quelles valeurs étaient significatives dans chaque matrice au moyen du logiciel R, en utilisant la fonction prop.test(i,30,x) avec i de 1 à 30 (nombre possible de réponses) et x le seuil du hasard, de 14% pour le sujet R, 17% pour les sujets S et M, 20% pour les sujets T et N, et 25% pour le sujet J. Nous obtenons ainsi un intervalle de significativité propre à chaque matrice (pour $p<0.05$).

Seuil du hasard	Intervalle de significativité
14%	9 à 30 (>30%)
17%	10 à 30 (>33%)
20%	11 à 30 (>37%)
25%	13 à 30 (>43%)

Figure 11 : Intervalle de significativité en fonction du seuil de hasard pour l'analyse des matrices de confusions.

[4] Annexe 4

6.2 Analyse des résultats

6.2.1 Sujet T

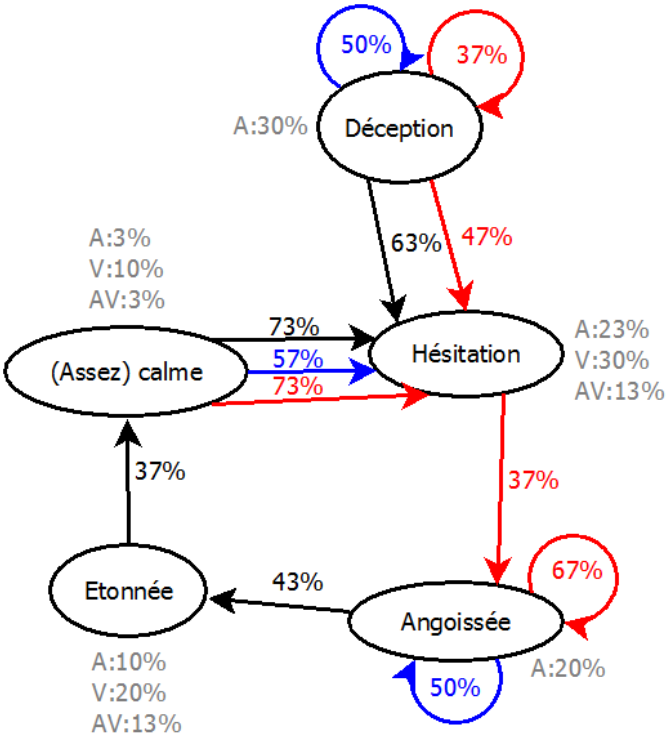

Figure 12 : Représentation graphique des réponses significatives A (noir), V (bleu), AV (rouge) et non significatives (gris) du sujet T.

Annotation	Stimuli	Nature Phonétique
Etonnée	TIP01	mmh
	TIP02	"hello"
Déception	TIP03	CVC_mheu_fff
	TIP04	euh_fff_chuchote_VC
(Assez) calme	TIP05	mmh_C
	TIP06	euh_+_exp_+_click
Angoissée	TIP07	inspiration_tremblante
	TIP08	mmh
Hésitation	TIP09	exp_avec_contrôle
	TIP10	occlusion

Figure 13 : Tableau récapitulatif des stimuli du sujet T

Le sujet T présente des résultats intéressants. En effet, on note que les stimuli tombant sous les étiquettes assez calme et déception sont des objets de type « euh » (interjection « euh », nasales « mmm » ou « nnn »). On remarque une confusion très forte avec l'étiquette hésitation, ce qui nous prouve qu'hors contexte, ce type d'objet renvoie spontanément à l'hésitation. Pour la déception, on remarque également qu'en condition

vidéo seule, on obtient un taux de reconnaissance acceptable de l'étiquette, et donc que l'audio vient perturber le jugement des juges pour les objets de type « euh ».

Autre résultat notable pour le sujet T, celui de l'étiquette angoissée, qui est bien identifié en vidéo, mal en audio, et très bien en audio-visuelle. Il semble ici que l'angoisse passe d'abord par le visuel mais se confirme avec l'appui de l'audio.

6.2.2 Sujet S

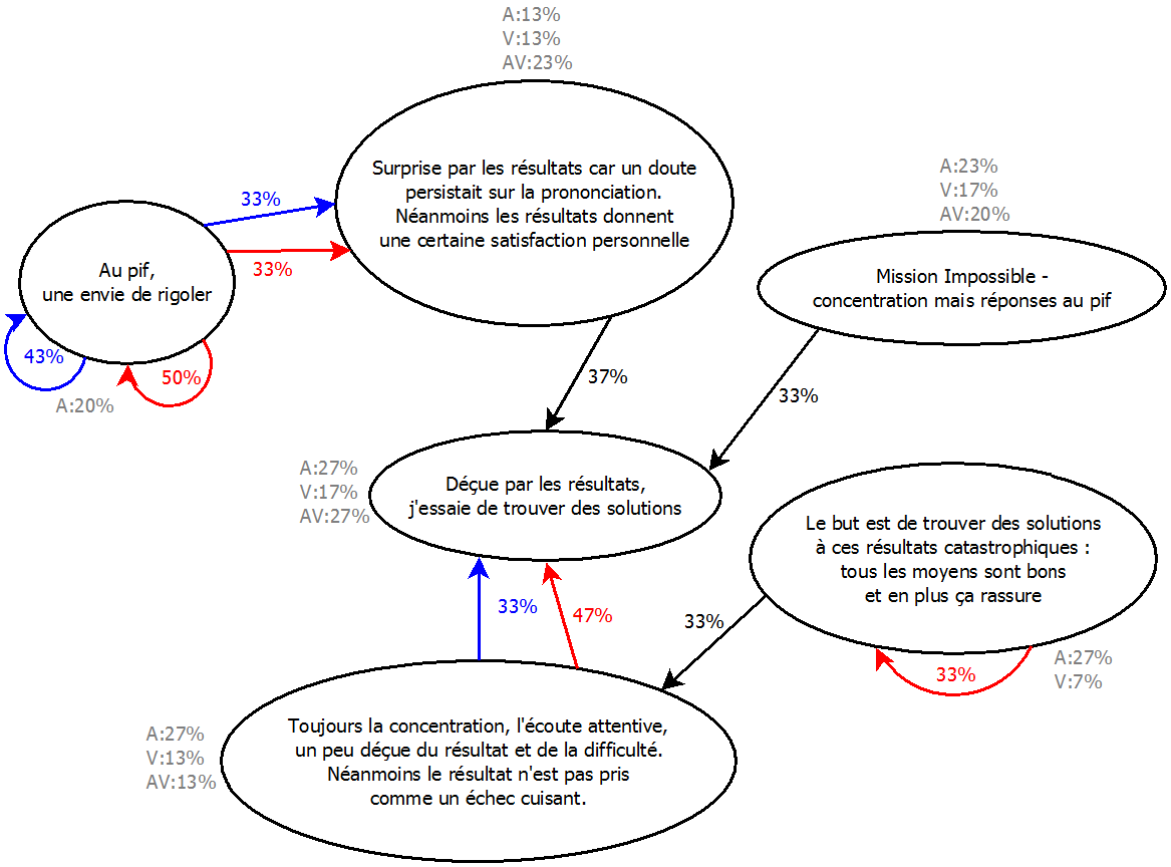

Figure 14 : Représentation graphique des réponses significatives A (noir), V (bleu), AV (rouge) et non significatives (gris) du sujet S.

Annotation	Stimuli	Nature Phonétique
Surprise par les résultats car un doute persistait sur la prononciation. Néanmoins les résultats donnent une certaine satisfaction personnelle	SAB01	occlusion-ingressive_inspiration
	SAB02	inspiration_(en_aspirant_sa_salive)
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.	SAB03	euh_fff_VC
	SAB04	ben_mmh_creaky_Comb-variante
Mission Impossible – concentration mais réponses au pif	SAB05	occlusion
	SAB06	inspiration_en_aspirant_sa_salive
Déçue par les résultats, j'essaie de trouver des solutions	SAB07	inspiration_en_aspirant_sa_salive
	SAB08	et ben euh Comb
Au pif, une envie de rigoler	SAB09	occlusion_ingressive_-_inspiration
	SAB10	Exp brutale continue
Le but est de trouver des solutions à ces résultats catastrophiques : tous les moyens sont bons et en plus ça rassure	SAB11	inspiration_+_exp
	SAB12	euh_mmh_VC

Figure 15 : Tableau récapitulatif des stimuli du sujet S

Pour le sujet S, c'est l'étiquette au pif, une envie de rigoler qui présente les résultats les plus marquants. En effet, les deux stimuli présentés appartenant à cette catégorie ne

sont pas du même type. L'un est une inspiration (plus de confusion), l'autre une expiration (meilleure reconnaissance).

On constate le report de « toujours la concentration » sur « déçue », présent en visuel et renforcé par l'ajout de l'audio. On notera également le report audio de « mission impossible », qui contient aussi le mot concentration dans son intitulé, sur « déçue ».

L'étiquette « le but est de trouver des solutions », typiquement appartenant à ce que nous définissons comme du FoT, est reporté sur « toujours la concentration » (qui contient un mot du registre émotionnel dans son intitulé : déçue) en audio seul alors que l'ajout de la dimension visuelle semble permettre aux juges de mieux discriminer les stimuli appartenant à cette étiquette.

6.2.3 Sujet M

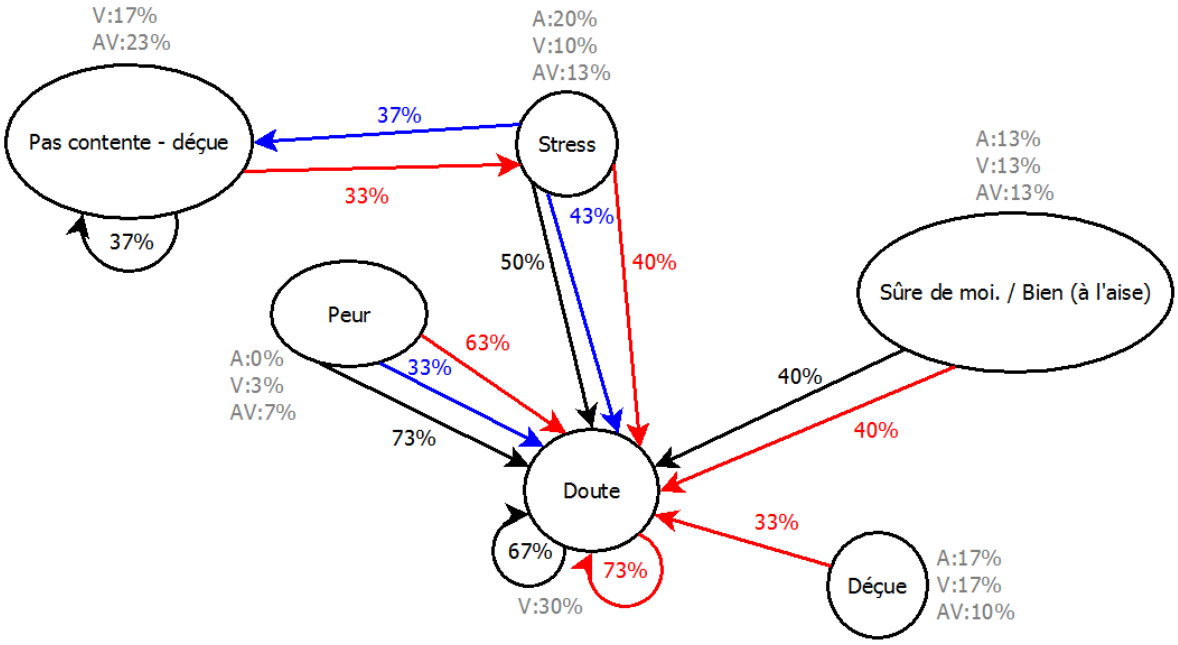

Figure 16 : Représentation graphique des réponses significatives A (noir), V (bleu), AV (rouge) et non significatives (gris) du sujet M.

Annotation	Stimuli	Nature Phonétique
Peur	MAR01_MAR02	V_"euh"
Stress	MAR03	V__"euh"
	MAR04	occlusion_ingressive_-_inspiration
Sûre de moi. / Bien (à l'aise)	MAR05	click
	MAR06	VC__"euh_mmh"
Doute	MAR07	VC_"euh_mmh"
	MAR08	V_"euh"
Pas contente - déçue	MAR09	V_"ouh"_+_prosodie!!
	MAR10	click_+_inspiration
Déçue	MAR11	friction_inspiree_mouillee
	MAR12	occlusion_ingressive_+_inspiration

Figure 17 : Tableau récapitulatif des stimuli du sujet M

Pour le sujet M, l'étiquette « pas contente - déçue » comprend 2 stimuli de nature différente. Une interjection type « ouh » bien reconnue en audio seul, et un click labiodental identifié comme du stress. Fait à souligner, le report de « stress » sur « pas contente » en vidéo seule est majoritairement présent lors de la présentation d'un stimulus contenant un clic labial.

Les quatre reports sur « doute » sont composés d'un stimulus type « euh » pour le stress et pour l'étiquette « sûre de moi », de deux stimuli type « euh » pour la peur

(identiques, nous rappelons que c’est celui-là qui a été dupliqué pour les besoins de l’expérience) et de deux clics labiaux pour « déçue ». Les deux stimuli présentés pour « doute » sont également des objets de type « euh ». On retrouve ici un effet lexical de l’interjection sur le choix des juges, l’hésitation n’étant pas un choix possible, le doute semble être un report logique.

6.2.4 Sujet J

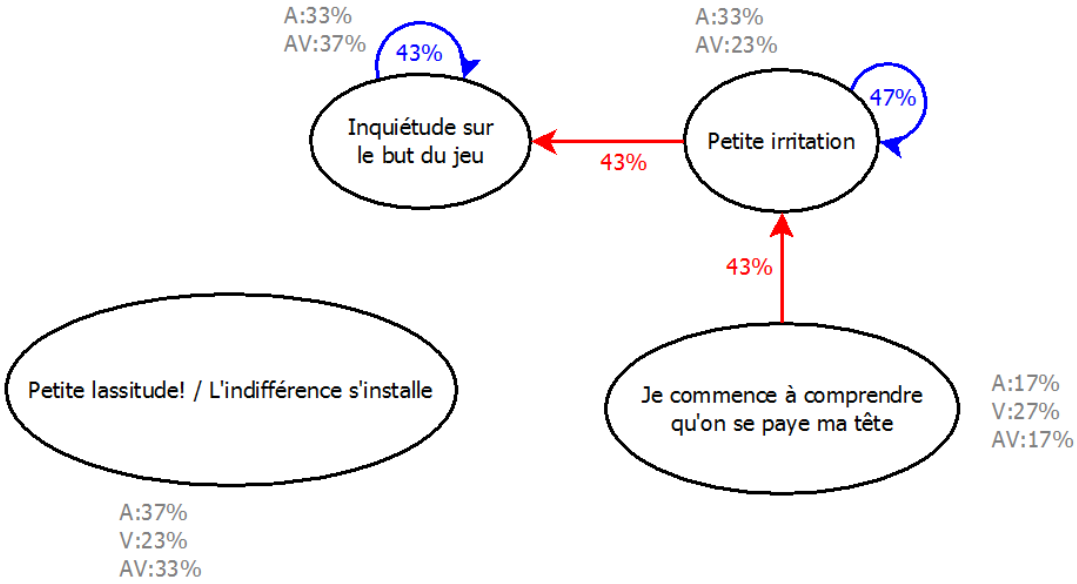

Figure 18 : Représentation graphique des réponses significatives A (noir), V (bleu), AV (rouge) et non significatives (gris) du sujet J.

Annotation	Stimuli	Nature Phonétique
Inquiétude sur le but du jeu	JEA01	inspiration
	JEA02	inspiration_(avec_controle)
Petite irritation	JEA03	VC_[Em]
	JEA04	click_+_inspiration
Petite lassitude! / L'indifférence s'installe	JEA05	relache_sa_respiration_+_interaction_langue-levres
	JEA06	inspiration
Je commence à comprendre qu'on se paye ma tête	JEA07	click_+_inspiration
	JEA08	inspiration

Figure 19 : Tableau récapitulatif des stimuli du sujet J

Ici, on observe un report intéressant de l'étiquette on se paye ma tête, typiquement du FoT, sur l'irritation, qui semble du coup une confusion logique hors contexte. La réaction normalement attendue d'un individu dont on se moque est qu'il s'énerve, ou du moins qu'il se sente irrité.

L'irritation est bien identifiée en vidéo seule, mais encore une fois on observe une perturbation de l'audio sur la vidéo avec un report sur l'inquiétude, cette dernière étant bien reconnue en vidéo seule uniquement.

6.2.5 Sujet N

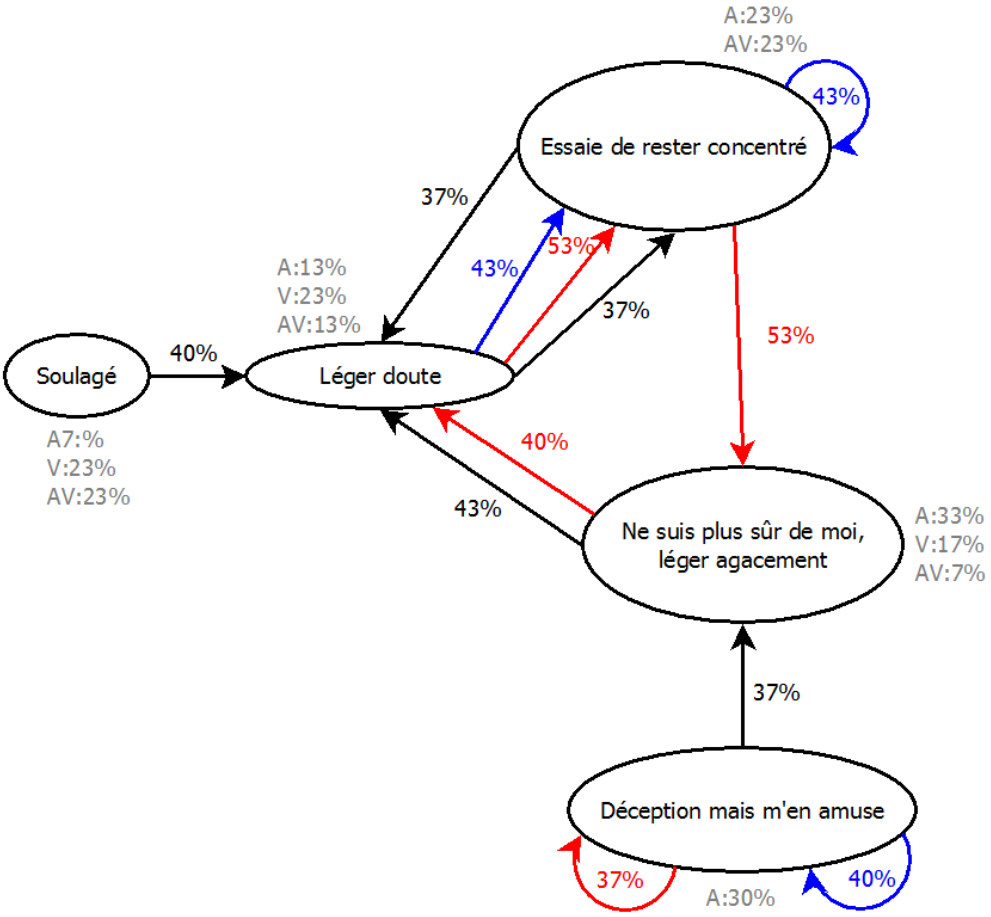

Figure 20 : Représentation graphique des réponses significatives A (noir), V (bleu), AV (rouge) et non significatives (gris) du sujet N.

Annotation	Stimuli	Nature Phonétique
Léger doute	NIC01	insp
	NIC02	friction ingressive
Ne suis plus sur de moi, léger agacement	NIC03	faible_expiration_voisee
	NIC04	V_"euh"
Essaie de rester concentré	NIC05	expiration_avec_contrôle
	NIC06	occlusion
Déception mais m'en amuse	NIC07	inspiration_controle++_prosodie
	NIC08	"pfff"
Soulagé	NIC09	VC_"euh_mmh"
	NIC10	CV-variante__"ben"_creaky

Figure 21 : Tableau récapitulatif des stimuli du sujet N

Pour le sujet N, on note une bonne reconnaissance de « déception mais m'en amuse » en condition vidéo seule, composé de deux stimuli à la nature assez proche : un

rire et une amorce de rire (relâchement nasal), et l'audio vient perturber ce jugement puisqu'il se reporte sur « ne suis plus sûr de moi » en audio seul.

La concentration est également bien reconnue en vidéo seule, alors que l'ajout du paramètre audio entraine une forte confusion avec « ne suis plus sûr de moi ». En audio seul la confusion se fait pourtant avec le doute.

Pour l'étiquette « léger doute », le stimulus le plus reporté sur la concentration est une inspiration longue suivie d'une expiration.

Les deux stimuli de « ne suis plus sûr de moi », reportés sur la déception, sont comme pour le sujet M, deux objets de type « euh ». On observe une certaine cohérence de la part des juges dans le choix de leurs réponses. De même, les deux stimuli de « soulagé » sont des objets de type « euh » qui se reportent sur le doute en audio seul.

6.2.6 Sujet R

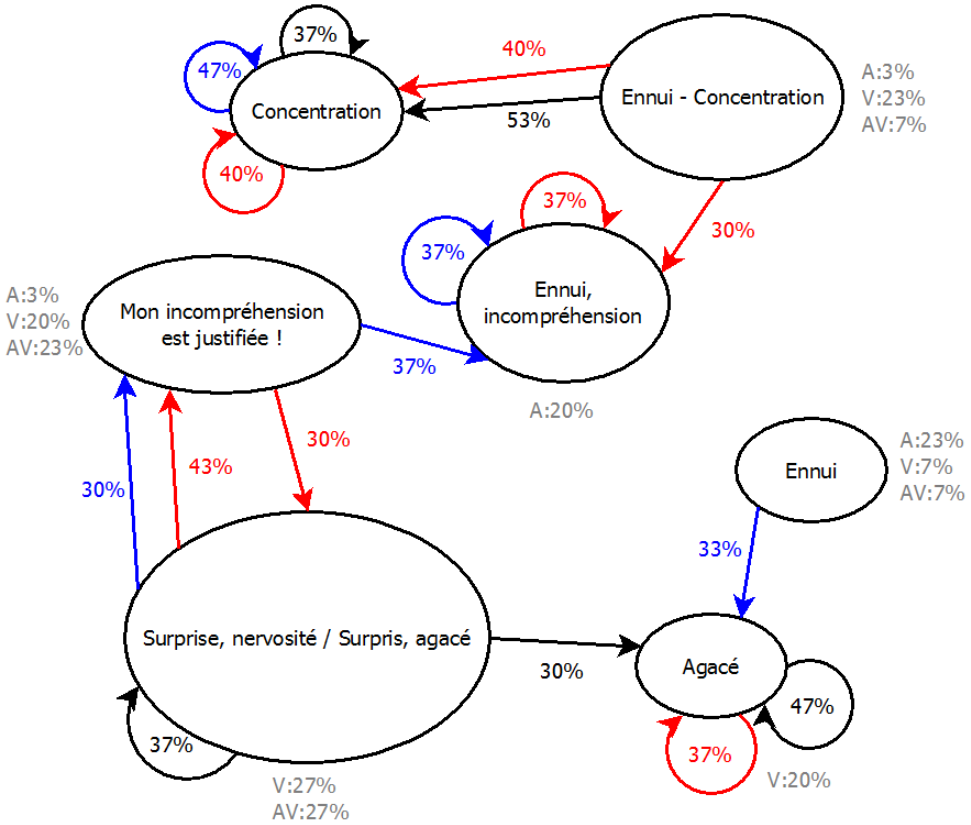

Figure 22 : Représentation graphique des réponses significatives A (noir), V (bleu), AV (rouge) et non significatives (gris) du sujet R.

Annotation	Stimuli	Nature Phonétique
Concentration	ROB01	LL
	ROB02	série_BB_insp_expr_click
Ennui - Concentration	ROB03	V_??
	ROB04	C-variante_"mmh"_creaky_+_exp
Ennui	ROB05	V_??_(alors_soupire)
	ROB06	V-variante_[ewe]_soupire
Ennui, incompréhension	ROB07	C-variante_"pff(ou)"_chuchote
	ROB08	C_"mmh"_+_friction_(trill_bilabial)
Surprise, nervosité / Surpris, agacé	ROB09	Comb_"ouah_ha_ha"
	ROB10	C-variante_"fff"_soupire
Agacé	ROB11	Comb-variante_"bah_pfff(ou)"_chuchote
	ROB12	VC_"euh_fff"
Mon incompréhension est justifiée !	ROB13	avale_+_raclement+_mmmh_insp
	ROB14	exp

Figure 23 : Tableau récapitulatif des stimuli du sujet R

Le sujet R comportait beaucoup d'étiquettes dont l'intitulé contenait au moins un mot en commun avec une autre étiquette. On observe ici une très bonne reconnaissance

dans toutes les modalités de la concentration. Des deux stimuli de l'étiquette « ennui - concentration », l'un est un objet de type « euh » (nasal « mmm » longue) qui s'est reporté sur la concentration, l'autre est un objet de type « euh » également (nasal dans une expiration forte et brève) qui s'est reporté sur « ennui - incompréhension ».

Les deux objets de l'étiquette « ennui - incompréhension » sont de type plosive + fricative, et ont bien été reconnus en vidéo seule, sans dégradation avec l'ajout de l'audio.

Les stimuli de l'étiquette « ennui » ont une nature visuelle différente. En effet, l'un est figé, l'autre est en mouvement de balancement sur l'axe gauche/droite du corps. Le deuxième est majoritairement reporté en vidéo seule sur « agacé ».

Les stimuli de l'étiquette « agacé » sont bien perçus en audio seul, mais sont visiblement dégradés par l'ajout de la vidéo. Les deux stimuli contiennent une expiration forte. Le stimulus reporté à l'audio sur « agacé » est de même nature, alors que le stimulus bien identifié de « surprise, nervosité / surpris, agacé » contient une amorce de rire nerveux. C'est ce dernier qui se reporte le plus sur « mon incompréhension est justifiée ! » lorsque la dimension vidéo est présente seule ou couplée avec l'audio.

Les reports de « mon incompréhension est justifiée ! » se font sur « ennui - incompréhension » pour le stimulus où le sujet pose sa main sur son menton, sourit à l'envers et baisse les yeux en condition vidéo seule, et sur « surprise, nervosité / surpris, agacé » pour le stimulus où le sujet fronce les sourcils, relâche sa mâchoire inférieure et réalise une inspiration forte et brève en condition audio-visuelle.

6.3 Conclusion sur les résultats

Tout d'abord, un petit récapitulatif comparatif sur la confusion et l'identification de certaines étiquettes qui méritent de retenir notre attention.

Pour les sujets T, M et N, on observe un report très importants des objets de type « euh » vers les étiquettes attractives « hésitation », « doute » et « léger doute » en audio seul, avec des scores différents d'un sujet à l'autre : très fort pour le sujet T (63% et 73%), moyen pour le sujet M (73%, 50%et 40%), plus faible pour le sujet N (40% et 43%).

Pour les sujets T et N, on observe un comportement semblable des étiquettes « déception » et « déception mais m'en amuse » quand on passe de la modalité vidéo seule à la modalité audio-visuelle : on a une chute du score de reconnaissance avec l'ajout de la

dimension audio plus forte pour le sujet T (de 50% à 37%) que pour le sujet N (de 40% à 37%).

On notera ensuite que tout ce qui est issu d'un processus conscient, à savoir les attitudes et les états mentaux, semble se baser principalement sur le visuel. Au contraire, les émotions, issues d'un processus inné non contrôlé, sont bien reconnues en audio. Il semble alors se dessiner un profil intéressant pour la perception des micro-événements vocaux. Tout ce qu'on met dans le FoT se voit avant de s'entendre, tout ce qui ne tombe pas dedans s'entend sans se voir. Dans les deux cas, le manque de contexte semble entrainer des confusions, ce qui tend à prouver non seulement que les informations audio et visuelles ne sont pas additives au sens propre du terme, mais également qu'il n'est pas possible d'extraire des fonctions primitives indépendantes du contexte d'énonciation.

6.4 Perspectives

A court terme, il serait opportun d'analyser en détail la confiance que les juges ont portée à leurs réponses grâce au degré de confiance présent dans le test. En effet, se pencher sur l'aspect inné/acquis du processus de reconnaissance du FoT et des émotions permettrait de renforcer l'opinion que nous nous faisons des résultats. Il serait aussi intéressant de faire une analyse par stimulus plutôt que par annotation. On pourra également réaliser une analyse en regroupant les stimuli par nature phonétique cette fois. Enfin, on pourra faire des méta-classes d'annotations pour observer les spécificités propres à chaque modalité A, V, et AV par classe d'annotation ou par sujet.

A plus long terme, c'est l'aspect interculturel qui va nous intéresser. L'expérience conduite dans ce mémoire doit être reconduite avec des juges originaires d'autres pays. Une comparaison avec une culture très différente de la notre (nous projetons entre autre de la réaliser avec le Japon) permettrait de rechercher s'il existe des constantes dans le type de confusion effectué par les juges, ou au contraire si un facteur culturel donne lieu à une perception différente que ce soit pour les attitudes ou pour les émotions.

On mesure bien l'importance de ce travail lorsque le moment sera venu de faire évoluer les ACA vers un modèle au plus près de l'humain. L'étude de l'inter-culturalité aura un intérêt tout particulier dans l'optique de permettre une expressivité émotive sans faille de l'ACA, quel que soit le contexte culturel et la langue utilisés dans l'interaction homme machine.

Conclusion

Nous avons montré à travers cette étude que l'humain extrait du sens dans des objets qui ne semblent à priori pas être une forme de langage au sens évolué du terme. Conscients ou inconscients, innés ou acquis, tous ces micro-événements audibles et/ou visibles contribuent à transmettre une information plus précise à notre interlocuteur.

On relèvera également que les informations audio et visuelles ne sont pas additives, puisque parfois on a même une dégradation du score de reconnaissance quand on ajoute la dimension manquante (c'est à dire quand on ajoute A à V ou V à A).

Il faudra encore du temps et de nombreuses expériences pour arriver à envisager une modélisation réaliste des émotions, des attitudes et des états mentaux, mais c'est un travail qui mérite d'être poursuivi. La dimension interculturelle que nous souhaiterions ajouter à cette étude devrait permettre de mettre en lumière d'autres vérités. Il restera ensuite à explorer le vaste champ du contexte d'énonciation, afin de confirmer ou d'infirmer la dépendance de ce dernier vis-à-vis des micro-événements audiovisuels observés, ainsi que celui de la rythmicité des événements, c'est-à-dire la façon dont ils s'organisent dans le temps, notion qui semble être tout aussi pertinente.

Bibliographie

Aristote (1990). *Éthique à Nicomaque*. Vrin, Paris.

Aubergé, V. (2002). *A gestalt morphology of prosody directed by functions : the example of a step by step model developed at ICP*. Speech Prosody 2002, Aix-en-Provence, France, 151–155.

Averill, J. (1980*). A constractivist view of emotion*. In Plutchik, R. et Kellerman, H., éditeurs : Emotion : Theory, research and experience, volume 1, pages 305–339. Academic Press, New York, USA.

Averill, J. (1998). *What are emotions, really ?* Cognition & Emotion, 12(6):849–855.

Buisine, S., Hartmann, B., Mancini, M. et Pelachaud, C. (2006). *Conception et évaluation d'un modèle d'expressivité pour les gestes des agents conversationnels*. Revue d'intelligence artificielle, 20(4-5):621–638.

Campbell, N. (2004). *Getting to the heart of the matter: speech as the expression of affect; rather than just text or language*. In *languages resources and evaluation* (Vol. 39, pp. 109-118).

Cosnier, J. (1996). *Les gestes du dialogue*. Psychologie de la motivation, 21:129–138.

Damasio, A. (1995). *L'erreur de Descartes : La raison des émotions*. Odile Jacob, Paris.

Damasio, A. (2003). *Spinoza avait raison : joie et tristesse, le cerveau des émotions*. Odile Jacob, Paris.

Darwin, C. (1872). *The expression of the emotions in man and animals*. John Murray, London, UK.

Descartes, R. (1649). *Les passions de l'âme*. Le Livre de Poche (1 octobre 1990). Classiques de la philosophie.

Ekman, P. (1973). *Cross-cultural studies of facial expression*. In Ekman, P., éditeur : Darwin and the facial expression, pages 169–222. Academic Press Inc.

Ekman, P. (2009). *Darwin's contributions to our understanding of emotional expressions*. Philosophical Transactions of the Royal Society B : Biological Sciences, 364(1535):3449–3451.

Ekman, P. et Friesen, W. (1969). *The Repertoire Of Nonverbal Behavior : Categories, Origins, Usage and Coding*. Semiotica, 1:49–98.

Hart, J. (1965). *Memory and the feeling-of-knowing experience*. Journal of Educational Psychology, 56:208–216.

Jakobson, R. (1963). *Closing statement : Linguistics and Poetics*. In *style in language*. T. A. Sebeok. New York.

James, W. (1884). *What Is An Emotion ?* Mind, 9(34):188–205.

Kendon, A. (2002). *Some uses of the heads shake*. Gesture, 2(2):147–182.

La Fontaine, J (1668-1678). *Epilogue*. In *Fables de la Fontaine*, Livre XI.

Lange, C. (1884). *La théorie de l'émotion*. L'Harmattan.

Lasswell, H. (1948). *Structures et fonctions de la communication dans la société*. in Balle Francis, Padioleau Jean-Gustave, dir., Sociologie de l'information, Hachette, 1973.

Lazarus, R. (1966). *Psychological Stress And The Coping Process*. McGraw Hill, New York, USA.

Loyau, F. (2007). *Expressions des états mentaux et émotionnels de l'humain en interaction : ébauches du « Feeling of Thinking »*. Thése en Sciences Cognitives, Spécialité Ingénierie pour le vivant : santé, cognition,

environnement. Département de Parole et Cognition du laboratoire GIPSA-Lab, Institut National Polythecnique, Grenoble, France.

Loyau F. et Aubergé, V. (2006). *Expressions outside the talk turn : ethograms of the feeling of thinking*. In 5th LREC, pages 47–50.

McNeill, D. (1992). *Hand and Mind*. University of Chicago Press, Chicago, USA.

Nhouyvanisvong, A. et Reder, L. M. (1998). *Rapid feeling-of-knowing : a strategy selection mechanism*. In Yzerbyt, V., Lories, G. et Dardenne, B., éditeurs : Metacognition : Cognitive and social dimensions, pages 35–52. Sage Publications Inc., London, Grande-Bretagne.

Reder, L. M. et Ritter, F. (1992). *What determines initial feeling of knowing ? familiarity with questions terms, not with the answer*. Journal of Experimental Psychology : Learning, Memory, and Cognition, 18(3):435–451.

Sander, D. et Scherer, K. (2009). *Traité de psychologie des émotions*. Dunod, Paris.

Sansonnet, J.-P., Leguern, K. et Martin, J.-C. (2005). *Une architecture médiateur pour des agents conversationnels animés*. Workshop WACA'*01*.

Sartre, J. (1930). *Esquisse d'une théorie des émotions*. Hermann, Paris, pages 43–66.

Scherer, K. (1994). *Affect bursts*. In van Goozen, S., van de Poll, N. et Sergeant, J., éditeurs : Emotions : Essays on emotion theory, pages 161–193. Lawrence Erlbaum Associates, Inc, Hillsdale, USA.

Shannon, C. et Weaver, W. (1949). *Théorie mathématique de la communication*. C.E.P.L. (1975)

Signorello, R. (2010). *La parole ne s'arrete pas aux mots : a la recherche d'indices de culture dans le backchannel audio-visuel*. Mémoire de Master 2 en Sciences du Langage, Université Stendhal Grenoble III – GIPSA-Lab, Grenoble, France.

Spinoza, B. (1677). *L'Ethique*. Gallimard (13 janvier 1994). Folio.

Swert, M. et Krahmer, E. (2005). *Audiovisual prosody and feeling of knowing*. Journal of Memory and Language, 53(1):81–94.

Vanpé, A. (2006). *Ethogrammes d'interaction personne-machine émotionnellement induites : premiers pas vers un modèle du "feeling of thinking"*. Master, Université Stendhal Grenoble III - ICP : Institut de la communication parlée, Grenoble, France.

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"*. Thèse en Langues, Littératures et Sciences Humaines. Laboratoire GIPSA-Lab, Grenoble, France.

Vanpé, A. et Aubergé, V. (2010). *Micro-gestures, vocal bursts and other non-lexical speech sounds in expressive communication*.

Wierzbicka, A. (1992). *Talking about emotions : Semantics, culture, and cognition*. Cognition and Emotion, 6:285–319.

Wundt, W. (1900). *Völkerpsychologie, eine Untersuchung der Entwicklungsgesetze von Sprache, Mythus, und Sitte*. Leipzig.

Table des annexes

ANNEXE 1 TABLEAU RECAPITULATIF DES AUTO-ANNOTATIONS RETENUES POUR L'EXPERIENCE 56
ANNEXE 2 LISTE DES STIMULI RETENUS AVEC LEUR NATURE PHONETIQUE 58
ANNEXE 3 EXTRAIT DU CODE SOURCE DE L'EXPERIENCE 60
ANNEXE 4 MATRICES DE CONFUSION EN EFFECTIF ET POURCENTAGE DES REPONSES DES JUGES AUX STIMULI DES DIFFERENTS SUJETS 67

Annexe 1
Tableau récapitulatif des auto-annotations retenues pour l'expérience

	Classe	Sujet T	Sujet S	Sujet M	Sujet J	Sujet N	Sujet R
1	Surprise	Etonnée	Surprise par les résultats car un doute persistait sur la prononciation. néanmoins les résultats donnent une certaine satisfaction personnelle				
2	Déception	Déception	Déçue par les résultats, j'essaie de trouver des solutions	Déçue			
3	Calme	(Assez) calme					
4	Stress	Angoissée		Stress	Inquiétude sur le but du jeu		
5	Hésitation	Hésitation					
6	Doute		Mission impossible - concentration mais réponses au pif	Doute		Léger doute	
7	Concentration		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			Essaie de rester concentré	Concentration
8	FoT		Le but est de trouver des solutions à ces résultats catastrophiques : tous les moyens sont bons et en plus ça rassure		Je commence à comprendre qu'on se paye ma tête		Mon incompréhension est justifiée !
9	Amusement		Au pif, une envie de rigoler				
10	Irritation				Petite irritation		Agacé

11	Lassitude/Ennui				Petite lassitude ! / L’indifférence s’installe		Ennui
13	Soulagement					Soulagé	
14	Peur			Peur			
15	Contentement			Sûre de moi / Bien (à l'aise)			
6.10	Doute+Irritation					Ne suis plus sûr de moi, léger agacement	
2.9	Déception+Amusement					Déception mais m'en amuse	
7.11	Concentration+Lassitude/Ennui						Ennui - Concentration
11.6	Lassitude/Ennui+Doute						Ennui, incompréhension
1.10	Surprise+Irritation						Surpris, agacé / Surprise, nervosité
2.10	Déception+Irritation			Pas contente - déçue			

Annexe 2
Liste des stimuli retenus avec leur nature phonétique

Stimuli	Nature Phonétique
TIP01	mmh
TIP02	"hello"
TIP03	CVC_mheu_fff
TIP04	euh_fff_chuchote_VC
TIP05	mmh_C
TIP06	euh_+_exp_+_click
TIP07	inspiration_tremblante
TIP08	mmh
TIP09	exp_avec_contrôle
TIP10	occlusion
SAB01	occlusion-ingressive_inspiration
SAB02	inspiration_(en_aspirant_sa_salive)
SAB03	euh_fff_VC
SAB04	ben_mmh_creaky_Comb-variante
SAB05	occlusion
SAB06	inspiration_en_aspirant_sa_salive
SAB07	inspiration_en_aspirant_sa_salive
SAB08	et ben euh Comb
SAB09	occlusion_ingressive_-_inspiration
SAB10	Exp brutale continue
SAB11	inspiration_+_exp
SAB12	euh_mmh_VC
MAR01_MAR02	V_"euh"
MAR03	V__"euh"
MAR04	occlusion_ingressive_-_inspiration
MAR05	click
MAR06	VC__"euh_mmh"
MAR07	VC_"euh_mmh"
MAR08	V_"euh"
MAR09	V_"ouh"_+_prosodie!!
MAR10	click_+_inspiration
MAR11	friction_inspiree_mouillee
MAR12	occlusion_ingressive_+_inspiration
JEA01	inspiration
JEA02	inspiration_(avec_controle)
JEA03	VC_[Em]
JEA04	click_+_inspiration
JEA05	relache_sa_respiration_+_interaction_langue-levres

JEA06	inspiration
JEA07	click_+_inspiration
JEA08	inspiration
NIC01	insp
NIC02	friction ingressive
NIC03	faible_expiration_voisee
NIC04	V_"euh"
NIC05	expiration_avec_contrôle
NIC06	occlusion
NIC07	inspiration_controle++_prosodie
NIC08	"pfff"
NIC09	VC_"euh_mmh"
NIC10	CV-variante__"ben"_creaky
ROB01	LL
ROB02	série_BB_insp_expr_click
ROB03	V_??
ROB04	C-variante_"mmh"_creaky_+_exp
ROB05	V_??_(alors_soupire)
ROB06	V-variante_[ewe]_soupire
ROB07	C-variante_"pff(ou)"_chuchote
ROB08	C_"mmh"_+_friction_(trill_bilabial)
ROB09	Comb_"ouah_ha_ha"
ROB10	C-variante_"fff"_soupire
ROB11	Comb-variante_"bah_pfff(ou)"_chuchote
ROB12	VC_"euh_fff"
ROB13	avale_+_raclement+_mmmh_insp
ROB14	exp

Annexe 3
Extrait du code source de l'expérience

Accueil, Initialisation :

```
on mouseUp
  global idSujet, numSujet, sexeSujet, nationSujet, repCourant, ficheRep, fichStimuli, listeStimuli, nbStimuli,
numStimuliCourant, fichRep

  -- on lit la liste des simuli dans le fichier des stimuli
  answer file "Selectionnez le fichier de stimuli, svp." -- with filter "*.txt"
  put it into fichStimuli
  open file fichStimuli
  read from file fichStimuli until EOF
  put it into listeStimuli
  close file fichStimuli


 -- on designe comme dossier par defaut le dossier contenant la liste des stimuli
  set the itemDel to "/"
  put the number of items in fichStimuli into nb
  put 1 into i
  put "" into repCourant
  repeat for nb-1 times
    put item i of fichStimuli after repCourant
    put "/" after repCourant
    put i+1 into i
  end repeat
  set the defaultFolder to repCourant

  -- on demande les infos sur le sujet
  ask "Donnez votre identifiant (01 à 30) :"
  put it into numSujet
  ask "Indiquez votre sexe (M ou F) :"
  put it into sexeSujet
  ask "Indiquez votre nationalité (FR ou JP) :"
  put it into nationSujet
  put numSujet & "_" & sexeSujet & "_"  & nationSujet into idSujet


  put repCourant & ".." & "/resultats/" & idSujet & ".txt" into fichRep
  -- on ouvre le fichier de resultats
  open file fichRep
  write "Test M2, resultats de " & idSujet & return to file fichRep
  write "Identifiant" & tab & "Sexe" & tab & "Nationalité" & tab & "Stimulus" & tab & "Locuteur" & tab &
"Condition" & tab & "Réponse" & tab & "Confiance" & return to file fichRep

end mouseUp
```

Accueil, Commencer :

```
on mouseUp
  global listeStimuli, nbStimuli, stimuliCourant, numStimuliCourant, fichRep
  global idSujet, numSujet, sexeSujet, nationSujet, repCourant, fichRep, fichStimuli, condition

    -- initialisation des variables
```

```
  set the itemDelimiter to return
  put the number of items of listeStimuli into nbStimuli
  put 1 into numStimuliCourant
  get item numStimuliCourant of listeStimuli
  put it into stimuliCourant
  put numStimuliCourant+1 into numStimuliCourant


    -- on affiche les consignes specifiques de la condition
  put stimuliCourant into condition
  if (stimuliCourant = "A")
  then put "Vous allez maintenant entendre uniquement la voix de la personne. Cliquez sur le bouton 'Suivant'
pour continuer." into field id 1022
  if (stimuliCourant = "V")
  then put "Vous allez maintenant uniquement voir la personne. Cliquez sur le bouton 'Suivant' pour
continuer." into field id 1022
  if (stimuliCourant = "AV")
  then put "Vous allez maintenant voir et entendre la personne. Cliquez sur le bouton 'Suivant' pour
continuer." into field id 1022

  go to card id 1021
  set the visible of field id 1022 to true
  set the visible of button id 1023 to true

end mouseUp
```

Changement, Suivant :

```
on mouseUp
  global repCourant, listeStimuli, nbStimuli, numStimuliCourant, stimuliCourant, condition, fichRep
  global idSujet, sexeSujet, nationSujet, stimuliCourant, locuteur, condition, reponse, confiance
global longueurVideo, rapport, dureeVideo

  -- on récupère le stimuli suivant
  set the itemDelimiter to return
  get item numStimuliCourant of listeStimuli
  put it into stimuliCourant
  put numStimuliCourant+1 into numStimuliCourant

  -- On met à jour l'affichage
  set the visible of field id 1022 to false
  set the visible of button id 1023 to false

  -- on va vers la carte du sujet correspondant
  if (stimuliCourant = "Suj1")
  then
  get item numStimuliCourant of listeStimuli
  put it into stimuliCourant
  put numStimuliCourant+1 into numStimuliCourant
  go to card id 1003
  --on met à jour l'affichage de la carte
  set the visible of button id 1004 to true
  set the visible of button id 1005 to true
  set the visible of button id 1006 to true
  set the visible of button id 1007 to true
  set the visible of button id 1008 to true
  set the visible of field id 1020 to true
  set the visible of scrollbar id 1010 to true
  set the visible of button id 1009 to true
  end if
```

```
if (stimuliCourant = "Suj2")
then
get item numStimuliCourant of listeStimuli
put it into stimuliCourant
put numStimuliCourant+1 into numStimuliCourant
go to card id 1024
--on met à jour l'affichage de la carte
set the visible of button id 1025 to true
set the visible of button id 1026 to true
set the visible of button id 1027 to true
set the visible of button id 1028 to true
set the visible of button id 1029 to true
set the visible of button id 1030 to true
set the visible of field id 1033 to true
set the visible of scrollbar id 1031 to true
set the visible of button id 1032 to true
end if
if (stimuliCourant = "Suj3")
then
get item numStimuliCourant of listeStimuli
put it into stimuliCourant
put numStimuliCourant+1 into numStimuliCourant
go to card id 1072
--on met à jour l'affichage de la carte
set the visible of button id 1073 to true
set the visible of button id 1074 to true
set the visible of button id 1075 to true
set the visible of button id 1076 to true
set the visible of button id 1077 to true
set the visible of button id 1078 to true
set the visible of field id 1081 to true
set the visible of scrollbar id 1079 to true
set the visible of button id 1080 to true
end if
if (stimuliCourant = "Suj4")
then
get item numStimuliCourant of listeStimuli
put it into stimuliCourant
put numStimuliCourant+1 into numStimuliCourant
go to card id 1082
--on met à jour l'affichage de la carte
set the visible of button id 1083 to true
set the visible of button id 1084 to true
set the visible of button id 1085 to true
set the visible of button id 1086 to true
set the visible of field id 1089 to true
set the visible of scrollbar id 1087 to true
set the visible of button id 1088 to true
end if
if (stimuliCourant = "Suj5")
then
get item numStimuliCourant of listeStimuli
put it into stimuliCourant
put numStimuliCourant+1 into numStimuliCourant
go to card id 1090
--on met à jour l'affichage de la carte
set the visible of button id 1091 to true
set the visible of button id 1092 to true
set the visible of button id 1093 to true
set the visible of button id 1094 to true
```

```
  set the visible of button id 1095 to true
  set the visible of field id 1098 to true
  set the visible of scrollbar id 1096 to true
  set the visible of button id 1097 to true
  end if
  if (stimuliCourant = "Suj6")
  then
  get item numStimuliCourant of listeStimuli
  put it into stimuliCourant
  put numStimuliCourant+1 into numStimuliCourant
  go to card id 1099
  --on met à jour l'affichage de la carte
  set the visible of button id 1100 to true
  set the visible of button id 1101 to true
  set the visible of button id 1102 to true
  set the visible of button id 1103 to true
  set the visible of button id 1104 to true
  set the visible of button id 1105 to true
  set the visible of button id 1106 to true
  set the visible of field id 1109 to true
  set the visible of scrollbar id 1107 to true
  set the visible of button id 1108 to true
  end if


  -- on lance la vidéo suivante
  create invisible player leJoueur
  -- on n'affiche pas la vidéo si condition audio seul
  if condition = "A"
  then
    set the location of player "leJoueur" to -1000,-1000
  else
    set the location of player "leJoueur" to 550,350
  end if
  -- on coupe le son si condition vidéo seule
  if condition = "V"
  then
    set the playLoudness of player "leJoueur" to 0
  else
    set the playLoudness of player "leJoueur" to 100
  end if
  -- on place la vidéo
  set the height of player "leJoueur" to 576
  set the width of player "leJoueur" to 784
  set the filename of player "leJoueur" to repCourant & ".." & "/stimuli/" & stimuliCourant

-- on calcule le temps du stimulus
  put the duration of player "leJoueur" into longueurVideo
  put the timeScale of player "leJoueur" into rapport
  put longueurVideo/rapport into dureeVideo

  -- on joue le stimulus
  start player "leJoueur"
  wait for dureeVideo + 1 sec
  delete player "leJoueur"
  wait for 0.5 sec

end mouseUp
```

Sujet T, Suivant (Chacun des six sujets fonctionne de la même manière) :

```
on mouseUp
  global repCourant, listeStimuli, nbStimuli, numStimuliCourant, stimuliCourant, condition, fichRep
  global idSujet, sexeSujet, nationSujet, stimuliCourant, locuteur, condition, reponse, confiance

  global longueurVideo, rapport, dureeVideo

  -- on enregistre les résultats
  put "Tiphaine" into locuteur
  if the hilite of button id 1004 is true
  then put "Etonnée" into reponse
  if the hilite of button id 1005 is true
  then put "Déception" into reponse
  if the hilite of button id 1006 is true
  then put "(Assez) calme" into reponse
  if the hilite of button id 1007 is true
  then put "Angoissée" into reponse
  if the hilite of button id 1008 is true
  then put "Hésitation" into reponse
  put the thumbPosition of scrollbar id 1010 into confiance
  write idSujet & tab & sexeSujet & tab & nationSujet & tab & stimuliCourant & tab & locuteur & tab &
condition & tab & reponse & tab & confiance & return to file fichRep

  -- on met à jour l'affichage
  unhilite button id 1004
  unhilite button id 1005
  unhilite button id 1006
  unhilite button id 1007
  unhilite button id 1008
  set the thumbPosition of scrollbar id 1010 to 0
  set the visible of button id 1004 to false
  set the visible of button id 1005 to false
  set the visible of button id 1006 to false
  set the visible of button id 1007 to false
  set the visible of button id 1008 to false
  set the visible of field id 1020 to false
  set the visible of scrollbar id 1010 to false
  set the visible of button id 1009 to false
  put empty into reponse

  -- on récupère le stimulus suivant
  set the itemDelimiter to return
  get item numStimuliCourant of listeStimuli
  put it into stimuliCourant

if (stimuliCourant = "A") or (stimuliCourant = "V")  or (stimuliCourant = "AV") or (stimuliCourant =
"Suj1") or (stimuliCourant = "Suj2") or (stimuliCourant = "Suj3") or (stimuliCourant = "Suj4") or
(stimuliCourant = "Suj5") or (stimuliCourant = "Suj6") or (stimuliCourant = "Fin")
then

  -- si on change de condition, on affiche les consignes specifiques de la condition
  if (stimuliCourant = "A") or (stimuliCourant = "V")  or (stimuliCourant = "AV")
  then put stimuliCourant into condition
  if (stimuliCourant = "A")
  then put "Vous allez maintenant entendre uniquement la voix de la personne. Cliquez sur le bouton 'Suivant'
pour continuer." into field id 1022
  if (stimuliCourant = "V")
  then put "Vous allez maintenant uniquement voir la personne. Cliquez sur le bouton 'Suivant' pour
continuer." into field id 1022
  if (stimuliCourant = "AV")
```

```
  then put "Vous allez maintenant voir et entendre la personne. Cliquez sur le bouton 'Suivant' pour
continuer." into field id 1022
  if (stimuliCourant = "A") or (stimuliCourant = "V")  or (stimuliCourant = "AV")
  then
  go to card id 1021
  set the visible of field id 1022 to true
  set the visible of button id 1023 to true
  put numStimuliCourant+1 into numStimuliCourant
  end if

  -- si on change de sujet, on va vers la carte changement de sujet
  if (stimuliCourant = "Suj1") or (stimuliCourant = "Suj2") or (stimuliCourant = "Suj3") or (stimuliCourant =
"Suj4") or (stimuliCourant = "Suj5") or (stimuliCourant = "Suj6")
  then
  put "changement de sujet" into field id 1022
  go to card id 1021
  set the visible of field id 1022 to true
  set the visible of button id 1023 to true
  end if

  -- si il n'y a plus de stimuli on finit
  if (StimuliCourant = "Fin")
  then
  close file fichRep
  put "fin de la manip" into field id 1022
  set the visible of field id 1022 to true
  set the visible of button id 1023 to false
  go to card id 1021
  end if

else

  -- si on ne change pas de carte on incrémente
  put numStimuliCourant+1 into numStimuliCourant

  -- on lance la vidéo suivante
  create invisible player "leJoueur"
  -- on n'affiche pas la vidéo si condition audio seul
  if condition = "A"
  then
    set the location of player "leJoueur" to -1000,-1000
  else
    set the location of player "leJoueur" to 550,350
  end if
  -- on coupe le son si condition vidéo seule
  if condition = "V"
  then
    set the playLoudness of player "leJoueur" to 0
  else
    set the playLoudness of player "leJoueur" to 100
  end if
  -- on place la vidéo
  set the height of player "leJoueur" to 576
  set the width of player "leJoueur" to 784
  set the filename of player "leJoueur" to repCourant & ".." & "/stimuli/" & stimuliCourant

  -- on calcule le temps du stimulus
  put the duration of player "leJoueur" into longueurVideo
  put the timeScale of player "leJoueur" into rapport
  put longueurVideo/rapport into dureeVideo
```

```
  -- on joue le stimulus
  start player "leJoueur"
  wait for dureeVideo + 1 sec
  delete player "leJoueur"
  wait for 0.5 sec

  --on met à jour l'affichage après le stimulus
  set the visible of button id 1004 to true
  set the visible of button id 1005 to true
  set the visible of button id 1006 to true
  set the visible of button id 1007 to true
  set the visible of button id 1008 to true
  set the visible of field id 1020 to true
  set the visible of scrollbar id 1010 to true
  set the visible of button id 1009 to true

end if

end mouseUp
```

Annexe 4
Matrices de confusion en effectif et pourcentage des réponses des juges aux stimuli des différents sujets

Légende

Reconnaissance non significative	Reconnaissance significative
Report non significatif	Report significatif

Sujet T audio

Stimulus \ Réponse	Etonnée	Déception	(Assez) calme	Angoissée	Hésitation
Etonnée	3 (10%)	3 (10%)	11 (37%)	4 (13%)	9 (30%)
Déception	1 (3%)	9 (30%)	1 (3%)	0 (0%)	19 (63%)
(Assez) calme	4 (13%)	2 (7%)	1 (3%)	1 (3%)	22 (73%)
Angoissée	13 (43%)	1 (3%)	3 (10%)	6 (20%)	7 (23%)
Hésitation	6 (20%)	6 (20%)	5 (17%)	6 (20%)	7 (23%)

Sujet T vidéo

Stimulus \ Réponse	Etonnée	Déception	(Assez) calme	Angoissée	Hésitation
Etonnée	6 (20%)	10 (33%)	2 (7%)	6 (20%)	6 (20%)
Déception	4 (13%)	15 (50%)	0 (0%)	7 (23%)	4 (13%)
(Assez) calme	2 (7%)	7 (23%)	3 (10%)	1 (3%)	17 (57%)
Angoissée	8 (27%)	2 (7%)	0 (0%)	15 (50%)	5 (17%)
Hésitation	1 (3%)	5 (17%)	8 (27%)	7 (23%)	9 (30%)

Sujet T audiovisuel

Stimulus \ Réponse	Etonnée	Déception	(Assez) calme	Angoissée	Hésitation
Etonnée	4 (13%)	7 (23%)	4 (13%)	5 (17%)	10 (33%)
Déception	2 (7%)	11 (37%)	1 (3%)	2 (7%)	14 (47%)
(Assez) calme	1 (3%)	4 (13%)	1 (3%)	2 (7%)	22 (73%)
Angoissée	6 (20%)	1 (3%)	1 (3%)	20 (67%)	2 (7%)
Hésitation	6 (20%)	2 (7%)	7 (23%)	11 (37%)	4 (13%)

Sujet S audio

Stimulus \ Réponse	Surprise par les résultats car un doute persistait sur la prononciation. Néanmoins les résultats donnent une certaine satisfaction personnelle	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.	Mission Impossible - concentration mais réponses au pif	Déçue par les résultats, j'essaie de trouver des solutions	Au pif, une envie de rigoler	Le but est de trouver des solutions à ces résultats catastrophiques : tous les moyens sont bons et en plus ça rassure
Surprise par les résultats car un doute persistait sur la prononciation. Néanmoins les résultats donnent une certaine satisfaction personnelle	4 (13%)	6 (20%)	4 (13%)	11 (37%)	4 (13%)	1 (3%)
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.	6 (20%)	8 (27%)	7 (23%)	7 (23%)	0 (0%)	2 (7%)
Mission Impossible - concentration mais réponses au pif	0 (0%)	3 (10%)	7 (23%)	10 (33%)	4 (13%)	6 (20%)
Déçue par les résultats, j'essaie de trouver des solutions	3 (10%)	5 (17%)	6 (20%)	8 (27%)	3 (10%)	5 (17%)
Au pif, une envie de rigoler	4 (13%)	4 (13%)	8 (27%)	6 (20%)	6 (20%)	2 (7%)
Le but est de trouver des solutions à ces résultats catastrophiques : tous les moyens sont bons et en plus ça rassure	0 (0%)	10 (33%)	3 (10%)	9 (30%)	0 (0%)	8 (27%)

Sujet S visuel

Stimulus \ Réponse	Surprise par les résultats car un doute persistait sur la prononciation. Néanmoins les résultats donnent une certaine satisfaction personnelle	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.	Mission Impossible - concentration mais réponses au pif	Déçue par les résultats, j'essaie de trouver des solutions	Au pif, une envie de rigoler	Le but est de trouver des solutions à ces résultats catastrophiques : tous les moyens sont bons et en plus ça rassure
Surprise par les résultats car un doute persistait sur la prononciation. Néanmoins les résultats donnent une certaine satisfaction personnelle	4 (13%)	9 (30%)	6 (20%)	7 (23%)	1 (3%)	3 (10%)
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.	2 (7%)	4 (13%)	9 (30%)	10 (33%)	0 (0%)	5 (17%)
Mission Impossible - concentration mais réponses au pif	8 (27%)	6 (20%)	5 (17%)	6 (20%)	3 (10%)	2 (7%)
Déçue par les résultats, j'essaie de trouver des solutions	5 (17%)	9 (30%)	3 (10%)	5 (17%)	4 (13%)	4 (13%)
Au pif, une envie de rigoler	10 (33%)	0 (0%)	3 (10%)	0 (0%)	13 (43%)	4 (13%)
Le but est de trouver des solutions à ces résultats catastrophiques : tous les moyens sont bons et en plus ça rassure	8 (27%)	8 (27%)	4 (13%)	5 (17%)	3 (10%)	2 (7%)

Sujet S audiovisuel

Stimulus \ Réponse	Surprise par les résultats car un doute persistait sur la prononciation. Néanmoins les résultats donnent une certaine satisfaction personnelle	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.	Mission Impossible - concentration mais réponses au pif	Déçue par les résultats, j'essaie de trouver des solutions	Au pif, une envie de rigoler	Le but est de trouver des solutions à ces résultats catastrophiques : tous les moyens sont bons et en plus ça rassure
Surprise par les résultats car un doute persistait sur la prononciation. Néanmoins les résultats donnent une certaine satisfaction personnelle	7 (23%)	7 (23%)	2 (7%)	7 (23%)	1 (3%)	6 (20%)
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.	1 (3%)	4 (13%)	7 (23%)	14 (47%)	2 (7%)	2 (7%)
Mission Impossible - concentration mais réponses au pif	4 (13%)	9 (30%)	6 (20%)	4 (13%)	4 (13%)	3 (10%)
Déçue par les résultats, j'essaie de trouver des solutions	5 (17%)	6 (20%)	5 (17%)	8 (27%)	1 (3%)	5 (17%)
Au pif, une envie de rigoler	10 (33%)	0 (0%)	4 (13%)	1 (3%)	15 (50%)	0 (0%)
Le but est de trouver des solutions à ces résultats catastrophiques : tous les moyens sont bons et en plus ça rassure	1 (3%)	8 (27%)	2 (7%)	5 (17%)	4 (13%)	10 (33%)

Sujet M audio

Réponse / Stimulus	Peur	Stress	Sûre de moi. / Bien (à l'aise)	Doute	Pas contente - déçue	Déçue
Peur	0 (0%)	3 (10%)	1 (3%)	22 (73%)	1 (3%)	3 (10%)
Stress	3 (10%)	6 (20%)	1 (3%)	15 (50%)	3 (10%)	2 (7%)
Sûre de moi. / Bien (à l'aise)	2 (7%)	5 (17%)	4 (13%)	12 (40%)	3 (10%)	4 (13%)
Doute	1 (3%)	2 (7%)	2 (7%)	20 (67%)	2 (7%)	3 (10%)
Pas contente - déçue	5 (17%)	3 (10%)	4 (13%)	4 (13%)	11 (37%)	3 (10%)
Déçue	4 (13%)	7 (23%)	3 (10%)	9 (30%)	2 (7%)	5 (17%)

Sujet M visuel

Réponse / Stimulus	Peur	Stress	Sûre de moi. / Bien (à l'aise)	Doute	Pas contente - déçue	Déçue
Peur	1 (3%)	1 (3%)	7 (23%)	10 (33%)	5 (17%)	6 (20%)
Stress	0 (0%)	3 (10%)	1 (3%)	13 (43%)	11 (37%)	2 (7%)
Sûre de moi. / Bien (à l'aise)	4 (13%)	7 (23%)	4 (13%)	6 (20%)	3 (10%)	6 (20%)
Doute	4 (13%)	5 (17%)	6 (20%)	9 (30%)	2 (7%)	4 (13%)
Pas contente - déçue	3 (10%)	5 (17%)	5 (17%)	6 (20%)	5 (17%)	6 (20%)
Déçue	0 (0%)	5 (17%)	9 (30%)	8 (27%)	3 (10%)	5 (17%)

Sujet M audiovisuel

Réponse / Stimulus	Peur	Stress	Sûre de moi. / Bien (à l'aise)	Doute	Pas contente - déçue	Déçue
Peur	2 (7%)	0 (0%)	3 (10%)	19 (63%)	2 (7%)	4 (13%)
Stress	0 (0%)	4 (13%)	1 (3%)	12 (40%)	8 (27%)	5 (17%)
Sûre de moi. / Bien (à l'aise)	5 (17%)	5 (17%)	4 (13%)	12 (40%)	3 (10%)	1 (3%)
Doute	0 (0%)	3 (10%)	3 (10%)	22 (73%)	2 (7%)	0 (0%)
Pas contente - déçue	1 (3%)	10 (33%)	4 (13%)	3 (10%)	7 (23%)	5 (17%)
Déçue	1 (3%)	6 (20%)	7 (23%)	10 (33%)	3 (10%)	3 (10%)

Sujet J audio

Stimulus \ Réponse	Inquiétude sur le but du jeu	Petite irritation	Petite lassitude! / L'indifférence s'installe	Je commence à comprendre qu'on se paye ma tête
Inquiétude sur le but du jeu	10 (33%)	5 (17%)	12 (40%)	3 (10%)
Petite irritation	7 (23%)	10 (33%)	7 (23%)	6 (20%)
Petite lassitude! / L'indifférence s'installe	8 (27%)	4 (13%)	11 (37%)	7 (23%)
Je commence à comprendre qu'on se paye ma tête	11 (37%)	12 (40%)	2 (7%)	5 (17%

Sujet J visuel

Stimulus \ Réponse	Inquiétude sur le but du jeu	Petite irritation	Petite lassitude! / L'indifférence s'installe	Je commence à comprendre qu'on se paye ma tête
Inquiétude sur le but du jeu	13 (43%)	5 (17%)	10 (33%)	2 (7%)
Petite irritation	9 (30%)	14 (47%)	5 (17%)	2 (7%)
Petite lassitude! / L'indifférence s'installe	7 (23%)	11 (37%)	7 (23%)	5 (17%)
Je commence à comprendre qu'on se paye ma tête	9 (30%)	7 (23%)	6 (20%)	8 (27%)

Sujet J audiovisuel

Stimulus \ Réponse	Inquiétude sur le but du jeu	Petite irritation	Petite lassitude! / L'indifférence s'installe	Je commence à comprendre qu'on se paye ma tête
Inquiétude sur le but du jeu	11 (37%)	6 (20%)	10 (33%)	3 (10%)
Petite irritation	13 (43%)	7 (23%)	4 (13%)	6 (20%)
Petite lassitude! / L'indifférence s'installe	9 (30%)	7 (23%)	10 (33%)	4 (13%)
Je commence à comprendre qu'on se paye ma tête	8 (27%)	13 (43%)	4 (13%)	5 (17%)

Sujet N audio

Réponse / Stimulus	Léger doute	Ne suis plus sur de moi, léger agacement	Essaie de rester concentré	Déception mais m'en amuse	Soulagé
Léger doute	4 (13%)	5 (17%)	11 (37%)	5 (17%)	5 (17%)
Ne suis plus sur de moi, léger agacement	13 (43%)	4 (13%)	10 (33%)	2 (7%)	1 (3%)
Essaie de rester concentré	11 (37%)	7 (23%)	7 (23%)	4 (13%)	1 (3%)
Déception mais m'en amuse	5 (17%)	11 (37%)	1 (3%)	9 (30%)	4 (13%)
Soulagé	12 (40%)	4 (13%)	10 (33%)	2 (7%)	2 (7%)

Sujet N visuel

Réponse / Stimulus	Léger doute	Ne suis plus sur de moi, léger agacement	Essaie de rester concentré	Déception mais m'en amuse	Soulagé
Léger doute	7 (23%)	10 (33%)	13 (43%)	0 (0%)	0 (0%)
Ne suis plus sur de moi, léger agacement	3 (10%)	5 (17%)	5 (17%)	8 (27%)	9 (30%)
Essaie de rester concentré	4 (13%)	10 (33%)	13 (43%)	3 (10%)	0 (0%)
Déception mais m'en amuse	4 (13%)	3 (10%)	2 (7%)	12 (40%)	9 (30%)
Soulagé	6 (20%)	6 (20%)	2 (7%)	9 (30%)	7 (23%)

Sujet N audiovisuel

Réponse / Stimulus	Léger doute	Ne suis plus sur de moi, léger agacement	Essaie de rester concentré	Déception mais m'en amuse	Soulagé
Léger doute	4 (13%)	9 (30%)	16 (53%)	0 (0%)	1 (3%)
Ne suis plus sur de moi, léger agacement	12 (40%)	2 (7%)	3 (10%)	7 (23%)	6 (20%)
Essaie de rester concentré	7 (23%)	16 (53%)	7 (23%)	0 (0%)	0 (0%)
Déception mais m'en amuse	3 (10%)	6 (20%)	1 (3%)	11 (37%)	9 (30%)
Soulagé	10 (33%)	4 (13%)	2 (7%)	7 (23%)	7 (23%)

Sujet R audio

Réponse / Stimulus	Concentration	Ennui - Concentration	Ennui	Ennui, incompréhension	Surprise, nervosité / Surpris, agacé	Agacé	Mon incompréhension est justifiée !
Concentration	11 (37%)	2 (7%)	5 (17%)	5 (17%)	3 (10%)	1 (3%)	3 (10%)
Ennui - Concentration	16 (53%)	1 (3%)	4 (13%)	2 (7%)	2 (7%)	1 (3%)	4 (13%)
Ennui	6 (20%)	5 (17%)	7 (23%)	5 (17%)	5 (17%)	1 (3%)	1 (3%)
Ennui, incompréhension	7 (23%)	3 (10%)	2 (7%)	6 (20%)	2 (7%)	6 (20%)	4 (13%)
Surprise, nervosité / Surpris, agacé	1 (3%)	1 (3%)	1 (3%)	3 (10%)	11 (37%)	9 (30%)	4 (13%)
Agacé	1 (3%)	3 (10%)	2 (7%)	2 (7%)	6 (20%)	14 (47%)	2 (7%)
Mon incompréhension est justifiée !	5 (17%)	6 (20%)	6 (20%)	4 (13%)	5 (17%)	3 (10%)	1 (3%)

Sujet R visuel

Réponse / Stimulus	Concentration	Ennui - Concentration	Ennui	Ennui, incompréhension	Surprise, nervosité / Surpris, agacé	Agacé	Mon incompréhension est justifiée !
Concentration	14 (47%)	4 (13%)	0 (0%)	3 (10%)	3 (10%)	3 (10%)	3 (10%)
Ennui - Concentration	7 (23%)	7 (23%)	2 (7%)	4 (13%)	2 (7%)	5 (17%)	3 (10%)
Ennui	3 (10%)	1 (3%)	2 (7%)	6 (20%)	7 (23%)	10 (33%)	1 (3%)
Ennui, incompréhension	1 (3%)	3 (10%)	4 (13%)	11 (37%)	6 (20%)	3 (10%)	2 (7%)
Surprise, nervosité / Surpris, agacé	1 (3%)	2 (7%)	0 (0%)	6 (20%)	8 (27%)	4 (13%)	9 (30%)
Agacé	3 (10%)	2 (7%)	2 (7%)	7 (23%)	5 (17%)	6 (20%)	5 (17%)
Mon incompréhension est justifiée !	3 (10%)	2 (7%)	0 (0%)	11 (37%)	5 (17%)	3 (10%)	6 (20%)

Sujet R audiovisuel

Réponse / Stimulus	Concentration	Ennui - Concentration	Ennui	Ennui, incompréhension	Surprise, nervosité / Surpris, agacé	Agacé	Mon incompréhension est justifiée !
Concentration	12 (40%)	5 (17%)	1 (3%)	3 (10%)	4 (13%)	4 (13%)	1 (3%)
Ennui - Concentration	12 (40%)	2 (7%)	0 (0%)	9 (30%)	1 (3%)	5 (17%)	1 (3%)
Ennui	6 (20%)	5 (17%)	2 (7%)	5 (17%)	6 (20%)	5 (17%)	1 (3%)
Ennui, incompréhension	1 (3%)	4 (13%)	5 (17%)	11 (37%)	1 (3%)	7 (23%)	1 (3%)
Surprise, nervosité / Surpris, agacé	0 (0%)	2 (7%)	0 (0%)	1 (3%)	8 (27%)	6 (20%)	13 (43%)
Agacé	0 (0%)	1 (3%)	2 (7%)	3 (10%)	8 (27%)	11 (37%)	5 (17%)
Mon incompréhension est justifiée !	5 (17%)	2 (7%)	0 (0%)	5 (17%)	9 (30%)	2 (7%)	7 (23%)

Table des illustrations

FIGURE 1: LES BIG SIX D'EKMAN, JOIE, SURPRISE, PEUR, COLERE, DEGOUT ET TRISTESSE (SANDER ET SCHERER, 2009, P.88). 12
FIGURE 2 : UNE AUTRE VISION DES EMOTIONS DE BASE : LE CERCLE DES EMOTIONS DE PLUTCHIK (NEIL, 2009). 12
FIGURE 3 : LE MODELE DE LASSWELL (LASSWELL, STRUCTURES ET FONCTIONS DE LA COMMUNICATION DANS LA SOCIETE, 1948) 19
FIGURE 4 : LE MODELE DE SHANNON ET WEAVER (SHANNON ET WEAVER, THEORIE MATHEMATIQUE DE LA COMMUNICATION, 1949)..... 19
FIGURE 5 : RILEY ET RILEY, COMMUNICATION DE MASSE ET SYSTEMES SOCIAUX, 1965 20
FIGURE 6 : SCHÉMA DE JAKOBSON (JAKOBSON, CLOSING STATEMENTS : LINGUISTICS AND POETICS, IN STYLE IN LANGUAGE, 1963) 21
FIGURE 7 : GRETA, UN ACA EN ANIMATION 3D (PELACHAUD, 2010)..... 25
FIGURE 8 : LA PLATE-FORME E-WIZ ET LE PARADIGME DU MAGICIEN D'OZ. (VANPE, 2006, P. 48) 32
FIGURE 9 : ECRAN D'ACCUEIL DE L'EXPERIENCE..... 35
FIGURE 10 : ECRAN REPONSE DU SUJET T. 36
FIGURE 11 : INTERVALLE DE SIGNIFICATIVITE EN FONCTION DU SEUIL DE HASARD POUR L'ANALYSE DES MATRICES DE CONFUSIONS. 39
FIGURE 12 : REPRESENTATION GRAPHIQUE DES REPONSES SIGNIFICATIVES A (NOIR), V (BLEU), AV (ROUGE) ET NON SIGNIFICATIVES (GRIS) DU SUJET T..... 40
FIGURE 13 : TABLEAU RECAPITULATIF DES STIMULI DU SUJET T..... 40
FIGURE 14 : REPRESENTATION GRAPHIQUE DES REPONSES SIGNIFICATIVES A (NOIR), V (BLEU), AV (ROUGE) ET NON SIGNIFICATIVES (GRIS) DU SUJET S. 42
FIGURE 15 : TABLEAU RECAPITULATIF DES STIMULI DU SUJET S..... 42
FIGURE 16 : REPRESENTATION GRAPHIQUE DES REPONSES SIGNIFICATIVES A (NOIR), V (BLEU), AV (ROUGE) ET NON SIGNIFICATIVES (GRIS) DU SUJET M..... 44
FIGURE 17 : TABLEAU RECAPITULATIF DES STIMULI DU SUJET M 44
FIGURE 18 : REPRESENTATION GRAPHIQUE DES REPONSES SIGNIFICATIVES A (NOIR), V (BLEU), AV (ROUGE) ET NON SIGNIFICATIVES (GRIS) DU SUJET J..... 46
FIGURE 19 : TABLEAU RECAPITULATIF DES STIMULI DU SUJET J 46
FIGURE 20 : REPRESENTATION GRAPHIQUE DES REPONSES SIGNIFICATIVES A (NOIR), V (BLEU), AV (ROUGE) ET NON SIGNIFICATIVES (GRIS) DU SUJET N. 47
FIGURE 21 : TABLEAU RECAPITULATIF DES STIMULI DU SUJET N 47
FIGURE 22 : REPRESENTATION GRAPHIQUE DES REPONSES SIGNIFICATIVES A (NOIR), V (BLEU), AV (ROUGE) ET NON SIGNIFICATIVES (GRIS) DU SUJET R. 49
FIGURE 23 : TABLEAU RECAPITULATIF DES STIMULI DU SUJET R 49

Sigles et abréviations utilisés

ACA :	Agent Conversationnel Animé
FoAK :	Feeling of Another's Knowing
FoK :	Feeling of Knowing
FoT :	Feeling of Thinking
ICP :	Institut de la Communication Parlée
IHM :	Interaction Homme Machine

Index

A
ACA, 24
Anaxagore, 9
Approche cognitive, 13
Approche darwinienne, 11
Approche linguistique, 15
Approche néo-darwinienne, 11
Approche physiologique, 12
Aristote, 9
Arnold, 13
Attitude, 28
Aubergé, 23, 28, 30
Audibert, 32
Averill, 14

B
Buisine, 24

C
Campbell, 28
Clément, 28
Communication, 17
Constructivisme social, 14
Cosnier, 21

D
Damasio, 14
Darwin, 11
Descartes, 9
Duchenne, 22

E
Ekman, 11
ELIZA, 24
Emotion, 8, 28
E-wiz, 32

F
FoAK, 28
FoK, 28
FoT, 28

G
GRETA, 25

H
Hart, 28
Hume, 10

J
Jakobson, 20
James, 12

K
Kendon, 23
Krahmer, 29

L
La Fontaine, 18
Langage, 17
Langage non verbal, 22
Langage verbal, 21
Lange, 13
Lasswell, 18
Lazarus, 13
Loyau, 22, 28, 30

M
McNeill, 22

N
Neil, 12
Nhouyvanisvong, 29

P
Pelachaud, 25
Platon, 9
Plutchik, 12

R
Reder, 28
Riley, 19
Rilliard, 32
Ritter, 28

S
Sansonnet, 25
Scherer, 14, 24
Shannon, 19
Shochi, 28
Signorello, 23, 24
Socrate, 9
Sound Teacher, 32
Spinoza, 10
Swerts, 29

V
Vanpé, 23, 24, 28, 30, 32, 33

W
Weaver, 19
Weizenbaum, 24
Wierzbicka, 15
Wundt, 24

Table des matières

REMERCIEMENTS **4**

SOMMAIRE **5**

INTRODUCTION **6**

PARTIE 1 ETAT DE L'ART 7

Chapitre 1 – Les émotions *8*

1.1 Les émotions dans la philosophie *9*

1.2 Les émotions vues par la science *11*

1.2.1 L'approche darwinienne et néo-darwinienne 11
1.2.2 L'approche physiologique 12
1.2.3 L'approche cognitive 13
1.2.4 Le constructivisme social 14
1.2.5 L'approche linguistique 15

1.3 Conclusion sur les émotions *15*

Chapitre 2 – Le langage *17*

2.1 La Communication *17*

2.1.1 Communiquer d'un point de vue historique/philosophique 18
2.1.2 Communiquer d'un point de vue scientifique 18

2.2 Langage verbal *21*

2.3 Langage non verbal *22*

2.3.1 Le non verbal visible 22
2.3.2 Le non verbal audible 23

2.4 Les Agents Conversationnels Animés (ACA) *24*

2.5 Conclusion sur le langage *25*

PARTIE 2 METHODOLOGIE EXPERIMENTALE 27

Chapitre 3 – Cadre théorique et enjeux de l'étude *28*

3.1 Cadre théorique *28*

3.1.1 Différence entre émotion et attitude 28
3.1.2 Evolution du Feeling of Knowing au Feeling of Thinking 28
3.1.3 Lien entre la prosodie expressive audiovisuelle et le FoK 29
3.1.4 Le FoT, une généralisation du FoK 30

3.2 Enjeux de l'étude *30*

Chapitre 4 – Du corpus à l'expérience *32*

4.1 Corpus Sound Teacher, plateforme E-wiz *32*

4.2 Sélection des stimuli de chaque sujet *33*

4.3 Les juges *34*

Chapitre 5 – Déroulement du test *35*

5.1 Interface d'expérimentation *35*

5.2 Procédure *35*

5.3 Limites de l'expérience *36*

PARTIE 3 RESULTATS DE L'EXPERIENCE 38

Chapitre 6 – Résultats et perspectives *39*

6.1 Traitement statistique *39*

6.2 Analyse des résultats *40*

6.2.1 Sujet T 40
6.2.2 Sujet S 42
6.2.3 Sujet M 44
6.2.4 Sujet J 46
6.2.5 Sujet N 47
6.2.6 Sujet R 49

6.3 Conclusion sur les résultats *50*

6.4 Perspectives *51*

CONCLUSION **52**
BIBLIOGRAPHIE **53**
TABLE DES ANNEXES **55**
TABLE DES ILLUSTRATIONS **76**
SIGLES ET ABREVIATIONS UTILISES **77**
INDEX **78**
TABLE DES MATIERES **79**

MOTS-CLÉS : langage, émotion, parole expressive, micro-événements

RÉSUMÉ

Notre travail, mené sur un corpus induit émotionnellement et restreignant l'interaction homme machine au minima langagier, s'inscrit dans la thématique de la communication des émotions de l'être humain en situation d'interaction face à face. Nous nous intéressons particulièrement aux messages contenus dans l'activité langagière non verbale présente hors du tour de parole. Nous présentons ici une partie théorique sur les émotions et le langage afin de placer le cadre de notre étude. Nous exposons ensuite l'expérience que nous avons mise en place, suivi d'une analyse des résultats obtenus auprès des quinze juges français qui ont passé le test.

KEYWORDS : language, emotion, expressive speech, micro events

ABSTRACT

Our work, carried out on an emotionally induced corpus that limits speech of the human-machine interaction to a bare minimum, is in the field of communication of human emotions in face to face interaction. We are particularly interested in messages contained in non verbal language activity out of the talk turns. We introduce our work with a theoretical part on emotion and language to place the framework of our study. Then, we talk about the test we conducted and expose an analysis of the results obtained from the fifteen French judges who took the test.

UNIVERSITE Stendhal GRENOBLE 3

Déclaration anti-plagiat

Document à scanner après signature
et à joindre au mémoire électronique

DECLARATION

1. Ce travail est le fruit d'un travail personnel et constitue un document original.
2. Je sais que prétendre être l'auteur d'un travail écrit par une autre personne est une pratique sévèrement sanctionnée par la loi.
3. Personne d'autre que moi n'a le droit de faire valoir ce travail, en totalité ou en partie, comme le sien.
4. Les propos repris mot à mot à d'autres auteurs figurent entre guillemets (citations).
5. Les écrits sur lesquels je m'appuie dans ce mémoire sont systématiquement référencés selon un système de renvoi bibliographique clair et précis.

NOM : DE BIASI PRENOM : Gilles

DATE : 27 juin 2011