THESE DE DOCTORAT

DE L'UNIVERSITÉ D'ÉVRY VAL D'ESSONNE

Présentée par

Mademoiselle Thi-Hai-Ha DANG

pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ D'ÉVRY VAL D'ESSONNE

Domaine

SCIENCE POUR L'INGÉNIERIE

Sujet de la thèse

D'ÉMOTION ET DE GRACE

VERS UN MODÈLE COMPUTATIONNEL UNIFIÉ DES ÉMOTIONS – APPLICATION À L'ÉCOUTE MUSICALE D'UN ROBOT DANSEUR

Thèse présentée et soutenue à Évry le 13 Janvier 2012 devant le jury composé de :

PESTY Sylvie	PR	Rapporteur
PELACHAUD Catherine	DR	Rapporteur
VIEYRES Pierre	PR	Rapporteur
NATKIN Stéphane	PR	Président du jury
SABOURET Nicolas	MdC HDR	Examinateur
HOPPENOT Philippe	PR	Directeur de thèse
HUTZLER Guillaume	MdC HDR	Co-tuteur
EHRHARDT Damien	HDR	Co-tuteur

Nom du Laboratoire : Informatics, Integrative Biology and Complex Systems – IBISC
Adresse : Bât. IBGBI – 3ème étage, 23, Bd de France 91034 – EVRY

Résumé

Les psychologues (comme A. Damasio, K. R. Scherer, P. Ekman) ont montré que l'émotion est un élément essentiel dans la prise de décision, dans l'évolution des capacités d'apprentissage et de création, dans l'interaction sociale. Il est donc naturel de s'intéresser à l'expression d'émotions dans le cadre de l'interaction homme-machine. Nous avons proposé dans un premier temps le modèle GRACE, modèle générique des émotions pour les applications computationnelles. Nous nous sommes basés en particulier sur la théorie psychologique de K. R. Scherer, qui cherche à produire une théorie des processus émotionnels qui soit modélisable et calculable. La pertinence de notre modèle a été vérifiée et validée via une comparaison avec différents modèles computationnels existants.

Si le modèle GRACE est générique, nous nous sommes attachés à montrer qu'il pouvait s'instancier dans un contexte particulier, en l'occurrence l'interaction homme-robot utilisant la modalité musicale. Nous nous sommes intéressés pour cela d'une part à la conception d'un module d'analyse du contenu émotionnel d'une séquence musicale, d'autre part à la conception de mouvements émotionnellement expressifs pour un robot mobile. Du point de vue de l'analyse musicale, la contribution principale de la thèse porte sur la proposition d'un ensemble réduit d'indicateurs musicaux et sur la validation du module d'analyse sur une base de données de grande taille conçue par un expert en musicologie. Du point de vue de la robotique, nous avons pu montrer expérimentalement qu'un robot avec des capacités expressives très limitées (déplacements, mouvements de caméra) pouvait néanmoins exprimer de manière satisfaisante un ensemble réduit d'émotions simples (joie, colère, tristesse, sérénité).

Abstract

Emotion, as psychologists argue (like A. Damasio, K. R. Scherer, P. Ekman), is an essential factor for human beings in making decision, learning, inventing things, and interacting with others. Based on this statement, researchers in Human-Machine Interaction have been interested in adding emotional abilities to their applications. With the same goal of studying emotional abilities, we propose, in our work, a model of emotions named GRACE, which helps in modelling emotions in computational applications. We based our model on the work of psychologist Klaus R. Scherer, who intensively searches to form a generic model of emotion applicable to computational domain (like informatics, robotics, etc.). We demonstrate the pertinence of our model by comparing it to other existing models of emotions in the field of informatics and robotics.

In this thesis, we also worked on the instantiation of GRACE, in particular the components *Cognitive Interpretation* and *Expression*. These two components have been developed to be applied in the context of interacting with users using music. To develop *Cognitive Interpretation*, we worked on the extraction of emotional content in musical excerpts. Our contribution consists in proposing a reduced number of musical features to efficiently extract the emotional content in music, and in validating them via a learning system with a large database designed by a musicologist. For *Expression*, we have worked on the design of emotional moves of a mobile robot. Through very limited moves (moves in space, camera moves), we have shown that with dance-inspired motions, the robot could efficiently convey basic emotions (i.e. happiness, sadness, anger, serenity) to people.

Remerciements

Je tiens tout d'abord à remercier l'ensemble des membres du jury pour m'avoir fait l'honneur de leur participation. Je sais combien leur temps est cher, surtout dans la période des soutenances et aussi à l'occasion de Noël 2011. Je les remercie également pour leur avis sur le travail de thèse et aussi les discussions que j'ai eues à la soutenance. Ces échanges m'ont été d'une aide inestimable pour l'amélioration de la qualité de ce mémoire.

Ce projet de thèse a été possible grâce à une collaboration interdisciplinaire de trois encadrants : M. Philippe HOPPENOT – professeur en robotique, M. Damien EHRHARDT – expert en musicologie, et M. Guillaume HUTZLER – expert en intelligence artificielle et systèmes multi-agents. Je leur suis très reconnaissante pour leur volonté de concevoir et de soutenir un projet de thèse aussi ambitieux.

Je désire remercier particulièrement M. Damien ERHRARDT pour son encadrement pendant la thèse. Sa passion pour la musique, son expertise sur l'interprétation musicale m'ont rassurée quant à la faisabilité de l'extraction du contenu émotionnel dans la musique. Grâce aux discussions que nous avons partagées, j'ai pu découvrir le monde de la musicologie, en particulier de l'interprétation musicale via le piano.

Mes compétences en recherche scientifique ont été beaucoup améliorées pendant la thèse. Ceci n'aurait jamais pu être possible sans l'encadrement intensif de M. Guillaume HUTZLER, envers qui je suis redevable. Je tiens à le remercier vivement pour sa supervision de très près, pour ses partages de connaissance sur le travail de thèse, pour ses raisonnements finement construits sur les divers sujets traités pendant la thèse, et pour ses corrections méticuleuses de mes fautes de langage écrit (à la fois en français et en anglais) dans mes documents (y compris ce mémoire). J'exprime aussi, à l'occasion de cette thèse, mon admiration pour sa passion pour la science, pour ses projets passionnants combinant la science et l'art.

Je tiens à exprimer ma gratitude à M. Philippe HOPPENOT – mon directeur de thèse pour son soutien et son assistance pendant ma thèse. C'était durant les séances de travail avec Philippe que j'ai appris la robotique pour l'assistance à la personne. Je le remercie pour cette formation. Philippe m'a beaucoup rassurée dans mon travail via ses écoutes, sa rapidité de répondre à mes questions, ses partages des idées, ses guides de réflexion sur les analyses que j'ai faites pendant ma thèse. Je lui suis redevable pour son assistance à mes travaux et pour sa volonté de former des doctorants pour la recherche scientifique.

L'évolution du travail de thèse a été accélérée par deux comités de thèse, l'un en 2009 et l'autre en 2010. Je tiens particulièrement à remercier M. Nicolas BRUNEL et M. François PACHET pour leur participation à ces comités de thèse. Je les remercie pour leurs commentaires et leurs conseils pour l'amélioration de mon travail. Je les remercie pour leur encouragement et l'intérêt qu'ils ont porté à mes recherches.

Je tiens à remercier également Hanna KLAUDEL, Jean-Marc DELOSME, et Jean-Christophe JANODET d'avoir participé activement à mes répétitions avant la soutenance de thèse. Leurs commentaires et leurs conseils m'ont permis d'améliorer la qualité de la présentation lors de ma soutenance du 13 Janvier 2012. Je voudrais aussi adresser mes sincères remerciements à Florence d'ALCHE-BUC pour ses

encouragements, et aussi pour ses attitudes qui m'ont fait rêver de devenir une dame comme elle. Je lui exprime mon admiration envers sa personnalité et ses réussites.

Un chemin a toujours un point de départ. Ma vie de recherche a commencé avec mon stage de M2 au laboratoire VALORIA de l'Université Bretagne-Sud. Je tiens donc à remercier M. Dominique DUHAUT – mon tuteur de stage. C'est grâce au stage avec M. DUHAUT que le modèle GRACE a été proposé. Et grâce à ce travail de stage, j'ai pu convaincre mes tuteurs de m'engager en thèse. Je lui suis très reconnaissante pour sa patience, son support, son encadrement pendant mon stage, et pour son grand intérêt pour mon projet de recherche.

Les résultats que j'ai pu présenté dans ce mémoire ont été rendus possible grâce à la participation de Roméo AGID – expert en musicologie. Il a fait le lourd travail d'annotation des 21 morceaux musicaux. Je me souviendrai pour toujours de cette contribution précieuse de Roméo, aussi bien que son enthousiasme et sa passion pour la musique, la philosophie, et l'informatique qu'il essaie lui-même de combiner dans son projet de thèse.

Une partie de cette thèse a été réalisée dans le cadre de TER d'Adel MEZINE et donc je tiens à lui adresser mes sincères remerciements. L'encadrement du TER m'a permis d'approfondir mes connaissances en apprentissage automatique, surtout au sujet des réseaux de neurones et des arbres de décision. Je lui en suis redevable et je lui souhaite bon courage pour son projet académique qu'il rêve de réaliser.

Ce projet de thèse aurait été impossible sans le support financier de la présidence de l'Université d'Evry Val d'Essonne pour des projets de recherche multidisciplinaire. Je suis très reconnaissante de ce support et j'espère que beaucoup d'autres personnes peuvent bénéficier de ce soutien pour réaliser des projets innovants.

Et il ne faut jamais oublier le rôle de l'environnement de travail très convivial dont j'ai eu la chance de profiter pendant ma thèse. Je tiens à remercier très sincèrement tous mes collègues du Laboratoire IBISC qui m'ont accueilli si aimablement au sein du laboratoire. J'adresse tout particulièrement mes remerciements à Dominique ANTONICELLI pour le service administratif efficace qu'elle a fourni pour mon séjour à IBISC. J'exprime mon affection à toutes les personnes travaillant à IBISC, pour leur politesse, leur gentillesse, leurs sens de l'humour. Je les remercie de m'avoir donné l'occasion d'apprécier la culture française, dont je rêvais quand je l'avais lue dans la littérature et vue à la télévision lors de mon enfance.

La complétion de cette formation doctorale est enracinée dans les rêves de mes parents. À leur époque, ils n'ont pas eu la chance de poursuivre les études supérieures et ils ont toujours rêvé d'en faire. Mon enfance a été enchantée par leurs conseils et leurs encouragements de poursuivre le parcours académique jusqu'au plus haut niveau d'étude. Ma réussite d'aujourd'hui est donc grâce à eux. Je les remercie de tout mon cœur, et je les aime avec tout mon cœur.

Je remercie aussi tous mes amis qui m'ont supportée pendant mes périodes difficiles de la thèse et aussi pendant mes moments dingues. Leur accompagnement m'est précieux pour la réalisation de mon rêve dont cette thèse n'est qu'un pas vers cet objectif. Le projet est encore loin d'être terminé…

Table des matières

Table des figures

Table des tableaux

Chapitre 1

Introduction

« Dark or Light is not a feeling, but a choice. »

La guerre des étoiles

Le rôle des émotions dans l'évolution biologique a été largement étudié par les biologistes. Les émotions comme la peur, la surprise, la colère sont importantes chez les animaux pour les aider à survivre et à évoluer. Par exemple, dans une situation de danger, le système physiologique de l'animal active les mécanismes de défense. Le sentiment de peur incitera l'animal à prendre la fuite, alors que la colère l'encouragera plutôt à se battre.

Chez l'être humain, les émotions jouent un rôle encore plus important. L'être humain est social, ce qui implique que l'interaction sociale est essentielle pour l'existence et l'évolution de l'espèce. L'émotion lui permet de caractériser les relations entre individus (l'empathie, l'amour, l'amitié, etc.). Dans les interactions sociales, l'émotion influence notamment la prise de décision. C'est un fait bien connu dans le marketing où certaines publicités sont conçues pour nous toucher pour nous encourager à acheter un produit ; ou dans la politique où les gestes et les discours des candidats sont pensés dans le but de gagner l'affection des gens ; ou bien dans l'art de la communication où l'on apprend à créer de l'affection chez l'auditeur via notre discours.

D'après certains psychologues (comme A. Damasio, K. Scherer, P. Ekman), l'émotion est considérée comme l'élément essentiel dans la prise de décision de l'humain, dans l'évolution des capacités d'apprentissage et de création, dans l'interaction sociale. Damasio et ses collègues montrent que quand il a subi un dommage au cortex préfrontal ventromédian, l'être humain n'est plus capable de prendre des décisions face à deux choix de même niveau d'avantage. La capacité à gérer ses émotions est aussi importante dans l'efficacité des tâches réalisées. C'est pour cette raison qu'il existe différents types de tests de personnalité qui permettent à une personne de reconnaître ses capacités émotionnelles (généralement regroupées sous le terme d'*Intelligence Émotionnelle*). Il est connu que si l'on arrive à bien gérer ses émotions, on peut améliorer sa performance dans le travail et ses relations sociales.

Etant donnée l'utilité de l'émotion dans la vie humaine, les psychologues ont consacré beaucoup d'efforts à la définition de l'émotion, afin de comprendre la composition essentielle d'un processus émotionnel chez l'humain. Plusieurs aspects de l'émotion ont été étudiés : les caractéristiques permettant de distinguer une émotion d'une autre ; les représentations dimensionnelles de l'émotion (comme la valence émotionnelle, l'activation) ; les différents types d'évaluation intervenant lors d'un processus émotionnel (comme l'évaluation cognitive, les réactions physiologiques, les changements mentaux). Leurs idées ont été reprises par les chercheurs dans le domaine de l'interaction homme – machine dans le but de faciliter l'utilisation des machines par l'humain.

Les domaines scientifiques les plus concernés par les travaux en psychologie des émotions sont les domaines de l'assistance à la personne, comme l'interaction avec

les enfants, avec les personnes âgées, avec les patients, avec les clients. Ces types d'interactions nécessitent de l'émotion pour être les plus naturels possible, ce qui doit rendre la communication plus agréable pour les utilisateurs. Pour ces derniers, la performance des machines qui interagissent avec eux n'est qu'un aspect de l'interaction. L'autre aspect, certainement le plus important, concerne la relation sociale, qui est liée à la capacité de comprendre l'émotion d'autrui et d'exprimer ses propres émotions. Ces capacités sont celles qui peuvent améliorer la facilité d'utilisation et donc le réconfort pour l'utilisateur humain.

Le nombre croissant des applications technologiques dans la vie de tous les jours induit aussi une plus grande interaction entre l'humain et les machines. Quand l'efficacité est suffisamment élevée, c'est la facilité d'interaction qui devient importante. Comme l'être humain est plus à l'aise avec les interactions émotionnellement convenables, il devient nécessaire que ces applications aient un certain niveau social. En effet, R. Picard a aussi remarqué que pour paraître intelligentes, les applications technologiques nécessitent non seulement la capacité d'accomplir leurs tâches efficacement mais aussi des capacités émotionnelles. Les agents virtuels ayant la capacité d'exprimer des émotions rendent l'interaction plus naturelle et agréable, les robots de service (comme les guides dans les musées ou les expositions, les jouets robotisés pour les enfants, etc.) deviennent plus acceptables et plus efficaces pour les utilisateurs.

Par conséquent, l'étude sur les différents aspects de l'émotion attire continuellement l'attention des recherches en informatique et en robotique. Certains chercheurs s'intéressent à l'expression des émotions pour rendre l'interaction plus agréable. D'autres sont intéressés par la simulation du processus émotionnel dans l'idée de rendre les réponses de leurs applications (comme les agents virtuels, les robots) plus compréhensibles à l'utilisateur.

Nous nous sommes intéressés pour notre part à la conception d'une interaction fluide entre un musicien et un robot, fondée sur la maîtrise de l'aspect émotionnel dans l'interaction homme-machine (en particulier homme-robot). L'idée d'utiliser la musique dans cette interaction vient du fait que la musique est l'une des voies de communication des émotions humaines parmi les plus universelles et populaires. L'ajout de la musique peut donc être considéré comme une modalité additionnelle pour exprimer les émotions du robot (en plus des modalités classiques, comme l'expression faciale, comportementale, vocale). Du côté de l'utilisateur, la familiarité de l'utilisation de la musique pour exprimer des émotions lui facilite évidemment l'expression de ses émotions. Le scénario d'interaction entre l'utilisateur (dans ce cas c'est un musicien/pianiste) et son robot personnel peut être vue comme un échange d'information de nature émotionnelle entre les deux partenaires de la communication (voir figure 1). La réalisation de ce scénario nécessite donc l'exploitation de trois domaines de recherche différents : (1) la perception du contenu émotionnel dans la musique, autrement dit la reconnaissance des émotions dans la musique jouée par l'interprète (i.e. le musicien), (2) l'expression émotionnelle robotisée pour communiquer avec le musicien, et (3) la représentation computationnelle des émotions pour le robot qui permet au robot de concevoir des stratégies d'actions appropriées en fonction des émotions qu'il perçoit dans la musique du musicien.

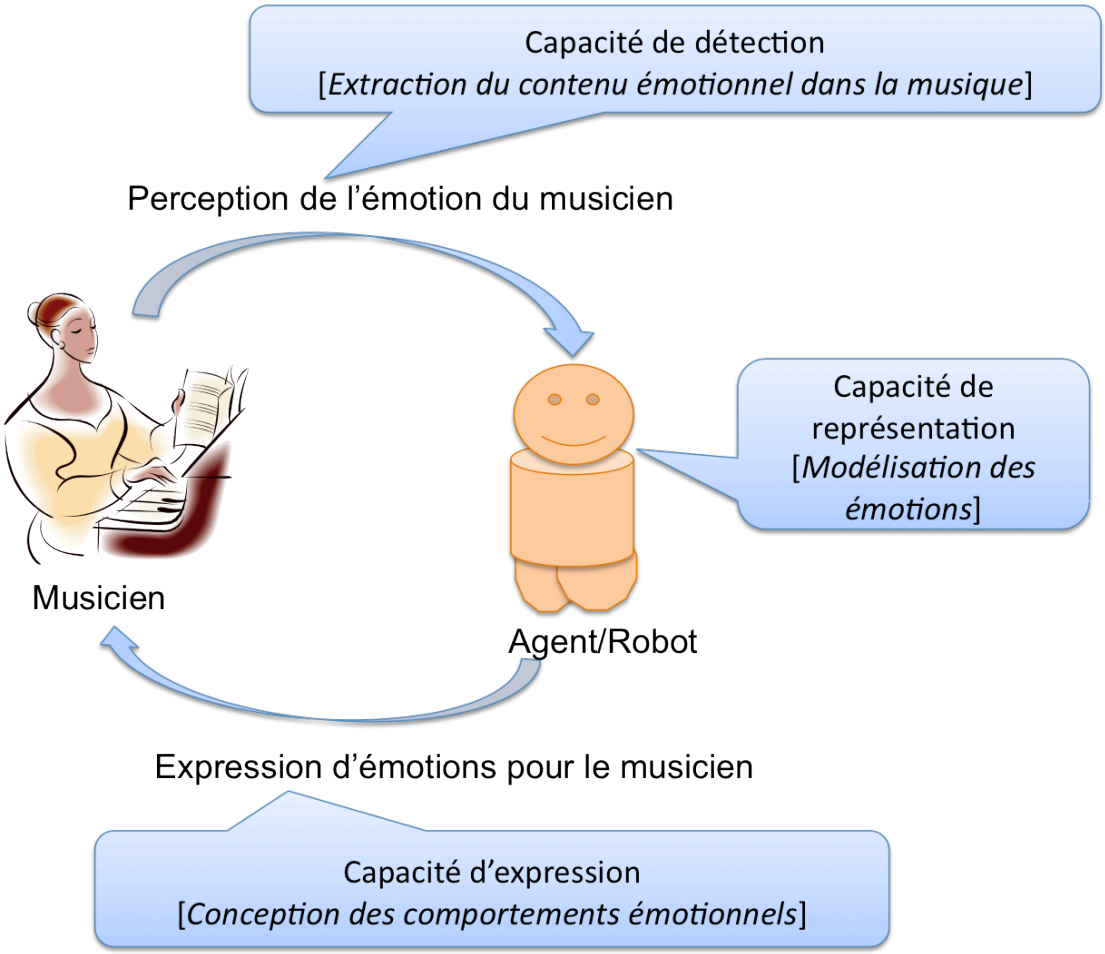


Figure 1 Scénario d'interaction entre un musicien et son robot

Ces problématiques constituent les trois axes de recherche de la thèse. Ils sont présentés dans trois chapitres séparés de ce mémoire. L'étude sur la représentation des émotions est présentée dans le chapitre 2, l'étude sur la perception des émotions dans la musique (i.e. la reconnaissance des émotions dans la musique) est traitée dans le chapitre 3, et l'étude sur l'expression robotisé des émotions lors de l'écoute musicale est synthétisée dans le chapitre 4. Le contenu de chaque chapitre est résumé dans les paragraphes qui suivent.

Axe 1 - La représentation des émotions : La représentation des émotions est décrite comme très complexe dans la littérature en psychologie. Il n'existe pas encore de consensus entre les psychologues sur la définition de l'émotion. Les modèles computationnels des émotions développés en informatique sont aussi très divers en termes de conception et de fonctionnement. L'utilité de ces modèles est limitée par les applications spécifiques car ils sont créés pour répondre à un contexte d'interaction prédéfini. Cela pose donc une certaine difficulté aux concepteurs d'une application pour trouver une architecture qui simule le processus émotionnel et qui s'adapte à l'application envisagée. Est-ce qu'il existe un consensus parmi les modèles computationnels des émotions existants dans ces domaines technologiques ? Ou bien, est-ce qu'il est possible de proposer un consensus sur la conception et la composition d'un processus émotionnel pour tous les modèles computationnels des émotions ? Nous traiterons ces questions dans le deuxième chapitre de la thèse en explicitant trois étapes de recherche : Identification – Modélisation - Validation. L'identification consiste à identifier ce qu'est une émotion humaine, et ce que sont les éléments

importants pour une modélisation informatique. Cette étape est faite via une étude bibliographique des travaux en psychologie sur la définition des émotions. Ce sont des travaux qui ont eu une grande influence dans le développement des modèles computationnels des émotions. L'identification est suivie par l'étape de modélisation. Nous présenterons ensuite notre propre modèle des émotions, nommé GRACE, qui s'inspire des théories psychologiques et s'adapte au contexte computationnel. La validation du modèle GRACE est faite de façon théorique et fonctionnelle. Sur l'aspect théorique, nous montrerons la généricité de notre modèle en le comparant avec les modèles computationnels existant en informatique et en robotique. L'aspect fonctionnel du modèle GRACE est validé partiellement via nos résultats de recherche sur les deux autres axes de recherche (i.e. sur la perception des émotions dans la musique et sur l'expression émotionnelle robotisée lors de l'écoute musicale). L'implémentation complète du modèle est envisagée pour la validation de notre modèle GRACE.

Axe 2 – La perception des émotions dans la musique : Un autre aspect traité dans cette thèse est l'extraction du contenu émotionnel dans la musique. En général, l'extraction de l'information émotionnelle est aussi importante pour la communication et pour la compréhension des sujets lors de l'interaction homme – machine. Ce type d'extraction peut concerner les domaines de la médecine, du service à la personne, du traitement du langage naturel, du traitement de documents écrits, de la performance artistique… Notre projet s'intéresse à ce dernier domaine, en particulier dans le domaine de l'écoute musicale. Notre motivation vient du fait que l'écoute musicale est l'un des moyens les plus utilisés par l'humain pour s'exprimer et réguler ses émotions. Pour un robot de service ou un agent virtuel qui interagit avec l'utilisateur humain, la musique est donc un bon moyen pour faciliter l'interaction. Notre troisième chapitre discutera les différents aspects de l'extraction automatique du contenu émotionnel dans la musique, qui passe donc de l'identification des descripteurs sonores et émotionnels dans la musique, à l'implémentation du système d'extraction automatique du contenu émotionnel dans la musique, et à la validation de notre système d'extraction. L'identification consiste en une étude bibliographique sur l'extraction du contenu émotionnel dans la musique, la constitution de descripteurs sonores et contextuels de la musique ; cette étape est présentée au début du chapitre. La construction d'un système d'extraction de l'information émotionnelle dans la musique est ensuite discutée. Nous faisons ensuite l'analyse des performances du réseau de neurones que nous avons développé par rapport à la littérature pour valider notre système.

Axe 3 – L'expression émotionnelle robotisée lors de l'écoute musicale d'un robot mobile : L'expression de l'émotion est un autre facteur essentiel dans l'interaction sociale. L'être humain exprime ses émotions pour soutenir ses paroles, sa position, ses standards, mais aussi faciliter l'interaction avec d'autres. Il a tendance à projeter l'action des autres sur le plan émotionnel lors de l'interaction. Les études sur l'interaction homme – machine ont aussi trouvé que l'expression émotionnelle améliore en général la participation des sujets, facilite l'interaction et rend ces sujets plus à l'aise et plus motivés pour communiquer avec la machine. Dans le but de développer un modèle computationnel complet des émotions, notre projet traite aussi la question de l'expression des émotions pour les robots de service. Nous présenterons notre travail sur ce sujet dans le dernier chapitre de ce mémoire, qui comprend également trois étapes : Identification – Modélisation - Validation. L'identification consiste en étude bibliographique sur l'expression émotionnelle des robots mobiles

dans le contexte de l'écoute musicale et les mouvements possibles de notre robot LINA. La modélisation correspond à la conception des mouvements émotionnels d'un robot mobile, inspirés des mouvements émotionnels des musiciens lors de leurs performances artistiques. Nous validons notre approche en présentant le résultat d'expérimentations que nous avons menées lors de la Fête de la Science en 2010. Le résultat des expérimentations montre qu'un robot, bien qu'ayant des capacités d'expression humanoïde très limitées, peut cependant être capable d'exprimer ses émotions correctement pour l'être humain.

Le mémoire se termine avec des perspectives sur le développement futur des trois axes de recherches pour une réalisation concrète du scénario d'interaction musicien-robot dans l'avenir.

Chapitre 2

Modélisation des émotions

L'ingénierie a, dès sa naissance, révolutionné la vie quotidienne des gens. L'assistance des machines a apporté une amélioration majeure à la productivité et à la qualité de vie humaine. D'abord conçues comme des outils destinés à aider à la production, les machines ont fait un grand pas dans la réduction de la distance avec l'humain, et elles sont alors sorties des usines pour se retrouver aujourd'hui dans notre maison (machine à laver, machine à café, télévision, ordinateur, etc.).

Cette évolution modifie également la philosophie de développement des machines. Initialement, ces machines étaient appréciées par leur capacité fonctionnelle, c'est-à-dire leur performance à réaliser des tâches techniques (production de voitures pour un bras robotique, capacité à laver le linge pour une machine à laver) au moindre coût. La proximité et l'interaction avec des humains ayant des connaissances très variées (et parfois aucune compétence de nature technique) ont fait surgir la nécessité de doter ces machines de capacités sociales. Cela passe en particulier par la possibilité de communiquer avec l'humain aisément, grâce à des capacités à comprendre son interlocuteur (par la vision, l'analyse de la parole, etc.) et à lui répondre (par la parole, les gestes de salutation, etc.). Depuis environs quinze ans, ces capacités ont été modulées par un élément fondamental dans la communication humaine à savoir l'émotion.

L'émotion jouant un rôle dans chacun des processus cognitifs humains, la prise en compte de l'émotion pour des machines doit se situer non seulement au niveau de l'interface (perception et action) mais également au niveau du contrôle (prise de décision et planification). Du point de vue de l'interface, il s'agit de doter les machines de capacités à reconnaître et à exprimer des émotions lors de la communication. Du point de vue du contrôle, il s'agit de prendre en compte les mécanismes de création d'une émotion, i.e. le processus émotionnel proprement dit. Dans ce processus émotionnel, l'émotion d'une part et les tâches d'autre part (dans leur organisation et leur réalisation) sont mélangées dans les modèles actuels des émotions. Autrement dit, la distinction entre processus émotionnel et processus de gestion des tâches est souvent difficile à établir. Le but de ce chapitre est de clarifier cette distinction, en précisant la frontière entre ce qui est de l'ordre de l'émotion et ce qui est de l'ordre de l'information et en explicitant les interactions entre les deux.

L'organisation de ce chapitre est la suivante. Partant des théories psychologiques connues sur la définition des émotions, nous présenterons la conception du modèle GRACE, modèle générique qui se fonde sur ces théories. Cette conception initiale est le résultat de mon stage de Master 2 effectué au laboratoire VALORIA, université de Bretagne-Sud. Pendant la thèse, nous avons fait évoluer le modèle GRACE en vue d'en faire un modèle computationnel, de manière à faciliter son implémentation et son instanciation dans différents contextes applicatifs. Les changements faits lors de la thèse sont présentés dans la section 3. La généricité du modèle GRACE est conservée lors de cette transformation, ce que nous montrons par la projection de différents

modèles des émotions existants dans le domaine de l'informatique et de la robotique sur notre modèle GRACE.

1. Processus émotionnel en psychologie

1.1. Introduction

La problématique de l'émotion humaine est très étudiée dans le monde de la psychologie. Depuis le travail de Darwin sur l'expression de l'émotion soulignant le fait que l'expression émotionnelle chez l'humain est héritée de l'ancêtre animal, un grand nombre de chercheurs ont étudié les processus émotionnels chez l'humain. Pendant la première moitié du 20ème siècle, Errol Bedfort, Magna Arnold et d'autres ont proposé différents éléments intervenant dans l'émotion humaine, dont l'intention, le contexte et les événements extérieurs. Une grande attention sur le sujet s'est manifestée au cours de la deuxième moitié du 20ème siècle. Jerome Neu (1977) ajoutait ainsi le désir et la croyance à la constitution de l'émotion. Magna Arnold (1966) soulignait le fait que l'émotion est le résultat de l'évaluation d'un événement. William Lyon, en 1980, a détaillé cette relation en indiquant que l'évaluation d'un événement commence par un changement physiologique anormal dû à cet événement. Ortony et ses collègues, en 1988, ont fait une proposition très importante sur l'évaluation de l'événement pendant le processus émotionnel. Cette proposition suggère que l'émotion humaine est une réaction affective lors d'une situation en réponse aux conséquences de l'événement, à l'action de l'agent, ou à l'aspect de l'objet. Robert Solomon, en 1980, propose que l'émotion participe au choix stratégique comportemental vis-à-vis des buts de protection de soi et de l'amélioration du respect pour soi. R. S. Lazarus supportait cette théorie de Solomon sur le choix stratégique et introduisait la stratégie d'adaptation lors du déroulement du processus émotionnel. Klaus Scherer, récemment, a construit une description complète sur le processus émotionnel chez l'humain. Il propose que le processus émotionnel chez l'humain consiste en l'interaction entre cinq composantes : *subjective feeling* est en charge de gérer l'expérience émotionnelle, *cognition* s'occupe de l'analyse cognitive, *motor expression* est en charge de l'expression gestuelle/comportementale de l'émotion, *action tendency or desire* contrôle la tendance à agir et les buts, et la dernière composante, *neurological process*, est en charge de contrôler l'état physiologique de l'humain.

Au niveau conceptuel, il y a trois grandes théories dans la recherche en psychologie sur l'émotion humaine, selon le psychologue Klaus Scherer. La première, dite *théorie de l'émotion basique*, considère que les émotions humaines sont réparties en familles d'émotions et chaque famille est caractérisée par des symptômes spécifiques. La *théorie constructiviste* considère l'émotion humaine comme une évaluation en valence-activation de la personne dans une situation donnée. Cette théorie *constructiviste* est assez connue en informatique pour sa capacité à distinguer les émotions en fonction de leur valeur d'activation et de valence. La troisième théorie considère l'*émotion comme un processus d'évaluation* qui associe une réaction émotionnelle appropriée à un événement (interne ou externe), en fonction de critères précis. Quelques travaux significatifs de chaque théorie sont présentés dans la suite.

1.2. Théories des familles d'émotions

Les théories des familles d'émotions, aussi connues sous le nom de « théories des émotions basiques », considèrent les émotions humaines comme des programmes prédéfinis ayant leurs propres stimuli d'incitation et leurs propres styles d'expression et de comportements. Les psychologues qui développent ces théories étudient notamment les caractéristiques qui permettent de distinguer ces programmes prédéfinis. Une liste de caractéristiques a notamment été proposée par Paul Ekman (Ekman, 1992) selon neuf critères (Table 1) permettant de distinguer les différentes émotions humaines. Cet auteur s'intéresse à la reconnaissance des émotions et non aux processus émotionnels. Cette idée est donc très intéressante, surtout pour les domaines de modélisation où l'on essaie de simuler les symptômes des émotions.

Table 1 Caractéristiques des familles des émotions, proposées par (Ekman, 1992)

No	Caractéristique	Description
1.	Signaux universels spécifiques	Expression faciale, symptôme physiologique, activité neuronale, etc.
2.	Présence chez d'autres animaux	L'expression émotionnelle est aussi présente dans l'animal et l'on pourrait la reconnaître d'une façon ou d'une autre
3.	Physiologie distincte	Il existe des patterns physiologiques spécifiques pour des émotions différentes, comme pour la colère, la peur, le dégoût, etc.
4.	Evénement antérieur universel	Un événement suscitant une émotion chez l'humain pourrait susciter cette émotion chez l'animal ; ceci s'appliqu aussi pour des personnes venant de cultures différentes
5.	Cohérence en réponse émotionnelle	Il y a une cohérence entre l'expression faciale et les symptômes physiologiques d'une émotion
6.	Activation rapide	L'émotion est activée rapidement après l'occurrence de l'événement d'activation
7.	Courte durée	L'émotion est présente pendant une courte durée
8.	Mécanisme d'évaluation automatique	L'évaluation d'un événement est automatique par un mécanisme d'évaluation, sans être remarquée par la personne
9.	Occurrence incontrôlable	L'émotion s'est involontairement exprimée, on ne peut pas la contrôler

Ortony et Tunner se sont interrogés sur la validité de la théorie des émotions basiques. Ces derniers arguent que le processus émotionnel nécessite de l'analyse cognitive tandis que dans les familles des émotions basiques, cet élément n'est pas souligné pour la distinction des émotions (Ortony & Turner, 1990). Carrolle E. Izard répond à la question d'Ortony et Tunner en montrant que la cognition n'est pas nécessairement présente lors d'une sensation émotionnelle (Izard, 1992). Ce dernier liste quatre types d'événements capables de susciter de l'émotion : *Neuronaux*, *Sensorimoteurs*, *Motivationnels*, et *Cognitifs*. Les événements de type *Neuronal* sont les événements créés par le cerveau, produits notamment par les activités neuronales, par les transmetteurs neuronaux, comme la création d'une idée et/ou d'un souvenir d'une situation qui fait naître un processus émotionnel. Les événements de type *Sensorimoteur* concernent les comportements, comme les traits du visage, les gestes du corps, les symptômes physiologiques qui déclenchent le processus émotionnel via

le retour proprioceptif. Les événements de type *Motivationnel* incluent les motivations quotidiennes comme la faim, la soif, etc. Les événements de type *Cognitif* concernent notamment les croyances, les désirs, les intentions, etc. de l'individu. Parmi ces catégories, seul le type *Cognitif* nécessite une analyse cognitive, les trois autres n'en ayant pas besoin.

Selon les théories des émotions basiques, les familles d'émotions les plus distinctes sont la colère, la peur, le dégoût, et la tristesse. Ekman souligne le fait que les émotions négatives sont plus faciles à classifier en familles que les émotions positives, comme l'amusement, le soulagement, la fierté, la satisfaction, etc. La raison principale, selon Ekman, en est que ces émotions positives ne possèdent pas de signaux suffisamment distinctifs (par exemple, les traits du visage pour exprimer du soulagement ou de la satisfaction ne diffèrent pas clairement de l'un à l'autre).

1.3. Théories constructivistes des émotions

Un processus émotionnel, selon les constructivistes, est un processus constitué de 5 concepts psychologiques de base : l'émotion de base (*core affect*), la qualité affective, l'affect attribué, la régulation d'affect, et l'objet. Ces concepts sont décrits dans la Table 2 :

Table 2 Définition des termes techniques d'un processus émotionnel proposé par (Russell, 2003, traduction personnelle)

Concept de base	**Définition**	**Commentaire**
Emotion de base	Un état neurophysiologique, consciemment accessible comme un sentiment simple et non réflexif, qui est une combinaison des valeurs d'hédonisme (plaisir - mécontentement) et d'activation (endormi - activé)	L'émotion de base elle-même est indépendante de l'Objet mais, via l'attribution, elle peut être dirigée vers un Objet. Le niveau de conscience est primaire
Qualité affective	La capacité à causer un changement dans l'émotion de base	Elle peut être décrite à l'aide des deux mêmes dimensions que l'émotion de base
Affect attribué	L'émotion de base attribuée à un Objet	(a) loin de tous jugements possibles sur la réalité de l'Objet (b) l'attribution est typiquement rapide et automatique mais peut aussi être délibérative
Régulation d'affect	Action visant directement à changer l'émotion de base	Ce processus ne dépend pas de l'Objet
Objet	Une personne, une condition, une chose, ou un événement auquel l'état mental est adressé	Un Objet est une représentation psychologique, et donc les états mentaux peuvent être dirigés vers des fictions, le futur, ou d'autres formes de réalité virtuelle

Les cinq concepts sont tous en lien, d'une manière ou d'une autre, avec l'*émotion de base*. Celle-ci peut être considérée comme l'état mental/émotionnel actuel de l'individu. La *qualité affective* est liée à un Objet et représente l'affectation objective

de l'Objet, indépendamment de la perception éventuelle de l'individu. Et l'affect attribué est le résultat de la perception de l'individu sur l'Objet, cette attribution déterminant si un changement dans l'*émotion de base* est nécessaire. Quand un changement de l'émotion de base est décidé, la régulation d'affect est mise en place pour la réaliser. De manière générale, les constructivistes considèrent l'émotion comme une composition des *émotions de base*. Un processus émotionnel est donc vu comme une suite de changements des *émotions de base* analysable dans les termes des concepts décrits ci-dessus.

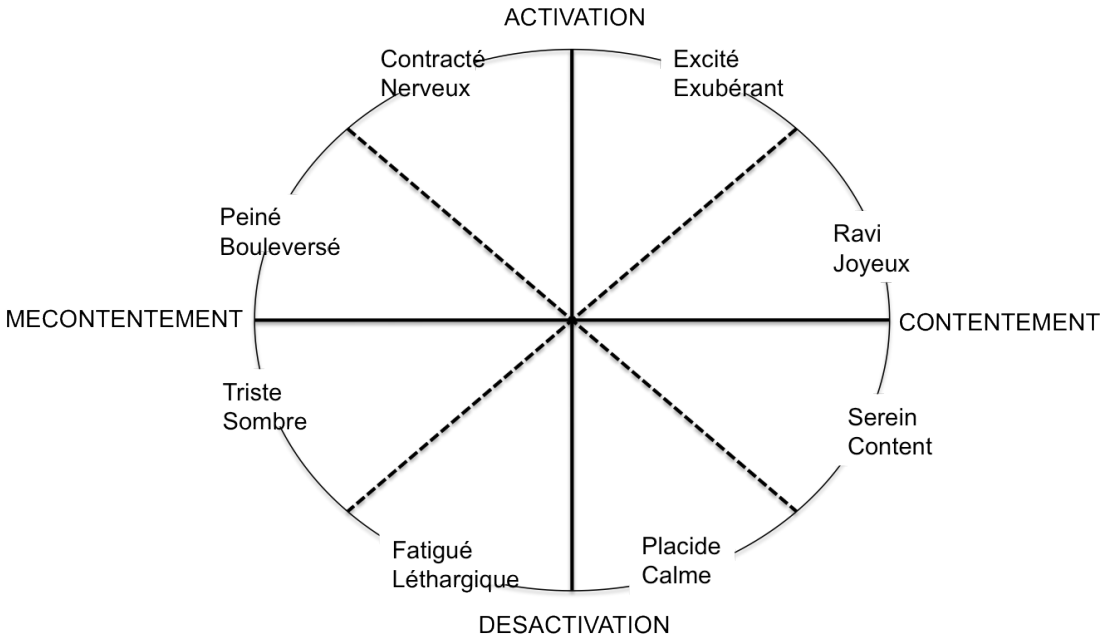


Figure 2 Espace de 2D de l'*émotion de base* proposée par (Russell, 2003)

Chaque *émotion de base* est représentée par deux valeurs : Valence et Activation (Figure 2). La valence mesure le niveau de plaisir (ou niveau de contentement) de l'émotion, et l'activation mesure le niveau d'énergie de l'émotion (ou intensité). L'*émotion de base* est considérée comme la brique élémentaire, à la base des différentes instances possibles de l'émotion, comme l'humeur, la perception, l'analyse cognitive, etc. De plus, l'introduction de *l'émotion de base* permet de décrire l'émotion indépendamment de l'objet et donc indépendamment du contexte, ce qui permet de la considérer comme un état mental sans chercher à en connaître les causes ou les conséquences. Ces dernières, en revanche, sont connues dans le processus émotionnel.

L'idée constructiviste de réduire toutes les évaluations possibles dans le processus émotionnel est intéressante mais ne convainc pas les tenants d'autres théories. Les psychologues qui défendent la théorie des émotions basiques considèrent par exemple que les émotions basiques sont les briques élémentaires de tous les phénomènes émotionnels chez l'humain et l'animal. Une autre tradition en psychologie des émotions, appelée la théorie des processus d'évaluation des événements, considère l'émotion comme un processus d'évaluation qui prend en compte davantage de critères que la Valence et l'Activation. La section suivante décrit cette tradition.

1.4. Théories des processus d'évaluation des événements

La troisième tradition en psychologie des émotions s'intéresse à l'évaluation des informations dans le processus émotionnel. Les psychologues qui défendent cette tradition abordent le processus émotionnel en analysant ce qui se passe lorsqu'un individu est confronté à des événements chargés émotionnellement, en identifiant les paramètres déterminant l'émotion ressentie en fonction de la situation. Nous allons voir dans cette section trois travaux en psychologie sur le processus d'évaluation des événements : il s'agit du modèle d'évaluation des événements proposé par Ortony et collègues, très apprécié des informaticiens, la théorie de la stratégie du « faire-face » (*coping*) lors d'une situation de stress proposé par Lazarus, qui explique une fonction très importante de l'émotion dans la vie humaine, et la théorie sur le processus d'évaluation complet proposé par Scherer. Cette dernière est aussi la base psychologique de notre proposition.

1.4.1. Modèle d'évaluation des événements

Dans l'idée de proposer une théorie des émotions adaptée aux applications informatiques, Ortony, Clore et Collins, dans leur ouvrage « The cognitive structure of emotions » en 1988 ont proposé un modèle des émotions s'appuyant sur l'interprétation cognitive de l'émotion (Ortony, Clore, & Collins, 1988). Selon leur théorie, l'émotion vis-à-vis d'un événement est déterminée en fonction de trois dimensions : les conséquences de l'*événement*, l'action de l'*agent*, et l'aspect de l'*objet*. L'interprétation des trois dimensions est présentée dans la Figure 3.

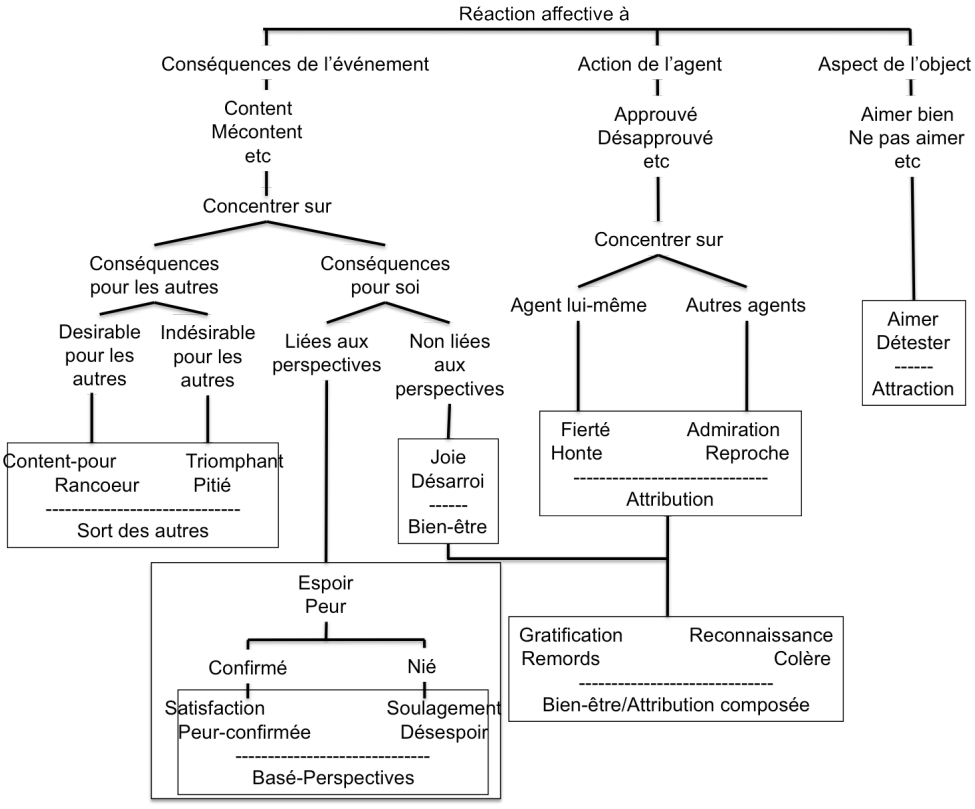


Figure 3 Théorie de l'évaluation émotionnelle d'Ortony, Clore et Collins

Du point de vue des *conséquences de l'événement*, l'évaluation concerne notamment le sort des autres, son propre bien-être, et les perspectives pour le futur. Du point de vue de l'*action de l'agent*, l'évaluation correspond à l'attribution d'une appréciation vis-à-vis de l'agent qui a réalisé l'action, qui peut moduler la sensation de bien-être. Du point de vue de l'*aspect* de l'objet, l'évaluation concerne une attraction ou répulsion pour l'agent/objet qui peut aussi susciter de l'émotion, comme l'amour ou la haine. Plusieurs émotions peuvent ainsi co-exister en même temps.

Par exemple, une personne entend une chanson d'enfance. Dans cet événement, l'« action » est de « chanter la chanson », l'objet est la chanson, et l'agent est le chanteur. En regardant les conséquences possibles de l'événement, entendre une chanson d'enfance donne normalement un sentiment positif, donc la personne pourrait ressentir de la satisfaction. Par rapport à l'action de l'agent, la personne pourrait ressentir de l'admiration ou de la jalousie pour ce chanteur. Quant à l'objet, la personne pourrait aimer cette chanson ou non selon ses préférences (les paroles, la mélodie, etc.).

L'intensité des émotions est calculée en fonction de trois variables (la *désirabilité*, la *valeur sociale*, et l'*attraction*) correspondant à trois branches d'évaluation. Les émotions liées aux conséquences de l'événement sont estimées en fonctions de la désirabilité - indiquant comment l'événement permet à l'agent d'atteindre ses buts. Les émotions liées à l'action de l'agent sont estimées en fonction de la valeur sociale - signifiant comment l'action est évaluée par rapport aux standards/normes sociaux. Les émotions liées à l'objet sont estimées par la variable d'attraction qui représente l'attitude ou les préférences de la personne.

Du fait de sa clarté et de sa simplicité, le modèle OCC a été implémenté par plusieurs modèles informatiques pour simuler les processus émotionnels. Nous reviendrons sur ces modèles informatiques dans la Section 4.

1.4.2. Théorie de Smith et Lazarus sur l'évaluation d'un événement

Smith et Lazarus, dans leur publication *Emotion and Adaptation* (Smith & Lazarus, 1990), ont détaillé leur théorie sur le processus émotionnel chez l'humain. Ce processus est, selon eux, réalisé en deux niveaux d'évaluation : primaire, et secondaire.

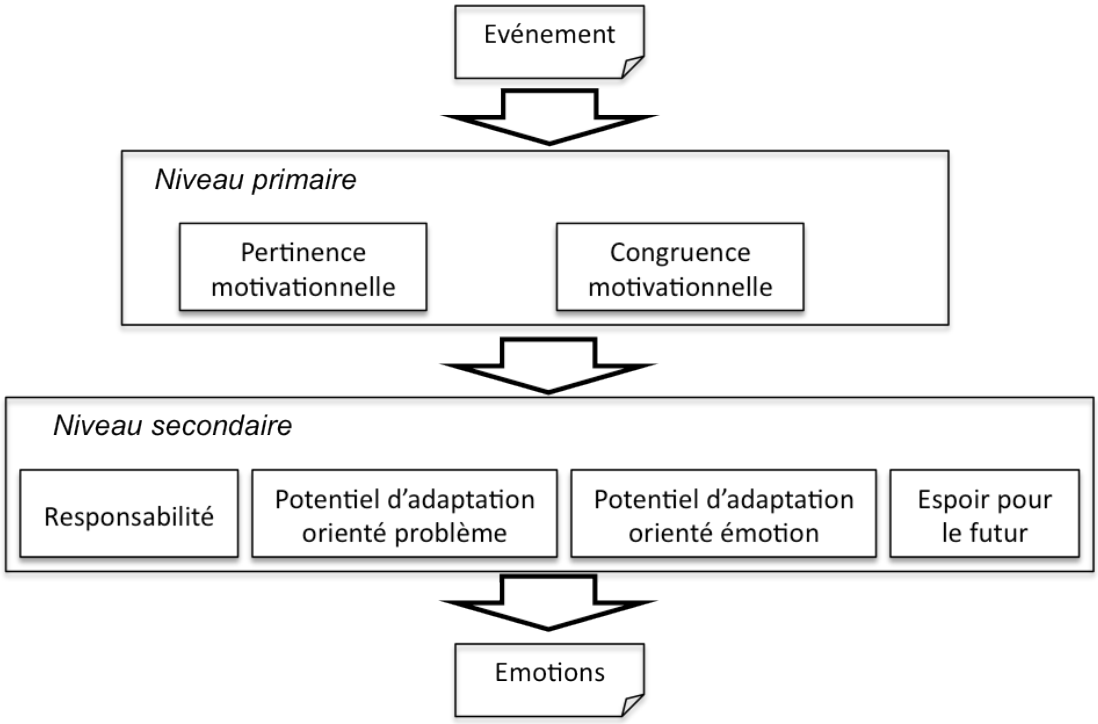


Figure 4 Processus émotionnel proposé par Lazarus et collègues

Au niveau primaire, l'événement est évalué en fonction de sa pertinence motivationnelle et en fonction de la congruence motivationnelle. La pertinence motivationnelle représente le lien entre l'événement et les buts/objectifs de la personne, autrement dit ce que la personne a l'intention d'accomplir. La congruence motivationnelle représente l'influence de l'événement sur l'accomplissement des buts/objectifs de la personne. Au niveau secondaire, il y a quatre critères d'évaluation : la *responsabilité*, le *potentiel d'adaptation orienté problème*, le *potentiel d'adaptation orienté émotion*, l'*espoir pour le futur*. Le critère de responsabilité permet de définir qui est responsable de la situation, autrement dit, qui devrait/pourrait recevoir une réprimande ou une récompense. Le potentiel d'adaptation orienté problème représente la capacité d'agir pour résoudre le problème survenu. Le potentiel d'adaptation orienté émotion, quant à lui, reflète la capacité à « changer d'avis », c'est-à-dire à ajuster l'émotion ressentie en fonction de l'événement. L'espoir pour le futur reflète la possibilité que de futurs changements psychologiques puissent affecter la congruence motivationnelle et dans quelle direction (bonne ou mauvaise).

Les événements pouvant susciter de l'émotion doivent toujours passer par l'évaluation primaire. Les événements qui ont peu de relations avec la motivation (déterminée via les deux critères de *pertinence motivationnelle* et de *congruence motivationnelle*) ne nécessitent plus de passer par le niveau secondaire de l'évaluation et donc il n'y a pas d'émotion en réponse à ces événements. En revanche, pour les événements ayant une forte relation avec la motivation, une évaluation secondaire est nécessaire pour déterminer quelle émotion associer à la réaction en réponse à ces événements.

Les auteurs ont essayé de définir le rôle des critères pour la production des émotions telles que *la colère, la culpabilité, l'anxiété, la tristesse,* et *l'espoir* pour

démontrer leur théorie (Table 3). Prenons l'exemple de la colère : selon le tableau, quand un événement est considéré comme défavorable à la motivation et que cet événement est sous la responsabilité des autres, la stratégie du faire-face (coping) pourrait générer soit un plan d'action du genre « enlever la source du dommage dans l'environnement afin de détruire la menace », soit une approche orientée émotion : il s'agit en fait de se mettre en colère parce que l'individu est victime d'une situation défavorable (i.e. accuser les autres de la mauvaise situation).

Table 3 Illustration de l'analyse fonctionnelle de quelques émotions

Emotion	**Plan d'action envisagé**	**Cause de l'émotion**	**Critères d'évaluation**
Colère	Enlever la source du dommage dans l'environnement afin de détruire la menace	Accuser les autres	1. Lié à la motivation
			2. Pas conforme à la motivation
			3. Responsabilité des autres
Culpabilité	Réparer le dommage pour les autres/Motiver le comportement social de responsabilité	Accuser soi-même	1. Lié à la motivation
			2. Pas conforme à la motivation
			3. Responsabilité de soi
Anxiété	Eviter les dommages potentiels	Menace/danger potentiel	1. Lié à la motivation
			2. Pas conforme à la motivation
			3. Potentiel d'adaptation (orienté émotion) faible/incertain
Tristesse	Demander de l'aide et du soutien pour faire face aux dommages/se débarrasser d'une perte	Perte irrévocable	1. Lié à la motivation
			2. Pas conforme à la motivation
			3. Potentiel d'adaptation (orienté problème) faible
			4. Espoir faible pour le futur
Espoir	Maintenir l'engagement et la	Possibilité d'amélioration/suc	1. Lié à la motivation

Espoir	stratégie d'adaptation	cès	1. Lié à la motivation
			2. Pas conforme à la motivation
			3. Espoir fort pour le futur

La théorie de Smith et Lazarus sur la stratégie d'adaptation via les deux critères de *potentiel d'adaptation orienté émotion* et de *potentiel d'adaptation orienté problème* est bien adaptée notamment pour expliquer l'attitude de l'humain dans les situations de stress.

1.4.3. Théories de Scherer sur l'évaluation d'un événement

Klaus Scherer, dans sa théorie (Scherer, 2009), propose que le processus émotionnel passe par quatre étapes d'évaluation consécutives, présentées dans la Figure 5. Le résultat de l'étape précédente doit être produit pour que l'étape suivante soit déclenchée. La première étape sert à déterminer si l'événement est pertinent pour l'intérêt de l'organe (comme les buts, les préférences, l'attention). Cette étape permet de décider si l'événement nécessite une analyse plus approfondie – ce qui implique de transmettre l'événement à l'étape suivante. La deuxième étape consiste à évaluer les conséquences de l'événement sur ses propres buts ou besoins. Si ces conséquences sont significatives et nécessitent la mise en place de stratégies d'adaptation, l'événement sera transmis à la troisième étape. A cette étape, en fonction des informations évaluées par les étapes précédentes sur l'événement, les stratégies d'actions sont proposées en fonction de la capacité de contrôle de l'organisme envers la situation. Si l'organisme n'est pas en mesure de contrôler/gérer les conséquences possibles de l'événement, il faut soit monter un plan d'action pour affecter l'environnement et donc changer les conséquences, soit réajuster son jugement sur l'événement afin d'accepter/nier ses conséquences. Cet ajustement possible est fait dans la quatrième étape où l'événement est évalué en fonction des normes personnelles et des normes sociales.

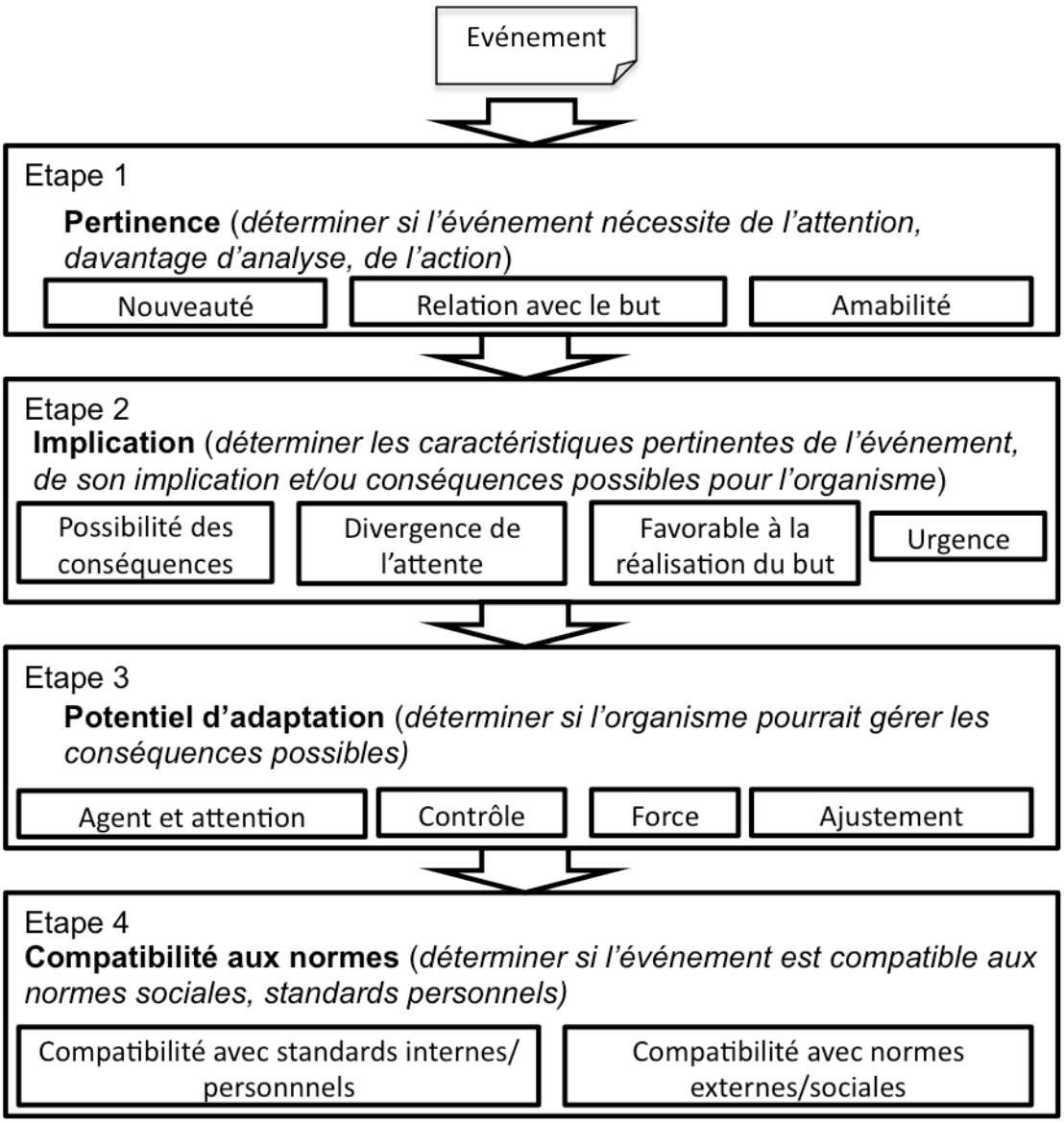


Figure 5 Quatre couches d'évaluation d'un processus émotionnel, (Scherer, 2009).

Pour chaque étape, des réactions physiques/physiologiques sont possibles en fonction de l'évaluation de l'étape. Par exemple, à l'étape 1, la valeur de l'amabilité est suffisante pour déclencher un geste de salutation/approche si l'événement est aimable. Ou bien dans l'étape 2, si l'événement est considéré comme urgent, un geste de recul ou de défense pourrait être déclenché.

Scherer a inclus dans sa proposition les critères proposés par Ortony et al. et ceux proposés par Smith et Lazarus. De plus, il a ajouté la couche de traitement *Compatibilité aux normes*, soulignant le fait que l'émotion est à la fois universelle et personnalisée.

1.5. Conclusion

Cette section résume les trois traditions en psychologie sur la définition de l'émotion humaine. Ces trois traditions peuvent être vues comme les études sur de

différentes facettes d'un seul objet : Emotion humaine. De manière générale, les théories sur les familles des émotions s'intéressent à construire une spécification des caractéristiques des émotions, i.e. les symptômes observables de l'émotion chez l'être vivant. Les théories des constructivistes, quant à elles, s'intéressent à la mise en correspondance entre les situations et les réactions émotionnelles, en particulière les mesures dites "affectives" de ces réactions émotionnelles. Les grandes dimensions affectives peuvent être citées comme la valence, l'activation. Les théories des processus émotionnels, par contre, investissent beaucoup plus sur la façon dont les informations sont traitées, sur les critères pris en compte lors d'une évaluation émotionnelle. Chaque tradition a ses propres avantages qui attirent des chercheurs dans les domaines d'applications, comme l'informatique, la robotique. Ces traditions participent au fondement des modèles computationnels des émotions.

Etant intéressé par l'aspect fonctionnel de l'émotion, i.e. le processus d'évaluation de l'émotion, nous nous basons notre modèle des émotions sur la troisième tradition "processus émotionnel chez l'humain" pour pouvoir modéliser de différents phénomènes émotionnels de l'humain (comme les reflex, l'humeur, la personnalité). La plupart des modèles informatiques sont inspirés de cette tradition pour la même raison. Les trois théories psychologiques les mieux citées dans les travaux informatiques sont celles présentées précédemment dans la section 1.4. Bien que le modèle OCC et la théorie de R. Lazarus ont été bien exploités dans les modèles informatiques existants, la théorie proposée par K. Scherer n'est utilisé que partiellement, i.e. l'utilisation de sous ensemble des critères d'évaluation listés par K. Scherer. Intéressée par la généricité de la théorie de K. Scherer, j'ai proposé, au cours de mon travail de stage de Master 2 recherche, un premier modèle computationnel des émotions tenant en compte les éléments essentiels du processus émotionnel abordés dans cette théorie . La description de ce modèle est présentée dans la section suivante.

2. Modèle GRACE – Conception initiale

GRACE est un modèle des émotions qui permet de simuler le processus émotionnel défini dans la théorie de Scherer. Selon cette théorie, ce processus émotionnel caractérise la réponse du corps humain à un événement.

Un *événement* est une perception par le corps humain d'un changement (ou d'une absence de changement) dans l'environnement ou d'un changement interne (ou d'une absence de changement) dans le corps humain. L'absence de changement peut engendrer une réponse émotionnelle quand un changement est attendu mais n'intervient finalement pas. Cette absence de changement va donc générer une réaction de l'individu à la situation inattendue.

Une *réponse du corps humain* correspond à un ou plusieurs changements internes (indétectable par un observateur extérieur) dans le corps humain, ou une expression ou posture du corps extérieur (détectable par un observateur) ou une absence de changement.

Donc, à partir d'une sensation qu'un événement a eu lieu, une première réaction consiste notamment à montrer son attention pour l'événement ou de répondre à la nouveauté de l'événement, ce qui correspond à la première étape dans la théorie de Scherer. L'analyse dans les étapes suivantes implique des traitements additionnels pour déterminer les conséquences de l'événement sur l'environnement, l'implication

de l'événement sur les buts, ce qu'on appelle souvent l'analyse cognitive. Cette analyse participe ensuite à la délibération de la réponse émotionnelle par l'individu. Cette réponse dépend non seulement de l'analyse cognitive mais aussi de l'état interne courant comme l'humeur, la personnalité, les motivations. La réponse émotionnelle et l'expression associée vont être enregistrées dans la mémoire comme l'expérience personnelle qui permet d'anticiper ou de s'adapter aux futures situations qui lui ressemblent. La capacité d'anticiper l'avenir permet aussi de créer des sensations virtuelles qui ressemblent à la façon dont l'imagination active des changements physiologiques (comme les battements du cœur, la température du corps) tandis que rien ne se passe dans la réalité. Nous considérons aussi cette capacité comme une partie du processus émotionnel à modéliser.

Sur la base de ces définitions, nous avons proposé un modèle d'émotions de huit composants synthétisé dans la figure suivante :

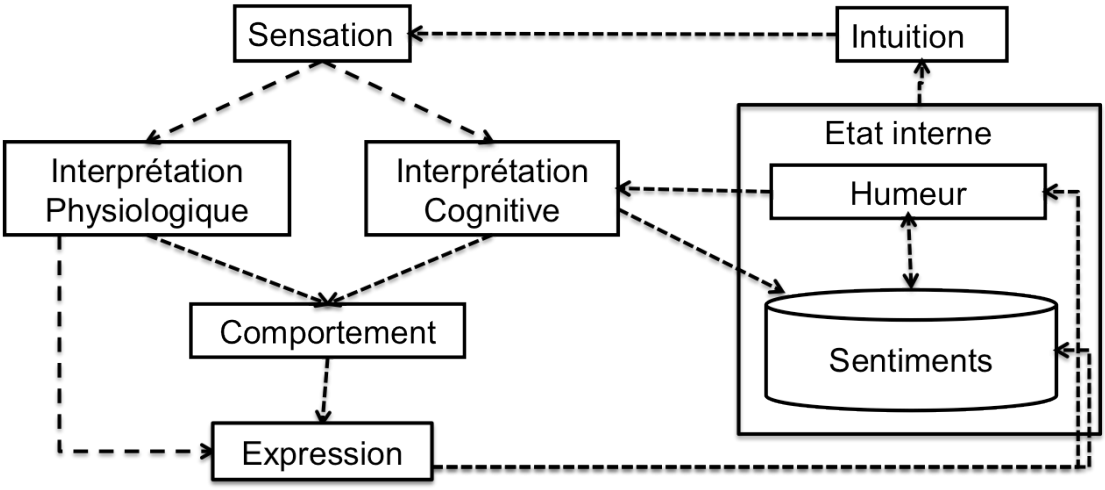


Figure 6 Architecture de GRACE

Processus émotionnel

Dans ce modèle, *Sensation* est le point de départ. La sensation est générée par un événement, quelque chose qui existe ou n'existe pas mais qui génère un changement physiologique dans le corps. Cette sensation est traitée en deux niveaux parallèles.

D'abord, *Interprétation Physiologique* transforme ce signal initial directement en réponse du corps (battement du cœur, pression sanguine, etc.) et alarme le niveau de *Comportement*.

Interprétation Cognitive transforme ce signal de sensation en information cognitive sur la situation de l'environnement qui est traitée par *Comportement*.

Comportement calcule la réponse à l'information venant du niveau des interprétations (physiologique et cognitive) en se basant sur *Etat Interne*. Cette réponse est ensuite envoyée à *Expression* où la réaction physique a lieu.

Connaissances/Expériences joue le rôle de la mémorisation et de la mise à jour des expériences personnelles – qui permettent à s'adapter aux nouvelles conditions d'interaction.

Description détaillée

A. Sensation

Ce composant est l'endroit où naissent les émotions. On peut considérer (selon l'idée de Scherer) que la composante *Sensation* scanne en permanence l'environnement du corps et l'état interne du corps. À un moment quelconque, un changement est détecté ; une sensation est née. Cette sensation peut venir (selon la théorie d'Ortony et al) d'un événement, d'une action d'un agent ou un aspect d'un objet. Cette sensation est envoyée ensuite à deux modules d'interprétation (physiologique et cognitive). Chacun de ces deux modules peut inhiber cette entrée si son niveau est considéré comme trop faible.

Le module *Sensation* peut être vu comme un représentant des cinq sens humains (la vue, le toucher, l'ouïe ou l'audition, l'odorat, le goût). Du point de vue d'une implémentation informatique, il s'agirait d'un système de récepteurs capable de capturer et de traduire plusieurs formes d'énergies (stimuli) et de les analyser pour en permettre la perception (ou la détection d'un changement dans l'environnement). Dans le contexte de notre application en robotique, les sens seront décrits classiquement par des capteurs.

La notion de sens couvre deux aspects bien différents suivant que l'on est en présence d'une communication immédiate (donc instinctive) ou médiate (donc rationnelle). Les sens ne sont pas uniquement des transducteurs permettant la mesure de paramètres. Les sens sont les instruments de la perception, c'est-à-dire le lien qui relie l'organisme au monde extérieur et qui lui permet de faire naître une sensation. La transformation entre l'activation d'un sens et la sensation générée dans le corps est ainsi potentiellement différente selon les personnes, en fonction de leur sensibilité.

B. Interprétation physiologique

En répondant à une sensation, une sortie au niveau du corps pourrait être immédiate. Le fonctionnement de cette composante correspond à l'évaluation du critère *Urgence* dans la deuxième étape décrite dans la théorie de Scherer. La perception est ici « instantanée », on peut considérer qu'il s'agit d'un événement externe ou d'une transformation interne d'un événement externe. Par exemple un son très violent déclenche une saturation du système de perception de l'oreille. Cette information provoque un sursaut de l'individu par son arc réflexe (ensemble constitué par la transmission d'une information sensitive (stimulation) vers un centre nerveux (notamment dans la moelle), ce centre et la transmission de la réponse (motrice notamment) de ce centre aux organes effecteurs).

Dans notre modèle, l'interprétation physiologique analyse l'information du module de *Sensation* pour calculer la réponse immédiate et fournir cette réponse au module *Expression.* Cette réponse peut aussi être transformée en information fournie comme entrée au module de *Comportement*.

C. Interprétation cognitive

L'interprétation cognitive est un filtrage des sensations. Elle transforme la sensation au niveau sémantique. Un sens peut être attaché à cette sensation. C'est la première partie de l'interprétation d'une sensation. La seconde partie est basée sur la croyance, la nouveauté, et la concordance avec la relation entre les standards et les buts personnels.

Ce type d'interprétation prend « du temps », on considère alors qu'un processus cognitif se met en place. Celui-ci va dépendre de plusieurs facteurs. Tout d'abord « l'état cognitif interne » de l'individu qui se caractérise par son humeur et ses sentiments. L'humeur dépend de l'histoire de l'individu, de sa fatigue et de sa concentration, il règle la perception de l'événement en fonction de sa valence. Les sentiments caractérisent un autre état interne de l'individu. Cet état est plus complexe et plus difficile à cerner. En effet bon nombre de choses dites sur les émotions sont applicables aux sentiments. Les sentiments vont avoir comme effet de perturber l'interprétation en donnant des valeurs positives ou négatives à des événements.

D. Comportement

Le *Comportement* calcule la réponse émotionnelle que le corps doit fournir à une perception. En effet, à partir des résultats venant de l'*Interprétation Physiologique* et de l'*Interprétation Cognitive*, ce composant détermine dans un premier temps une valeur émotionnelle comme réponse à l'événement. Cette valeur émotionnelle reflète en général l'émotion de l'individu en réponse à la situation, en fonction de sa connaissance (via les *Interprétations*) et de son état actuel (dont l'état mental, les motivations, la personnalité). Dans un deuxième temps, ce composant réalise la stratégie de « faire-face » (coping) proposée par Smith et Lazarus en fonction de la valeur déterminée initialement. Si l'événement est en faveur de l'individu (en rapport notamment avec ses intentions et ses préférences), le *Comportement* produira une émotion positivement corrélée avec l'état désiré de l'individu. Si l'événement n'est pas en faveur de l'individu, l'émotion en réponse sera négativement corrélée avec l'état désiré de l'individu. Le dernier cas déclenchera le processus de coping, soit pour déclencher une action en vue de changer la situation, soit pour changer un standard personnel, un intérêt, ou des préférences de manière à avoir une évaluation plus positive sur la situation.

E. Humeur

C'est là que l'état émotionnel courant est stocké. Elle a une influence sur la décision prise par le *Comportement*. Elle inclut la prise de position (combattre, s'enfuir, aider, aimer, etc.), les états mentaux (motivation, intérêts, extraversion/introversion, etc.), les états physiques (fatigue, anxiété, etc.).

F. Sentiments

C'est le méta-niveau dans lequel l'interprétation cognitive, le comportement et l'action du corps sont mémorisés. Ce niveau analyse et construit l'expérience personnelle. En fonctionnement, cette composante pourrait donner de l'information au autres composantes, comme un sentiment d'une situation qui est déjà expérimentée,

un sentiment que c'est une bonne direction, un sentiment inconfortable car la situation n'est pas celle espérée, un sentiment joyeux car tout est contrôlable, etc. Ces informations aident à calculer les états affectifs appropriés par rapport à une situation donnée.

G. Intuition

Ce module est utilisé pour créer une sensation quand rien ne s'est réellement passé dans l'environnement. Cette intuition est basée sur l'état cognitif interne et sur les expériences personnelles. L'intuition peut être vue comme une conséquence des sentiments. Les sentiments analysent la situation et prédisent une sensation via les connaissances acquises. L'intensité de cette prédiction peut générer une sensation réelle. Du point de vue informatique, ce niveau pourra être implémenté par une analyse statistique des sensations déjà détectées dans un contexte spécifique, couplée à un algorithme de prédiction. Un deuxième niveau d'intuition sera obtenu par l'association des séquences de sensation avec une transformation homomorphique sur les sensations passées.

H. Expression

L'*Expression* est le lieu où l'émotion est exprimée. Cette expression peut être interne au corps et être responsable par exemple de l'augmentation du rythme du cœur et de la pression du sang, parmi d'autres changements physiologiques, traduisant l'excitation ressentie à mesure de l'augmentation de l'adrénaline. L'expression peut aussi être externe et se traduire par des changements dans l'expression du visage, de la voix, la posture, la sueur, etc.

3. GRACE : vers un modèle computationnel

A l'instanciation du modèle GRACE initial, plusieurs modifications de la structuration du modèle se sont avérées nécessaires. Dans le cadre de la thèse, nous nous sommes intéressés plus particulièrement à l'interprétation musicale et à l'expression robotique des émotions, ce qui imposait d'instancier les modules d'*Interprétation cognitive* et d'*Expression* en fonction de ces domaines applicatifs particulier. Il nous est alors apparu la nécessité de séparer clairement dans le modèle les mécanismes spécifiques de modalités particulières de perception ou d'expression de ceux qui sont communs à tout processus émotionnel. Tandis que l'instanciation des premiers devra tenir compte du contexte applicatif visé, on peut espérer proposer à terme un modèle et une implémentation « universels » pour les seconds. Nous avons donc proposé des modifications dans la structuration de GRACE pour lui donner plus de flexibilité et de généricité par rapport à sa conception initiale. Ces modifications sont présentées dans les paragraphes suivants.

Le terme de *Sentiments* dans le modèle GRACE initial a été repris tel quel de la proposition de Scherer sur le processus émotionnel chez l'humain. Le composant *Sentiments*, selon Scherer, est en charge de mémoriser le passé, de construire l'expérience émotionnelle, de développer la capacité d'être conscient de l'émotion vécue. Du point de vue de l'implémentation informatique, il s'agit pourtant moins

d'une conscience que d'une mémoire de l'expérience passée. Ce qu'un programme informatique peut faire, c'est mémoriser les événements et les suites de traitements effectuées par le système, puis actualiser, grâce à un processus d'apprentissage, une connaissance cohérente avec ses comportements dans le passé. Pour ces raisons, le nom de ce composant est devenu *Connaissance/Expériences*.

L'*Humeur* dans le modèle GRACE initial est aussi le nom repris depuis la proposition de Scherer sur le processus émotionnel chez l'humain. Ce composant s'occupe de l'état affectif courant, de la prise de position, de la caractéristique individuelle (i.e. de la personnalité). Pour faciliter l'implémentation, le nom de ce composant est devenu *Etat interne*.

L'*Intuition* dans le modèle GRACE initial représente la capacité de création/anticipation des événements dans l'avenir. Le composant *Intuition* peut ainsi créer des événements imaginaires de manière à ressentir l'émotion avant que l'événement réel ne se passe. Par exemple, quand un enfant attend que son père le récupère après l'école, et si le père est en retard, l'enfant peut se sentir triste car l'événement attendu ne survient pas. De plus, l'enfant peut aussi ressentir de la peur s'il imagine des choses terribles qui pourraient arriver à son père. Dans ce dernier cas, il n'y a pas d'événements réels mais l'imagination a déclenché le traitement d'information du processus émotionnel. L'implémentation de la création de l'événement imaginaire est beaucoup plus compliquée que l'implémentation de l'effet affectif de l'événement. De plus, l'implémentation de l'effet affectif de l'événement imaginaire simplifie l'étape *Perception* parce que du point de vue de l'implémentation, l'effet affectif de cet événement est connu. Pour ces raisons, une *Intuition* est de même nature qu'une *Perception* dont on a analysé l'effet affectif, et est donc maintenant redirigée directement vers la composante *Comportement*.

Il y a aussi un changement dans le fonctionnement du composant *Interprétation Cognitive*. A la conception, ce composant était destiné à filtrer/évaluer l'événement capturé par le composant *Sensation*. Cette évaluation prend en compte l'humeur pour simuler l'influence de l'état affectif courant et la personnalité sur la perception d'un événement. Cet effet est doublé dans la connexion entre le composant *Comportement* et le composant *Etat interne*. La connexion entre le composant *Interprétation Cognitive* et le composant *Etat interne* est donc enlevée.

Dans la version initiale de GRACE, le composant *Interprétation Cognitive* envoie aussi sa sortie au composant *Sentiments* pour la mise à jour des expériences et l'enregistrement dans la mémoire. Etant donné que la réponse émotionnelle est déterminée par le composant *Comportement* et que l'*Interprétation Cognitive* ne fait qu'une assistance à la décision du *Comportement*, cet envoi est remplacé par l'envoi des résultats du composant *Comportement* au composant *Connaissances/Expériences* (auparavant nommé *Sentiments*).

Dans le but de simplifier l'implémentation, une unification des informations échangées entre les composants est aussi proposée. Comme ont proposé les constructivistes, le résultat de l'analyse affective des différents aspects de l'émotion, comme l'humeur, l'interprétation cognitive, la réponse émotionnelle (ou *core affect*), peut être représenté par le couple Valence - Activation. Du point de vue computationnel, cette proposition est intéressante car elle permet un développement indépendant d'un composant à l'autre sans devoir s'inquiéter du format des messages échangés entre les différents composants. De même, du point de vue de la représentation du contenu émotionnel dans la musique, c'est le couple Valence –

Activation qui est le plus souvent utilisé pour la représentation des émotions musicales. Du point de vue des mouvements robotiques enfin, le couple Valence – Activation facilitera l'implémentation des déplacements réels du robot en fonction de l'émotion, surtout dans le contexte de l'écoute musicale. Pour les trois raisons énumérées ci-dessus, nous avons décidé d'unifier les messages échangés entre les composants en couples Valence – Activation. Par exemple, pour le composant *Interprétation physiologique,* l'analyse affective sur l'aspect « Urgence » peut être évaluée en répondant aux questions du niveau de favorabilité et du niveau d'activation auquel l'événement affecte l'individu. Pour le composant *Interprétation Cognitive,* l'analyse de l'événement donne une mesure de la favorabilité de l'événement par rapport aux buts, standards, préférences de l'individu, et une mesure du niveau d'intensité avec lequel l'événement influence ces aspects. La sortie de *Comportement* peut être aussi représentée en employant cette représentation des émotions en deux dimensions Valence – Activation. Une expérience émotionnelle peut quant à elle être représentée comme une suite de couples (Valence, Activation) représentant l'état de sortie de chaque composant dans le processus émotionnel lors du traitement d'un événement.

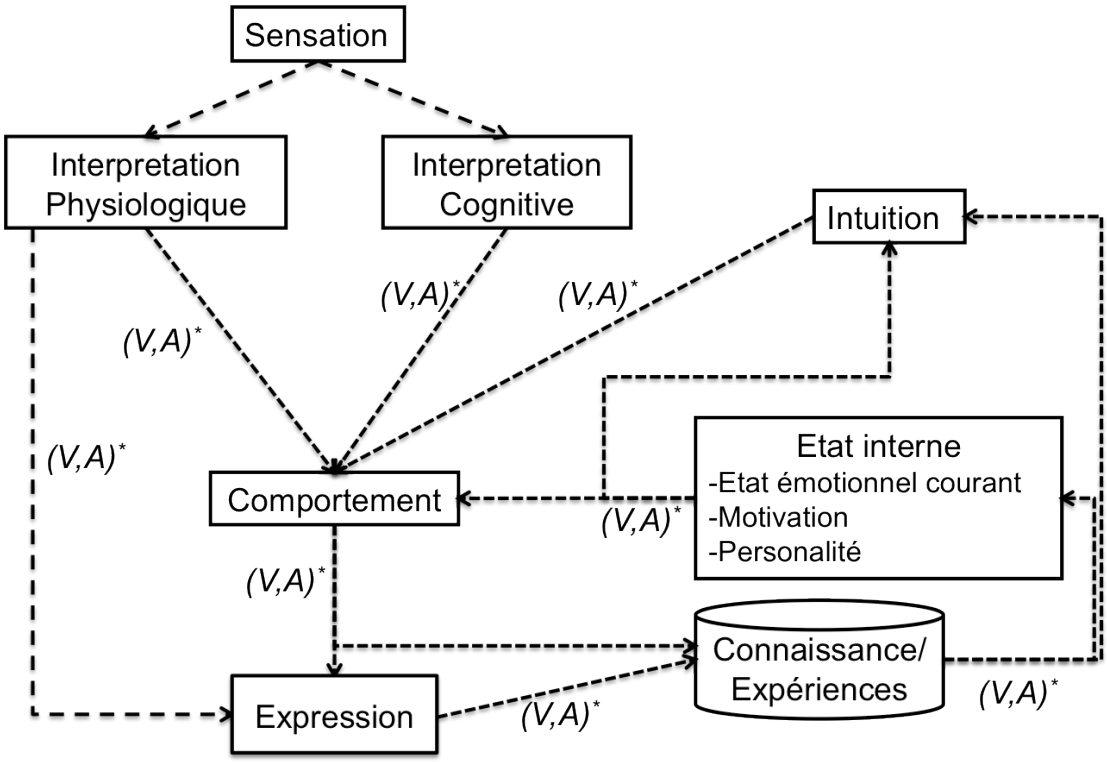


Figure 7 GRACE à l'état actuel

L'unification des informations échangées entre les composantes clarifie aussi le processus émotionnel que modélise la version initiale de GRACE. Les psychologues ont bien souligné le rôle indispensable de l'émotion pour la survie et le développement individuel. Ce rôle réside dans le fait que l'émotion, au cours de son processus d'évaluation des événements, influence (et parfois change) le plan d'action de l'individu. Pourtant, il n'est pas clair si c'est le résultat du processus émotionnel qui affecte la planification ou si c'est le résultat de la planification qui affecte le

processus émotionnel. Avec l'adaptation que nous avons apportée dans le modèle GRACE, nous ne traitons que l'aspect émotionnel, ce qui implique que le processus émotionnel dans GRACE ne s'occupe pas de la planification ni de tout ce qui est lié à la prise de décisions. L'information sur la planification est implicitement fournie là où le processus émotionnel en a besoin, par exemple dans le composant *Comportement* pour faire du coping, dans le composant *Etat interne* pour déterminer la motivation, ou dans le composant *Intuition* pour anticiper l'avenir.

Dans l'ambition de construire un modèle générique, englobant un grand nombre de modèles existants, nous allons ensuite établir une comparaison entre le modèle GRACE et différents modèles computationnels pour examiner de quelle manière les processus émotionnels associés à ces modèles peuvent être pris en compte par GRACE (pour simplifier l'écriture, dans la suite du mémoire, le mot GRACE est utilisé pour indiquer le modèle GRACE dans sa version actuelle, i.e. celui qui contient des modifications effectuées lors de la thèse).

4. GRACE par rapport aux modèles informatiques récents

Cette section décrit la comparaison entre GRACE et différents modèles computationnels des émotions présentés par ordre chronologique. Ce sont les modèles le plus souvent cités dans les recherches sur la modélisation des émotions en informatique. Ce n'est pas une liste exclusive mais la comparaison donne les premiers arguments en faveur de la généricité du modèle GRACE par rapport à ce qui existe dans la littérature.

4.1. Affective Reasoner

Affective Reasoner est une plate-forme de simulation ayant pour but de simuler l'interaction sociale, implémentée dans un programme d'entrainement pour les nouveaux vendeurs (Elliott, 1993). Ce programme simule plusieurs clients virtuels interagissant avec le vendeur humain.

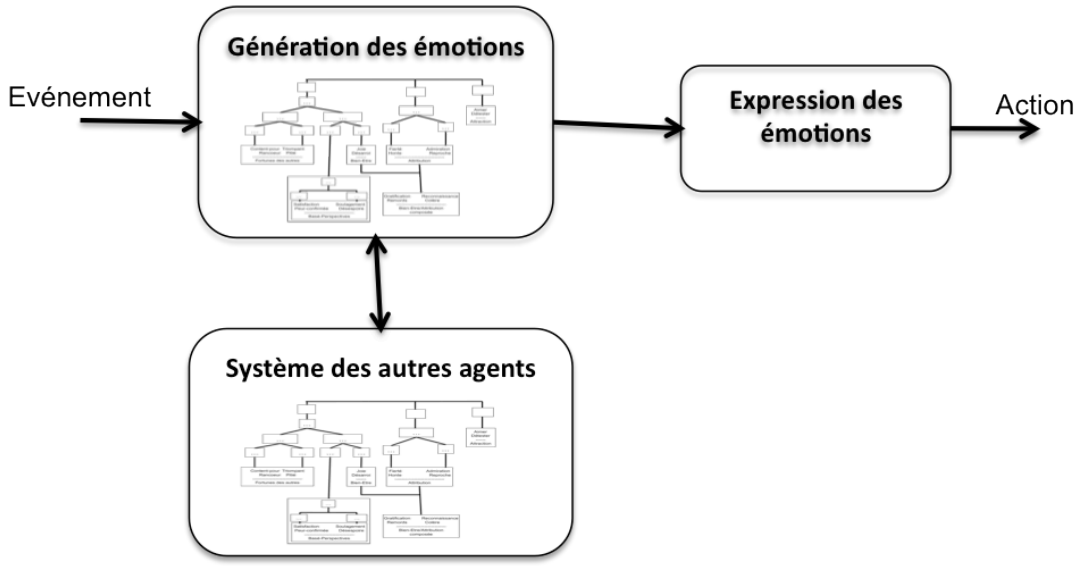


Figure 8 Principe de fonctionnement d'Affective Reasonner (Elliott, 1993)

Affective Reasoner est reconnu comme le premier travail qui intègre totalement le modèle d'Ortony et al., qu'il utilise pour construire un ensemble de règles de production de l'émotion. Un agent est constitué de trois composantes principales : le *générateur de l'émotion*, l'*expression émotionnelle*, et le *système des autres agents*. Le *générateur de l'émotion* utilise un ensemble de règles de production de l'émotion (EECRs – Emotion Eliciting Condition Relations en anglais) basées sur la proposition d'Ortony, Clore et Collins. L'*expression émotionnelle* est la composante en charge de relier l'émotion aux comportements appropriés. Le *système des autres agents* prend l'émotion exprimée par les autres agents pour reconstruire les personnalités/préférences des autres agents, et donc aider le *générateur de l'émotion* dans le processus d'application des règles (dans la proposition d'Ortony et al, un agent a besoin de ces informations pour faire le raisonnement sur *les conséquences de l'événement* et *l'action de l'agent*).

Chaque agent *client* intègre une instance d'Affective Reasoner permettant d'exprimer ses émotions au cours de l'interaction avec l'humain. Chaque agent a aussi des buts, des standards, des préférences, et une humeur représentant le personnage qu'il joue au cours de la simulation.

Affective Reasoner pourrait facilement être implémenté dans GRACE (Figure 9). Le processus de mise en correspondance entre l'événement et l'émotion appropriée correspond au travail du module *Interprétation Cognitive*. Le fonctionnement de ce composant est donc l'évaluation de l'événement en fonction de trois aspects : ses conséquences, les actions liées, et les caractéristiques de l'événement. L'application des règles EECRs s'effectue dans le composant Interprétation Cognitive sans prendre en compte l'impact de la personnalité de l'agent (exprimée via ses buts, ses préférences, et ses standards). La personnalité de l'agent est gérée par le composant *Etat interne* et est utilisée dans le composant *Comportement* pour filtrer l'émotion à exprimer. L'émotion résultant de *Comportement* est ensuite transmise au composant *Expression* pour choisir l'action appropriée et l'exécuter.

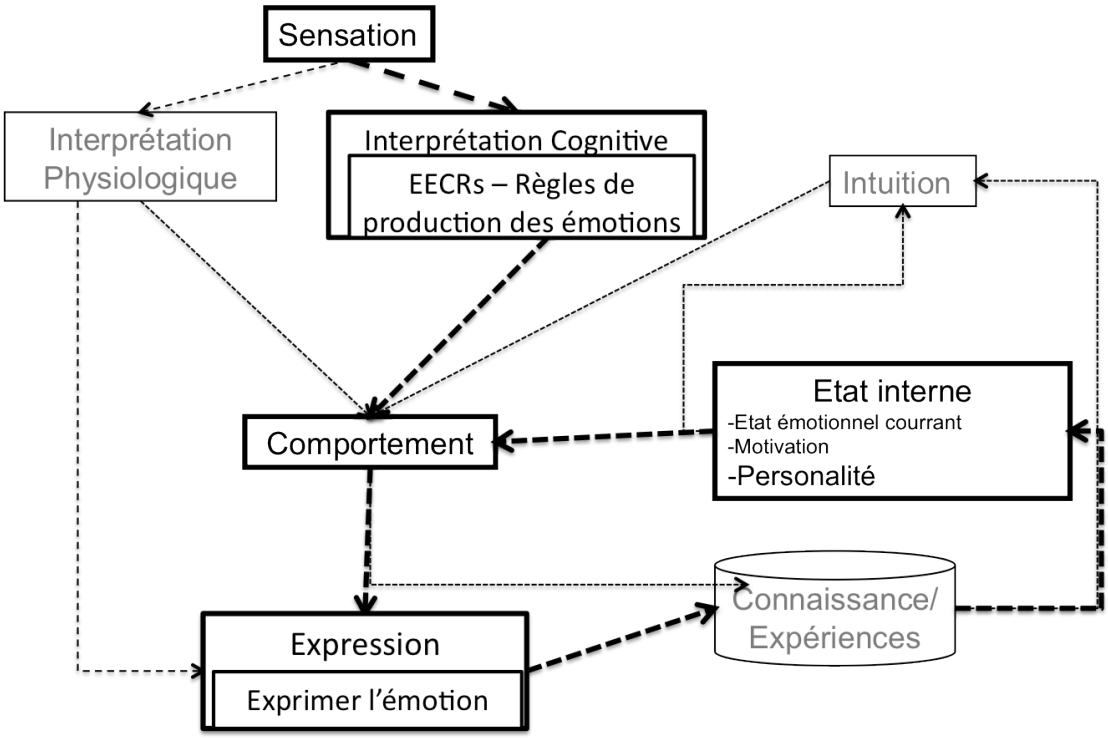


Figure 9 Adaptation de *GRACE* pour simuler *Affective Reasoner*

On peut remarquer la différence dans le principe de mise en correspondance des émotions dans les fonctionnements de GRACE et d'Affective Reasoner. Dans Affective Reasoner, la personnalité (exprimée via les buts, les préférences et les normes) est impliquée dans les règles de mise en correspondance entre l'événement et l'émotion. Par exemple, avec Affective Reasoner, un personnage dominant (i.e. une personne qui préfère des décisions logiques et rapides) est défini comme quelqu'un qui s'intéresse aux conséquences de l'événement mais pas aux deux autres aspects définis dans OCC. Ce personnage peut facilement se mettre en colère si l'événement n'est pas en faveur de ses buts. Un personnage social considère l'événement plutôt en fonction de l'action des autres (i.e. l'aspect *action de l'agent*) et l'aspect de l'objet. GRACE peut prendre en charge les différentes caractéristiques d'Affective Reasoner par le fait que le composant *Interprétation Cognitive* mène une analyse objective sur l'événement – i.e. l'événement est évalué selon les trois aspects du schéma OCC. Le composant *Comportement* est ensuite chargé de faire le choix de l'aspect le plus en adéquation avec la personnalité simulée. Dans GRACE, c'est donc le moment du choix de la réponse émotionnelle qui est modifié, pas la réponse elle-même.

De plus, l'adaptation de GRACE est plus flexible car elle permet de prendre en compte l'état affectif interne ainsi que les réflexes (i.e. réactions immédiates) qui ont de l'influence sur la réponse émotionnelle.

4.2. Cathexis

Cathexis est un modèle des émotions basé sur la théorie des émotions basiques (Velasquez, 1997). Dans Cathexis, les émotions sont classifiées en familles, chaque famille se caractérisant par des événements spécifiques d'activation, des comportements appropriés et des seuils d'activation différents.

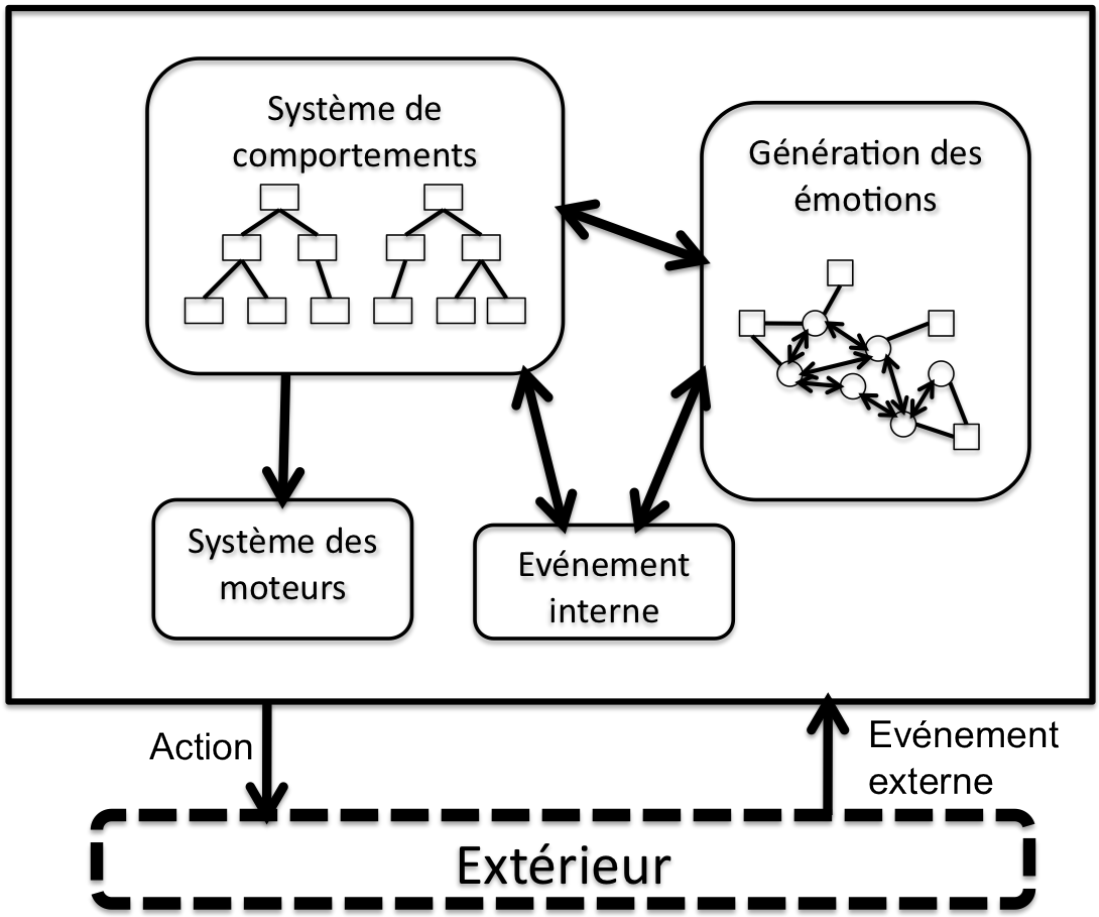


Figure 10 Architecture de Cathexis (Velasquez, 1997)

L'architecture de Cathexis se compose d'un système de génération des émotions, d'un système de comportement, d'un système moteur, et d'un système de perception des stimuli. Le système de génération des émotions est en charge de calculer l'intensité des émotions en fonction des stimuli que le système de perception a détectés. Le système de comportement active ensuite les comportements correspondant aux émotions activées par le système de génération des émotions. Enfin, le système moteur est en charge d'exécuter le(s) comportement(s) choisi(s) par le système de comportement. A un instant donné, le système de génération des émotions peut gérer plusieurs émotions selon les événements capturés (par exemple, dans une situation de stress, le système peut générer à la fois la colère et la peur). Cathexis met également en œuvre l'effet inter-émotionnel, c'est-à-dire que deux émotions peuvent s'inhiber/se renforcer. La colère inhibe la joie, et peut renforcer la peur. Comme il est possible que plusieurs émotions coexistent, le système de comportement peut mettre les comportements suscités par des émotions coexistantes en compétition les uns avec les autres, et choisir le gagnant comme comportement à exécuter par le système moteur.

Dans Cathexis, les événements suscitant de l'émotion sont classifiés en quatre catégories : *Neuronal*, *Sensorimoteur*, *Motivationnel*, et *Cognitif*. Cette idée est inspirée de la proposition d'Izard sur les types d'événements suscitant de l'émotion. L'implémentation de ces quatre types d'événements dans Cathexis donne une grande flexibilité au système pour simuler différentes situations dans lesquelles sont suscitées les émotions humaines.

Le terme « événement » utilisé correspond au type de traitement appliqué pour faire naître une émotion. Les caractéristiques de ces types d'événements/traitements motivent une adaptation de GRACE pour implémenter Cathexis comme suit (Figure 11) :

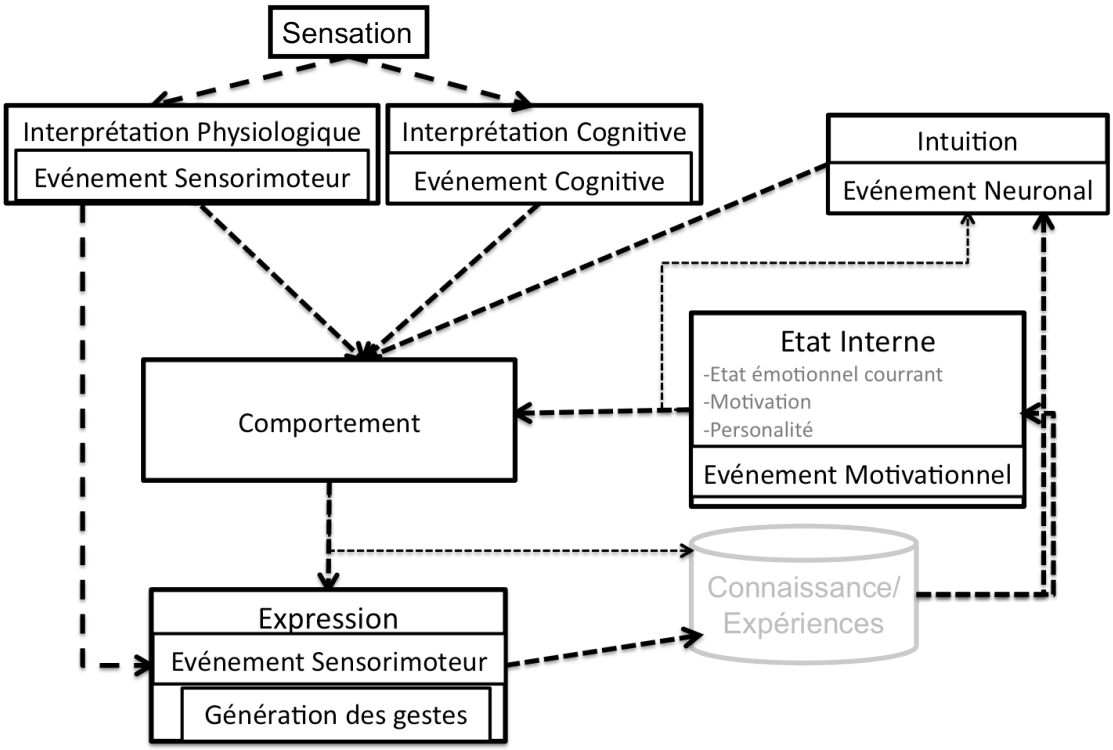


Figure 11 Adaptation de *GRACE* pour simuler *Cathexis*

Les événements de type neuronal sont générés/simulés par la composante *Intuition* de GRACE – qui est en charge de simuler/anticiper les stimulations futures - ce qui ressemble à l'activation neuronale. Par exemple, quand un bébé a faim, et qu'il voit sa maman, il s'agit pour lui d'un événement positif. Cet événement engendre une réponse émotionnelle positive du bébé. De plus, cet état émotionnel positif peut activer l'anticipation du bébé que la maman va lui apporter quelque chose à manger. Cette anticipation aura lieu dans *Intuition* qui envoie à la composante *Comportement* un état positif qui est en fait imaginaire. Cette anticipation est l'événement de type neuronal.

Les événements de type Sensorimoteur correspondent à l'activation des mécanismes proprioceptifs (incluant donc les réflexes et le retour proprioceptif). Ces événements sont donc intégrés dans GRACE par la connexion du composant *Sensation* jusqu'au composant *Expression* via le composant *Interprétation Physiologique*, et par la connexion de *Comportement* vers Etat interne.

Les événements de type motivationnel sont implémentés dans la composante *Etat interne* comme défini dans la conception de GRACE.

Les événements cognitifs correspondent à l'analyse cognitive de GRACE – ce qui se traduit par la connexion du composant *Sensation* au composant *Interprétation Cognitive*.

Le composant *Comportement* de GRACE s'occupe ensuite du calcul de l'intensité de l'émotion en sortie. Le conflit entre les émotions en concurrence est aussi à résoudre par *Comportement*. Seule l'émotion gagnante va être transférée au composant *Expression*.

Avec GRACE, il est possible de simuler différentes personnalités en changeant la façon dont chaque type d'événement affecte le calcul de l'intensité émotionnelle du composant *Comportement*. Cet avantage facilite significativement la conception d'agents de personnalités différentes.

4.3. FLAME

FLAME est un modèle des émotions pour agents virtuels proposé par (El-Nars, Yen, & Ioerger, 2000). Le modèle est composé de trois parties : Emotion (EC - Emotional Component en anglais), Comportement (DC - Decision-making Component en anglais), et Apprentissage (LC - Learning Component). Quand un événement se produisant dans l'environnement est capturé par la composante Comportement, il est transmis aux composantes Emotion et Apprentissage pour être analysé. La composante Apprentissage met à jour les attentes et les associations Evénement-But de l'agent en fonction de la situation actuelle (perçue via l'événement). La composante Emotion analyse l'événement en fonction des buts de l'agent pour déterminer le comportement émotionnel à sélectionner pour répondre à la situation. Ce comportement est ensuite transmis à la composante Comportement pour l'exécution.

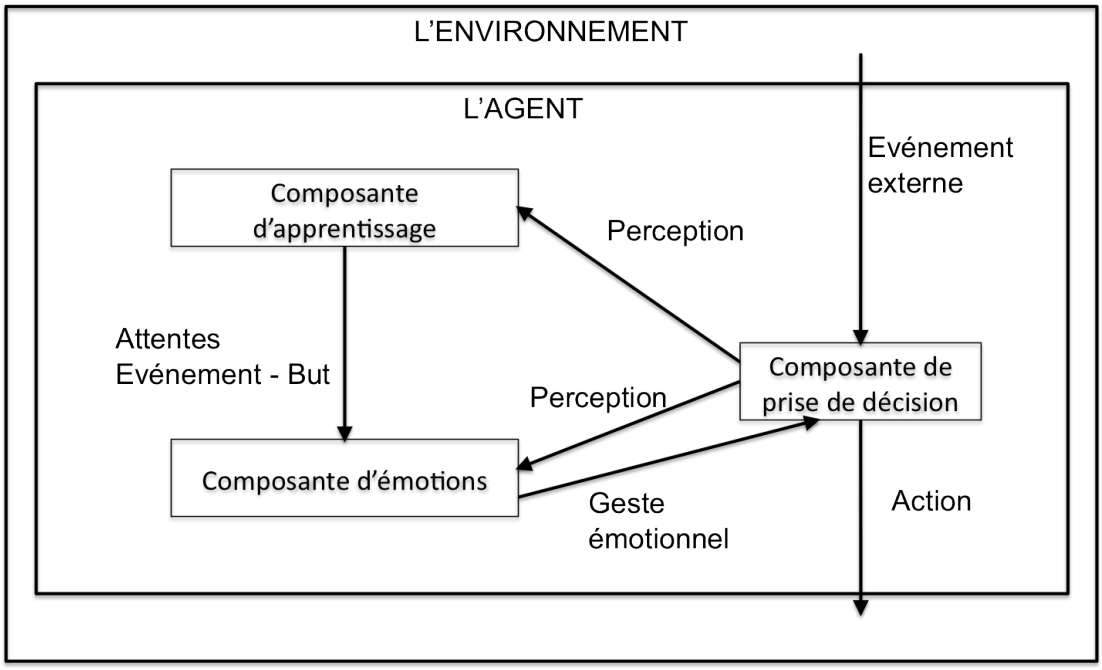


Figure 12 Architecture de FLAME (El-Nars, Yen, & Ioerger, 2000)

Le processus émotionnel dans la composante Emotion (voir Figure 13) commence par l'évaluation du niveau de désirabilité de l'événement en fonction des buts de l'agent.

Puis une mise en correspondance est déclenchée pour déterminer la(les) émotion(s) appropriée(s) à l'événement. Cette mise en correspondance est basée sur le modèle OCC. Les émotions en sortie de la mise en correspondance sont ensuite filtrées en fonction des besoins physiologiques (i.e. les états motivationnels, comme la faim, la soif, le sommeil) et en fonction de l'humeur actuelle de l'agent. Un état motivationnel fort peut inhiber une émotion ou vice versa. Par exemple, une faim très forte peut inhiber la peur d'être dans un lieu qui n'est pas familier. Et inversement, si la faim est faible, la peur peut être suscitée facilement. Après avoir choisi l'émotion appropriée à l'événement, une réaction est choisie pour exprimer l'émotion. Le choix de l'action dépend aussi de l'intensité de l'émotion, de l'événement et de l'interlocuteur. Différentes actions sont définies notamment en fonction de l'intensité de l'émotion et en fonction de l'association de différentes émotions. Une peur très forte sera ainsi traduite en une action différente selon qu'elle s'exprime seule ou en association avec une colère modérée. Dans FLAME, l'émotion diminue également au cours du temps. FLAME définit deux taux de diminution, l'un pour les émotions positives, et l'autre pour les émotions négatives, avec l'idée que les émotions négatives sont plus persistantes que les émotions positives.

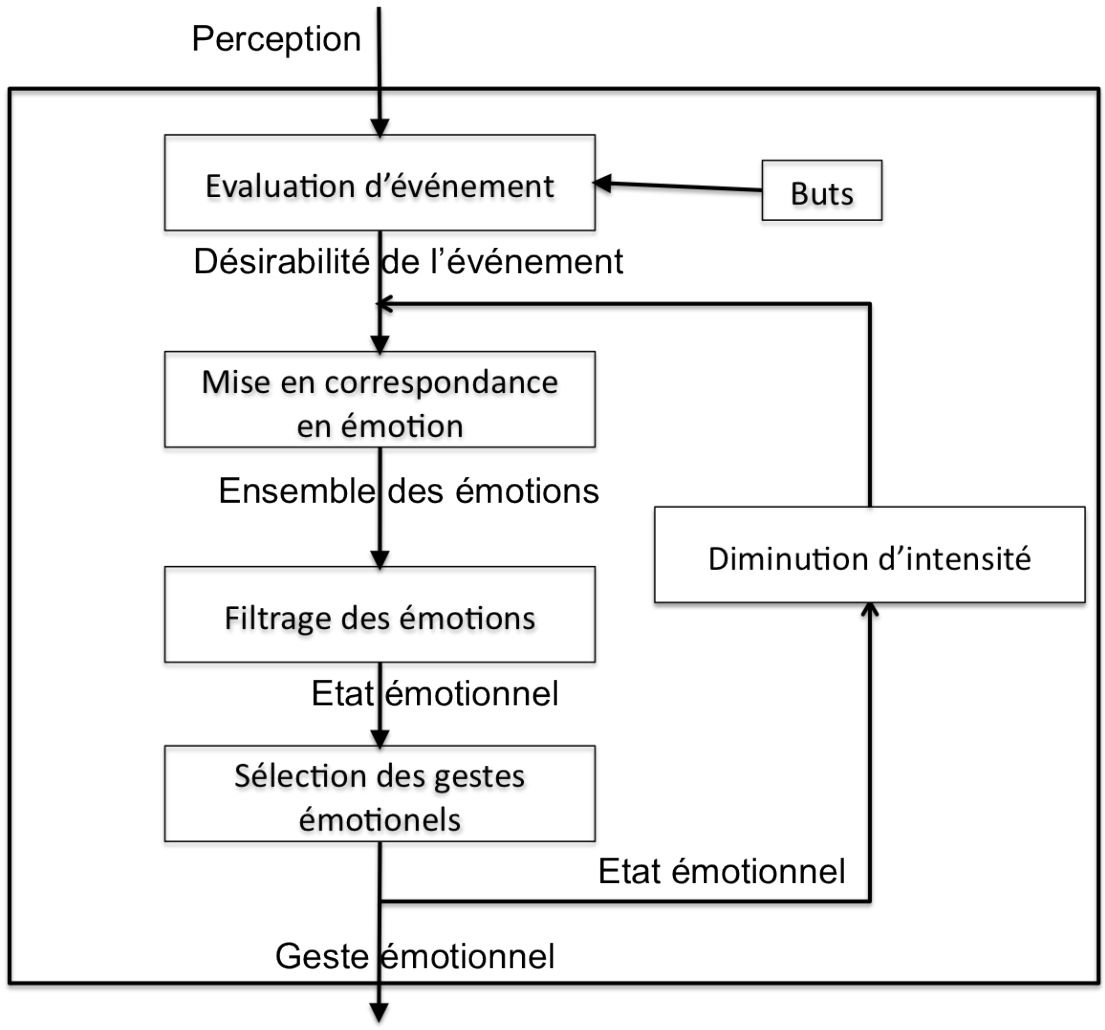


Figure 13 Détail de la composante émotionnelle de FLAME

L'originalité de ce travail réside notamment dans l'utilisation de la logique floue pour évaluer le niveau de pertinence (i.e. désirabilité) de l'événement avec le(s) but(s) de l'agent, et pour déterminer le comportement à sélectionner en fonction de l'intensité de l'émotion ressentie. De plus, les auteurs ont développé une capacité d'apprentissage qui permet à l'agent de s'adapter aux changements dans l'environnement. Supposons que l'agent perçoive un événement qui lui fait du mal, être attaqué par un chien, par exemple. La prochaine fois que l'agent perçoit un chien qui s'approche de lui, l'agent peut prédire le mal et donc ressentir peut-être la peur. Mais s'il voit un chien qui s'approche sans lui faire de mal, l'agent peut mettre à jour sa connaissance en diminuant la possibilité d'avoir mal quand un chien s'approche. Grâce à sa capacité de mettre à jour la possibilité d'associer une conséquence à un événement, FLAME permet à l'agent virtuel de s'adapter aux comportements de l'utilisateur et d'avoir ainsi des comportements proches de ceux d'un humain.

L'adaptation de GRACE pour FLAME est présentée dans la Figure 14. Dans FLAME, le filtrage de l'événement et son évaluation correspondent à l'application de règles OCC pour déterminer les émotions candidates pour la réponse effective. Ce filtrage est traduit dans GRACE par l'analyse cognitive menée par le composant *Interprétation Cognitive*. La gestion des états physiologiques (la faim, la soif) est réalisée par le composant *Interprétation Physiologique*. L'humeur est gérée par le composant *Etat interne*. Le filtrage de l'émotion dans FLAME est pris en charge par le composant *Comportement*. L'émotion renvoyée par le composant *Comportement* est ensuite exprimée par le composant *Expression* – qui s'occupe de la détermination de l'action à réaliser en fonction de l'intensité de l'émotion.

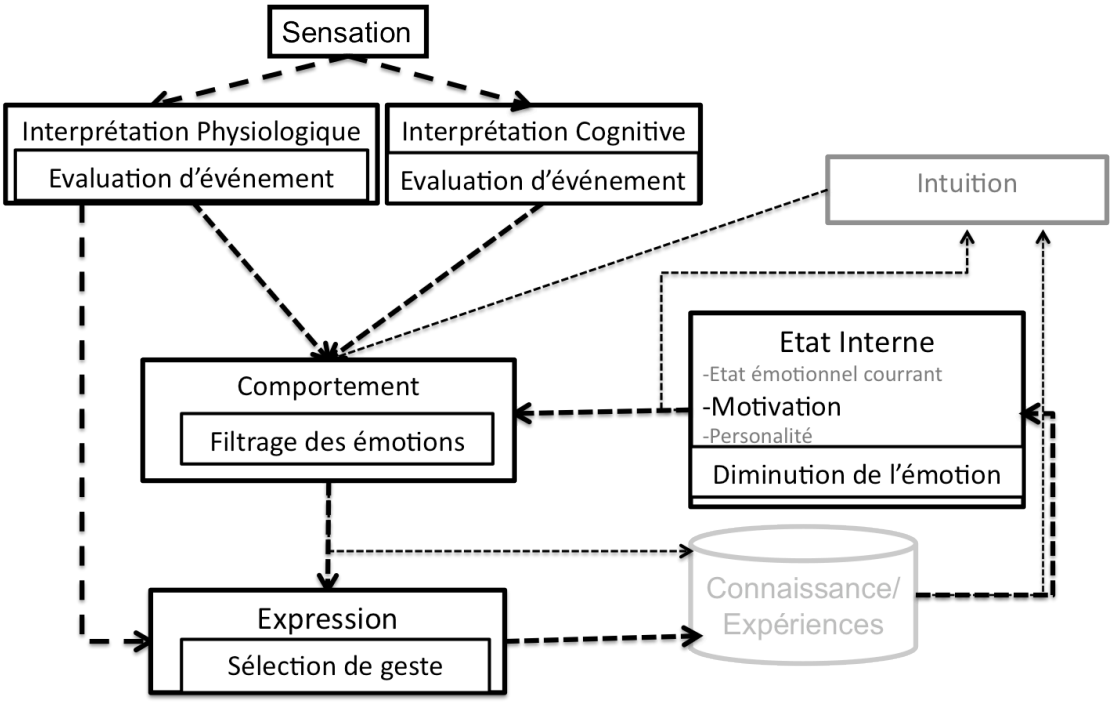


Figure 14 GRACE et la composante Emotion de FLAME

Comme la fonction d'apprentissage de FLAME prend la réponse émotionnelle (i.e. la valeur Valence – Activation en sortie du *Comportement* de GRACE) comme une sorte de poids pour mettre à jour les buts et le plan d'action de l'agent, cette partie de

l'apprentissage n'est donc pas à la charge de GRACE. L'idée est que GRACE ne s'occupe que la partie 'Emotion' dans l'architecture globale de l'agent. De même, la composante *Comportement* de FLAME n'est pas non plus incluse dans GRACE. C'est parce que GRACE est exclusivement dédié à simuler le processus émotionnel et ne prend aucune part à la génération des actions ou des plans d'actions de l'agent sauf la génération des gestes émotionnels. La composante *Expression* de GRACE ne fait qu'interfacer le processus émotionnel avec le composant d'action. Donc, la projection de FLAME vers GRACE ne contient pas la composante *Comportement* de FLAME. L'architecture globale de FLAME en remplaçant sa composante Emotion par GRACE peut être vue comme dans la Figure 15.

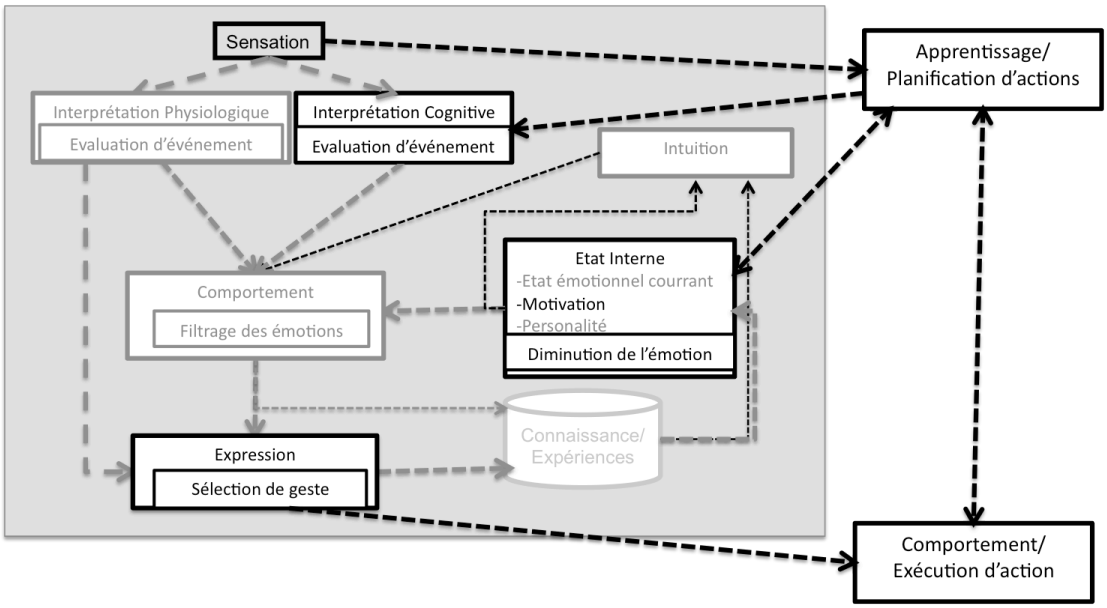


Figure 15 GRACE dans l'architecture globale de l'agent FLAME

La composante *Apprentissage/Planification* fournit de l'information à la composante *Interprétation Cognitive* de GRACE pour qu'elle puisse faire l'évaluation des événements en fonction de préférences et de plans actuels de l'agent. Cette composante *Apprentissage/Planification* interagit aussi avec la composante *Etat interne* de GRACE pour échanger de l'information sur l'état émotionnel et motivationnel actuel de l'agent géré par le processus émotionnel et de l'information sur le plan d'action et des préférences pour mettre à jour la motivation de l'agent. La composante *Comportement/Exécution d'action* est en charge d'exécuter les gestes choisis par la composante *Expression* de GRACE et aussi les actions dans le plan d'action pour achever les buts de l'agent. L'adaptation de GRACE à l'architecture globale de l'agent FLAME permet de bien distinguer entre le processus émotionnel et d'autres processus d'un agent (comme la planification, l'exécution d'actions, l'apprentissage), ce qui donne de la flexibilité pour l'implémentation de différentes caractéristiques (e.g. émotionnelle, rationnelle) d'un agent virtuel.

4.4. ParleE

ParleE est un modèle des émotions qui implémente aussi la théorie d'Ortony et al pour l'évaluation des événements (Bui, Heylen, Poel, & Nijholt, 2002). Ce modèle se fonde sur l'utilisation des probabilités pour estimer l'impact d'un événement sur la réalisation des buts de l'agent et donc l'émotion correspondante ; le choix d'une action pour répondre à l'événement, ainsi que l'estimation de l'émotion des autres

agents (dans un environnement multiagent) et des actions que ces derniers pourraient entreprendre sont également fondés sur des calculs probabilistes. Ces informations sur les autres agents servent à l'évaluation de l'événement, comme proposé par la théorie d'Ortony. ParleE incorpore le modèle de personnalité de Rousseau (Rousseau, 1996), dont chaque aspect affecte différentes composantes du modèle. Par exemple, l'aspect *Sentiment* reflétant la sensibilité détermine le seuil d'activation des émotions ; les aspects *Perception* et *Raisonnement* affectent la façon dont ParleE calcule l'attente et l'impact des événements sur les buts de l'agent ; l'aspect *Apprentissage* détermine la capacité d'apprentissage, et donc la capacité d'adaptabilité de l'agent.

Le processus émotionnel modélisé dans ParleE est présenté à la Figure 16. Le processus émotionnel se déclenche quand ParleE détecte un événement. Ce dernier est évalué tout d'abord par la composante d'évaluation des émotions (Emotion Appraisal Component en anglais) pour attribuer la valeur de probabilité des six variables suivantes : (1) Importance aux buts – estimer l'importance de l'événement par rapport à la réalisation des buts ; (2) Probabilité d'atteindre le(s) but(s) ; (3) Probabilité d'un événement à venir – estimer la probabilité qu'un événement se réalise dans l'avenir ; (4) Impact de l'événement sur le(s) but(s) de l'agent – estimer si l'événement est favorable à la réalisation des buts ou pas ; (5) Valeur sociale – pour évaluer l'action de l'agent ; (6) Niveau d'amabilité – évaluer l'attractivité de l'objet ou de l'agent qui effectue l'action. En fonction de ces variables, un vecteur d'impulsion des émotions est calculé (EIV – Emotion Impulse Vector, en anglais). Ce vecteur contient l'intensité de toutes les émotions qui sont susceptibles d'être produites, déterminées selon le modèle OCC. L'intensité de chaque émotion est associée à une formule de mise à jour différente. Par exemple, l'intensité de l'espoir et de la peur est calculée en fonction de l'importance du ou des but(s) (variable 1) et de la probabilité d'atteindre ces buts (variable 2), tandis que la joie et la tristesse sont fortement reliées à l'impact de l'événement sur les buts (variable 4), à l'importance des buts (variable 1), et à la possibilité d'un événement à venir (variable 3). Le résultat de la composante d'évaluation des émotions est donc un vecteur EIV des émotions avec les intensités appropriées.

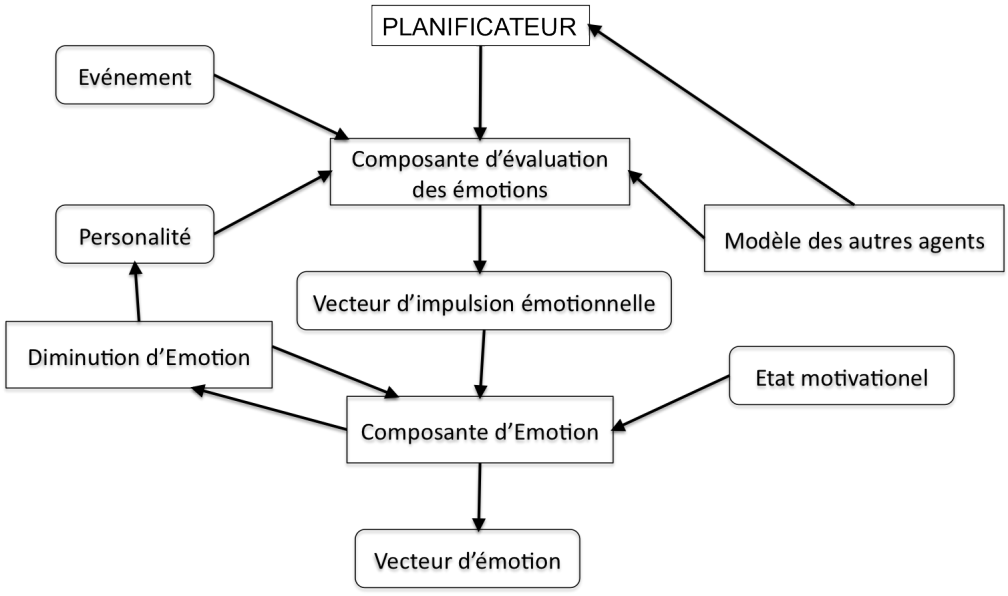


Figure 16 Architecture de ParleE

Ce vecteur EIV est ensuite transmis à la composante Emotion (Emotion Component en anglais) pour mettre à jour l'état émotionnel de l'agent. Chaque émotion dans ParleE possède deux seuils : le seuil d'activation – déterminer à quelle intensité l'émotion se manifeste, et le seuil de saturation – déterminer à quelle intensité l'émotion cesse d'augmenter. La mise à jour des intensités des émotions est affectée par la personnalité, l'état émotionnel précédant, le vecteur EIV et par l'état motivationnel. L'état motivationnel de ParleE est géré par la composante Etat Motivationnel (Motivational State en anglais). L'état motivationnel sert à modifier le seuil d'activation des émotions. Par exemple, la fatigue baisse le seuil d'activation des émotions négatives et augmente celui des émotions positives tandis que la faim augmente le seuil de toutes les émotions. La sortie de la composante Emotion est donc un vecteur contenant l'intensité de toutes les émotions après la mise à jour des intensités.

Une composante intéressante de ParleE est le composant *Planner* où les auteurs implémentent un algorithme de planification utilisant les techniques probabilistes. Cet algorithme permet de construire (et de reconstruire si nécessaire) le plan d'action de l'agent en fonction de l'évolution dans l'environnement et en fonction de priorités des buts à atteindre.

ParleE dispose aussi de la capacité d'apprentissage pour être plus adaptative. Deux éléments dans ParleE implémentent cette capacité, ce sont la mise à jour (1) de la variable 3 (Possibilité d'un événement à venir) dans la composante d'évaluation des émotions (Emotion Appraisal Component en anglais) et (2) de la valeur des actions (pour représenter le comportement standard). Le premier permet à ParleE d'anticiper le futur tandis que le second permet à ParleE de s'adapter aux préférences des utilisateurs au cours de l'interaction.

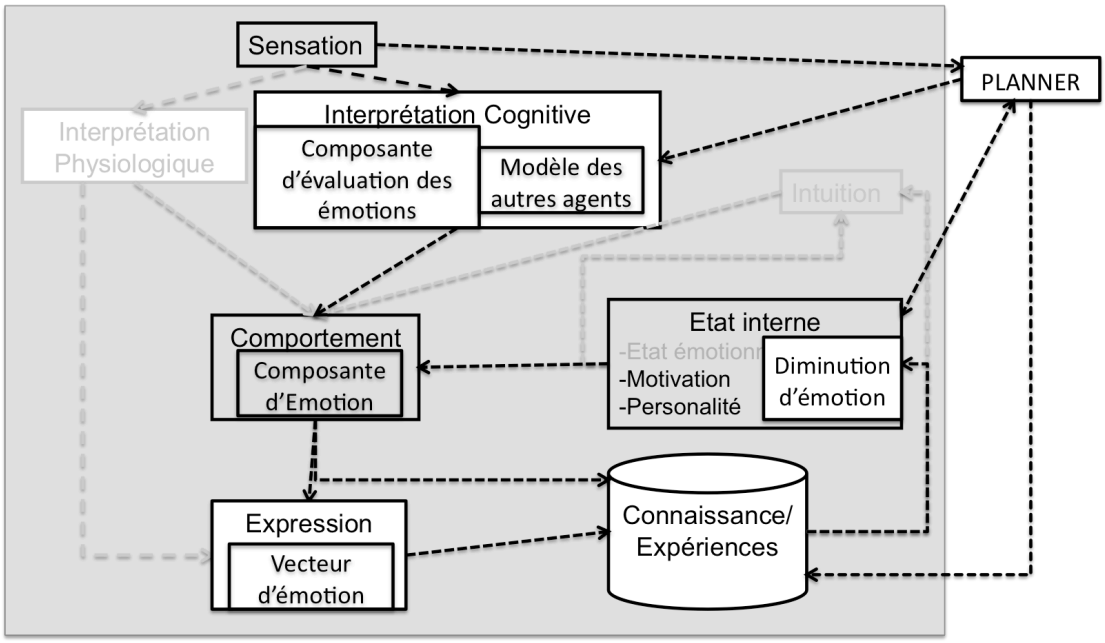


Figure 17 GRACE et ParleE

La Figure 17 explicite la manière dont GRACE simule les processus de ParleE. Dans GRACE, l'évaluation menée par le composant *Composante d'évaluation des émotions* est à la charge du composant *Interprétation Cognitive*. L'évaluation dans *Interprétation Cognitive* va produire un vecteur des valeurs d'intensité des émotions

élicitées par l'événement. Ce vecteur est ensuite passé au composant *Comportement* où une seule émotion va être choisie en fonction de l'intensité et du seuil d'activation et de saturation prédéfini. Cette émotion choisie est ensuite transmise au composant *Expression*. La diminution de l'intensité et la gestion de la personnalité sont prises en charge par le composant *Etat interne* de GRACE.

Dans cette adaptation, on voit aussi la distinction entre la partie Emotion et la partie Planification. Dans ParleE, la mise à jour des 6 variables affectives est faite dans la *Composante d'évaluation des émotions*. Dans notre adaptation, cette phase de mise à jour est déportée dans le *Planner*. La raison en est que cette mise à jour est liée à la mise à jour du plan d'actions, ce qui implique un traitement rationnel. Pour garder la distinction entre la partie émotionnelle et la partie rationnelle d'un agent, et donc garder la flexibilité de GRACE, la mise à jour de ces 6 variables est donc faite par PLANNER qui se trouve en haut à droite dans la Figure 17, donc à l'extérieur de la partie émotionnelle, i.e. à l'extérieur du modèle GRACE. L'information sur le plan mis à jour sert aussi à la mise à jour de l'état motivationnel de l'agent (faite dans la composante *Etat interne*) et à la mise à jour de l'expérience émotionnelle dans la composante *Connaissance/Expériences.*

4.5. Greta

Greta est un agent émotionnel virtuel qui incorpore dans son fonctionnement un mécanisme d'activation de l'émotion en fonction de l'événement que l'agent a perçu (de Rosis, Pelachaud, Valeria, & De Carolis, 2003). Chaque agent dispose de trois composantes : un *cerveau* qui gère la connaissance, les émotions et la personnalité de l'agent, via des *Réseaux Dynamiques de Croyances* ; un *marqueur de langue* qui traduit les émotions en balises expressives intégrées dans la parole de l'agent ; un *corps* qui exécute l'action de l'agent, c'est-à-dire parler et exprimer l'émotion de la manière définie par les balises.

Fondé sur le modèle agent BDI (Belief – Desire - Intention) (Rao & Georgeff, 1991), le cerveau de Greta est organisé de façon à représenter ses croyances, ses buts, et son plan d'action. L'émotion de Greta lors d'un événement est déterminée selon l'évaluation proposée par Ortony et al en prenant en compte le but de l'agent et sa personnalité. En fonction de son but, l'agent évalue ainsi l'événement en fonction des *conséquences de l'événement*, de *l'action de l'agent*, ou de *l'aspect de l'objet* pour déterminer quelle(s) émotion(s) susciter. La personnalité aide à définir la priorité que l'agent va associer à ces buts. Un agent égoïste donne ainsi une priorité élevée au but d'*« atteindre son bien-être dans l'avenir »* et active donc facilement la détresse et la peur. Par contre, un agent généreux donne une grande priorité au but d'« *atteindre le bien-être des autres* » et active donc facilement l'espoir, la joie/tristesse envers les autres.

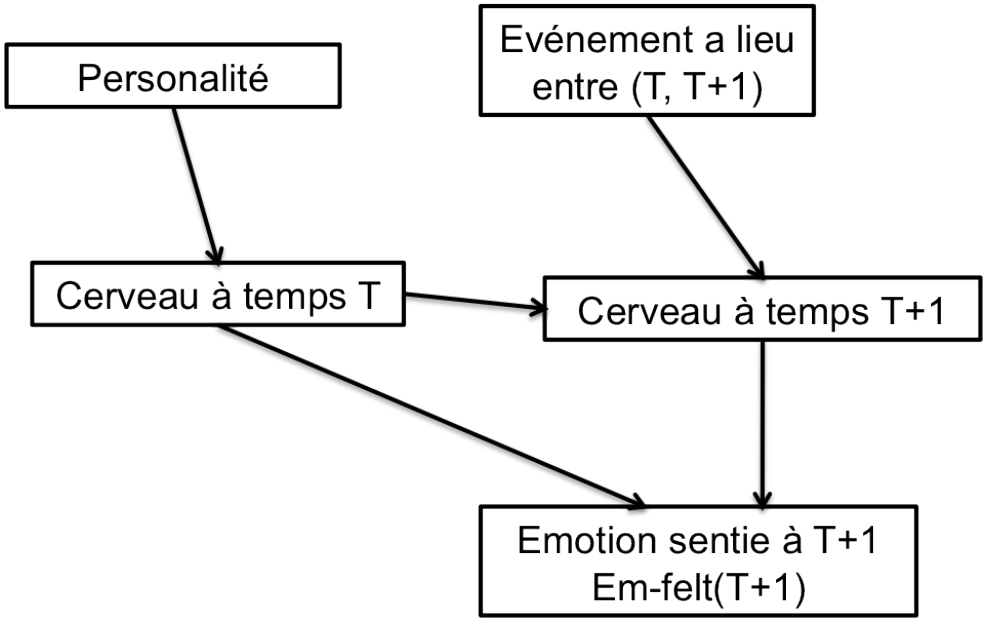


Figure 18 Architecture des réseaux dynamiques de croyance de Greta

Dans Greta, plusieurs émotions peuvent coexister en même temps. A chaque pas de temps, chaque émotion met à jour son intensité en fonction de l'événement perçu. Cette opération prend en compte deux critères : l'*incertain* indiquant la possibilité d'achever ses buts ; et l'*utilité* associée à l'achèvement de chaque but. Les émotions déjà existantes mais pas impactées par l'événement vont diminuer en intensité en fonction du type d'émotion (par exemple la joie diminue plus vite que la tristesse) et en fonction de la personnalité de l'agent.

Dans Greta, l'événement perçu passe d'abord dans la composante *Cerveau* pour l'analyse – et correspond à une évaluation en termes des possibilités d'achèvement de buts. Si l'événement est favorable aux buts de l'agent, alors l'évaluation sera positive, et vice versa. Cette évaluation se base sur le modèle OCC pour déterminer l'émotion en réponse à la situation, et l'intensité de cette émotion est calculée en fonction de son aspect plus ou moins favorable aux buts de l'agent.

Dans GRACE, le fonctionnement de la composante *Cerveau* de Greta est implémenté dans la composante *Interprétation Cognitive*. L'émotion de sortie de ce composant est transmise au composant *Comportement* – qui est en charge de calculer l'intensité de cette émotion en fonction de la personnalité prédéfinie. La composante *Expression* exprime ensuite cette émotion avec l'intensité appropriée. La diminution de l'intensité de l'émotion a lieu dans le composant *Etat interne*.

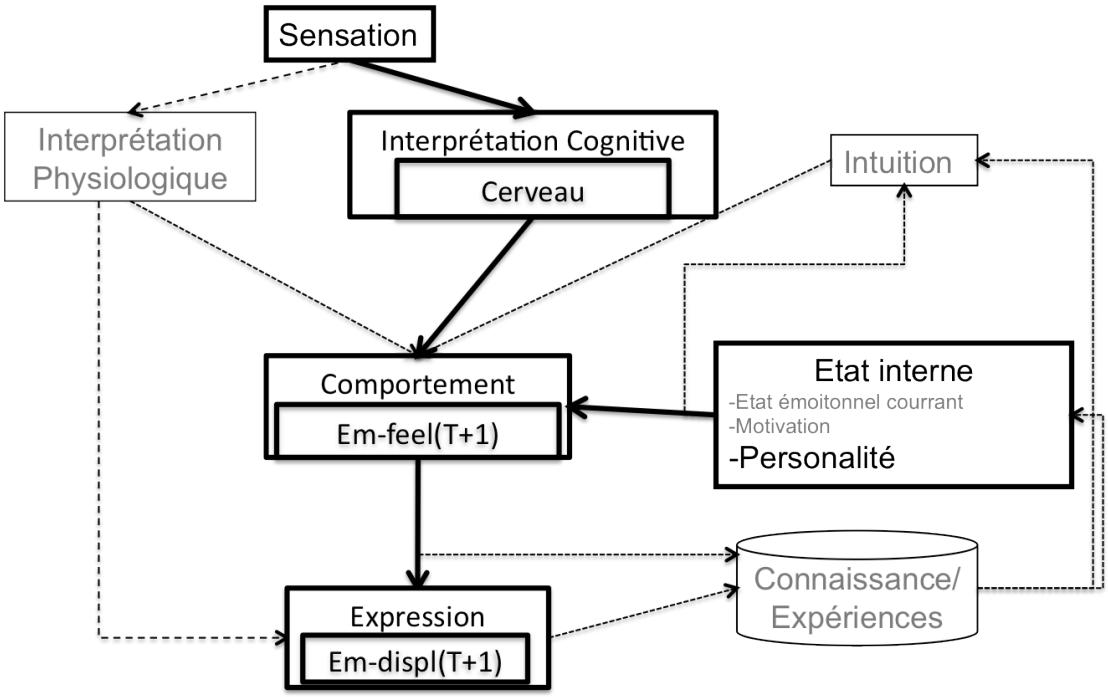


Figure 19 Adaptation de GRACE pour simuler Greta

La généricité de GRACE permet d'englober de manière satisfaisante les différents éléments de ces réseaux dynamiques. De plus, là où la composante *Cerveau* de Greta analyse seulement la croyance et les buts de l'agent par rapport à la situation actuelle, nous proposons d'implémenter l'état cognitif interne en tenant compte de plusieurs aspects dans la vie de la personne affectant ses comportements émotionnels, tels que l'état physiologique ou encore l'intuition.

4.6. EMA

Le modèle EMA (Gratch & Marsella, 2006) est inspiré du travail de Smith et Lazarus sur le processus émotionnel, en particulier de la stratégie du faire face dans le cadre de situations stressantes. Ce modèle implémente un processus émotionnel en deux étapes : (1) évaluer la relation individu - environnement, et (2) choisir une stratégie d'action face à la situation. Tandis que la première étape sert à évaluer l'événement en fonction des buts, des croyances et du plan d'action de l'agent, la deuxième consiste à débarrasser l'agent des situations dans lesquelles ses buts sont en conflit. Par exemple, dans un incendie, la sécurité personnelle et la tâche d'éteindre le feu sont souvent en conflit pour un pompier, et il devra choisir d'éxecuter l'une et donc réévaluer/ré-organiser ses priorités des buts pour se débarrasser de l'autre.

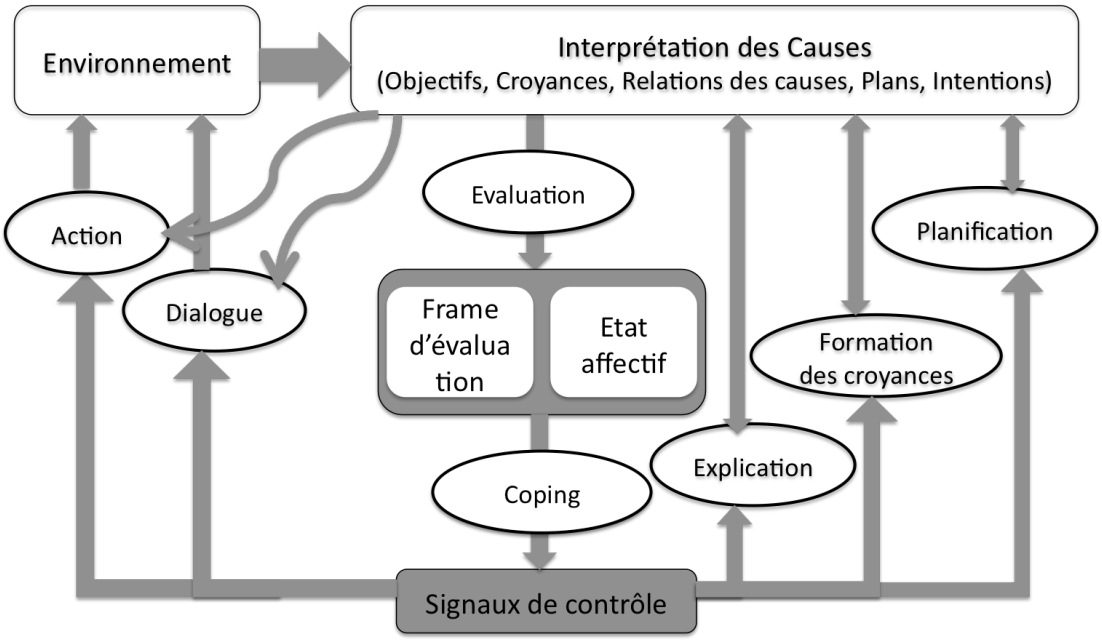


Figure 20 Architecture d'EMA (Gratch & Marsella, 2006)

La Figure 20 présente le mécanisme de fonctionnement d'un agent incorporé EMA. Chaque événement perçu par l'agent va d'abord passer la première étape : évaluer la relation individu - environnement. Dans cette étape, l'agent évalue l'événement en fonction de sept variables :

- *perspectives* pour mesurer l'impact de l'événement sur lui-même et les autres agents dans la situation ;
- *pertinence* mesure comment l'événement est relié aux buts de l'agent ;
- *désirabilité* mesure la valence de l'événement en fonction des préférences de l'agent ;
- *vraisemblance* mesure la certitude de l'événement ;
- *attribution causale* détermine qui est en charge de l'occurrence de l'événement (pour accuser/admirer) ;
- *contrôlabilité* mesure la capacité de l'agent à résoudre le problème associé à l'événement ;
- *volatilité* mesure la possibilité que la situation va changer sans intervention de la part de l'agent.

L'attribution de valeurs à ces variables donne à l'agent une vue globale sur la situation actuelle. L'événement déjà évalué via les six variables est ensuite utilisé par l'agent pour choisir une stratégie afin de faire face à la situation (i.e. la stratégie de coping), pour résoudre le problème associé à l'événement. Normalement, si l'événement est favorable pour l'agent et son plan d'action, l'agent n'a qu'à exécuter son plan. Mais si l'événement impacte négativement la réalisation des buts de l'agent ou affecte négativement les autres agents, l'agent lui-même doit faire une ré-évaluation de la situation, soit orientée émotions, soit orientée problème pour se débarrasser de la situation de stress.

EMA a été déployé dans un simulateur des situations de stress (comme les incendies, les opérations médicales, les missions de secours lors des catastrophe naturelles, etc). Les agents dans ce simulateur sont implémentés avec EMA pour avoir une capacité de reproduire les comportements humains dans ces situations.

Pour implémenter EMA dans GRACE, nous proposons d'implémenter l'évaluation primaire dans le composant *Interprétation Cognitive*. Le résultat de l'analyse de cette composante représente en général la compréhension de l'agent sur l'environnement. L'évaluation de cette compréhension va ensuite mettre en correspondance les émotions associées à l'événement (i.e. à la situation actuelle). Ces émotions sont ensuite transmises à la composante *Comportement* de GRACE pour faire la partie coping des émotions, c'est-à-dire choisir l'émotion la plus appropriée selon les préférences/motivation de l'agent dans l'état courant. L'émotion sélectionnée va être exprimée dans la composante *Expression*. La partie coping orientée problème est purement rationnelle et est prise en charge par une composante à l'extérieur de GRACE. On l'appelle pour l'instant *Planification d'actions* et elle se trouve en haut à droite de la Figure 21. Celle-ci s'occupe de la planification et de re-planification en cas de conflit des buts, ou en cas de mauvaise situation (ce qui s'exprime aussi en état émotionnel négatif). L'*Etat interne* est en charge d'alerter la composante *Planification* en cas de situation défavorable. Elle met à jour aussi l'état motivationnel et les préférences de l'agent en fonction du plan construit par *Planification*. Avec cette adaptation de GRACE, deux stratégies de coping sont bien distinguées. L'approche émotionnelle est prise en charge par GRACE tandis que l'approche rationnelle est prise en charge par la partie de planification. Une autre composante extérieure de GRACE, *Exécution d'actions* (qui se trouve en bas à droite de la Figure 21), est en charge d'exécuter les actions selon le plan et les expressions émotionnelles venant de la composante *Expression* de GRACE.

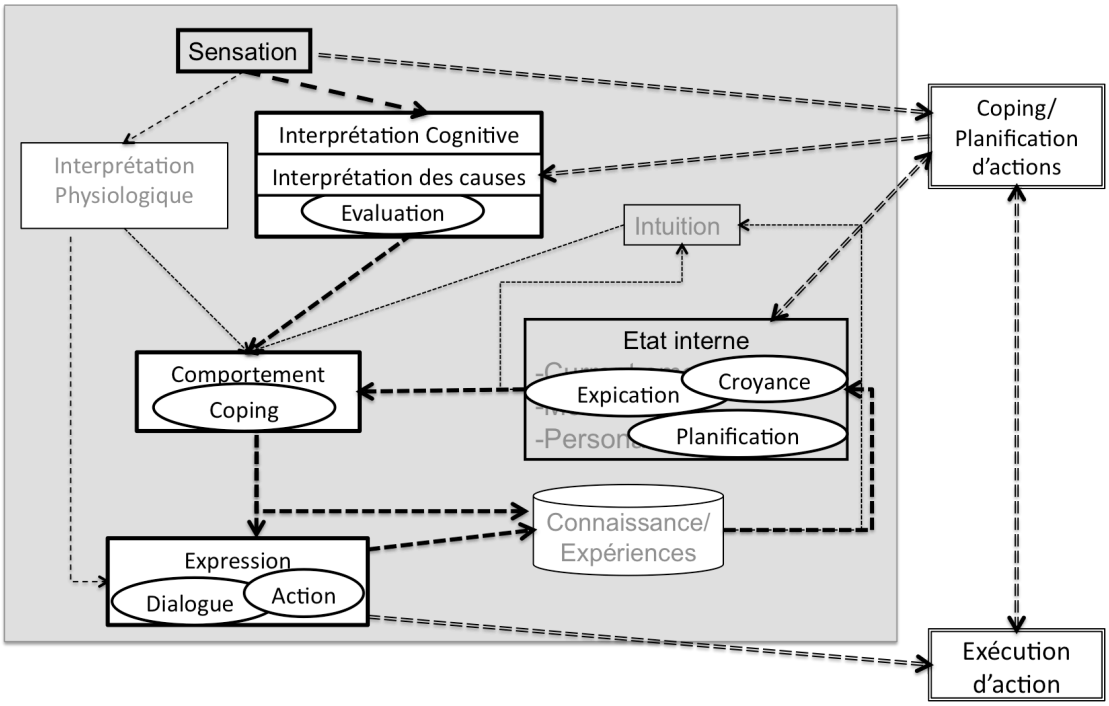


Figure 21 GRACE et EMA

Quand il y a des événements défavorisant le plan de l'agent, le composant *Comportement* a pour rôle de déclencher le processus de coping émotionnel. La focalisation sur l'émotion consiste à refaire l'évaluation cognitive dans laquelle l'agent devrait mettre à jour ses préférences, ses motivations, ses buts pour accepter la réalité et au pire abandonner ses buts. Cette orientation conduit notamment à une expression émotionnelle visible. La personnalité et l'humeur – qui sont gérées par le composant *Etat interne*, participent aussi au processus de coping comme décrit dans le paragraphe précédent. Et la *Planification*, quant à elle, revisite le plan afin de chercher une solution adaptative pour achever le but final.

L'aspect intéressant que GRACE pourrait apporter à EMA est la possibilité d'avoir des reflexes émotionnels sans passer par le processus de coping. Ceci est possible grâce à l'analyse physiologique – ce qui est aussi un élément important dans le processus émotionnel. De plus, le fait de re-évaluer peut être vu comme l'activation neuronale, ou l'activité du cerveau, grâce auxquelles l'agent peut générer des événements imaginaires pour surmonter le stress.

4.7. ALMA – A layer Model of Affect

Dans le but de développer des agents conversationnels capables d'interagir avec l'humain de façon réaliste, (Gebhard, 2005) a proposé un modèle des émotions appelé ALMA (A Layered Model of Affect en anglais) pour intégrer dans les agents conversationnels animés. Ce travail est intégré dans le projet VirtualHuman portant sur l'étude des concepts et des techniques pour améliorer la conversation multimodale (en combinant les aspects textuel, vocal et gestuel) entre l'humain et des agents conversationnels animés. L'aspect émotionnel a donc pour but de renforcer l'efficacité de l'interaction en rendant les réactions des agents les plus naturelles possibles (Figure 22).

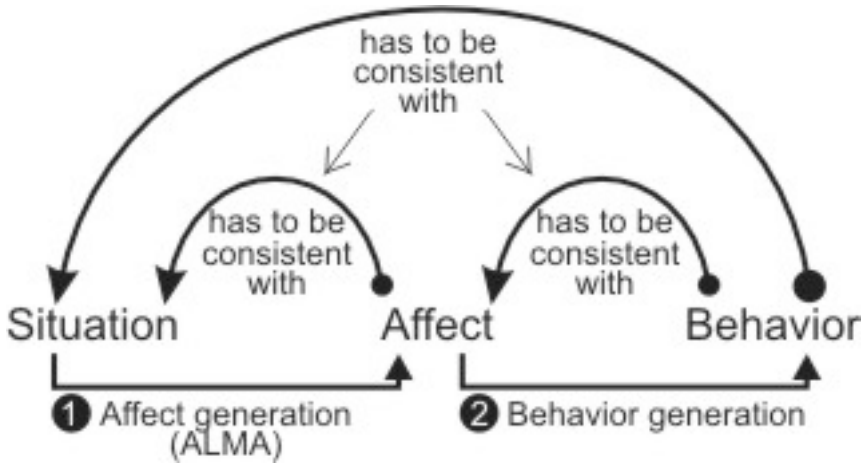


Figure 22 Principe de génération des comportements en consistance avec l'état affectif du modèle ALMA (Gebhard & Kipp, 2006)

Le modèle ALMA[1] considère l'état affectif de l'agent en trois niveaux : émotion, humeur et personnalité. L'émotion est l'état affectif à court terme et qui a pour objet de répondre à un événement, une action ou un objet. L'humeur est l'état affectif à moyen terme qui ne se rattache ni à un événement, ni à une action, ni à un objet spécifique. L'humeur est stable pendant une certaine durée de temps (généralement prédéfinie par le concepteur) et est influencée par le résultat d'analyse cognitive de l'agent (i.e. l'émotion ressentie lors de la perception d'un événement). L'humeur dans ce modèle est définie comme la moyenne de l'état émotionnel de l'agent à travers diverses situations. La personnalité est l'état affectif à long terme. Elle représente les caractéristiques mentales de l'agent et détermine comment l'émotion de l'agent va évoluer en fonction de l'information perçue et de son humeur. ALMA utilise le modèle de personnalité Big Five (proposé par (McCrae & John, 1992)) pour les agents. Ce modèle permet de définir les traits de personnalité en fonction de 5 dimensions : Extraversion, Névrose, Agréabilité, Conscience, Ouverture à l'expérience.

L'émotion de l'agent envers un événement est déterminée en utilisant le modèle OCC. L'intensité de cette émotion est calculée en se basant sur le trait de personnalité. Puis l'humeur va être modifiée en fonction de l'émotion ressentie. Pour la représentation des émotions, l'auteur utilise la représentation dimensionnelle avec trois dimensions : Valence, Activation, et Dominance. Il a utilisé la proposition de (Mehrabian, 1996) pour placer les émotions OCC dans l'espace Valence-Activation-Dominance. Il se base aussi sur le travail de (Mehrabian, 1996) pour calculer l'humeur de l'agent connaissant son trait de personnalité dans le modèle Big-Five.

Ce modèle peut être vu comme une conception concrète de la relation entre la personnalité, l'humeur et l'émotion. Il permet de valider, du point de vue computationnel, la modélisation de ces trois aspects sur les agents virtuels. Pourtant, du point de vue de la modélisation des émotions artificielles, il n'est pas clair dans ce

[1] Le modèle ALMA a été implémenté et est téléchargable à l'adresse http://www.dfki.de/~gebhard/alma/index.html#download.

travail comment il prend en compte l'aspect *Réflexe* (qui corespond à l'*Interprétation physiologique* de GRACE pour répondre à une situation de danger/urgence) et l'aspect *Rappel de la mémoire* (qui corespond à l'*Intuition* de GRACE pour anticiper l'avenir, pour planifier à l'avance ses comportements). De plus, dans ce modèle, il n'est pas clair non plus comment l'action de l'agent va modifier l'état émotionnel pour prendre en compte le mécanisme de *retour physiologique* rapporté dans les théories psychologiques des émotions. Ce dernier est important car il permet la contagion émotionnelle via l'imitation gestuelle et l'augmentation de l'acceptabilité lors de l'interaction sociale.

4.8. FAtiMA – Fearnot AffecTIve Mind Architecture

Afin de modéliser les comportements émotionnels des agents, (Dias, Mascarenhas, & Paiva, 2011) ont proposé FAtiMA – Fearnot AffecTIve Mind Architecture). Cette architecture a été étendue dans plusieurs travaux de recherche pour simuler différents phénomènes émotionnels pour les agents virtuels. Comme ces utilisations sont de différentes natures, l'implémentation concrète de l'architecture est différée d'un travail à un autre (FearNot! (Paiva, et al., 2005), ORIENT (Ruth, Vannini, Andre, Paiva, Enz, & Hall, 2009), Model of Empathy (Rodrigues, Mascarenhas, Dias, & Paiva, 2009), Cultural Behaviour (Mascarenhas, Dias, Prada, & Paiva, 2010), Drives (Lim, Dias, Ruth, & Paiva, 2011)). Les chercheurs trouvent qu'il est nécessaire de re-définir l'essentiel de l'architecture pour modéliser la capacité émotionnelle pour les agents intelligents. Ils ont donc clarifié, dans (Dias, Mascarenhas, & Paiva, 2011), les fonctionnalités importantes du processus modélisé dans FAtiMa, appelé FAtiMA Core (Figure 23).

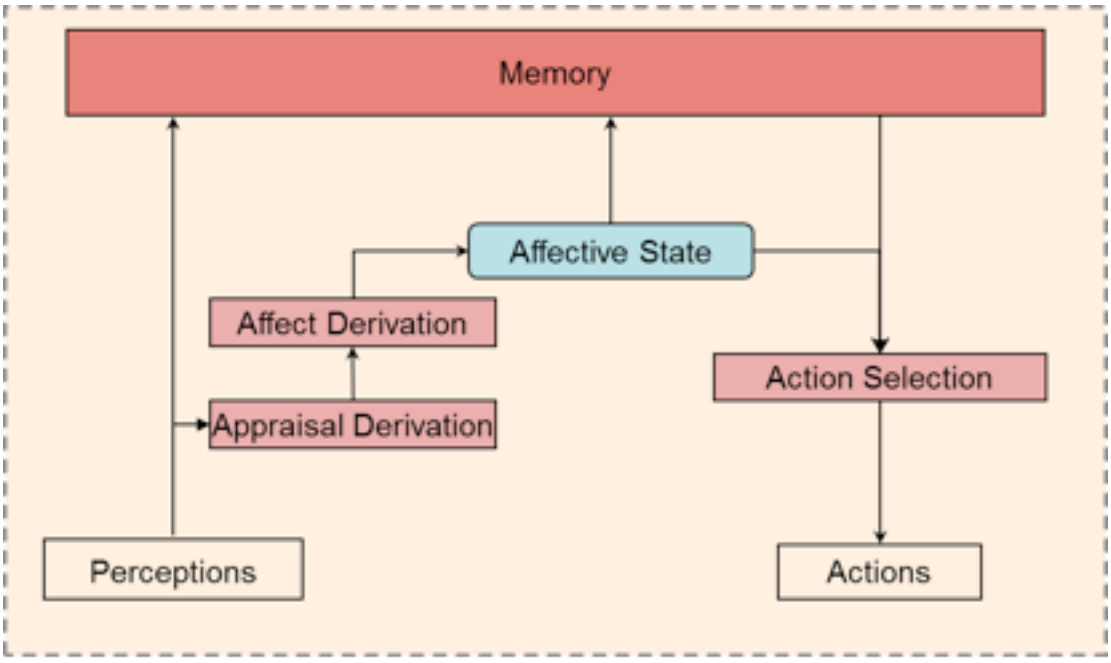


Figure 23 Architecture FAtiMA Core (Dias, Mascarenhas, & Paiva, 2011)

FAtiMA Core modélise le processus émotionnel en deux étapes : Dérivation Cognitive (Appraisal Derivation en anglais), et Dérivation Affective (Affect Derivation en anglais). La Dérivation Cognitive est en charge de faire l'association de l'événement perçu avec des variables de l'évaluation cognitive, comme celles proposées dans OCC ou bien les theories de K. Scherer sur des émotions. Les

variables de cette évaluation peuvent être le niveau de désirabilité, de nouveauté, de danger, etc. Ensuite, la Dérication Affective détermine l'émotion en fonction des valeurs des variables données par la Dérivation Cognitive. La deuxième phase d'association peut se baser aussi sur le modèle OCC ou de Scherer. Par exemple, si le niveau de désirabilité d'un événement est élevé, la joie peut être reproduite.

Dans FAtiMA Core, la Dérivation Cognitive peut faire aussi de la ré-évaluation sur l'événement déjà traité, qui va donc changer la valeur des variables cognitives précédemment attribuées à l'événement. Ce changement déclenchera une ré-évaluation au niveau de la Dérivation Affective, qui entrainera peut-être des changements de réponse émotionnelle envers un événement. Par exemple, un événement initialement évalué comme indésirable peut être ré-évalué comme désirable. Dans ce cas, la détresse initialement ressentie peut être changée en joie. D'ailleurs, l'évaluation de chaque phase du processus émotionnel dans FAtiMA Core est différentielle, c'est-à-dire que l'événement est évalué en boucle et le résultat de l'évaluation (i.e. les valeurs des variables cognitive ou l'émotion en sortie) est généré à différents moments. Ce dernier point explique aussi l'effet de ré-évaluation présenté précédemment.

Du point de vue de l'implémentation, FAtiMA Core est intégré dans une architecture appelée FAtiMA Modular. FAtiMA Modular est conçue en ajoutant des composants à FAtiMA Core. Par exemple, dans le cas de FearNot!, six composants ont été ajoutés. La Dérivation Cognitive a été assistée par deux composants : Reactive Component et Deliberative Component. Le premier composant analyse l'événement en fonction des variables Désirabilité, DésirabilitéPourLesAutres, Admirabilité, Amabilité. Le deuxième composant analyse l'événement en fonction des variables suivantes : EtatDesButs, FavorabilitéAuxButs, ProbabilitéAtteinteButs. Le composant OCCDerivationAffective est ajouté pour prendre en charge la généralisation des émotions en fonction des valeurs des variables cognitives résultant de l'analyse cognitive (conduite par les deux composants Reactive Component et Deliberative Component). De plus, Motivational Component, Theory of Mind Component, et Cultural Component sont aussi ajoutés pour modéliser respectivement l'influence de la motivation, de l'humeur, et du standard social sur le processus émotionnel de l'agent virtuel.

4.9. Psi – Emotion en fonction de l'intention

La théorie psychologique appelé Psi de (Dörner & Hille, 1995) suggère qu'une émotion est une modulation du processus cognitif pour faire face à certaines conditions/situations. Un processus cognitif dans ce cas comprend l'intention, la planification d'action et la perception. L'apparition d'un besoin ou d'une envie (s'il s'agit d'une motivation) déclenchera une intention de satisfaire ce besoin/envie. L'intention provoque ensuite la planification d'actions et l'exécution d'actions pour satisfaire les besoins/envies. L'intention influence aussi la perception en la renforçant ou la diminuant et donc influence l'attention de l'individu envers l'environnement lors des différents états émotionnels.

Par exemple, la colère peut être caractérisée par les détails suivants :

- Un niveau d'activation élevée qui implique une vitesse élevée du processus de traitement de l'information.

- Un seuil élevé de sélection : il s'agit d'une tendance forte pour concentrer sur une seule intention et d'un niveau de sensibilité faible envers les stimulus non concernés par cette intention.

- Un niveau faible de résolution : il s'agit d'être limité sur la façon de voir/percevoir l'environnement, de planifier, et de prendre des décisions. Autrement dit, l'individu qui est en colère ne prend pas en compte les conditions nécessaires et/ou des conséquences possibles de ses actions quant au processus cognitif. Cela résulte en général en des comportements « inconditionnels » et en sur-planification.

- Un faible taux de mise à jour de l'image de la situation actuelle : quelqu'un en colère ne prend pas en compte beaucoup de détails sur la situation ou sur ses changements.

Chaque émotion est donc un processus cognitif caractérisé par des motifs spécifiques de la motivation qui engendre l'intention (comme le niveau d'activation, le seuil de sélection), de la planification (i.e. niveau de résolution) et l'exécution d'actions et de la perception (i.e. taux de mise à jour de l'image de la situation actuelle).

L'adaptation computationnelle de leur théorie, aussi présentée dans (Dörner & Hille, 1995), est appelée Âme Artificielle (Artificial Soul en anglais), permet de démontrer la pertinence de la modulation des paramètres définis dans leur théorie Psi (i.e. niveau d'activation, seuil de sélection d'une intention, niveau de résolution, et taux de mise à jour). Avec cette adaptation, ils ont pu simuler plusieurs émotions, comme la peur, l'espoir, la colère, la résignation, la dépression, et d'autres encore. Ils peuvent aussi simuler différentes personnalités (i.e. quatre types de personnalités proposés par (Eysenck & Rachman, 1965)) avec leur Âme Artificielle. Ils proposent aussi des paramètres à régler pour rendre un robot autonome et naturel, qui sont :

- *« un motif de curiosité pour acquérir des connaissances sur l'environnement au cas possible et nécessaire,*

- *un mécanisme d'action ou une sélection de motifs pour déterminer automatiquement quoi faire en fonction de la situation,*

- *des paramètres internes qui déterminent le trait de personnalité et peuvent engendrer des types de comportements différents pour les robots aussi bien que pour les tâches variées,*

- *des émotions en terme des possibilités de générer des comportements adaptifs convenant à la situation courante*. »

La spécification de la théorie Psi sur les paramètres à régler a facilité la conception des comportements dits émotionnels des agents virtuels ou robotiques. Elle a donné lieu à différentes adaptations scientifiques en agent virtuels ou robots autonomes. Par exemple, (Lim, Aylett, & Jones, 2005) implémente un agent virtuel jouant le rôle d'un guide touristique intégré dans un assistant personnel numérique (Figure 24). Cette application reste fidèle à Psi dans le sens où l'architecture logicielle de l'agent respecte le processus de traitement de l'information décrit dans Psi, i.e. comprenant, entre autres, la Généralisation de l'intention, la sélection de l'intention, l'exécution de l'intention (incluant la planification et l'exécution d'actions), et la perception. L'objectif de l'agent est de faire visiter les touristes selon une trajectoire prédéfinie.

Au cours de la visite, son état émotionnel change en fonction de ses compétences à satisfaire les demandes (besoin/envie) du touriste et de maintenir la trajectoire de visite. Ce travail de (Lim, Aylett, & Jones, 2005) peut être considéré comme une instance fonctionnelle de la théorie Psi de (Dörner & Hille, 1995).

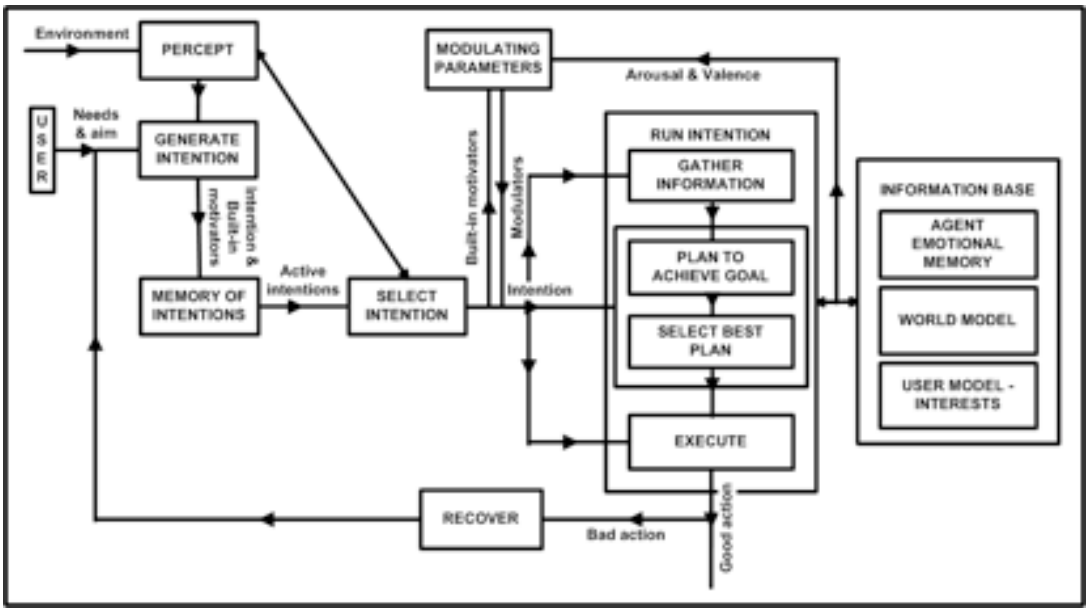


Figure 24 Adaptation de Psi faite par (Lim, Aylett, & Jones, 2005) pour les agents émotionnels

Un autre exemple d'adaptation est le résultat du projet MicroPsi (Figure 25) géré par Josche Bach au Centre des Sciences Intégratives de la Vie (Center for Integrative Life Sciences) à l'Université Humboldt de Berlin (Bach, 2009). Le projet a développé un environnement de simulation des agents intelligents intégrant des capacités rationelles et émotionnelles. L'aspect émotionnel de MicroPsi est fondé sur la théorie Psi de (Dörner & Hille, 1995). Les envies cognitives évoquant le processus cognitif de l'agent MicroPsi sont les six suivantes : Energie, Eau, Intégrité, Affiliation, Certitude, et Compétence. A tout moment, l'agent évalue ces six envies cognitives afin de planifier ses comportements pour satisfaire les envies surgies ou bien rassurer l'exécution du plan actuel quant à la diminution des envies en question. MicroPsi a été dévéloppé comme un plug-in Eclipse fournissant un éditeur graphique pour concevoir le modèle cognitif Psi de l'agent, un simulateur de modèle Psi permettant de tracer l'évolution cognitive des agents, un éditeur et un simulateur pour simuler l'environnement, et une vue 3D pour les rendus graphiques du monde simulé. MicroPsi a été utilisé non seulement pour créer des agents virtuels cognitifs mais aussi pour gérer l'autonomie de robots Khepera (e.g. imitant la perception et les comportements des souris dans un labyrinthe).

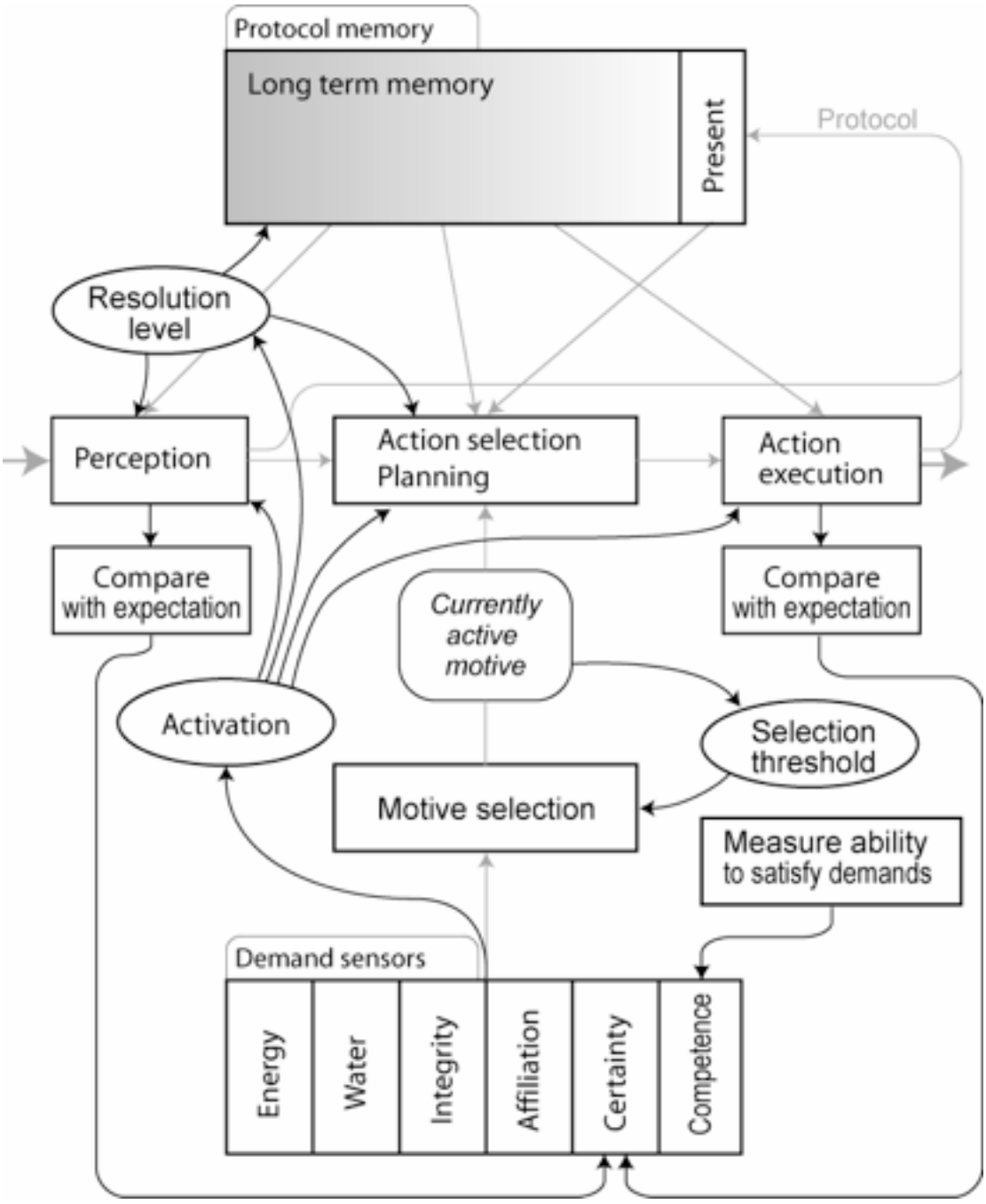


Figure 25 Adaptation de Psi proposée par (Bach, 2009)

5. Elements pour l'implémentation

Même si le modèle GRACE n'est pas encore complètement implémenté, certains modules le sont (voir chapitres suivants). Il est aussi possible de donner des éléments pour comprendre de quelle manière les autres modules seront implémentés.

L'implémentation de l'interprétation physiologique consiste à mettre en œuvre les réflexes émotionnels, comme les réflexes de la surprise ou de la peur. Cette interprétation peut être considérée comme une évaluation des critères Nouveauté et Urgence comme propose K. L. Scherer dans sa théorie. Ce type d'évaluation produit

notamment une valeur d'activation élevée et une valence proche de neutre. Prenons le cas d'un robot personnel : quand un objet apparaît soudainement proche du robot, la distance entre le robot et l'objet peut être utilisée pour calculer la valeur d'activation de l'événement. Si la distance passe sous un certain seuil (prédéfini par le concepteur), la composante *Interprétation Physiologique* peut envoyer son résultat d'évaluation à la composante *Expression* pour déclencher une expression de surprise. Ce résultat de *Interprétation Physiologique* doit être envoyé aussi à la composante *Comportement* pour alerter d'une surprise pour que le *Comportement* soit au courant de ces réflexes et qu'il puisse faire une évaluation appropriée de l'événement.

L'implémentation de l'interprétation cognitive consiste à monter un mécanisme de mise en correspondance entre un événement et des valeurs affectives. Ceci peut être considéré comme une connaissance à priori comme dans le cas de l'agent Affective Reasoner, de l'agent Greta ou dans l'interprétation des causes d'EMA. Nous avons travaillé aussi sur l'implémentation de cette composante *Interprétation Cognitive* pour déduire des valeurs émotionnelles (i.e. la valence et l'activation) dans la musique. L'idée est de mettre en œuvre un système d'extraction automatique des valeurs émotionnelles lors de l'écoute musicale, seconde par seconde. La mise en œuvre d'un tel système fait partie de notre projet de thèse. Nous discutons la problématique scientifique du développement d'un tel système dans le chapitre qui suit.

L'implémentation de la composante *Comportement* repose sur la manière de faire le choix entre plusieurs émotions et la mise à jour de l'intensité des émotions existantes. Cette implémentation dépend de l'objectif du concepteur. Comme les sorties des composantes *Interprétation Cognitive* et *Interprétation Physiologique* sont des couples Valence – Activation, la composante *Comportement* doit combiner ces couples de valeurs pour en produire un couple de valeur représentant l'émotion choisie pour répondre à l'événement. Par exemple, pour le choix de l'émotion en réponse à la situation, si GRACE implémente un agent d'Affective Reasoner, il n'y a qu'une seule émotion à la fois, donc il n'a pas besoin de faire le choix entre des émotions concurrentes. En revanche, si GRACE implémente un agent de Cathexis, il peut y avoir plusieurs émotions à un moment donné, et il faut donc faire le choix en fonction de la motivation et de préférences de l'agent. Mais, si GRACE implémente l'agent Greta, le *Comportement* n'a pas besoin de choisir l'émotion à sélectionner pour répondre à la situation, parce que Greta a la capacité d'exprimer des émotions complexes, autrement dit une combinaison d'émotions, sur son visage. La mise à jour de l'intensité des émotions, quant à elle, prend en compte l'état émotionnel courant venant de la composante *Etat interne*. Une implémentation simple de cette composante peut être de prendre la moyenne des entrées de la composante, donc la sortie des composantes *Interprétation Physiologique, Interprétation Cognitive, Etat interne,* et *Intuition*.

L'implémentation de la composante *Etat interne* varie en fonction de l'application. Comme cette composante gère le changement d'état mental au cours du temps, une implémentation simple est de faire diminuer la valence et l'activation d'une quantité fixe par unité de temps. Il est possible aussi d'implémenter une diminution différée pour la valence et l'activation en définissant deux taux de changement différents. Cette composante s'occupe aussi de la personnalité, qui peut être traduite par la façon dont l'état émotionnel change au cours du temps. Par exemple, un personnage optimiste a tendance à facilement changer positivement son état émotionnel, tandis que quelqu'un de pessimiste a plus tendance à changer négativement son état émotionnel vis-à-vis d'un mauvais événement ou de l'absence d'événement. Un

personnage stable, par contre, possède un taux de changement d'état émotionnel très faible, ce qui peut être vu comme une résistance forte à son propre état émotionnel dans toutes les circonstances. Là on peut voir que la personnalité décide la fonction de diminution des émotions. La gestion de la motivation est aussi à la charge de la composante *Etat interne*. La motivation représente l'état désiré, qui, en quelque sorte, dépend de l'objectif et donc du plan d'action. Un exemple très simple de la motivation est l'état neutre, ce qui signifie qu'il n'y a pas d'état de préférence. Au final, une formule pour calculer l'état émotionnel peut être :

$$Etat_{A\_courant} = \frac{1}{2}\left(Fonc\_Per\left(Etat_{A\_courant}, Sortie\_Comportement_{A}\right) + Motiv_{A}\right)$$

$$Etat_{V\_courant} = \frac{1}{2}\left(Fonc\_Per\left(Etat_{V\_courant}, Sortie\_Comportement_{V}\right) + Motiv_{V}\right)$$

où V représente la Valence, A l'Activation, Fonc_Per est la fonction de la personnalité qui détermine le changement de l'état émotionnel (soit une diminution, soit une augmentation en fonction de l'état précédent ($Etat_{A\_courrant}$, $Etat_{V\_courrant}$) et de la sortie de la composante *Comportement* ($Sortie\_Comportement_{A}$, $Sortie\_Comportement_{V}$), $Motiv_{A}$ et $Motiv_{V}$ sont la valence et l'activation de l'état désiré.

L'implémentation de la composante *Expression* est fortement liée au système sur lequel est installé le modèle. Par exemple, si c'est un modèle graphique comme dans le cas de l'agent Greta, l'expression est liée aux dispositifs du modèle graphique, à savoir l'expression faciale fournie par Greta. Mais si GRACE est implémenté sur un robot personnel, l'expression des émotions varie en fonction de la capacité d'action du robot, comme la capacité à parler, simuler les traits de visage, ou à se déplacer dans l'environnement. Nous avons implémenté une instance de cette composante sur un robot. L'expression du robot consiste à faire des déplacements et des mouvements de la caméra (pour simuler le regard). Ces mouvements sont inspirés des mouvements des musiciens lors de leurs performances artistiques. La présentation complète de cette implémentation et des discussions scientifiques correspondantes se trouve dans le chapitre 4 de ce mémoire.

De manière générale, la complexité de l'implémentation de chaque composante de GRACE varie en fonction de l'objectif du concepteur. Cela peut être la mise en œuvre d'une formule très simple pour calculer la moyenne de toutes les entrées, mais cela peut être aussi la mise en œuvre d'un système d'apprentissage, ou bien un système d'expert pour faire du raisonnement en fonction de l'événement capturé. De plus, il existe encore des désaccords entre les psychologues sur le fonctionnement et même sur l'organisation des composantes dans un processus émotionnel, et l'implémentation des composantes abordées dans ce mémoire pourrait aboutir à des problématiques scientifiques pluridisciplinaires. Notre contribution dans cette thèse réside dans la proposition d'une architecture qui permet non seulement d'unifier les modèles computationnels existants mais aussi de rassembler les composantes du processus émotionnel mises en avant par les psychologues.

6. Conclusion

Tout au long de ce chapitre, nous avons balayé les trois traditions dans les théories psychologiques sur les émotions. Nous avons ensuite abordé la conception du modèle GRACE qui repose sur la théorie proposée par K. Scherer. Bien que la théorie de K. Scherer soit complète et couvre les différents éléments importants d'un processus émotionnel, la conception initiale de GRACE se prêtait difficilement à une implémentation. Dans le but d'instancier le modèle GRACE dans des applications computationnelles (comme un agent conversationnel animé, une assistance robotisée, la réalité virtuelle, etc.) nous avons effectué des modifications majeures dans la structuration du modèle GRACE. La généricité de cette nouvelle version du modèle GRACE est montrée via la projection de différents modèles computationnels des émotions existant vers le modèle GRACE.

La contribution de la thèse sur la modélisation du modèle GRACE consiste en des modifications majeures que nous avons apportées pendant la thèse. Ces modifications permettent d'isoler le développement de chaque composant du modèle, qui donne donc plus de flexibilité au niveau de l'implémentation informatique. L'unification des messages échangés entre les composants est également importante : elle permet de rendre le fonctionnement des composants plus indépendant les uns des autres. D'ailleurs, la comparaison entre cette nouvelle version de GRACE et les différents modèles computationnels des émotions existants montre aussi qu'avec GRACE, on arrive à bien isoler le processus émotionnel du processus de planification et de contrôle. Cela permet de rendre la modélisation des émotions plus indépendante des deux autres parties, qui permet de rendre l'implémentation de la partie « émotion » plus homogène sur des systèmes de natures différents (comme des robots, des agents conversationnels).

La généricité du modèle pourra être encore plus complète si l'on peut montrer comment le modèle GRACE peut s'interfacer avec les architectures générales d'agents (comme l'architecture BDI, l'architecture de subsomption, l'architecture hybride). Ces architectures s'intéressent en effet à toute la partie contrôle et planification d'actions dont nous avons souligné qu'elle ne devait pas être intégrée à GRACE car elle ne relevait pas directement des processus émotionnels. Cependant, il est bien connu que les émotions peuvent influencer les mécanismes de contrôle, et qu'à l'inverse l'activité de contrôle et de planification peut influencer les processus émotionnels. Ces architectures ont été utilisées pour le développement des différents agents et/ou robots émotionnels, comme l'agent Greta, le robot Kismet, les agents d'EMA, etc. Comme nous avons pu montrer dans ce chapitre qu'il était possible d'établir une projection de GRACE vers ces exemples, il semble raisonnable de penser qu'on puisse de même définir une interface entre GRACE d'une part et ces architectures générales d'agents d'autre part, ce qui reste à réaliser et à valider.

Chapitre 3

Indicateurs Musicaux

Nous allons présenter dans ce chapitre l'implémentation de la composante Interprétation Cognitive pour l'extraction des valeurs émotionnelles dans les événements musicaux. Nous discuterons de la problématique de la construction d'un tel extracteur en balayant la littérature du domaine de traitement de signaux musicaux. Nous présenterons aussi notre démarche pour construire notre propre extracteur et les résultats de recherche obtenus. Nous discuterons aussi les différents travaux en cours et nous conclurons le chapitre par quelques perspectives.

1. Introduction

Dans le rapport de (Rentfrow & Gosling, 2003), écouter de la musique a été listée parmi les activités de loisir les plus populaires. Cette dominance est en fait expliquée par le réconfort émotionnel que la musique offre à ses auditeurs (Rentfrow & Gosling, 2003). D'ailleurs, la musique est assez reconnue comme un médicament thérapeutique efficace pour guérir les désordres émotionnels ou bien pour réguler l'humeur ou l'émotion (Zentner, Grandjean, & Scherer, 2008). La recherche dans le domaine musical connaît ainsi un grand engouement non seulement en sciences sociales mais aussi en science de l'information et de l'ingénieur (Hu, 2010). En informatique, beaucoup de travaux de recherche portent sur l'extraction d'information musicale à partir des signaux sonores (pour une revue de ces travaux, voir (Wiering, 2007)), sur la synthèse de ces signaux pour la reproduction automatique (Weinberg & Driscoll, 2006), et sur l'utilisation de la musique pour l'assistance à la personne.

(Juslin & Västfjäll, 2008) proposent un cadre théorique intégrant six mécanismes additionnels (à côté de l'évaluation cognitive) qui décrivent l'induction émotionnelle par la musique. Ce sont (a) le réflexe du tronc cérébral, (b) le conditionnement évaluatif, (c) la contagion émotionnelle, (d) l'imagination visuelle, (e) la mémoire épisodique, et (f) l'attente musicale.

Le mécanisme *Réflexe du tronc cérébral* est le processus où l'émotion musicale est ressentie due à la détection des caractéristiques sonores fondamentales signalant un événement potentiellement important et urgent. Ce mécanisme est relié au processus d'audition de l'être vivant. Par exemple, des sons qui sont soudains, forts, dissonants, ou caractérisés par des patterns temporels rapides évoquent des sensations de désagrément aux auditeurs. Les émotions évoquées par ce mécanisme sont donc rapides et automatiques.

Le mécanisme *Conditionnement évaluatif* concerne le processus émotionnel où la musique est souvent associée avec un stimulus positif ou négatif. Par exemple, un morceau musical est joué à chaque fois que Jean-Jacques voit son meilleur ami. Avec le temps, via cette co-occurrence répétée, ce morceau peut évoquer de la joie à Jean-Jacques même si son ami n'est pas présent.

Comme son nom l'indique, le mécanisme de *Contagion émotionnelle* représente le cas où l'auditeur perçoit de l'émotion dans la musique et l'imite de manière interne

(physiquement, physiologiquement, etc.), ce qui lui fait ressentir la même émotion que celle dans la musique. Ce mécanisme a été beaucoup reporté dans les études psychologiques sur l'induction des émotions par la musique.

L'*imagination visuelle* décrit l'association d'une image virtuelle à la musique lors de l'écoute musicale. L'émotion ressentie est donc le résultat de l'interaction entre la musique et l'image virtuelle. Il existe des « thèmes » dans l'imagination visuelle en écoute musicale, les plus souvent cités sont des scènes de la nature (comme le soleil, le ciel, l'océan) et des expériences hors-du-corps (comme flotter dans le ciel). La nature de cette association n'est pas encore clairement décrite en psychologie, pourtant, l'effet de ce mécanisme est très souvent constaté et utilisé dans la thérapie musicale.

Le mécanisme de *Mémoire épisodique* correspond à l'évocation des émotions via une musique significativement liée à un événement important de l'auditeur, comme son mariage, sa première rencontre amoureuse, une victoire sportive, etc. Ce mécanisme se distingue clairement du mécanisme de *Conditionnement évaluatif*. Bien que les deux mécanismes soient des rappels du passé, la *Mémoire épisodique* demande un rappel conscient d'un événement passé qui préserve beaucoup d'information contextuelle. De plus, la *Mémoire épisodique* est organisée sous la forme d'une structure hiérarchique en trois niveaux : expérience unique, événements généraux, connaissances spécifiques d'un événement.

Le sixième mécanisme est l'*Attente musicale*. Il s'agit du processus où l'émotion est évoquée quand une caractéristique de la musique viole, retarde, ou confirme l'attente de l'auditeur sur la continuation de la musique. L'attente musicale concerne le genre d'attente qui engage la relation syntactique entre différentes parties d'une structure musicale. La musique, dans ce cas, est considérée comme un langage de communication qui doit donc respecter un certain nombre de règles de production.

Du point de vue de leur application dans GRACE, ces six mécanismes sont bien présents. Le *Réflexe du tronc cérébral* est pris en charge par l'*Interprétation Physiologique*, l'*Attente musicale* et le *conditionnement évaluatif* sont pris en compte par l'*Interprétation cognitive*. L'*imagination visuelle* peut être intégrée dans l'*Intuition*, la *mémoire épisodique* construit donc l'*expérience personnelle*, tandis que la *contagion émotionnelle* est exécutée dans l'*état interne*.

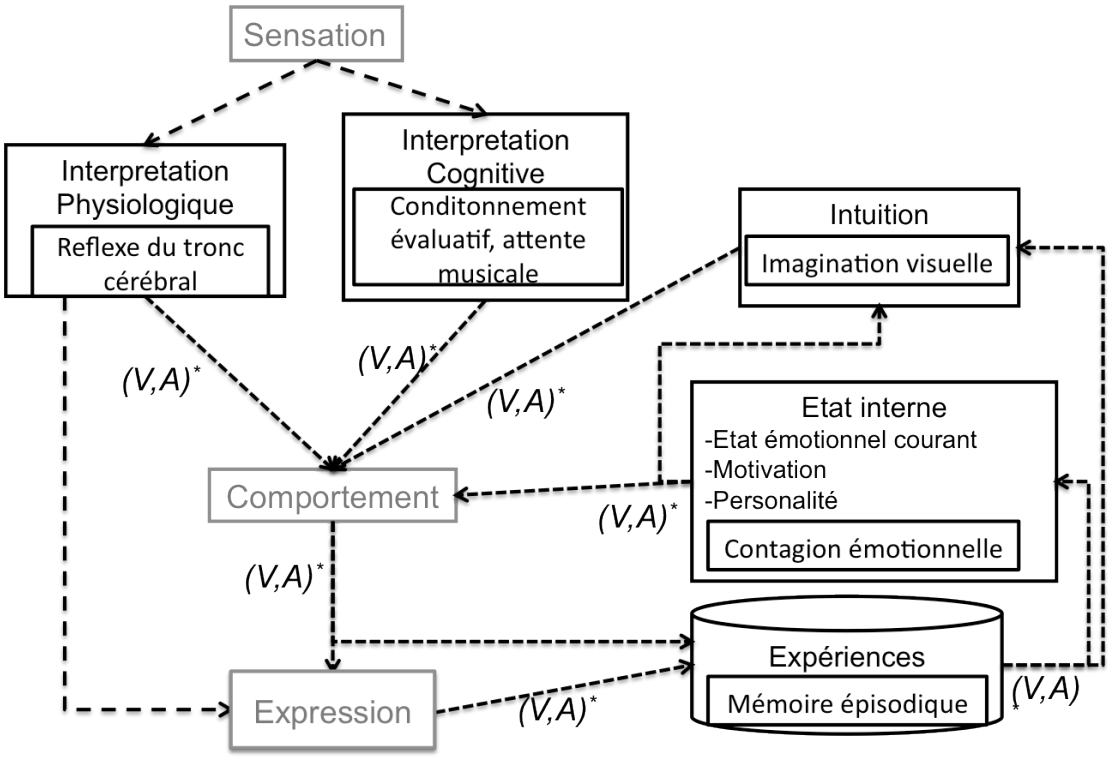


Figure 26 Intégration de six mécanismes d'induction des émotions par la musique

L'étude de l'extraction de contenu émotionnel dans la musique menée dans le cadre de la thèse est la mise en œuvre de l'*attente musicale* consistant à retrouver la sémantique émotionnelle dans la musique via ses descripteurs sonores.

Le niveau de l'information dans la musique peut varier de très élémentaire (bas niveau ou niveau physique) comme les caractéristiques sonores (la fréquence, le spectre, l'intensité, etc.) à très complexe (haut niveau ou niveau abstrait) comme l'émotion exprimée. Il existe un cadre conceptuel pour décrire le contenu dans la musique, informations distinguées par niveaux de représentation (voir Figure 27) (Leman, Vermeulen, De Voogdt, Taelman, Moelants, & Lesaffre, 2004). Ce cadre considère la musique en deux grandes classes descriptives : non contextuelle et contextuelle. La classe non contextuelle contient les descripteurs de bas niveau - ceux qui servent à décrire la musique en fonction de caractéristiques locales des signaux sonores, parmi lesquelles la fréquence, la durée du son, le spectre, l'intensité, etc.

<table>
<tr><th colspan="2">STRUCT</th><th colspan="2">NIVEAU CONCEPTUEL</th><th colspan="5">DESCRIPTEURS DE CONTENU MUSICAL</th></tr>
<tr><td rowspan="4">CONTEXTUEL</td><td rowspan="2">global >3 sec</td><td>HAUT II</td><td>EXPRESSIF</td><td colspan="5">Cognition ı émotion ı affectif = concept syntaxique+sémantique</td></tr>
<tr><td>HAUT I</td><td>FORMEL</td><td>mélodie
clé

profil</td><td>harmonie
tonalité

cadence</td><td>rythme
patterns rythmiques
tempo</td><td>source
instrument

voix</td><td>dynamique
trajet

articulation</td></tr>
<tr><td>global < 3 sec</td><td>MID</td><td>PERCEPTUEL</td><td>pattern d’intervalles successifs</td><td>pattern d’intervalles simultanés</td><td>battement

IOI</td><td>enveloppe spectrale</td><td>rang dynamique
niveau du son</td></tr>
<tr><td rowspan="3">NON-CONTEXTUEL</td><td>local + spatial</td><td>BAS II</td><td>SENSORIEL</td><td colspan="2">ton</td><td>temps</td><td>timbre</td><td>volume</td></tr>
<tr><td>local + temporel</td><td>BAS I</td><td>PHYSIQUE</td><td colspan="2">ton périodique
divergence du ton
fréquence fondamentale</td><td>durée de note

onset

offset</td><td>Rugosité

flux spectral

centroide spectral</td><td>energie neuronnale

sommet</td></tr>
<tr><td colspan="2"></td><td colspan="2">fréquence</td><td>durée</td><td>spectre</td><td>intensité</td></tr>
</table>

Figure 27 Différents niveaux de description de contenu musical

La classe contextuelle contient les descripteurs de niveau moyen et de haut niveau. Les descripteurs de niveau moyen considèrent les caractéristiques globales de la musique, comme les patterns sonores, la mélodie, le tempo, etc. En général, les descripteurs de niveau moyen sont estimés sur un intervalle d'environ 3 secondes, tandis que les descripteurs de bas niveau sont estimés sur un intervalle de moins de 3 secondes. Les descripteurs de haut niveau reflètent notamment l’aspect cognitif, émotionnel et/ou affectif de la musique. Ce niveau est pourtant flou et entraine de l'ambiguïté pour les travaux de l'extraction d'information. S'agit-il de cognition ? d'émotion ? d'affection ? Comment la musique exprime-elle ces aspects aux auditeurs ? (Zentner, Grandjean, & Scherer, 2008) suggère que l'humain est plus à l'aise pour annoter la musique en fonction des émotions basiques (comme la joie, la tristesse, la colère, etc.). Selon (Leman, Vermeulen, De Voogdt, Taelman, Moelants, & Lesaffre, 2004), les émotions dans la musique sont facilement distinguées via leur représentation dimensionnelle, notamment sur l'espace de Valence - Activation (voir la Figure 28).

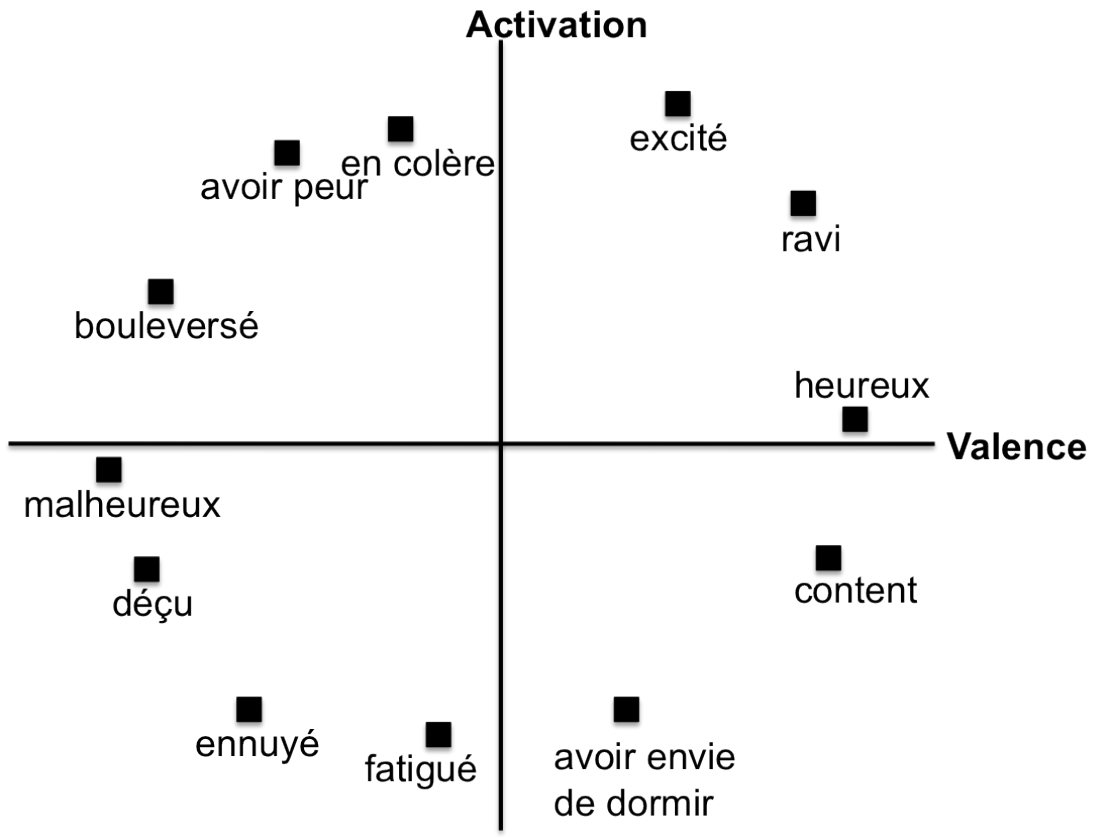


Figure 28 Distribution des émotions sur la surface Valence-Activation

Dans le contexte du travail de la thèse, nous nous intéressons à l'extraction du contenu émotionnel dans la musique, et donc à la construction d'un modèle permettant d'anticiper le contenu émotionnel à partir des descripteurs sonores (i.e. de bas niveau et de niveau moyen).

Dans ce chapitre, nous commençons par une étude de la littérature sur l'acquisition du contenu émotionnel dans la musique, puis nous passerons à la construction de notre système d'extraction de l'émotion véhiculée dans la musique à partir de six descripteurs : nombre de notes, nombre d'impulsions, fréquence, intensité, durée des notes jouées et valeur affective. Nous présenterons finalement notre contribution à la littérature du domaine et les perspectives sur le développement de notre travail.

2. Etat de l'art

Dans cette section nous allons voir différents travaux de recherche sur l'extraction du contenu émotionnel dans la musique. Dans un premier temps, trois travaux seront présentés pour justifier l'utilisation de la représentation dimensionnelle Valence – Activation pour modéliser le contenu émotionnel dans la musique. Ce sont des travaux en psychologie et en informatique qui cherchent à évaluer les effets mentaux (comme l'état émotionnel, la capacité de raisonnement) que l'écoute musicale apporte à l'auditeur. Parmi d'autres points, ils ont tous trouvé que l'effet émotionnel de la musique envers l'auditeur peut être représenté en deux dimensions : Valence et Activation. Cette représentation est en fait souvent utilisée dans les travaux informatiques sur l'extraction du contenu émotionnel dans la musique, dont quelques-uns sont présentés dans la suite de la section (i.e. la sous-section 2.2). Dans la sous-

section 2.2, nous allons voir quelques travaux portant sur la proposition des descripteurs musicaux qui permettent d'extraire automatiquement le contenu émotionnel dans la musique, soit pour construire des systèmes de recommandation de musique adaptés à l'état émotionnel de l'utilisateur, soit pour construire des systèmes d'aide à l'interprétation du contenu émotionnel dans la musique. Bien que ces travaux aient obtenu des résultats intéressants, différentes problématiques restent à étudier : nous les aborderons dans la conclusion de la section.

2.1. Choix des descripteurs musicaux

2.1.1. Effet du tempo et le mode musical à l'état émotionnel de l'auditeur

(Husain, Thompson, & Schellenberg, 2001) étudie le rôle de l'écoute musicale sur la capacité spatiale de visualisation[2] de l'humain et sur son état émotionnel. Les auteurs ont mené une expérimentation dans laquelle ils faisaient varier le tempo et le mode (majeur, mineur) de la musique pour examiner le changement dans la capacité spatiale, l'humeur, et l'activation de la personne. Le mot 'humeur' utilisé dans ce travail représente la valence, une des deux dimensions de la représentation 2D Valence - Activation de l'émotion, proposée par Russell (Russell, 1980). Pour leur expérimentation, le premier mouvement de la sonate de Mozart K.448 a été utilisé. Cet extrait est enregistré en format MIDI et reproduit en 4 versions : majeur - tempo rapide, mineur - tempo rapide, majeur - tempo lent, mineur – tempo lent. L'objectif était d'examiner l'influence du mode et du tempo sur le changement dans l'état émotionnel et dans la capacité spatiale de l'auditeur.

36 étudiants (8 hommes, 28 femmes ; âgés de 18 à 27 ans) ont participé à l'expérimentation. Ils sont testés individuellement. Chaque participant passait tout d'abord un test simplifié POMS (Profile Of Mood States en anglais) pour s'assurer qu'il ne souffrait pas d'une dépression clinique. Puis on demande à chacun d'évaluer son état émotionnel en Valence - Activation sur le plan 2D *Affect Grid*. Puis, il écoutait une des 4 versions de la sonate (aléatoirement choisie par un programme informatique). L'écoute dure environ 10 minutes. Après avoir fini l'écoute, il passait un test de capacité spatiale nommé PF&C conçu par Nantais et Schellenberg en 1999 (Nantais & Schellenberg, 1999). Il remplissait finalement à nouveau le plan Valence - Activation pour indiquer son état émotionnel et aussi indiquer comment il aimait l'extrait qu'il a écouté sur une échelle de 7 points.

Leur résultat relève que, parmi d'autres points, le mode (majeur/mineur) affecte la valence ressentie par l'auditeur en écoutant la musique. La musique en mode majeur hausse l'humeur des gens et la musique en mode mineur les rend plus négatifs. De son côté, le tempo affecte clairement le changement d'activation ressentie. Quand la musique est jouée avec un tempo rapide, l'activation ressentie après l'écoute monte par rapport à celle avant l'écoute. Et inversement, quand la musique est jouée à un tempo lent, l'activation diminue après l'écoute musicale. Leur résultat suggère aussi qu'il n'y a pas de relation entre le changement de mode musical et l'activation ressentie, et qu'il n'y a pas non plus de relation entre le changement de tempo et l'humeur. En ce qui concerne la capacité spatiale, ils ont trouvé que si l'auditeur

[2] La capacité spatiale de visualisation est la capacité de manipuler mentalement les figures 2D et 3D. Elle est typiquement mesurée via des tests cognitifs.

aimait l'extrait musical joué, il obtenait un meilleur résultat de test de capacité spatiale – ce qui affirme ce qu'ils appellent « l'effet Mozart » dans la psychologie.

2.1.2. Corrélation entre les descripteurs sonores et contextuels

L'objectif du travail de (Leman, Vermeulen, De Voogdt, Taelman, Moelants, & Lesaffre, 2004) est de vérifier la relation entre les descripteurs sonores et les descripteurs contextuels de la musique. Trois expérimentations ont été faites : la première pour calibrer les dimensions représentant la qualité affective/émotionnelle de la musique ; la deuxième analyse la relation entre les descripteurs sonores annotés par l'humain et ceux calculés automatiquement ; la troisième exploite la relation entre les valeurs affectives annotées par l'humain et les descripteurs sonores annotés par l'humain et ceux extraits automatiquement.

Dans la première expérimentation, 60 extraits (de divers genres : pop, rock, ethnique, jazz, classique) ont été annotés en fonction de leur qualité affective/émotionnelle. Dans ce travail, 15 adjectifs bipolaires représentant l'aspect sémantique/affectif de la musique ont été utilisés. 100 étudiants (73 filles et 27 garçons, d'âge moyen de 21 ans) participaient à l'expérimentation, chaque étudiant annotait 24 extraits aléatoirement sélectionnés parmi les 60 extraits, ce qui fait que chaque extrait est annoté 40 fois. En utilisant la méthode d'analyse factorielle (notamment l'estimation du maximum de vraisemblance avec la rotation varimax), les auteurs ont trouvé trois dimensions principales. La première dimension est la *valence* et exprime 19.5% des données. La deuxième dimension est l'*activité* et exprime 19% des données. La troisième dimension est l'*intérêt* qui exprime 18% de données. Au total, les trois dimensions expriment 56.5% de qualité affective des extraits annotés.

Pour la deuxième expérimentation, 11 descripteurs sonores extraits automatiquement et 6 descripteurs annotés manuellement sont utilisés pour le test de corrélation. Les 11 descripteurs sonores sont les moyennes et les écarts-types de 4 variables : la volume, la dureté (si la musique est brutale ou pas), la fréquence (i.e. la hauteur), le centroïde (i.e. le centre de gravité du volume des différentes instruments dans la musique), et 3 autres descripteurs - la durée moyenne entre des impulsions, le nombre d'impulsions pendant 30 secondes, et l'écart-type de la fréquence moyenne par mesure. Les 6 descripteurs annotés manuellement sont les moyennes et les écarts-types des trois variables suivantes : le tempo, l'harmonie perçue (consonances / dissonances) et le volume (faible / fort). Dix musicologues ont annoté les 60 extraits utilisés dans la première expérimentation selon les 6 descripteurs extraits manuellement décrits ci-dessus. L'analyse de corrélation montre que :

- l'harmonie moyenne est corrélée avec la dureté et la durée moyenne entre des impulsions ;
- le volume annoté manuellement est corrélé avec le volume et la dureté extraits automatiquement ;
- le tempo est corrélé avec la durée moyenne entre les impulsions et le nombre d'impulsions.

La troisième expérimentation consiste à analyser la relation entre les descripteurs sonores extraits automatiquement et les descripteurs affectifs retirés de la première expérimentation. Un groupe de 8 personnes a annoté les 60 extraits musicaux selon les 15 adjectifs utilisés dans la première expérimentation. Puis leurs annotations ont été mises en correspondance dans l'espace à 3 dimensions *valence – activité – intérêt*. En utilisant l'analyse de régression, les auteurs ont analysé la corrélation entre les

annotations affectives et les descripteurs proposés dans les deux expérimentations précédentes. Leur résultat montre que, pour les descripteurs sonores extraits automatiquement, la dimension *valence* a une corrélation significative avec le nombre d'impulsions et la durée moyenne entre les impulsions, la dimension *activité* est corrélée avec la dureté, et la dimension *intérêt* n'est bien corrélé avec aucun descripteur. Et pour les descripteurs annotés manuellement, la *valence* est corrélée avec l'harmonie et le tempo, l'*activité* est corrélée avec le volume annoté manuellement, et la dimension *intérêt* n'est corrélée avec aucun descripteur.

Leurs expérimentations permettent de conclure que l'annotation humaine sur la qualité affective/émotionnelle est cohérente d'une personne à l'autre, et qu'il existe une connexion entre les descripteurs sonores extraits automatiquement et la qualité affective/émotionnelle de la musique. Cela signifie donc qu'à partir des descripteurs sonores, on pourrait récupérer la qualité affective/émotionnelle de la musique. De plus, ce travail montre aussi que la représentation 2D en Valence – Activation de l'émotion est compatible pour représenter l'émotion véhiculée dans la musique. Ceci est donc un indice intéressant pour les travaux de recherche sur la construction des systèmes d'extraction automatique du contenu émotionnel dans la musique.

2.1.3. Corrélation entre les descripteurs sonores et émotionnels

(Bresin & Friberg, 2011) s'intéressent à vérifier les descripteurs musicaux qui permettent la communication des émotions dans la musique. Pour ce faire, ils ont mené une expérimentation pour tester le rôle des 7 descripteurs musicaux suivants : tempo, volume, articulation, phrase musicale, registre, timbre et vitesse d'attaque. Pendant l'expérimentation, ils ont demandé à 20 artistes de manipuler ces 7 descripteurs pour communiquer 5 émotions (neutre, joie, tristesse, peur, et sérénité) sur 4 partitions différentes. Ces partitions sont composées pour communiquer les quatre émotions (joie, tristesse, peur, calme) respectivement. La musique était reproduite en utilisant un synthétiseur en temps-réel ce qui permet au participant de savoir toute de suite comment ses modifications sont prises en compte dans la musique. Ils pouvaient donc varier les valeurs des descripteurs autant de fois qu'ils voulaient pour finalement exprimer l'émotion en question via la musique.

Les modifications possibles sur les 7 descripteurs sont décrites dans les paragraphes suivants.

- Tempo : Le tempo de l'extrait en question varie entre 10 fois plus lent que le tempo original de l'extrait jusqu'à 4 fois plus rapide que l'original.
- Volume : le volume d'un extrait peut être modifié de 62 dB à 82 dB
- Articulation : Durant l'expérimentation, le participant peut faire varier l'articulation de l'extrait de *legato* à *staccatissimo*.
- Phrase musicale : Typiquement dans la musique, le musicien tend à jouer la musique plus vite (*accelerando*) et plus fort (*crescendo*) vers le milieu de la phrase, et il tend à jouer la musique plus lentement (*rallentando*) et plus doucement (*decrescendo*) vers la fin de la phrase musicale. Il est indiqué dans leur travail que *forward phrasing* (i .e. accelerando-crescendo suivi par rallentando-decrescendo) peut être utilisé pour communiquer la tristesse et la tendresse, tandis que *reverse phrasing* (i.e. rallentando-decrescendo suivi par accelerando-crescendo) peut être utilisé pour communiquer une performance

agressive. Pendant l'expérimentation, le participant pouvait faire varier la phrase musicale en le changeant entre *forward phrasing* et *reverse phrasing*.

- Durée d'attaque : Pendant l'expérimentation, le participant pouvait ajuster la durée d'attaque des notes de très lentement (75% de la durée jouée de la note mais n'excède pas 1 seconde) à très rapide (instantané).

- Registre : Pendant l'expérimentation, le participant pouvait modifier la tonalité entre [-24 demi-ton, +24 demi-ton] par rapport à la valeur du registre original pour communiquer l'émotion en question.

- Timbre : En ce qui concerne le timbre, la musique est reproduite avec le son du piano comme accompagnement, puis un autre instrument pour jouer la mélodie. L'instrument pour jouer la mélodie est soit un cor d'harmonie, soit une flute, soit une trompette.

Leur résultat d'expérimentation confirme la corrélation entre différents descripteurs et les émotions communiquées par la musique. Plus précisément, le tempo, le volume, l'articulation sont corrélés significativement avec l'activation émotionnelle (i.e. l'émotion est représentée en 2D Valence-Activation) ; le registre est corrélé avec la valence émotionnelle ; concernant le choix de l'instrument, la trompette est principalement choisie pour communiquer la joie, le cor d'harmonie est principalement choisi pour la peur et la tristesse, tandis que la flute est souvent choisie pour exprimer le neutre, la sérénité et la tristesse. Les modifications sur les descripteurs sont indépendantes de l'extrait utilisé.

Leur travail est intéressant non seulement grâce à sa confirmation de la corrélation entre les descripteurs musicaux et les émotions exprimées dans la musique mais il permet aussi de déterminer les valeurs exactes de ces descripteurs pour une manipulation automatique par les machines. Grâce aux valeurs précises découvertes lors de leur travail, la musique synthétique peut être reproduite de façon fiable pour communiquer différentes émotions, ce qui est très intéressant à appliquer dans les travaux sur l'interaction homme-machine.

2.2. Extraction des valeurs émotionnelles dans la musique

Contrairement à la représentation simple du contenu émotionnel dans la musique (i.e. en Valence – Activation), les descripteurs musicaux à choisir pour bien extraire ce contenu émotionnel ne sont pas si simples à déterminer. Nous allons voir dans la suite trois travaux récents portant sur le choix des descripteurs musicaux pour extraire l'émotion véhiculée dans la musique. Leurs résultats sont intéressants, mais il reste aussi des problématiques difficiles à résoudre.

2.2.1. Système de recommandation musicale

Ce travail consiste à construire un système de classification des extraits musicaux en fonction de l'émotion exprimée (représentée en Valence - Activation, ou bien en valeurs affectives) à partir des descripteurs musicaux. Leur travail consistait à construire une base des extraits monotones en termes de l'émotion exprimée (c'est-à-dire n'exprimant qu'une seule émotion sur toute la durée de l'extrait), et à sélectionner les descripteurs pertinents pour la classification des extraits en fonction des émotions.

La construction de la base musicale est faite à l'aide du modèle LBDM – Local Boudary Detection Model. Pour un morceau musical quelconque (en codage MIDI), LBDM positionne les marqueurs le long du morceau pour marquer les grandes variations musicales (comme la fréquence, le rythme et les silences). Ils arguent que les extraits avec différents descripteurs musicaux peuvent exprimer différentes valeurs affectives. Pour avoir des extraits monotones en termes d'émotion exprimée, les chercheurs ont découpé les morceaux en fonction de ces marqueurs. Au final, ils ont obtenu 96 morceaux de musique occidentale, chacun durant de 30 à 60 secondes. Ils demandaient ensuite à 80 auditeurs d'annoter en ligne la valeur affective qu'ils percevaient de chaque morceau. Cette base est utilisée comme base d'apprentissage et de test pour le développement de leur système de classification.

Pour déterminer les descripteurs pertinents pour la classification des extraits musicaux, les auteurs ont examiné 146 descripteurs unidimensionnels et 3 descripteurs multidimensionnels, classés en 6 catégories : instrumentation (20 descripteurs), textures (15), rythme (39), dynamique (4), mélodie (68), et harmonie (13).

Ces descripteurs sont premièrement évalués en termes de leur corrélation avec les valeurs affectives associées (celles de la base des extraits). Chaque descripteur est associé à deux poids représentant leur corrélation avec les valeurs affectives : la valence et l'activation (voir Figure 28). Deuxièmement, une phase de régression utilisant la méthode SVM (Machine à Vecteurs de Support – ou Support Vector Machine en anglais) est appliquée pour raffiner les poids des descripteurs. Puis, un algorithme de sélection de descripteurs est mis en place (via le logiciel Weka) pour réduire le nombre de descripteurs et améliorer la classification des extraits. Après cette phase, 26 descripteurs sont retenus pour la valence et 23 descripteurs sont retenus pour l'activation. Dans l'ensemble des descripteurs retenus, les auteurs effectuaient une sélection manuelle du meilleur groupe de descripteurs importants retenus à l'étape précédente et qui leur permet de transformer facilement les extraits afin de changer l'émotion exprimée. A l'issue de cette étape, 5 descripteurs sont retenus pour la valence, et 5 descripteurs pour l'activation. Avec la validation croisée, le coefficient de corrélation et le coefficient de détermination pour la valence sont de 71.5% et 51.12% respectivement ; ceux pour l'activation sont de 79.14% et 62.63%. Les formules pour calculer la valence et l'activation sont les suivantes :

valence = -0.41*durée moyenne des notes + 0.17*écart dominante[3] + 0.41*tempo initial – 0.18*mode clé + 0.24*position de climax

activation = -0.56*durée moyenne des notes + 0.24*tempo initial + 0.11*position de climax + 0.37*fréquences identiques consécutives + 0.58*densité des notes

Ces chercheurs ont aussi mené la même démarche sur d'autres types musicaux (pop et r&b). Dans cette étude, ils ont 16 extraits musicaux, annotés en ligne par 53 auditeurs. L'ensemble des descripteurs de départ était constitué de 106 descripteurs calculés par les logiciel JSymbolic et MIDI toolbox. Ces 106 descripteurs sont unidimensionnels et sont classés en 7 catégories : mélodie, rythme, instrumentation, harmonie, dynamique, fréquence, et texture. Pour la sélection des descripteurs, la même démarche est appliquée. Au final, il y a 3 descripteurs retenus pour la valence et 5

[3] Nombre maximal des octaves consécutives jouées.

descripteurs retenus pour l'activation. Avec la validation croisée, le coefficient de corrélation et le coefficient de détermination pour la valence sont de 75.98% et 57.73% respectivement ; ceux pour l'activation sont de 81.85% et 66.99%. Les formules de calcul pour la valence et l'activation sont les suivantes :

valence = -0.45*temps entre les impulsions + 0.11*densité des notes – 0.54*variabilité de durée des notes

activation = -0.56*durée moyenne des notes – 0.31*temps moyen entre les impulsions – 0.45*importance du registre élevé + 0.06*densité des notes + 0.05*variation de la dynamique

De manière générale, à partir d'un ensemble de descripteurs, les auteurs ont trouvé différents sous-ensembles de descripteurs pour extraire différents types de musique. Les taux de corrélation sont élevés et montrent donc la performance de leur système de recommandation. Pourtant, l'extraction du contenu émotionnel dans la musique doit être faite en temps réel pour une interaction naturelle avec l'humain, ce qui fait que ce travail nécessite une validation en temps réel (i.e. seconde par seconde par exemple) avec l'utilisateur pour s'adapter au contexte d'interaction naturelle. Nous allons voir maintenant deux travaux qui se positionnent dans ce cadre, mieux adapté au contexte d'interaction naturelle.

2.2.2. Anticipation de l'émotion musicale

Ce travail s'intéresse à la construction d'un système mesurant le contenu émotionnel dans la musique. Les auteurs désiraient construire un système qui soit capable de mesurer, à un instant quelconque, l'émotion exprimée en se basant sur les caractéristiques acoustiques dans la musique jouée.

L'émotion est annotée seconde par seconde en utilisant *EmotionSpace Lab* – une interface dédiée à cette effet. Chaque émotion est un couple de valeurs (Valence, Activation). Chaque valeur varie de -100 à 100. Les auditeurs humains annotent l'émotion perçue tout au long de l'écoute, ce qui donne une base d'annotation plus précise et plus complexe à traiter. Pour leur projet, 6 extraits de musique classique sont utilisés, qui font au total 18 minutes et 38 secondes. Il y avait 35 volontaires participant à l'annotation de ces extraits musicaux. Pour la construction du système, l'annotation médiane (parmi les 35 annotations) de chaque extrait est utilisée.

Les descripteurs musicaux utilisés dans leur projet sont inspirés d'un travail antérieur de Schubert (Schubert, 1999). Ce sont 18 descripteurs (voir Figure 29 ci-dessous) représentant 7 caractéristiques de la musique : la dynamique, la fréquence moyenne, la variation de la fréquence, le timbre, l'harmonie, le tempo, et la texture. A l'exception du tempo qui est extrait manuellement selon la méthode décrite dans le travail de thèse de Schubert (Schubert, 1999), ces descripteurs sont extraits via le logiciel PsySound et l'extracteur de Fourrier de MARSYAS.

No.	Musical Property	Musical Feature	Extraction Method
1	Dynamics	Loudness Level	PsySound
2		Short Term Max. Loudness	PsySound
3	Mean Pitch	Power Spectrum Centroid	PsySound
4		Mean STFT Centroid	MARSYAS
5	Pitch Variation	Mean STFT Flux	MARSYAS
6		Std. Dev. STFT Flux	MARSYAS
7		Std. Dev. STFT Centroid	MARSYAS
8	Timbre	Timbral Width	PsySound
9		Mean STFT Rolloff	MARSYAS
10		Std. Dev. STFT Rolloff	MARSYAS
11		Sharpness (Zwicker and Fastl)	PsySound
12	Harmony	Spectral Dissonance (Hutchinson and Knopoff)	PsySound
13		Spectral Dissonance (Sethares)	PsySound
14		Tonal Dissonance (Hutchinson and Knopoff)	PsySound
15		Tonal Dissonance (Sethares)	PsySound
16		Complex Tonalness	PsySound
17	Tempo	Beats per Minute	Schubert's method
18	Texture	Multiplicity	PsySound

Figure 29 Listes des descripteurs utilisés dans le travail de Korhonen et al.

Deux modèles linéaires ont été testés pour leur système : ce sont l'autorégression avec l'entrée additionnelles (AutoRegression with eXtra inputs - ARX) et le modèle d'espace d'états (state-space model structure). Ils ont construit et testé 12 modèles d'espace d'état et 45 modèles ARX, avant de retenir un modèle ARX surclassant les autres. Ce modèle produit un coefficient de détermination[4] R^2 de 21.9% pour la valence et 78.4% pour l'activation. Ce modèle utilise 16 descripteurs parmi les 18 descripteurs listés dans la figure ci-dessus. Les deux descripteurs qui ne sont pas retenus sont Spectral Dissonance (Hutchinson and Knopoff) et Tonal Dissonance (Hutchinson and Knopoff). En analysant leur modèle, ils concluaient aussi que la valence pourrait dépendre de l'activation, tandis que l'activation pourrait être calculée à partir des descripteurs acoustiques eux-mêmes, indépendamment de la valence.

2.2.3. Anticipation de la valence musicale

Etant intéressés par l'interprétation de la valence dans la musique, les auteurs ont mené une recherche sur la relation entre la valence et les descripteurs musicaux (Fornari & Eerola, 2009). Leur but est aussi de construire un système qui pourrait mesurer/prédire la valence dans la musique. Il faut noter qu'ici, la valence à apprendre est la valence annotée seconde par seconde tout au long des extraits musicaux – à distinguer des travaux où la valence prise en compte est une valeur associée de manière globale à chaque extrait.

Ils arguaient de la difficulté de mesurer la valence constatée dans la littérature car ces travaux se basaient sur des descripteurs de bas niveau tandis que la valence est fortement liée aux descripteurs contextuels. Par conséquent, ils ont proposé 8

[4] Le coefficient de détermination R^2 est un indicateur permettant de juger la qualité d'une régression linéaire. Sa valeur est inférieur ou égal à 1 reflète l'adéquation entre le modèle et les données observées. (voir : http://www.jybaudot.fr/Correlations/coeffdeterm.html).

descripteurs de haut niveau, dans le cadre de leur projet *Braintuning*, pour estimer la valence. Les huit descripteurs sont :

- *clarté d'impulsion* (Pulse Clarity) mesure la clarté de la perception humain sur des impulsions musicales ; elle varie de « impulsion non percevable » à « perception forte des impulsions ».
- *clarté du ton* (Key Clarity) mesure la clarté de la perception de la tonalité dans la musique ; sa valeur est comprise entre 0 (atonal) et 1 (tonal).
- *complexité harmonique* (Harmony Complexity) mesure la perception de l'entropie des sons dans la musique. En fait, la complexité musicale peut être considérée comme l'entropie des sons. La *complexité harmonique* mesure donc la perception de cette entropie, pas l'entropie elle-même. Par exemple, un bruit blanc est un son complexe en terme d'acoustique, mais en terme de perception auditoire, il est simple et constant. La valeur de la *complexité harmonique* varie entre 0 (pas de complexité harmonique perceptible) à 1 (perception forte de l'harmonie complexe).
- *articulation* varie entre 0 (il y a des pauses perceptibles entre les notes jouées) et 1 (pas de pause perceptible entre des notes jouées).
- *répétition* détecte la présence des patterns répétitifs dans la musique, qui peuvent être mélodiques, harmoniques ou rythmiques. Sa valeur varie entre 0 (pas de répétition remarquable) et 1 (répétition d'un pattern clairement remarquée).
- *mode* fait référence au mode de la musique (i .e. majeur/mineur). Sa valeur varie entre 0 (mode mineur) et 1 (mode majeur). Son utilisation est plutôt pour distinguer entre les extraits majeurs et les extraits mineurs.
- *densité des événements* reflète le nombre d'événements musicaux par mesure (ou unité de temps) que l'on peut percevoir, ces événements pouvant être mélodiques, harmoniques ou rythmiques. Sa valeur varie entre 0 (perception d'un seul événement) et 1 (perception du nombre maximal d'événement possibles par un auditeur).
- *Brightness* estime la sensation de vivacité dans la musique, souvent corrélée avec les fréquences élevées, l'articulation, des impulsions, etc. Sa valeur varie entre 0 (pas de vivacité dans la musique) et 1 (l'extrait est vivace).

Pour la prédiction de la valence, ils ont construit un modèle linéaire de régression multiple avec ces 8 descripteurs contextuels. Ce modèle a été calibré pour pouvoir estimer la valence d'un extrait musical (le « Concerto d'Aranjuez » de Joaquin Rodrigo) qui a été traité dans deux travaux antérieurs (celui de (Schubert, 1999) et celui de (Korhonen & Clausi, 2006)). Leur résultat montre que leur modèle peut prédire 42% de la valence de cet extrait – un résultat majorant ceux de deux travaux antérieurs sur cet extrait.

	Modèle de Schubert	Modèle de Korhonen	Modèle de Fornari &Eerola	Densité d'événement
Type du modèle	OLS	ARX	Régression multiple	Descripteur
R^2	33%	-88%	42%	35%

De plus, le descripteur *Densité des événements* semble avoir beaucoup d'influence sur l'interprétation de la valence de cet extrait. D'ailleurs, en analysant la corrélation entre les 8 descripteurs et la valence de cet extrait, le *mode* et la *clarté du ton* sont les

moins corrélés avec la valence, ce qui est contraire à ce qui est dit dans la littérature du domaine.

Ce travail ne traitant que d'un seul extrait de 2 minutes et 45 secondes, leur résultat, bien qu'intéressant, demande une validation sur une base de données plus grande pour justifier leur proposition.

2.3. Conclusion

L'extraction de l'émotion dans la musique à partir des descripteurs sonores présente encore beaucoup de défis. Tandis que l'extraction de l'émotion dans un extrait donné de 30 secondes ou plus donne de bons résultats comme rapporté dans le travail de (Oliveira & Cardoso, 2008), l'extraction de l'émotion à chaque seconde de la musique ne semble pas évidente. Le choix des descripteurs varie d'un travail à l'autre, et même le nombre des descripteurs varie fortement (par exemple, plus de 100 descripteurs dans (Oliveira & Cardoso, 2008), 18 descripteurs dans le travail de (Korhonen & Clausi, 2006), 8 descripteurs dans le travail de (Fornari & Eerola, 2009)). L'utilisation de différents logiciels pour l'extraction des descripteurs ajoute aussi à la difficulté de ré-utilisation de ces résultats. Combien des descripteurs faut-il pour l'extraction de l'émotion dans la musique ? Est-ce que cela nécessite des descripteurs de haut niveau pour extraire la valence comme le travail de (Leman, Vermeulen, De Voogdt, Taelman, Moelants, & Lesaffre, 2004) et de (Fornari & Eerola, 2009) le proposent ? Nous allons traiter ces questions dans la suite.

3. Extraction automatique de valeur émotionnelle dans la musique

3.1. Méthodologie de recherche

Notre objectif est de construire un système d'extraction automatique des émotions véhiculées par la musique. Pour ce faire, il nous faut un système qui répond aux exigences suivantes :

- L'émotion dans la musique est extraite seconde par seconde. Ceci est dû au fait que l'émotion ressentie lors de l'écoute musicale varie au cours du temps. Ceci est dû aussi au fait que le système d'extraction final va être intégré dans le modèle des émotions GRACE pour des interactions avec l'humain, ce qui demande donc de pouvoir répondre à des informations capturées en temps réel.
- Les descripteurs musicaux doivent être choisis d'une telle façon que la performance du système d'extraction soit la meilleure possible mais que le nombre d'indicateurs soit le minimal. En fait, dans les travaux présentés dans la section précédente, de grands nombres de descripteurs sont utilisés pour construire les systèmes d'analyse. De plus, les descripteurs suggérés en fonction du résultat d'analyse des systèmes en question diffèrent de l'un à l'autre. L'utilisation d'un grand nombre de descripteurs permettrait plus de précision dans l'extraction du contenu émotionnel. Pourtant, cette utilisation apporte un risque de sur-apprentissage, c'est-à-dire qu'on aurait un ensemble des descripteurs qui fonctionnent bien avec la base de données sur laquelle on les a utilisés mais qu'ils n'assurent pas une bonne performance de reconnaissance sur les données qui ne sont pas dans la base de données de départ. L'utilisation d'un nombre réduit de descripteurs est donc souhaitable. Le résultat du travail de (Fornari & Eerola, 2009) montre qu'avec des descripteurs bien définis, l'extraction des valeurs

émotionnelles dans la musique peut donner de bons résultats en utilisant peu de descripteurs (huit descripteurs dans le cas du travail de (Fornari & Eerola, 2009)). Notre objectif dans ce projet est donc d'essayer de définir un ensemble de descripteurs musicaux, aussi réduit et efficace que possible pour l'extraction du contenu émotionnel dans la musique.

- Le système doit être construit sur une base de données cohérente, ce qui implique l'utilisation de morceaux musicaux d'un seul genre ou un seul compositeur. Cette exigence a pour but de limiter la diversité dans le style musical de différents compositeurs et différents styles de musiques (comme le classique, la pop, le r&b, etc).
- Le système doit pouvoir généraliser sur le style musical sur lequel il est construit. Cette capacité de généralisation a pour but de s'assurer que les indicateurs musicaux choisis sont représentatifs pour l'extraction de l'émotion pour ce style musical en général et non pas seulement pour la base de données sur laquelle l'apprentissage est effectué.

Pour construire un tel système, nous avons suivi les étapes suivantes :

- Choisir les descripteurs musicaux à utiliser : cette phase consiste donc à définir l'ensemble des descripteurs musicaux à utiliser pour l'extraction du contenu émotionnel dans la musique. Elle comprend aussi la détermination de la représentation du contenu musical à utiliser, soit la représentation dimensionnelle en valence et activation, soit la représentation en émotions basiques.
- Définir le protocole pour obtenir les entrées et les sorties du système. Cette phase consiste à déterminer les extraits musicaux choisis pour la base de données, à extraire des descripteurs musicaux de ces extraits, et à obtenir l'annotation émotionnelle de l'expert sur ces extraits pour construire la base d'apprentissage et de test pour le système.
- Conduire le processus d'obtention des sorties du système. Il s'agit du processus d'annotation des extraits par un expert.
- Choisir une structure du système et les critères d'évaluation. A cette étape, il faut déterminer la méthode d'apprentissage qui permet qu'à partir des valeurs des descripteurs en entrée, le système arrive à reproduire la valeur émotionnelle annotée (en valence et activation) par l'expert humain. L'évaluation de la performance du système consiste à évaluer la précision du système sur la base d'apprentissage et sur la base de test.
- Entrainer le système.
- Evaluer la validité du système. Cette évaluation comprend l'évaluation des descripteurs proposés et l'évaluation de la performance de notre système par rapport à des résultats existants dans la littérature.

Ces étapes sont détaillées dans les sections suivantes.

3.2. Protocole pour obtenir les entrées et les sorties du système

La construction d'un système d'extraction des émotions véhiculées nécessite deux types de données : des entrées, et des sorties. Les entrées sont les valeurs des descripteurs musicaux, et les sorties sont les valeurs émotionnelles à reproduire par le système. Nous allons décrire, dans cette section, notre motivation pour le choix des entrées et des sorties de notre système, et ainsi la procédure d'obtention de ces données pour la construction de notre système.

3.2.1. Les entrées : les descripteurs musicaux choisis

Nous avons tout d'abord construit un ensemble de descripteurs dont nous faisons l'hypothèse qu'ils sont suffisants pour caractériser le contenu émotionnel de la musique. Comme décrit dans la section bibliographique, le choix et même le nombre des descripteurs varie fortement d'un travail à l'autre. Nous avons donc collaboré avec des musicologues (Damien Erhrardt et Roméo Agid) pour construire cette liste de descripteurs : le nombre de notes, le nombre d'impulsions, la hauteur des notes, la valeur affective (qui représente le mode musical, i.e. majeur ou mineur), l'intensité des notes (i.e. le volume), et la durée des notes. La définition de chaque descripteur est présentée comme suit :

Nombre de notes : c'est le nombre de notes jouées pendant l'intervalle, ce qui inclut les notes qui commencent dans l'intervalle ainsi que les notes commencées avant le début de l'intervalle et qui se prolongent au cours de l'intervalle, voire après l'intervalle.

Nombre d'Impulsions : c'est le nombre d'impulsions sur l'instrument pendant l'intervalle considéré. A titre d'exemple, un accord de trois notes jouées simultanément sera comptabilisé comme trois notes pour le descripteur Nombre de notes mais comme une seule impulsion.

Durée : c'est le pourcentage de la durée moyenne d'une note jouée pendant l'intervalle considéré. Par exemple, considérons un intervalle de n secondes avec nb notes jouées, chaque note i est jouée pendant d_i secondes dans l'intervalle. La valeur du descripteur Durée est calculée selon la formule suivante :

$$Durée = \frac{\frac{\sum_{i=1}^{nb} d_i}{nb} * 100}{n}$$

Hauteur : c'est la hauteur moyenne des notes jouées pendant l'intervalle. Dans le cadre de la thèse, la hauteur moyenne est calculée à partir des hauteurs des notes jouées pendant l'intervalle pondérées par la durée jouée de chaque note dans l'intervalle. Considérons un intervalle de n secondes avec nb notes jouées, la note i a une hauteur de h_i et est jouée pendant d_i secondes dans l'intervalle. La valeur du descripteur Hauteur est calculée selon la formule suivante :

$$Hauteur = \frac{\sum_{i=1}^{nb} \left(h_i * \frac{d_i}{n} \right)}{nb}$$

Intensité : c'est l'intensité des notes jouées pendant l'intervalle. De même, le calcul de l'intensité utilise aussi l'intensité de chaque note jouée, pondérée par sa durée dans l'intervalle. Considérons un intervalle de n secondes avec nb notes jouées, la note i a une intensité de I_i et est jouée pendant d_i secondes dans l'intervalle. La valeur du descripteur Intensité est calculée selon la formule suivante :

$$Intensité = \frac{\sum_{i=1}^{nb} \left(I_i * \frac{d_i}{n} \right)}{nb}$$

Valeur affective : Pour calculer la valeur affective, on considère le nombre des accords joués pendant l'intervalle en question. Chaque accord est associé à une valeur émotionnelle (voir la Table 24 dans l'Annexe 2) et à une durée de présence. Cette durée de présence est la durée pendant laquelle l'accord est joué sur l'intervalle, compris entre 0-*non_joué* jusqu'à 1_*jouée tout au long de l'intervalle*, et considéré comme un coefficient de poids. La valeur affective d'un intervalle est donc la moyenne des valeurs émotionnelle de ces accords pondérés par les durées de présence.

$$ValAff = \frac{\sum_i \left(ValEmo\_Accord_i * poids\_présence_i\right)}{nbAccord}$$

Pour la hauteur et l'intensité, nous ajoutons aussi leurs écarts-types pour représenter la dispersion de ces caractéristiques dans l'intervalle considéré.

Les morceaux utilisés dans le projet sont en codage MIDI, ce qui nous permet d'avoir les valeurs acoustiques (comme la hauteur, l'intensité, la durée jouée) exactes de chaque note jouée dans le morceau. Un exemple de calcul de ces descripteurs est présenté en Annexe 2.

Pour chaque extrait musical, nous avons la possibilité de calculer ces descripteurs en différents intervalles de temps. On a aussi la possibilité d'ajouter la dérivée première et la dérivée seconde pour chacun des descripteurs. En fait, (Fornari & Eerola, 2009) propose que pour déterminer la valence, il faut considérer les intervalles de temps plus longs que 1 seconde, et il est donc important de tester avec quel intervalle du temps on arrive le mieux à extraire la valence dans la musique.

Par ailleurs, nous allons également tester notre système avec les descripteurs estimés en fonction du temps musical – qui se base sur le tempo de la musique. L'utilisation de l'unité de temps musical sert à prendre en compte l'influence du tempo dans la musique. En fait, avec une même partition, le changement de tempo lors de l'interprétation de la partition entraîne des changements dans la durée (en secondes, minutes, etc.) de l'extrait musical tandis que les caractéristiques des notes jouées restent les mêmes. Ceci amène la question suivante : si on arrive à conserver ces caractéristiques, est-ce que la performance du système d'extraction sera meilleure ou pas ? Nous allons donc ajouter ces descripteurs en temps musical comme un autre type d'entrées du système à tester. La transformation depuis l'unité classique vers l'unité de temps musical est faite pour les descripteurs : nombre de notes, nombre d'impulsions, durée des notes jouées en fonction de l'intervalle. Les formules de transformation sont les suivantes :

$$nbNotes_{enTempsMusical} = \frac{nbNotes_{enUniteClasique}}{\text{int}ervalle} * \frac{60}{tempo}$$

$$nb\text{Im}pls_{enTempsMusical} = \frac{nb\text{Im}pls_{enUniteClasique}}{\text{int}ervalle} * \frac{60}{tempo}$$

$$durée_{enTempsMusical} = \frac{duréeToutesNotes_{enUniteClasique}}{nbNotes_{enUniteClasique}} * \frac{tempo}{60}$$

En termes d'échelle des descripteurs, la fréquence, le volume et leurs écarts-types respectifs varient dans l'échelle du codage MIDI, donc entre 0 à 127. Le nombre de notes, le nombre d'impulsions varient en fonction des notes et des impulsions dans la musique, et il n'y a donc pas une échelle particulière. La durée moyenne des notes est la valeur moyenne des durées jouées de toutes les notes considérées pendant un intervalle du temps. Cette valeur moyenne est exprimée en pourcentage par rapport à l'intervalle de temps considéré. La valeur affective, par définition, n'a pas d'échelle précise, elle dépend seulement du nombre d'accords joués pendant l'intervalle considéré.

Pour obtenir les descripteurs des morceaux musicaux[5], nous avons développé un programme C++ qui analyse le flux d'événements MIDI d'une séquence musicale et calcule, pour chaque intervalle successif, la valeur des descripteurs que nous venons de définir.

3.2.2. La sortie : la représentation de l'émotion véhiculée dans la musique

La représentation de l'émotion utilisée dans ce projet est l'espace 2D Valence – Activation, comme proposé dans (Leman, Vermeulen, De Voogdt, Taelman, Moelants, & Lesaffre, 2004) et (Zentner, Grandjean, & Scherer, 2008). Notre système d'extraction sera donc en charge de déduire ces deux valeurs de Valence et d'Activation à partir des valeurs des descripteurs, elles-mêmes calculées par l'analyse du flux d'événements MID. Au vu des travaux listés dans la section précédente, la représentation en Valence – Activation est fréquemment utilisée pour représenter l'émotion véhiculée dans la musique, surtout pour les travaux sur l'extraction du contenu émotionnel dans la musique en temps réel. Pourtant, ce n'est pas la seule raison qui motive notre choix. Comme présenté dans le chapitre précédent sur notre modèle des émotions GRACE, la représentation Valence – Activation a été choisie comme la représentation universelle des informations échangées entre les composants du modèle. L'extraction du contenu émotionnel d'une séquence musicale étant à la charge du composant *Interprétation Cognitive* du modèle GRACE, qui est censé fournir en sortie un couple de valeurs (Valence, Activation), le choix de cette représentation comme résultat de l'extraction s'est donc imposé de manière assez naturelle.

L'interprétation émotionnelle d'un morceau musical dépend de plusieurs critères. (Zentner, Grandjean, & Scherer, 2008) décrit une expérimentation sur la classification des compositeurs de musique classique en fonction de l'émotion fréquemment véhiculée dans leur musique. Le résultat montre qu'avec différents modèles des émotions (dont les émotions basiques, les émotions dimensionnelles, les émotions

---

[5] Les morceaux musicaux sont en codage MIDI. Chaque morceau en codage MIDI se compose des messages MIDI. Un message MIDI contient notamment l'information sur les notes, l'intensité, le volume, etc. par exemple : *note-on* (début de note), *note-off* (fin de note), volume, *pitch-bend* (modulation de la hauteur de la note) et des signaux de modulation codés avec un identificateur de canal (il peut y en avoir jusqu'à 16). Dans le contexte de notre recherche, on s'intéresse aux messages des notes (message *note-on* et *note-off*). Ces messages MIDI donnent la possibilité de calculer beaucoup de descripteurs musicaux présentés dans la figure 1.

musicales), on obtient différentes classifications. Ceci montre que la façon dont on interprète un style musical varie non seulement en fonction des compositeurs mais aussi en fonction du modèle des émotions utilisé.

Pour construire un système d'extraction de l'émotion en musique, il nous faut donc des styles qui se ressemblent et des interprétations suffisamment cohérentes. Nous avons donc décidé de travailler sur les extraits musicaux de Robert Schumann et d'utiliser l'annotation d'un seul expert, étudiant de Master en musicologie (Roméo Agid).

Pour obtenir l'annotation sur l'ensemble des morceaux choisis, nous avons mis en œuvre une interface NetLogo pour faciliter l'annotation de l'expert.

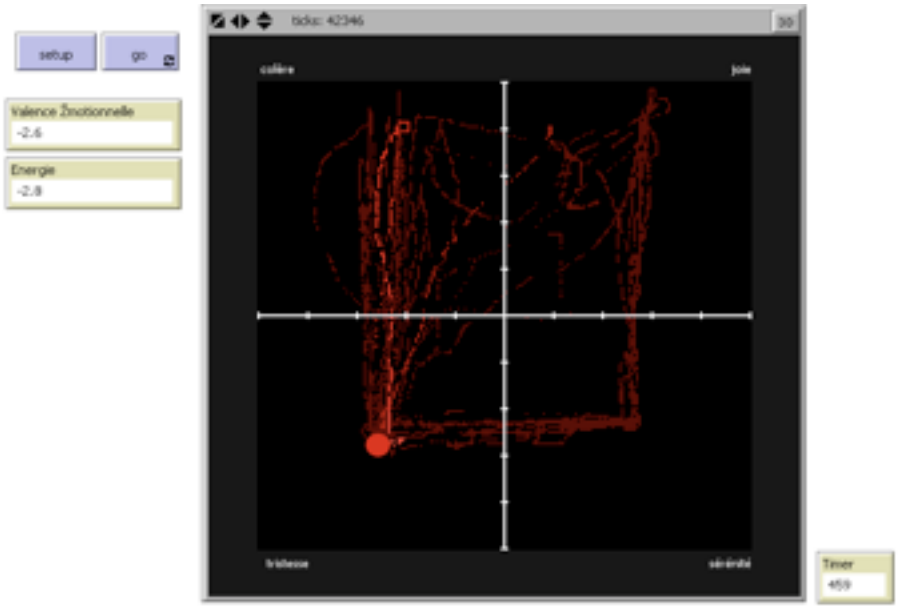


Figure 30 Interface NetLogo pour annoter les morceaux musicaux

Au cours de l'écoute musicale, l'expert place la souris sur la partie noire de l'interface, dont les coordonnées correspondent à un repère orthonormé (Valence, Activation), pour indiquer l'émotion actuellement exprimée par la musique et son intensité. La position de la souris est enregistrée toutes les dixièmes secondes, ce qui permet d'avoir une très bonne précision pour l'analyse ultérieure. L'émotion annotée est donc représentée par deux valeurs Valence et Activation comprises dans l'intervalle [0..10] mesurées tous les dixièmes de secondes.

Nous avons travaillé sur un ensemble de 21 morceaux de musique de Schumann annotés par l'expert (la liste complète est présentée dans l'annexe 3), ce qui représente un total de 3587 secondes de musique.

3.2.3. Structure du système et critères d'évaluation

Avec la base d'annotation humaine obtenue, nous avons décidé d'utiliser la technique de l'apprentissage automatique pour construire notre système d'extraction automatique. Afin de permettre au système d'anticiper la valence et l'activation de la musique, nous construisons deux réseaux de neurones, un pour chaque dimension émotionnelle, pour apprendre à reproduire l'annotation humaine (Figure 31).

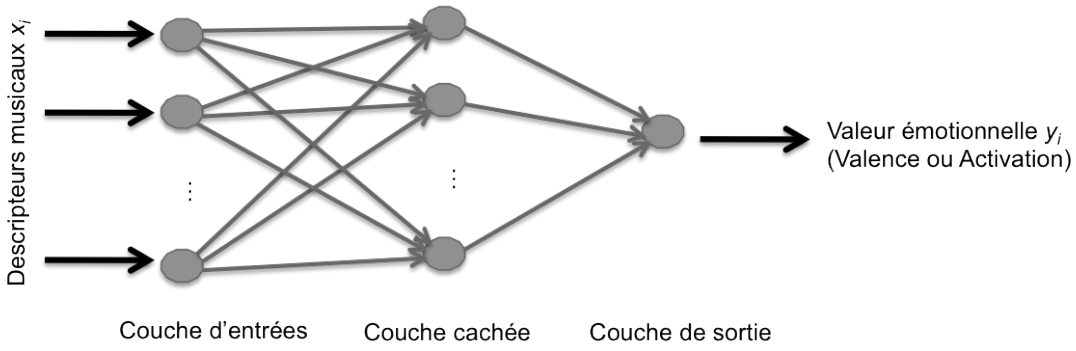


Figure 31 Structure du réseau de neurones pour l'extraction du contenu émotionnel dans la musique

Techniquement, notre réseau de neurones dispose en entrée de 8 valeurs acoustiques et 1 biais. La sortie du réseau de neurones est une valeur (pour la valence ou l'activation). Quand on ajoute la dérivée des descripteurs, la dimension des entrées du réseau atteint 17 (16 descripteurs et 1 biais) ; la dimension de sortie reste de un. Notre réseau de neurones dispose d'une couche cachée. La taille de cette couche est de 10 neurones quand la dimension des entrées est de 9, et cette taille est de 17 quand la dimension des entrées est de 17. La fonction de transition utilisée dans chaque neurone est une sigmoïde.

Pour chaque système, nous employons un apprentissage supervisé. Les entrées sont les descripteurs musicaux présentés dans la section précédente, la sortie associée est la valence ou l'activation annotée par le musicologue. Ces valeurs de valence ou d'activation sont ramenées entre 0 et 1. Les entrées sont également normalisées avant d'être placées en entrée du réseau de neurones. La formule de normalisation est la suivante:

$$d_{tinorm} = \frac{d_{ti} - m_i}{ecart - type_{di}} \qquad (1)$$

avec d_{ti} la $t^{ième}$ valeur du descripteur i, d_{tinorm} est la valeur normalisée de la $t^{ième}$ valeur du descripteur i, m_i est la valeur moyenne du descripteur i, ecart-$type_{di}$ est l'écart type du descripteur i.

L'erreur d'apprentissage est l'erreur moyenne au carré, comme exprimée dans la formule suivante :

$$emc = \frac{1}{n}\sum\left(y_i - f\left(x_i\right)\right)^2 \qquad (2)$$

avec y_i la $i^{ième}$ valeur de sortie désirée (valence ou activation), $f(x_i)$ est la sortie du réseau de neurones pour la $i^{ième}$ entrée (i.e. l'ensemble des descripteurs musicaux associés à la $i^{ième}$ sortie), n est le nombre d'individus utilisés pour l'apprentissage. L'apprentissage vise à minimiser cette erreur *emc* pour reproduire une annotation la plus proche possible de celle fournie par l'expert.

Pour estimer la performance du système, nous calculons le coefficient de détermination au carré R^2 qui a été utilisé par les travaux antérieurs (Fornari & Eerola, 2009 ; Korhonen & Clausi, 2006 ; Oliveira & Cardoso, 2008). Ce coefficient exprime la performance du système, et doit être proche de 1 si le système anticipe

exactement la sortie désirée. Le calcul de ce coefficient se base sur la formule suivante :

$$R^2 = 1 - \frac{emc}{\frac{1}{n}\sum y_i^{\,2}} \qquad (3)$$

avec *emc* est l'erreur moyenne au carré calculée à partir de la formule (2).

3.3. Performance et validité du système

3.3.1. Validité des descripteurs choisis

3.3.1.1. En fonction des performances du réseau de neurones

Afin d'estimer la capacité du système à reproduire l'annotation émotionnelle, nous avons testé plusieurs scénarios possibles, c'est-à-dire utilisé différents types de données en entrée pour le système. Comme nous l'avons abordé dans la section sur les descripteurs musicaux disponibles, nous disposons de 4 types d'entrée différents : descripteurs (en unités classiques), descripteurs (en unités classiques) avec les dérivées premières, descripteurs en unités musicales, descripteurs en unités musicales avec les dérivées premières. Nous évaluons notre système en utilisant la validation croisée. Les 21 morceaux ont été divisés en quatre groupes pour des raisons de performance de l'ordinateur utilisé.

Les valeurs de R^2 du réseau de neurones sur la valence et l'activation sont présentées dans les tableaux 4 (pour la valence) et 5 (pour l'activation) et aussi dans les figures 32 (pour la valence) et 33 (pour l'activation).

Table 4 R^2 du réseau de neurones pour la valence

Groupe	NotTempo_ NotDerive	Not_Tempo_ Derive	Tempo_N otDerive	Tempo_Der ive	Moyenne
G1	0.8372	0.8106	0.7832	0.7707	0.8004
G2	0.3386	0.3351	0.3222	0.3114	0.3268
G3	0.7315	0.7471	0.7491	0.7645	0.7480
G4	0.4276	0.4856	0.5030	0.5822	0.4996
Moyenne	0.5837	0.5946	0.5894	0.6072	**0.5937**

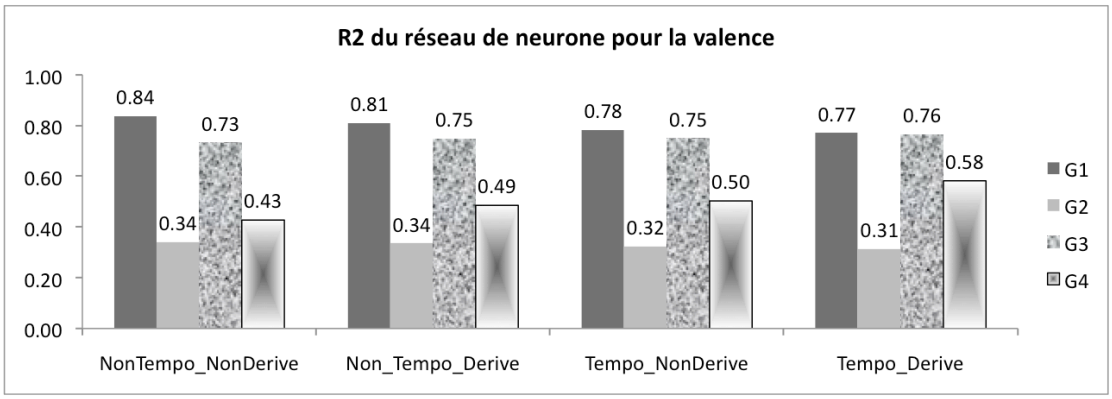


Figure 32 R^2 du réseau de neurones sur les 4 groupes de morceaux pour la valence

Table 5 R^2 du réseau de neurones pour l'activation

Groupe	NotTempo_ NotDerive	Not_Tempo _Derive	Tempo_N otDerive	Tempo_Derive	Moyenne
G1	0.6854	0.6719	0.6914	0.6154	0.6660
G2	0.7623	0.7824	0.7186	0.7850	0.7621
G3	0.1073	0.5608	0.2381	0.3921	0.3246
G4	0.7359	0.8067	0.7377	0.7373	0.7544
Moyenne	0.5727	0.7055	0.5965	0.6325	**0.6268**

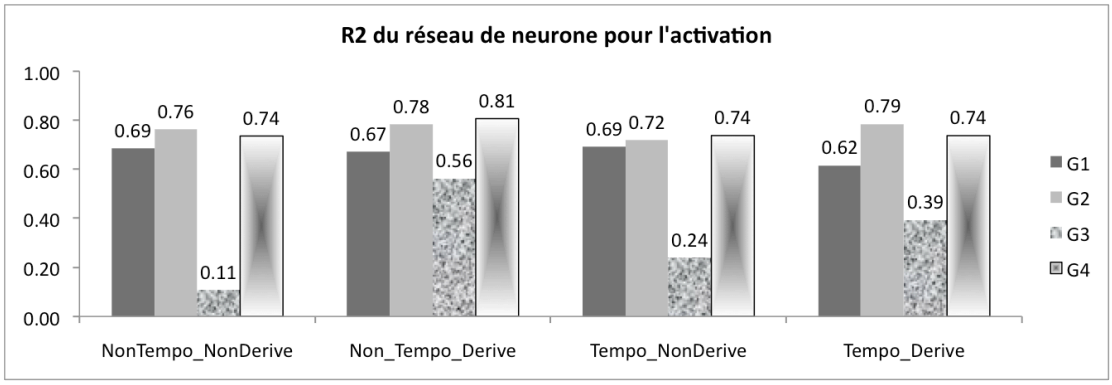


Figure 33 R^2 du réseau de neurones pour l'activation

En comparaison avec les travaux antérieurs, notre système propose une meilleure simulation pour la valence. Tandis que le système de (Korhonen & Clausi, 2006) pouvait simuler 21.9% de la valence, et celui de (Fornari & Eerola, 2009) 42%, notre proposition atteint une hauteur moyenne de plus de 50%. Ceci signifie que les descripteurs que nous avons proposés sont bons pour l'extraction de la valence dans la musique.

Pour ce qui concerne l'activation, le réseau de neurones semble mieux la simuler que la valence. On constate aussi le rôle de la dérivée première à l'amélioration du R^2 du réseau de neurones, aussi bien pour la valence que pour l'activation. Le passage en

unités musicales par contre n'améliore pas la performance du système d'extraction, ni pour la valence, ni pour l'activation.

Le résultat du réseau de neurones pour les 4 groupes suggère que les six indicateurs que nous avons proposés sont cohérents. La valeur de R^2 pour les quatre types de données est supérieure à 0.5 (voir tableaux 4 et 5), ce qui montre que notre système pourrait simuler plus de 50% d'information sur les valeurs émotionnelles dans la musique.

A travers la performance du réseau de neurones sur les 4 groupes, on constate que les morceaux du groupe 1 sont les mieux appris. Ce sont des morceaux courts (dont la durée n'excède pas 2 minutes). L'apprentissage de la valence rencontre des difficultés pour les groupes 2 et 4, et l'apprentissage de l'activation a des difficultés avec le groupe 3. De manière générale, les morceaux de durée supérieure à 2 minutes sont plus difficiles à apprendre. En fait, les caractéristiques de ces morceaux sont plus complexes que les morceaux courts. Ceci montre que bien que les six descripteurs utilisés permettent au réseau de neurones de surclasser la performance des modèles existants, il semble qu'il faille tenir compte d'autres descripteurs pour perfectionner l'extracteur du contenu émotionnel dans la musique.

3.3.2. Réseau de neurones ou arbre de décision

Nous avons dans ce travail testé la performance de notre système en utilisant deux techniques statistiques différentes : le réseau de neurones et l'arbre de décision. Cette action a pour but de trouver la meilleure technique pour construire notre système. Le travail a été réalisé dans le cadre du TER (Travail d'Etude et de Recherche) d'Adel Mezine, étudiant en M1 à l'université d'Evry Val d'Essonne.

Les morceaux utilisés dans ce travail sont les morceaux qui sont annotés plusieurs fois par le musicologue. Ce choix sert à s'assurer de la cohérence dans l'annotation du musicologue et donc de s'assurer de la cohérence de la base de données pour le système d'apprentissage. L'annotation moyenne est utilisée.

Ces morceaux ont été divisés en deux groupes : le premier constitue la base d'apprentissage, le second la base de test.

Base d'apprentissage (1134sec)	Base de test (748sec)
1. Album pour la jeunesse F	1. Album pour la jeunesse (I, L, G)
2. Arabesque	2. Kreisleriana
3. Dans la nuit	3. Mondnacht
4. Toccata	4. Papillons
	5. Petite fugue

Etant donnée la difficulté lors de l'extraction de la Valence (et parfois de l'Activation) présentée dans la section précédente, nous allons tout d'abord comparer la performance des deux techniques sur la classification des émotions en musique avant de comparer leurs performances pour l'extraction de l'émotion en musique. Une autre raison de développer les deux types de systèmes (classification et extraction) est de

voir s'il est possible de combiner les types pour avoir une meilleure performance en extraction du contenu émotionnel.

3.3.2.1. La classification des émotions en musique

La classification sur la valence a été faite avec le réseau de neurones et l'arbre de décision. Les classes de valence à déterminer sont « négative » pour les valeurs de 0 à 5, et « positive » pour les valeurs de 5 à 10 (en sachant que la valeur de valence varie entre 0 et 10). En comparant la performance des deux méthodes, le réseau de neurones semble mieux classifier la valence. Les matrices de confusion pour la classification de la valence en utilisant le réseau de neurones sont les suivantes :

Matrice de confusion de la classification en apprentissage :

Valence	**Négative classée par le réseau de neurones**	**Positive classée par le réseau de neurones**
Négative selon l'expert	360	85
Positive selon l'expert	80	613

Matrice de confusion de la classification sur l'ensemble de validation

Valence	**Négative classée par le système**	**Positive classée par le système**
Négative selon l'expert	288	84
Positive selon l'expert	70	306

La classification sur l'activation a été faite avec le réseau de neurones et l'arbre de décision. Les classes d'activation à déterminer sont « faible» pour les valeurs de 0 à 5, et « forte» pour les valeurs de 5 à 10 (en sachant que la valeur d'activation varie entre 0 et 10). En comparant la performance de deux méthodes, le réseau de neurones semble mieux. Les matrices de confusion pour la classification de l'activation en utilisant le réseau de neurones sont suivantes :

Matrice de confusion de la classification en apprentissage :

Activation	**Faible classée par le réseau de neurones**	**Forte classée par le réseau de neurones**
Faible selon l'expert	989	78
Forte selon l'expert	151	1049

Matrice de confusion de la classification sur l'ensemble de validation

Activation	**Faible classée par le réseau de neurones**	**Forte classée par le réseau de neurones**
Faible selon l'expert	419	105
Forte selon l'expert	3	322

De manière générale, on voit que la classification se passe mieux pour la valence que pour l'activation.

3.3.2.2. Extraction de valeur émotionnelle dans la musique

L'extraction de la valeur émotionnelle est aussi utilisée pour comparer la performance du réseau de neurones et l'arbre de décision. Elle sert à s'approcher le mieux possible de l'annotation humaine. En comparant le résultat obtenu, le réseau de neurones semble là encore plus performant pour la tâche.

Pour l'extraction de la valeur émotionnelle, plusieurs façons de traiter les données musicales ont été proposées. Dans notre travail, nous avons testé avec la normalisation, la technique centrer-réduire, la projection ACP, et la combinaison de plusieurs intervalles de temps. La combinaison des intervalles est due au fait que la perception de la musique varie dans le temps et que l'influence de la musique passée sur la perception de la musique est ambiguë. La prise en compte de plusieurs intervalles pourrait aider à déterminer s'il existe cette relation entre la musique du moment précédent et la musique du moment présent sur la perception humaine. Les trois autres techniques (la normalisation, la technique centrer-réduire, et la projection en ACP) sont destinées à s'assurer que l'échelle des valeurs dans la base de données est valide pour alimenter le réseau de neurones.

Le R^2 de régression pour l'activation est présenté dans le tableau suivant :

Table 6 Tableau de régression pour l'activation en utilisant le réseau de neurones

Coefficients de déterminant Apprentissage / Validation							
Intervalle d'observation		Données normalisées		Données centrées et réduites		Cinq composantes principales	
1s	1	0.91	0.88	0.92	0.89	0.91	0.89
2s	2s	0.93	0.89	<u>0.94</u>	<u>0.89</u>	0.93	0.89
	1s+1s	0.92	0.89	0.93	0.89	0.93	0.89
	3s	0.89	0.88	0.94	0.89	0.93	0.89
3s	2s+1s	0.92	0.89	0.93	0.89	0.93	0.89
	1s+1s+1s	0.89	0.88	0.91	0.88	0.93	0.89

Les résultats sont quasiment les mêmes partout. Cependant on a des résultats aussi bons avec 5 données d'entrées pour l'ACP qu'avec 27 entrées (3x1seconde). Donc au vu de ces résultats, réaliser une ACP peut être intéressant, ne serait-ce que pour augmenter la vitesse d'apprentissage. Le résultat de régression de l'ACP sur les données est présenté dans les 4 figures suivantes.

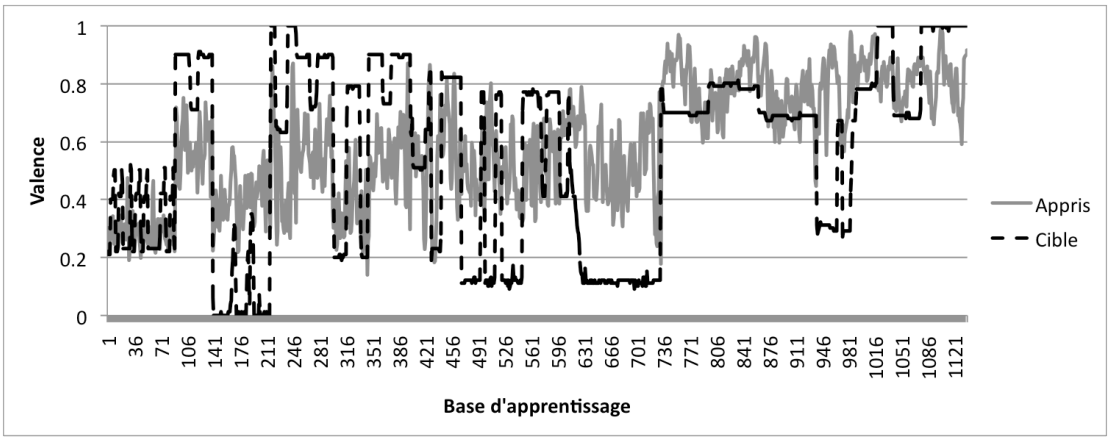


Figure 34 Performance de l'extracteur sur la base d'apprentissage pour la valence (R^2 = 0.80)

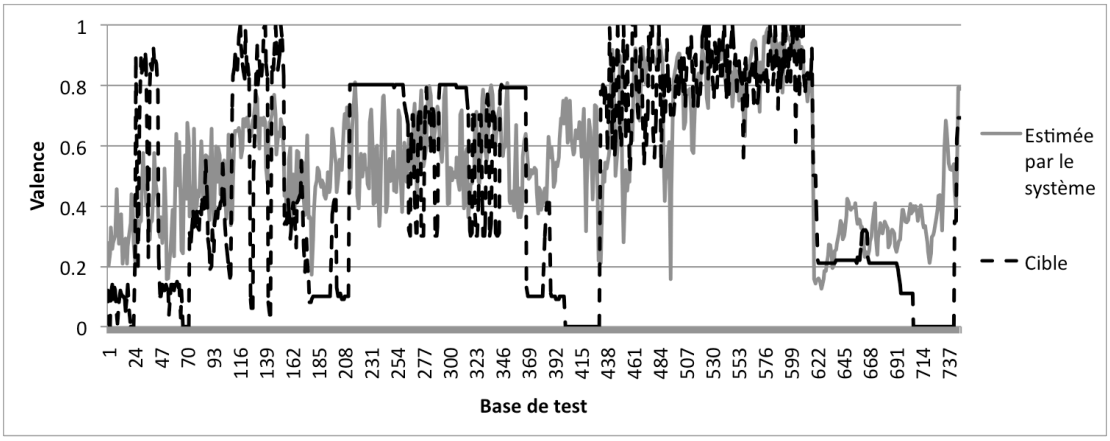


Figure 35 Performance de l'extracteur sur la base de test pour la valence (R^2 = 0.77)

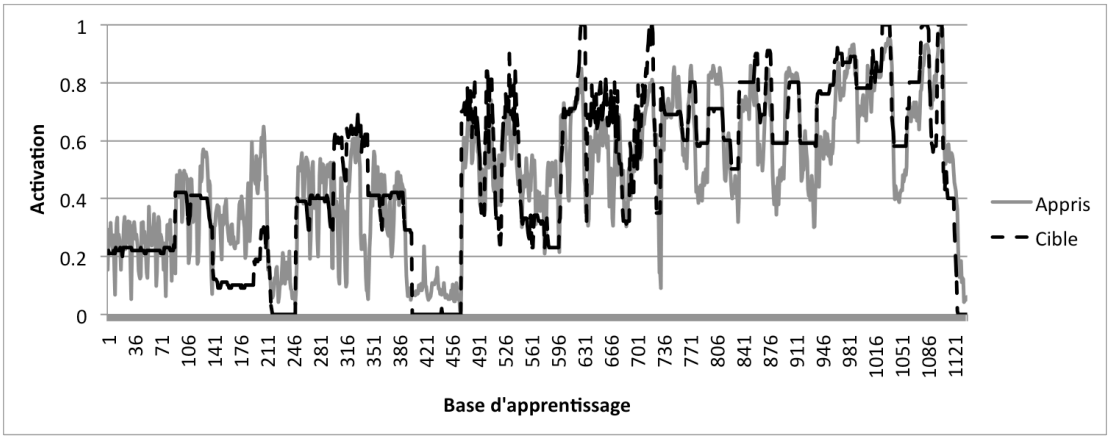


Figure 36 Performance de l'extracteur sur la base d'apprentissage sur l'activation (R^2 = 0.93)

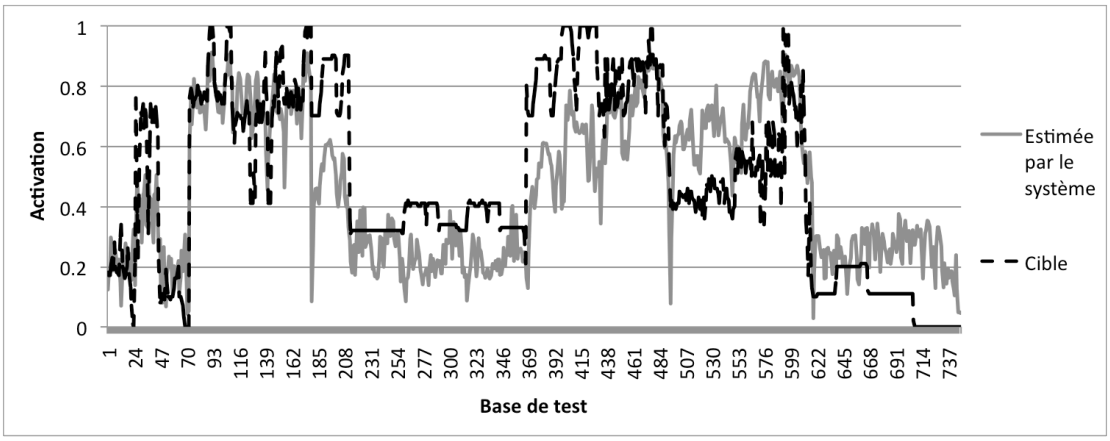


Figure 37 Performance de l'extracteur pour la base de test pour l'activation (R^2 = 0.88)

En comparant les résultats d'apprentissage et de test entre la valence et l'activation, il est clair que la régression sur l'activation fonctionne mieux que celle pour la valence.

4. Discussion

Avec les six descripteurs que nous avons choisis (nombre des notes, nombre d'impulsions, valeur affective, hauteur, intensité, durée), nous arrivons à simuler plus de 50% de la valence d'une base de musique importante. Ceci nous permet de dire qu'il existe une forte relation entre les descripteurs de bas niveau et les descripteurs de haut niveau dans la musique.

Du point de vue théorique, notre liste des descripteurs reste cohérente avec la littérature. En effet, parmi les travaux listés qui s'intéressent à l'extraction de valeurs émotionnelles en musique seconde par seconde (ceux de (Korhonen & Clausi, 2006) et de (Fornari & Eerola, 2009)), les descripteurs sont bien explicités et sont voisins de ceux que nous avons sélectionnés.

Cependant, un nombre réduit de ces descripteurs pourrait être établi tout en conservant la majorité des informations musicales considérées. A notre avis, les 16 descripteurs proposés par Korhonen peuvent être représentés via seulement nos 8 descripteurs comme présenté dans le tableau qui suit : fréquence moyenne, écart-type de la fréquence, volume moyen, écart-type du volume, valeur affective, nombre de notes, nombre d'impulsions, durée des notes.

Table 7 Une mise en correspondance entre les 16 descripteurs proposés par Korhonen et par notre travail

No	Propriété musical	Descripteurs proposés par Korhonen	Descripteurs proposés dans notre travail
1	Dynamique	Niveau du volume	Volume moyen Nombre des notes
		Volume maximal à court terme	Ecart-type du volume
2	Fréquence	Centroïd du spectre de puissance	Fréquence moyenne

	moyenne	Moyenne du centroïde STFT	
3	Variation de la fréquence	Moyen du flux STFT Ecart-type du flux STFT Ecart-type du centroïde STFT	Ecart-type de la fréquence
4	Timbre	Largeur du timbre Netteté Moyenne du transfert STFT Ecart-type du transfert STFT	
5	Harmonie	Dissonance spectrale (Hutchinson et Knopoff) Dissonance tonale (Hutchinson et Knopoff) Dissonance spectrale (Sethares) Dissonance tonale (Sethares) Complexité du ton	Valeur affective Nombre d'impulsions Durée moyenne des notes
6	Tempo	Battement par munite	
7	Texture	Multiplicité	Nombre des notes

Le volume moyen et l'écart-type du volume sont conservés pour la dynamique. De plus, le nombre des notes et la durée moyenne des notes jouées affectent aussi la dynamique de la musique. Donc, nous l'ajoutons dans notre liste des descripteurs. La fréquence moyenne et l'écart-type de la fréquence dans notre travail sont calculés directement depuis les événements MIDI, ces deux valeurs correspondent aux 5 descripteurs proposés par Korhonen. Notre travail s'intéresse à des extraits musicaux sur un seul instrument (i.e. le piano), donc il n'y a pas de descripteur pour le timbre. La texture dans la liste de Korhonen est en fait le nombre de notes jouées dans un son donc nous proposons de le conserver. Pour le tempo, nous avons testé notre extracteur avec l'unité musicale (expliqué dans la section 9.2.1). Pour évaluer l'harmonie d'un extrait, (Husain, Thompson, & Schellenberg, 2001) propose de combiner le mode et le tempo. La valeur du mode est appelée la valeur affective dans notre travail. La valeur affective est calculée à partir des accords joués. Dans notre projet, chaque accord correspond à une valeur affective (voir le tableau des valeurs affectives de tous les accords dans l'Annexe). A un instant de musique donné, sa valeur affective est la moyenne des valeurs affectives des accords joués à cet instant. Pour calculer la valeur affective d'un intervalle, les valeurs affectives des accords joués sont pondérées par leurs durées de présence[6] respectives (i.e. de ces accords) avant le calcul de la moyenne.

[6] La durée de présence d'un accord est la durée dans laquelle l'accord est joué.

La contribution principale de la thèse du point de vue de l'analyse émotionnelle de séquences musicales consiste en la proposition d'un nombre réduit des descripteurs qui permet de bien extraire le contenu émotionnel dans la musique grâce au réseau de neurones. La validité des descripteurs proposés est aussi renforcée car nous travaillons sur une base de données plus importante que les bases comparables dans la littérature.

Plusieurs pistes restent à exploiter sur cette problématique. Par exemple, sur la performance de l'extraction de valeurs émotionnelles, il serait intéressant de voir la performance du réseau de neurones sur tous les morceaux annotés qui sont utilisés dans la validation des descripteurs choisis (section 9.3.1). D'ailleurs, il serait aussi intéressant de comparer la performance du réseau de neurones et de l'arbre de décision sur tous les morceaux annotés. Cette comparaison pourrait donner des indices pour une meilleure extraction.

Une autre piste consiste à améliorer la performance du système d'extraction. Bien que la performance de notre système soit bonne en comparaison avec les travaux existants, on peut encore espérer des gains de performance. De plus, il semble aussi que l'extracteur fonctionne moins bien sur les morceaux longs (plus de 2 minutes de longueur). Plusieurs solutions sont possibles. Par exemple, on peut ajouter d'autres descripteurs (par exemple ceux proposés dans la littérature) pour le traitement de la valence, ou analyser les morceaux musicaux séquence par séquence. D'ailleurs, peut-être que l'émotion véhiculée dans les morceaux musicaux déjà écoutés affecte la perception des gens. Si c'est le cas, il faut prendre en compte les valeurs de Valence et Activation des minutes (ou secondes) précédentes dans l'extraction de la valeur émotionnelle à l'instant courant. Une autre piste possible est de refaire l'annotation des morceaux. En effet, l'annotation utilisée dans ce projet a été faite par une seule personne, et la plupart des morceaux ne sont annotés qu'une seule fois. Comme l'humeur de cette personne n'était pas clairement contrôlé lors de l'annotation, il est possible que la base d'annotation que l'on a utilisée ne soit pas complètement cohérente en elle-même. Une phase de ré-annotation est ainsi à envisager pour l'utilisation future de la base de données.

Chapitre 4

Expression émotionnelle d'un robot en réaction à la musique

1. Introduction

La recherche sur les robots compagnons ou en robotique sociale investigue depuis longtemps les caractéristiques anthropomorphiques pour améliorer l'interaction homme-robot, comme l'apparence anthropomorphique (robot humanoïde, tête robotique avec jeux, bouches, oreilles, etc.), ou les capacités de communication (par exemple : voir, parler, écouter). Au cours du développement de l'interaction homme-robot, l'émotion est apparue comme un élément essentiel à traiter pour que les robots gagnent plus de confiance de leurs partenaires humains et améliorent l'efficacité de l'interaction. En effet, l'aspect émotionnel dans l'interaction avec l'humain a été confirmé par les psychologues comme un des éléments importants pour faciliter la communication et renforcer l'acceptabilité (Mayer, Salovey, & Caruso, 2008) (Scherer, 2009). Les expérimentations robotiques ont montré par ailleurs que la capacité émotionnelle (comme l'expression émotionnelle, la perception de l'émotion, la personnalité) contribue à l'acceptabilité des robots par l'humain et par conséquent rend l'assistance robotisée plus efficace (Forlizzi, Disalvo, & Gemperle, 2004) (Tapus, Tapus, & Matarié, 2008).

La conception des réactions émotionnelles est donc très bien connue dans la recherche technologique sur l'interaction homme – machine (qui traite non seulement de l'interaction homme – robot mais aussi de l'interaction homme – agent virtuel). Dans la plupart des cas, ce sont des développements de l'expression faciale qui sont étudiés. Les robots Kismet ou iCat, l'agent Greta, etc. sont tous équipés d'une capacité d'expression faciale de l'émotion, s'inspirant du système de codage des actions faciales (Facial Action Coding System) proposé par Paul Ekman et W.V. Friesen depuis les années 1970 (Ekman, Friesen, & Hager, 2002). L'expression émotionnelle via l'action du corps, par contre, reçoit une attention limitée (Fong, Nourbakhsh, & Dautenhahn, 2003).

Pour but de développer une interaction active entre un musicien et un robot (ou bien un agent virtuel pour une possibilité d'utilisation plus large), nous nous intéressons donc à des façons efficaces pour exprimer les émotions du robot envers le musicien. Selon les résultats de recherche d'A. Camurri sur l'écoute active à la musique, il est claire que l'auditeur s'exprime ses émotions lors de l'écoute musicale via ses gestes ((voir le chapitre « Musique et Emotions » (Glowinski & Camurri, 2010) du livre « Système d'interaction émotionnelle » de C. Pelachaud (Pelachaud, 2010) pour une vue globale sur les résultats et le domaine de recherche concerné). Le robot, dans notre scénario d'étude, joue le rôle de l'auditeur lors de l'interaction musicien - robot. Il est donc intéressant d'étudier la possibilité d'exprimer les émotions du robot via ses mouvements corporels pendant l'interaction.

Dans le cadre de notre projet sur l'implémentation d'une expression émotionnelle robotique dans un contexte musical, nous cherchons à développer des codes de

mouvements permettant au robot d'exprimer de l'émotion via ses mouvements. Après une étude bibliographique, nous détaillons notre conception et l'expérimentation que nous menons pour valider notre concept.

Nous commençons le chapitre par une revue de littérature sur la conception des expressions émotionnelles pour des robots personnels ou des agents virtuels animés. Les travaux abordés seront les comportements émotionnels simulant un enfant du robot Kismet, les comportements sociaux du robot iCat pour interagir avec les personnes âgées, les expressions émotionnelles de l'agent virtuel animé Greta pour traduire l'empathie lors d'une conversation avec un adolescent, et les mouvements émotionnels du regard dans le cadre de GWT (Gaze Wrapping Transformation) pour intégrer dans des agents virtuels.

Ensuite, nous présenterons les travaux sur l'évaluation des mouvements émotionnels qui nous aident soit à concevoir les mouvements émotionnels de notre robot soit à valider notre conception. Ce sont l'évaluation de l'expression faciale du robot EDDIE, l'évaluation des mouvements du corps des musiciens pour interpréter le contenu émotionnel dans la musique jouée, et l'évaluation des mouvements du robot MEX pour interpréter musicalement des émotions.

Après avoir achevé la présentation des travaux antérieurs, nous passerons à la conception des mouvements émotionnels de notre robot qui s'appelle LINA. Comme LINA est un robot mobile ayant une capacité limitée d'expression (il n'a pas de bras robotisé, pas de visage robotisé), l'expression émotionnelle de LINA se limite donc a des déplacements dans l'espace et à la simulation du regard via sa caméra.

Les mouvements de LINA ont été validés via une expérimentation que nous avons menée en Octobre 2010 lors de la Fête de la Science organisée par l'Université d'Evry val d'Essonne. Cette expérimentation nous a permis de tester, avec un grand nombre de participants, notre conception sur le robot réel. Nous présenterons les détails de l'expérimentation dans la section 5. Nous développerons ensuite une discussion sur les résultats d'expérimentation et sur notre conception actuelle selon les conclusions tirées des résultats de l'expérimentation.

2. Expression émotionnelle dans l'interaction homme – machine

La recherche de mise en œuvre des capacités d'expression des émotions chez des robots personnels ou des agents virtuels animés vient du fait que l'interaction avec un tel agent (i.e. un robot personnel émotif, un agent virtuel émotif) rend l'interaction plus naturelle et plus efficace. Nous allons voir dans cette section quatre travaux portant sur l'ajout des capacités d'expression des émotions : deux pour les robots (Kismet et iCat) et deux pour les agents virtuels animés (Greta et GWT).

2.1. Kismet avec l'expression faciale d'un enfant

Dans le but de développer un système robotique capable d'interagir naturellement avec l'humain, C. Breazeal (Breazeal, 2001) a mis en œuvre une tête robotique s'appelant Kismet qui est capable d'exprimer des émotions à travers ses expressions faciales. La réponse émotionnelle de Kismet est caractérisée par les éléments suivants :

- un événement qui sollicite une réaction émotionnelle
- une analyse affective de l'événement
- une expression caractéristique (visage, voix, geste)
- une tendance d'action qui motivera une réponse comportementale.

Pour faire fonctionner Kismet, les chercheurs ont implémenté un système d'*émotion* qui simule les réactions d'un enfant. A partir d'un événement capturé, une analyse affective est faite pour déterminer quelle réaction émotionnelle lui associer. Cette réaction émotionnelle se décompose en deux phases : l'expression émotionnelle et le plan d'action associé. Chaque phase correspond à un seuil prédéfini. L'expression émotionnelle a un seuil plus bas que le plan d'action. L'idée de deux seuils est que l'expression émotionnelle entraîne une réponse comportementale appropriée. Ceci permet à l'humain d'interpréter et de comprendre le comportement du robot. L'expression émotionnelle est l'expression faciale de Kismet. Le plan d'action est prédéfini en fonction de l'émotion. Par exemple, quand un événement est considéré comme une menace, la peur est générée. Une expression faciale de la peur est déclenchée et un plan pour s'enfuir est mis en action.

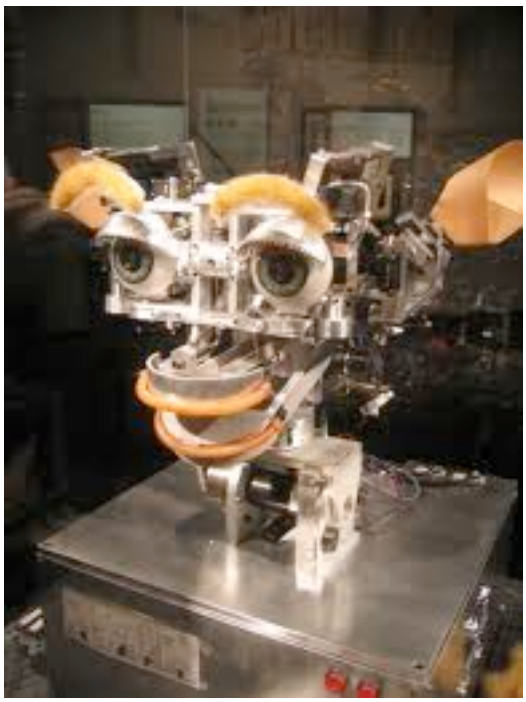


Figure 38 La tête robotique Kismet

Les émotions de Kismet sont représentées sur 3 dimensions : Valence – Activation – Position. Sur ces trois dimensions, les chercheurs ont défini neuf prototypes d'expressions faciales qui couvrent l'espace de cette représentation (Figure 39). Les neuf états affectifs sont la joie, la tristesse, la colère, la peur, la surprise, la fatigue, le dégout, l'air fermé (*stern* en anglais), l'ouverture (*accepting* en anglais).

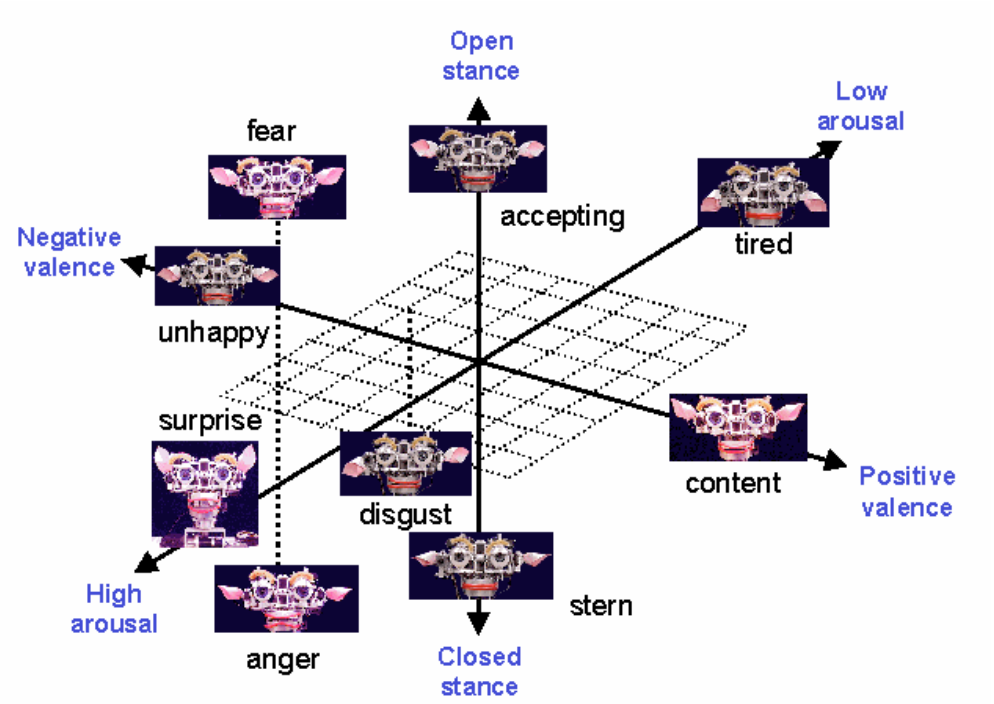


Figure 39 Neuf prototypes d'expressions faciales des émotions de Kismet

La capacité émotionnelle de Kismet a été validée via une expérimentation avec cinq sujets féminins. Aucune parmi les cinq n'avait interagit avec Kismet avant l'expérimentation, c'est-à-dire qu'elles n'avaient aucune idée de la capacité de Kismet. Le scénario d'interaction est qu'une nounou joue avec un enfant. Chaque sujet est sollicité pour faire quatre tâches : amuser Kismet, le mettre en colère, l'alerter, et le calmer.

Les résultats d'observation montrent que les sujets se basent naturellement sur l'expression émotionnelle de Kismet pour déduire quand le robot le « comprend ». Les sujets ont aussi tendance à alterner leurs comportements pour améliorer l'interaction avec le robot (par exemple : changer le volume de la voix, répéter des phrases). Les sujets ont même exprimé de la culpabilité quand elles ont réussi à mettre le robot en colère.

L'expérimentation avec Kismet montre que si le robot est capable de communiquer en respectant les standards sociaux, l'interaction homme – robot pourrait être très agréable et efficace. Pour le robot, de telles capacités lui permettent d'obtenir la réaction désirée de la part de l'humain lors de l'interaction avec lui. De la part de l'humain qui interagit avec d'un tel robot, l'humain n'a pas besoin d'entrainement pour pouvoir communiquer avec le robot et l'interaction avec un tel robot promet de lui apporter des expériences agréables.

2.2. iCat

iCat (Figure 40) est un robot immobile fourni par la compagnie Philips. C'est un robot qui peut simuler les traits de visage, les signes de tête, et qui peut émettre des sons. M. Heerink et al ont fait une expérimentation avec iCat pour évaluer l'acceptabilité, pour des personnes âgées, d'un robot ayant une capacité d'interaction sociale (Heerink, Krose, Evers, & Wielinga, 2006). Elle est définie par les caractéristiques suivantes :

1. coopérer

2. exprimer de l'empathie
3. montrer l'affirmation de soi
4. exposer la maîtrise de soi
5. montrer la responsabilité
6. gagner la confiance d'autrui
7. montrer sa compétence

Figure 40 Robot iCat

Etant donné le fonctionnement de iCat, les chercheurs ont traduit ces caractéristiques sociales en un certain nombre de fonctionnements pour le robot :

- écouter attentivement les interlocuteurs, par exemple en les regardant et leur faisant un signe de tête (1, 2),
- être gentil et aimable, par exemple en souriant pendant l'interaction avec les interlocuteurs (1, 2, 7),
- rappeler des petits détails personnels des interlocuteurs, comme leur nom (6, 7),
- être expressif en utilisant l'expression faciale (2, 3),
- avouer l'erreur (5, 6).

Avec cette conception des comportements d'iCat, Heerink et ses collègues ont mené une expérimentation avec les personnes âgées pour évaluer si les capacités sociales d'iCat contribuent à l'amélioration de l'interaction. Ils ont invité les sujets à interagir avec deux versions d'iCat et à demander de l'information à iCat. Une version d'iCat est capable de se comporter selon les caractéristiques sociales présentées précédemment, et l'autre se comporte tout simplement en donnant les indications

demandées. Les informations demandées à chaque iCat sont : régler une alarme, donner des instructions pour aller au supermarché le plus proche, donner de l'information sur la météo du lendemain. Il y a 17 personnes qui interagissent avec l'iCat social, et 19 personnes qui interagissent avec l'iCat neutre. Chaque interaction entre un sujet et le robot dure 10 minutes.

Le résultat de l'expérimentation montre que l'interaction avec l'iCat social donne du confort aux sujets, et que l'interaction avec l'iCat neutre donne peu du confort et même du désagrément aux sujets. Quant à l'interaction avec l'iCat social, les sujets donnent plus d'expression communicative positive envers le robot, comme faire des signes de tête, s'approcher ou s'écarter du robot, sourire au robot, saluer verbalement. Ce résultat confirme donc le rôle des caractéristiques sociales et émotionnelles du robot pour pouvoir apporter du confort aux utilisateurs et par conséquent augmenter l'efficacité de l'interaction.

2.3. Greta

Ayant pour but de créer un agent virtuel réaliste capable de communiquer en temps réel de façon expressive et crédible, de Rosis et collègues ont développé un agent conversationnel animé qui s'appelle Greta (de Rosis, Pelachaud, Valeria, & De Carolis, 2003). Le projet de développement de Greta consiste en trois grande parties : (1) la représentation mentale de l'agent (*Agent Mind* en anglais) pour gérer les états affectifs/émotionnels de Greta au cours d'une conversation – cette partie a été discutée dans le chapitre sur les modèles des émotions de ce mémoire, (2) un langage de marqueurs pour définir les gestes de communication que Greta doit faire lors de son discours, (3) la transition entre les marqueurs des gestes de communication de Greta à l'expression faciale de l'agent, incluant notamment la direction du regard, la forme des sourcils, les directions de la tête, les signes de tête.

Les gestes de communication de Greta sont définis via un langage de marqueurs qui s'appelle « Affective Presentation Markup Language » (APML), proposé par De Carolis et al en 2002. L'APML est un langage basé sur XML qui se compose de balises décrivant les gestes de communication possible dans une conversation.

L'apparence de Greta (Figure 41) est un agent graphique en 3D programmable qui permet aux chercheurs de mettre en œuvre différentes expressions émotionnelles similaires à celles d'un humain. De Rosis et al. ont aussi suggéré quatre qualités que Greta doit apporter dans ses gestes de communication. Ce sont ses *croyances*, ses *intentions*, son *état affectif (i.e. son état émotionnel)*, et ses *actes de raisonnements*. Par exemple, les *croyances* sont exprimées via les gestes de hausser les sourcils, de faire des signes de tête, ou de regarder l'interlocuteur ; les *intentions* sont exprimées dans la façon dont l'agent dit ses phrases, par exemple suggérer, approuver, demander ; l'*état affectif* est exprimé via l'expression faciale ; les *actes de raisonnement* sont pour simuler l'acte de raisonner/mémoriser/se souvenir de l'humain lors d'une conversation, comme ne pas regarder l'interlocuteur pour se concentrer sur ses propre pensées, regarder vers le bas pour se souvenir de quelque chose dans le passé, ou regarder vers le haut pour raisonner sur les informations données.

Figure 41 Apparence de l'agent Greta

De Rosis et al. ont fait des simulations sur ordinateur pour simuler plusieurs personnalités sur Greta et pour valider la cohérence dans le changement d'état affectif de Greta au cours d'une conversation. Ils ont aussi testé la partie graphique de Greta (i.e. l'apparence 3D de Greta) pour voir la dynamique dans le changement d'état expressif de Greta au cours d'une conversation prédéfinie. Une validation avec les humains n'est pas encore rapportée, mais l'implémentation de Greta montre que l'utilisation des gestes de communication est réaliste et réalisable pour rendre l'interaction homme – machine la plus agréable possible.

2.4. Gestes du regard pour l'expression de l'émotion

Spécialement intéressés par le regard dans l'expression émotionnelle chez l'humain, B. Lance et S. Marsella se posaient la question non seulement des directions du regard mais aussi des caractéristiques des mouvements du regard (Lance & Marsella, 2010). Constatant qu'il y a peu de travaux scientifiques décrivant explicitement comment le changement du regard se passe à chaque expression émotionnelle, ni dans la littérature psychologique, ni dans la littérature informatique, B. Lance et S. Marsella (Lance & Marsella, 2010) cherchent à construire un système d'indices pour la constitution des regards émotionnels pour les agents virtuels animés. Les indices en question sont les caractéristiques des mouvements de l'humain pour constituer le regard à chaque expression émotionnelle. Par exemple, à quelle vitesse faut-il changer la direction du regard pour exprimer la colère ? Dans quelle direction ? Comment la posture du corps humain change-t-elle ? Le comportement émotionnel[7] du regard se décompose en trois comportements partiels : la posture de la tête, la posture du tronc, et la direction du regard.

Pour construire la mise en correspondance entre le comportement du regard et la représentation émotionnelle perçue par l'humain, les chercheurs ont généré 150

[7] Un comportement émotionnel est ici une posture statique de toutes les parties concernées. Un mouvement émotionnel correspond à des changements des parties concernées pour exprimer une émotion en question.

simulations de comportements possibles susceptibles d'apporter l'impression émotionnelle selon la littérature psychologique et sociale. Ensuite, ils ont mené deux expérimentations pour vérifier la pertinence de 150 comportements virtuels générés.

La première expérimentation est pour classifier les émotions interprétées par les 150 simulations. Cette classification permet de voir comment les comportements simulés couvrent l'espace de représentation des émotions choisi pour faire l'analyse sur les caractéristiques des mouvements émotionnels du regard. L'espace de représentation des émotions choisi est un espace à 3 dimensions : Valence – Activation – Dominance (en anglais : *Pleasure – Arousal - Dominance*). 31 participants ont évalué 20 simulations choisies aléatoirement parmi les 150 simulations. Chaque simulation a donc été évaluée en moyenne 4 fois. L'analyse des résultats obtenus montre que les 150 simulations représentent 10 grandes catégories d'émotions (voir Table 8).

Table 8 Les 10 catégories des émotions classifiées auprès la première expérimentation de (Lance & Marsella, 2010)

Catégories d'émotions	
Colère	Mépris
Incrédulité	Excitation
Peur	Séduction (en anglais *flirtatious*)
Culpabilité	Tristesse
Mystère	Surprise

La deuxième expérimentation cherche à découvrir la correspondance entre les caractéristiques des mouvements émotionnels et les 3 dimensions représentant les émotions (Valence – Activation – Dominance). Chaque personne parmi les 100 participants regardait 15 simulations aléatoirement choisies dans les 150 simulations initiales. Chaque personne devait évaluer ces simulations en fonction de 6 questions (voir Table 9) sur l'échelle de 0 (pas du tout) à 5 (entièrement d'accord).

Table 9 Questions d'évaluation utilisées dans la deuxième expérimentation de (Lance & Marsella, 2010)

Dimension émotionnelle	**Question**
Dominance élevée	Le personnage est dominant
Dominance faible	Le personnage est soumis
Activation élevée	Le personnage est agité
Activation faible	Le personnage est détendu
Valence élevée	Le personnage est content
Valence faible	L personnage est mécontent

Leur résultat d'analyse révèle, parmi d'autres conclusions, qu'il existe une corrélation entre les dimensions émotionnelles et les caractéristiques des mouvements émotionnels du regard.

Table 10 Corrélation entre les caractéristiques de mouvements du regard et les dimensions AVD

Catégorie d'émotion	Dimension émotionnelle (VAD)	Corrélation entre les caractéristiques de mouvements du regard et les dimensions AVD		
Mépris	-V-A-D	Tête vers le haut +D	Tronc neutre -V	Vitesse normale -A
Excitation	+V+A+D	Tête neutre +D	Tronc incliné +V	Vitesse élevée +A
Culpabilité	-V+A-D -V-A-D	Tête vers le bas -D	Tronc neutre -V	Vitesse normale Vitesse faible -V
Tristesse	-V-A-D	Tête vers le bas -D	Tronc neutre -V	Vitesse élevée +A Vitesse normale -A

Leur tableau de corrélation entre les dimensions émotionnelles et les gestes émotionnels est intéressant pour la conception des mouvements émotionnels pour les agents virtuels ou les robots personnels. Pourtant, étant donnée l'incohérence sur l'aspect vitesse des mouvements pour les émotions Tristesse et Culpabilité, les auteurs admettent que la vitesse des mouvements émotionnels conçus n'était pas très proprement calibrée, et cela laisse ouverte la question sur la corrélation Activation - Vitesse.

2.5. Conclusion

Les expressions émotionnelles aident à améliorer la qualité d'interaction et aussi offrent un certain niveau de confort à la personne qui interagit avec les machines ayant de telles capacités. Pourtant, ces travaux ne s'intéressent pas à vérifier si l'expression émotionnelle est bien perçue par la personne ou pas. Une bonne reconnaissance des émotions exprimées assurera que l'information communiquée par le robot/l'agent est bien interprétée par l'humain. Et ceci est crucial pour le développement d'une interaction efficace et d'une relation à long-terme entre le robot/l'agent et la personne. Nous allons voir dans la section suivante quelques travaux scientifiques traitant l'évaluation de la perception de l'humain lors des comportements émotionnels : un travail sur l'évaluation des expressions émotionnelles d'une tête robotisée, un travail sur l'évaluation des mouvements

émotionnels lors des performances des musiciens ; et un travail sur l'évaluation des mouvements émotionnels d'un robot mobile avec l'aide de la musique en arrière plan. Ce sont les trois travaux qui nous aident aussi à effectuer la validation de notre conception des expressions émotionnelles pour le robot LINA.

3. Evaluation de l'expressivité des agents émotionnels

3.1. Evaluation de l'expression faciale des robots

L'expression faciale est un des éléments les plus fréquents dans l'implémentation de capacités émotionnelles pour les robots et les agents virtuels. Pour construire ces expressions, les chercheurs se basent en général sur le cadre FACS (Facial Action Coding System) proposé par des psychologues (Ekman, Friesen, & Hager, 2002). Bien que ce cadre soit bien utilisé, évaluer si l'humain perçoit bien l'émotion exprimée par les robots et les agents virtuels semble recevoir très peu d'attention.

Kolja Kuhnlenz et ses collègues à l'Institute of Automatic Control Engineering s'intéressent à établir un cadre d'évaluation pour la conception des émotions faciales des robots (Kuhnlenz, Sosnowski, & Buss, 2007). Selon eux, l'évaluation de l'expression faciale des robots devrait se faire dans un cadre d'évaluation approprié. A l'heure actuelle, la plupart des conceptions d'expression faciale se basent sur le cadre FACS, qui s'intéresse à l'expression des émotions individuelles. L'évaluation de la perception par l'humain de ces expressions est faite en demandant d'indiquer quelle est l'émotion exprimée. Ce genre d'évaluation ne permet pas de déduire comment modifier l'implémentation des actionneurs pour pouvoir améliorer l'expression faciale des robots. Pour relever ce défi, les auteurs ont proposé de se baser sur la représentation dimensionnelle des émotions pour faire l'évaluation. Pour démontrer leur idée, ils ont mené deux expérimentations pour évaluer l'expression faciale d'une tête robotique nommée EDDIE (Figure 42).

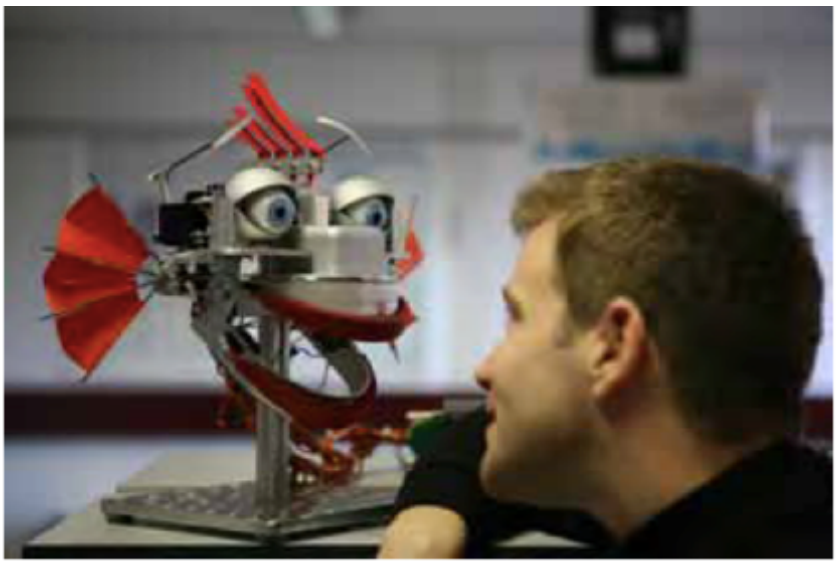


Figure 42 La tête robotique EDDIE (Kuhnlenz, Sosnowski, & Buss, 2007)

EDDIE est conçu pour imiter les expressions faciales de l'humain. Il peut réaliser 13 expressions parmi les 21 expressions émotionnelles définies dans FACS. EDDIE

possède non seulement les éléments d'un visage, mais aussi une crête de plumes de cacatoès et deux oreilles ressemblant à celles des lézards.

Les auteurs ont fait deux expérimentations pour évaluer les expressions faciales en fonction des émotions basiques exprimées d'une part, et pour évaluer les expressions faciales en fonction des dimensions émotionnelles exprimées d'autre part.

Dans la première expérimentation, il y a eu 30 participants (11 femmes, 19 hommes, 30 ans d'age moyen). On présente aux participants 6 expressions faciales d'EDDIE correspondant à sept émotions de base (Joie, Surprise, Inquiétude, Tristesse, Colère, Dégoût et Neutre). En regardant chaque expression, les participants doivent indiquer quelle émotion EDDIE exprime parmi les sept émotions données.

Figure 43 Résultat de reconnaissance des émotions lors de la première expérimentation de (Kuhnlenz, Sosnowski, & Buss, 2007)

Les résultats de reconnaissance des expressions faciales de EDDIE sont présentés dans la Figure 43. Les auteurs concluent que ce résultat montre qu'EDDIE a bien exprimé des émotions mais ne permet pas de déduire comment faire pour améliorer l'expression de EDDIE.

Dans la deuxième expérimentation, Kuhnlenz et al. proposent d'utiliser la représentation dimensionnelle Valence – Activation - Dominance pour évaluer les expressions émotionnelles de EDDIE. Pour cette expérimentation, ils ont 50 participants (15 femmes, 35 hommes, 25 ans d'âge moyen). Chaque participant évalue 30 expressions émotionnelles de EDDIE, ces expressions sont aléatoirement choisies. Chaque expression de EDDIE est évaluée sur 3 dimensions : Valence – Activation – Dominance sur une échelle de 9 points (de -4 à 4).

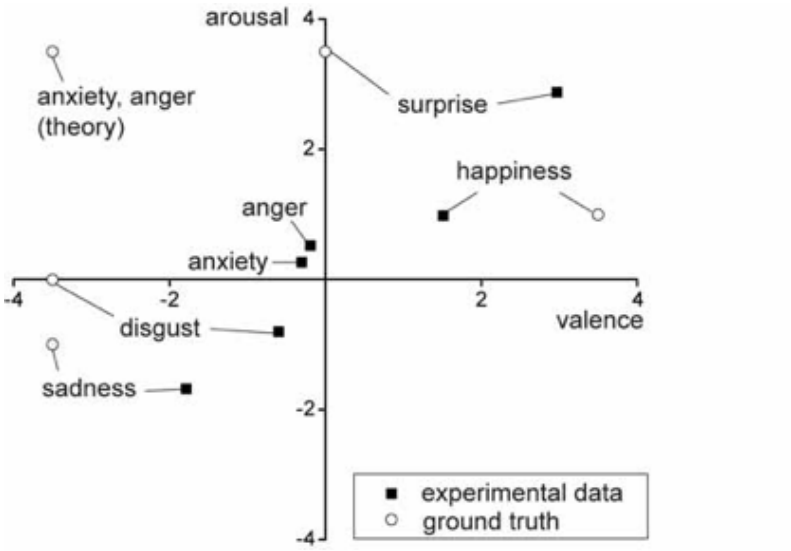


Figure 44 Résultats de reconnaissance obtenus dans la deuxième expérimentation de (Kuhnlenz, Sosnowski, & Buss, 2007), sans la dimension Dominance

Le résultat de reconnaissance de la deuxième expérimentation (Figure 44) montre que l'expression de la Valence était plutôt positive dans la plupart des cas (saut pour la joie). Les auteurs expliquent que ceci est dû au fait que les lèvres de EDDIE à l'état neutre donnent déjà une impression joyeuse, ce qui peut affecter l'efficacité des expressions négatives de EDDIE. De plus, la Colère et l'Inquiétude sont perçues comme l'état neutre, ce qui signifie que l'implémentation de ces deux émotions n'est pas bonne. Selon les auteurs, ceci est dû à la limite technique de quelques actionneurs du robot, ce qui fait que les expressions de la colère et de l'inquiétude ne sont pas bien présentées.

Bien que les auteurs essaient de mettre en valeur l'évaluation via la représentation dimensionnelle des émotions, leurs arguments ne sont pas très convaincants. En fait, à partir du résultat de la première expérimentation, on peut déjà constater que la colère et l'inquiétude sont parfois perçues comme neutre, et que la valence perçue est plutôt positive en moyenne via les taux de reconnaissance de la surprise, la colère, et l'inquiétude. De plus, les deux expérimentations sont faites avec deux groupes différents de participants avec des conditions différentes (la première expérimentation évalue 6 expressions, tandis que la seconde évalue 30 expressions), ceci pose aussi la question de l'égalité des informations qu'ils ont utilisées à l'appui de leurs arguments.

3.2. Mouvements humains lors d'une performance musicale

Dans le travail de (Dalh & Friberg, 2007), les auteurs visent deux objectifs: (1) estimer l'expressivité des mouvements corporels (i.e. exclure l'aspect sonore) dans les performances musicales; et (2) trouver des critères (en terme de caractéristiques des mouvements) pour qualifier cette expressivité. Pour atteindre ces deux objectifs, ils définissent trois questions auxquelles répondre lors de l'expérimentation avec les sujets:

1. Comment les mouvements globaux (i.e. de toutes les parties du corps) ont réussi à communiquer l'émotion envisagée ?

2. Est-ce que la perception par les sujets des émotions exprimées dans la musique change en fonction du performeur ou de parties visibles du performeur ?
3. Comment l'émotion exprimée est-elle décrite en termes de caractéristiques des mouvements ?

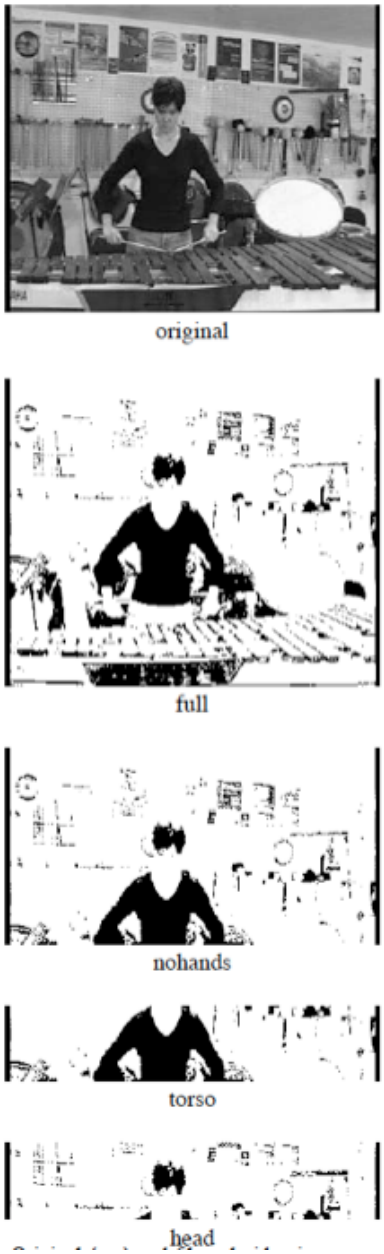


Figure 45 Différentes parties visibles utilisées dans l'expérimentation de (Dalh & Friberg, 2007)

Deux expérimentations ont été menées.

La première expérimentation consiste à évaluer la performance d'une percussionniste qui joue au xylophone. On a demandé à cette percussionniste d'interpréter un extrait neutre de Morris Goldenberg, "Melodic study in sixteens", en exprimant les quatre émotions : joie, tristesse, colère, peur. Les quatre interprétations durent chacune de 30 à 50 secondes. Sa performance est filmée et puis filtrée en différentes parties pour l'évaluation. Ces différentes parties sont : *tout le corps, la tête, la tête et le corps sans les mains, le corps sans la tête ni les mains* (Figure 45). 20 sujets ont participé à cette première expérimentation. On les a invité à regarder les vidéos sans le son et demandé d'estimer les émotions présentées via les mouvements du musicien. L'estimation est représentée sous la forme de quatre valeurs (chacune varie entre 0 (pas du tout) et 6 (très forte)) correspondant à 4 émotions (joie, tristesse, colère, peur). Ces sujets devaient aussi donner un avis sur les caractéristiques des mouvements, dont *la quantité* (nombre de mouvements)*, la vitesse, la fluidité, et la régularité*. Le résultat de cette première expérimentation révèle que la tristesse est la plus reconnue, puis la joie et la colère. La peur n'est pas bien reconnue, elle était confondue avec les trois autres émotions. Ils ont trouvé aussi que le mouvement de la tête facilite l'expression de la tristesse (i.e. la tristesse est mieux reconnue quand elle est présentée avec les mouvements de la tête parmi d'autres). La colère est la mieux reconnue quand présentée via toutes les parties disponibles (dont la vidéo avec toutes les parties présentes).

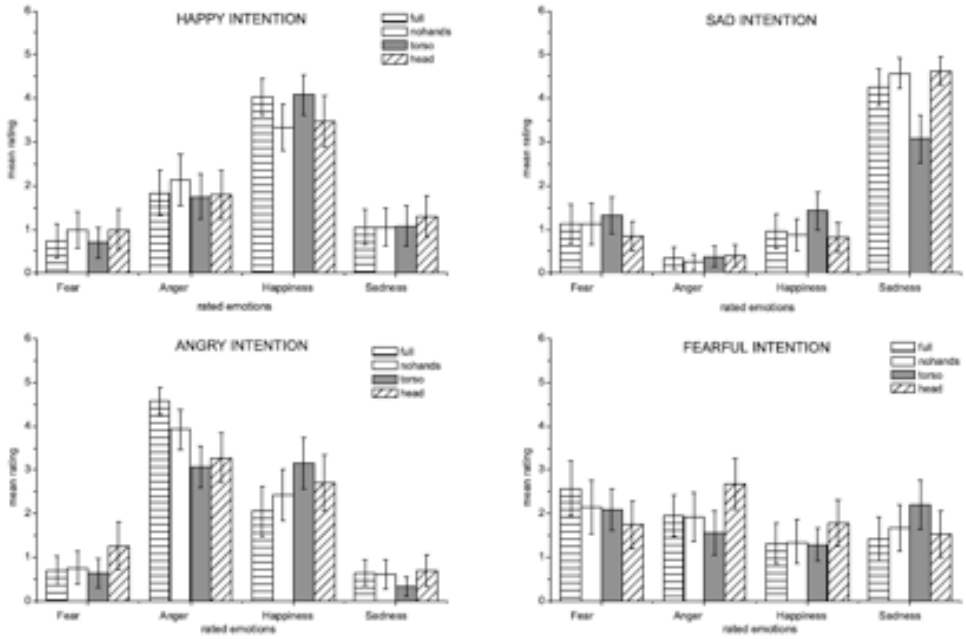


Figure 46 Résultat de reconnaissance de la première expérimentation de (Dalh & Friberg, 2007)

Les caractéristiques des mouvements ont aussi été identifiées. Les caractéristiques des mouvements pour la joie et la colère se ressemblent fortement, i.e. de grandes quantités, de grandes vitesses, et les mouvements de la colère sont moins réguliers et moins fluides que ceux de la joie. Les mouvements de la peur sont caractérisés par des mouvements rares, rapides et saccadés. Ceux de la tristesse sont caractérisés par des mouvements lents, fluides, et réguliers.

Pour affiner l'évaluation de l'expressivité des mouvements des musiciens, les chercheurs ont mené une autre expérimentation qui suit le même protocole d'expérimentation que la première. Les différences dans la deuxième expérimentation sont (1) l'utilisation d'instruments à vent (dont un saxophone soprano et un basson), (2) l'ajout d'extraits pour la performance des musiciens (4 extraits différents), et (3) seul le filtrage de tout le corps dans les vidéos est utilisé pour l'expérimentation. L'objectif de cette expérimentation est de voir comment les émotions sont communiquées quand les mouvements du musicien sont limités et sont connectés à l'instrument. 20 personnes ont participé à l'expérimentation. Le résultat de cette deuxième expérimentation révèle le même phénomène de reconnaissance des émotions que dans la première expérimentation : la tristesse, la joie et la colère sont bien reconnues et la peur n'était pas significativement exprimée. Les gens ont toujours tendance à confondre la joie avec l'expression de la colère. Les caractéristiques des mouvements pour les émotions ont eu aussi le même vote. Ce qui veut dire que la colère et la joie sont caractérisées par une grande quantité de mouvements, et ce sont des mouvements rapides et réguliers, et que les mouvements de la colère sont plus saccadés que ceux de la joie. La tristesse est caractérisée par une petite quantité de mouvements, et ce sont des mouvements lents, fluides, réguliers.

De manière générale, à travers ces deux expérimentations, les réponses pour les trois questions posées sont :

1. La joie, la tristesse et la colère sont bien communiquées, mais la peur ne l'est pas.

2. L'identification des émotions varie en général très peu selon les conditions de vue (i.e. tout le corps, seulement la tête, le corps sans la tête, le corps avec la tête mais sans le bas), tandis qu'il y a des cas où le tête est un indice important.
3. Les mouvements du musicien permettent de caractériser les émotions qu'il veut communiquer.

3.3. Mouvements robotiques lors d'une écoute musicale

L'objectif principal du travail de (Burger & Bresin, 2010) est de concevoir des mouvements robotiques pour accompagner le flux musical, autrement dit d'exprimer les émotions musicales via les mouvements d'un robot. En fait, la littérature sur le comportement humain montre qu'il existe une relation forte entre les émotions et les mouvements corporels de l'humain et qu'il existe aussi une relation forte entre la musique et les mouvements corporels, d'autant plus que la musique est bien connue comme un moyen d'exprimer/stimuler de l'émotion chez l'humain.

Pour mettre en œuvre leur idée, ils ont conçu un robot s'appelant MEX – un petit robot autonome qui peut être contrôlé via un ordinateur à distance. Le robot MEX peut faire des déplacements dans l'environnement via ses roues, bouger ses deux bras, et éviter les obstacles grâce à son capteur d'ultrason (qui ressemble à un genre de tête avec deux grands jeux) (voir Figure 47).

Figure 47 Robot MEX dans le travail de (Burger & Bresin, 2010)

Dans l'expérimentation avec MEX, les auteurs ont décidé de mettre en œuvre les mouvements de trois émotions : la joie, la tristesse, la colère. Ces émotions sont choisies à partir du travail de (Dalh & Friberg, 2007) où la peur est montrée comme la moins reconnue parmi les quatre émotions.

Les caractéristiques des mouvements émotionnels de MEX sont présentées dans la Figure 48. Les mouvements de joie sont en général à vitesse élevée, de trajectoire ronde, régulière, et avec les bras dirigés vers le haut. La colère est caractérisée par des mouvements rapides, saccadés, irréguliers, et avec les bras bougeant dans tous les sens. La tristesse est traduite par des mouvements lents, réguliers, non saccadés, et avec les bras bougeant vers le bas.

Movement cue	Emotion		
	Happiness	Anger	Sadness
Amount of gesture	Large	Large	Small
Speed	Fast	Fast	Slow
Fluency	Fluent	Jerky	Fluent
Regularity	Regular, circular	Irregular	Regular
Direction of arm movements	Upwards	Fast up and down	Slow up and down

Figure 48 Principe de mouvements émotionnels du robot MEX (Burger & Bresin, 2010)

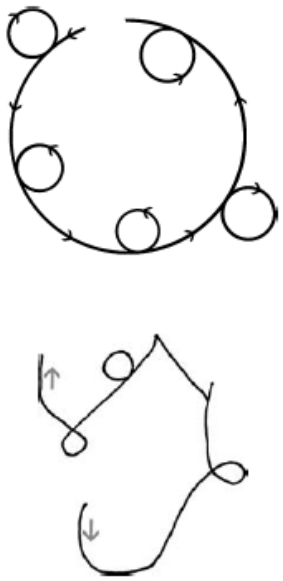


Figure 49 Exemples de mouvements de MEX pour la joie (la figure en haut) et la colère (la figure en bas) de (Burger & Bresin, 2010)

Pour valider leur conception, les auteurs ont mené deux expérimentations pour évaluer l'expressivité de MEX seul et l'expressivité de MEX accompagné par la musique. Le but était de voir si l'écoute de la musique affecte la perception des gens sur les mouvements du robot.

La première expérimentation est conduite en Finlande, avec la participation de 15 personnes (dont 10 hommes et 5 femmes d'âge moyen de 28 ans). Les mouvements de MEX sont présentés d'abord sans musique en arrière plan, puis avec musique. L'ordre des émotions présentées est alterné d'une condition à l'autre. Chaque mouvement émotionnel dure environ 30 secondes. Les participants doivent observer les mouvements et marquer sur une grille d'évaluation leur perception sur les mouvements. Chaque mouvement est évalué en 5 catégories (joie, tristesse, colère, neutre, et expressivité) et sur une échelle de type Likert variant de *rien du tout…* à *très…* L'analyse sur les grilles d'évaluation des participants dans cette expérimentation montre que la tristesse est la mieux reconnue avec ou sans la musique en arrière plan. La joie et la colère sont assez bien exprimées, pourtant l'ajout de la musique affecte différemment la perception de ces deux émotions. Tandis que l'ajout de la musique permet de mieux reconnaître la joie, les sujets ont eu tendance à percevoir la colère comme la joie quand les mouvements de colère de MEX sont présentés avec la musique de colère.

La deuxième expérimentation sur les mouvements de MEX est conduite en Allemagne avec 36 participants (19 hommes, 16 femmes, et un participant ne donne pas l'information sur son sexe). Un changement dans la vitesse pour la tristesse est fait pour rendre les mouvements plus réguliers ; un autre changement dans les mouvements des bras pour renforcer l'intensité de la colère. La procédure d'expérimentation est la même que celle de la première expérimentation. L'analyse des évaluations des participants confirme les effets retrouvés dans la première expérimentation, incluant la meilleure perception de la tristesse, l'impact positif de la musique pour la joie et l'impact négatif de la musique pour la colère.

De manière générale, (Burger & Bresin, 2010) montre que l'expression de l'émotion via les mouvements d'un robot mobile sans visage est possible. L'ajout de la musique en arrière plan pourrait renforcer l'expressivité du robot. Bien qu'ils aient mis en œuvre les mouvements des bras, il n'est pas bien clairement établi comment ces mouvements affectent la perception des gens sur les émotions exprimées par le robot. De plus, ils ont trouvé un effet lié à l'ordre de présentation des émotions sur la perception des sujets, ce qui met en question l'efficacité de leur expérimentation.

3.4. Conclusion

Les travaux présentés ci-dessus montrent que l'expression émotionnelle est en général bien perçue par l'humain. Pourtant, il existe toujours des confusions d'une émotion à l'autre, même dans le cas de l'expression humaine. Nous présenterons dans la section suivante notre conception qui permet d'améliorer ce taux de reconnaissance en utilisant aussi peu de mouvements que possible de notre robot LINA. A noter que ces mouvements sont développés pour être compatibles avec le contexte musical de notre projet, ce qui oriente notre attention vers deux (parmi les trois) travaux présentés dans cette section.

4. Mouvements de LINA – Notre conception

Pour notre robot LINA, les mouvements correspondant aux quatre émotions Joie, Tristesse, Colère, Sérénité sont implémentés. Comme décrit dans le chapitre sur les descripteurs musicaux, ce sont les quatre émotions les plus exprimées dans la musique. De plus, elles sont éloignées les unes des autres dans l'espace Valence - Activation de l'émotion. La représentation Valence – Activation semble la plus pertinente pour mettre en correspondance une émotion donnée et l'action corporelle pour exprimer cette émotion.

Le robot utilisé dans notre expérimentation s'appelle LINA – un robot mobile développé par Droids Company (http://www.droids-company.com/topic/index-en.html). LINA dispose d'un système d'actionneurs et de capteurs qui lui permet de se déplacer dans l'environnement et d'éviter des obstacles. Il a deux moteurs de déplacement et peut se déplacer dans toutes les directions. LINA dispose aussi d'un système de contrôle pour commander le robot à distance, soit manuellement via un joystick, soit automatiquement via un programme. Un simulateur de LINA est aussi fourni qui nous permet de tester les mouvements de LINA sans utiliser le vrai robot.

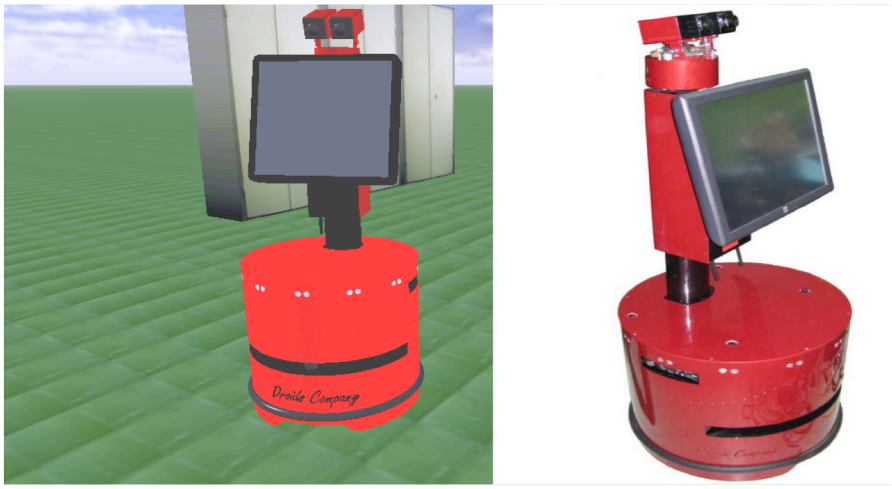


Figure 50 Robot LINA - Le vrai robot à droite et le simulateur à gauche

Dans le contexte d'expression de l'émotion dans la musique, les mouvements de LINA sont inspirés des travaux de (Burger & Bresin, 2010) et de (Dalh & Friberg, 2007). (Dalh & Friberg, 2007) expliquent que l'humain a l'habitude d'interpréter le contenu émotionnel dans la musique via les mouvements du corps de l'artiste (i.e. le musicien, le danseur, etc.). Tandis que S. Dalh et A. Friberg exploitent les mouvements des différentes parties du corps humain, le robot MEX de (Burger & Bresin, 2010) implémente des déplacements dans l'espace et des mouvements de bras. Comme LINA ne dispose pas de bras, nous avons décidé de mettre en œuvre des déplacements dans l'espace et des changements de direction de la caméra de LINA, cette dernière donnant au spectateur l'illusion que le robot possède un regard.

La représentation des émotions selon les deux axes Valence-Activation permet de définir des caractéristiques de déplacements liées aux roues de LINA. L'axe Valence est associé à la forme géographique des trajectoires. Les émotions de haute valence sont exprimées par des trajectoires rondes, tandis que celles de basse valence sont associées à des trajectoires saccadées. L'Activation de l'émotion est associée avec la vitesse de déplacement. Quand l'activation est élevée, LINA se déplace plus vite et change de direction plus brusquement. Quand l'activation est basse, LINA se déplace lentement, et change de direction de temps en temps.

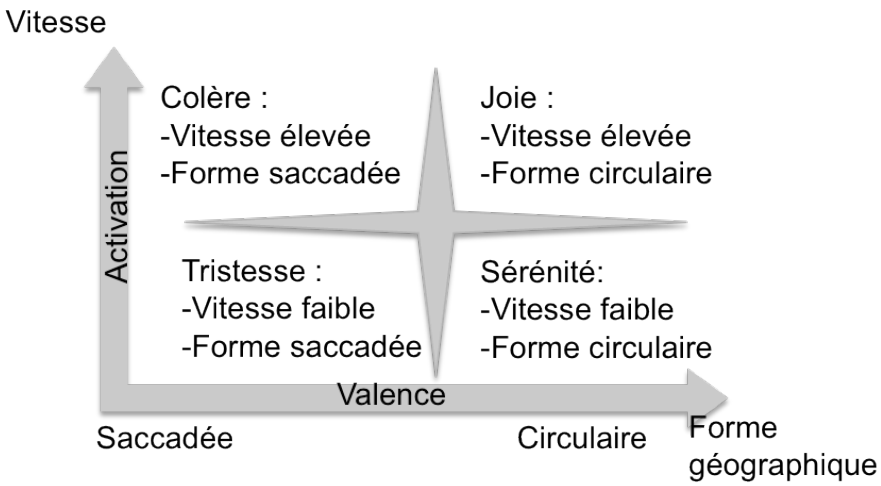


Figure 51 Principe de déplacements de LINA en fonction de la valence et l'activation

Quelques exemples de déplacements de LINA sont présentés dans la Figure 52. La joie est exprimée par une trajectoire ronde à vitesse modérée (trajet b). La colère est exprimée via une trajectoire saccadée et rapide avec de grands changements de direction (trajet a). La tristesse est associée avec un trajet saccadé mais à vitesse réduite et avec des arrêts (trajet c). La sérénité est exprimée via une trajectoire arrondie typiquement en forme de huit (trajet d) à vitesse modérée.

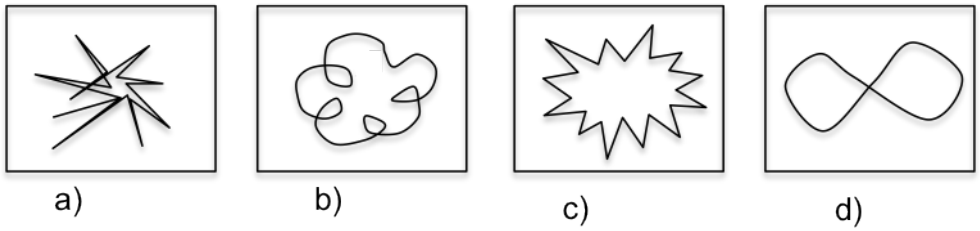


Figure 52 Exemples de déplacement de LINA. a) la colère, b) la joie, c) la tristesse, d) la sérénité

Comme présenté dans la section 2.4 sur le travail de B. Lance et G. Marsella, l'émotion peut aussi s'exprimer via le regard. La direction du regard (vers le haut ou vers le bas) indique normalement la valence de l'émotion. La vitesse, par ailleurs, peut être associée à l'activation de l'émotion. En outre, S. Dalh et A. Friberg suggèrent que la colère et la tristesse sont associées à des mouvements saccadés du regard (ou bien de la tête en général). La joie est exprimée via des mouvements assez réguliers et non saccadés. Pour LINA, la tristesse est associée avec un regard vers le bas avec quelques changements en direction (i.e. vers la gauche, vers la droite) pour éviter l'inactivité. La joie est associée à un regard vers le haut et avec quelques changements en direction (i.e. vers la gauche, vers la droite) pour éviter l'inactivité. Pour la colère, le regard change assez souvent sa direction pour exprimer le mécontentement. Pour la sérénité, le regard de LINA fait un trajet en forme de huit à vitesse modérée.

5. Expérimentation

Nous présentons maintenant une expérimentation menée lors de la Fête de la Science pour valider notre conception pour le robot LINA, comprenant l'objectif de l'expérimentation, la description sur le lieu de l'expérimentation, les participants et enfin la procédure de l'expérimentation.

5.1. Objectifs de l'expérimentation

L'objectif de l'expérimentation est d'estimer l'expressivité de LINA via les mouvements que nous avons présentés. A travers l'expérimentation, nous cherchons à vérifier (1) si les émotions de LINA exprimées via ses déplacements et les mouvements de caméra peuvent être reconnues par les sujets, et (2) si la musique facilite la reconnaissance des émotions de LINA. Pour cela, il faut estimer la manière dont les sujets perçoivent l'expression émotionnelle du robot, et estimer comment la musique affecte la perception par ces sujets de l'expression émotionnelle du robot. Les émotions utilisées pour cette expérimentation sont la joie, la tristesse, la colère, et la sérénité.

Pour ce faire, nous avons défini trois conditions dans lesquelles nous demandons aux sujets d'évaluer l'émotion exprimée :

- Condition Musique Seule : dans cette condition, nous présentons quatre extraits musicaux représentatifs des quatre émotions. Cette condition est pour s'assurer de la représentativité des extraits utilisés. Ce sont des extraits musicaux de Robert Schumann qui sont utilisés. Nous avons préparé deux groupes d'extraits, chacun contenant quatre extraits représentant les quatre émotions. Chaque extrait dure environ 30 secondes.
- Condition Robot Seul : dans cette condition, nous présentons aux sujets les quatre mouvements du robot (déplacement dans l'espace et mouvement de la caméra) représentant les quatre émotions. Cette condition est pour tester l'expressivité des mouvements que nous avons définis.
- Condition Robot Plus Musique : dans cette condition, les mouvements du robot sont présentés avec la musique en arrière plan. Des quatre émotions exprimées par le robot, deux sont accompagnées par une musique exprimant la même émotion, deux autres sont accompagnées par une musique exprimant une émotion différente. Pour simplifier la gestion et l'analyse de ces cas, dans le cas où la musique est en discordance avec l'émotion du robot, nous associons des émotions opposées dans le repère Valence / Activation. La colère du robot est accompagnée par un extrait de musique sereine, et quand le robot exprime la sérénité, c'est une musique qui exprime la colère qui l'accompagne. Le même principe est appliqué pour la joie et la tristesse : quand le robot exprime la tristesse, il est accompagné par une musique joyeuse, et quand le robot exprime la joie, une musique triste est jouée en arrière plan.

5.2. Configuration de l'espace d'expérimentation

Notre expérimentation a eu lieu pendant la Fête de la Science organisée par l'Université d'Evry-Val d'Essonne, le 13 Octobre 2010. Elle s'est déroulée dans un

espace de 16 m^2. L'organisation de l'espace d'expérimentation est présentée dans la Figure 53. LINA est situé au milieu de la salle, il dispose d'une espace de 2m x 2m pour ses déplacements. Autour de cet espace, nous disposons des chaises et des tables pour que les participants puissent s'asseoir en regardant l'expression de LINA.

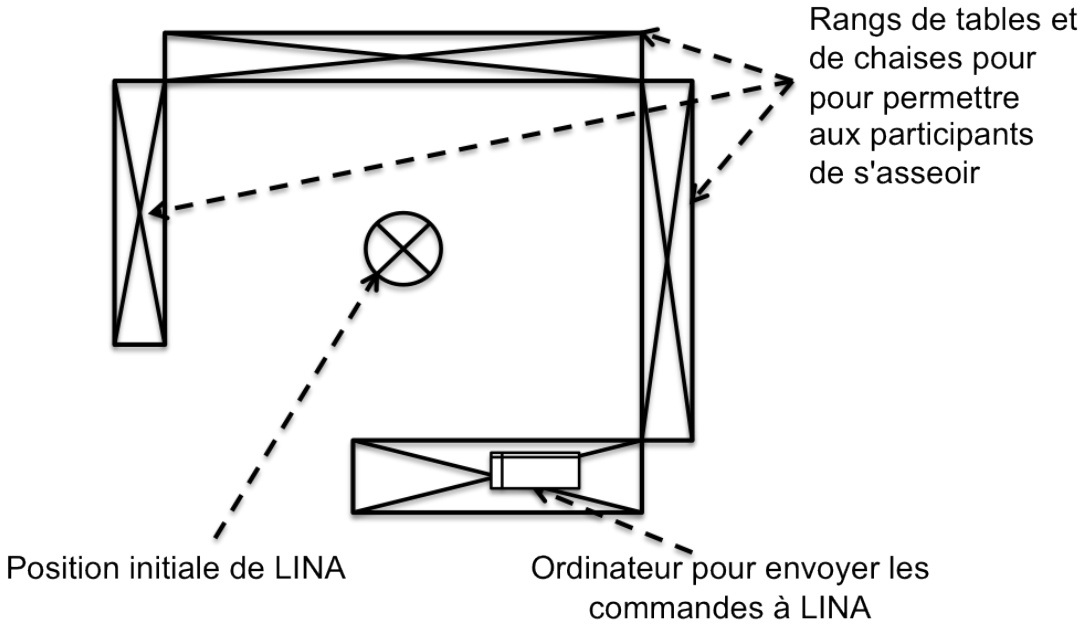


Figure 53 Organisation de l'espace d'expérimentation

5.3. Participants

Pendant la fête de la science, nous avons accueilli plusieurs groupes de lycéens et de collégiens. Ils avaient entre 10 et 17 ans, mélangés de manière à peu près homogène entre filles et garçons. Chaque groupe est accompagné par un professeur, dont l'âge est compris entre 30 et 50 ans environ. L'expérimentation avec chaque groupe dure environ 30 minutes.

5.4. Procédure expérimentale

Quand un groupe arrive au stand, les sujets sont informés qu'ils vont participer à une expérimentation scientifique pour évaluer l'expression émotionnelle d'un robot dans un contexte d'écoute musicale. Nous présentons tout d'abord le contexte de recherche en général, et distribuons les feuilles d'évaluation (voir en Annexe 1.1 le format de la feuille d'évaluation). Nous expliquons les termes présentés dans la feuille. Après cette introduction, ces participants (incluant les élèves et le(s) professeur(s)) sont mis en situation dans les trois conditions et remplissent la feuille d'évaluation. Les feuilles sont rassemblées après que les trois conditions aient été présentées. Nous expliquons les réponses correctes pour chaque condition et terminons par une discussion informelle.

Comme indiqué précédemment, les trois conditions sont *Musique Seule*, *Robot Seul*, et *Robot Plus Musique*. Pour éviter le biais lié à l'ordre de présentation de ces conditions, l'ordre de présentation est changé d'un groupe d'élèves à l'autre. De plus,

l'ordre de présentation des émotions dans chaque condition est aussi alterné d'une condition à l'autre et d'un groupe à l'autre.

Pour chaque expression de l'émotion, le participant devait choisir l'émotion la plus appropriée parmi les quatre émotions proposées (i.e. la joie, la tristesse, la colère, la sérénité). Pour la condition *Robot Plus Musique*, nous laissons la possibilité aux participants de choisir plus d'une émotion, car la musique et le robot sont parfois en discordance du point de vue des émotions exprimées. Entre chaque présentation, les participants ont 10 secondes pour indiquer leur choix sur la feuille.

6. Résultats et discussions

Tout au long de la journée, nous avons reçu 10 groupes d'élèves, ce qui représente un total de 161 participants. L'analyse des réponses des participants est réalisée suivant les critères suivants.

Pour les conditions *Robot Seul* et *Musique Seule*, le choix correct est celui correspondant à l'émotion exprimée. Par exemple, si le robot exprime la joie, nous donnons 1 point pour le choix Joie et 0 sinon. Il y a des participants qui ont donné deux choix ou plus pour la même expression, ce que nous appelons des réponses confuses. Ces réponses confuses sont considérées comme fausses et sont donc associées au score 0. Par exemple, si on présente une musique joyeuse, et que le participant a choisi à la fois la joie et la sérénité dans sa réponse, cette réponse est systématiquement considérée comme fausse, même si la joie est bien présente dans la réponse du participant.

Dans la condition *Robot Plus Musique*, nous traitons des réponses composées de deux parties : une pour l'expression du robot, et une autre pour la musique. Dans le cas où la musique et le robot sont en concordance du point de vue de l'émotion exprimée, nous donnons 1 point pour la bonne réponse, et 0 sinon. Si le robot et la musique expriment en parallèle la même émotion, par exemple la joie, si le participant a choisi la joie seule, on lui attribue le score 1 (0.5 pour le robot et 0.5 pour la musique) ; si le participant a choisi la joie et une autre émotion, on lui attribue le score 0.5 car il a quand même reconnu la joie ; si la joie n'est pas choisie dans la réponse du participant, on lui attribue le score 0. Dans le cas où on présente le robot avec une musique en discordance (e.g. le robot exprime la joie et la musique exprime la tristesse), si le participant choisit la joie et la tristesse, on lui attribue le score 1 ; si le participant ne choisit qu'une seule des deux émotions, on lui attribue le score 0.5 ; si le participant ne choisit aucune des deux émotions, on lui attribue le score 0.

6.1. Analyse des résultats obtenus

Dans cette section, nous allons faire l'évaluation des résultats de l'expression de LINA. Cela consiste à évaluer le taux de reconnaissance de l'expression de LINA sans la musique, le taux de reconnaissance avec la musique en arrière-plan, et pour chaque condition nous ajoutons la comparaison avec les travaux dans la littérature présentés dans la section 3.2.

6.1.1. Expression du robot sans la musique en arrière plan

L'expression de LINA semble bien reconnue par les sujets, comme présenté dans la Figure 54. Toutes les expressions ont été reconnues largement au-dessus du taux de

hasard (i.e. 25%). Ce résultat montre que l'expression du robot est bien compréhensible par les sujets. En fait, exprimer de l'émotion via un robot sans les traits du visage est un défi, parce que l'humain a l'habitude de regarder le visage pour deviner l'émotion exprimée. Dans notre expérimentation, il y a des participants qui confirment que reconnaître l'émotion exprimée du robot sans trait de visage est vraiment difficile. Et pourtant, notre résultat montre que notre robot a bien exprimé des émotions de manière identifiable. Ce résultat est aussi prometteur pour les robots qui n'ont pas de possibilité de simuler le visage.

Table 11 Taux de reconnaissance pour la condition Robot Seul

		Taux de reconnaissance				
		Joie	Tristesse	Colère	Sérénité	Pas reconnue
Emotion exprimée	Joie	49.1	8.7	11.8	25.5	5.0
	Tristesse	4.3	74.5	6.2	11.8	3.1
	Colère	27.3	2.5	67.1	0.0	3.1
	Sérénité	7.5	17.4	10.6	59.6	5.0

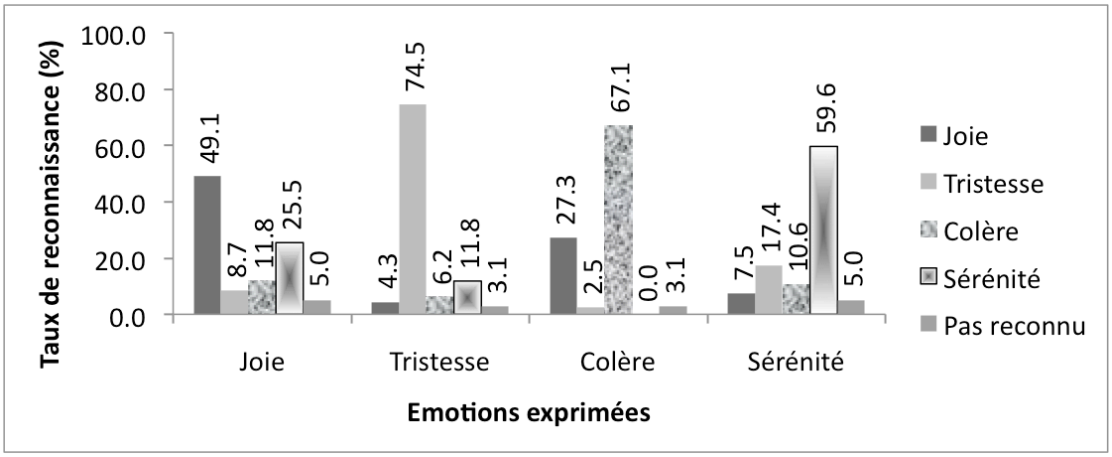


Figure 54 Taux de reconnaissance pour la condition Robot Seul

On peut aussi présenter le résultat obtenu sur forme de matrice de confusion, comme présenté dans le Table 12. Pour chaque émotion, on considère l'expression positive et l'expression négative. L'expression positive est l'expression de l'émotion en question, l'expression négative est l'expression des autres émotions. Par exemple, l'expression négative de la joie peut être l'expression de la tristesse, de la colère ou de la sérénité. Dans la matrice de confusion pour la joie, les valeurs de l'expression positive sont donc calculées à partir des pourcentages obtenus de la joie ; les valeurs des expressions négatives sont la somme des pourcentages des expressions de la tristesse, de la colère et de la sérénité. Le même principe est appliqué pour les autres matrices de confusion.

Table 12 Matrice de confusion pour la condition Robot Seul.

Joie		Reconnaissance	
		Positive	Négative
Expression	Positive	49.1	50.9
	Négative	39.1	260.9

Tristesse		Reconnaissance	
		Positive	Négative
Expression	Positive	74.5	25.5
	Négative	28.6	271.4
Colère		Reconnaissance	
		Positive	Négative
Expression	Positive	67.1	32.9
	Négative	28.6	271.4
Serenity		Reconnaissance	
		Positive	Négative
Expression	Positive	59.6	40.4
	Négative	37.3	262.7

Table 13 Statistique sur les matrices de confusion de la condition Robot Seul. (Les formules pour calculer ces valeurs sont présentées dans l'Annexe 1.4)

Condition Robot Seul				
	Joie	Tristesse	Colère	Sérénité
Accuracy (AC)	0.77	0.86	0.85	0.81
Recal or true positive rate (TP)	0.49	0.75	0.67	0.60
False Positive Rate (FP)	0.13	0.10	0.10	0.12
True Negative Rate (TN)	0.87	0.90	0.90	0.88
False Negative Rate (FN)	0.51	0.25	0.33	0.40
Precision (P)	**0.56**	**0.72**	**0.70**	**0.62**

En analysant les matrices de confusion des quatre émotions exprimées, la reconnaissance de la joie semble la plus défavorable. Avec la précision de 0.56, il semble qu'un sujet sur deux confond l'expression de la joie de LINA avec une autre émotion.

Cependant, en comparant avec (Burger & Bresin, 2010) avec le robot MEX, dans la condition Robot Seul, nous avons de meilleurs résultats. Tandis que l'expression de MEX donnait de la confusion entre la joie et la colère, notre expérimentation montre que LINA ne subit pas cette confusion. Cette dominance de notre résultat pourrait être due à l'ajout du regard dans les mouvements de LINA. En fait, comme suggéré dans (Lance & Marsella, 2010), le regard joue un rôle important pour renforcer l'expressivité des comportements émotionnels des agents artificiels. En appliquant cette suggestion pour notre robot, il semble que cet élément (i.e. le regard) soit aussi important pour les robots.

Il y a une grande tendance à implémenter une expression faciale pour les agents émotionnels et les robots émotionnels. Ces expressions sont synthétisées par exemple dans le cadre du système FACS – Facial Action Code System développé par (Ekman, Friesen, & Hager, 2002). Récemment, (Kuhnlenz, Sosnowski, & Buss, 2007) ont essayé d'évaluer comment les gens perçoivent les expressions émotionnelles d'un visage robotisé. Les résultats de ce travail permettent de faire une comparaison entre

l'expressivité du visage et des mouvements corporels (i.e. sans l'expression faciale). La comparaison est faite sur les trois émotions communes dans les deux travaux : la joie, la tristesse, la colère. Pour la tristesse, l'expression faciale a le meilleur taux de reconnaissance (90% par rapport à 74.5% dans notre travail). Par contre, l'expression corporelle de la colère et de la joie dans notre travail reçoit de meilleurs taux de reconnaissance : la joie dans notre cas est reconnue à un taux de 49.1% contre seulement 25% pour leur travail ; la colère dans notre cas est reconnue par 67.1% des participants contre 50% dans le travail de K. Kuhnlenz.

6.1.2. Expression du robot avec la musique en arrière plan

Pour cette condition, la musique est aussi un élément qui affecte l'expression émotionnelle présentée. Nous cherchons tout d'abord à déterminer si la musique utilisée est bien représentative pour l'expression de l'émotion via la condition Musique Seule.

Table 14 Taux de reconnaissance pour la condition Musique Seule

		Taux de reconnaissance				
		Joie	Tristesse	Colère	Sérénité	Pas reconnue
Emotion exprimée	Joie	62.1	0.6	17.4	18.0	1.9
	Tristesse	1.2	93.8	0.0	4.3	0.6
	Colère	15.5	2.5	74.5	6.2	1.2
	Sérénité	23.6	6.2	1.9	65.2	3.1

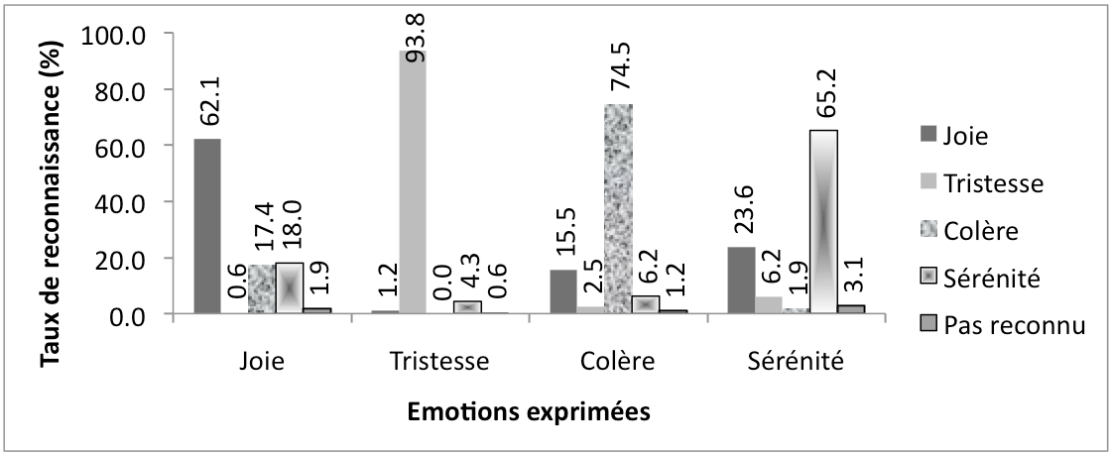


Figure 55 Taux de reconnaissance pour la condition Musique Seule

De manière générale, la musique est bien reconnue par les participants (Figure 55). Nous remarquons aussi que les émotions négatives (i.e. la tristesse et la colère) sont mieux reconnues que les émotions positives (i.e. la joie et la sérénité). Ce résultat confirme certaines expérimentations en psychologie (Zentner, Grandjean, & Scherer, 2008) qui montrent que les sujets semblent mieux reconnaître les émotions négatives que les émotions positives. De plus, en prenant en compte les taux de reconnaissance des émotions dans la condition Robot Plus Musique en concordance (Figure 56), il apparait que les émotions sont mieux exprimées par la musique seule que par les mouvements du robot ou par les mouvements du robot avec la musique en arrière plan. Il y a deux raisons principales pour ce phénomène. La première raison est liée à la popularité de la musique dans la vie humaine. L'omniprésence de la musique dans

la vie de tous les jours affecte/améliore la perception des contenus musicaux de l'humain. De plus, comme la musique est aussi un des moyens les plus utilisés pour l'humain pour exprimer ses émotions, l'aisance dans la perception du contenu émotionnel pour un humain est évidente. La deuxième raison est liée à l'utilisation du robot pour exprimer en même temps la même émotion. On voit bien que quand le robot exprime de l'émotion sans la musique, le taux de reconnaissance est le plus faible parmi les trois conditions. Quand on ajoute l'expression émotionnelle du robot à la musique, il est raisonnable que l'expression combinée soit perturbée et donc moins bonne que l'expression de la musique seule. Notre objectif dans ce projet est d'évaluer notre conception des expressions émotionnelles pour le robot, et nous considérons donc dans notre expérimentation que la musique est un élément supplémentaire qui pourrait aider le robot à mieux exprimer ses émotions. Nous analysons dans la suite les résultats obtenus pour révéler le rôle de la musique pour l'expression du robot.

Passons maintenant à la condition Robot Plus Musique en concordance (Figure 56) pour voir si la musique aide LINA à mieux exprimer ses émotions. Si le taux de reconnaissance augmente quand LINA est accompagné par la musique appropriée, et que ce taux diminue quand LINA est accompagné par une musique en discordance, ceci permet de valider notre choix de conception. La Figure 56 présente le taux de reconnaissance de la condition *Robot Plus Musique* quand ces deux parties présentent les mêmes émotions.

Table 15 Taux de reconnaissance pour la condition Robot Plus Musique en concordance

		Taux de reconnaissance				
		Joie	Tristesse	Colère	Sérénité	Pas reconnue
Emotion exprimée	Joie	46.0	3.0	11.6	31.8	7.6
	Tristesse	0.0	92.9	0.0	7.1	0.0
	Colère	19.5	5.8	68.8	3.2	2.6
	Sérénité	20.1	10.4	2.6	64.3	2.6

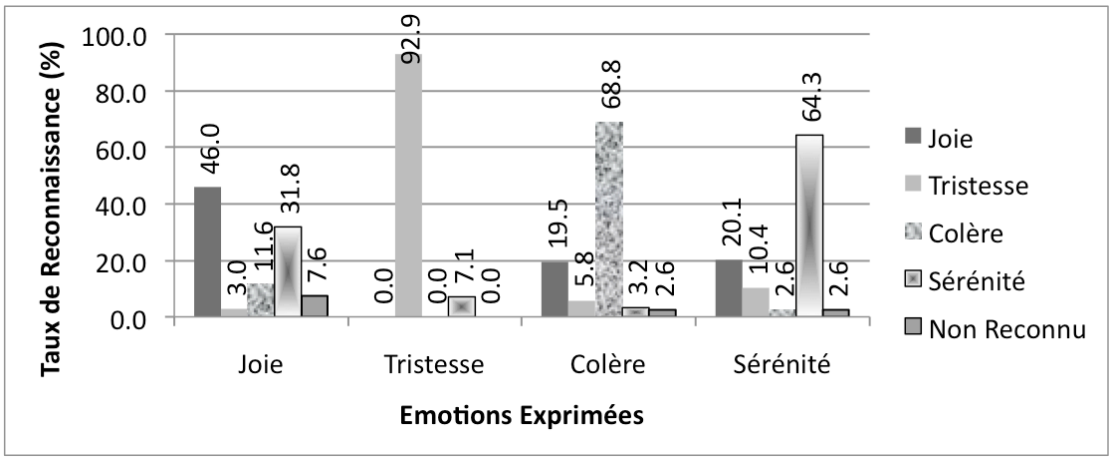


Figure 56 Taux de reconnaissance de la condition Robot Plus Musique en concordance

Table 16 Matrices de confusion pour la condition Robot Plus Musique en concordance

Joie		Reconnaissance	
		Positive	Negative
Expression	Positive	46.0	54.0
	Negative	39.6	260.4
Tristesse		Reconnaissance	
		Positive	Negative
Expression	Positive	92.9	7.1
	Negative	19.3	280.7
Colère		Reconnaissance	
		Positive	Negative
Expression	Positive	68.8	31.2
	Negative	14.2	285.8
Serenity		Reconnaissance	
		Positive	Negative
Expression	Positive	64.3	35.7
	Negative	42.1	257.9

Table 17 Statistique sur les matrices de confusion pour la condition Robot Plus Musique en concordance

Condition Robot Plus Musique en Concordance				
	Joie	Tristesse	Colère	Sérénité
Accuracy (AC)	0.77	0.93	0.89	0.81
Recal or true positive rate (TP)	0.46	0.93	0.69	0.64
False Positive Rate (FP)	0.13	0.06	0.05	0.14
True Negative Rate (TN)	0.87	0.94	0.95	0.86
False Negative Rate (FN)	0.54	0.07	0.31	0.36
Precision (P)	**0.54**	**0.83**	**0.83**	**0.60**

La figure 57 ci-dessous présente une synthèse sur la confusion des quatre expressions pour les deux conditions : Robot Seul et Robot Plus Musique en concordance. Les valeurs de True Positive et False Positive montre que l'ajout de la musique a amélioré significativement la reconnaissance de la tristesse, légèrement pour la reconnaissance de la colère et la sérénité, et diminué la reconnaissance de la joie.

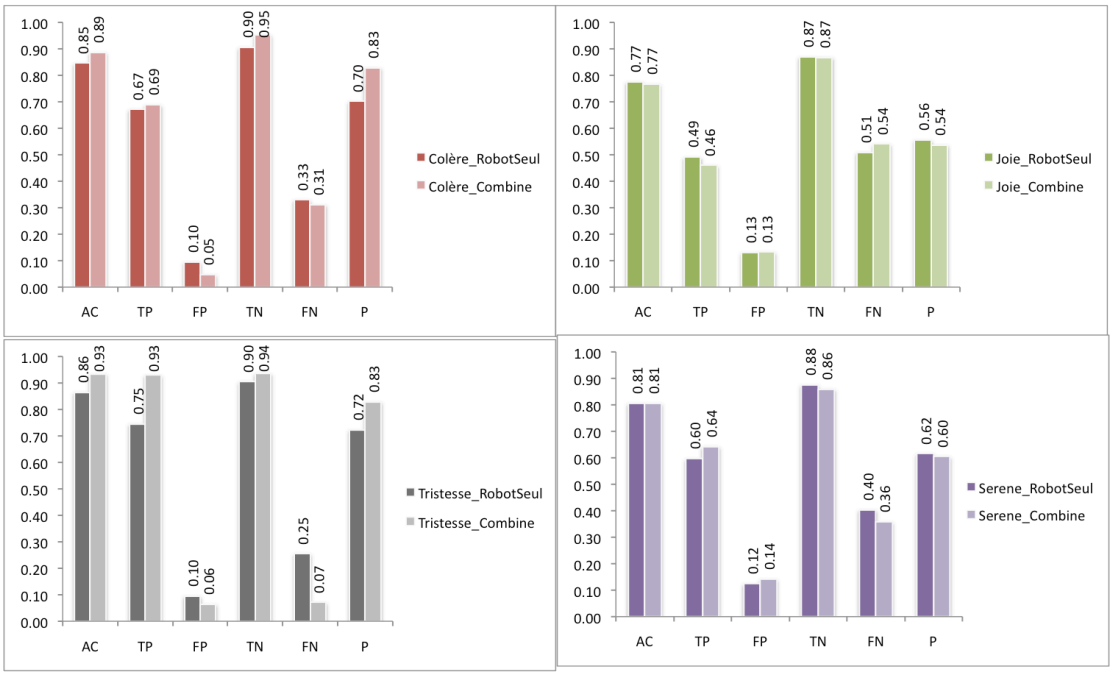


Figure 57 Statistique des matrices de confusion pour les deux conditions Robot Seul et Robot Plus Musique en concordance. AC = Accuracy, TP = True Positive, FP = False Positive, TN = True Negative, FN = False Negative, P = Precision.

Même si l'ajout de la musique permet d'améliorer l'expression de trois des quatre émotions, l'analyse des taux de reconnaissance en fonction de la représentation dimensionnelle des émotions relève des points intéressants. Les taux de reconnaissance des expressions en fonction de Valence - Activation sont présentés dans la figure 58.

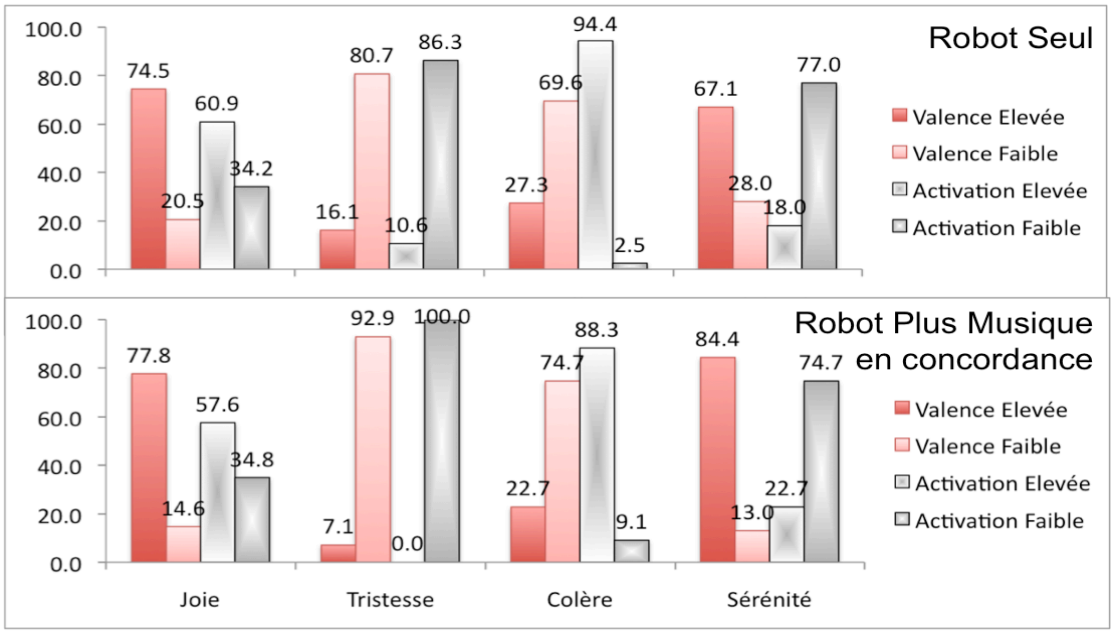


Figure 58 Taux de reconnaissance en fonction de Valence-Activation pour la condition Robot Seul et Robot Plus Musique en concordance

La joie dans la condition *Robot Seul* est reconnue comme de basse valence à 20.5% (8.7% pour la tristesse et 11.8% pour la colère), et dans la condition *Robot Plus Musique en concordance* ce taux de fausse reconnaissance diminue à 14.6%. Le taux de fausse reconnaissance pour la valence de la tristesse diminue de 16.1%, dans la condition *Robot Seul*, à 7.1% dans la condition *Robot Plus Musique en concordance*. Celui de la colère diminue de 27.3%, dans la condition *Robot Seul*, à 22.7% dans la condition *Robot Plus Musique en concordance*. Et celui de la sérénité diminue de 28%, dans la condition *Robot Seul*, à 13% dans la condition *Robot Plus Musique en concordance*. Cette observation suggère que l'utilisation de la musique pourrait renforcer l'expression de la valence des émotions que les robots doivent communiquer aux gens.

Comparé aux résultats de (Burger & Bresin, 2010), nous avons obtenu un meilleur résultat. Le travail de B. Burger et R. Bresin évalue les trois émotions Joie, Tristesse, Colère. Nous allons donc comparer seulement ces trois émotions dans les deux expérimentations. Les deux expérimentations ont bien réussi à exprimer la tristesse. Dans notre expérimentation, l'ajout de la musique semble augmenter l'expressivité de cette émotion, ce qui n'est pas le cas dans le travail de B. Burger et R. Bresin. Dans leur expérimentation, quand le robot exprime la colère avec la musique de colère en arrière-plan, l'expression est reconnue comme la joie. Dans notre expérimentation, la confusion entre la colère et la joie est diminuée quand les mouvements robotisés de ces émotions sont accompagnés par la musique.

La comparaison de notre résultat avec l'expérimentation sur l'expression des musiciens dans (Dalh & Friberg, 2007) donne une validation pour notre approche. Comme montré par nos résultats, la tristesse semble l'émotion la mieux reconnue, tandis que la joie est parfois confondue avec la colère, et ces phénomènes sont constatés également dans le travail de S. Dahl et A. Frieberg. Ceci indique que la conception des mouvements de LINA est dans la bonne direction. Pourtant, la confusion de la joie avec la sérénité semble élevée, ce qui implique que la conception

des deux émotions n'est pas très bien faite. Nous discuterons les améliorations possibles pour limiter cette confusion dans la section suivante.

Pourtant, parmi les quatre émotions exprimées dans la condition *Robot Plus Musique* dans notre expérimentation, la joie semble la moins bien reconnue. Nous discuterons de ce phénomène dans la section suivante.

Finalement, il est clair que quand la musique n'est pas en concordance avec l'expression émotionnelle du robot LINA, les participants ont des difficultés à identifier les émotions exprimées (Figure 59). Peu de sujets arrivaient à identifier les deux émotions exprimées (0.8% pour la combinaison Joie & Tristesse et 0.6% pour la combinaison Colère & Sérénité). La combinaison de la joie et la tristesse est généralement perçue comme de la tristesse (à 39.5%), et/ou de la sérénité (à 30.6%). L'humain perçoit la combinaison de la colère et de la sérénité comme de la colère à 47.6% et comme de la joie à 23.8%. Pourtant, il n'est pas très clair si c'est le robot ou la musique qui affecte principalement la décision des sujets quant à l'émotion exprimée.

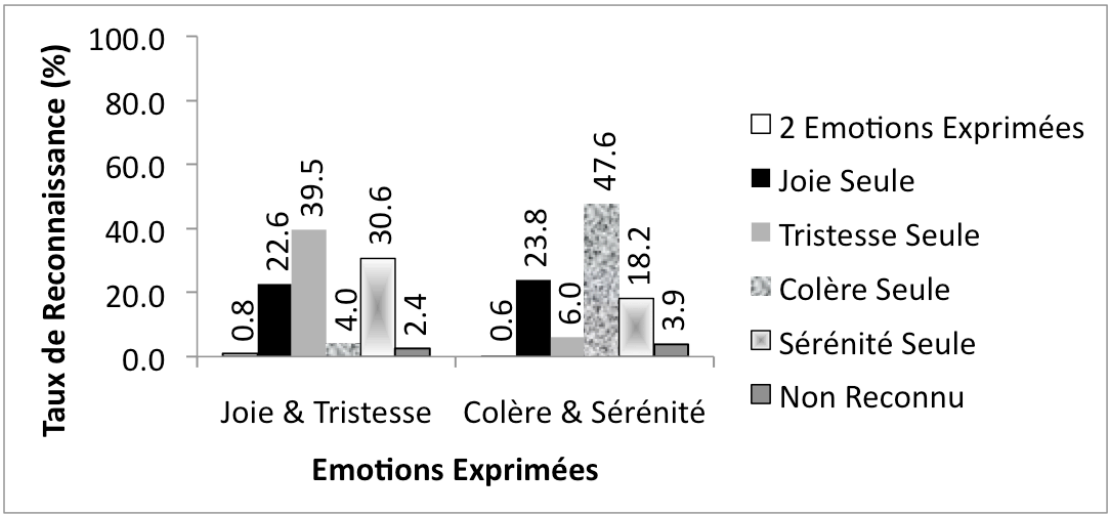


Figure 59 Taux de reconnaissance de la condition Robot Plus Musique en discordance

6.2. Discussion et Perspectives

Au travers du résultat dans la condition *Robot Seul*, l'expression de la joie de LINA est la moins bien reconnue. La joie est confondue avec la colère dans la condition *Robot Seul*, et confondue avec la sérénité dans la condition *Robot Plus Musique en concordance*. Bien que ce genre de confusion soit aussi constaté dans la littérature (e.g. robot MEX de (Burger & Bresin, 2010), l'expression des musiciens dans (Dalh & Friberg, 2007)), nous avons identifié la raison de cette confusion dans notre contexte. La raison principale est liée à la vitesse linéaire du robot. Nous souhaitons associer la vitesse linéaire à la valeur de l'activation de l'émotion. Pour les émotions d'activation élevée (i.e. la joie et la colère), nous aurions aimé mettre en œuvre un déplacement plus rapide et avec plus de changements de direction. En pratique, le robot ne peut pas se déplacer à de telles vitesses dans un environnement aussi contraint (2mx2m) sans risque de basculer. Nous devons donc limiter la vitesse

linéaire, ce qui entraîne que le déplacement pour la joie ne donne pas une impression suffisamment active.

Par ailleurs, comme notre analyse sur la condition *Robot Plus Musique* en concordance le suggère, la musique pourrait aider à exprimer la valence. De plus, du fait de l'impact positif que nous avons pu mettre en évidence de la caméra de LINA sur l'expression émotionnelle, nous avons au moins deux possibilités pour améliorer l'expression des émotions. Par exemple, le concepteur peut se concentrer sur la modélisation de l'aspect Activation dans les déplacements du robot et implémenter l'aspect Valence pour les parties hautes (comme le regard dans notre cas) ou employer un autre élément dans l'environnement (comme la musique dans notre cas).

Pour les expérimentations à venir sur le même sujet, il semble préférable d'utiliser une grille d'évaluation de type Likert. Comme les deux autres travaux référencés dans notre analyse (celui de (Burger & Bresin, 2010) et celui de (Dalh & Friberg, 2007)) utilisaient ce type de grille d'évaluation, il en a résulté une certaine difficulté à comparer nos résultats avec les leurs. Cependant, ce n'est pas un inconvénient majeur car comme nous avons discuté dans la section 3.1 de l'évaluation des expressions faciales du robot EDDIE, un questionnaire à choix multiples donne à peu près la même information qu'un questionnaire utilisant une échelle de type Likert.

Une autre élément à explorer dans les prochaines expérimentations est la distinction dans le rôle des déplacements et le rôle du regard sur l'expressivité du robot. En fait, notre présente expérimentation permet de dire que le regard renforce l'expressivité, mais ne permet pas de préciser à quel niveau, ni d'expliciter quel est son impact. Ce sont les questions auxquelles nous ne savons pas encore répondre et dont la réponse permettra de clarifier si le rôle du regard est vraiment incontournable.

La comparaison de l'expressivité entre le robot LINA et le robot MEX relève aussi des questions intéressantes,. Ces questions sont liées à la différence de taille et de forme de deux robots. Tandis que LINA est grand et plus anthropomorphique, MEX est de petit taille et ressemble plus un animal. Est-ce que la taille et la forme des objets affectent son expressivité émotionnelle ? L'étude sur les mouvements des musiciens (Dalh & Friberg, 2007) et l'étude sur les formes des objets pour l'expression des émotions (Isbister, Köök, Sharp, & Laaksolahti, 2006) ne discutent pas de cet aspect, ce qui reste donc une question à étudier pour le développement pertinent des capacités d'expression des émotions chez les robots.

Chapitre 5

Conclusion et Perspectives

A travers ce mémoire, nous avons abordé trois grands axes de l'étude sur l'interaction homme - machine : la modélisation du processus émotionnel pour les agents intelligents, l'extraction du contenu émotionnel dans la musique et la conception des mouvements émotionnels inspirés des mouvements des musiciens.

Le chapitre sur la modélisation du processus émotionnel montre qu'il est possible d'unifier différents modèles computationnels des émotions en une seule architecture. Notre proposition offre une vue computationnelle globale sur le processus émotionnel et des exemples de mise en œuvre des différents composants. Cela est rendu possible par la décomposition des différents aspects d'un processus émotionnel dans GRACE, comme les réflexes, l'interprétation cognitive, l'impact de l'humeur, l'impact de la personnalité, etc. Ce sont les aspects qui sont listés par les psychologues comme les éléments participant au processus émotionnel. Ce sont aussi les éléments pris en compte par certains modèles computationnels, mais le plus souvent de manière partielle. La prise en compte de ces aspects dans GRACE est nécessaire pour construire une architecture complète de l'émotion en informatique. L'analyse de la conception de GRACE et la comparaison faite pour démontrer sa généricité sont les premières étapes du processus de développement d'un tel modèle. La suite du développement sera l'implémentation de chaque composant et la validation du modèle en fonction des divers scénarios d'interaction.

L'implémentation de chaque composante de GRACE se décompose en deux tâches distinctes : la logique de l'association des valeurs affectives aux événements détectés, et l'implémentation technique de cette logique de l'association. Par exemple, l'implémentation de l'interprétation cognitive a été réalisée avec diverses techniques dans les modèles des émotions existants, comme la logique floue dans FLAME, le raisonnement sémantique dans Greta, l'association des valeurs affectives aux événements dans EMA, etc. L'implémentation de l'interprétation cognitive dans GRACE est donc un choix de technique à utiliser en fonction des compétences techniques. L'association des valeurs affectives aux événements, quant à elle, dépend de la complexité du processus émotionnel à mettre en œuvre. Par exemple, cela peut être une association simple d'un événement à une émotion comme dans le cas de Cathexis, mais cela peut très bien être l'application d'un raisonnement selon le modèle OCC, ou bien l'association d'un événement à plusieurs variables affectives comme suggéré dans EMA.

L'échange de l'information entre des composants dans GRACE est simple grâce à l'unification du flux d'informations échangées entre des composantes. L'utilisation des couples Valence - Activation simplifie l'interaction entre les composantes. En fait, les modèles computationnels des émotions existants sont très divers au niveau des informations échangées entre les composantes dans leurs modèles, ce qui fait qu'il existe de l'incohérence d'un modèle à l'autre comme nous l'avons analysé dans le chapitre 2. L'unification du flux d'information de GRACE enlève cette incohérence. Elle permet, par conséquent, une plus ample compréhension des différents aspects d'un processus émotionnel mis en œuvre dans un modèle computationnel et une plus

simple implémentation d'un tel modèle. L'implémentation d'une composante est bien isolée de celle d'autres composantes, à condition que ses entrées respectent la spécification prédéfinie par l'architecture GRACE.

Dans le cadre de la thèse, cette séparation nous a permis de focaliser notre attention sur le développement de deux composantes dans GRACE, i.e. l'*Interprétation cognitive* et l'*Expression* sans avoir besoin de nous préoccuper du fonctionnement des autres composantes. La seule condition est que les entrées et les sorties des deux composantes respectent la spécification de l'architecture GRACE. Dans le cas de l'*Interprétation cognitive* qui consiste en l'extraction du contenu émotionnel dans la musique, les entrées de la composante sont une séquence d'événements musicaux qui peuvent être décrits par l'intermédiaire d'un ensemble de descripteurs musicaux, et les sorties sont les couples Valence – Activation qui représentent l'émotion véhiculée dans la musique. Dans le cas de l'*Expression*, les entrées sont les couples Valence – Activation représentant l'émotion en réponse à l'événement capturé par la *Sensation*, et les sorties sont les instructions permettant de faire exécuter des mouvements au robot, i.e. la vitesse linéaire et la vitesse angulaire de LINA.

La construction de l'extracteur du contenu musical, à implémenter dans le composant *Interprétation Cognitive* du modèle GRACE, est présentée dans le chapitre 3. Notre analyse sur les extraits musicaux de R. Schumann, à la fois avec un réseau de neurones et avec un arbre de décision suggère que le réseau de neurones est un bon candidat pour l'apprentissage des interprétations émotionnelles sur la musique. Les premiers résultats sur l'extraction de la valence et de l'activation sont encourageants et à exploiter dans l'avenir. Une autre piste de recherche pour l'avenir concerne la nécessité d'une base de données musicale validée. Lors de notre travail, nous avons eu beaucoup de difficultés pour trouver des bases de données publiques pour faire la validation des descripteurs musicaux proposés, et aussi pour valider notre système d'extraction de l'information émotionnelle dans la musique. Les travaux antérieurs sont pour la plupart validés sur leurs propres bases de données, qui sont en général petites en taille et différentes d'un travail à l'autre. Il est donc difficile de comparer leurs résultats et de les ré-utiliser pour un développement ultérieur. Vu le développement croissant de ce domaine, de telles bases de données seront nécessaires pour tous les systèmes d'extraction proposés dans l'avenir.

Toujours dans la perspective d'étudier l'utilisation de l'émotion lors de l'interaction homme-machine, nous avons mené des études sur l'expression émotionnelle pour les robots. Comme cité dans les études en sciences sociales (psychologie, sociologie, philosophie), l'expression des émotions joue un rôle très important dans la vie sociale de l'être humain. Cet élément est aussi important dans l'interaction homme – machine comme présenté dans les expérimentations avec les robots iCat et Kismet et l'agent virtuel animé Greta. Nous avons proposé une conception des mouvements émotionnels inspirée des mouvements des musiciens lors de leurs performances artistiques. L'expérimentation avec les adolescents que nous avons menée lors de la Fête de la Science valide notre conception en montrant un taux de reconnaissance convaincant par les sujets humains. De plus, notre conception montre aussi l'utilité du regard simulé et de la musique pour améliorer l'expressivité émotionnelle du robot. Pour développer davantage cette direction de recherche, quelques pistes pourraient être exploitées : (1) tester ces mouvements sur d'autre types de robots (les robots de plus petite taille, les robot humanoïdes), et les agents virtuels animés ; (2) comparer ces mouvements appliqués soit à un robot soit à un agent virtuel ; (3) analyser la perception des sujets de différentes tranches d'âge pour évaluer l'universalité de ces

mouvements ; (4) aborder l'harmonisation entre la fluidité des mouvements et la dynamique des émotions au cours de l'interaction pour l'intégration de ces mouvements émotionnels dans le modèle des émotions lors de l'interaction avec les humains.

L'implémentation du composant *Interprétation Cognitive* sur la musique et du composant *Expression* sur les mouvements dansants n'est que le début du projet de développement du modèle computationnel GRACE. La validation complète du modèle GRACE ne pourra en effet être obtenue que quand nous disposerons d'une instance fonctionnelle complète de GRACE (i.e. tous les modules de GRACE), ce qui permettra également une diffusion plus large. La validation globale d'un tel modèle pourra se faire via la reprise des scénarios étudiés dans les travaux antérieurs, comme pour FLAME, Greta, EMA, ParleE, etc. Elle pourra se faire également par des comparaisons avec les autres modèles des émotions dit génériques comme le modèle EMA actuellement développé par J. Gratch et ses collègues. Est-ce que ces modèles dit génériques se ressemblent ? Est-ce qu'il vaut mieux avoir une normalisation des modèles des émotions qu'avoir des modèles bien cadrés ? Est-ce que, si on arrive à se mettre en accord sur un seul modèle computationnel des émotions, ce modèle pourrait aider à la recherche de la définition de l'émotion des chercheurs en psychologie ? Ces questions restent à étudier dans une recherche multidisciplinaire à long terme.

Références

Arbib, M. A., & Fellous, J.-M. (2004). Emotions: from brain to robot. *TRENDS in Cognitive Science , 8* (12), 554-561.

Arkin, R. C., Fujura, M., Takagi, T., & Hasegawa, R. (2003). An ethological and emotional basis for human-robot interaction. *Robotics and Autonomous Systems , 42*, 191-201.

Bach, J. (2009). *Principles of Synthetic Intelligence PSI: An Architecture of Motivated Cognition.* New York: Oxford University Press.

Baumgartner, T., Esslen, M., & Jänke, L. (2005). From emotion perception to emotion experience: Emotions evoked by pictures and classical music. *Internaltional Journal of Psychophysiology , 60*, 34-43.

Becker-Asano, C. (2008). *WASABI: Affect Simulation for Agents with Believable Interactivity.* University of Bielefeld, Facuty of Technology. IOS Press.

Bour, G., Hutzler, G., & Gortais, B. (2004). Ambient Cognitive Environments and the Distributed Synthesis of Visual Ambiences. *Engineering Self-Organizing Applications 2004 . 3464*, pp. 69-83. Berlin: Springer Verlag.

Breazeal, C. (2001). Affective Interaction between Humans and Robots. *2001 European Conference on Artificial Life*, (pp. 582-591).

Breazeal, C. (2004). Function Meets Style: Insights From Emotion Theory Applied to HRI. *IEEE Transactions on Systems, Man, and Cybernetics—Part C: Applications and Reviews , 34* (2), 187-194.

Breazeal, C. (2003). Towards social robots. *Robotics and Autonomous Systems , 42*, 167-175.

Breazeal, C., & Scassellati, B. (1999). How to build robots that make friends and influence people. *1999 IEEE/RAS Intervational Conference on Intelligent Robots and Systems*, (pp. 858-863).

Bresin, R., & Friberg, A. (2011). Emotion rendering in music: Range and characteristic values of seven musical variables. *CORTEX , 47* (9), 1068-1081.

Bui, T. D., Heylen, D., Poel, T., & Nijholt, A. (2002). ParleE: An Adaptive Plan Based Event Appraisal Model of Emotions. *25th Annual German Conference on Artificial Intelligence*, (pp. 16-20).

Burger, B., & Bresin, R. (2010). Communication of musical expression by means of mobile robot gestures. *Multimodal User Interfaces , 3*, 109-118.

Camurri, A. (2004). Multimodal Interfaces for Expressive Sound Control. *7th International Conference on Digital Audio Effects.* Naples.

Camurri, A., & Coglio, A. (1998). An Architecture for Emotional Agents. *IEEE MultiMedia , 5* (4), 24-33.

Camurri, A., Coletta, P., Massari, A., Mazzarino, B., Peri, M., Ricchetti, M., et al. (2004). Toward real-time multimodal processing: EyesWeb 4.0. *AISB 2004 Convention: Motion, Emotion and Cognition.* Leeds.

Canamero, L. D., & Fredslund, J. (2000). *How does it feel? Emotional Interaction with a humanoïd LEGO robot.* LEGO-Lab. AAAI.

Dörner, D., & Hille, K. (1995). Artificial Souls: Motivated Emotional Robots. *International Conference on Systems, Man and Cybernetics 1995*, (pp. 3828-3832).

Dalh, S., & Friberg, A. (2007). Visual perception of expressiveness in musicians' body movements. *Music Perception , 24* (5).

Dang, T. H., & Duhaut, D. (2009). Experimentation with GRACE, the Generic Model of Emotions For Computational Applications. *2nd Mediterranean Conference on Intelligent Systems and Automation.*

Dang, T. H., Hutzler, G., & Hoppenot, P. (2011). Emotion modeling for intelligent agents - Towards a unifying framework. *2011 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT)*, (pp. 70-73).

Dang, T. H., Hutzler, G., & Hoppenot, P. (2010). How to gain emotional rewards during human-robot interaction using music? Formulation and propositions. *10th International Conference on Artificial Intelligence and Soft Computing (ICAISC-2010)*, (pp. 247-255).

Dang, T. H., Hutzler, G., & Hoppenot, P. (2011). Mobile Robot Emotion Expression with Motion base on MACE-GRACE Model. *15th International Conference On Advanced Robotics (ICAR 2011)*, (pp. 137-142).

Dang, T. H., Letellier-Zarshenas, S., & Duhaut, D. (2008). Comparison of recent architectures of emotions. *10th International Conference on Control, Automation, Robotics and Vision – ICARCV 2008*, (pp. 1976-1981).

Dang, T. H., Letellier-Zarshenas, S., & Duhaut, D. (2008). Grace – Generic Robotic Architecture To Create Emotions. *11th International Conference on Climbing and Walking Robots and the Support Technologies for Mobile Machines – CLAWAR 2008.*

de Rosis, F., Pelachaud, C. P., Valeria, C., & De Carolis, B. (2003). From Greta's Mind to Her Face: Modeling the Dynamics of Affective States in a Conversational Embodied Agent. *Internaltional Journal of Human-Computer Studies , 59*, 81-118.

Dias, H., Mascarenhas, & Paiva, A. (2011). FAtiMA Modular: Towards an Agent Architecture with a Generic Appraisal Framework. *Workshop 'Standard in Emotion Modeling'.*

Eerola, T., & Toiviainen, P. (2004). Mir in matlab: The midi toolbox. *5th International Conference on Music Information Retrieval*, (pp. 22-27).

Egermann, H., Nagel, F., Altenmüller, E., & Kopiez, R. (2009). Continuous Measurement of Musically-Induced Emotion: A Web Experiment. *International Journal of Internet Science , 4* (1), 4-20.

Ekman, P. (1992). *An Argument for Basic Emotions.* (N. a. Stein, Ed.) Hove, UK: Lawrence Erlbaum.

Ekman, P., Friesen, W. V., & Hager, J. C. (2002). *Facial Action Coding System.* (W. &. Nicolson, Ed.) Research Nexus eBook.

Elliott, C. (1993). Using the Affective Reasoner to support social simulations. *Thirteenth International Joint Conference Artificial Intelligence*, (pp. 194-200). France.

El-Nars, M. S., Yen, J., & Ioerger, T. (2000). FLAME - A Fuzzy Logic Adaptive Model of Emotions. *Journal of Autonomous Agents and Multi-Agent Systems , 3*, 219-257.

Eschrich, S., Münte, T. F., & Altermüller, E. O. (2008). Unforgettable film music: The role of emotion in episodic long-term memory for music. *BMC Neuroscience* .

Eysenck, H. J., & Rachman, S. (1965). *The causes and cures of neurosis.* San Diego: Robert R. Knapp.

Fellous, J.-M. (2004). From human emotions to robot emotions. (E. Hudlicka, Ed.) *Architectures for Modeling Emotion: Cross-disciplinary Foundations* , 37-47.

Fong, T., Nourbakhsh, I., & Dautenhahn, K. (2003). A survey of socially interactive robots. *Robotics and Autonomous Systems , 42*, 143-166.

Forlizzi, J., Disalvo, C., & Gemperle, F. (2004). Assistive robotics and an ecology of elders living independently in their homes. *Human-Computer Interaction , 19* (1), 25-59.

Fornari, J., & Eerola, T. (2009). The Pursuit of Happiness in Music: Retrieving Valence with Contextual Music Descriptors. *International Synposiun on Computer Music Modeling and retrieval. 5493*, pp. 119-133. Springers Berlin/Heigelberg.

Friberg, A. (2004). A fuzzy analyzer of emotional expression in music performance and body motion. In J. Sundberg, & B. Brunson (Ed.), *Music and Music Science 2004.*

Gebhard, P. (2005). ALMA - A Layered Model of Affect. *The 2005 Autonomous Agents and Multiagent Systems Conference* (pp. 29-36). ACM.

Gebhard, P., & Kipp, K. H. (2006). Are computer-generated emotions and moods plausible to humans? *6th International Conference on Intelligent Virtual Agents (IVA'06)*, (pp. 343-356).

Glowinski, D., & Camurri, A. (2010). Musique et Emotions. In C. Pelachaud, *Système d'interaction émotionnelle* (pp. 317-340). Hermès Science Publications Lavoisier.

Gockley, R., Forlizzi, J., & Simmons, R. (2006). Interactions with a moody robot. *1st ACM SIGCHI/SIGART conference on Human-robot interaction*, (pp. 186-193). New York.

Goetz, J. K., & Powers, A. (2003). Matching robot appearance and behavior to tasks to improve human-robot cooperation. *12th IEEE workshop on Robot and Human Interactive Communication, ROMAN 2003* (pp. 55-60). IEEE.

Gratch, J., & Marsella, S. (2004). A Domain-independant Framework for Modeling Emotion. *Journal of Cognitive Systems Research , 5* (4), 269-306.

Gratch, J., & Marsella, S. (2006). Evaluating a computational model of emotion. *Journal of Autonomous Agents and Multiagent Systems (Special issue on the best of AAMAS 2004) , 11* (1), 23-43.

Heerink, M., Krose, B., Evers, V., & Wielinga, B. (2006). The Influence of a Robot's Social Abilities on Acceptance by Elderly Users. *RO-MAN 2006* (pp. 521-526). Hertfordshire.

Hoffman, G., & Weinberg, G. (2010). Gesture-based Human-Robot Jave Improvisation. *IEEE International Conference on Robotics and Automation.*

Hu, X. (2010). Music and Mood: Where Theory and Reality Meet. *5th iConference.* Champaign.

Hunter, P. J., & Schellenberg, E. G. (2010). Feelings and Perceptions of Happiness and Sadness Induced by Music: Similarities, Differences, and Mixed Emotions. *Psychology of Aesthetics, Creativity, and the Arts , 4* (1), 47-56.

Husain, G., Thompson, W. F., & Schellenberg, E. G. (2001). Effects of Musical Tempo and Mode on Arousal, Mood, and Spacial Abilities. *Music Perception: An Interdisciplinary Journal , 20* (2), 151-151.

Isbister, K., Köök, K., Sharp, M., & Laaksolahti, J. (2006). The Sensual Evaluation Instrument: Developing an Affective Evaluation Tool. *Conference on human factors in computing systems - CHI 2006*, (pp. 1163-1172).

Izard, C. E. (1992). Basic emotions, relations among emotions, and emotion-cognition relation. *Psychological Review , 99* (3), 561-565.

Juslin, P. N., & Västfjäll, D. (2008). Emotional Responses to Music: The Need to Consider Underlying Mechanisms. *Behavioural and Brain Sciences , 31* (5), 559-575.

Kiesler, S., & Hinds, P. (2004). Introduction to This Special Issue on Human-Robot Interaction. *Human-Computer Interaction , 19*, 1-8.

Korhonen, M. D. (2004). *Modeling Continuous Emotional Appraisals of Music Using System Identification.* Master Thesis, University of Waterloo, Waterloo, Ontario, Canada.

Korhonen, M. D., & Clausi, D. A. (2006). Modeling Emotional Content of Music Using System Identification. *IEEE Transaction on Systems, Man, and Cybernetics - Part B: Cybernetics , 36* (3), 588-599.

Kuhnlenz, K., Sosnowski, S., & Buss, M. (2007). Evaluating emotion expressing robots in affective space. *Advanced Robotic Systems & I-Tech* .

Lance, B. J., & Marsella, S. C. (2008). A model of gaze for the purpose of emotional expression in virtual embodied agents. *7th international joint conference on Autonimous agents and multiagent systems*, *1*.

Lance, B., & Marsella, S. (2010). Glances, Glares, and Glowering: How should a virtual human express emotion through gaze? *Journal of Autonomous Agents and Multiagent System , 20* (1), 50-69.

Lazarus, R. S. (1993). From psychological stress to the emotions: A history of changing outlooks. *Annual Review of Psychology , 44*, 1-21.

Leman, M., Vermeulen, V., De Voogdt, L., Taelman, J., Moelants, D., & Lesaffre, M. (2004). Correlation of gestural musical audio cues and perceived expressive qualitites. *GW 2003: Internaltional geture workshop N° 15. 2915*, pp. 40-54. Springer Berlin.

Lesaffre, M., Leman, M., De Voogdt, L., De Baets, B., De Meyer, H., & Martens, J.-P. (2006). A user-dependent approach to the perception of high-level semantics of music. *9th International Conference on Music Perception and Cognition*, (pp. 1003-1008).

Lesaffre, M., Leman, M., Tanghe, K., De Baets, B., De Meyer, H., & Martens, J.-P. (2003). User-Dependent Taxonomy of Musical Features as a Conceptual Framework for Musical Audio-Mining Technology. In R. Bresin (Ed.), *Stockholm Music Acoustics Conference*, (pp. 635-638). Stockholm.

Lim, M. Y., Aylett, R., & Jones, C. M. (2005). Emergent Affective and Personality Model. *The 5th International Working Conference on Intelligent Virtual Agents (IVA'05)*, *LNAI 3661*, pp. 371-380.

Lim, M. Y., Dias, J., Ruth, A., & Paiva, A. (2011). Creating Adaptive Affective Autonomous NPCs. *Special Issue Journal of Autonomous Agents and Multiagent Systems* .

Livingstone, S. R., Schubert, E., & Loehr, J. D. (2009). Emotional arousal and the automatic detection of musical phrase boundaries. In A. Williamon, S. Pretty, & R. Buch (Ed.), *International Symposium on Performance Science* (pp. 445-450). Utrecht: The Netherlands: European Association of Conservatoires.

Malfaz, M., & Salichs, M. A. (2004). A new architecture for autonomous robots based on emotions. *5th IFAC Symposium on Intelligent Autonomous Vehicles.* IFAC 2004.

Mancini, M., Bresin, R., & Pelachaud, C. (2005). From acoustic cues to an expressive agent. *Gesture Workshop 2005.* Vannes.

Mancini, M., Varni, G., Kleimola, J., Volpe, G., & Camurri, A. (2010). Human movement expressivity for mobile active music listening. *Journal of Multimodal User Interfaces , 4*, 27-35.

Marsella, S., & Gratch, J. (2009). EMA: A computational model of appraisal dynamics. *Journal of Cognitive Systems Research , 10* (1), 70-90.

Marsella, S., Gratch, J., & Petta, P. (2010). Computational Models of Emotion. (K. R. Scherer, T. Bänziger, & E. Roesch, Eds.) *A blueprint for a affective computing: A sourcebook and manual* .

Martinez-Miranda, J., & Aldea, A. (2005). Emotions in human and artifical intelligence. *Computers in Huma Behavior , 21*, 323-342.

Mascarenhas, S. F., Dias, J., Prada, R., & Paiva, A. (2010). A dimensional Model for cultural behaviour in virtual agents. *Applied Artificial Intelligence , 24* (6), 552-574.

Mayer, J. D., Salovey, P., & Caruso, D. R. (2008). Emotional intelligence: New ability or eclectic traits? *American Psychologist , 63* (6), 503-517.

McCrae, R. R., & John, O. P. (1992). An introduction to the five factor and its applications. *Journal of Peronality , 60*, 171-215.

McKay, C. (2004). *Automatic Genre Classification of MIDI Recordings.* Master Thesis, McGill University, Department of Theory, Faculty of Music, Montreal.

McKay, C., & Fujinaga, I. (2006). jSymbolic: A Feature Extractor for MIDI Files. *2006 International Computer Music Conference*, (pp. 302-305).

Mehrabian, A. (1996). Analysis of the Big-five Personality Factors in Terms of the PAD Temperament Model. *Autralian Journal of Psychology , 48* (2), 86-92.

Metaphysics Research Lab, CSLI, Stanford University. (n.d.). *Emotion*. Retrieved from Stanford Encyclopedia for Philosophy: http://plato.stanford.edu/entries/emotion/

Michalowski, M. P., Sabanovic, S., & Kozima, H. (2007). A Dancing Robot for Rhythmic Social Interaction. *ACM/IEEE international conference on Human-robot interaction (HRI '07)* (pp. 89-96). New York: ACM.

Miwa, H., Itoh, K., Matsumoto, M., Zecca, M., Takanobu, H., Roccella, S., et al. (2004). Effective Emotional Expressions with Emotion Expression Humanoïd Robot WE-4RII. *2004 IEEE/RSJ International Conference on Intelligent Robots and Systems*, (pp. 2203-2208).

Murata, K., Nakadai, K., Takeda, R., Okuno, H. G., Torii, T., Hasegawa, Y., et al. (2008). A Beat-Tracking Robot for Human-Robot Interaction and Its Evaluation. *2008 8th IEEE-RAS International Conference on Humanoid Robots* (pp. 79-84). IEEE.

Nantais, K. M., & Schellenberg, E. G. (1999). The Mozard Effect: An Artifact of Preference. *Psychological Science , 10* (4), 370-373.

Oliveira, A. P., & Cardoso, A. (2008). Modeling Affective Content of Music: A Knowledge Base Approach. *Sound and Music Computing Conference.*

Oliveira, A. P., & Cardoso, A. (2008). Towards Bi-dimensional Classification of Symbolic Music by Affective Content. *2008 International Comuter Music Conference.*

Ortony, A., & Turner, T. J. (1990). What's Basic About Basic Emotions? *Psychological Review , 97* (3), 315-331.

Ortony, A., Clore, G., & Collins, A. (1988). *The cognitive structure of emotions.* Cambridge: Cambridge University Press.

Paiva, A., Dias, J., Sobral, D., Aylett, R., Woods, S., Hall, L., et al. (2005). Learning by Feeling: Evoking Empathy with Synthetic Characters. *Applied Artificial Intelligence , 19* (3), 235-266.

Pelachaud, C. (2009). Studies on Gesture Expressivity for a Virtual Agent. *Speech Communication , 51* (7), 630-639.

Pelachaud, C. (2010). *Système d'interaction émotionnelle.* Hermes Science Publications Lavoisier.

Pelachaud, C., & Poggi, I. (2002). Multimodal Embodied Agents. *Workshop on Multimodal Communication and Context in Embodied Agents.*

Pereira, G., Dimas, J., Prada, R., Santos, P. A., & Paiva, A. (2011). A Generic Emotional Contagion Computational Model. *Affective Computing and Intelligent Interaction 2011. LNCS 6974*, pp. 256-266. Springer-Verlag Berlin Heidelberg.

Rao, A. S., & Georgeff, M. P. (1991). Modeling Rational Agents within a BDI-Architecture. *In Proceedings of the 2nd International Conference on Principles of Knowledge Representation and Reasoning*, (pp. 473-484).

Rentfrow, P. J., & Gosling, S. D. (2003). The Do Re Mi's Everyday Life: The structure and Personality Correlates of Music Preferences. *Journal of Personality and Social Psychology , 84*, 1236-1256.

Rodrigues, S. H., Mascarenhas, S. F., Dias, J., & Paiva, A. (2009). I can feel it too! Emergent empathetic reactions between synthetic characters. *The 3rd International Conference on Affective Computing and Intelligent Interaction* (pp. 1-7). IEEE.

Rousseau, D. (1996). Personality in Computer Characters. *1996 AAAI Workshop on Entertainment and AI/A-Life* (pp. 38-43). Portland, Oregon: AAAI Press.

Russell, J. A. (1980). A Circumplex Model of Affect. *Journal of Personality and Social Psychology , 39* (6), 1161-1178.

Russell, J. A. (2003). Core Affect and the Psychological Construction of Emotion. *Psychological Review , 110* (1), 145-172.

Russell, J. A., Weiss, A., & Mendelsohn, G. A. (1989). Affective Grid: A Single-Item Scale of Pleasure and Arousal. *Journal of Personality and Social Psychology , 57* (3), 493-502.

Ruth, A., Vannini, N., Andre, E., Paiva, A., Enz, S., & Hall, L. (2009). But that was in another country: agents and intercultural empathy. *The 8th International Conference on Autonomous Agents and Multiagent Systems*, (pp. 329-336).

Sarmento, L., Moura, D., & Oliveira, E. (2004). Fighting fire with fear. *2nd European Workshop on Multi-Agent Systems.*

Scheirer, J., & Picard, R. W. (1999). *Affective Objects.* Technical Report, MIT Media Laboratory Perceptual Computing Section.

Scherer, K. R. (2001). Appraisal considered as a process of multi-level sequential checking. In K. R. Scherer, A. Schorr, T. Johnstone, K. R. Scherer, A. Schorr, & T.

Johnstone (Eds.), *Appraisal process in emotion: Theory, MEthods, Research* (pp. 92-120). New York and Oxford: Oxford University Press.

Scherer, K. R. (1999). Appraisal theory. In T. Dalgleish, M. Power, T. Dalgleish, & M. Power (Eds.), *Handbook of cognition and emotion* (pp. 637-663). Chichester: Wiley.

Scherer, K. R. (2009). Emotions are emergent processes: They requite a dynamic computational architecture. *Philosophical Transactions of the Royal Society, series B , 364*, 3459-3474.

Scherer, K. R. (2009). Emotions are emerging processes: They require a dynamic computational architecture. *Philosophical Transactions of the Royal Society , 364* (B), 3459-3474.

Scherer, K. R. (1984). On the nature and Function of Emotion: A component Process Approach. (K. R. Scherer, & P. Ekman, Eds.) *Approaches to Emotion* , 293-317.

Scherer, K. R. (2005). What are emotions? How can they be measured? *Social Science Information , 44* (4), 695-729.

Scherer, K. R. (2004). Which emotions can be induced by music? What are the underlying Mechanisms? And how can we measure them? *Journal of New Music research , 33* (3), 239-251.

Scherer, K. R., & Zentner, M. R. (2001). Emotional effects of music: Production rules. In P. N. Juslin, J. A. Sloboda, P. N. Juslin, & J. A. Sloboda (Eds.), *Music and emotion: Theory and research* (pp. 361-392). Oxford: Oxford University Press.

Scherer, K. R., Johnstone, T., & Sangsue, J. (1998). L'état émotionnel du locuteur: facteur négligé mais non négligeable pour la technologie de la parole. *Actes des XXIIèmes Journées d'Etudes sur la Parole.* Martigny, Swizerland.

Schubert, E. (1999). *Measurement and Time Series Analysis of Emotion in Music.* PhD Thesis, University of New South Wales, New South Wales.

Schubert, E., & Dunsmuir, W. T. (2004). Introduction to interrupted time series analysis of emotion in music: the case of arousal, valence and points of rest. *8th International Conference on Music Perception and Cognition*, (pp. 445-448).

Smith, C. A., & Lazarus, R. S. (1990). Emotion and adaptation. In L. Pervin (Ed.), *Handbook of Personality: Theory and Research* (pp. 609-637). New York: Guilford.

Sosnowski, S., Bittermann, A. K., & Buss, M. (2006). Design and Evaluation of Emotion-Display EDDIE. *2006 IEEE/RSJ International Conference on Intelligent Robots and Systems* (pp. 3113-3118). IEEE.

Suzuki, K., Ohashi, T., & Hashimoto, S. (1999). Interactive Multimodal Mobile Robot for Musical Performance. *1999 International Computer Music Conference*, (pp. 407-410).

Svetlicic, I. (2004). *robotics, Emotion based subsumption architecture for autonomous mobile.* Master Thesis, Miami University, Computer Science and Systems Analysis, Oxford.

Tapus, A., Tapus, C., & Matarié, M. J. (2008). User-Robot Personality Matching and Robot Behavior Adaptation for Post-Stroke Rehabilitation Therapy. *Intelligent Service Robotics, Special Issue on Multidisciplinary Collaboration for Socially Assistive Robotics* , 169-183.

Timmers, R., Camurri, A., & Volpe, G. (2003). Performance cues for listeners' emotional engagement. *Stokholm Music Acoustics Conference.*

Velasquez, J. D. (1998). A computational Framework for Emotion-Based Control. *Workshop on Grounding Emotions in Adaptive Systems, International Conference on SAB.*

Velasquez, J. D. (1997). Modeling emotions and other motivations in synthetic agents. *Fourteenth National Conference on Artificial Intelligence* (pp. 10-15). Rhode Island: The MIT Press.

Wada, K., & Shibata, T. (2007). Social and Physiological Influences of Living with Seal Robots in an Elderly Care House for Two Months. *2007 IEEE International Conference on Robotics and Automation*, (pp. 1250-1255). Roma.

Wada, K., & Shibata, T. (2009). Social Effects of Robot Therapy in a Care House – Change of Social Network of the Residents for One Year –. *Journal of Advanced Computational Intelligence and Intelligent Informatics , 13* (4), 386-392.

Weinberg, G., & Driscoll, S. (2006). Toward Robotic Musicianship. *Computer Music Journal , 30* (4), 28-45.

Weinberg, G., Blosser, B., Mallikarjuna, T., & Raman, A. (2009). The creation of multi-human, multi-robots interactive jam section. *International Conference on New Instruments for Music Expression*, (pp. 70-73). Pittsburgh.

Wiering, F. (2007). Can human benifit from music information retrieval. In AMR'06 (Ed.), *4th International Conference on Adaptive Multimedia Retrieval: User, Context, and Feedback.* Springer-Verlag.

Yang, Y.-H., Lin, Y.-C., Cheng, H.-T., & Chen, H. (2008). Mr. Emo: Music Retrieval in the Emotion Plane. *16th ACM International Conference on Multimedia*, (pp. 13-21).

Zentner, M., Grandjean, D., & Scherer, K. R. (2008). Emotions Evoked by the Sound of Music: Characterization, Classification, and Measurement. *Emotion , 8* (4), 494-521.

Annexe

1. Matériels pour l'expérimentation lors de la Fête de la Science 2010

1.1. La grille d'évaluation

Questionnaire pour le stand « Le robot danseur »

Numéro de la séance :

Noter : à chaque Condition, vous avez le droit de cocher plusieurs fois la même émotion.

Condition 1 : Robot – Musique – Robot + Musique

Séquence 1 : ☐ Joie ☐ Tristesse ☐ Colère ☐ Sérénité

Séquence 2 : ☐ Joie ☐ Tristesse ☐ Colère ☐ Sérénité

Séquence 3 : ☐ Joie ☐ Tristesse ☐ Colère ☐ Sérénité

Séquence 4 : ☐ Joie ☐ Tristesse ☐ Colère ☐ Sérénité

Condition 2 : Robot – Musique – Robot + Musique

Séquence 1 : ☐ Joie ☐ Tristesse ☐ Colère ☐ Sérénité

Séquence 2 : ☐ Joie ☐ Tristesse ☐ Colère ☐ Sérénité

Séquence 3 : ☐ Joie ☐ Tristesse ☐ Colère ☐ Sérénité

Séquence 4 : ☐ Joie ☐ Tristesse ☐ Colère ☐ Sérénité

Condition 3 : Robot – Musique – Robot + Musique

Séquence 1 : ☐ Joie ☐ Tristesse ☐ Colère ☐ Sérénité

Séquence 2 : ☐ Joie ☐ Tristesse ☐ Colère ☐ Sérénité

Séquence 3 : ☐ Joie ☐ Tristesse ☐ Colère ☐ Sérénité

Séquence 4 : ☐ Joie ☐ Tristesse ☐ Colère ☐ Sérénité

1.2. La liste des extraits musicaux

Groupe	Extrait	Emotion exprimée selon l'annotation du musicologue
Groupe 1	Knecht Ruprecht, A minor (00 :38 – 01 :08)	Joie
	Petite fugue (00 :00 – 00 :29)	Tristesse
	Knecht Ruprecht, A minor (00 :00 – 00 :30)	Colère
	Melody (00 :00 -00:30)	Sérénité
Groupe 2	Moonlight (00 :00 – 00 :31)	Joie
	Folksong (00 :00 – 00 :24)	Tristesse
	Kreisleriana Op.16 No 3 (00 :00 – 00 :32)	Colère
	Litle Piece (00 :00 – 00 :29)	Sérénité

1.3. Données sur les participants

Le tableau ci-dessus contient le nombre des participants de chaque groupe d'élèves. La colonne **Ordre de conditions** explique l'ordre des conditions appliquées pour chaque groupe. La lettre C vaut pour *Combinaison Robot Plus Musique*, la lettre M vaut pour *Musique Seule*, la lettre R vaut pour *Robot Seul*. Un ordre des conditions MRC veut dire que le groupe est présenté d'abord pour la condition Musique Seule, puis la condition Robot Seul, puis la condition Robot Plus Musique.

No	Scenario	Ordre de conditions	Nombre de Participants
G1	S1	MRC	18
G2	S2	CMR	12
G3	S3	CRM	20
G4	S4	RCM	18
G5	S5	MRC	13
G6	S6	CMR	16
G7	S7	CRM	16
G8	S8	RCM	15
G9	S1	MRC	17
G10	S2	CMR	16
Total			**161**

Le résultat d'évaluation en nombre de réponse est présenté dans les tableaux suivants :

Table 18 Résultat pour la condition 'Musique Seule'

No	Scénario	Ordre de conditions	Nb de Participants	Taux de bonne reconnaissance pour Musique Seule			
				J	T	C	S
G1	S1	MRC	18	11	18	18	15
G2	S2	CMR	12	5	12	7	6
G3	S3	CRM	20	16	20	15	17
G4	S4	RCM	18	12	18	12	11
G5	S5	MRC	13	11	13	13	11
G6	S6	CMR	16	15	14	15	14
G7	S7	CRM	16	14	16	13	13
G8	S8	RCM	15	1	13	0	8
G9	S1	MRC	17	4	16	15	4
G10	S2	CMR	16	12	11	13	7
Total			**161**	**101**	**151**	**121**	**106**

Table 19 Résultat pour la condition 'Robot Seul'

No	Scénario	Ordre de conditions	Nb de Participants	Taux de bonne reconnaissance pour Robot Seul			
				J	T	C	S
G1	S1	MRC	18	6	14	14	13
G2	S2	CMR	12	4	3	6	5
G3	S3	CRM	20	14	14	14	12
G4	S4	RCM	18	12	14	16	14
G5	S5	MRC	13	1	13	0	7
G6	S6	CMR	16	12	15	13	12
G7	S7	CRM	16	8	15	14	10
G8	S8	RCM	15	7	7	9	9
G9	S1	MRC	17	10	16	13	11
G10	S2	CMR	16	7	11	8	7
Total			**161**	**81**	**122**	**107**	**100**

Table 20 Résultat pour la condition ‘Robot Plus Musique’

No	Emotions en concordance	Nb de Participants	Concordance				Discordance			
			J	T	C	S	J	T	C	S
G1	JT	18	3	14	0	0	0	0	0	0
G2	CS	12	0	0	8	6	0	0	0	0
G3	JT	20	14	20	0	0	0	0	1	0
G4	CS	18	0	0	12	15	2	5	0	0
G5	JT	13	5	13	0	0	0	0	0	2
G6	CS	16	0	0	13	8	1	1	0	0
G7	JT	16	9	16	0	0	0	0	3	2
G8	JTCS	15	9	14	9	12	0	0	0	0
G9	JT	17	9	15	0	0	0	0	1	0
G10	CS	16	0	0	11	9	2	5	0	0
Total		**161**	**49**	**92**	**53**	**50**	**5**	**11**	**5**	**4**
Nombre de participants selon les cas			**99**	**99**	**77**	**77**	**62**	**62**	**84**	**84**

1.4. Formules pour calculer les statistiques des matrices de confusion

Supposons que l'on a une matrice de confusion comme suivante :

Table 21 Exemple de matrice de confusion

Emotion		Reconnaissance	
		Positive	Négative
Expression	Positive	a	b
	Négative	c	d

Les statistiques sur cette matrice sont calculées comme suivante :

$$Accuracy(AC) = \frac{a+d}{a+b+c+d}$$

$$True\_positive\_rate(TP) = \frac{a}{a+b}$$

$$False\_positive\_rate(FP) = \frac{c}{c+d}$$

$$True\_negative\_rate(TN) = \frac{d}{c+d}$$

$$False\_negative\_rate(FN) = \frac{b}{a+b}$$

$$\Pr ecision(P) = \frac{a}{a+c}$$

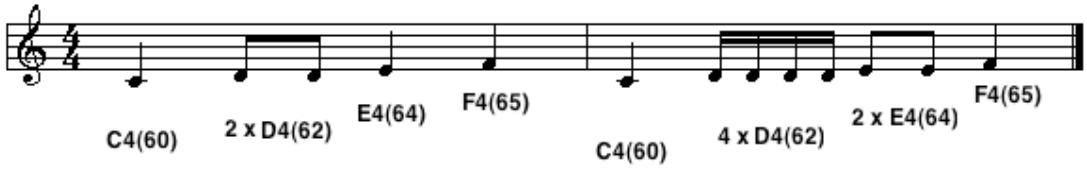
2. Calcul des descripteurs musicaux à partir des messages MIDI

Les messages MIDI de notes vont du C1 (note 0, située 5 octaves en dessous du C5 situé sous la partition en clé de sol, soit 8,175 Hz) au G10 (note 127, soit 5 octaves au-dessus du sol moyen soit 12 557 Hz) avec une résolution d'un 1/2 ton. Ci-dessous est un tableau des notes MIDI complet :

Octave #	Note Numbers											
	C	C#	D	D#	E	F	F#	G	G#	A	A#	B
-1	0	1	2	3	4	5	6	7	8	9	10	11
0	12	13	14	15	16	17	18	19	20	21	22	23
1	24	25	26	27	28	29	30	31	32	33	34	35
2	36	37	38	39	40	41	42	43	44	45	46	47
3	48	49	50	51	52	53	54	55	56	57	58	59
4	60	61	62	63	64	65	66	67	68	69	70	71
5	72	73	74	75	76	77	78	79	80	81	82	83
6	84	85	86	87	88	89	90	91	92	93	94	95
7	96	97	98	99	100	101	102	103	104	105	106	107
8	108	109	110	111	112	113	114	115	116	117	118	119

Le codage MIDI nous permet de facilement calculer les descripteurs musicaux de bas niveau sans passer par le traitement du signal sonore. A partir des messages MIDI, nous calculons les 8 descripteurs de bas niveau : nombre de notes, nombre d'impulsions, fréquence, volume, écart-type de la fréquence, écart-type du volume, durée des notes jouées, et valeur affective. Les descripteurs sont explicités dans la section 3.2.1.

Prenons l'exemple d'un extrait musical simple suivant :



La signature du temps est 4/4, chaque mesure contient 4 temps. Le tempo de l'extrait détermine la durée d'un temps. Selon le codage MIDI, les notes jouées sont dans la quatrième octave, de C4 (60) à F4(65). Supposons que le volume de chaque note est C4 (40), D4 (50), E4 (60), F4 (70) en codage MIDI. Considérons que le tempo de l'extrait est 60. Donc chaque mesure dure 4 secondes. Les valeurs des descripteurs pour la première mesure sur la fenêtre d'une seconde vont être comme suit :

Table 22 Exemple - Valeurs des descripteurs pour la première mesure, l'intervalle d'une seconde

Descripteur	**Seconde 1**	**Seconde 2**	**Seconde 3**	**Seconde 4**
Nombre des notes	1	2	1	1
Nombre d'impulsion	1	2	1	1
Valeur affective	0	0	0	0
Fréquence moyenne	60	62	64	65
Ecart-type de fréquence	0	0	0	0
Volume moyen	40	50	60	70
Ecart-type de volume	0	0	0	0
Durée des notes	100	50	100	100

Maintenant supposons que l'extrait est joué au tempo de 120, ce qui veut dire que chaque mesure dure 2 secondes. Considérons la première mesure : sur une fenêtre d'une seconde, les descripteurs sont calculés comme suivant :

Table 23 Exemple - Valeurs des descripteurs pour l'intervalle de 2 secondes

Descripteur	Seconde 1	Seconde 2
Nombre des notes	3	2
Nombre d'impulsion	3	2
Valeur affective	0	0
Fréquence moyenne	61.33	64.5
Ecart-type de fréquence	1.15470	0.707106
Volume moyen	46.66666	65
Ecart-type de volume	5.773502	7.071067
Durée des notes	33.33333	50

Table 24 Valeur émotionnelle des accords

Accord	Valeur affective
Mineur	-1
Majeur	1
5ème diminué	-2
5ème augmenté	0
7ème majeur	1
7ème mineur	-1
7ème dominante	0
7ème diminué	-2
9ème dominante majeur	1
9th dominante mineur	-1
5ème 7ème diminué	0
Non reconnu	0

3. Liste des morceaux musicaux utilisés

La liste des morceaux est présentée dans le tableau suivant :

Table 25 Liste des morceaux utilisés pour l'extraction du contenu émotionnel dans la musique

No	**Nom du morceau**	**Alias**	**Durée (en second)**
1	Theme etude symphonique	Th	86
2	Album für die Jungend-a	Al-a	63
3	Album für die Jungend-b	Al-b	49
4	Album für die Jungend-c	Al-c	60
5	Album für die Jungend-d	Al-d	82
6	Album für die Jungend-e	Al-e	61
7	Album für die Jungend-f	Al-f	56
8	Album für die Jungend-g	Al-g	65
9	Album für die Jungend-i	Al-i	73
10	Album für die Jungend-l	Al-l	108
11	Kreis-c	Kr	255
12	Papillons. Op.2	Pa	126
13	Petite fugue	Pe	132
14	Schumann_68311a_kreisleriana_(nc)smythe	Sc	119
15	Toccata.Op7	To	61
16	Mondnacht	Mo	62
17	Arabesque	Ar	380
18	Dans la nuit	Da	263
19	Fantaisie	Fa	338
20	Kinderzsenen	Ki	351
21	Pianoconcerto	Pi	797
Total			3587