



Ingénieur 3ème année, Spécialisation Robotique et Apprentissage

Projet Semestriel : État de l'art

travail réalisé par

**DURAND Clément
GARCIA Amandine
JAFFRÉ Anaelle**

Rapport de projet : Génération de mouvements expressifs sur Reachy Mini

Année universitaire 2025 - 2026

Encadrement

M. Vincent Padois
M. Rémi Fabre

Directeur de Recherche
Ingénieur roboticien

Résumé

L'efficacité d'une interaction homme-robot dépend à la fois de la capacité du robot à répondre avec justesse aux besoins exprimés par l'utilisateur et à la fois de ses compétences sociales perçues. Ce travail constitue un état de l'art servant d'étape préparatoire à l'enrichissement du Reachy Mini. Il s'agit d'un robot assistant de bureau, en capacités de génération de mouvements expressifs et de sons. Cela lui permet de communiquer des états affectifs de manière crédible et adaptée au contexte. Dans le cadre de cet état de l'art, des modèles d'expressivité sont explorés, dans le but de servir d'espace latent où les états émotionnels peuvent être représentés. L'analyse portent ensuite sur des architectures de génération et d'adaptation de trajectoires, qui permettent de traduire les variables d'états émotifs en mouvements adaptés aux contraintes morphologiques du robot. Parallèlement, la production sonore, qui inclut parole, onomatopées, éléments paralinguistiques et musique, est analysée pour renforcer la personnalité et l'émotion perçue du robot. Cette approche permet de penser un robot qui réagit de façon dynamique, cohérente et expressive, et qui améliore ainsi l'engagement et la fluidité de l'interaction sociale.

Mots-clés : *robotique sociale, Reachy Mini, émotions, mouvements expressifs, génération sonore, interaction homme-robot, reciblage, expressivité, synthèse vocale, adaptation morphologique*

Introduction

Dans de nombreux secteurs comme l'enseignement, l'assistance ou l'accompagnement, l'efficacité d'une interaction dépend autant de la machine que de sa capacité à être "humaine" aux yeux de l'utilisateur. Les compétences sociales des robots semblent favoriser l'engagement, la confiance et la fluidité interactionnelle. Ce sont les gestes, les expressions, et les réponses adaptées qui permettent à l'humain de se projeter, de comprendre et d'anticiper plus facilement le comportement de l'interface. En rendant l'interaction plus naturelle et intuitive, ces capacités sociales transforment un simple outil en véritable partenaire. Elles contribuent également à réduire la charge cognitive, soutenir l'attention, améliorer la qualité des échanges et faciliter l'acceptation de l'agent par l'utilisateur.

Les états expressifs jouent un rôle central dans la communication des animaux sociaux et constituent un vecteur puissant pour enrichir les interactions avec un robot. L'expression émotionnelle ne se limite pas aux gestes et à la posture, elle inclut également la voix, les onomatopées, les éléments para linguistiques et même la musique, qui peuvent renforcer la personnalité perçue du robot et rendre ses réactions plus crédibles et engageantes. La capacité d'un robot à traduire efficacement un état affectif en mouvement et en son est donc un enjeu majeur pour la robotique sociale.

Le Reachy Mini est un robot assistant de bureau conçu pour converser avec l'utilisateur tout en étant incarné physiquement. Il peut exprimer des émotions via des sons et des mouvements grâce à son corps minimal : un tronc, une tête et deux antennes. Il constitue ainsi une plateforme concrète pour explorer les interactions humain-robot enrichies par des compétences sociales. Le Reachy Mini est déjà capable de générer des mouvements expressifs, mais ceux-ci manquent de spontanéité et ne sont pas toujours pertinents par rapport au contexte. Les réponses émotionnelles et gestuelles restent prédéfinies et statiques.

Il existe des travaux assez diverses dans le domaine, qui présentent une modélisation de mouvements expressifs chez des robots à faible degrés de liberté. Parmi les études les plus pertinentes, on trouve notamment celles d'Osorio *et al.* et de Bretan *et al.*, qui ont intégré l'émotion dans la gestuelle de robots non-humanoïdes avec succès [3, 21].

Cependant, leur méthode n'est pas forcément adaptée à un robot comme le Reachy Mini : Osorio *et al.* proposent une méthode efficace mais nécessitant une trajectoire fonctionnelle prédéfinie. Le Reachy Mini est destiné à rester posé à un endroit fixe et ne peut effectuer de tâche physique fonctionnelle particulière. Quant-à Bretan *et al.*, leur méthodologie implique une cartographie et une définition de chaque pose possible à la main, elle est non-évolutive et ne permet pas la créativité.

C'est ici que le présent état de l'art prend place. Il vise à fournir les connaissances nécessaires, en s'aidant des travaux existants qui traitent, au moins en parti, du problème posé, pour proposer une solution adaptée.

Table des matières

| | |
|--------------------------------------------------------------------------------|-----------|
| Introduction | v |
| Table des figures | ix |
| 1 Modélisation et évaluation des émotions | 1 |
| 1.1 Modèles de classification | 1 |
| 1.1.1 La roue des émotions de Plutchik | 1 |
| 1.1.2 Les six émotions d'Ekman | 2 |
| 1.1.3 Le modèle circonplexe de Russell | 3 |
| 1.1.4 Modèle PAD ou Valence-Activation-Domination | 3 |
| 1.1.5 Théorie des émotions TDRL (Time Difference Recurrent Learning) | 4 |
| 1.2 Détection et évaluation des émotions | 5 |
| 1.2.1 Limites de ces modèles de détection et de classification | 5 |
| 1.3 Optimisation des réponses sociales du robot | 6 |
| 1.3.1 Intelligence sociale perçue | 6 |
| 1.3.2 Présence sociale du robot | 6 |
| 2 Modélisation d'un mouvement expressif | 7 |
| 2.1 Transfert de l'émotion vers le mouvement | 7 |
| 2.1.1 Modèles de mouvements | 7 |
| 2.1.1.1 Laban | 8 |
| 2.1.1.2 Ekman et Friesen | 9 |
| 2.1.2 Architectures de transfert | 10 |
| 2.1.2.1 Génération de mouvements | 10 |
| 2.1.2.2 Adaptation d'un mouvement | 10 |
| 2.2 Crédibilité du mouvement | 11 |
| 2.2.1 Mise en valeur de l'expressivité dans le mouvement | 12 |
| 2.2.2 Adaptation du mouvement à une morphologie réduite | 12 |
| 3 Utilisation et génération du son pour la robotique sociale | 13 |
| 3.1 Caractérisation du son dans la robotique sociale | 13 |
| 3.1.1 Différents domaines d'utilisation | 13 |
| 3.1.1.1 Utilisation fonctionnelle du son | 13 |
| 3.1.1.2 Langage et expression orale | 14 |
| 3.1.1.3 Musique | 14 |
| 3.1.2 Paramètres du robot influençant l'émotion que le son produit | 14 |
| 3.2 Interprétation sonore de l'émotion | 15 |
| 3.2.1 Corrélation entre le son et l'émotion ressentie | 15 |
| 3.2.2 Production de son | 16 |
| 3.2.2.1 Les biais | 16 |

| | |
|-----------------------------------------------------------|-----------|
| 3.2.2.2 Générations connues et expérimentations | 16 |
| Conclusion & perspectives | 19 |
| Bibliographie | 24 |

Table des figures

| | | |
|-----|---------------------------------------------------------------------------------------------------------------------------------------------------|----|
| 1.1 | Roue des émotions de Plutchik | 2 |
| 1.2 | Modèle Valence-Activation de Russell | 3 |
| 1.3 | Illustration de la théorie TDRL des émotions, tirée de <i>A generic self-learning emotional framework for machines</i> [11] | 4 |
| 2.1 | Graphe d'effort de Laban | 8 |
| 3.1 | Catégorisation haut-niveau du son dans les interactions Homme-Robot. Source : [28] | 13 |
| 3.2 | Architecture du système de mémorisation de l'étude [22] | 14 |
| 3.3 | Évaluations de la taille et de la rapidité du robot, notées sur une échelle de 1 à 7, à partir d'un enregistrement audio. Source : [20] | 15 |
| 3.4 | Résultats d'analyse de cinq intentions et deux emotions dans les films "Star Wars" et "Wall-E". Source : [13] | 17 |

1 Modélisation et évaluation des émotions

Dans une interaction sociale, le locuteur adapte sa réponse émotionnelle en fonction de l'émotion perçue chez l'interlocuteur. Afin de recréer chez un système robotique ce comportement qui participe intégralement d'une communication réussie, il est nécessaire d'avoir un modèle qui rend le mieux compte, de façon exhaustive et significative, du panel d'émotions que l'on peut rencontrer chez l'être humain. C'est à partir d'un tel modèle qu'il est possible de développer des méthodes de détection et/ou de classification d'émotions de l'humain et également de proposer une réponse adaptée, qui correspond à une des émotions du modèle.

1.1 Modèles de classification

Il existe plusieurs manières de modéliser ou classifier les émotions selon l'objectif et la discipline concernée par la modélisation. La notion même d'émotion est difficile à circonscrire. Il s'agit en fait d'un concept multidimensionnel. L'un des premiers à s'être intéressé à cette question est Charles Darwin, dans son livre *L'Expression des émotions chez l'homme et les animaux*, paru en 1872 [8]. Il y démontre la fonction évolutive des émotions. Il n'est pas possible de développer ici une liste exhaustive des modèles qui ont vu le jour par la suite, mais la théorie des émotions a été alimentée par la Psychologie, la Sociologie, les Sciences cognitives ainsi que par des réflexions liées à la psychanalyse de Freud [26]. La plupart de ces approches gardent comme point de départ que les émotions sont fonctionnelles, et donc qu'elles se manifestent dans le comportement et dans les réponses physiologiques des animaux. Seuls les modèles qui rendent compte des émotions utiles à la communication dans un contexte d'interaction sociale ont été retenus. Ainsi le modèle philosophique de Schopenhauer qui veut que l'être humain oscille entre trois états : l'ennui, le désir et la souffrance a par exemple été écarté. De même, les modèles plus scientifique mais qui ne s'intéressent qu'à un contexte spécifique d'interactions n'ont pas été retenus. Par exemple, d'un point de vue académique, les seules émotions utiles à classifier sont l'ennui, la confusion, l'engagement et la frustration [18], ce qui est trop réducteur.

1.1.1 La roue des émotions de Plutchik

La théorie psycho évolutionnaire des émotions de Plutchik se base sur 10 principes, parmi lesquels : il n'existe qu'un nombre restreint d'émotions primaires et toutes les autres émotions sont des mélanges ou des dérivées de ces émotions primaires. Les émotions primaires peuvent être conceptualisées comme des paires opposées symétriquement. Chaque émotion peut varier en degré d'intensité ou en niveau d'activation. Plutchik a

identifié 8 émotions primaires : la joie et la tristesse, l'attriance et le dégoût, l'anticipation et la surprise, et enfin la peur et la colère [25]. La figure Fig. 1.1 montre une représen-

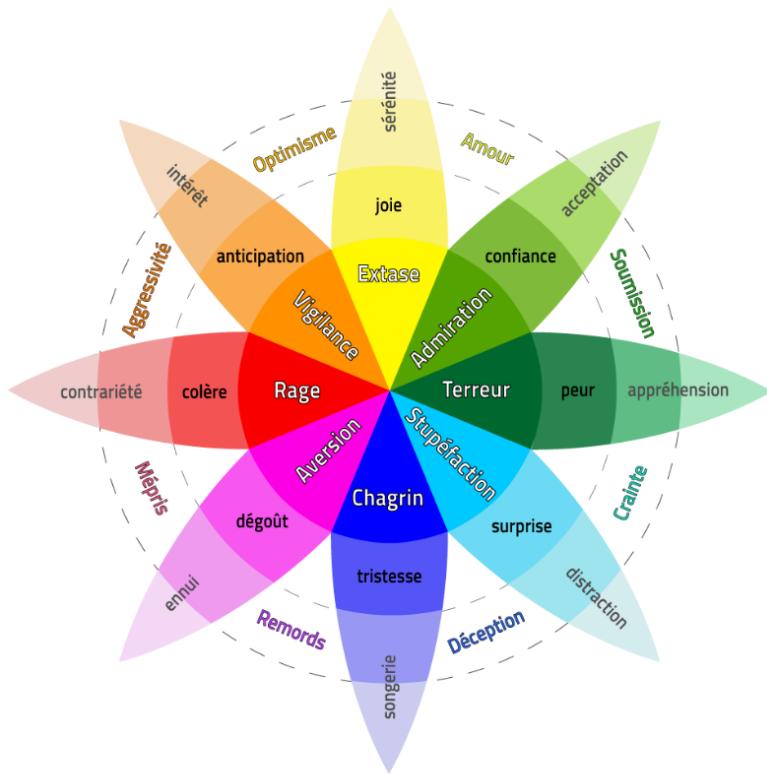


FIGURE 1.1 – Roue des émotions de Plutchik

tation 2D du modèle de Plutchik. À noter qu'il existe également une représentation 3D de cette roue, en forme de cône, qui permet de modéliser le niveau d'activation d'une émotion : plus nous sommes à la base du cône, plus nous sommes neutre, plus l'intensité est élevée, plus le cône s'élargit et plus l'émotion ressentie est forte.

1.1.2 Les six émotions d'Ekman

Ekman propose, dans les années 70, un modèle qui réduit le nombre d'émotion primaires à six. Pour différencier les émotions primaires il se base sur neuf critères parmi lesquels : il y a des signaux universels associés à chacune d'elles en terme d'expressions faciales qui les distinguent, ainsi que des réponses physiologiques différentes. Elles peuvent être rattachées à des causes universelles. Elles se déclenchent rapidement, et ne durent pas longtemps [12]. Ekman laisse ainsi de coté l'attriance et l'anticipation car ces émotions ne correspondent pas aux critère de durée et peuvent être dérivées des autres émotions primaires.

1.1.3 Le modèle circonplexe de Russell

Le quadrant Valence-Activation est d'abord apparu avec le modèle proposé par Russell, dans *Affective space is bipolar*, en 1979 [29]. Il y met en évidence l'aspect dichotomique des émotions dès lors qu'on essaye de les classer. Russell ira plus loin dans *A circumplex model of affect*, en 1980, [1] où il mettra au point ce quadrant qui représente 28 émotions. L'axe vertical montre le niveau d'éveil/d'activation tandis que l'axe horizontal détermine la valence, qu'elle soit positive ou négative. La Fig. 1.2, montre que les 28 émotions sont

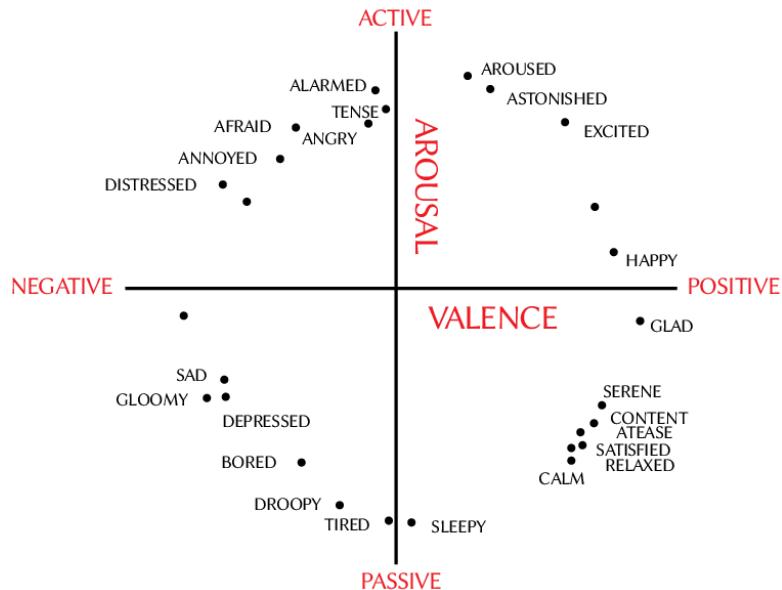


FIGURE 1.2 – Modèle Valence-Activation de Russell

placées plus ou moins près du point d'origine du cadran et qu'elles forment un cercle, à l'origine du nom que l'on attribue également à ce modèle de modèle circonplexe ou circulaire. La position exacte de chaque émotion se décide selon la manière dont les données sont récoltées. Par exemple, il est possible de se baser sur des données déclaratives, avec tout ce que cela implique de subjectif. Chaque personne obtient donc un cercle un peu différent. En 1977, Russell a également travaillé sur la dimension soumission/dominance, qui peut en partie causer la variance inter-individuelle, et qui rentrerait dans une troisième dimension [30].

1.1.4 Modèle PAD ou Valence-Activation-Domination

C'est de cette étude que va naître le modèle PAD qui est donc tridimensionnel. Ce modèle permet d'appuyer sur l'aspect situé socialement des interactions. L'échelle de dominance/soumission représente le sentiment de contrôle d'un individu. Par exemple, les sentiments de peur et de colère sont tous les deux des sentiments négatifs à forte activation mais ils se différencient sur cette échelle de domination car la peur est associée à un sentiment de perte de contrôle tandis que la colère est associée à une prise de contrôle. Ce modèle permet de rendre compte des trois composantes de l'action identifiées en psychologie [24] :

Affectivité : *valence*. Chacun est plus ou moins sensible à certains sujets.

Cognition : *activation*. C'est du jugement et de l'évaluation de l'information que dépend le niveau d'excitation.

Conation : *dominance*. L'intention d'un individu dans un contexte donné dépend de ce qu'il est, consciemment ou non.

Il a été utilisé dans la construction d'agents animés qui manifestent des émotions comme Avatars.

1.1.5 Théorie des émotions TDRL (Time Difference Recurrent Learning)

La théorie TDRL s'inspire des sciences computationnelles pour proposer une compréhension des mécanismes qui sous-tendent les émotions. Son objectif principal n'est pas de modéliser directement les émotions humaines, mais elle fournit des concepts utiles pour évaluer l'état interne d'un individu en fonction de son expérience passée, présente et anticipée. Cette approche pourrait également permettre à un modèle d'apprentissage de déterminer quelles émotions il "devrait" ressentir dans un contexte donné, rendant ainsi les mouvements expressifs générés motivés et cohérents.

Dans ce cadre, une émotion est considérée comme une réaction à un changement interne ou externe, et elle conserve l'idée de valence, positive ou négative. Certains ressentis, comme la surprise ou la confusion, ne sont pas inclus car ils ne sont pas associés à une valence claire. Les émotions sont ancrées dans les sensations corporelles ou homéostatiques, mais des sensations corporelles seules, comme le dégoût gustatif, ne constituent pas une émotion. De même, des pulsions biologiques immédiates telles que la rage ou le désir sexuel ne sont pas considérées comme des émotions [4].

Les émotions fonctionnent comme des signaux de rétroaction :

- **Rétroaction anticipée** : espoir ou peur.
- **Rétroaction réfléchie, reflétant une situation** : déception ou soulagement.

Ainsi, une émotion manifeste un changement de comportement futur. Sa valence et son intensité déterminent son pouvoir motivationnel, conformément à la conception de l'émotion comme signal d'approche ou d'évitement. En résumé, une émotion est l'évaluation de l'adaptation comportementale appropriée à un événement.

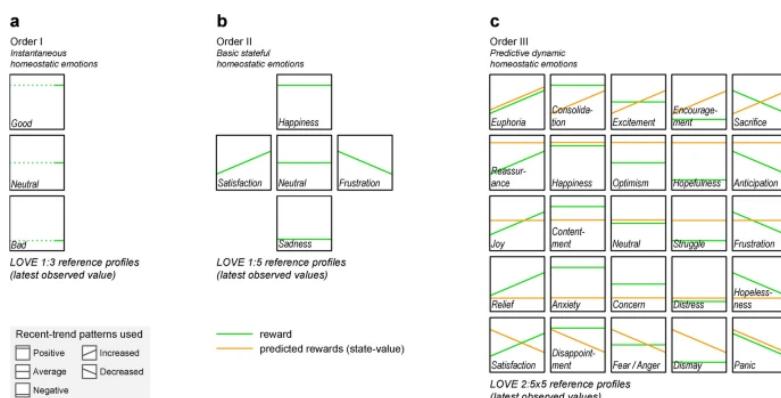


FIGURE 1.3 – Illustration de la théorie TDRL des émotions, tirée de *A generic self-learning emotional framework for machines* [11]

La Fig. 1.3 montre qu'en comparant la récompense réelle et la récompense attendue, un modèle d'émotion complet est obtenu. Les émotions ont été labellisées, permettant ainsi de créer un modèle d'émotion basé sur cette théorie.

1.2 Détection et évaluation des émotions

Comment détecter l'état interne d'une personne ? Les émotions modulent directement certains paramètres physiologiques et elles s'expriment surtout à travers un large éventail de comportements sociaux, c'est-à-dire de signaux non verbaux. Les individus mobilisent simultanément la posture, la gestuelle, le regard, la voix, les micro-expressions ou encore les schémas rythmiques de la parole pour transmettre — parfois involontairement — des indications sur leurs intentions, ou leur état interne. Un système intelligent peut exploiter ces multiples modalités pour inférer l'émotion d'un interlocuteur.

Parmi ces modalités, le visage constitue l'un des vecteurs les plus riches d'informations. L'observation des unités d'action faciales, comme celles décrites par le FACS, un système de codage des expressions faciales développé par Ekman et Friesen en 1978 [9], permet d'associer certains mouvements musculaires — contraction du front, élévation des commissures, plissement des paupières — à des états émotionnels typiques. Des modèles d'analyse d'images ou de séquences vidéo, notamment ceux reposant sur des réseaux convolutifs, apprennent à détecter automatiquement ces configurations. Certaines approches comparent également l'expressivité d'un individu à une référence « neutre » pour évaluer l'intensité émotionnelle.

La voix constitue un autre canal essentiel. Le timbre, l'intonation, la hauteur, la vitesse d'élocution ou encore l'amplitude des variations sonores évoluent de façon caractéristique selon l'état affectif. Des modèles supervisés peuvent classifier ces signaux à partir de leurs propriétés acoustiques, tandis que l'analyse de schémas de parole plus larges, tels que le rythme, les hésitations, les discontinuités, les ruptures, offre des indices complémentaires, parfois plus robustes que la simple prosodie.

1.2.1 Limites de ces modèles de détection et de classification

En combinant ces différentes modalités — visage, voix, rythme, posture, gestuelle — un système peut approcher une compréhension plus nuancée de l'état émotionnel. Toutefois, cette démarche présente plusieurs limites. Les manières de communiquer ses émotions ne sont pas homogènes d'une communauté à l'autre. Les expressions émotionnelles sont profondément façonnées par la culture : ce qui semble « naturel » pour certains ne l'est pas nécessairement pour d'autres. Les modèles risquent d'enfermer la compréhension des émotions humaines dans un cadre culturel dominant, en transformant des conventions sociales particulières en critères prétendument universels. Ce glissement normatif soulève des questions, tant pour la validité des systèmes que pour leurs implications sociales.

On peut imaginer une alternative plus ouverte : un apprentissage à visée auto-organisée ou autotélique, inspiré de la manière dont un enfant comprend progressivement les émotions des adultes en s'imprégnant de leurs comportements, en les imitant et en intégrant les associations contextuelles qui leur donnent sens. Un tel système pourrait potentiellement dépasser les catégories imposées par les approches classiques. Néanmoins, une telle

perspective relève d'un effort de recherche beaucoup plus large et sort du cadre de cet état de l'art.

1.3 Optimisation des réponses sociales du robot

Dans ce chapitre, après avoir cherché à modéliser une émotion afin d'identifier celles qu'un robot pourrait être amené à reproduire, il apparaît prudent d'adopter un regard critique. En effet, le robot peut générer des réponses qui paraissent séduisantes grâce à leur caractère innovant alors même qu'elles ne sont pas réellement adaptées aux intentions de l'utilisateur. Comment savoir si la solution apportée est réellement pertinente ? Quelle est la valeur ajoutée d'un robot "social" par rapport à un agent conversationnel désincarné ? Pour répondre à ces questions, des outils de mesure, souvent basés sur des questionnaires utilisateurs, ont été développés spécifiquement pour l'intelligence émotionnelle et sociale du robot. En effet, la robotique sociale n'en étant encore qu'à ses balbutiements, il serait incohérents d'évaluer les performances sociales d'un robot avec les mêmes outils que pour un être humain.

1.3.1 Intelligence sociale perçue

Dans l'étude *Mesurer l'intelligence sociale perçue des robots* [2] les auteurs ont développé les échelles d'intelligence sociale perçues (PSI), un ensemble de 20 échelles conçues pour évaluer la perception de l'intelligence sociale des robots de manière exhaustive, quel que soit leur type de corps ou leurs fonctionnalités. Ces échelles mesurent quatre dimensions principales : (1) la compétence sociale générale, (2) la capacité à identifier les personnes (présence, individus, groupes), (3) les aptitudes à reconnaître, s'adapter et prédire les comportements, cognitions et émotions humaines, et (4) la présentation sociale du robot, c'est-à-dire sa capacité à apparaître comme un partenaire social désirable : amical, fiable et prosocial.

1.3.2 Présence sociale du robot

Un autre outil est proposé dans *Développement et validation d'une échelle de mesure de la présence sociale des robots* [6], les auteurs partent de la notion de **présence sociale** humaine et proposent un modèle théorique de la présence sociale des robots. Ils retiennent cinq critères pour évaluer cette présence sociale :

- **Présence** : perception de la présence physique et consciente du robot
- **Allocation de l'attention** : attention portée au robot et perçue de la part du robot
- **Expression interactive et compréhension de l'information**
- **Interdépendance émotionnelle perçue** : influence mutuelle des états émotionnels et attitudes
- **Perception des comportements d'interaction**

Une échelle de mesure avec 17 questions a été développée et testée via des questionnaires, interviews et analyses statistiques (EFA, CFA). Les résultats montrent que le modèle est fiable, cohérent, et valide pour mesurer la présence sociale des robots.

2 Modélisation d'un mouvement expressif

Afin de permettre la modélisation de mouvements expressifs sur le Reachy Mini, il faut se pencher sur la traduction des émotions en gestuelle, en prenant en compte les degrés de liberté du robot. Il existe certains modèles de transcription du mouvement, tels que le modèle d'effort de Laban, souvent utilisé en robotique à cet effet [15, 21]. Ces modèles peuvent être exploités dans des architectures de transfert, souvent plus complexes. Ces architectures peuvent être divisées en deux grandes familles. Premièrement, celles qui permettent une génération de mouvement, par exemple à partir d'une entrée textuelle correspondant à l'émotion cible [3, 27]. Deuxièmement, celles qui adaptent un mouvement cible pour le rendre expressif [10, 21, 31] ; par exemple, attraper un objet donné en étant influencé par une émotion particulière.

Ces modèles et architectures peuvent servir de base pour modéliser un mouvement expressif. Cependant, dans le cadre du projet sur le Reachy Mini, il est intéressant d'approfondir certains points permettant d'améliorer la crédibilité du mouvement. Cette crédibilité passe par l'augmentation de l'expressivité : il existe des concepts fondamentaux tirés de l'animation qui sont aujourd'hui utilisés dans le domaine de la robotique, à cet effet [17, 3]. Enfin, il est nécessaire de prendre en compte la contrainte de la morphologie du robot. Le Reachy Mini ne possède que neuf degrés de liberté, il est donc nécessaire de savoir s'y adapter pour transcrire correctement l'émotion.

2.1 Transfert de l'émotion vers le mouvement

Afin de passer de la commande d'une émotion à l'action du mouvement, il est nécessaire d'établir un algorithme permettant d'établir une relation entre l'état affectif et les paramètres de mouvement [15]. Pour cela, il est possible de passer par des modèles de transfert, puis de les utiliser pour générer ou adapter des trajectoires pour le robot.

2.1.1 Modèles de mouvements

Les modèles de mouvements sont divisibles en deux catégories : structurels et fonctionnels. Les modèles structurels permettent de poser des marqueurs sur des mouvements existants, et de les évaluer par rapport à certains paramètres, qui diffèrent en fonction du modèle. Ces paramètres peuvent prendre des valeurs dans leur espace dédié, qu'on pourra qualifier d'espace affectif [15] puisqu'il permet de représenter les émotions selon des variables liées au mouvement. Ensuite, selon leur nature, ils peuvent être traduits en commandes motrices du robot. Quant aux modèles fonctionnels, ils permettent de catégoriser chaque mouvement en fonction de son objectif, et donc les paramètres qui en découlent. Ils prennent en compte le contexte et la signifiance.

Afin de rendre compte de la théorie derrière la retranscription du mouvement, deux modèles sont étudiés dans cette section : le modèle de Laban, de nature structurelle,

et le modèle d'Ekman et Friesen, à vocation fonctionnelle. Ils proposent tous deux une approche différente des mouvements, l'un permettant de les transcrire, l'autre de les classifier. Il est pertinent d'évaluer leur utilisabilité.

2.1.1.1 Laban

Le modèle dit de Laban appartient à la famille des modèles structurels, tout comme le modèle BAP, lui aussi utilisé dans le domaine de la robotique. Ils permettent de retranscrire des marqueurs spatiaux en un nouvel espace dépendant de variables paramétriques. Il est possible d'en créer un modèle computationnel, et de les utiliser pour traduire une émotion en un mouvement.

Plus spécifiquement, le modèle de Laban est un système de notation originellement créé pour la danse. Il permet d'écrire et d'analyser la structure et l'expressivité des mouvements lors d'une chorégraphie [15, 31]. Il est basé sur une approche structurelle. Il comporte quatre éléments principaux :

1. **Le corps** | Parties actives du corps et séquence de leur implication dans un mouvement.
2. **L'espace** | Endroit où se déroule un mouvement, directions du corps et des parties du corps.
3. **L'effort** | Attitude intérieure envers l'utilisation de l'énergie selon quatre composantes bidirectionnelles [31], illustrées sur la figure 2.1 :
 - a) l'espace, de indirect à direct ;
 - b) le poids, de léger à fort ;
 - c) le temps, de soutenu à soudain ;
 - d) le flux, de libre à lié.
4. **La forme** | Comprend le flux de forme, la directionnalité et le sculptage, qui décrivent tous les changements dynamiques dans la forme du mouvement.

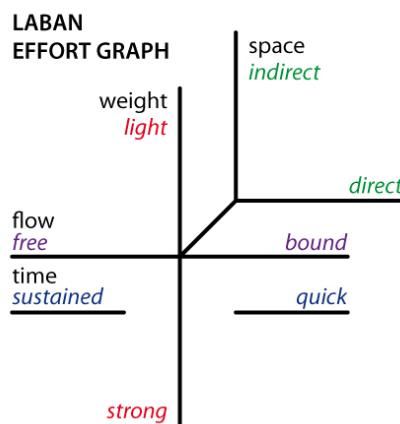


FIGURE 2.1 – Graphe d'effort de Laban

Le modèle de Laban permet de décrire initialement les mouvements du corps, non de les relier à des états affectifs. Cependant, c'est un modèle fréquemment utilisé lors de la

transcription d'émotions vers une trajectoire [15], comme dans les études de Samadani *et al.* [31] ou Osorio *et al.* [21].

Il est possible d'établir une relation entre expression affective et descripteurs de mouvement. Pour cela, Karg *et al.* proposent deux options [15] :

1. Créer manuellement un ensemble de relations possibles. Pour ce faire, il est conseillé de :
 - s'appuyer sur une relation entre l'état affectif et le mouvement issue d'études psychologiques ;
 - utiliser uniquement la composante *Effort*, ou bien utiliser à la fois les composantes *Effort* et *Forme*.
2. Créer une correspondance entre composantes de Laban et angles et vitesses articulaires à spécifier pour la cinématique du robot.

2.1.1.2 Ekman et Friesen

Le modèle d'Ekman et Friesen, défini en 1969 dans leur ouvrage *The Repertoire of Nonverbal Behavior : Categories, Origins, Usage, and Coding*, permet de rendre compte de la fonction communicationnelle du mouvement. Il regroupe les comportements kines-thésiques selon cinq catégories distinctes [15] :

1. **Emblèmes** | Gestes codés culturellement, ayant signification précise et compréhensible sans contexte verbal.
2. **Illustrateurs** | Gestes qui accompagnent le discours, pour illustrer, qui n'ont pas de sens seuls. Par exemple, dans leur étude, Fernandez-Fernandez *et al.* génèrent des trajectoires stylisées, qui peuvent servir d'illustrateurs [10].
3. **Manifestations affectives** | Mouvements ou postures qui expriment un état émotionnel interne. C'est ce qui est recherché dans les études [3] et [27], qui génèrent directement des mouvements expressifs.
4. **Régulateurs** | Gestes ou signaux qui régulent l'interaction sociale : prise de parole, écoute, attention, synchronisation.
5. **Manipulateurs** | Gestes réflexifs ou auto-centrés, souvent inconscients, liés à la gestion de l'état interne comme le stress et le confort, ou à des ajustements posturaux, souvent non consciemment communicatifs. Choi *et al.* présentent un système de reciblage qui se rapproche, fonctionnellement, de l'adaptation corporelle, et qui peut y être assimilé [7].

Contrairement à Laban ou à d'autres modèles structurels comme le BAP, le modèle d'Ekman et Friesen n'est pas un modèle qui permet un transfert direct entre émotion et mouvement. Cependant, il est possible de s'en servir pour construire un modèle exploitable pour la génération de mouvements, grâce à son approche exhaustive. Le fait d'apporter un vocabulaire spécifique à certains éléments permet d'en ignorer d'autres afin d'y voir plus clair.

Dans le cadre du Reachy Mini, il serait pertinent de se concentrer sur la gestuelle appartenant uniquement aux manifestations affectives, et non aux autres formes de gestes.

2.1.2 Architectures de transfert

Afin de passer de l'émotion à la modélisation du mouvement, il existe deux approches possibles. La première, c'est la génération totale du mouvement : elle prend en entrée l'émotion cible. La seconde, c'est l'adaptation d'un mouvement : elle prend en entrée une trajectoire fonctionnelle associée à une émotion cible.

2.1.2.1 Génération de mouvements

La génération directe de mouvements exprimant des émotions comprend les étapes suivantes :

1. Sélection de l'état émotionnel souhaité.
2. Traduction de l'état émotionnel en mouvement via des paramètres appartenant à deux espaces distincts.
3. Génération de la trajectoire associée.

Il existe différentes approches pour générer un mouvement à partir d'une émotion cible : des approches faites à la main, d'autres automatisées.

Il est possible d'utiliser un modèle émotif comme base de classification sur un espace discrétilisé. Bretan *et al.* utilisent le modèle de Valence et Activation [3]. Il est possible de reproduire leur expérience en utilisant un autre espace similaire, comme le PAD, ou le modèle de Plutchik. A contrario, d'autres études comme Rawal *et al.* utilisent directement un réseau de neurones convolutif, pour extraire des marqueurs faciaux d'images de visages expressifs classifiés [27].

L'espace construit sur ces paramètres sert de base pour l'état émotif recherché : des régions distinctes sont définies, correspondant à des émotions. Chaque point de l'espace est associé à un ensemble de variables correspondant aux mouvements possibles du robot. Ces variables prennent des valeurs différentes en fonction de la région dans laquelle elles se trouvent, et de leur place dans cette région. Dans leur cas, Bretan *et al.* ont utilisé la table de mouvements accompagnant les émotions définie par Walbott en 1998, issue des observations comportementales de Darwin en 1916 [3]. Ils ont défini les mouvements du robot à la main à l'aide de la table, puis les ont évalué et corrigé à l'aide de tests utilisateurs.

Cependant, il est aussi possible de traduire directement l'espace vers des configurations articulaires du robot via un générateur optimisé. C'est le cas de l'étude de Rawal *et al.*, qui ajuste la configuration des joints du robot pour que l'émotion prédite par le classificateur corresponde à l'émotion cible. Pour cela, une recherche globale par grille est combinée à une optimisation locale par rétropropagation. Le résultat est un intervalle de configurations articulaires capables de produire l'expression désirée [27].

2.1.2.2 Adaptation d'un mouvement

La deuxième façon de construire un mouvement émotif est d'adapter une trajectoire entreprise par le robot afin d'accomplir un but [15]. Des paramètres qui transforment le mouvement prédéfini pour le rendre expressif y sont injectés. Il y a donc deux entrées : un ordre de mouvement et un ordre d'émotion.

Dans le cadre de Reachy Mini, cela pourrait s'illustrer lorsqu'une action de communication non-verbale est effectuée. Par exemple, lorsque le robot tourne la tête afin d'observer une personne, lorsqu'il hoche la tête pour acquiescer ou encore s'il pointe quelque chose à l'aide de ses antennes.

L'adaptation automatique de mouvements exprimant des émotions comprend les étapes suivantes :

1. Sélection du type de mouvement.
2. Sélection de l'état émotionnel souhaité.
3. Modulation du mouvement par les paramètres émotionnels traduits dans son espace.
4. Génération de la trajectoire.

Il est possible d'automatiser ce processus, en définissant les variables adéquates. Cela peut se faire via un processus d'encodage-décodage. C'est le cas de l'étude [10], qui utilise un transfert de style basé sur des démonstrations humaines expressives. Chacune d'entre elles est encodée en un vecteur latent "*style*", et la trajectoire de base en un vecteur "*content*". Ensuite, les deux vecteurs sont combinés via un algorithme d'apprentissage par renforcement, pour générer des trajectoires robotisées qui reproduisent l'émotion de la démonstration.

1. Un **VAE** encode les mouvements humains expressifs : signaux issus de la cinématique, comme l'accélération et la vitesse angulaire de parties du corps. Ces signaux sont corrélés aux composantes de l'Effort de Laban : temps, espace, poids et flux.
2. Un **VAE** encode la configuration de base du robot en fonction de la tâche donnée.
3. Récupération de l'espace latent : combinaison des deux sources grâce à des **mécanisme d'attention**, pour injecter l'expressivité humaine dans la trajectoire du robot. Un gain expressif permet de moduler l'intensité de l'expressivité.
4. Un décodeur de type **GAN** prend ces paramètres en entrée pour générer la trajectoire finale.
5. Durant l'entraînement antagoniste, le discriminateur du GAN juge cette trajectoire et calcule une perte, pour superviser la prédiction des valeurs de l'Effort de Laban. Cela force l'encodeur à organiser l'espace latent selon des dimensions expressives significantes.

Le transfert de style offre un contrôle précis basé sur des démonstrations, tandis que la combinaison VAE/GAN se généralise plus globalement.

2.2 Crédibilité du mouvement

Afin de permettre la modélisation de mouvements crédibles chez Reachy Mini, deux paramètres additionnels peuvent être pris en compte. Le premier est la mise valeur de l'expressivité dans le mouvement. Il existe des principes qui peuvent être pris en compte dans l'algorithme de transfert émotif à ces fins. Le second est l'adaptation du mouvement à une morphologie réduite : Reachy Mini ne possède que neuf degrés de liberté, ce qui est drastiquement plus faible qu'un humain. Pour autant, la trajectoire finale se doit d'être réaliste et crédible.

2.2.1 Mise en valeur de l'expressivité dans le mouvement

En 1987, les animateurs des studios Walt Disney et Pixar ont proposé un ensemble de 11 principes de conception pour créer des personnages crédibles [17], qui sont devenus les piliers de l'animation traditionnelle.

Parmi ces 11 principes de conception, certains sont associés à l'expression d'états affectifs et peuvent être pertinents pour améliorer la crédibilité d'un agent virtuel ou d'un robot.

1. **Temporalité appropriée (n°2)** : un mouvement rapide ou lent influence l'expression. Dans [3], un temps T spécifique est initialisé pour chaque émotion, de 1 à 4 secondes.
2. **N'exprimer qu'un seul état (n°4)** affectif clairement défini à la fois. Par exemple, Bretan *et al.* utilisent une fonction de croissance et de décroissance pour passer d'un état discret à un autre [3]. Cela permet aussi de respecter le principe n°8 qui évoque les transitions fluides entre les différents états.
3. **Exagération (n°9)** : peut être utilisée pour améliorer la visibilité d'une expression. Certaines études l'implémentent, comme Bretan *et al.* qui ajoutent un facteur d'exagération lors de la génération du mouvement [3]. Quant à Osorio *et al.*, ils proposent l'implémentation d'un gain expressif. Ce dernier peut être modulé pour agir sur l'exagération du mouvement et donc son expressivité.
4. **État affectif comme action secondaire (n°10)** : séparer l'affect de l'action principale. Par exemple, exprimer de la joie en buvant [15]. L'observateur ne peut se concentrer que sur une seule action à la fois, donc une action secondaire simultanée peut passer inaperçue.
5. **Attrait (n°11)** : créer un mouvement que les spectateurs aiment regarder. Reachy Mini aura plus de chances d'être apprécié par ses utilisateurs si ses mouvements sont attrayants.

2.2.2 Adaptation du mouvement à une morphologie réduite

Afin d'adapter un mouvement à la morphologie de Reachy Mini, il faut veiller à ce que son faible nombre de degrés de liberté ne soit pas une barrière à la crédibilité de son expression.

Une solution naïve, comme celle adoptée par l'article [27], est de cartographier directement les caractéristiques émotionnelles repérées sur l'objet qui sert de modèle vers les articulations du robot disponibles. Cependant, sur un robot à faibles degrés de liberté, cela réduirait drastiquement les mouvements possibles.

Une solution plus raisonnable est de procéder par reciblage. Il s'agit de l'objet de l'étude de Choi *et al.*. Un reciblage est effectué, pour passer d'un émetteur humain à un corps animable humanoïde récepteur comportant moins de degrés de liberté [7]. Les positions des articulations du récepteur sont calculées pour que l'ensemble du corps reproduise au mieux la forme et le style global du mouvement de l'émetteur.

Enfin, une autre solution est celle qu'utilisent Osorio *et al.*. Leur méthode modifie simplement la trajectoire initiale envoyée en commande du robot [21]. Son avantage est donc qu'elle fonctionne indépendamment de sa morphologie.

3 Utilisation et génération du son pour la robotique sociale

Bien que le langage non verbal joue un rôle essentiel dans la communication des émotions et des ressentis, le son que produit un robot a un impact considérable selon l'information à transmettre. Ce chapitre aborde en premier lieu les différentes utilisations du son dans la robotique sociale, puis à la manière dont l'informatique permet la compréhension de ce dernier afin de le générer de manière adéquate.

3.1 Caractérisation du son dans la robotique sociale

Le *Sound Design* est le processus de création sonore afin d'introduire au mieux l'identité sonore du robot. Cette partie explore à la fois les domaines où le son est utile, puis ce qu'il faut prendre en compte dans ce processus.

3.1.1 Différents domaines d'utilisation

Le son occupe diverses fonctions dans la robotique sociale, selon le type de robot concerné et les objectifs qui lui sont assignés.

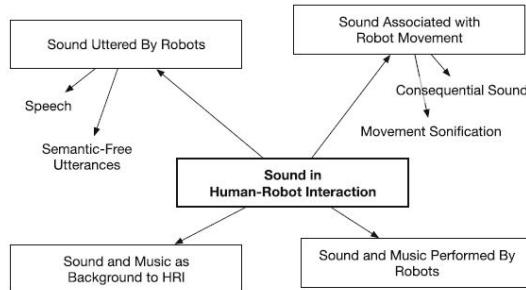


FIGURE 3.1 – Catégorisation haut-niveau du son dans les interactions Homme-Robot. Source : [28]

Ces dernières sont représentées sur le schéma 3.1 et vont être présentées dans cette partie.

3.1.1.1 Utilisation fonctionnelle du son

Il existe également des utilisations moins évidentes du son. Il peut, dans certains cas, permettre de localiser les robots. Ces derniers, pouvant être surprenants par leur taille ou leur forme, il est toujours rassurant de savoir où ils se situent. Le son peut également servir à masquer les bruits physiques du robot ou, à l'opposé, comme c'est le cas avec le

robot **Vector**, il est aussi possible d'utiliser les bruits du moteur, qui sont influencés par le mouvement et la direction, pour moduler le son et amplifier l'émotion exprimée , tel un son aigu du moteur qui rappellerait la vitesse du robot [28].

3.1.1.2 Langage et expression orale

La parole n'est pas seulement composée de texte, mais également de l'intonation et de la modulation de la voix. Le rire correspond par exemple à une altération du timbre et à l'élévation de la fréquence fondamentale [33]. Ces altérations peuvent par ailleurs être utilisées en même temps que le mouvement pour accentuer le sens voulu. Elles peuvent également créer des nuances pour un même mouvement. C'est par exemple le cas chez l'humain, qui peut taper du pied lorsqu'il est impatient, mais également lorsqu'il s'ennuie. La difficulté à distinguer l'émotion avec seul le mouvement est d'autant plus forte que le robot est restreint dans ses capacités motrices. Le défi est donc d'utiliser les variations possibles, qu'il s'agisse des onomatopées, des éléments paralinguistiques ou du ton employé, pour transmettre l'intention du robot. Ben Gabaldon, le *sound designer* du robot social **Cozmo**, explique que l'objectif est de créer un personnage convaincant plutôt qu'un personnage fidèle à son environnement [28]. Il est parfois nécessaire d'exagérer les expressions vocales du robot pour lui donner une personnalité intrigante. Ce schéma de pensée est d'ailleurs le même que pour le mouvement dans l'animation 2D ou encore pour le *character design* des personnages de fiction.

3.1.1.3 Musique

Plusieurs tentatives d'utiliser la musique afin de mettre à l'aise l'interlocuteur ou exprimer l'état émotionnel réactif à la discussion existent. L'une d'entre elles utilise la reconnaissance vocale et faciale afin d'évaluer le système d'expression de son robot. Même si ce dernier est principalement basé sur la génération de parole, il peut également choisir une musique à jouer en fonction de la discussion [22].

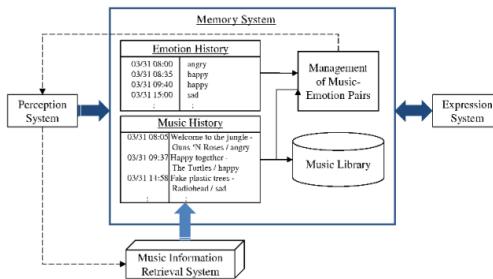


FIGURE 3.2 – Architecture du système de mémorisation de l'étude [22]

Comme le montre la figure 3.2, le robot de cette étude récupère l'historique de la conversation, que ce soit les émotions ou la musique jouée, pour prendre sa décision.

3.1.2 Paramètres du robot influençant l'émotion que le son produit

Une étude sur l'interprétation du son par le cerveau humain permet de déterminer que ce dernier différencie la langue et l'environnement. Cette étude s'intéresse notamment à

la langue japonaise, qui utilise les bruits d'animaux pour définir ces derniers, comme "*ka-ka*" pour le croassement du corbeau, plutôt que le nom de l'animal lui-même, "*karasu*", provenant de l'anglais "*crow*" [14].

Si l'objectif est également de donner une personnalité à un robot, certains facteurs vocaux ont une influence considérable. Par exemple, pour obtenir une voix jugée "mignonne", ou "*kawaii*", la voix doit être douce, proche de l'humain et le moins artificielle possible [32].

D'autres facteurs sont à prendre en compte pour caractériser un son. Certains sons ont une connotation de rapidité ou de taille, comme le montre la figure 3.3.

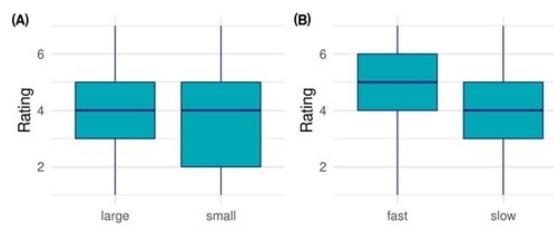


FIGURE 3.3 – Évaluations de la taille et de la rapidité du robot, notées sur une échelle de 1 à 7, à partir d'un enregistrement audio. Source : [20]

Il est donc possible de renforcer l'image de notre robot en modulant le son.

La position de la source sonore est également déterminante. Si celle-ci se trouve derrière le robot, l'interlocuteur ne percevra pas que le robot lui parle par sa bouche, lorsqu'il en possède une. De plus, la qualité sonore sera affectée. Cependant, il est possible de positionner délibérément les sources sonores et de jouer sur leur placement. C'est le cas de **Cozmo**, où Gabaldon a décidé de séparer la source des musiques, qui est diffusée sur les côtés droit et gauche du robot, de la source des paroles, qui se situe à l'avant [28]. Il serait également possible de produire le son d'un battement de cœur au niveau du corps, ou les sons mécaniques à l'arrière...

3.2 Interprétation sonore de l'émotion

Avec l'arrivée de l'intelligence artificielle pour assimiler informatiquement le monde qui nous entoure, les chercheurs se sont de plus en plus intéressés à la production cohérente de son, plutôt qu'à la sélection d'échantillons dans une base de donnée.

3.2.1 Corrélation entre le son et l'émotion ressentie

Avant de générer du son, il est important de comprendre comment le son et l'émotion sont corrélés.

La recherche sur la corrélation entre l'émotion générée et la musique n'est pas nouvelle. Bien que de nombreuses techniques de machine learning aient été utilisées, notamment les réseaux convolutionnels (CNN) pour l'extraction de propriétés de spectrogrammes, les réseaux de neurones récurrents (RNN) pour capturer les motifs mélodiques, rythmiques et les progressions harmoniques associées à l'émotion ressentie, ainsi que les Deep Belief

Networks (DBNs), qui ont montré des résultats prometteurs, ces approches présentent encore certaines limites [23].

De nouvelles méthodes étudiant les ondes cérébrales par l'intermédiaire d' électroencéphalographie (EEG) pour une évaluation plus scientifique ont été proposées, mais le bruit généré, la quantité limitée d'échantillons et le post-traitement nécessaire restent des obstacles à ces études [23].

L'utilisation de transformées par Zhou et Lian permet une précision de 97,1 % pour une classification tri-partie (neutre, négative, positive) et de 97,3 % pour une classification binaire (négative, positive) [23].

Une autre étude a également établi une corrélation entre l'émotion perçue et les instruments joués. Pour cela, 137 enregistrements d'instruments provenant de la bibliothèque symphonique de Vienne ont été utilisés, puis analysés à l'aide de la régression des moindres carrés partiels (PLSR) [19].

3.2.2 Production de son

La production sonore connaît un essor avec l'arrivée de l'intelligence artificielle. Il faut toutefois s'assurer de connaître son sujet. Différentes technologies ont été testées afin d'améliorer ou de générer du son.

3.2.2.1 Les biais

Tout comme pour la génération et l'utilisation de mouvements, le son n'est pas forcément interprété de la même manière par tous. Les onomatopées, par exemple, sont à 79 % préservées à travers les cultures, selon une étude réalisée avec 16 000 participants provenant des États-Unis, de Chine, d'Inde, d'Afrique du Sud et du Venezuela [5].

3.2.2.2 Générations connues et expérimentations

Un essai de génération vocale pour un robot professeur d'anglais s'est inspiré de **R2D2** et **Wall-E**. **R2D2**, selon le créateur de sa voix, Ben Burtt, utilisait un synthétiseur pour modifier sa propre voix. Tout comme la voix humaine, les notes aiguës sont courtes tandis que les notes graves sont continues. Sur le même principe et avec le même créateur, **Wall-E** fut inspiré par les onomatopées d'un enfant en bas âge, telles que "Oh", "Hm ?", "Huh". Pour ces deux robots, l'utilisation initiale d'une voix humaine afin d'obtenir un langage plus robotique permet d'avoir un rendu naturel. Il a également été conclu que les fréquences appropriées à un son confortable se situaient entre 100 et 1 500 Hz, et que le timbre détermine le caractère du robot [13].

Une étude [33] a tenté d'introduire le rire et l'hésitation dans des enregistrements audio. Pour ce faire, les auteurs utilisent la dernière consonne précédant l'élément para-linguistique comme point de départ pour l'annonce de l'élément. Ils déterminent dans un premier temps quels liquides, consonnes comme /l/, /r/, /m/, /n/ dont le flux d'air est relativement libre, sont considérés comme segmentables ou non. ils ont utilisé un dataset contenant 130 000 groupes de souffle, composé de sous-titres de films, de pièces de théâtre contemporaines, de recettes de cuisine et d'articles du Monde, ainsi que l'enregistrement

| Expression | Intonation | Pitch (Hz) | Timbre |
|---------------|----------------------------------------------------|------------|----------------|
| Affirmation | Descending progression of short sixteenth note | 262–1,175 | Whistle |
| | Arch motion of ascending and descending motion | | Synthesizer |
| Denial | Arch shape of ascending and descending progression | 587–1,396 | Whistle |
| | Connection of dotted rhythm | | |
| Encouragement | Ascending and descending progression | 262–1,175 | 2 Synthesizers |
| | 3 sixteenth notes | 294–494 | Whistle |
| Introduction | Ascending and descending progression | | siren |
| | | | Synthesizer |
| Question | Ascending three notes over background F# chord | 185–466 | Voice |
| | | | Synthesizer |
| Happiness | 4 sixteenth notes | 294–1,397 | Whistle |
| | Ascending motion of glissando | | Synthesizer |
| Sadness | Short sixteenth note and triple note | 294–1,175 | Ocarina |
| | Background sound of flute | | Flute |
| | | | Synthesizer |
| | | | Voice |

FIGURE 3.4 – Résultats d'analyse de cinq intentions et deux émotions dans les films "Star Wars" et "Wall-E". Source : [13]

de 200 prompts de rire pour entraîner leur modèle. L'évaluation de leurs résultats par le MOS a conclu à la réussite de l'inclusion de ces éléments[33].

Un modèle capable de modifier l'émotion associée à un enregistrement audio est également fonctionnel. Celui-ci peut être utilisé pour moduler les sorties de TTS et générer des vocalisations non verbales (rires, soupirs), rendant la voix plus réaliste [16].

Conclusion & perspectives

Dans le premier chapitre, il a été question de la modélisation et de l'évaluation des émotions dans le contexte des interactions sociales homme-robot. Différents modèles théoriques permettant de classifier les émotions humaines ont été présentés. La détection des émotions repose sur l'exploitation de signaux multiples, dont il a été question rapidement. Enfin, l'optimisation des réponses sociales des robots a été discutée à travers des outils permettant de mesurer la qualité des réponses apportées.

Le deuxième chapitre, portant sur le transfert des émotions vers le mouvement, permet de rendre compte des différentes architectures possibles. Le processus se fait généralement par une cartographie des états émotifs dans un premier espace, auquel sont associés des correspondances en termes de mouvements, ainsi que leurs paramètres. Afin de le rendre plus crédible, l'expressivité du mouvement peut être influencée par des variables additionnelles. Par ailleurs, il existe des techniques pour garder cette expressivité sur un robot à faibles degrés de liberté comme le Reachy Mini.

Enfin, le troisième chapitre concerne la production sonore pour la robotique sociale. De nombreuses études sont en cours, afin d'améliorer l'intégration d'éléments naturels de langage à la synthèse vocale. Le phénomène s'accentue actuellement, par le biais des grands modèles de langage, comme ChatGPT, qui sont utilisés comme support direct à la communication.

Mis en corrélation, ces trois grands axes de recherche permettent d'envisager un algorithme réaliste pour l'intégration de comportements expressifs crédibles chez le Reachy Mini.

Certaines problématiques restent cependant levées pour la suite de ce projet :

- Très peu d'études existent sur la génération totale de mouvements et de sons expressifs, et aucun ne correspond réellement au cas de Reachy Mini. Pour y pallier, il est possible d'établir une solution en synergie avec les approches les plus adaptées.
- Dans de nombreux articles, les mouvements générés sont majoritairement prédéfinis ou statiques, avec une adaptation dynamique limitée au contexte.
- La production sonore se limite souvent à des échantillons fixes ou à un nombre restreint d'émotions, sans intégration systématique avec les gestes.

Bibliographie

- [1] James A RUSSEL. *A circumplex model of affect*. en. Déc. 1980. URL : https://www.researchgate.net/publication/235361517_A_circumplex_model_of_affect (visité le 24/11/2025).
- [2] Kimberly A. BARCHARD et al. “Measuring the Perceived Social Intelligence of Robots”. en. In : *ACM Transactions on Human-Robot Interaction* 9.4 (déc. 2020), p. 1-29. ISSN : 2573-9522. DOI : [10.1145/3415139](https://doi.org/10.1145/3415139). URL : <https://dl.acm.org/doi/10.1145/3415139> (visité le 07/12/2025).
- [3] Mason BRETN, Guy HOFFMAN et Gil WEINBERG. “Emotionally expressive dynamic physical behaviors in robots”. en. In : *International Journal of Human-Computer Studies* 78 (juin 2015), p. 1-16. ISSN : 10715819. DOI : [10.1016/j.ijhcs.2015.01.006](https://doi.org/10.1016/j.ijhcs.2015.01.006). URL : <https://linkinghub.elsevier.com/retrieve/pii/S1071581915000191> (visité le 17/10/2025).
- [4] J. BROEKENS. “A Temporal Difference Reinforcement Learning Theory of Emotion: unifying emotion, cognition and adaptive behavior”. In : *ArXiv* (juill. 2018). URL : <https://www.semanticscholar.org/paper/A-Temporal-Difference-Reinforcement-Learning-Theory-Broekens/a4a8adb9d25076a9f3b69c446619ab4eb455eb76> (visité le 07/12/2025).
- [5] Jeffrey BROOKS et al. “Deep learning reveals what vocal bursts express in different cultures”. In : *Nature Human Behaviour* 7 (déc. 2022), p. 1-11. DOI : [10.1038/s41562-022-01489-2](https://doi.org/10.1038/s41562-022-01489-2).
- [6] Na CHEN et al. “Development and validation of a robot social presence measurement dimension scale”. en. In : *Scientific Reports* 13.1 (fév. 2023). Publisher: Nature Publishing Group, p. 2911. ISSN : 2045-2322. DOI : [10.1038/s41598-023-28817-4](https://doi.org/10.1038/s41598-023-28817-4). URL : <https://www.nature.com/articles/s41598-023-28817-4> (visité le 07/12/2025).
- [7] Sungjoon CHOI et Joohyung KIM. “Towards a Natural Motion Generator: a Pipeline to Control a Humanoid based on Motion Data”. en. In : *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. Macau, China : IEEE, nov. 2019, p. 4373-4380. ISBN : 978-1-7281-4004-9. DOI : [10.1109/IROS40897.2019.8967941](https://doi.org/10.1109/IROS40897.2019.8967941). URL : <https://ieeexplore.ieee.org/document/8967941/> (visité le 07/12/2025).
- [8] Charles DARWIN. “L’expression des émotions chez l’homme et les animaux”. fr. In : () .
- [9] Paul EKMAN et Wallace V. FRIESEN. *Facial Action Coding System*. en. Google-Books-ID: 7pqFtQAACAAJ. Consulting Psychologists Press, 1978.

- [10] Raul FERNANDEZ-FERNANDEZ et al. "Transferring human emotions to robot motions using Neural Policy Style Transfer". In : *Cognitive Systems Research* 82 (déc. 2023). ISSN : 1389-0417. DOI : [10.1016/j.cogsys.2023.05.010](https://doi.org/10.1016/j.cogsys.2023.05.010). URL : <https://www.sciencedirect.com/science/article/pii/S1389041723000499> (visité le 17/10/2025).
- [11] Alberto HERNÁNDEZ-MARCOS et Eduardo Ros. "A generic self-learning emotional framework for machines". en. In : *Scientific Reports* 14.1 (oct. 2024). Publisher: Nature Publishing Group, p. 25858. ISSN : 2045-2322. DOI : [10.1038/s41598-024-72817-x](https://doi.org/10.1038/s41598-024-72817-x). URL : <https://doi.org/10.1038/s41598-024-72817-x> (visité le 07/12/2025).
- [12] David INGLIS, John BONE et Rhoda WILKIE. *Nature: The nature of human nature*. en. Google-Books-ID: xGFOq8JKgCYC. Taylor & Francis, 2005. ISBN : 978-0-415-33306-1.
- [13] Eun-Sook JEE et al. "Sound design for emotion and intention expression of socially interactive robots". en. In : *Intelligent Service Robotics* 3 (juin 2010), p. 199-206. ISSN : 1861-2784. DOI : [10.1007/s11370-010-0070-7](https://doi.org/10.1007/s11370-010-0070-7). URL : <https://doi.org/10.1007/s11370-010-0070-7> (visité le 06/12/2025).
- [14] Junko KANERO et al. "How Sound Symbolism Is Processed in the Brain: A Study on Japanese Mimetic Words". In : *PLoS ONE* 9.5 (mai 2014). ISSN : 1932-6203. DOI : [10.1371/journal.pone.0097905](https://doi.org/10.1371/journal.pone.0097905). URL : <https://doi.org/10.1371/journal.pone.0097905> (visité le 19/11/2025).
- [15] Michelle KARG et al. "Body Movements for Affective Expression: A Survey of Automatic Recognition and Generation". In : *Affective Computing, IEEE Transactions on* 4 (nov. 2013), p. 341-359. DOI : [10.1109/T-AFFC.2013.29](https://doi.org/10.1109/T-AFFC.2013.29).
- [16] Felix KREUK et al. *Textless Speech Emotion Conversion using Discrete and Decomposed Representations*. arXiv:2111.07402 [cs]. Déc. 2022. DOI : [10.48550/arXiv.2111.07402](https://doi.org/10.48550/arXiv.2111.07402). URL : [http://arxiv.org/abs/2111.07402](https://doi.org/10.48550/arXiv.2111.07402) (visité le 19/11/2025).
- [17] John LASSETER. "Principles of traditional animation applied to 3D computer animation". en. In : *Proceedings of the 14th annual conference on Computer graphics and interactive techniques*. ACM, août 1987, p. 35-44. ISBN : 978-0-89791-227-3. DOI : [10.1145/37401.37407](https://doi.org/10.1145/37401.37407). URL : <https://doi.org/10.1145/37401.37407> (visité le 06/12/2025).
- [18] Jeniffer Xin-Ying LEK et Jason TEO. "Academic Emotion Classification Using FER: A Systematic Review". en. In : *Human Behavior and Emerging Technologies* (mai 2023). eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1155/2023/9790005>. ISSN : 2578-1863. DOI : [10.1155/2023/9790005](https://doi.org/10.1155/2023/9790005). URL : <https://doi.org/10.1155/2023/9790005> (visité le 24/11/2025).
- [19] Stephen MCADAMS, Chelsea DOUGLAS et Naresh N. VEMPALA. "Perception and Modeling of Affective Qualities of Musical Instrument Sounds across Pitch Registers". English. In : *Frontiers in Psychology* 8 (fév. 2017). Publisher: Frontiers. ISSN : 1664-1078. DOI : [10.3389/fpsyg.2017.00153](https://doi.org/10.3389/fpsyg.2017.00153). URL : <https://doi.org/10.3389/fpsyg.2017.00153>.

- frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2017.00153/full (visité le 07/12/2025).
- [20] Bastian ORTHMANN et al. “Sounding Robots: Design and Evaluation of Auditory Displays for Unintentional Human-robot Interaction”. In : *J. Hum.-Robot Interact.* 12.49 (2023), p. 1-26. DOI : [10.1145/3611655](https://doi.acm.org/doi/10.1145/3611655). URL : <https://dl.acm.org/doi/10.1145/3611655> (visité le 07/12/2025).
- [21] Pablo OSORIO et al. “A Generative Model to Embed Human Expressivity into Robot Motions”. In : *Sensors (Basel, Switzerland)* 24.2 (jan. 2024), p. 569. ISSN : 1424-8220. DOI : [10.3390/s24020569](https://doi.pmc.ncbi.nlm.nih.gov/articles/PMC10819644/). URL : <https://pmc.ncbi.nlm.nih.gov/articles/PMC10819644/> (visité le 17/10/2025).
- [22] Jeong-Sik PARK, Gil-Jin JANG et Yong-Ho SEO. “Music-aided affective interaction between human and service robot”. en. In : *EURASIP Journal on Audio, Speech, and Music Processing* 2012.5 (jan. 2012). ISSN : 1687-4722. DOI : [10.1186/1687-4722-2012-5](https://doi.org/10.1186/1687-4722-2012-5). URL : <https://doi.org/10.1186/1687-4722-2012-5> (visité le 06/12/2025).
- [23] Sheetal PATIL et al. “Review on music emotion analysis using machine learning: technologies, methods, datasets, and challenges”. en. In : *Discover Applied Sciences* 7.692 (juill. 2025). ISSN : 3004-9261. DOI : [10.1007/s42452-025-07178-9](https://doi.org/10.1007/s42452-025-07178-9). URL : <https://doi.org/10.1007/s42452-025-07178-9> (visité le 19/11/2025).
- [24] *Pleasure, Arousal, Dominance: Mehrabian and Russell revisited*. en. URL : https://www.researchgate.net/publication/265439455_Pleasure_Arousal_Dominance_Mehrabian_and_Russell_revisited (visité le 07/12/2025).
- [25] Robert PLUTCHIK. “A psychoevolutionary theory of emotions”. EN. In : *Social Science Information* 21.4-5 (juill. 1982). Publisher: SAGE Publications Ltd, p. 529-553. ISSN : 0539-0184. DOI : [10.1177/053901882021004003](https://doi.org/10.1177/053901882021004003). URL : <https://doi.org/10.1177/053901882021004003> (visité le 07/12/2025).
- [26] Robert PLUTCHIK et Henry KELLERMAN. *Theories of Emotion*. en. Google-Books-ID: Ie1GBQAAQBAJ. Academic Press, oct. 2013. ISBN : 978-1-4832-7001-2.
- [27] Niyati RAWAL et al. “ExGenNet: Learning to Generate Robotic Facial Expression Using Facial Expression Recognition”. en. In : *Frontiers in Robotics and AI* 8 (jan. 2022). ISSN : 2296-9144. DOI : [10.3389/frobt.2021.730317](https://doi.org/10.3389/frobt.2021.730317). URL : <https://www.frontiersin.org/articles/10.3389/frobt.2021.730317/full> (visité le 17/10/2025).
- [28] Frederic Anthony ROBINSON, Oliver BOWN et Mari VELONAKI. “Designing Sound for Social Robots: Candidate Design Principles”. en. In : *International Journal of Social Robotics* 14.6 (juin 2022), p. 1507-1525. ISSN : 1875-4805. DOI : [10.1007/s12369-022-00891-0](https://doi.org/10.1007/s12369-022-00891-0). URL : <https://doi.org/10.1007/s12369-022-00891-0> (visité le 24/11/2025).
- [29] James A. RUSSELL. “Affective space is bipolar”. In : *Journal of Personality and Social Psychology* 37.3 (1979). Place: US Publisher: American Psychological Association, p. 345-356. ISSN : 1939-1315. DOI : [10.1037/0022-3514.37.3.345](https://doi.org/10.1037/0022-3514.37.3.345).

-
- [30] James A. RUSSELL et Albert MEHRABIAN. “Evidence for a three-factor theory of emotions”. In : *Journal of Research in Personality* 11.3 (1977). Place: Netherlands Publisher: Elsevier Science, p. 273-294. ISSN : 1095-7251. DOI : [10.1016/0092-6566\(77\)90037-X](https://doi.org/10.1016/0092-6566(77)90037-X).
 - [31] Ali SAMADANI, Rob GORBET et Dana KULIC. *Affective Movement Generation using Laban Effort and Shape and Hidden Markov Models*. arXiv:2006.06071 [cs]. Juin 2020. DOI : [10.48550/arXiv.2006.06071](https://doi.org/10.48550/arXiv.2006.06071). URL : <http://arxiv.org/abs/2006.06071> (visité le 05/12/2025).
 - [32] Katie SEABORN et al. “Can Voice Assistants Sound Cute? Towards a Model of Kawaii Vocalics”. In : *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*. arXiv:2304.12809 [cs]. Avr. 2023, p. 1-7. DOI : [10.1145/3544549.3585656](https://doi.org/10.1145/3544549.3585656). URL : <http://arxiv.org/abs/2304.12809> (visité le 24/11/2025).
 - [33] Lionel SÉGALEN et Didier CADIC. “Introduction d’éléments paralinguistiques en synthèse vocale”. fr. In : () .