



Disponible en ligne sur

ScienceDirect
www.sciencedirect.com

Elsevier Masson France

EM|consulte
www.em-consulte.com

*neuropsychiatrie
de l'enfance
et de l'adolescence*

Neuropsychiatrie de l'enfance et de l'adolescence 65 (2017) 21–32

Article original

Rééducation de l'expression émotionnelle chez l'enfant avec trouble du spectre autistique grâce aux supports numériques : le projet JEMImE

The reeducation of emotional expressions for children with autism spectrum disorders thanks to information communication technologies: JEMImE project

C. Grossard^{a,*}, S. Hun^c, S. Serret^c, O. Grynszpan^b, P. Foulon^d, A. Dapogny^b, K. Bailly^b,
L. Chaby^{b,e}, D. Cohen^{a,b}

^a Service de psychiatrie de l'enfant et de l'adolescent, GHU Pitié-Sapètrière Charles-Foix, 47-83, boulevard de l'Hôpital, 75013 Paris, France

^b CNRS UMR 7222, institut des systèmes intelligents et robotiques, université Pierre-et-Marie-Curie, 4, place Jussieu, 75005 Paris, France

^c Cognition-Behaviour-Technology (CoBTek), EA 7276, université de Nice Sophia-Antipolis, 10, rue Molière, 06100 Nice, France

^d Groupe e-Genious, 3 ter, rue des Pins, 34000 Montpellier, France

^e Institut de psychologie, université Paris-Descartes, Sorbonne-Paris-Cité, 71, avenue Edouard-Vaillant, 92100 Boulogne-Billancourt, France

Résumé

L'utilisation de jeux sérieux pour favoriser l'émergence de capacités d'interactions sociales plus abouties chez les personnes avec autisme est de plus en plus explorée. Après une revue de la littérature concernant les jeux sérieux focalisés sur la reconnaissance des émotions ayant été proposés et utilisés avec des personnes autistes, nous décrivons l'architecture d'un nouveau jeu sérieux JEMImE dont l'objectif est de travailler la production émotionnelle uni et bimodale en contexte social. Nous décrivons le principal verrou technologique, à savoir la programmation d'un algorithme de reconnaissance émotionnelle capable de donner en temps réel un feed-back au joueur. Cet algorithme sera développé suivant des principes d'apprentissage machine sur une base de données d'émotions produites par des enfants tout-venant recrutés dans des écoles et spécifiquement créée pour le projet.

© 2016 Elsevier Masson SAS. Tous droits réservés.

Mots clés : Jeu sérieux ; Algorithme automatisé ; Émotion ; Trouble du spectre autistique

Abstract

The use of serious games to promote social skills in children and adolescents with autism spectrum disorder (ASD) has been a focus of interest during the past 10 years. We begin with a review of the literature regarding serious games proposed for individuals with ASD that focus on emotion processing. We found a total of 16 games but their overall quality was often limited both in terms of design/playability and clinical interest/efficacy. Additionally the study of emotion production was absent with the games being focused on emotion recognition. Second, we describe the design principles and the computational architecture of a novel serious game, JEMImE, that aims to promote uni and bimodal emotion production in individuals with ASD. We detail the technological gap and the need to compute an automated algorithm able to recognize emotion in real time in order to give feed-back to the player. This algorithm will be designed based on machine learning principles. To achieve this goal, we created a video database including more than 150 children and adolescents with typical development who were producing emotion.

© 2016 Elsevier Masson SAS. All rights reserved.

Keywords: Serious game; Emotion; Autism spectrum disorder; Machine learning

* Auteur correspondant.

Adresse e-mail : charline.grossard@aphp.fr (C. Grossard).

<http://dx.doi.org/10.1016/j.neurenf.2016.12.002>

0222-9617/© 2016 Elsevier Masson SAS. Tous droits réservés.

1. Introduction

Les troubles du spectre autistique (TSA) sont des troubles neuro-développementaux qui se caractérisent, entre autres, par des difficultés dans les interactions sociales, notamment concernant la compréhension des émotions d'autrui et la production d'expressions en accord avec les émotions ressenties [1,2]. Ces symptômes retentissent sur l'intégration des personnes avec TSA à l'école, dans le monde professionnel et dans la société de façon plus générale. Aider les personnes avec TSA à développer leurs habiletés sociales est donc un enjeu important dans leur prise en charge.

Ces dernières années, de nombreux chercheurs se sont intéressés aux technologies numériques dans la prise en charge de ce trouble (voir pour revue, Durkin, 2010) [3]. Certaines études semblent indiquer que les supports numériques permettraient d'améliorer l'intérêt des enfants avec TSA, plus encore que chez les enfants typiques [4]. De plus, ces supports ont l'avantage de présenter les informations de manière séquentielle : le déroulé est prévisible et donc rassurant [5,6]. Enfin, l'outil numérique est un support permettant de recréer facilement des situations de vie réelle et ainsi fournit au thérapeute de nombreuses situations de travail [7]. Quelques revues récentes se sont intéressées à l'efficacité de ces nouvelles approches basées sur les nouvelles technologies de l'information et de la communication (TIC) [8]. Celles-ci semblent indiquer que ce type d'intervention est prometteur auprès d'un public avec TSA. Parmi les TIC, on distingue 3 domaines principaux :

- les applications qui sont des programmes pour petits systèmes (téléphone, tablette) dont l'objectif est limité et porte sur un facilitateur de la vie quotidienne ;
- Les *serious games* ou jeux sérieux que l'on peut décrire comme des jeux sur supports numériques ayant une visée éducative allant au-delà du divertissement [9]. Ceux-ci existent sur des supports multiples : ordinateurs, tablettes, etc ;
- enfin, les robots.

L'utilisation des nouvelles technologies pourrait donc être un support intéressant pour la rééducation de la compréhension et de la production des expressions faciales et vocales émotionnelles chez la personne avec TSA. Dans ce domaine, de nombreux jeux sérieux ont été créés. Cependant, malgré un fort engouement pour ce type de jeu, leur utilisation reste limitée [10] ; la plupart des jeux proposés ne peuvent actuellement être utilisés que selon certaines conditions qui limitent leur utilisation sur un large panel de personnes avec TSA (niveau cognitif, âge, modalité...). Enfin, la méthodologie des études cliniques portant sur ces jeux ne permet pas de conclure à leur efficacité, et notamment concernant le transfert de compétences en situation réelle.

Beaucoup de jeux sérieux proposés ont pour but d'améliorer la compréhension des émotions chez les personnes avec TSA. C'est le cas de JeStiMule, dont l'utilité a été validée dans une étude pilote [11]. Par contre, peu de jeux proposent de travailler la production des émotions chez les personnes avec TSA. Dans

cette optique, un projet intitulé JEMImE, construit sur le modèle de JeStiMule, vise à développer un jeu sérieux permettant de travailler les expressions faciales et vocales émotionnelles chez les enfants avec TSA tant en compréhension qu'en production. Ce jeu intégrera un module algorithmique capable de reconnaître les émotions faciales et vocales produites par l'enfant en temps réel. Il devrait ainsi permettre de passer un cap supplémentaire dans le développement des jeux sérieux en offrant un feed-back instantané à l'enfant sur la qualité de ses expressions.

Nous proposons dans cet article, après une brève revue de la littérature existante, de décrire succinctement le jeu JeStiMule et enfin de détailler les principes généraux et les verrous technologiques attendants à la réalisation de JEMImE dans les conditions actuelles d'utilisation.

2. Les jeux existants

Dans une revue plus détaillée, nous avons colligé les jeux sérieux ayant pour objectif de travailler les aptitudes sociales des personnes TSA [12]. Parmi ces jeux, nous avons recensé 16 jeux sérieux visant à améliorer la compréhension et l'expression des émotions chez les personnes avec TSA (Tableau 1). La plupart de ces jeux cherchent uniquement à travailler la reconnaissance des émotions (12/16). Ceux-ci s'appuient sur des supports comme des photos, des dessins, des enregistrements audio, des vidéos, des personnages 3D. Cependant, bien que la reconnaissance des émotions soit multimodale par nature, la moitié des jeux (6/12) ne proposent qu'un apprentissage sur support fixe (dessins, photographies...) et unimodal (le plus souvent visuel). Le support audio n'est que rarement proposé et seul un jeu s'intéresse à la posture corporelle comme indice de l'émotion ressentie. Par contre, la plupart de ces jeux proposent de travailler la reconnaissance des émotions en contexte. Seuls 4 jeux (*LifeIsGame*, *CopyMe*, *SmileMaze*, le *serious game* de Park et al. [2012]) proposent de travailler la production des émotions [9,13–15]. Parmi eux, seul le jeu *LifeIsGame* propose un support non fixe (avatar 3D) [13]. Il incite également à travailler la production émotionnelle uniquement grâce au contexte : le joueur est amené à produire une expression faciale à partir d'un contexte mais sans le modèle d'un avatar.

À partir de grilles standardisées, nous avons essayé dans cette revue d'appréhender 2 dimensions :

- la première concerne la qualité des études cliniques rapportant une efficacité clinique éventuelle d'un jeu ; la grille est proche des standards de l'*evidence based medicine* [population, type d'étude (c.à.d. étude randomisée contrôlée vs. étude de cas), généralisation, confiance dans les résultats] [16] ;
- l'ergonomie informatique du jeu pour laquelle nous avons créée une échelle basée sur les travaux de Yusoff (2010) [17].

Cette échelle propose une liste d'attributs structurant un jeu sérieux en se basant sur des théories cognitivistes, constructivistes et comportementales. Ces attributs incluent l'apprentissage progressif, la linéarité (évolution progressive entre les niveaux), les capacités attentionnelles, l'aide apportée au joueur, le transfert des acquis, l'interaction, le contrôle

Tableau 1
Résumé des jeux travaillant les compétences émotionnelles chez les patients avec TSA.

| Projet (Référence) | Participants ciblés | Habilités ciblées | Support | Score à l'échelle de Connolly | Score à l'échelle de Yussof |
|---|--|--|-----------------------------------|-------------------------------|-----------------------------|
| Aprende con zapo (Lozano et al., 2011) | Enfants avec TSA entre 8 et 18 lecteurs | Reconnaissance d'expressions faciales avec et sans contexte | Dessins et photos | + | + |
| CMotion (Finkelstein et al., 2009) | Individus avec un haut niveau de fonctionnement | Reconnaissance d'expressions faciales en contexte | Agent virtuel en 3D | + | ++ |
| CopyMe (Tan et al., 2013) | Enfants avec TSA | Mime rune expression faciale à partir d'un modèle | Photos | + | + |
| Computer-based program (Bölte et al., 2002) | Adolescents et adulte avec un haut niveau de fonctionnement ou syndrome d'Asperger | Reconnaissance des expressions faciales associées aux émotions de bases | Photos | ++ | + |
| The junior detective training program (Beaumont et al., 2008) | Enfants avec syndrome d'Asperger (7,5 à 12 ans) | Reconnaissance d'expressions faciales et habiletés sociales en contexte | Jeu en 3D | +++ | +++ |
| Emotion Trainer (Silver & Oakes, 2001) | Individus avec syndrome d'Asperger | Reconnaître et prédire les expressions faciales émotionnelles | Photos | + | + |
| Facial expression recognition serious game (Christinaki et al., 2013) | Enfants avec TSA (2 à 6 ans) | Reconnaissance d'expressions faciales avec et sans contexte | Photos | + | ++ |
| FaceSay2 (Hopkins et al., 2011) | Enfants et adolescent avec TSA avec un bas et un haut niveau de fonctionnement | Reconnaissance d'expressions faciales et travail de l'attention conjointe | Photos et dessins | +++ | ++ |
| JeStimule (Serret and al., 2014) | Enfants et adolescents avec un bas et un haut niveau de fonctionnement | Reconnaissance d'expressions faciales avec Jeu en 3D et sans contexte | Jeu en 3D | +++ | +++ |
| Let's face it (Tanaka et al., 2010) | Enfants, adolescents et adultes avec TSA ou syndrome d'Asperger | Reconnaissance d'expressions faciales | Photos | +++ | ++ |
| LifeIsGame (Fernandes et al., 2011) (Alves et al., 2013) | Individus avec TSA | Reconnaissance d'expressions faciales avec et sans contexte | Avatar en 3D (ordinateur et Ipad) | + | + |
| Mind Reading (Golan et Baron- Cohen, 2006) | Individus avec TSA de haut niveau de fonctionnement | Reconnaissance d'expressions faciales et vocales avec et sans contexte | Films, enregistrements, écrit | +++ | +++ |
| SmileMaze (Cockburn et al., 2008) | Enfants avec TSA | Reconnaissance et production d'expressions faciales avec et sans contexte | Smileys | + | + |
| Theory driven serious game framework (Park et al., 2012) | Individus avec TSA lecteurs | Reconnaissance et production d'expressions faciales avec et sans contexte | Photos et écrit | + | + |
| The transporters (Golan et al., 2010) | Enfants avec TSA (3 à 8 ans) | Comprendre des émotions en contexte | Vidéo d'animation en 3D | +++ | + |
| Virtual reality in second life (Kandalaf et al., 2013) | Individus avec syndrome d'Asperger | Reconnaissance d'expressions faciales en contexte et compétences communicationnelles | Jeu en 3D | ++ | +++ |

Pour Connolly : += score entre 5 et 8/++= score entre 9 et 11/+++ = score entre 12 et 15. Pour Yussof : += score entre 9 et 12/++= score entre 13 et 16/+++ = score entre 17 et 21.

du joueur, la pratique, les feed-backs intermittents, les récompenses, les situations de vie réelle et la personnalisation par le joueur (Tableau 1).

Il ressort de ce travail de revue, de nombreux biais qui méritent d'être soulignés. Concernant la validation clinique, quatre jeux n'ont même jamais été proposés à des patients avec TSA et 2 n'ont testés que l'utilisabilité du jeu, sans réelle évaluation des bénéfices pour les patients. Dans les études d'efficacité, la taille des échantillons est souvent trop restreinte (< 30 enfants par groupe) et peu d'études ont inclus un groupe contrôle. Une seule étude (portant sur le jeu *Let's Face It!*) inclus à la fois un groupe contrôle d'enfants au développement typique et un groupe contrôle d'enfants avec TSA appariés par âge développemental et diagnostic au groupe suivant le jeu [18]. Quant aux effets cliniques, plusieurs jeux démontrent que les patients avec TSA présentent de meilleures habiletés sociales après l'entraînement, mais aucune étude ne permet d'affirmer que ces résultats sont généralisés à la vie. Or, de nombreux thérapeutes s'accordent sur le fait qu'une thérapie efficace doit produire des résultats visibles et généralisables dans les situations de vie réelle [11]. Cette revue permet cependant de faire ressortir des études à la méthodologie plus robuste pour les jeux suivants : *The junior detective program* [19], *Emotion Trainer* [20], *FaceSay2* [21], *JeStiMule* [11], *Let's Face it* [18], *Mind reading* [22] et *The Transporters* [23].

Concernant l'ergonomie des jeux sérieux, les jeux incluent généralement différents attributs. Trois attributs sont rarement présents (la prise en compte du temps d'attention, la mise en œuvre de récompense et la personnalisation par le joueur). Il existe de grandes disparités, chaque jeu s'avérant très différent, dépendant de l'orientation prise par son créateur. Enfin, de façon générale, il apparaît que les jeux les plus robustes au niveau méthodologique ne sont que rarement ceux qui ont produit un effort au niveau de l'ergonomie et la jouabilité.

En termes de public cible, la moitié des jeux (8/16) nécessitent a minima de bonnes capacités de lecture et ne sont destinés qu'à des patients avec un haut niveau de fonctionnement. Ils ne sont donc pas accessibles pour de nombreux patients TSA qui présentent une déficience intellectuelle associée ou un trouble de la compréhension du langage oral ou écrit. Certains jeux tentent de contourner ce problème en proposant des consignes données à l'oral ; cependant, beaucoup de personnes avec TSA présentent également des difficultés de compréhension orale, ce qui complique également l'accès à ces jeux. Peu de jeux sérieux proposent d'utiliser un support visuel (autre que du langage écrit) comme mode de communication. En conséquence, une minorité de jeux est utilisable auprès d'une large population d'enfants avec TSA.

Une autre difficulté se pose quant à la définition d'un jeu sérieux. La plupart des études retrouvées se concentrent sur les objectifs thérapeutiques, au détriment de l'aspect ludique. La plupart des jeux proposés ressemblent donc à des exercices, sans posséder les caractéristiques d'un jeu vidéo. D'autre part, on retrouve peu d'informations sur l'architecture du jeu alors que les résultats sont mis en avant. Ainsi, les jeux les mieux évalués en termes d'utilité clinique au niveau méthodologique ne sont pas nécessairement les jeux les plus intéressants en termes de

jouabilité et d'ergonomie. Une interprétation de ces résultats pourrait être que les champs de la psychologie du développement et de la psychiatrie n'interagissent pas facilement avec celui de l'ingénierie.

Ce manque de prise en compte dans la structuration du jeu va de paire avec l'absence de guide décrivant comment construire un jeu sérieux pour travailler les habiletés sociales chez les personnes avec TSA. Plusieurs guides visent à aider les équipes dans la création de jeux sérieux [17,24,25]. Cependant, ces guides ne sont pas spécifiques aux patients avec TSA ou à l'apprentissage des habiletés sociales. Récemment, Khowaja et Salim (2014) ont tenté de résumer les principaux composants d'un jeu sérieux à destination d'un public avec TSA [26].

Enfin, une dernière critique concerne l'utilisation des plateformes numériques dans ces recherches. La plus-value apportée par les nouvelles technologies réside aussi dans la possibilité de mettre en place des activités qui ne pourraient pas exister entre un patient et son thérapeute dans une prise en charge classique. Or, de nombreux jeux utilisent des photos ou des images qui pourraient être présentés dans un format plus classique de type papier/crayon [18,21,27]. Les chercheurs doivent désormais viser à exploiter au maximum les possibilités offertes par l'outil informatique. Ces supports permettent de créer des environnements 3D proches de la réalité, ainsi que de récupérer des informations objectives et en temps réel sur le comportement de l'enfant (par exemple sur l'utilisation du regard, les temps de traitement) qui sont souvent de recueil difficile par les thérapeutes.

Malgré ces limites, l'utilisation des jeux sérieux semble pouvoir offrir de réelles opportunités pour travailler auprès des patients avec TSA. Une récente méta-analyse se concentrant sur les remédiations basées sur les nouvelles technologies qui ont été testées auprès de patients TSA indépendamment de la cible thérapeutique mais dans des études de bonne qualité d'un point de vue de l'*evidence based* ($N_{\text{études}} = 14$), montre que ce type de prise en charge est efficace [8].

3. JeStiMule : présentation

JeStiMule [11] est un jeu sérieux visant à travailler la reconnaissance de 7 émotions (joie, colère, tristesse, surprise, dégoût, peur) et de la douleur grâce aux expressions faciales mais également grâce aux gestes. Ce jeu s'adresse à des enfants et adolescents avec TSA, quels que soient leur niveau de fonctionnement, leurs compétences intellectuelles et langagières. Le jeu se divise en deux phases :

- une phase d'apprentissage ayant pour but d'apprendre aux enfants à associer une expression faciale et des gestes à une émotion matérialisée par un code couleur ou grâce à du vocabulaire émotionnel (mots ou expressions idiomatiques) ;
- une phase de jeu dans laquelle le joueur est immergé dans un environnement 3D et pendant laquelle il assiste à des scènes dans lesquels il doit reconnaître ou anticiper une émotion en interprétant la situation. Cette phase comporte 5 lieux différents (parc, théâtre, magasin, restaurant, jardin) dans

lesquels l'avatar de l'enfant se déplace et rencontre différentes situations sociales.

Afin de pouvoir proposer ce jeu à un large panel d'enfants avec TSA, JeStiMulE propose 3 modalités de réponses (Fig. 1). En fonction des capacités de l'enfant, le thérapeute peut donc choisir de lui proposer d'associer une émotion à :

- un code non verbal (code couleur) ;
- un mot émotionnel (ex : joie, content) ;
- une expression idiomatique (ex : il est aux anges, avoir la chair de poule).

Ces associations sont apprises durant la phase d'apprentissage et réutilisées dans la phase de jeu. Il a ainsi été possible de proposer le jeu à des patients non lecteurs en utilisant le code couleur comme modalité de réponse.

Afin de respecter la définition du terme jeu sérieux (e.g. apprendre en s'amusant), deux aspects ludiques ont été insérés. Le premier concerne une quête de bonbons disséminés dans l'environnement 3D afin d'encourager l'enfant à explorer les 5 lieux. Le second concerne la récompense lors de la reconnaissance correcte de l'émotion en situation. En cas de réussite, le joueur gagne une pièce de puzzle et peut choisir comment réagir face à cette situation. Le jeu est terminé quand l'enfant a reconstitué le puzzle (Fig. 2).

Trente-trois enfants, entre 6 et 17 ans et ayant tous un diagnostic d'autisme ont été inclus dans une étude d'évaluation de 1) l'adaptabilité et 2) l'efficacité du jeu. La population est hétérogène au niveau du QI et intègre aussi bien des patients de bas que de haut niveau de fonctionnement avec différents niveaux de compétences langagières et de lecture. Au plan de la jouabilité, tous les enfants ont pu jouer à la phase d'apprentissage et utiliser une modalité de réponse adaptée. Un seul enfant n'a pas pu apprendre à associer les émotions et le code couleur et n'a pas pu jouer à la phase de jeu. Tous les autres ont pu apprendre l'association et évoluer dans le jeu.

Concernant l'évaluation de l'efficacité du jeu sur la reconnaissance des émotions (faciales, gestuelles et en situation), il a été proposé aux patients de les reconnaître sur les avatars de JeStiMulE (pour évaluer les apprentissages effectués) et sur des photos de personnages réels (pour évaluer la généralisation des apprentissages à des stimuli réels) avant et après l'entraînement avec JeStiMulE. Les résultats montrent une progression significative des résultats dans ces deux conditions que ce soit pour la reconnaissance faciale, gestuelle ou en situation. JeStiMulE permet donc d'améliorer la reconnaissance des expressions faciales aussi bien pour les enfants ayant un bas ou un haut niveau de fonctionnement, et ce pas seulement sur les avatars servant à l'entraînement mais également sur des photos de visages sur lesquels ils n'ont pas été entraînés.

4. JEMImE (jeu éducatif multimodal d'imitation émotionnelle) : principes généraux

Le projet JEMImE s'inscrit dans la continuité du projet JeStiMulE. Le jeu a pour objectif d'apprendre aux enfants

TSA à produire une émotion de qualité correcte en termes d'adaptabilité en contexte. Cette production concerne trois émotions (joie, tristesse, colère). Le prototype du jeu s'adresse aux enfants TSA sans déficience intellectuelle âgés de 6 à 12 ans. La nouveauté technologique réside dans la tentative d'apporter un feed-back en temps réel sur l'expression émotionnelle produite par l'enfant en intégrant une mesure de l'émotion produite.

JeMImE est construit sur le même principe que JeStiMulE et comprend donc deux phases de jeu (Fig. 3) :

- la phase d'apprentissage dont l'objectif est d'apprendre aux enfants avec TSA à produire des expressions faciales et vocales de bonne qualité sur :
 - imitation (l'enfant produit une émotion à partir d'un modèle en 3D),
 - mime (l'enfant produit une émotion sans modèle). Ces deux situations sont proposées à l'enfant sans contexte social puis en situation sociale ;
- la phase d'entraînement dont l'objectif est de mettre l'enfant avec TSA en situation la plus écologique possible, afin qu'il généralise les apprentissages effectués lors de la première phase. L'enfant en situation de jeu devra :
 - produire spontanément une émotion en réaction à une action (scénario 1),
 - comprendre le contexte social et produire spontanément l'émotion attendue (scénario 2),
 - comprendre une situation plus complexe (par exemple de fausse croyance. . .) et produire l'émotion attendue dans le contexte (scénario 3).

Afin de pouvoir donner un retour objectif sur les productions de l'enfant durant le jeu, il est nécessaire de pouvoir évaluer la qualité des expressions faciales réalisées en temps réel. Dans ce but, l'un des leviers technologiques du projet sera la création d'un outil de détection et d'évaluation automatique de la production des expressions émotionnelles faciales et vocales. Ce prototype utilisera un algorithme basé sur l'apprentissage machine construit à partir d'une base de données d'expressions émotionnelles vocales et faciales chez l'enfant au développement typique (Fig. 3).

5. Création d'une base de données d'émotions produites par des enfants

Si l'on retrouve de nombreuses bases de données d'expressions faciales chez l'adulte, il en existe très peu chez l'enfant. Or, il semble délicat de comparer les expressions faciales d'un adulte à celles d'un enfant du fait de la transformation morphologique et des différences dans l'expression des émotions [28]. Par ailleurs, les algorithmes de reconnaissance entraînés chez l'adulte ne peuvent être utilisés chez l'enfant dont les visages sont en général plus ronds et moins saillants que les adultes. Enfin, même si les émotions sont universelles, elles sont empreintes de dimensions culturelles dont il faut tenir compte [29]. Le Tableau 2 répertorie les bases de données existantes, comprenant des enfants dans leur cohorte et en donne les principales caractéristiques. Seules 7 bases de données avaient été

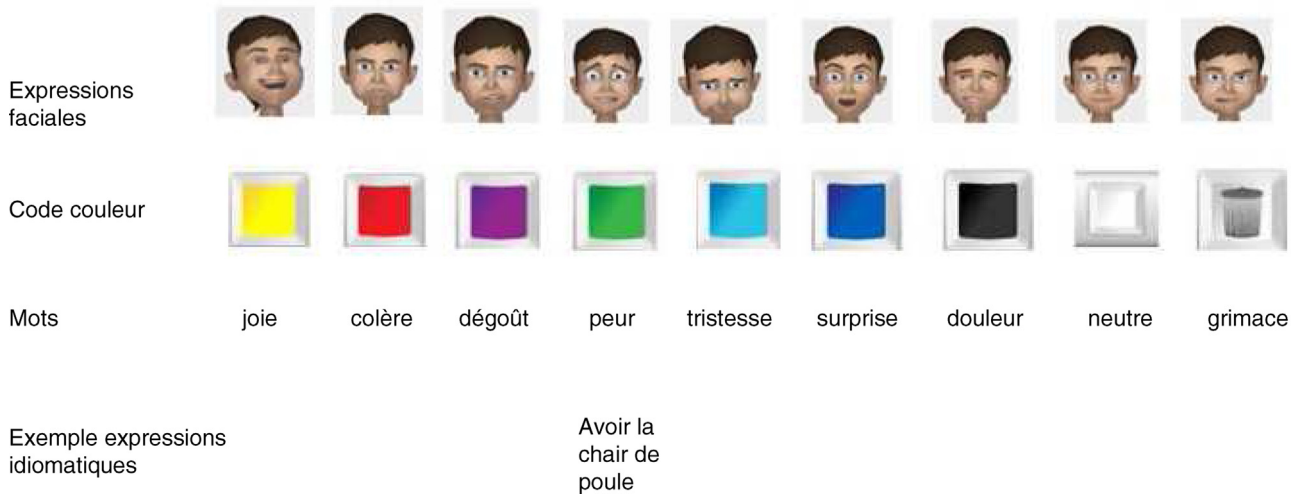


Fig. 1. Modalités de réponse proposées à l'enfant dans JeStiMule.

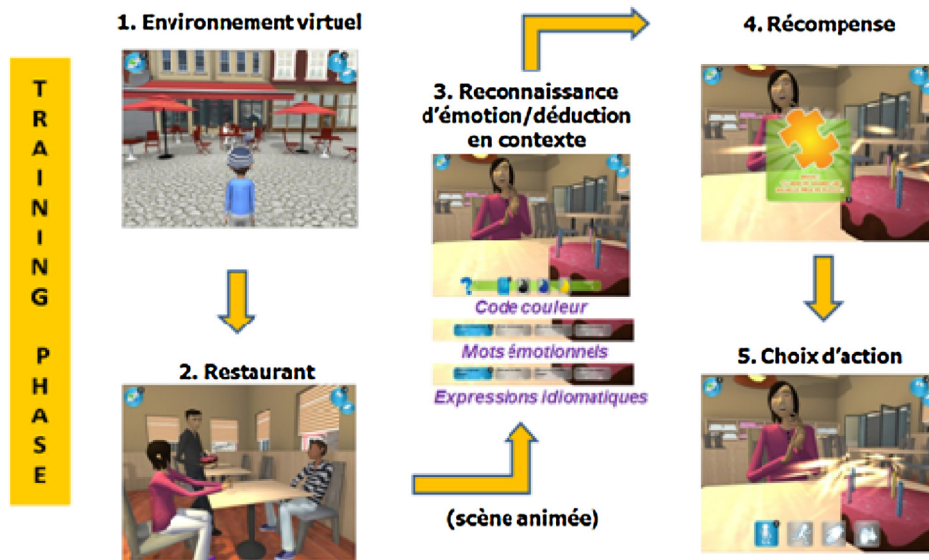


Fig. 2. Déroulé de la phase de jeu dans JeStiMule.

retrouvées grâce une recherche par mot clé sur PubMed. Les données recueillies dans ces bases sont insuffisantes pour la création d'un algorithme robuste nécessaire au jeu. En effet, la majorité de ces bases (6/7) sont construites à partir de photos [30–35] ce qui ne permet pas de rendre efficacement compte de la production des expressions faciales en situation écologique. Elles ne prennent pas en compte la notion de dynamique et de mouvement, forcément présente lors d'expressions faciales qui résultent du déplacement des composantes du visage. De plus, ces bases ne rendent compte que d'une vision en 2D des expressions faciales (photos et vidéos 2D) et ne contiennent donc pas les informations nécessaires au traitement 3D de l'émotion qui ne peuvent être recueillies que par une caméra adaptée. Cao, Weng, Zhou, Tong & Zhou (2013) ont réalisé des vidéos en 3D pour créer une base de données d'expressions faciales [36]. Cependant, cette base n'est pas conçue spécifiquement sur les enfants. La proportion d'enfants enregistrés n'est pas mentionnée, ne permettant pas de savoir si le nombre d'enfants vus

permet une fiabilité statistique. De plus, pour la plupart, les expressions capturées ne découlent pas d'émotions mais correspondent seulement à des déplacements d'éléments du visage (yeux fermés, bouche à gauche, etc.).

Au total, aucune des bases de données retrouvées ne permet d'obtenir des informations 2D et 3D sur le mouvement produit lors de la production d'expressions faciales chez un panel suffisant d'enfants. Afin de créer un système de détection robuste des expressions faciales chez l'enfant, il a donc été indispensable de construire au préalable une base de données d'expressions faciales chez l'enfant typique. Deux centres hospitaliers (le CHU de Nice et le CHU de la Pitié-Salpêtrière) sont impliqués dans la réalisation des enregistrements de productions émotionnelles nécessaires à la création de cette base de données (Tableau 2).

Le recrutement a permis d'inclure 153 enfants typiques âgés de 6 à 12 ans présentant des caractéristiques physiques diverses avec divers accessoires (lunettes, bijoux...) et différentes coupes de cheveux (mèche, frange, etc.). Une majorité de garçons ont

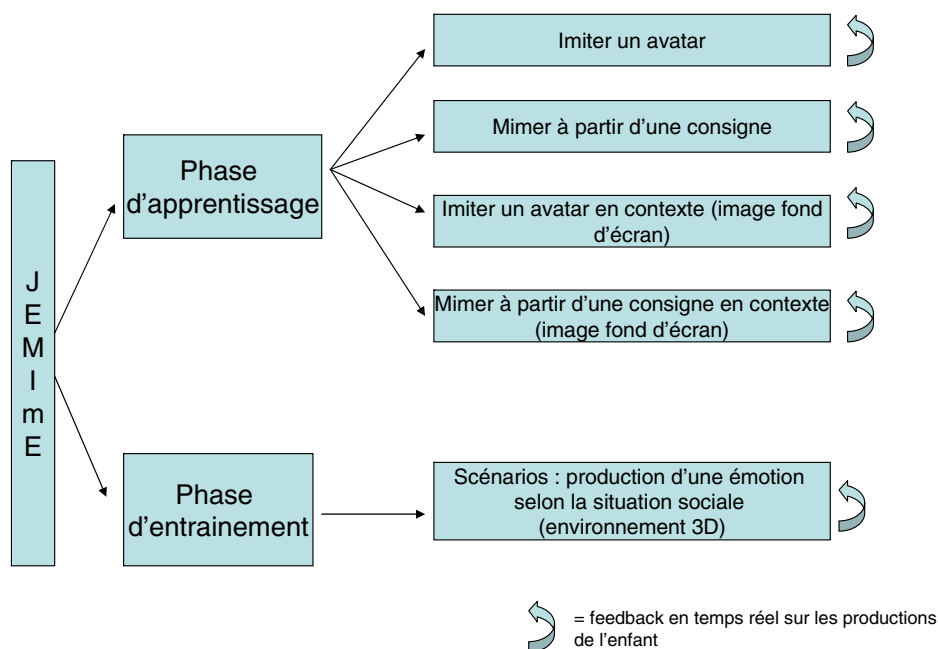


Fig. 3. Architecture globale de JEMImE.

Tableau 2
Principales caractéristiques des bases de données d'expressions faciales chez l'enfant.

| Bases de données | Population | Expression faciale | Support |
|--|--|--|--------------------------------|
| NIMH-ChEFS [31] | 39 filles et 20 garçons de 10 à 17 ans | Peur, colère, joie, tristesse, neutre | 482 photos |
| Dartmouth database [30] | 40 filles et 40 garçons caucasiens de 6 à 16 ans | Neutre, contentement, joie, tristesse, colère, peur, surprise, dégoût | Photos (nombre non communiqué) |
| Facewarehouse [36] | 150 individus de 7 à 80 ans (proportion d'enfants non précisée) | Bouche ouverte, front bas, front relevé, mâchoire à gauche et à droite, mâchoire en avant, bouche à gauche et à droite, menton relevé, lèvres plissées, bisous, tristesse, sourire, joie, souffle, yeux fermés, lèvres roulées, fossettes relevées | 3000 vidéos 3D |
| Facial expression database of children [32] | 53 garçons et 54 filles japonais de 11 à 13 ans | Neutre, joie, surprise, tristesse, colère | 535 photos |
| Slides depicting facial expression of affect [35] | 3 garçons (9 à 11 ans) et 3 filles (8 à 12 ans) | Colère, dégoût, peur, joie, surprise, neutre | Photos (nombre non communiqué) |
| CAFE : the child affective facial expressions [34] | 90 filles et 64 garçons âgés de 2 à 8 ans, de différentes origines | Colère, peur, tristesse, joie, surprise, dégoût | 1192 Photos |
| Radboud faces database [33] | 49 personnes caucasiennes dont 10 enfants (6 filles) de 7 à 11 ans | Neutre, colère, tristesse, peur, dégoût, surprise, joie, contentement | 1200 photos pour les enfants |

été inclus dans l'étude afin d'approcher le ratio garçon/fille de la population avec autisme. L'accord de la CNIL, celui des écoles, des parents et des enfants ont été obtenus pour le projet. Il est également prévu d'inclure des enfants avec TSA afin de pouvoir tester l'efficacité de l'algorithme auprès de cette population. Les patients inclus devront avoir entre 6 et 12 ans et ne pas présenter de retard cognitif. L'inclusion est en cours actuellement après accord de l'étude par le comité de protection des personnes du CHU de Nice (N°C 2015-A00929-40 du 30/09/2015).

Les enregistrements de productions émotionnelles ont été réalisés grâce à un outil d'acquisition créé spécifiquement pour le projet (Fig. 4). Les enfants ont été filmés par courtes séquences (environ 30 sec à 1 min) grâce à une caméra 3D/2D. Chaque séquence filmée contient une émotion. Trois flux (une vidéo couleur, une séquence contenant l'enchaînement des cartes de profondeur et le flux audio) sont extraits et serviront à la création de l'algorithme. Trois tâches (mime d'émotion, imitation d'émotion et récit d'une histoire spontanée à partir du

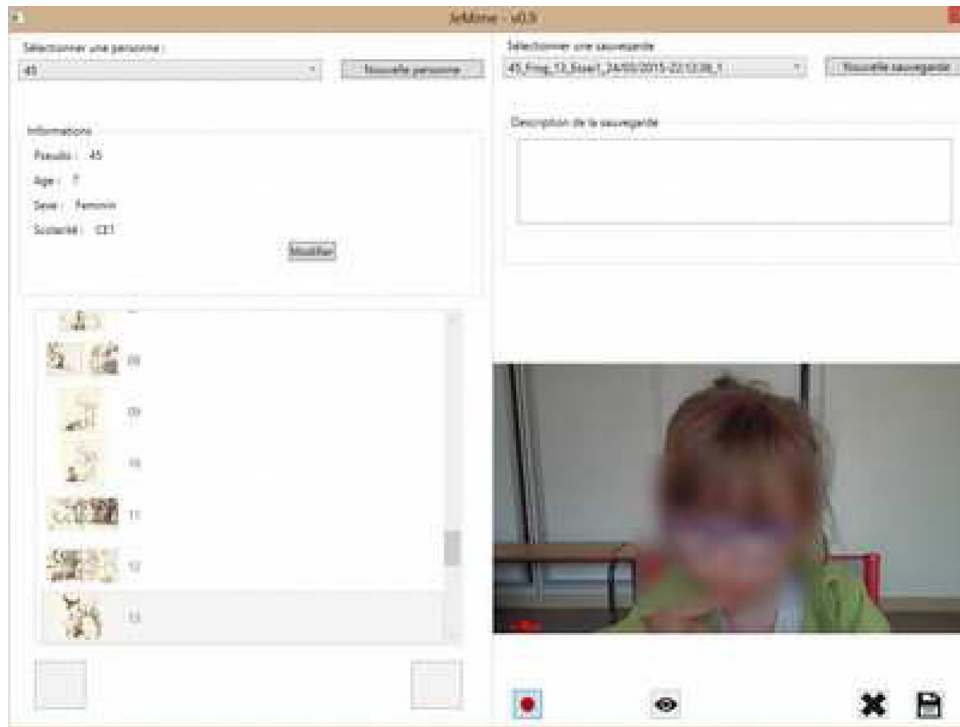


Fig. 4. Outil d'acquisition. Il permet à l'examinateur de contrôler les stimuli présentés à l'enfant et d'avoir un retour en temps réel des productions de l'enfant.

livre d'images "Frog where are you?") ont été proposées aux enfants. Ces trois tâches permettent de recueillir des productions avec et sans modèle ainsi qu'une production spontanée afin de pouvoir obtenir plus de diversité dans les productions d'émotion. Les tâches de mime et d'imitation ont d'abord été proposées en condition visuelle seule, puis en condition audiovisuelle (faciale et vocale). Ces deux modalités ont été introduites afin de prendre en compte les productions faciales avec et sans bouche ouverte selon qu'un son est produit ou non. Ces activités travaillent 4 expressions faciales : joie, colère, tristesse et neutre. Chaque enfant réalise chaque émotion 2 fois en mime et 4 fois en imitation. La dernière condition, le récit d'histoire, permet d'enregistrer les productions spontanées faciales et vocales des enfants. Chaque image est classée comme porteuse d'une émotion positive, négative ou neutre [37]. Lors des enregistrements, les tâches de mime et d'imitation ont été contrebalancées afin d'éviter de démarrer systématiquement par la tâche d'imitation et ainsi de créer un effet d'apprentissage. À l'intérieur de chaque tâche, les émotions sont présentées de façon aléatoire. En revanche, comme dans nos analyses, les productions sur stimulus visuel et audiovisuel sont traitées séparément, il n'était pas nécessaire de les contrebalancer. La stimulation audiovisuelle étant plus complexe, nous l'avons systématiquement proposée en second. Le récit d'histoire intervenait quant à lui toujours en dernier. Enfin, l'ordre de présentation des tâches a été contrebalancé en fonction du sexe et de l'âge des enfants. On obtient ainsi par enfant 16 enregistrements en imitation, 8 en mime et 24 lors du récit (Fig. 4).

6. Annotation de la base de données

Afin d'entraîner l'algorithme de reconnaissance, il est nécessaire que les vidéos soient auparavant annotées par des juges. Différents outils d'annotation [30,31] ont déjà été créés pour la notation de base de données d'expressions faciales chez l'enfant. Tous demandent en premier lieu d'identifier l'émotion produite par l'enfant, puis de juger de l'intensité de cette émotion. Cependant, dans la création de la nouvelle base de données, il a été choisi d'écarter la notion d'intensité, JEMImE ayant pour but d'apprendre aux enfants avec TSA à produire une expression faciale naturelle, adaptée et cohérente avec le contexte. Pour cela, l'émotion produite n'a pas besoin d'être intense ; elle doit seulement être facilement reconnaissable. Il a donc été nécessaire de se pencher sur la définition de ce qu'est une émotion de qualité. Cette notion de qualité de l'émotion ne fait pas consensus chez les auteurs à l'origine des bases de données de l'enfant. Ainsi, Egger et al. (2011) propose au cotateur de noter « à quel point l'expression est bien faite » par l'enfant [31]. Mazurski et al. (1993) demandent quant à eux à l'annotateur de juger à quel point il est certain que l'émotion qu'il a reconnue est bien celle produite par l'enfant [35]. Chez l'adulte, on retrouve encore d'autres possibilités de cotation, comme dans la base GEMEP Corpus pour laquelle il faut juger de l'authenticité et de la plausibilité de l'expression faciale [38].

Aucune définition de qualité ne se dégage réellement. Pour coter les vidéos de la base de données, nous avons choisi d'évaluer la qualité comme une combinaison de la reconnaissance et de la crédibilité de l'expression faciale produite.

Postulant qu'une émotion non reconnue ne peut pas être considérée comme crédible, il est possible d'envisager un continuum entre reconnaissance et crédibilité. Chaque expression peut donc être annotée sur une échelle allant de l'expression « non reconnue » à « reconnue et crédible » en passant par « reconnue mais non crédible ». Comme les autres outils existants, cette échelle permet donc de juger quelle émotion est produite (« non reconnue » versus « reconnue ») et en même temps de la crédibilité de l'émotion (« non crédible » versus « crédible »). Un outil d'annotation a été créé spécialement pour juger de la qualité des expressions faciales formant notre base de données. Il est composé de 4 échelles d'évaluation, soit une échelle par émotion (joie, tristesse, colère, neutre). Afin de vérifier l'utilisabilité de cet outil, deux cent quarante vidéos ont été annotées par deux juges indépendamment après une période d'apprentissage sur quelques vidéos. Les deux juges sont des femmes appartenant chacune aux deux équipes cliniques impliquées. Chaque vidéo donnait lieu à une quadruple notation (une note par émotion) allant de 0 à 8 (où 0 correspond à « émotion non reconnue », 4 à « émotion reconnue mais non crédible » et 8 à « émotion reconnue et crédible ») : une note sur la qualité de la joie, de la colère, de la tristesse et de l'expression neutre. Du fait de la nature de la notation (allant de 0 à 8), nous avons mesuré l'accord interjuge à l'aide des coefficients de corrélations intra-classe (ICC) pour chaque émotion. On obtient ainsi les valeurs suivantes pour la joie : ICC=0,93, la colère : ICC=0,92, la tristesse : ICC=0,93, et l'expression neutre : ICC=0,93. La reproductibilité interjuges est donc très bonne (>0,8) quelle que soit l'émotion considérée.

7. Création d'un algorithme de reconnaissance émotionnelle

Afin d'évaluer automatiquement la production émotionnelle des enfants, il est nécessaire de développer un algorithme d'analyse faciale et vocale des émotions. Cet algorithme s'appuie sur une méthode d'apprentissage statistique à partir d'exemples issus d'une base de données. Pour l'algorithme de JEMImE, nous avons donc utilisé la base de données d'expressions faciales que nous avons créée. Celle-ci est composée d'enfants typiques et leurs productions correspondront donc aux références à atteindre par les enfants avec TSA. On distingue donc deux phases, la phase d'entraînement (ou d'apprentissage) de l'algorithme d'une part et la phase de prédiction d'autre part.

Durant phase d'entraînement, l'algorithme va apprendre à prédire la qualité d'une expression faciale. Cela consiste dans un premier temps à extraire un ensemble de caractéristiques à partir des flux 2D, 3D et audio de la base de données : par exemple, pour le visage, des mesures d'angles et de distances entre des points caractéristiques localisés sur le visage (commissures des lèvres et des yeux, extrémité des sourcils...) ou des changements de textures liés à l'apparition de rides. Pour l'audio, on extrait par exemple des variations d'intensité associées à certaines plages de fréquence de la voix. Dans un second temps, l'algorithme d'apprentissage va construire un modèle prédictif capable de sélectionner parmi ces caractéristiques celles qui sont pertinentes et d'y associer une mesure de qualité de l'émotion

produite. Cela est réalisé en ajustant les paramètres du modèle qui minimisent les erreurs de prédiction par rapport aux annotations fournies par les juges (Fig. 5A).

Lors de la seconde phase de prédiction, lorsque l'enfant joue avec le jeu sérieux, l'algorithme d'analyse des expressions faciales récupère en temps réel les flux issus du capteur multimodal, extrait les caractéristiques pertinentes et évalue la qualité de la production émotionnelle à l'aide du modèle prédictif appris précédemment. L'information est alors retransmise au jeu sérieux dont le scénario s'adapte en fonction de la performance de l'enfant (Fig. 5B). L'algorithme d'analyse des expressions faciales est décrit en détail dans Dapogny et al. (2016) [39], celui des flux audio dans Ringeval et al. (2016) [40] (Fig. 5).

8. Architecture globale de JEMImE : détails

Dans cette partie, nous détaillons l'architecture générale du jeu, considérant que nous avons un outil automatisé de mesure des productions émotionnelles (faciales et vocales) rapide et performant. Par rapport à JeStiMule, on conserve la construction du jeu en 2 phases. L'objectif de la première phase (phase d'apprentissage) est « d'enseigner » à l'enfant avec TSA à produire des expressions faciales et vocales de bonne qualité en accord avec l'émotion attendue. Elle comporte 2 jeux principaux, subdivisés ensuite selon les paramètres de modalité (visuelle ou audiovisuelle) et la présence ou non d'un contexte. Cette phase devrait proposer 2 façons d'évoluer parmi ces jeux : un mode par défaut avec un enchaînement des jeux prédéfinis et un mode personnalisé pendant lequel le joueur pourra choisir un mini-jeu et le configurer selon ses objectifs. Lors de chaque jeu, la production d'une émotion se fera selon le déroulement suivant : dans un premier temps, l'enfant regarde l'avatar (imitation) ou le mot (mime) qui s'affiche à l'écran ; dans un second temps, l'enfant produit l'expression attendue dans le temps imparti. À ce moment, l'enfant voit apparaître sur l'écran 3 jauges (1 jaune pour la joie, 1 rouge pour la colère, 1 bleue pour la tristesse) qui montent ou descendent en fonction de l'émotion reconnue par l'algorithme. Il a ainsi un retour visuel sur ses productions qui lui permet de s'ajuster en temps réel. Une jauge sous l'écran (sous forme de barre) lui permet de visualiser le temps restant pour produire l'émotion. Trois autres jauges lui permettent de suivre sa progression dans le jeu en visualisant ses réussites par émotion : chaque émotion correctement produite engendre le gain d'une pièce de monnaie ; il est nécessaire d'atteindre un seuil de réussite (de gain de pièce) pour valider le jeu. À chaque transition, une fenêtre de dialogue apparaît. Les boutons de validation sont toujours verts et placés à droite de la fenêtre, ceux de refus toujours rouges et placés à gauche de la fenêtre de dialogue. Ainsi, à la fin de la phase d'apprentissage, l'enfant avec TSA devrait être en capacité de produire correctement les 3 émotions travaillées.

Comme dans JeStiMule, la deuxième phase a pour objectif de mettre l'enfant avec TSA en situation la plus écologique possible afin qu'il utilise les apprentissages effectués lors de la première phase. Cette phase se déroulera dans un environnement semblable à celui présent dans JeStiMule (une plateforme avec plusieurs lieux comme la ville, le jardin, le restaurant).

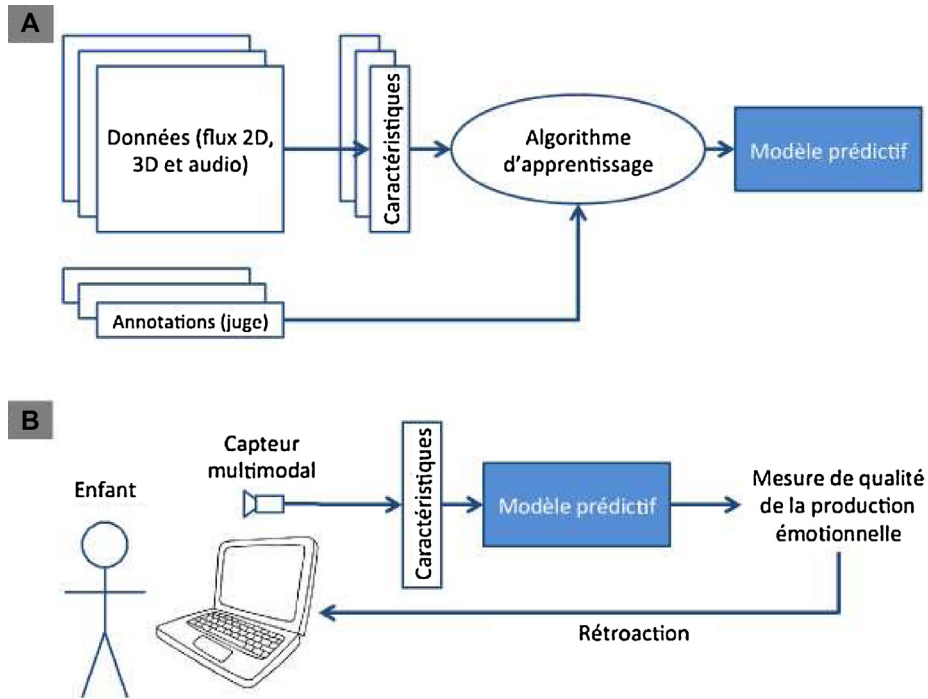


Fig. 5. Principes régissant la création de l'algorithme de reconnaissance émotionnelle : A) phase d'apprentissage ; B) phase de prédiction.

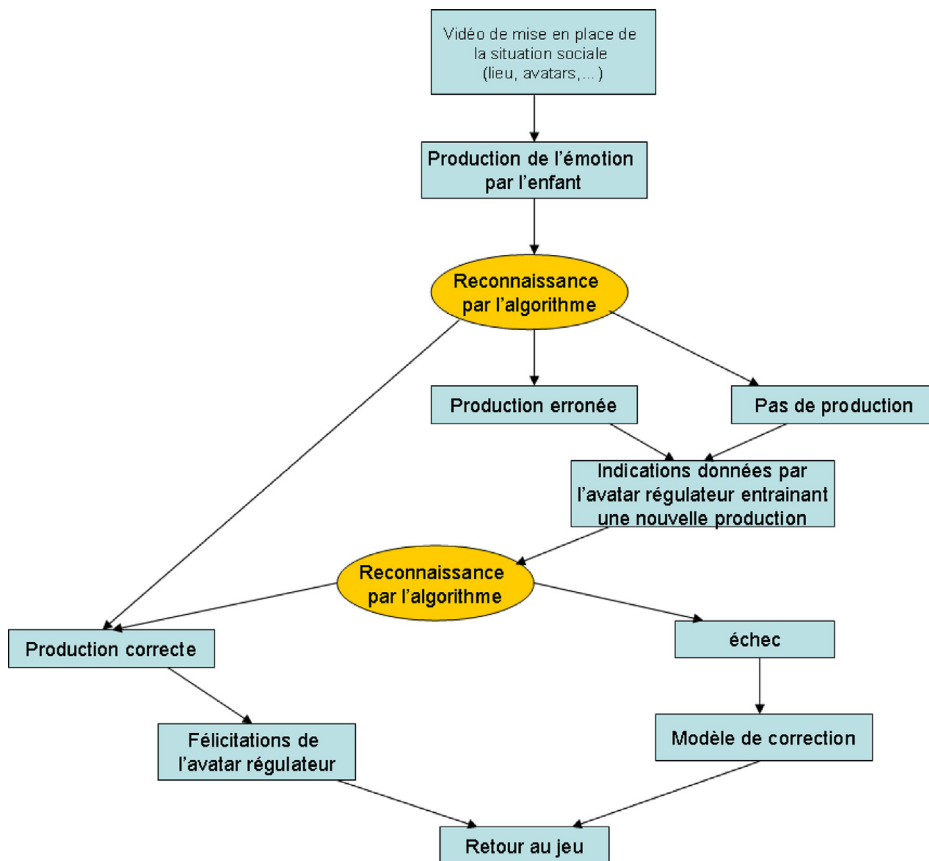


Fig. 6. Architecture des scénarios de JEMImE.

L'enfant avec TSA pourra y déplacer son propre avatar qu'il aura au préalable personnalisé. Dans l'environnement, l'enfant sera confronté à deux types d'interactions :

- les interactions avec des objets non pertinents pour augmenter l'aspect ludique du jeu. Ces évènements inattendus motivent les enfants à revenir et les encouragent à explorer la plateforme. Cependant, afin d'éviter que l'enfant ne passe trop de temps à les manipuler, ces objets ne sont activables qu'un nombre limité de fois ;
- les interactions avec des avatars non joueurs. L'enfant déclenche alors la mise en route de scénarios.

Les étapes des scénarios sont représentées dans la Fig. 6. L'enfant se retrouve confronté à une situation sociale. Il est alors amené à produire une expression faciale et/ou vocale. Durant sa production, l'enfant reçoit un feedback en temps réel de l'expression qu'il produit. Quelques ajouts ont été apportés par rapport à JeStiMulE. Ainsi, un avatar « régulateur », sorte de modérateur, accompagne l'enfant dans les 2 phases du jeu. L'avatar régulateur (choisit par le joueur parmi plusieurs propositions) donne les consignes au joueur, lui signale quand le jeu va démarrer, l'incite à répondre, le félicite quand il a gagné. En outre, plusieurs statistiques sont récupérées durant le jeu. Durant toute la phase 1, il sera ainsi possible de savoir :

- le pourcentage de réussites pour chaque émotion ;
- le temps moyen mis par l'enfant pour produire la bonne expression pour chaque émotion ;
- le nombre de fois où l'enfant a effectué un mini-jeu.

Durant toute la phase 2, le jeu récupère des statistiques en fonction des actions de l'enfant. Il est ainsi possible de savoir :

- le nombre de fois où une émotion est travaillée pour chaque scénario ;
- le nombre de réussites pour une émotion pour chaque scénario ;
- le temps global passé sur le jeu ;
- le temps passé sur les scénarios ;
- le score total (réussites/nombre de passation) pour chaque émotion tous scénarios confondus ;
- le temps de latence avant production de la bonne émotion ;
- le nombre de fois et le temps que les joueurs passent à manipuler des éléments non pertinents par rapport aux objectifs du jeu, par exemple actionner des éléments du décor (gonfler les ballons) (Fig. 6).

9. Perspectives

Une fois mis au point l'algorithme pour évaluer en temps réel les productions émotionnelles, nous pourrions étudier dans une population d'enfants avec TSA si des différences sont présentes avant entraînement à partir des bases de données créées. Pour ce faire, il est désormais nécessaire de proposer les tâches de mime, d'imitation et de lecture d'histoire à des patients avec TSA. Les passations sont en cours et les vidéos recueillies permettront

d'appliquer l'algorithme sur les productions de ces patients dans des conditions analogues à celles des enfants typiques.

Lorsque le développement du jeu sera complété (choix des avatars, écriture des scénarios, design...) auquel nous intégrerons l'algorithme de reconnaissance, nous pourrions tester sa fiabilité et son ergonomie dans une étude pilote (notamment son bon fonctionnement quant au feedback donné par l'algorithme) et effectuer les ajustements nécessaires en terme de programmation ; pour ensuite tester son utilité thérapeutique dans une étude clinique contrôlée auprès d'une population de patients avec TSA.

Remerciements

Ce travail est financé par l'Agence nationale pour la recherche (ANR) dans le cadre de son programme CONTINT (JEMImE, numéro de projet ANR-13-CORD-0004).

Déclaration de liens d'intérêts

Les auteurs Charline Grossard, Stéphanie Hun, Sylvie Serret, Ouriel Grynszpan, Arnaud Dapogny, Kevin Bailly, Laurence Chaby, David Cohen déclarent ne pas avoir de liens d'intérêts.

Pierre Foulon est directeur général d'e-Genious, société privée qui développe JeMime. Il n'est pas prévu de commercialisation en l'absence d'étude de validité.

Références

- [1] Uljarevic M, Hamilton A. Recognition of emotions in autism: a formal meta-analysis. *J Autism Dev Disord* 2013;43(7):1517–26.
- [2] McIntosh DN, Reichmann-Decker A, Winkelman P, Wilbarger JL. When the social mirror breaks: deficits in automatic, but not voluntary, mimicry of emotional facial expressions in autism. *Dev Sci* 2006;9(3):295–302.
- [3] Durkin K. Videogames and young people with developmental disorders. *Rev Gen Psychol* 2010;14(2):122–40. <http://dx.doi.org/10.1037/a0019438>.
- [4] Bernard-Opitz V, Sriram N, Nakhoda-Sapuan S. Enhancing social problem solving in children with autism and normal children through computer-assisted instruction. *J Autism Dev Disord* 2001;31(4):377–84.
- [5] Mitchell P, Parsons S, Leonard A. Using virtual environments for teaching social understanding to 6 adolescents with autistic spectrum disorders. *J Autism Dev Disord* 2007;37(3):589–600. <http://dx.doi.org/10.1007/s10803-006-0189-8>.
- [6] Knight V, McKissick BR, Saunders A. A review of technology-based interventions to teach academic skills to students with autism spectrum disorder. *J Autism Dev Disord* 2013;43(11):2628–48. <http://dx.doi.org/10.1007/s10803-013-1814-y>.
- [7] Josman N, Ben-Chaim H, Friedrich S, Weiss P. Effectiveness of virtual reality for teaching street-crossing skills to children and adolescents with autism. *Int J Disabil Hum Devel* 2008;7(1):49–56.
- [8] Grynszpan O, Weiss PLT, Perez-Diaz F, Gal E. Innovative technology-based interventions for autism spectrum disorders: a meta-analysis. *Autism* 2014;18(4):346–61.
- [9] Park JH, Abirached B, Zhang Y. A framework for designing assistive technologies for teaching children with ASDs emotions. In: CHI'12 extended abstracts on human factors in computing systems. New York: ACM; 2012. p. 2423–8.
- [10] Queirós C, Alves S, Marques AJ, Oliveira M, Orvalho V. Serious games and emotion teaching in autism spectrum disorders: a comparison with LIFEisGAME project; 2012 [Available from: URL: <https://repositorio-aberto.up.pt/bitstream/10216/64635/2/92142.pdf>, serial online, cited 2016 August 2].

- [11] Serret S, Hun S, Iakimova G, Lozada J, Anastassova M, Santos A, et al. Facing the challenge of teaching emotions to individuals with low-and high-functioning autism using a new serious game: a pilot study. *Mol Autism* 2014;5(1):1.
- [12] Grossard C, Grynspan O, Serret S, Jouen AL, Bailly K, Cohen D. Serious games to teach social interactions and emotions to individuals with autism spectrum disorders (ASD). Submitted.
- [13] Fernandes T, Alves S, Miranda J, Queirós C, Orvalho V, LIFEisGAME: A facial character animation system to help recognize facial expressions. In: Enterprise information systems. Berlin: Springer; 2011. p. 423–32.
- [14] Tan CT, Harrold N, Rosser D. Can you CopyMe?: an expression mimicking serious game. In: SIGGRAPH Asia 2013 symposium on mobile graphics and interactive applications. New York: ACM; 2013. p. 73.
- [15] Cockburn J, Bartlett M, Tanaka J, Movellan J, Pierce M, Schultz R, et al. A tutoring system in real-time facial expression perception and production in children with autism spectrum disorder. In: ECAG 2008 workshop facial and bodily expressions for control and adaptation of games. 2008.
- [16] Connolly TM, Boyle EA, MacArthur E, Hainey T, Boyle JM. A systematic literature review of empirical evidence on computer games and serious games. *Comput Educ* 2012;59(2):661–86.
- [17] Yusoff A. A conceptual framework for serious games and its validation [Thèse de doctorat d'université]. Southampton: University of Southampton; 2010. Available: <http://eprints.soton.ac.uk/171663/1.hasCoversheetVersion/Thesis.017b.pdf> dernier accès le 2 août 2016.
- [18] Tanaka JW, Wolf JM, Klaiman C, Koenig K, Cockburn J, Herlihy L, et al. Using computerized games to teach face recognition skills to children with autism spectrum disorder: the Let's Face It! program. *J Child Psychol Psychiatry* 2010;51(8):944–52. <http://dx.doi.org/10.1111/j.1469-7610.2010.8.x>.
- [19] Beaumont R, Sofronoff K. A multi-component social skills intervention for children with Asperger syndrome: the junior detective training program. *J Child Psychol Psychiatry* 2008;49(7):743–53.
- [20] Silver M, Oakes P. Evaluation of a new computer intervention to teach people with autism or Asperger syndrome to recognize and predict emotions in others. *Autism* 2001;5(3):299–316.
- [21] Hopkins IM, Gower MW, Perez TA, Smith DS, Amthor FR, Wimsatt FC, et al. Avatar assistant: improving social skills in students with an ASD through a computer-based intervention. *J Autism Dev Disord* 2011;41(9):1543–55.
- [22] Golan O, Baron-Cohen S. Systemizing empathy: teaching adults with Asperger syndrome or high-functioning autism to recognize complex emotions using interactive multimedia. *Dev Psychopathol* 2006;18(02):591–617.
- [23] Golan O, Ashwin E, Granader Y, McClintock S, Day K, Leggett V, et al. Enhancing emotion recognition in children with autism spectrum conditions: An intervention using animated vehicles with real emotional faces. *J Autism Dev Disord* 2010;40(3):269–79.
- [24] Copenhagen Marfisi-Schottman I, George S, Tarpin-Bernard F. Proceedings of the 4th European Conference on games based learning ECGBL. Tools and methods for efficiently designing serious games. 2010. p. 226–34.
- [25] Marne B, Wisdom J, Huynh-Kim-Bang B, Labat JM. The six facets of serious game design: a methodology enhanced by our design pattern library. In: 21st century learning for 21st century skills. Berlin, Heidelberg: Springer; 2012. p. 208–21.
- [26] Khowaja K, Salim SS, editors. International symposium on simulation and serious games 2014. Design components for serious games of children with autism spectrum disorder (ASD). 2014. <http://dx.doi.org/10.3850/978-981-09-0463-0.061>.
- [27] Bolte S, Hubl D, Feineis-Matthews S, Prvulovic D, Dierks T, Poustka F. Facial affect recognition training in autism: can we animate the fusiform gyrus? *Behav Neurosci* 2006;120(1):211–6. <http://dx.doi.org/10.1037/0735-7044.120.1.211>.
- [28] Houstis O, Kiliaridis S. Gender and age differences in facial expressions. *Eur J Orthod* 2009;31(5):459–66.
- [29] Riviello MT, Chetouani M, Cohen D, Esposito A. On the perception of emotional “voices”: a cross-cultural comparison among American, French and Italian subjects. In: Springer, editor. Analysis of verbal and nonverbal communication and enactment. The processing issues Berlin: Springer; 2011. p. 368–77.
- [30] Dalrymple KA, Gomez J, Duchaine B. The Dartmouth database of children's faces: acquisition and validation of a new face stimulus set. *PLoS One* 2013;8(11):e79131.
- [31] Egger HL, Pine DS, Nelson E, Leibenluft E, Ernst M, Towbin KE, et al. The NIMH child emotional faces picture set (NIMH-ChEFS): a new set of children's facial emotion stimuli. *Int J Methods Psychiatr Res* 2011;20(3):145–56.
- [32] Komatsu S, Hakoda Y. Construction and evaluation of a facial expression database of children. *Shinrigaku Kenkyu* 2012;83(3):217–24.
- [33] Langner O, Dotsch R, Bijlstra G, Wigboldus DH, Hawk ST, Van Knippenberg A. Presentation and validation of the Radboud faces database. *Cogn Emotion* 2010;24(8):1377–88.
- [34] LoBue V, Thrasher C. The child affective facial expression (CAFE) set: validity and reliability from untrained adults. *Front Psychol* 2014;5:1532.
- [35] Mazurski EJ, Bond NW. A new series of slides depicting facial expressions of affect: a comparison with the pictures of facial affect series. *Aust J Psychol* 1993;45(1):41–7.
- [36] Cao C, Weng Y, Zhou S, Tong Y, Zhou K. Facewarehouse: a 3d facial expression database for visual computing. *IEEE Trans Vis Comput Graph* 2013;20(3):413–25.
- [37] Ringeval F, Demouy J, Szaszák G, Chetouani M, Robel L, Xavier J, et al. Automatic intonation recognition for the prosodic assessment of language-impaired children. *IEEE Trans Audio Speech Lang Process* 2011;19(5):1328–42.
- [38] Bänziger T, Mortillaro M, Scherer KR. Introducing the Geneva multimodal expression corpus for experimental research on emotion perception. *Emotion* 2012;12(5):1161.
- [39] Dapogny A, Bailly K, Dubuisson S. Pairwise conditional random forests for facial expression recognition. *IEEE Int Conf Comput Vision* 2015:3783–91.
- [40] Ringeval F, Marchi E, Grossard C, Xavier J, Chetouani M, Cohen D, et al. Automatic analysis of typical and atypical encoding of spontaneous emotion in the voice of children. *Compare* 2016;4366(397):1610.