

TABLE DES MATIÈRES

	Page
INTRODUCTION.....	1
CHAPITRE 1 ÉTAT DE L'ART.....	3
1.1 Description et mesure de l'expression faciale.....	3
1.2 Expressions faciales liées à la parole.....	4
1.3 Expressions faciales liées aux émotions.....	4
1.4 Expressions faciales liées à l'activité physique.....	6
CHAPITRE 2 ACQUISITION DES DONNÉES.....	7
2.1 Capture de mouvement.....	7
2.2 Capture des expressions faciales.....	9
CHAPITRE 3 SYNTHÈSE DE L'EXPRESSION FACIALE.....	11
3.1 Observations sur les données acquises.....	12
3.2 Modèle biomécanique.....	16
3.3 Apprentissage machine des expressions faciales.....	17
3.3.1 Prédiction des paramètres métaboliques.....	17
3.3.2 Prédiction des composantes de l'expression faciale.....	22
CHAPITRE 4 DISCUSSION.....	27
4.1 Capture de mouvement.....	27
4.1.1 Système de mocap utilisé.....	27
4.1.2 Costumes et marqueurs utilisés.....	28
4.1.3 Matériel d'enregistrement des paramètres métaboliques.....	29
4.2 Capture faciale.....	32
4.3 Acquisition manuelle des expressions faciales.....	34
4.4 Prototype.....	35
4.5 Limitations.....	36
CONCLUSION.....	41
LISTE DE RÉFÉRENCES.....	42

LISTE DES TABLEAUX

	Page
Tableau 3.1	Combinaisons d'entrées / sorties évaluées (modèle métabolique) 17
Tableau 3.2	Tableau des paramètres d'expressivité utilisés dans la figure 3.7 24
Tableau 3.3	Combinaisons d'entrées / sorties évaluées (modèles des <i>blend shapes</i>) 25
Tableau 3.4	Erreur RMSE des modèles de prédiction des <i>blend shapes</i> 26

LISTE DES FIGURES

	Page
Figure 2.1	Types de captures effectuées 8
Figure 3.1	Vue générale de l'approche..... 11
Figure 3.2	Outils utilisés pour l'analyse des données 13
Figure 3.3	Exemples d'expressions faciales induites par l'effort physique 15
Figure 3.4	Vue générale des modèles 18
Figure 3.5	Comparaison entre le rythme cardiaque réel et prédit..... 20
Figure 3.6	Comparaison entre une composante de l'expression faciale réelle et prédite 23
Figure 3.7	Résultats obtenus avec différents paramètres d'expressivité..... 24
Figure 3.8	Résultats obtenus avec différents niveaux d'entraînement..... 25
Figure 4.1	Studio de mocap 28
Figure 4.2	Différentes configurations de costumes et systèmes de fixation 30
Figure 4.3	Caméra témoin et caméras de capture de mouvement 31
Figure 4.4	Caméras et systèmes de fixation utilisés 33
Figure 4.5	Système utilisé pour l'animation en poses clés 35
Figure 4.6	<i>Blend shapes</i> utilisées dans l'implémentation 36
Figure 4.7	Prototype réalisé 37
Figure 4.8	Comparaison entre les expressions faciales réelles et prédites 38

LISTE DES ABRÉVIATIONS, SIGLES ET ACRONYMES

10-fold	Technique de validation croisée
AU	<i>Action Units</i>
CDM	Centre de masse
CPU	<i>Central Processing Unit</i> ou processeur
ECG	Électrocardiogramme
EPOC	<i>Excess Post Oxygen Consumption</i>
ÉTS	École de technologie supérieure
FACES	<i>Facial Expression Coding System</i>
FACS	<i>Facial Action Coding System</i>
FAP	<i>Facial Action Parameter</i>
GPU	<i>Graphics Processing Unit</i> ou processeur graphique
HRR	<i>Heart Rate Reserve</i> ou réserve de rythme cardiaque
IMC	Indice de masse corporelle
LARTech	Laboratoire de recherche-cr�ation en technochor�graphie
libSVM	Librairie pour les machines � vecteurs de support (SVM)
mocap	<i>Motion capture</i>
p	Puissance m�canique : d�riv�e du travail m�canique
RBF	<i>Radial Basis Function</i>
RMSE	<i>Root Mean Square Error</i>
R-R	Intervalle entre deux ondes R du signal �lectrique d�tect� par un ECG
SETT	<i>Subtil Expression Training Tool</i>
SVM	<i>Support Vector Machine</i>
w	Travail m�canique : variation de l'�nergie m�canique du syst�me

INTRODUCTION

Que cela soit dans le domaine du jeu vidéo, du cinéma d'animation ou des effets spéciaux, l'animation 3D gagne de plus en plus de réalisme notamment avec l'utilisation de la capture de mouvement (mocap). Cependant, l'animation faciale reste l'un des aspects les plus délicats, coûteux et nécessitant le plus de temps d'élaboration. L'expression faciale est difficile à modéliser car non seulement elle implique un grand nombre de groupes musculaires mais aussi qu'elle est induite par divers facteurs :

- Émotions : joie, tristesse, surprise, *etc*;
- Mouvements de la bouche : parole, déglutition, respiration profonde, *etc*;
- Mouvements des yeux et des paupières : clignement, direction du regard, réaction à une luminosité intense, *etc*;
- Physiologiques : effort, fatigue, douleur, *etc*.

Il existe différentes approches permettant d'automatiser les expressions faciales liées aux émotions ou à la parole, mais il n'y a pas de méthodes permettant d'automatiser les expressions liées à l'activité physique. Dans les communautés d'effets visuels et de l'animation par ordinateur, les animations faciales sont générées par clés ou en utilisant la capture de mouvements. Bien que cette dernière technique soit relativement complexe et coûteuse, il s'agit le plus souvent d'un choix logique pour les personnages principaux. Pour les personnages secondaires par contre, une foule par exemple, l'animation faciale liée à l'activité physique est le plus souvent souvent négligée.

Actuellement dans l'industrie du jeu vidéo, bien que les personnages 3D doivent souvent fournir des efforts physiques importants, l'animation faciale liée à ce facteur est quelque peu négligée. Elle est parfois présente lors des séquences de cinématiques, mais pendant le jeu, elle est le plus souvent tout simplement négligée ou elle souffre d'une grossière approximation. Les animations faciales liées à l'activité physique les plus élaborées pendant le jeu reposent sur de très simples approches *ad hoc* basées sur des fonctions linéaires et des seuils ; ceci reste

loin de la complexité de l'expression faciale liée à l'activité physique que les personnages 3D pratiquent.

Nous proposons à travers cette étude une nouvelle approche basée sur l'apprentissage machine permettant de simuler les expressions faciales liées à l'activité physique, afin d'améliorer le réalisme des animations 3D. L'approche est basée sur l'analyse des données de capture de mouvement acquises lors de séances d'exercices réels. Selon le type d'activité contenu dans les données capturées, des techniques d'apprentissage machine ont été sélectionnées afin de permettre la synthèse des expressions faciales correspondant à ces activités. Les modèles développés dans ce travail fournissent également des données métaboliques pouvant être utilisées afin d'automatiser des aspects autres que l'expression faciale tels que le niveau de transpiration, de mouvement d'expiration, *etc.* Les principales contributions de l'approche proposée peuvent être résumées comme suit :

- Synthèse de résultats réalistes à partir des modèles d'apprentissage machine entraînés sur des données réelles ;
- Identification et analyse des principales caractéristiques de l'expression faciale liée à l'activité physique ;
- Conception de modèles pouvant être utilisés en temps réel ainsi que pour de l'animation faciale hors-ligne ;
- Contrôle sur la stylisation grâce à la possibilité d'ajuster ou remplacer les données d'entraînement ;
- Contrôle sur l'expressivité grâce aux paramètres simples ayant un impact sur l'expression faciale du personnage 3D ;
- Prédiction des paramètres métaboliques pouvant être utilisés pour d'autres aspects tels que la respiration ou la transpiration.

CHAPITRE 1

ÉTAT DE L'ART

L'animation faciale est parmi les tâches les plus difficiles en infographie. Cela est principalement dû à la complexité des structures musculaires du visage humain et au fait que nous sommes habitués au quotidien à analyser, consciemment ou inconsciemment, les changements subtils dans les expressions faciales. Animer un modèle 3D d'un visage humain devient alors un réel défi tant sur le plan technique qu'artistique. Les animations faciales les plus réalistes sont effectuées en utilisant diverses techniques de capture faciale. Le coût et la complexité de la capture de performance ont tendance à la laisser hors de la portée de la plupart des entreprises. De plus, la capture de performance est plus difficile à utiliser dans les jeux vidéo, où l'interactivité implique le calcul en temps réel de transitions non-triviales entre les données des mouvements capturés (Zhao *et al.*, 2009). En outre, les besoins en mémoire et la puissance de traitement sont des facteurs importants à prendre en compte.

1.1 Description et mesure de l'expression faciale

Diverses approches objectives et systématiques permettant de coder les expressions faciales ont été proposées. Bien que les expressions faciales soient dues à un large éventail de facteurs, la plupart du temps, seules les modifications dues aux émotions, intentions ou communication sociale sont prises en compte, (Li *et al.*, 2005).

Divers systèmes de codage ont été développés principalement pour les études psychologiques tels que FACES (*Facial Expression Coding System*) (Kring et Sloan, 2007) et FACS (*Facial Action Coding System*) (Ekman et Friesen, 1978); ils sont examinés dans les travaux de Mauss et Robinson (2009). FACS est un système de codage basé sur l'anatomie du visage humain. Il rassemble les groupes musculaires du visage en unités d'action (*Action Units* ou AU) dont la combinaison permet d'obtenir toute expression possible, (Sagar, 2006). FACS a été utilisé avec succès par des studios d'animation tels que Weta Digital dans des films tels que *Monster House*, *King Kong* et *Avatar*. Le standard MPEG-4 propose une approche similaire en utilisant

des paramètres d'action faciale (*Facial Action Parameter* ou FAP) qui a été utilisée dans divers projets de recherche. Dans ce mémoire, les expressions faciales sont construites en utilisant des *blend shapes* correspondants aux groupes musculaires tels que ceux du système FACS.

1.2 Expressions faciales liées à la parole

La simulation de l'animation faciale liée à la parole est un domaine très actif. La synthèse d'animation se fait généralement par l'analyse d'une entrée audio et l'extraction des phonèmes, puis l'animation des visèmes du visage du modèle 3D (équivalents visuels de phonèmes) (Kshirsagar et Magnenat-Thalmann, 2003). Différentes approches ont été également mises au point afin d'améliorer le réalisme de l'animation tels que mélange d'animation liée à la parole avec celle induite par l'émotion (Kshirsagar *et al.*, 2001) et en utilisant des systèmes dont la structure est basée sur l'anatomie (Kshirsagar *et al.*, 2001; Sifakis *et al.*, 2006). D'autres travaux se sont concentrés sur l'amélioration de l'aspect visuel des animations liées à la parole (Cao *et al.*, 2005). Toutes ces approches permettent d'améliorer sensiblement le réalisme des expressions faciales liées à la parole et réduisent considérablement le temps d'animation faite manuellement. Néanmoins, ces approches fonctionnent correctement dans le cas où l'on dispose d'une entrée (audio) regroupant toute l'information requise pour ajuster l'animation faciale. Dans une situation d'activité physique intense, l'animation du personnage virtuel implique divers segments du corps ce qui rend les données plus difficiles à analyser. De plus, les réactions du personnage peuvent être le résultat du mouvement qu'il effectue à cet instant mais aussi des mouvements effectués dans les minutes précédentes. Finalement, le mouvement effectué par le personnage 3D peut déclencher des expressions dans des zones du visage autres que celles sollicitées lors de la parole, comme la région autour des yeux. Pour ces raisons, les travaux d'automatisation d'expression faciale liée à la parole ne peuvent résoudre la problématique de la synthèse d'expressions faciales liées à l'activité physique.

1.3 Expressions faciales liées aux émotions

Dans le domaine de la psychologie, les émotions sont très importantes et les chercheurs ont proposé des bases de classification sur un nombre restreint d'émotions. Pour simplifier davan-

tage la relation entre les différentes émotions, elle ont été représentées dans des espaces 2D simples (Russell, 1980; Thayer, 1989). Les chercheurs dans le domaine de l'infographie se sont basés sur ces approches et ont développé des interfaces bi-dimensionnelles facilitant le mélange d'émotions et permettant l'obtention de transitions réalistes entre les différentes émotions (Ruttkay *et al.*, 2003). D'autres approches ont été basées sur des systèmes de codage tels que FACS ou MPEG-4 afin d'obtenir des transitions réalistes entre les émotions (Arya *et al.*, 2009). Bien que ces travaux fournissent des approches intéressantes pour le mélange et les transitions entre les différentes expressions, elles fonctionnent lorsque les émotions sont connues et quand les émotions de base sont définies dans le modèle 3D. Diverses classifications des émotions de base ont été proposées dont la première date du premier siècle avant J.-C. (Russell et Mehrabian, 1977). La plupart des travaux de recherche tels que celui de Ruttkay *et al.* (2003) sont basés sur les six états émotionnels universels définis dans le système FACS : la joie, la surprise, la peur, la tristesse, la colère et le dégoût. Afin d'animer les expressions faciales dans un modèle 3D, ce dernier doit contenir un squelette de déformation paramétré en relation avec les émotions de base ou être composé d'un ensemble de *blend shapes*. Dans ce dernier cas, une version du maillage du modèle 3D doit être définie pour chaque émotion de base. La création de ces différentes versions du maillage requiert généralement un long travail manuel. La création de la déformation de maillage correspondant à un ensemble d'émotions a été étudiée dans le travail de Puklavage *et al.* (2010). Divers logiciels automatisant la création de visages 3D existent tels que *FaceGen* de *Singular Inversion* ou *Facial Studio* de *di-o-matic*. Les *blend shapes* d'émotions de base créées par ces logiciels sont celles des six émotions de base proposée par le système FACS.

Trouver quelle émotion déclencher dans le modèle 3D et quand la déclencher reste un problème complexe, ce qui est similaire au défi impliqué dans ce travail : développer une approche pouvant prédire l'expression faciale à partir d'observations et de modèles définissant comment un sujet humain réagit à différentes circonstances.

1.4 Expressions faciales liées à l'activité physique

Bien que les personnages 3D aient souvent à accomplir des activités physiques intenses, nous n'avons pas trouvé de recherche automatisant la synthèse d'expressions faciales réalistes liées à l'activité physique. En dehors de la littérature dans le domaine de l'infographie, le travail de McKenzie (1905) décrit les différentes expressions faciales liées à l'effort intense, l'essoufflement et la fatigue.

Dans le domaine de l'infographie, la littérature concernant l'expression faciale liée à l'effort physique se trouve principalement dans des ouvrages destinés aux artistes (Faigin, 1990; Goldfinger, 1991). Ces ouvrages contiennent diverses expressions faciales dont certaines sont liées à l'activité physique. Bien que ces ouvrages présentent des informations utiles pour la création de visages 3D expressifs et pour leur animation, ils ne sont pas utiles dans un contexte de synthèse d'animation faciale à partir du mouvement d'un personnage.

Bien qu'il y ait un nombre important de travaux sur l'animation faciale liée à la parole ou à l'émotion, ces travaux ne sont pas adaptés pour la synthèse d'expressions liées à l'activité physique. Basée sur l'apprentissage machine, l'approche proposée peut générer des animations faciales en lien avec le mouvement du personnage et les connaissances acquises à partir des données capturées.

CHAPITRE 2

ACQUISITION DES DONNÉES

Tel que mentionné précédemment, l'expression faciale liée à l'activité physique est le plus souvent négligée, principalement à cause du coût relié à l'animation manuelle ou à la capture de mouvement. Les techniques dynamiques employées actuellement dans les jeux vidéos pour animer l'expression faciale liée à l'activité physique sont basées sur de simples fonctions linéaires ou des seuils. Ces approches sont trop simples pour pouvoir simuler la complexité des animations faciales humaines en relation avec les efforts physiques fournis. Le but de ce projet est d'aller plus loin en termes de réalisme et de contrôle artistique en fournissant des outils pour automatiser cette tâche complexe tout en laissant un certain contrôle sur le résultat final.

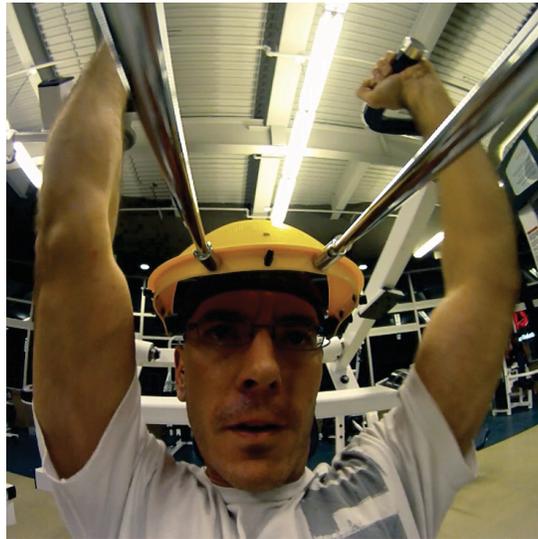
Afin de capturer entièrement la complexité des expressions faciales résultant d'efforts physiques, des données réelles étaient nécessaires. Diverses séances de captures ont été effectuées afin de collecter les données nécessaires au développement d'un modèle fournissant des résultats réalistes. Le choix de ces types de captures était motivé par l'hypothèse que durant un effort physique, le mouvement effectué, les données métaboliques, ainsi qu'une mesure des expressions faciales, fourniraient les informations appropriées. La figure 2.1 montre deux types de captures effectuées dans ce projet. Le présent projet de recherche a été soumis au comité d'éthique de la recherche et a reçu son approbation. Tout le nécessaire a été fait afin de garantir la sécurité des participants lors des séances de capture ainsi que la protection de leur données personnelles. La prochaine section détaille ces séances de captures.

2.1 Capture de mouvement

Les séances de capture de mouvement du corps entier ont été effectuées dans une grande salle où les participants pouvaient s'exercer librement sans aucun appareil d'entraînement. En raison de la gamme étendue des mouvements que les participants pouvaient pratiquer, il n'y a pas eu d'enregistrement d'animation faciale lors de ces séances. Néanmoins, ce n'était pas un problème étant donné que ces séances avaient pour but de fournir les données permettant d'établir



(a)



(b)

Figure 2.1 Deux types de captures ont été effectuées pour ce projet : (a) capture de mouvement du corps entier avec rythme cardiaque et (b) capture faciale avec rythme cardiaque

un lien entre le mouvement effectué et les paramètres métaboliques tels que le rythme cardiaque. Une vidéo témoin a toutefois été enregistrée pour faciliter l'analyse préliminaire ainsi que l'ajustement des données de capture de mouvement.

Il a été demandé à un total de 15 participants d'effectuer des séries d'exercices qu'ils pratiquent régulièrement et ce sans dépasser leur niveau d'intensité habituel. Ce nombre de participants nous permet d'avoir des données valides à $\pm 25\%$ 9 fois sur 10. Les rythmes cardiaques ont été contrôlés en temps réel et les participants ont été invités à alterner entre différents niveaux d'intensité pour chaque type d'exercice, afin de garantir une variabilité suffisante dans les données. À la fin de chaque session, les participants ont été invités à ralentir de façon à garantir l'acquisition de données de récupération. Chaque participant s'est entraîné de 7 à 13 minutes. L'âge des participants variait de 20 à 46 ans (moyenne 30.4 ans). Le niveau d'entraînement des participants variait entre activité physique légère au moins une fois toutes les deux semaines et un entraînement intensif presque tous les jours pour un total d'environ 10 heures par semaine. L'échelle ayant servi à mesurer le niveau d'entraînement est celle définie dans la documentation du matériel utilisé pour enregistrer le rythme cardiaque et les autres paramètres métaboliques.

Pendant que les participants s'entraînaient, le mouvement du corps entier a été capturé et les données de rythme cardiaque ont été enregistrées. Le matériel ainsi que le logiciel ayant servi à enregistrer le rythme cardiaque ont fourni une estimation d'autres indicateurs métaboliques tels que la consommation d'énergie, la fréquence respiratoire et l'EPOC (*Excess Post Oxygen Consumption*) Le matériel utilisé ainsi que le détail des procédures sera discuté dans la section 4.1.

2.2 Capture des expressions faciales

La seconde session de captures a été faite dans des salles d'entraînement d'un centre sportif, où les participants étaient invités à effectuer un entraînement cardiovasculaire ou de musculation en utilisant les divers appareils ou avec des poids libres. Le but de cette session de capture était d'établir le lien entre le mouvement effectué, le rythme cardiaque, et l'expression faciale. Lors de cette session de capture, 17 participants ont été invités à s'entraîner. Ce nombre de participants nous permet d'avoir des données valides à $\pm 20\%$ 9 fois sur 10. De même que pour la session de capture de mouvement, les participants devaient faire des exercices auxquels ils étaient habitués, à une intensité ne dépassant pas leur intensité habituelle.

Pour les exercices cardiovasculaires, les candidats étaient invités à changer l'intensité de l'exercice durant la session. Une période de récupération a été enregistrée à la fin de chaque session. Pour les exercices de musculation, il y avait une période de récupération entre les différents exercices enregistrés.

Chaque participant s'est entraîné durant 19 à 43 minutes. L'âge des participants variait entre 20 et 43 ans (moyenne de 28.1 ans). Le niveau d'entraînement des participants variait entre activité physique légère au moins une fois toutes les deux semaines et un entraînement intensif presque tous les jours pour un total d'environ 10 heures par semaine.

Le rythme cardiaque ainsi que les autres indicateurs métaboliques ont été enregistrés suivant la même procédure que les sessions de capture de mouvement. L'expression faciale a été capturée en enregistrant une vidéo du visage du participant pendant l'exercice (voir Fig. 2.1(b)).

Bien que du travail manuel ait été requis pour récupérer les données d'animation sous forme de courbes d'animation de *blend shapes* basiques, ce processus a grandement simplifié l'acquisition des données par rapport à l'utilisation de méthodes de capture faciale classique dans cet environnement. Le processus de récupération des données d'animation faciale est détaillé dans la section 4.4. Le matériel utilisé ainsi que le détail des procédures sera discuté dans la section 4.2.

En plus de la taille et du poids du participant, les poids utilisés dans les différents appareils de musculation ou lors des exercices de poids libre ont été notés afin de permettre une bonne approximation du travail mécanique et des forces impliquées lors de l'entraînement.

Les différentes informations constituant cette base de données ont été utilisées comme entrées dans différentes techniques d'apprentissage machine tel que décrit dans le prochain chapitre.

CHAPITRE 3

SYNTHÈSE DE L'EXPRESSION FACIALE

L'approche proposée pour générer des expressions faciales est axée sur les données. Des techniques d'apprentissage machine sont appliquées aux données capturées afin d'obtenir des modèles pouvant prédire l'expression faciale d'un personnage 3D selon son activité physique et son niveau de fatigue. Comme le montre la figure 3.1, cette approche est divisée en une phase d'apprentissage hors-ligne et une phase de synthèse en-ligne.

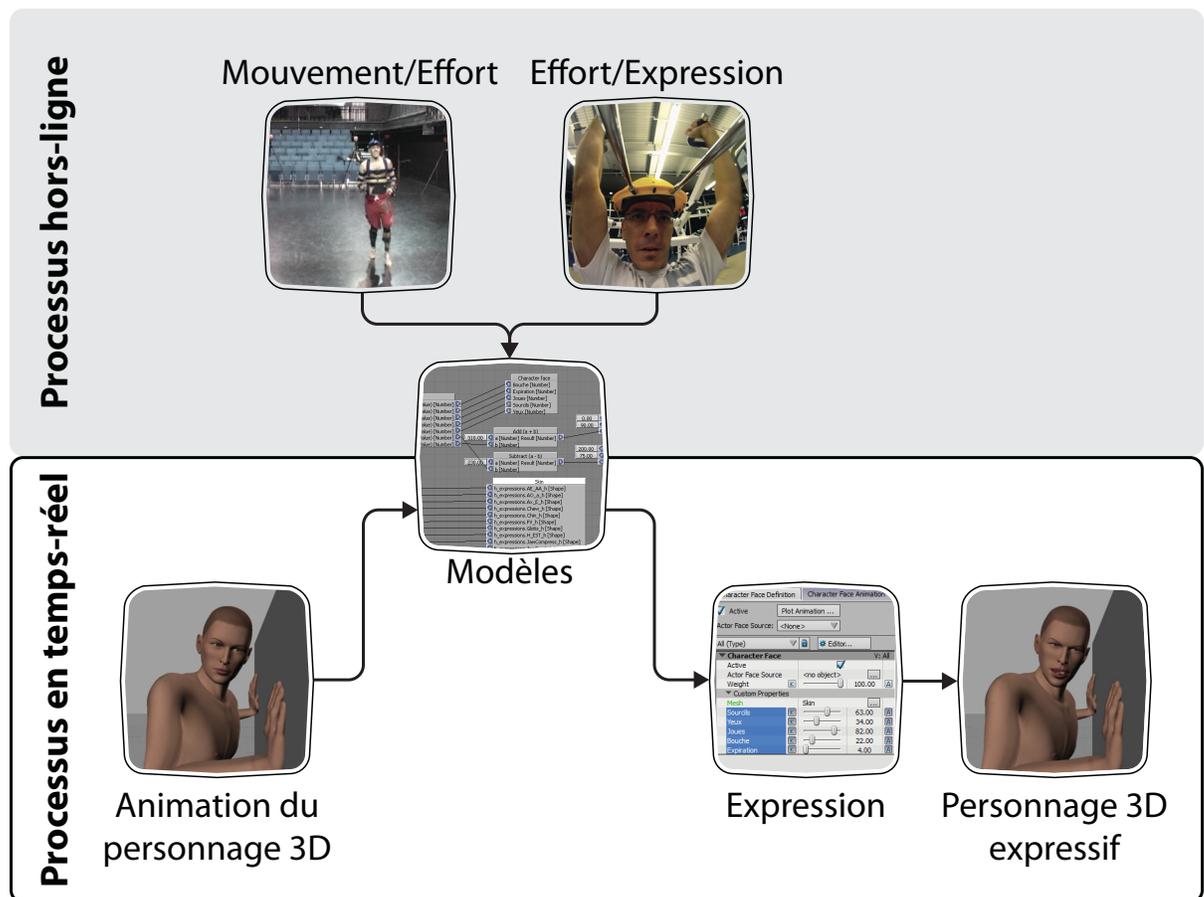


Figure 3.1 Vue générale de l'approche

Afin de mieux comprendre les mécanismes sous-jacents et les relations entre les exercices physiques et les expressions faciales, une analyse préliminaire a été réalisée. Cette analyse a permis de choisir de façon éclairée les méthodes d'apprentissage machine appropriées à nos



données et pouvant modéliser l'évolution ainsi que la variation des expressions faciales lors d'activité physique modérée à intense.

Un résumé des résultats obtenus à partir de cette analyse est tout d'abord présenté dans la section suivante, tandis que l'approche d'apprentissage machine sera présentée dans les sections subséquentes.

3.1 Observations sur les données acquises

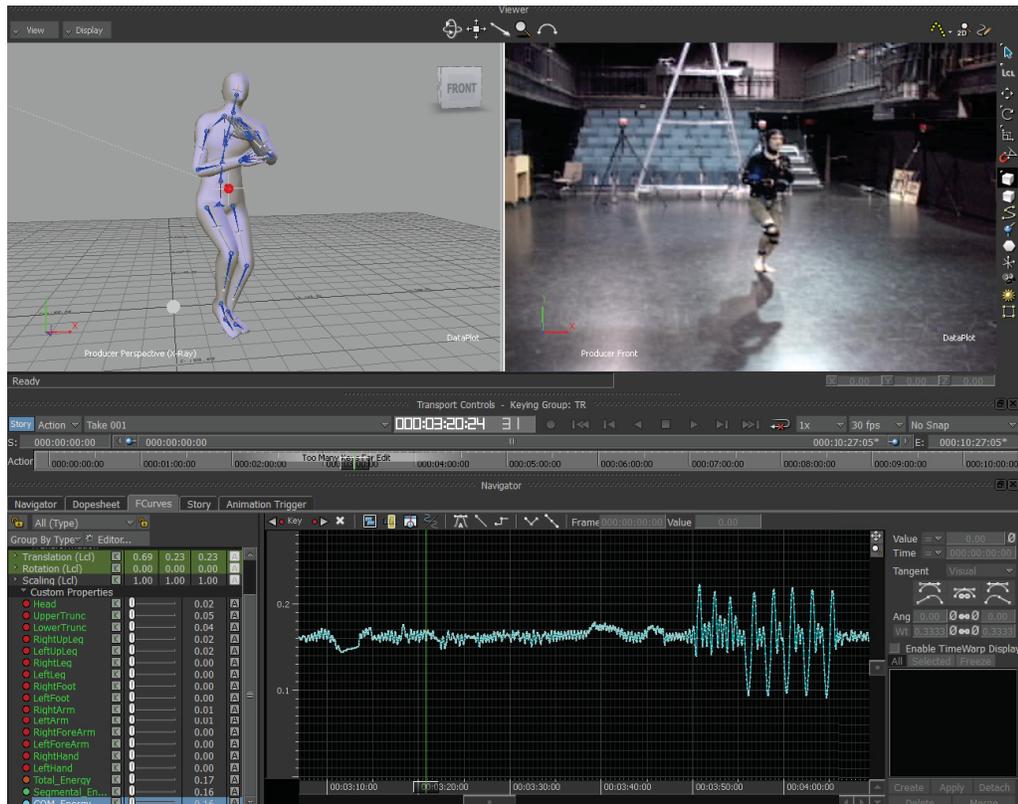
Une analyse des données acquises (capture de corps entier, rythme cardiaque et vidéo) a été réalisée en utilisant diverses interfaces utilisateur spécialement développées pour ce projet (voir fig. 3.2). Ceci a conduit à une série d'observations clés facilitant la compréhension des relations complexes entre l'activité physique et l'expression faciale.

Bien que les données acquises fournissent la plupart des informations nécessaires, il a été constaté que d'autres paramètres métaboliques étaient importants : intensité de l'effort, niveau de fatigue, *etc.* Bien que ces paramètres soient influencés par des facteurs mesurables tels que la vitesse, cadence, poids soulevé, EPOC ou la durée de l'exercice, nous n'avons pas pu déterminer un moyen simple de les approximer. Il était donc logique d'utiliser l'apprentissage machine pour prédire les expressions faciales.

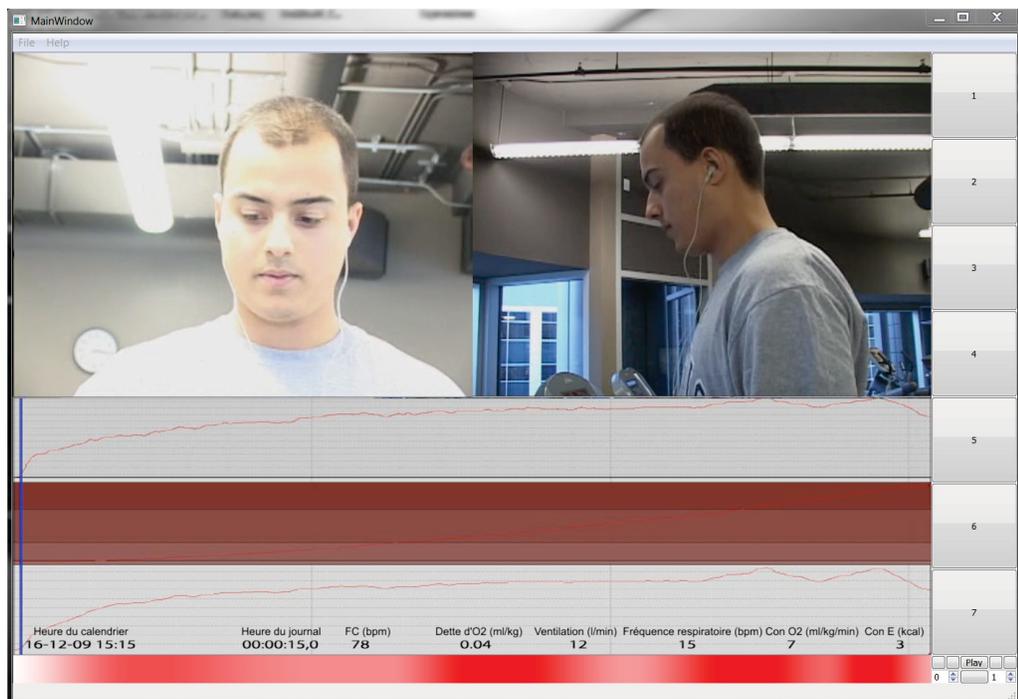
Sur la base des données acquises ainsi que les autres indicateurs, cette section détaille les relations observées entre les aspects mécaniques, métaboliques ainsi que les paramètres d'expressions faciales des données capturées. Deux types de relations ont été découverts : des relations générales valables pour toutes les composantes de base des expressions faciales, et des relations spécifiques s'appliquant plutôt à une composante spécifique de l'expression faciale.

Observations générales

Il a été déterminé que les zones du visage les plus sollicitées lors d'efforts physiques étaient concentrées autour des yeux et de la bouche. Les modèles de synthèse développés dans ce travail servent donc à automatiser les paramètres de ces zones spécifiques. Il a aussi été déterminé que



(a) Mouvement vs Métabolique



(b) Métabolique vs Expression

Figure 3.2 Outils utilisés pour l'analyse des données

les expressions faciales liées à l'activité physique sont liées à deux composantes principales : l'effort instantané et la fatigue. L'analyse des données a abouti à une liste de règles de base liant l'activité physique à l'expression faciale :

- L'intensité de l'expression faciale est proportionnelle à la masse déplacée et inversement proportionnelle à la masse musculaire utilisée ;
- L'évolution de l'intensité de l'expression faciale est proportionnelle à la variation du rythme cardiaque et à la variation d'énergie métabolique dépensée ;
- L'expression faciale liée à l'effort instantané est proportionnelle à la puissance mécanique ;
- Le temps de récupération est proportionnel à l'intensité de l'effort et inversement proportionnelle au niveau d'entraînement.

Observations spécifiques

Des règles spécifiques à chacune des différentes composantes des expressions faciales liées à l'activité physique ont aussi été observées. Concernant les expressions de base de la bouche :

- L'étirement de la bouche est induit par deux facteurs : l'effort instantané et la fatigue (voir fig. 3.3(a)) ;
- La bouche reste fermée au début de la session d'exercice. Après un certain temps, elle commence à s'ouvrir et cette ouverture est liée à deux facteurs : le rythme respiratoire et la fatigue (voir fig. 3.3(b)) ;

D'autres observations concernaient la région des yeux et des sourcils :

- Le plissement des yeux est principalement lié à l'effort instantané et ce jusqu'à un certain niveau de fatigue mesuré dans nos données par l'EPOC (voir fig. 3.3(c)). À un niveau de



Figure 3.3 L'effort physique induit diverses expressions faciales

fatigue supérieur, les yeux ont tendance à se relâcher en relation avec la fatigue. Le plissement ne retourne toutefois pas à l'état de repos ; une légère valeur constante subsiste (voir fig. 3.3(d)) ;

- Le comportement des sourcils est une combinaison de mouvement descendant lié à l'effort instantané (voir fig. 3.3(e)) et un mouvement ascendant lié à la fatigue (voir fig. 3.3(f)).

Finalement, des observations ont été faites concernant la respiration et la déglutition :

- La fréquence du mouvement d'expiration forte et profonde est liée à deux facteurs : la fréquence respiratoire et la fatigue ;

- La fréquence du mouvement de déglutition est liée à deux facteurs : la fréquence respiratoire et la fatigue.

3.2 Modèle biomécanique

Une des entrées principales aussi bien pour la phase hors-ligne (construction des modèles) que la phase en-ligne (prédiction des paramètres métaboliques et des expressions faciales) est le travail mécanique résultant du mouvement des personnes capturées et des personnages 3D.

Afin d'approximer le travail mécanique effectué, diverses approches ont été évaluées : énergie potentielle, énergie cinétique de translation, et énergie cinétique de rotation ont été calculées aussi bien pour le centre de masse du personnage 3D que pour chaque segment. Comme la phase d'apprentissage machine peut prendre n'importe quel nombre d'entrées, divers configurations d'entrées / sorties ont été testées : travail mécanique du centre de masse (CDM), bas/haut du corps, et finalement le travail total des segments. Le but de ces tests était de trouver l'approche qui serait calculée efficacement et qui donnerait de bons résultats aussi bien pour la phase d'apprentissage que pour la phase de synthèse. Le tableau 3.1 montre les différentes combinaisons évaluées. Bien qu'avoir en entrée le travail mécanique de chaque segment semblait intuitivement l'approche qui donnerait les meilleures connaissances sur le type d'exercice effectué, les résultats obtenus en termes d'animation faciale étaient bruités avec des valeurs de *blend shapes* changeant trop rapidement en comparaison avec les données réelles. Une explication à ce phénomène est que bien que diverses séances de captures aient été réalisées, le nombre d'instances de données d'entraînement disponibles reste insuffisant pour capturer les interrelations complexes entre les différents segments et les différentes composantes de l'expression faciale. Finalement, ce qui a donné les meilleurs résultats était la somme du travail mécanique calculé à partir de l'énergie potentielle, cinétique de translation et cinétique de rotation, pour tous les segments.

3.3 Apprentissage machine des expressions faciales

L'apprentissage machine a été utilisé pour modéliser le comportement complexe des diverses composantes de l'expression faciale en relation avec l'activité physique. Étant donné le type de données qui ont été acquises et le type de prédiction requis, les techniques de régression paraissent les plus prometteuses. Afin de sélectionner la meilleure approche, divers modèles ont été entraînés en utilisant différentes combinaisons de données comme entrées, puis la qualité de ces modèles a été évaluée sur des ensembles de données de test. De même, les paramètres optimaux des modèles ont été sélectionnés en utilisant une validation croisée.

Les arbres de régression se sont avérés être la meilleure approche pour apprendre la relation entre le travail mécanique et les paramètres métaboliques alors que les machines à vecteurs de support (SVM) ont été plus précises dans la prédiction des expressions faciales à partir du travail mécanique. Le flux de données, l'approche d'apprentissage ainsi que les modèles sont présentés dans la figure 3.4.

3.3.1 Prédiction des paramètres métaboliques

Pour prédire le rythme cardiaque à partir du mouvement 3D du personnage, diverses techniques d'apprentissage ont été testées avec différentes combinaisons de données en entrée. Afin de prendre en compte l'aspect temporel, des tests ont été réalisés en injectant de multiples instances de données décalées dans le temps. Le nombre d'instances utilisées ainsi que le décalage ($\Delta(t)$) permettaient de couvrir une fenêtre temporelle allant de 0 à 4 secondes. Ces combinaisons sont résumées dans le tableau 3.1. Puisque les diverses techniques d'apprentis-

Tableau 3.1 Combinaisons d'entrées / sorties évaluées (modèle métabolique)

Entrée(s)	Énergie mécanique	Nombre d'instances	$\Delta(t)$	Sortie
Hanche	$E_p, E_{p+E_t},$ $E_{p+E_t+E_r}$	1, 3, 5	$1/60s, 1/30s,$ $1/4s, 1s$	rcn, $\Delta(rcn)$
CDM total				
CDM haut, CDM bas				
CDM segments				
Σ CDM segments				

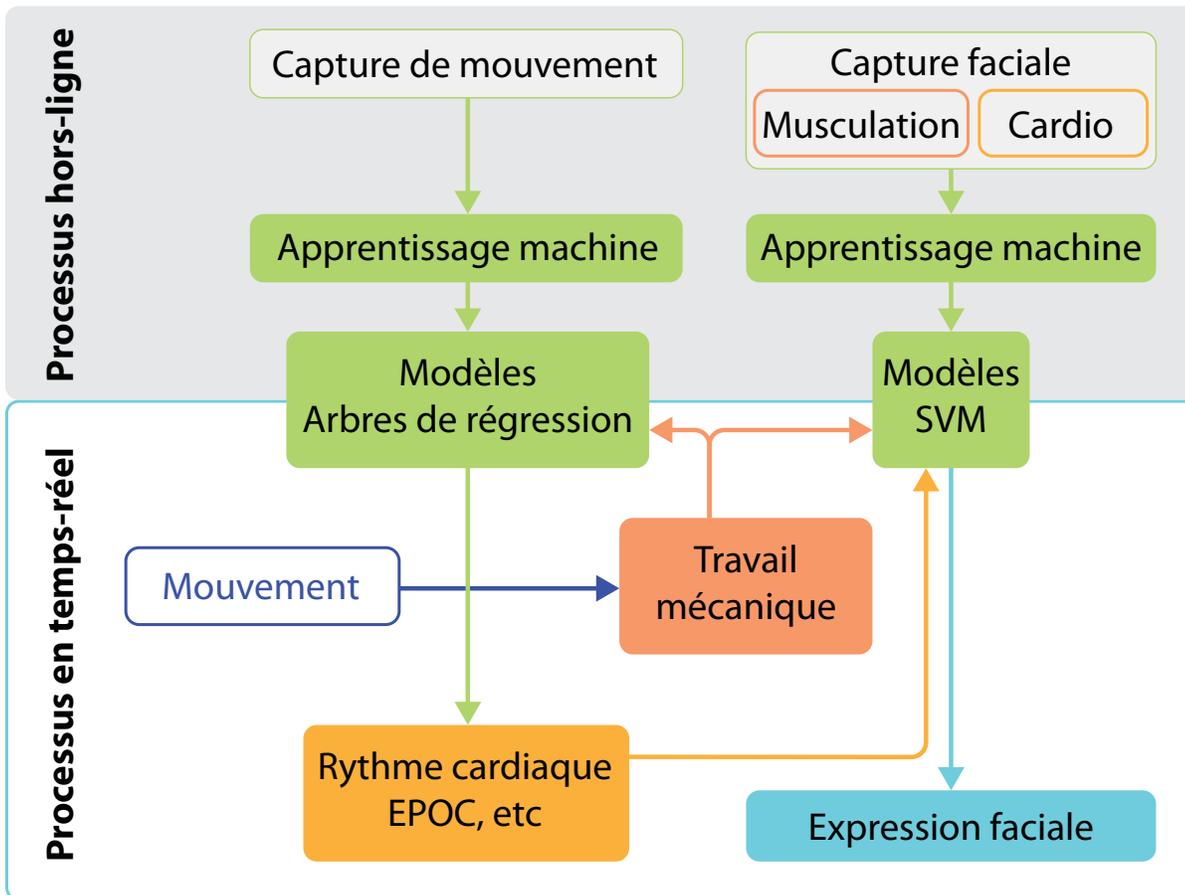


Figure 3.4 Approche utilisant l'apprentissage machine pour la modélisation des expressions faciales

sage sont affectées par l'intervalle des données en entrée, et parce que l'intervalle de rythme cardiaque varie d'un candidat à l'autre, les données de rythme cardiaque ont été normalisées dans l'intervalle $[0, 1]$ et ce à l'intérieur de la HRR (*Heart Rate Reserve*) ou réserve de rythme cardiaque qui est calculée comme la différence entre le rythme cardiaque maximal et le rythme cardiaque au repos. Elle constitue un indicateur de la condition physique.

Le rythme cardiaque normalisé (rcn) est calculé comme suit :

$$\text{rcn} = \frac{\text{rythme cardiaque actuel} - \text{rythme cardiaque au repos}}{\text{rythme cardiaque maximal} - \text{rythme cardiaque au repos}}$$

Le rythme cardiaque maximal est calculé à partir de l'âge de la personne en utilisant une formule standard d'entraînement. Le rythme cardiaque au repos est quant à lui calculé à partir

du niveau d'entraînement de la personne en utilisant une formule dérivée des tableaux standards d'entraînement.

Le rythme cardiaque se comporte dans deux modes différents : rythme cardiaque croissant et rythme cardiaque décroissant. Ceci est basé sur l'intensité de l'exercice : pour chaque personne, il y a un certain seuil d'intensité de l'exercice qui résulte en une augmentation ou diminution des besoins en oxygène et du même coup du rythme cardiaque.

Un modèle d'arbre de régression a été construit en utilisant le *UserClassifier* dans le logiciel de forage de données Weka (Hall *et al.*, 2009). Diverses combinaisons d'entrées / sorties ont été testées telles que le travail mécanique en entrée et le rythme cardiaque en sortie, le travail mécanique en entrée et la variation du rythme cardiaque en sortie, *etc.* Des différents modèles testés, celui donnant les meilleurs résultats était de prédire la variation du rythme cardiaque à partir du rythme cardiaque actuel et du travail mécanique.

Il a été déterminé qu'un modèle entraîné sur les données d'un seul participant pouvait prédire correctement le rythme cardiaque de ce même participant (coefficient de corrélation de 0.88 et une erreur RMSE de 19 %). Cependant, combiner les données de plusieurs candidats pour générer un seul modèle donnait un taux d'erreur trop important (coefficient de corrélation de 0.21 et une erreur RMSE de 79 %). Comme on peut le constater dans la figure 3.5, les modèles donnent une meilleure estimation pour les exercices cardiovasculaires que pour les exercices de musculation. Ceci s'explique par la variabilité moins importante dans le rythme cardiaque normalisé durant les exercices cardiovasculaires. Afin d'évaluer les modèles, deux tiers des données étaient utilisées pour l'entraînement et un tiers des données était utilisé pour le test. Les modèles finaux ont été évalués en utilisant la méthode *10-fold*. Les segments de données d'entraînement et de test étaient continus.

En utilisant les paramètres de régularisation appropriés (nombre minimum d'instances, élagage) pour limiter la complexité de la régression, cela a donné des arbres de régression simples à seulement deux feuilles, tout en prédisant le rythme cardiaque des participants avec les coefficients de corrélation et les erreurs RMSE montrés plus haut. En analysant les arbres de

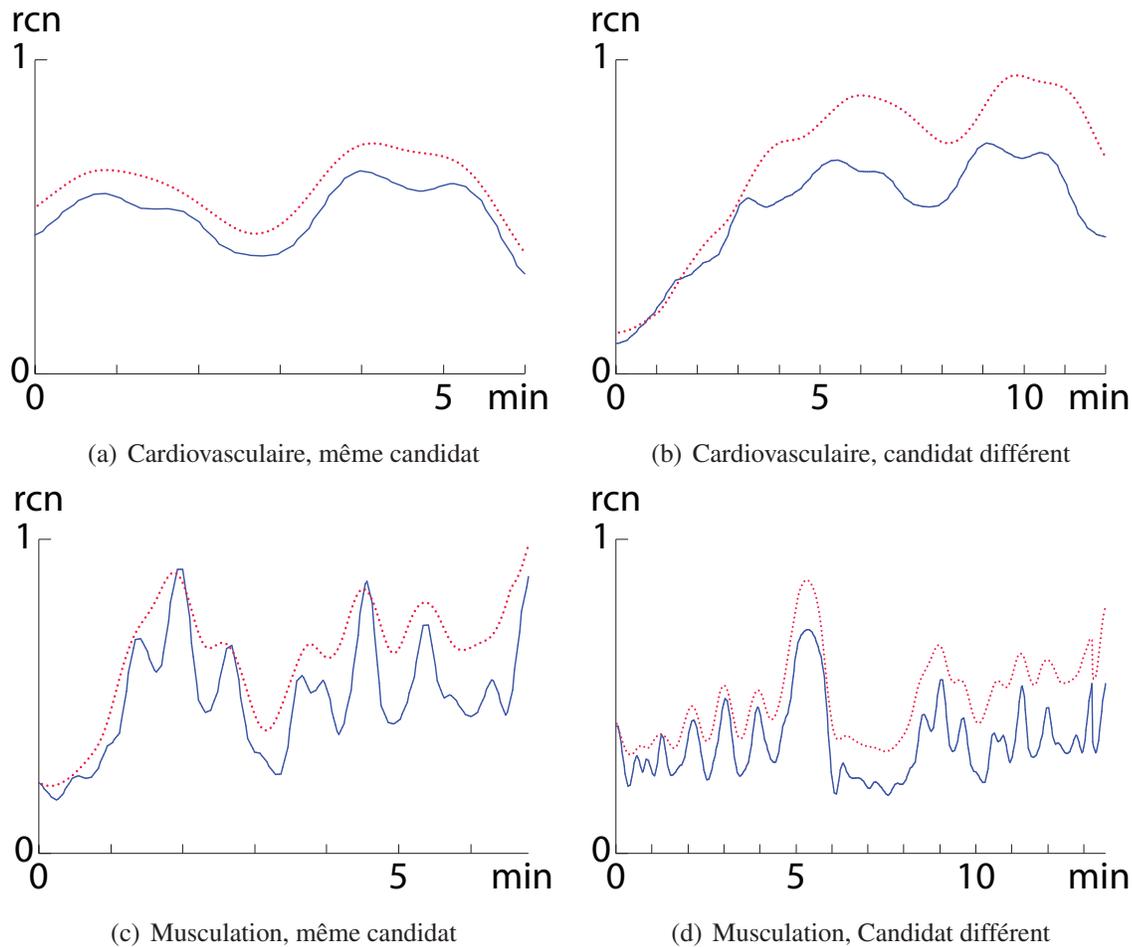


Figure 3.5 Comparaison entre le rythme cardiaque normalisé réel (bleu pointillé) en fonction du temps, pour (a-b) un exercice cardiovasculaire et (c-d) un exercice de musculation, aussi bien pour (a,c) le même candidat que pour (b,d) un candidat différent

régression générés pour chaque participant, une structure commune a émergé : Tous les arbres de régression avaient un test identique au nœud racine. Ce test compare le travail mécanique w à la valeur d'un seuil s_{racine} , qui divise ensuite la prédiction en deux formules de rythme cardiaque : ascendant ou descendant. De plus, il a été déterminé que la différence principale entre ces modèles personnalisés était la valeur de ce seuil utilisé dans le test, et que ces valeurs dépendaient grandement du niveau d'entraînement du participant. Afin de construire un modèle unique, une courbe de tendance a été calculée avec les valeurs des modèles individuels.

Ceci a permis de déterminer le seuil et les constantes du modèle d'arbre de régression suivant :

$$s_{racine} = 7.13 + 0.42 \times \text{niveau d'entraînement}$$

$$\Delta(\text{rcn}) = \begin{cases} c_{\text{asc}} \times (w - s_{racine}) \times (1 - \text{rcn})^2 & s_{racine} < w \\ c_{\text{déc}} \times (w - s_{racine}) \times \text{rcn}^2 & w \leq s_{racine} \end{cases}$$

$$c_{\text{asc}} = 0.0056 - 0.00043 \times \text{niveau d'entraînement}$$

$$c_{\text{déc}} = 0.0009 + 0.00025 \times \text{niveau d'entraînement}$$

Dans ce modèle, ces constantes représentent des facteurs de croissance et décroissance du rcn. L'utilisation de ce modèle a donné des prédictions presque aussi bonnes que celles en utilisant les modèles générés pour chaque participant avec un coefficient de corrélation de 0.87 et une erreur RMSE de 24 %. De plus, le niveau d'entraînement constitue un élément intuitif pouvant être utilisé par les animateurs comme paramètre permettant de contrôler le niveau de réponse du personnage 3D aux différents types d'exercices.

Une fois le rythme cardiaque normalisé prédit, il est utilisé pour estimer d'autres paramètres métaboliques :

- VO_2 ;
- EPOC ;
- Fréquence respiratoire ;
- Consommation d'énergie métabolique.

Comme pour le rythme cardiaque, le VO_2 , la fréquence respiratoire et la consommation d'énergie métabolique doivent être normalisés entre leurs valeurs minimales et maximales.

À partir du rythme cardiaque prédit, le VO_2 est estimé d'une façon directe car il est proportionnel au rythme cardiaque normalisé (Swain et Leutholtz, 1997). De même, il a été constaté que cette approche appliquée à la fréquence respiratoire donnait des résultats suffisamment précis

pour améliorer sensiblement le réalisme de l'animation du personnage 3D. En utilisant le VO_2 estimé, l'EPOC est approximé en utilisant une formule dérivée du modèle d'accumulation de l'EPOC, (Firstbeat Technologies, 2007a). Ces estimations du rythme cardiaque normalisé et de l'EPOC, en plus du travail mécanique et de la puissance mécanique, seront ensuite utilisées dans la prédiction des expressions faciales tel qu'il sera décrit dans la prochaine section.

3.3.2 Prédiction des composantes de l'expression faciale

Afin d'animer l'expression faciale d'un personnage 3D, le poids correspondant à chaque *blend shape* associé à un groupe musculaire particulier doit être déterminé à partir du mouvement de ce personnage. Comparativement aux données métaboliques, les données d'expressions faciales issues de nos captures contiennent des changements plus brusques et varient plus fréquemment. Afin d'apprendre la relation dynamique et complexe liant les données des sessions de captures et les données de *blend shapes*, il a été déterminé que la régression en utilisant des machines à vecteurs de supports (SVM) donnait de bons résultats. Des tests ont été effectués pour divers participants, divers exercices ainsi que pour des participants uniques et exercices uniques (voir fig. 3.6). Pour un participant unique, le fait d'entraîner le modèle sur les données d'un seul participant et sur un seul exercice donnait des prédictions décentes de valeurs de *blend shape* pour ce même participant, sur des exercices non utilisés pour l'entraînement du modèle. Néanmoins, il y avait des différences entre les valeurs capturées et celles prédites comme le montre la figure 3.6(d). L'apprentissage fait sur un plus large répertoire d'exercices diminuait cette erreur. Comparé à ce qui a été observé concernant la prédiction de paramètres métaboliques, la prédiction sur des exercices de musculation (Fig. 3.6(a,c,d)) est assez bien alignée avec les données réelles. La prédiction pour des exercices cardiovasculaires suit la forme globale de la courbe sans pouvoir suivre les petits détails (Fig. 3.6(b)). Pour des participants multiples, l'entraînement fait sur un seul participant donnait des résultats relativement bons pour un même exercice fait par un autre participant non utilisé dans les données d'entraînement (voir fig. 3.6(c)). Cependant, l'entraînement des modèles effectué sur les données de tous les participants et exercices donnait des résultats relativement mauvais. Encore une fois, les données ont dû être normalisées, mais cette fois à l'égard de l'expressivité du participant. Ceci

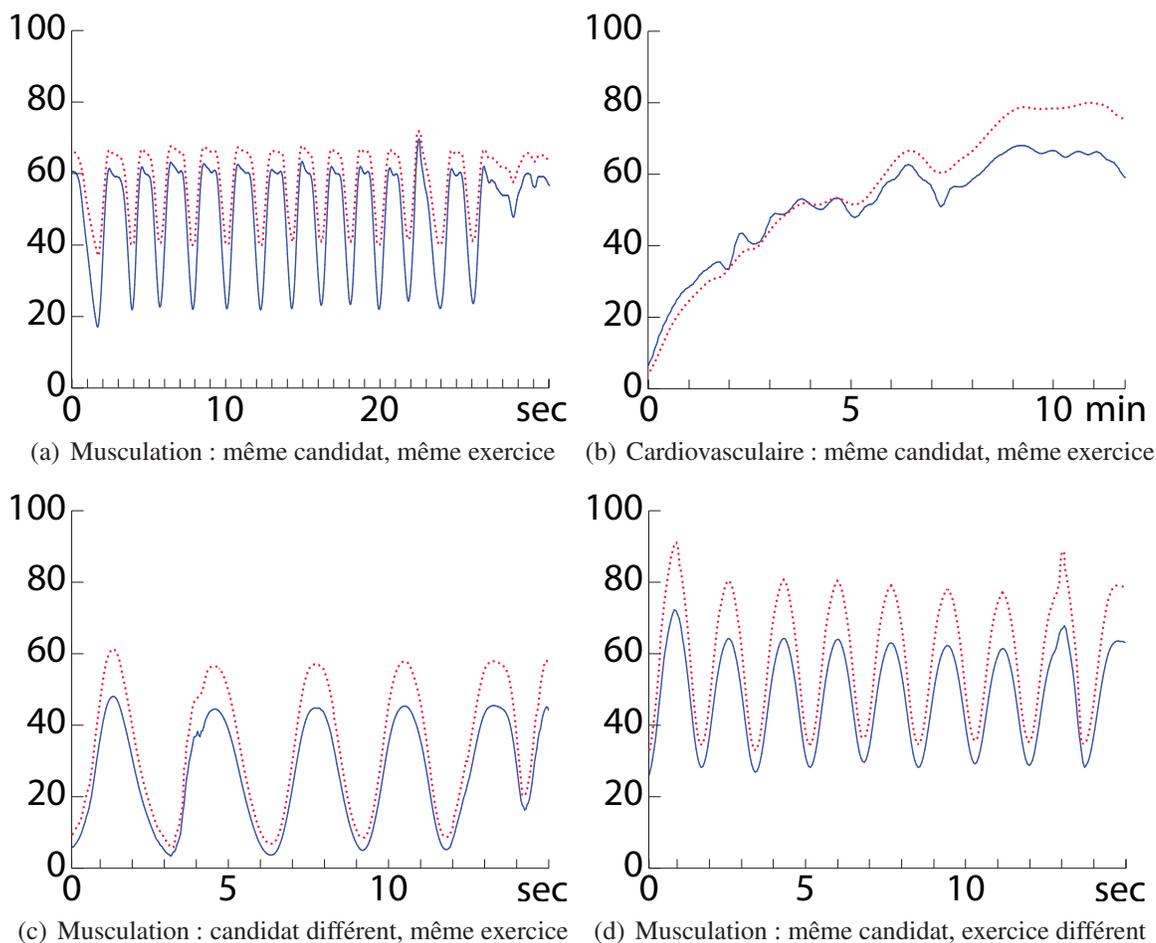


Figure 3.6 Comparaison entre la valeur du poids dans l'intervalle $[0, 100]$ d'une *blend shape* correspondant à une composante de l'expression faciale réelle (bleu) et prédite (rouge pointillé) pour (a,c,d) exercice de musculation et (b) exercice cardiovasculaire. Les courbes dans (c) sont celles d'un candidat différent pour un même exercice, et (d) la prédiction pour un même candidat sur un exercice différent

peut se voir sur la figure 3.6(c) comme les courbes sont bien alignées, mais les valeurs des *blend shapes* sont à une autre échelle. Certains participants pouvaient endurer des efforts extrêmes tout en gardant une expression faciale relativement neutre alors que d'autres participants affichaient des expressions très prononcées pour un effort moins important. Malheureusement, aucun lien n'a pu être établi entre l'expressivité et les paramètres recueillis sur chaque participant (âge, niveau d'entraînement, taille, poids). Il a été possible cependant de calculer un paramètre d'expressivité de chaque participant. Ceci a été fait en trouvant la valeur maximale pour chaque *blend shape* du participant, résultant en un vecteur d'expressivité à cinq dimen-

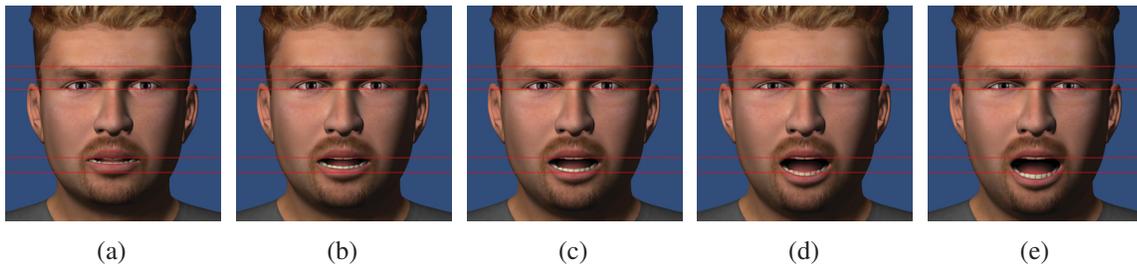


Figure 3.7 Résultats obtenus avec différents paramètres d’expressivité (voir tableau 3.2) : les différentes *blend shapes* sont adoucies ou exagérées indépendamment les unes des autres

sions. Les valeurs des *blend shapes* des participants ont donc été normalisées de façon à ce que la valeur maximale atteinte pour chaque *blend shape* corresponde à un (1). Ces vecteurs d’expressivité permettent aux animateurs de créer des personnages présentant des expressions faciales distinctives en augmentant ou en diminuant les valeurs d’expressivité. La figure 3.7 présente des résultats obtenus au même instant avec différents paramètres d’expressivité et la figure 3.8 montre les résultats obtenus avec différents niveaux d’entraînement après cinq minutes de course à un rythme moyen.

Tableau 3.2 Tableau des paramètres d’expressivité utilisés dans la figure 3.7

<i>blend shape</i>	Bouche/Ouverture	Bouche/Expiration	Bouche/Étirement	Yeux	Sourcils
3.7(a)	0.86	0.95	0.72	0.96	0.73
3.7(b)	0.94	0.95	0.93	0.94	0.95
3.7(c)	1.0	1.0	1.0	1.0	1.0
3.7(d)	1.05	1.08	1.08	1.10	1.05
3.7(e)	1.14	1.02	1.16	1.15	1.06

En utilisant ces données normalisées de *blend shape*, la totalité des données (tous les participants et tous les exercices) a été utilisée dans la phase d’apprentissage des modèles de SVM. Pour sélectionner la combinaison d’entrées et de paramètres de modèle la plus appropriée, diverses combinaisons ont été évaluées sur un ensemble de données de test en utilisant une validation croisée. Le tableau 3.3 montre différentes combinaisons d’entrées / sorties qui ont été évaluées. Les combinaison utilisant de multiples instances décalées dans le temps donnaient des résultats convenables à l’entraînement mais souffraient d’une importante accumulation d’erreur les rendant inutilisables. Les modèles utilisant des données mécaniques seulement

Tableau 3.3 Combinaisons d'entrées / sorties évaluées (modèles des *blend shapes*)

Entrée(s)	Nombre d'instances	$\Delta(t)$	Sortie
w	1, 3, 5	1/30s, 1/4s, 1s	<i>blend shape</i> , $\Delta(\textit{blend shape})$
w, p			
w, p, rcn			
w, p, EPOC			
$w, p, \text{rcn}, \text{EPOC}$			

donnaient de bons résultats dans les tests réalisés sur de courts exercices mais ne pouvaient prédire correctement l'effet de la fatigue. Finalement, la combinaison qui a donné les meilleurs résultats était le travail mécanique, puissance mécanique, rythme cardiaque normalisé et EPOC en entrée, et la valeur de *blend shape* en sortie. En ce qui concerne les paramètres des SVM,



(a) Niveau d'entraînement = 0

(b) Niveau d'entraînement = 5

(c) Niveau d'entraînement = 10

Figure 3.8 Résultats obtenus avec différents niveaux d'entraînement après cinq minutes de course à un rythme moyen

un noyau à fonction de base radiale (RBF) a été sélectionné et différentes combinaisons de paramètres ont été testées : Le paramètre de régularisation et le gamma de la RBF variaient indépendamment de 10^{-10} à 10^{10} à un intervalle de 0.5 de l'exposant dans une première phase de calculs. Une fois les valeurs les plus proches déterminées, une optimisation a été réalisée en limitant les intervalles de recherche autour de ces valeurs trouvées. Comme les SVM donnent une seule sortie, plusieurs modèles devaient être générés, un pour chacune des quatre *blend shapes*. Les valeurs du paramètre de régularisation et de gamma ayant donné les meilleurs résultats étaient différentes pour chaque *blend shape*. Pour le paramètre de régularisation, cela variait entre 10^3 et 10^6 , pendant que la valeur du gamma de la RBF était dans un intervalle

de 10^{-5} à 10^{-2} . L'erreur RMSE à l'entraînement des modèles finaux était dans l'intervalle [17 %, 26 %] (voir tableau 3.4). Les modèles finaux permettent une bonne généralisation des

Tableau 3.4 Erreur RMSE des modèles de prédiction des *blend shapes*

<i>blend shape</i>	Bouche/Ouverture	Bouche/Étirement	Yeux	Sourcils
RMSE	17.14	17.83	21.6	25.78

données capturées, ce qui indique que notre méthode pourrait être utilisée pour générer des animations faciales réalistes sur d'autres types de mouvements.

CHAPITRE 4

DISCUSSION

Dans cette section, le matériel utilisé lors des captures et les procédures suivies sont discutés. Ensuite, le prototype de cette approche et les évaluations expérimentales sont discutés dans le contexte d'utilisation prévu de ces modèles.

4.1 Capture de mouvement

4.1.1 Système de mocap utilisé

La capture de mouvement a été réalisée dans les locaux du LARTech (Laboratoire de recherche-cr ation en technochor graphie) en utilisant un syst me de capture de mouvement optique passif de marque *MotionAnalysis*. La figure 4.1 montre une session de capture. Le syst me contenait 20 cam ras de type *Hawk*. Le nombre important de cam ras permettait d'obtenir une bonne pr cision de capture de mouvement. Toutefois, la r solution de ces derni res ne permettait pas de faire des captures faciales en m me temps de que celles du mouvement du corps. Le logiciel *MotionAnalysis EVaRT* a  t  utilis  pour la capture de mouvement tandis que le nettoyage des donn es ainsi que l'animation des personnages 3D ont  t  r alis s dans *Autodesk MotionBuilder*. Le nettoyage des donn es de capture de mouvement est une op ration typique qui consiste   corriger les erreurs dues aux occlusions des marqueurs (ceux-ci sont parfois cach s par le candidat lui-m me lors de certains mouvements ou positions) et le bruit principalement d  aux changement des cam ras qui voient un marqueur dans l'espace de capture. La phase d'animation consiste principalement   calculer l'animation du squelette du personnage 3D selon le mouvement des marqueurs captur s. Les donn es ont  t  captur es   une fr quence de 60Hz. Cette fr quence permettait d'offrir une quantit  suffisante de donn es tout en n cessitant un temps raisonnable de nettoyage des donn es.



Figure 4.1 Studio de capture de mouvement

4.1.2 Costumes et marqueurs utilisés

Les costumes de capture de mouvement fournis par le fabricant du système permettent d'obtenir des résultats visuellement convaincants et sont bien adaptés à une utilisation dans un contexte d'animation 3D pour le cinéma ou le jeu vidéo. Toutefois, les déformations du tissu dues aux plis engendrés par les mouvements lors des exercices ne permettaient pas d'obtenir le niveau de précision recherché. De plus, étant donné que les costumes de capture de mouvement sont chauds et que les participants avaient à faire des activités physiques moyennement intenses, la transpiration n'aurait pas permis de réutiliser les costumes par plusieurs personnes. Diverses solutions ont donc été testées : adhésif médical, bandes élastiques, costume de mocap. Les meilleurs résultats étaient généralement obtenus en collant les bases des marqueurs directement sur la peau en utilisant des bandes d'adhésif médical (voir fig.4.2(a)). Cette so-

lution était toutefois moins précise pour les candidats ayant un important volume de graisse corporelle. De plus, pour les personnes poilues, l'adhésif n'était pas une solution retenue car le retrait des marqueurs pouvait être douloureux. Dans ces situations, un système à base de bandes élastiques et de bandes auto-adhérentes a été utilisé.

Ce système développé à l'aide de bandes élastiques et de bandes auto-adhérentes (voir fig. 4.2(b)) s'est avéré une solution bien plus précise que l'utilisation du costume de mocap. De plus, étant donné que les participants pouvaient le porter sur la peau en étant peu vêtus leur permettait d'éviter une transpiration excessive. Toutefois, le système était moins stable au niveau des genoux que les autres solutions testées. Il a nécessité quelques arrêts pour ajustements lors de certaines séances de capture.

Le costume de mocap est celui qui offre le moins de précision car même lorsqu'il est correctement ajusté, il n'adhère pas correctement au corps et le tissu glisse légèrement sur la peau lors des flexions. Toutefois, il restait une solution pratique pour les situations où il n'était pas possible d'utiliser les autres solutions, particulièrement pour les candidats de sexe féminin (voir fig. 4.2(c)). De plus, les costumes de mocap étant très chauds, ils occasionnent une transpiration excessive et doivent donc être lavés après chaque utilisation ce qui aurait limité le nombre de candidats qu'il était possible de capturer en une journée. Finalement, lors des séances de capture, les trois systèmes ont été combinés selon la situation et le participant. Un total de 36 marqueurs ont été utilisés lors des captures. Étant donné qu'une seule personne était capturée à chaque fois et que les mouvements effectués ne provoquaient pas trop souvent des occlusions de marqueurs, ce nombre était suffisant pour capturer fidèlement les mouvements effectués sans toutefois augmenter considérablement le temps nécessaire au nettoyage des données.

4.1.3 Matériel d'enregistrement des paramètres métaboliques

Les données métaboliques ont été capturées en utilisant une ceinture thoracique de type *Suunto Comfort Belt* à transmission sans fil ANT. Cette dernière était reliée à une montre cardiofréquence-mètre *Suunto t6c* et à un ordinateur portable muni d'une clé sans fil ANT *Suunto PC-Pod*. La montre cardiofréquence-mètre enregistrait les données à une fréquence d'un enre-

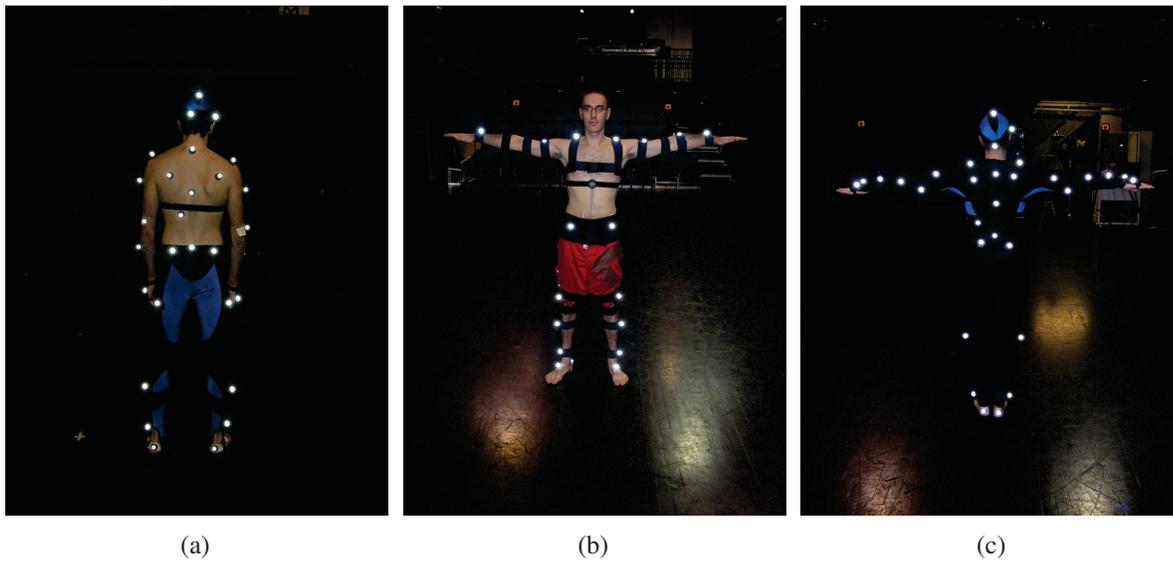


Figure 4.2 Différentes configurations de costumes et systèmes de fixation

gistrement toutes les deux secondes et calculait tous les paramètres métaboliques en temps réel pendant que le logiciel *Suunto Monitor* enregistrerait les données de rythme cardiaque à une fréquence d'un enregistrement par seconde et affichait un graphique ainsi que les zones de rythme cardiaque. L'utilisation de *Suunto Monitor* permettait non seulement le suivi de l'évolution du rythme cardiaque durant la session de capture afin d'assurer la variabilité dans les données, mais aussi la disponibilité de données dans les différents zones d'exercice.

Les données métaboliques fournies par la montre Suunto en temps réel et calculées à partir du rythme cardiaque par le logiciel *Suunto Training Manager* étaient :

- Fréquence cardiaque : Nombre de battements par minute ;
- R-R : Intervalle R-R en millisecondes entre deux battements cardiaques consécutifs ;
- Ventilation : Quantité d'oxygène entrant dans les poumons en litres par minute ;
- VO_2 : Consommation d'oxygène en millilitre par kilogramme par minute ;
- Fréquence respiratoire : Nombres de respirations par minute ;
- Consommation d'énergie : Énergie métabolique consommée en kilocalories par minute.

Une vidéo témoin a été enregistrée pour chaque séance de capture en utilisant une caméra *miniDV* reliée à la station de capture fournissant ainsi un enregistrement synchronisé avec les données de capture de mouvement (voir fig. 4.3). Ceci facilitait le travail de nettoyage des données ainsi que l'analyse préliminaire de celles-ci. Afin de faciliter la synchronisation des données métaboliques avec les données de capture de mouvement, l'horloge de la station de capture ainsi que celle de l'ordinateur portable ayant servi à enregistrer les données métaboliques ont été synchronisées avec la montre cardiofréquence-mètre. De plus, un décompte a été effectué avant le démarrage de chaque capture afin que les personnes chargées de lancer les divers enregistrements puissent les lancer simultanément.



Figure 4.3 Caméra témoin et caméras de capture de mouvement

Le poids des candidats ainsi que d'autres indicateurs corporels ont été mesurés à l'aide d'un moniteur de composition corporelle de type *Omron HBF516*. Cet appareil fournissait sept indicateurs corporels :

- Poids ;
- Métabolisme au repos ;
- Pourcentage de masse grasseuse ;
- Indice de masse corporelle (IMC) ;
- Niveaux de graisse viscérale ;
- Pourcentage de muscles squelettiques ;
- Âge corporel.

Seules les données de poids et de métabolisme au repos ont été utilisées pour l'élaboration des modèles de prédiction des paramètres métaboliques. Le poids des participants a été mesuré au début et à la fin de la session et une diminution moyenne de 0.2kg a été observée. Étant donnée que cette variation représentait moins de 0.3 % du poids des candidats, elle a été considérée comme négligeable et a été ignorée lors des calculs d'énergie mécanique. Seul le poids mesuré avant l'exercice a été utilisé.

4.2 Capture faciale

Les séances de capture faciale ont été effectuées au centre sportif de l'ÉTS. Afin d'enregistrer l'expression faciale, diverses configurations et caméras ont été testées. Dans un premier temps, des captures à deux caméras ont été réalisées avec deux types de matériel différent. La première configuration comprenait des caméras *miniDV* et la deuxième utilisait des caméras *PlayStation Eye* de Sony. Bien que la résolution des premières soit supérieure, la qualité obtenue en utilisant les caméras *PlayStation Eye* était supérieure étant donné que la compression était de meilleure qualité et que l'image était enregistrée en mode progressif, à l'inverse des caméras *miniDV* qui enregistraient en mode entrelacé souffraient de la compression *DV*. De plus, les deux pistes vidéo filmées avec les caméras *PlayStation Eye* étaient enregistrées directement sur ordinateur et étaient parfaitement synchronisées, ce qui évitait de devoir effectuer une synchronisation manuelle des deux pistes vidéo. L'enregistrement a été réalisé en utilisant le logiciel *iPi Recorder* de *iPi Soft*.

L'enregistrement vidéo était lancé en premier lieu puis un décompte était fait pour lancer l'enregistrement des données métaboliques sur la montre cardiofréquence-mètre et l'ordinateur portable simultanément. Un signal sonore émis lors du début de l'enregistrement des paramètres métaboliques permettait de synchroniser les données métaboliques avec les vidéos capturées.

Ayant constaté que la caméra filmant de face était suffisante pour capturer les expressions faciales, une seconde phase de capture a été réalisée avec une seule caméra miniature de type *GoPro HD* montée sur un casque. Deux types de casques ont été utilisés et étaient choisis selon l'utilisateur et les mouvements effectués : Un casque de vélo adapté pour pouvoir recevoir le système de fixation de la caméra, et un casque de protection adapté pour pouvoir y fixer une caméra *PlayStation Eye* ou la caméra *GoPro HD*. La figure 4.4 montre les caméras et les systèmes de fixations utilisés lors des captures. La caméra *GoPro HD* donnant une meilleure



Figure 4.4 Caméras et systèmes de fixation utilisés

qualité et résolution d'image comparé aux autres caméras utilisées, et les systèmes de fixation la maintenant fixe par rapport au visage, ceci simplifiait grandement le travail manuel d'animation

décrit dans la section 4.3. De plus, étant donné qu'elle enregistre directement sur une carte mémoire, l'utilisation d'un ordinateur ou de cassettes n'était pas nécessaire, ce qui simplifiait grandement le processus d'acquisition. Les candidats étaient aussi plus libres dans leurs gestes, étant donné qu'il n'y avait pas de fils reliant le casque à un ordinateur. Finalement, autant pour les exercices cardiovasculaires que ceux de musculation, le matériel et les logiciels utilisés pour la capture des paramètres métaboliques sont ceux décrits dans la section 4.1.

4.3 Acquisition manuelle des expressions faciales

Tel qu'indiqué plus tôt, afin de simplifier les sessions de capture, en plus des données métaboliques, seul un enregistrement vidéo a été réalisé lors des sessions de capture faciale. Afin de récupérer des données objectives pouvant être utilisées par les approches d'apprentissage machine, un visage de personnage 3D a été animé de façon à faire correspondre ses expressions faciales à celles du participant et ce tout au long de la vidéo. Ceci a été fait en enregistrant des poses clés des valeurs de chacune des quatre *blend shapes* du visage du personnage virtuel. Afin de faciliter et d'accélérer ce processus, un système a été développé où l'utilisateur peut voir côte à côte le visage virtuel et celui du participant (voir fig. 4.5 (1)).

Afin de permettre à l'utilisateur de défiler aisément à travers la ligne de temps, et ce à vitesse variable, une souris 3D de type *3Dconnexion Space Pilot Pro* a été convertie en dispositif de navigation de type *Jog/Shuttle* (voir fig. 4.5 (2)). L'utilisateur ajuste ensuite les intensités de chaque *blend shape* en utilisant une surface de contrôle de type *KORG nanoKONTROL* utilisée classiquement pour le mixage sonore (voir fig. 4.5 (3)). Les clés d'animations sont créées automatiquement lors des changements des valeurs d'intensité des *blend shapes* ce qui permet de simplifier le processus de création des clés d'animation. Ce processus a été répété pour chaque participant. Afin de s'assurer que les résultats sont reproductibles, les valeurs d'intensité des *blend shapes* ont été fixées une à une et ce dans le même ordre à chaque fois. De plus, afin de mesurer l'efficacité de notre processus d'entrée, trois personnes ont indépendamment ajusté les intensités de *blend shape*, sur une sélection de 18 poses représentatives. Bien que les valeurs n'aient pas été identiques, l'erreur était limitée à 11 % en moyenne. Ce taux d'erreur relativement bas a été considéré suffisant pour les besoins de ce travail.



Figure 4.5 Système utilisé pour l'animation en poses clés

Le choix des *blend shapes* utilisées a été grandement inspiré des groupes musculaires du visage humain tel que décrit dans le système FACS (Ekman et Friesen, 1978). Le prototype réalisé active un total de quatre *blend shapes*, basées sur les groupes musculaires importants identifiés dans la section 3.1. La figure 4.6 présente les quatre expressions faciales de base correspondant à la sélection de *blend shapes*.

4.4 Prototype

Un prototype a été développé en utilisant le moteur de jeu *Unity 3D* afin de tester l'approche en temps réel et de valider visuellement les prédictions résultantes. L'arbre de régression ainsi que les modèles de SVM entraînés ont été utilisés pour animer les expressions faciales du visage d'un personnage 3D. En utilisant une implémentation de *libSVM* pour effectuer les pré-

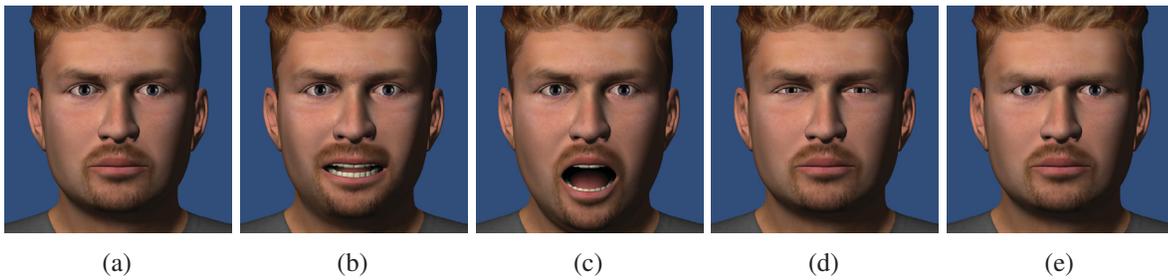


Figure 4.6 Les quatre *blend shapes* utilisées dans notre implémentation : (a) Neutre, (b) Étirement de la bouche, (c) Ouverture de la bouche, (d) Plissement des yeux, (e) Sourcils

dictions, cela a montré que les temps de calcul minimes permettaient une application en temps réel. La charge du CPU/GPU était en majorité dédiée au différents calculs de rendu et de déformation de maillage ce qui permettait dans nos tests d’atteindre des fréquences d’affichage allant de 60 à 600 images par seconde (selon la qualité de rendu et la complexité du maillage). Les tests ont été effectués sur un ordinateur portable muni d’un CPU *Intel iCore i7-2630QM* cadencé à 2.0GHz et d’un GPU *Nvidia GeForce GT555*. Le prototype peut être utilisé en déclenchant des animations pré-enregistrées ou en utilisant une entrée temps réel à partir d’un dispositif *Kinect*. L’utilisateur peut spécifier l’âge, la taille, le poids, le niveau d’entraînement et le vecteur d’expressivité du personnage 3D, et peut ajouter des poids portés dans chaque main ou sur le dos. L’expression faciale change en lien avec le mouvement du personnage et les paramètres spécifiés. L’utilisateur peut changer les paramètres du personnage et obtenir une rétroaction en temps réel. La figure 4.7 présente deux captures d’écran montrant l’utilisation du prototype avec des animations pré-enregistrées (Fig. 4.7(a)) ou une entrée temps-réel avec un dispositif *Kinect* (Fig. 4.7(b)).

4.5 Limitations

Tel qu’on peut le voir dans la figure 4.6, le modèle de *blend shapes* utilisé dans ce travail est suffisant, mais ne couvre pas la totalité des expressions faciales pouvant être observées. Par exemple, le cou ainsi que les joues pourraient être des régions où il serait pertinent de rajouter des *blend shapes*. De plus, il pourrait être pertinent d’animer séparément les paupières supérieures et inférieures.



(a)



(b)

Figure 4.7 Prototype réalisé : (a) mouvements pré-enregistrés (b) mouvements capturés avec un dispositif *Kinect*

Comme chaque *blend shape* a un modèle de prédiction indépendant et qu'il est donc calculé séparément des autres, il pourrait y avoir une certaine inconsistance dans le résultat global obtenu. La résolution de ces éventuelles inconsistances ainsi que la généralisation des résul-

tats d'animation pourraient être réalisés en utilisant un système de propagation contrainte des influences tel que celui décrit par Qing et Zhigang (2008).

Bien que le système utilisé pour l'acquisition manuelle des données ainsi que celui de l'implémentation soient basés sur des modèles 3D utilisant des *blend shapes*, le système pourrait être aussi facilement utilisé avec d'autres systèmes de contrôle tels que le contrôle par squelette de déformation.

Les expressions faciales générées sont une généralisation de toutes les données observées. De ce fait, elles correspondent à des valeurs moyennes et manquent parfois d'expressivité en comparaison avec les données capturées (voir fig. 4.8). De plus, l'approche proposée enchaîne deux modèles de prédiction. Ceci pourrait être la source de l'atténuation au niveau des résultats obtenus. Ceci peut aussi augmenter la possibilité d'accumulation d'erreur entre les deux modèles. Toutefois, les résultats obtenus durant les tests restaient stables pour des durées dépassant les dix minutes ce qui est supérieur aux durées de niveaux de jeu généralement rencontrés.



Figure 4.8 Comparaison entre les expressions faciales réelles et prédites d'une personne à une autre : (a) générées, (b-c) candidat différent faisant un même exercice. Les différences d'ouverture de la bouche et du plissement des yeux sont perceptibles mais les expressions restent ressemblantes

D'un autre côté, les modèles donnent parfois des résultats loin de la moyenne des données observées. De tels phénomènes ont été observés lors des tests réalisés (voir fig. 3.6(c)), particulièrement lorsqu'un dispositif *Kinect* est utilisé. Dans le cas de la *Kinect*, ils sont dûs aux

erreurs d'interprétation du positionnement du squelette ce qui se traduit en un travail mécanique nettement supérieur à celui humainement possible. Un phénomène semblable a été observé avec les animations pré-enregistrées lorsque les transitions entre les cycles d'animations ne sont pas parfaitement fluides. Comme ces erreurs de prédiction arrivent peu fréquemment et que les prédictions subséquentes sont dans la moyenne des données capturées, elles peuvent donc facilement être filtrées.

Étant donné que le modèle de prédictions métaboliques utilise la dernière prédiction comme entrée, il est donc sujet à l'accumulation d'erreur et peut diverger des données observées à travers le temps. Ceci peut devenir perceptible dans les animations de longue durée (voir fig. 3.6(b)). Bien que dans la plupart des cas, les niveaux de jeu ou les séquences de film soient de courte durée, dans des situations d'animations continues de longues durées, il serait nécessaire d'utiliser des approches permettant de remettre les données dans l'intervalle de celles observées.

L'utilisation de la régression SVM s'est avérée une solution efficace dans la prédiction d'expressions faciales. Cependant, elle reste une technique boîte noire qui ne laisse pas beaucoup d'options de personnalisation. Si les paramètres pouvant être contrôlés par les animateurs deviennent insuffisants, il serait difficile d'ajuster davantage la synthèse de l'animation.

CONCLUSION

La création d'animations faciales réalistes basées sur l'activité physique est une tâche difficile. En analysant deux ensembles de données capturées, ce mémoire révèle plusieurs observations importantes sur ce qui déclenche des expressions faciales spécifiques. Basée sur les données capturées, une combinaison de deux techniques d'apprentissage machine a été utilisée afin de déterminer automatiquement certains paramètres métaboliques et de synthétiser l'animation faciale d'un personnage 3D.

Bien qu'elle soit automatique, cette approche fournit des paramètres concrets que les animateurs peuvent ajuster afin d'obtenir des résultats d'animations faciales réalistes et convaincants qui s'ajustent automatiquement aux mouvements du personnage 3D et à son état de fatigue. De plus, les paramètres métaboliques fournis par cette approche pourraient également être utilisés afin d'animer d'autres aspects tels que la respiration ou la transpiration. Enfin, l'approche peut être utilisée dans des applications en temps réel ainsi que du rendu hors-ligne de haute qualité. Elle pourrait servir comme première passe pour l'animation manuelle allégeant ainsi grandement le travail des animateurs. L'approche devrait permettre d'obtenir des personnages plus réalistes tout en réduisant le coût lié à la capture faciale ou le fardeau de l'animation manuelle des expressions faciales liées à l'activité physique.

Tout au long de l'élaboration de l'approche proposée, certaines limites ont été remarquées. Les approches précédentes décrites dans la section 1 permettent de générer un seul aspect de l'expression faciale (celui lié à la parole, ou à l'émotion). De même, notre approche permet de générer les expressions faciales liées uniquement à l'activité physique. Un travail pertinent serait de fournir un cadre global qui pourrait permettre la fusion et la transition entre l'ensemble des éléments influençant les expressions faciales. Comme dans le travail de Arya *et al.* (2009), l'utilisation des unités d'action de FACS comme base devrait être un bon point de départ pour élaborer un tel cadre.

Bien que l'approche proposée donne des résultats réalistes, elle est de nature déterministe : pour des entrées identiques de données de mouvement, elle donnerait toujours une expression

faciale identique. Une avenue intéressante pour la recherche future, serait d'intégrer une nature probabiliste et stochastique se rapprochant plus de la nature des réactions humaines.

LISTE DE RÉFÉRENCES

- Arya, A., S. DiPaola, et A. Parush. 2009. « Perceptually Valid Facial Expressions for Character-based Applications ». *International Journal of Computer Games Technology*, vol. 2009.
- Byun, M. et N.I. Badler. 2002. « FacEMOTE : Qualitative Parametric Modifiers for Facial Animations ». In *Proc. of 2002 ACM SIGGRAPH/Eurographics Symposium on Computer Animation*. p. 65–71. ACM.
- Cao, Y., W.C. Tien, P. Faloutsos, et F. Pighin. October 2005. « Expressive Speech-driven Facial Animation ». *ACM Trans. Graph.*, vol. 24, p. 1283–1302.
- Chuang, E., F. Deshpande, et C. Bregler. 2002. « Facial Expression Space Learning ». In *Proc. of 10th Pacific Conference on Computer Graphics and Applications 2002*. p. 68–76.
- De Martino, J., L.P. Magalhães, et F. Violaro. 2006. « Facial Animation Based on Context-dependent Visemes ». *Computers & Graphics*, vol. 30, n° 6, p. 971–980.
- Deng, Z. et X. Ma. 2008. « Perceptually Guided Expressive Facial Animation ». In *Proc. of 2008 ACM SIGGRAPH/Eurographics Symposium on Computer Animation*. p. 67–76.
- Deng, Z. et U. Neumann. 2008. « Expressive Speech Animation Synthesis with Phoneme-Level Controls. ». *Computer Graphics Forum*, vol. 27, n° 8, p. 2096–2113.
- Ekman, P. et W.V. Friesen, 1978. *Facial Action Coding System : Investigator's Guide*. Consulting Psychologists Press.
- Ezzat, T., G. Geiger, et T. Poggio. July 2002. « Trainable Videorealistic Speech Animation ». *ACM Trans. Graph.*, vol. 21, p. 388–398.
- Faigin, G., 1990. *The Artist's Complete Guide to Facial Expression*. Watson-Guptill Publications.
- Firstbeat Technologies. February 2007a. « Indirect EPOC Prediction Method Based on Heart Rate Measurement ». White paper, Firstbeat Technologies. <http://www.firstbeat.net/files/white_paper_epoc.pdf>. Available online (5 pages).
- Firstbeat Technologies. February 2007b. « EPOC Based Training Effect Assessment ». White paper, Firstbeat Technologies. <http://www.firstbeat.net/files/white_paper_training_effect.pdf>. Available online (5 pages).
- Goldfinger, E., 1991. *Human Anatomy for Artists : The Elements of Form*. Oxford University Press.
- Gutierrez-Osuna, R., P.K. Kakumanu, A. Esposito, O.N. Garcia, A. Bojorquez, J.L. Castillo, et I. Rudomin. feb. 2005. « Speech-driven Facial Animation with Realistic Dynamics ». *Multimedia, IEEE Transactions on*, vol. 7, n° 1, p. 33–42.

- Hall, M., E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, et I.H. Witten. November 2009. « The WEKA Data Mining Software : An Update ». *SIGKDD Explor. Newsl.*, vol. 11, p. 10–18.
- Koch, R., M.H. Gross, et A.A. Bosshard. 1998. « Emotion Editing using Finite Elements ». *Computer Graphics Forum*, vol. 17, n° 3, p. 295–302.
- Kring, A. et D.M. Sloan. JUN 2007. « The Facial Expression Coding System (FACES) : Development, validation, and utility ». *Psychological Assessment*, vol. 19, n° 2, p. 210–224.
- Kshirsagar, S. et N. Magnenat-Thalmann. 2003. « Visyllable Based Speech Animation. ». *Computer Graphics Forum*, vol. 22, n° 3, p. 631–639.
- Kshirsagar, S., T. Molet, et N. Magnenat-Thalmann. 2001. « Principal Components of Expressive Speech Animation ». In *Proc. of Computer Graphics International 2001*. p. 38–44.
- Li, S., A.K. Jain, Y.-L. Tian, T. Kanade, et J.F. Cohn. 2005. Facial expression analysis. *Handbook of Face Recognition*, p. 247–275. Springer New York.
- Mauss, I. et M.D. Robinson. 2009. « Measures of Emotion : A Review ». *Cognition & Emotion*, vol. 23, n° 2, p. 209–237.
- McKenzie, R. October 1905. « The Facial Expression of Violent Effort, Breathlessness, and Fatigue ». *Journal of Anatomy and Physiology*, vol. 40, p. 51–56.
- Osipa, J., 2008. *Facial Modeling And Animation : Stop Staring, 2nd Ed (W/Cd)*. Wiley India Pvt. Ltd.
- Puklavage, C., A. Pirela, A.J. Gonzalez, et M. Georgiopoulos. May 2010. « Imitating Personalized Expressions through an Avatar using Machine Learning ». In *Proc. of 23rd International Florida Artificial Intelligence Research Society Conference (FLAIRS)*. p. 68–73. Association for the Advancement of Artificial Intelligence.
- Qing, L. et D. Zhigang. 2008. « Orthogonal-Blendshape-Based Editing System for Facial Motion Capture Data ». *Computer Graphics and Applications, IEEE*, vol. 28, n° 6, p. 76–82.
- Rusko, H., A. Pulkkinen, S. Saalasti, E. Hynynen, et J.K. KIHU. 2003. « Pre-Prediction of Epoc : A Tool for Monitoring Fatigue Accumulation During Exercise ? ». *Medicine and Science in Sports and Exercise*, vol. 35, n° 5, p. S183.
- Russell, J. Décembre 1980. « A Circumplex Model of Affect. ». *Journal of Personality and Social Psychology*, vol. 39, n° 6, p. 1161–1178.
- Russell, J. A. et Albert Mehrabian. 1977. « Evidence for a three-factor theory of emotions ». *Journal of Research in Personality*, vol. 11, n° 3, p. 273–294.

- Ruttkay, Z., H. Noot, et P. Ten Hagen. 2003. « Emotion Disc and Emotion Squares : Tools to Explore the Facial Expression Space ». *Computer Graphics Forum*, vol. 22, n° 1, p. 49–53.
- Sagar, M. 2006. « Facial Performance Capture and Expressive Translation for King Kong ». In *ACM SIGGRAPH 2006 Sketches*. (New York, NY, USA 2006).
- Sifakis, E., A. Selle, A. Robinson-Mosher, et R. Fedkiw. 2006. « Simulating Speech with a Physics-based Facial Muscle Model ». In *Proc. of 2006 ACM SIGGRAPH/Eurographics symposium on Computer animation*. p. 261–270.
- Swain, D. et B.C. Leutholtz. March 1997. « Heart Rate Reserve is Equivalent to %VO₂ Reserve, Not to %VO₂max. ». *Med Sci Sports Exerc*, vol. 29, p. 410–414.
- Thayer, R., 1989. *The Biopsychology of Mood and Arousal*. Oxford University Press.
- Unuma, M., K. Anjyo, et R. Takeuchi. 1995. « Fourier Principles for Emotion-based Human Figure Animation ». In *Proc. of 22nd annual conference on Computer graphics and interactive techniques*. p. 91–96. ACM.
- Winter, D., 2005. *Biomechanics and Motor Control of Human Movement*. John Wiley & Sons.
- Winter, D., 2009. *Biomechanics and Motor Control of Human Movement*. Wiley.
- Witten, I. et E. Frank, 2005. *Data Mining : Practical Machine Learning Tools and Techniques*. Morgan Kaufmann series in data management systems. Morgan Kaufman.
- Witten, I., E. Frank, et M.A. Hall, 2011. *Data Mining : Practical Machine Learning Tools and Techniques*. The Morgan Kaufmann Series in Data Management Systems. Elsevier Science.
- Zhao, L., A. Normoyle, S. Khanna, et A. Safonova. 2009. « Automatic Construction of a Minimum Size Motion Graph ». In *Proc. of 2009 ACM SIGGRAPH/Eurographics Symposium on Computer Animation*. p. 27–35.