

Université de Montréal

Composition automatique d'expressions
faciales synthétiques à partir de
capture de postures faciales

par

Cyriaque Kouadio

Département d'informatique et de recherche opérationnelle

Faculté des arts et des sciences

Mémoire présenté à la faculté des études supérieures
en vue de l'obtention du grade de Maître ès sciences (M.Sc.)
en informatique

Janvier 1997

© Cyriaque Kouadio, 1997

Université de Montréal
Faculté des études supérieures

Ce mémoire de maîtrise intitulé

Composition automatique d'expressions
faciales synthétiques à partir de
capture de postures faciales

présenté par
Cyriaque Kouadio

a été évalué par un jury composé des personnes suivantes:

Président:	Neil Stewart
Directeur de recherche:	Pierre Poulin
Membre:	Jean Meunier

Sommaire

L'animation de personnages synthétiques par ordinateur constitue un domaine d'intérêt grandissant non seulement pour la communauté scientifique en infographie mais aussi pour les artistes qui veulent exprimer leurs talents par cet outil qu'est l'ordinateur. La particularité du visage en tant que véhicule de communication de sensations mais aussi pour son anatomie, fait qu'il lui est réservé un traitement à part en animation de personnages. Plusieurs approches spécialisées de modélisation et de contrôle des déformations du visage sont proposées de nos jours, partant de la paramétrisation du visage, à la modélisation des mécanismes physiologiques des tissus faciaux, en passant par l'interpolation des expressions faciales.

L'approche d'animation des expressions faciales présentée dans ce mémoire se base sur des concepts existant dans un système d'animation commercial, i.e. *TAARNA*, et sur l'hypothèse suivante: en définissant un système de coordonnées pour les expressions d'un visage synthétique donné, il est possible de synthétiser toutes les expressions que ce visage peut prendre. Vu la gamme des visages à animer et le type d'expressions désiré (expressions possiblement exagérées) dans le contexte de production qu'est celui de *TAARNA*, la méthode adoptée s'intègre bien dans l'environnement du système d'animation déjà existant. Le choix du système de coordonnées est essentiel pour permettre la détermination exacte des coordonnées d'expressions faciales. Pour ce faire, une bibliothèque d'expressions élémentaires de référence est utilisée pour définir ce référentiel. Par la suite, une vérification de la bibliothèque d'expressions est faite pour en éliminer les inconsistences, suivie d'un traitement dynamique des expressions afin d'obtenir un référentiel dans lequel les directions sont le plus possible indépendantes les unes des autres. Si le référentiel est bien choisi, il devrait être possible de déterminer les coordonnées d'une expression faciale donnée par résolution d'un système d'équations dont les inconnues sont ces-dites coordonnées.

Une façon de spécifier l'expression faciale désirée est de la capturer sur le visage d'un acteur réel grâce à un système de capture d'expressions faciales. L'animation du visage d'un personnage synthétique peut ainsi être contrôlée par les expressions faciales d'un acteur réel. L'outil d'animation facial développé dans le cadre du projet de maîtrise faisant l'objet de mémoire permet d'assurer ce contrôle d'animation. Les coordonnées d'expression sont déterminées automatiquement. Pour améliorer la synthèse de l'expression faciale obtenue à partir des coordonnées, le contrôle des paupières et des accessoires sont ajoutés. Une trame sonore enregistrée lors de la capture des expressions faciales de l'acteur réel est enfin mise en présence de la séquence des expressions faciales synthétiques pour valider les synthèses effec-

tuées (veiller à ce qu'il y ait correspondance entre le son et les déformations du visage synthétique). La méthode de détermination automatique adoptée pour les coordonnées d'expression permet de trouver celles-ci en un temps de calcul suffisamment court pour contrôler l'animation des expressions faciales de synthèse en temps réel. La qualité des expressions faciales obtenues pour une banque d'expressions bien choisies sont remarquables.

L'approche développée dans ce mémoire apporte des améliorations notables à la méthode d'animation faciale à partir d'une bibliothèque d'expressions de base et au système d'animation faciale de *TAARNA*. Elle est utilisable en temps réel, permet d'animer la face de divers types de personnages synthétiques (humains, animaux, *cartoons*) et étend la gamme des expressions faciales recomposables au delà de la matrice initiale des expressions de la banque. Par ailleurs, une formulation mathématique de la composition des expressions synthétiques est proposée. L'intervention de l'artiste est réduite sans toutefois lui enlever l'opportunité d'imprimer son talent dans le résultat final de l'animation.

Mots-clés:

animation, animation faciale, expression faciale, capture de mouvement, moindres carrés, espace vectoriel.

Table des matières

1	Introduction	13
2	Animation faciale	16
3	Travaux antérieurs en animation faciale	19
3.1	Méthodes de modélisation	21
3.1.1	Bibliothèque d’expressions et interpolation	21
3.1.2	Modélisation paramétrique	21
3.1.3	Modélisation des muscles de la face	22
3.1.4	Modélisation des tissus faciaux à partir de la physique des corps déformables	24
3.2	Contrôle de l’animation finale	26
3.2.1	Contrôle d’animation à base de règles	26
3.2.2	Contrôle d’animation à base d’analyse	26
3.2.3	Contrôle d’animation à base de performance	26
3.3	Le système d’animation de <i>TAARNA</i>	27
3.3.1	L’animation en général	27
3.3.2	Animation faciale	28
4	Approche par une banque d’expressions	30
4.1	La banque d’expressions	32
4.2	Capture des postures faciales (<i>Tracking</i>)	35
4.2.1	Système de capture à base de réflexion lumineuse	36
4.2.2	Répartition des marqueurs	37
4.2.3	Problèmes relatifs à la capture des postures faciales	39

4.3	Association des marqueurs aux points de contrôle de la face	41
4.3.1	Modes d'association	43
4.4	Choix des expressions nécessaires de la banque	49
4.5	Subdivision des expressions	51
4.6	Contraintes sur les coordonnées admissibles des expressions	51
4.7	Le système d'animation faciale	52
5	Modèle mathématique	54
5.1	Données de la banque d'expressions	55
5.2	Expression faciale désirée	56
5.3	Mise à l'échelle	56
5.3.1	Calcul des facteurs d'échelle des marqueurs	56
5.3.2	Déplacement final	58
5.4	Formulation mathématique du problème	59
5.5	Méthodes de résolution	60
5.5.1	Résolution pseudo-directe	61
5.5.2	Résolution par algorithme de minimisation de fonction	63
5.5.3	Résolution par moindres carrés linéaires	63
5.6	Choix des expressions nécessaires de la banque	66
5.7	Utilisation du résultat de la résolution	67
6	Résultats	69
6.1	Validation de la méthode utilisée	69
6.1.1	Illustration de la condition nécessaire de consistance	70
6.1.2	Illustration de l'influence du domaine de définition du système d'équations	74
6.1.3	Recomposition d'expression de la banque	77
6.1.4	Correspondance avec le son	82
6.2	Ajout des accessoires	83
7	Amélioration des résultats	87
7.1	Filtrage	87
7.1.1	Filtrage des résultats de capture	87

<i>TABLE DES MATIÈRES</i>	v
7.1.2 Filtrage des résultats de la minimisation	89
7.2 Traitement des paupières	90
8 Améliorations potentielles	93
8.1 Améliorations à court terme	93
8.1.1 Mouvement des yeux	93
8.1.2 Capture de mouvement en 3D	94
8.2 Améliorations nécessitant un effort d'implantation	95
8.2.1 Automatisation	95
8.2.2 Coordonnées d'expression et expression de synthèse	96
8.2.3 Déplacement hiérarchique des points liés	97
8.2.4 Synchronisation du son	97
9 Conclusion	99
Bibliographie	101

Liste des tableaux

3.1 Valeurs typiques des constantes de raideur des couches [7]	26
--	----

Table des figures

3.1	Muscles faciaux (tiré de [20])	23
3.2	Couches de tissus faciaux (tiré de [7])	25
4.1	Exemple de banque d'expressions (modèles)	33
4.2	Expressions vectorielles - Expressions et combinaisons	34
4.3	Actrice réelle avec le casque	37
4.4	Répartition des marqueurs virtuels et réels	38
4.5	Difficulté dans la répartition des marqueurs	38
4.6	Mélange de marqueurs	39
4.7	Perte de marqueurs	40
4.8	Réflexion parasite	40
4.9	Marqueur de référence et répartition des marqueurs réels	41
4.10	La face portant le masque	42
4.11	Association point à points	44
4.12	Association par zone d'influence	45
4.13	Fonction de poids des marqueurs	45
4.14	Influences de plusieurs marqueurs (deux) sur un point	45
4.15	Association polygone à points	47
4.16	Coordonnées barycentriques (triangle)	48
4.17	Calcul de la composante en Z	49
4.18	Subdivision d'expression	52
5.1	Mise à l'échelle de mouvements de marqueurs	55
5.2	Boîtes englobantes des marqueurs et des points liés	58

5.3	Expressions $[e]$ et $[e_c]$	68
6.1	Consistence	71
6.2	Banque d'expressions ne contenant aucun mouvement de la bouche	71
6.3	Consistence	72
6.4	Banque d'expressions	73
6.5	Expression composée avec banque de la figure 6.4	74
6.6	Expression composée à partir d'un domaine de définition insuffisant	76
6.7	Expression composée avec $[0.0] \leq [\lambda] \leq [1.0]$	76
6.8	Expression "F" de la banque de référence de la figure 6.4 ($[\lambda] = [0, 1, 0, 0, \dots, 0]$)	78
6.9	Reconstitution de l'expression "F" de la banque de référence de la figure 6.4 à partir de l'expression "E"	78
6.10	Reconstitution de l'expression "F" de la banque de référence de la figure 6.4 à partir de trois expressions	79
6.11	Reconstitution de l'expression "F" de la banque de référence de la figure 6.4 à partir de quatre expressions	79
6.12	Reconstitution de l'expression "F" de la banque de référence de la figure 6.4 à partir de toutes les expressions exceptée "F"	80
6.13	Reconstitution de l'expression "F" de la banque de référence de la figure 6.4 à partir de toutes les expressions	80
6.14	Expressions de la banque à reconstituer	81
6.15	Reconstitution de l'expression "E" de la banque de référence de la figure 6.4	81
6.16	Reconstitution de l'expression "kiss" de la banque de référence de la figure 6.4	82
6.17	Exemple d'accessoires	83
6.18	Accessoires (dents,bouche,langue) ajoutés au visage	83
6.19	Banque de postures de référence des dents	84
6.20	Définition des liens entre les postures de référence d'accessoires et de la face	84
7.1	Filtre passe-bas (équation 7.1)	88
7.2	Processus de filtrage	89
7.3	Exemple de banque de paupières (oeil gauche du point de vue du personnage)	91
7.4	Correspondance entre variations des paramètres des paupières supérieures	91

TABLE DES FIGURES

7.5	Correspondance entre variations des paramètres des paupières inférieures	91
8.1	Expression à prédominance en Z	94

Remerciements

Je témoigne ma très sincère reconnaissance à mon directeur de recherche, le professeur Pierre Poulin qui a eu confiance en moi. Je le remercie pour son soutien, sa disponibilité, ses brillantes suggestions à toutes les étapes de la maîtrise, mais aussi pour sa grande expérience en infographie dont j'ai eu la chance de bénéficier.

J'exprime toute ma gratitude au directeur de la compagnie *TAARNA Studios inc.*, monsieur Pierre Lachapelle pour m'avoir donné l'opportunité de réaliser mon projet au sein d'une entreprise, dans un environnement dynamique et stimulant. Je lui suis reconnaissant pour ses directives et suggestions qui témoignent de sa grande expérience dans le domaine de l'infographie et qui m'ont évité de m'égarer sur le sentier de la connaissance.

J'aimerais également adresser mes remerciements à toute l'équipe de *TFX Animation* pour l'aide très appréciable qu'elle m'a apportée en effectuant les tests du projet. Particulièrement à Patricia Pawlak, Grégory Roy, Stephen Menzies et Frank Ritlop.

Je remercie toute l'équipe de développement de *Digits'n Art Software inc.* pour les suggestions et discussions qui m'ont permis de mieux comprendre le système d'animation à l'intérieur duquel ce projet a été réalisé. J'adresse des remerciements particuliers à Benoît Lafleur pour ses explications et suggestions pour ce qui est des méthodes numériques.

Je tiens à remercier l'ensemble du personnel de *TAARNA Studios inc.*, *Digits'n Art Software inc.* et *TFX Animation* pour l'atmosphère chaleureuse, dynamique et motivante. Je remercie également Kaveh Kardan pour ses explications et son travail initial sur les expressions faciales dans le système d'animation de *TAARNA*.

Je remercie vivement tous(tes) mes amis(es) pour le soutien qu'ils représentent pour moi. Particulièrement Soro Nongolougo pour l'amitié qu'il me témoigne depuis tant d'années et dont les suggestions pendant la révision du mémoire m'ont permis d'en améliorer la qualité.

J'aimerais témoigner toute ma gratitude à mon amie Diane El Abidi pour son amour, son soutien et ses encouragements pendant les moments difficiles. Je lui suis très reconnaissant de m'avoir poussé à choisir ce projet et pour ses interventions judicieuses pendant la révision du mémoire.

Je remercie les membres du laboratoire d'infographie pour leur contribution à y maintenir une atmosphère agréable et dont l'enthousiasme et le dévouement à l'infographie incite au travail et à pousser plus loin l'activité de recherche. Un merci tout particulier à Normand Brière avec qui les discussions m'ont fait voir certains aspects de la vie d'un autre oeil et aller plus loin dans les idées.

Je témoigne ma profonde gratitude au gouvernement de la république de Côte d'Ivoire pour m'avoir donné l'opportunité d'effectuer mes études de second cycle en m'octroyant une bourse d'étude.

À ma très chère famille qui m'a toujours soutenu.

À Diane qui partage ma vie.

À mes amis; vous êtes chers à mon coeur.

Je vous aime très fort.

Chapitre 1

Introduction

Le défi de construire et animer par ordinateur des personnages humains ou animaux est rendu complexe par la nécessité de produire et contrôler des déformations *douces* et réalistes de la surface de la peau; notamment au voisinage des articulations du squelette. En outre, un observateur devient très critique et a de grandes attentes lorsqu'il est placé en présence d'un personnage synthétique dont l'apparence est moindrement réaliste.

Dans ce contexte de complexité de modélisation des êtres vivants animés, la face constitue une particularité, tant dans sa structure anatomique que son aspect visible. En effet, non seulement le nombre d'éléments spécifiques de la structure osseuse de la face ainsi que les muscles et tissus cutanés faciaux sont très élevés, mais aussi, les mécanismes d'interaction entre muscles et os ou intermusculaires sont compliqués. Ces interactions sont communément appelées expressions faciales.

Le projet de maîtrise qui fait l'objet de ce mémoire a consisté à développer une approche d'animation des expressions faciales de personnages synthétiques complexes. Ces personnages peuvent être de toutes sortes, allant du visage humain à celui d'un *cartoon*, en passant par des animaux.

La modélisation et l'animation du visage de personnages synthétiques intéressent les chercheurs en infographie pour la raison principale qu'il est difficile de créer des déformations expressives de surface. Un modèle de la face se définit comme étant une abstraction mathématique qui capture à un certain niveau d'exactitude la forme et la fonction de la face, qu'elle soit humaine ou pas, de sorte que le modèle soit utilisable pour l'application spécifique qui en est prévue.

L'animation faciale est une activité pluridisciplinaire, faisant appel à des connaissances acquises dans des domaines de recherche scientifique tels que la médecine, la biologie (anatomie), la psychologie, les mathématiques et l'informatique (modélisation de surface), mais aussi à la capacité de pouvoir modé-

liser l'état émotif de la face. Elle trouve son application également dans plusieurs secteurs de l'activité humaine, dont l'éducation, la narration, l'amusement, la médecine, la criminologie et la psychologie pour n'énumérer que ceux-là.

Pour être convainquante et réaliste, l'animation des expressions faciales doit correspondre à ce qu'un observateur attend du comportement d'un visage réel. Le système permettant de produire ces animations se doit donc d'avoir plusieurs caractéristiques jugées importantes, en fonction du contexte d'utilisation. Entre autres, un système doit offrir:

- la possibilité de modélisation du visage (avec ou sans expression),
- la possibilité de modélisation d'expressions émotives.

Il doit aussi permettre le contrôle de la succession des expressions faciales dans le temps, le contrôle de zones particulières du visage, et cela éventuellement de manière indépendante. La synchronisation des expressions faciales avec le son doit également être possible.

Il faudrait aussi pouvoir visionner la séquence d'animation pour d'éventuelles retouches.

Un feed-back en temps réel de modifications apportées au modèle ou à des paramètres de l'animation est essentiel pour permettre à l'utilisateur d'ajuster la modélisation et/ou l'animation.

Le contexte de développement et d'utilisation du présent projet est celui d'une entreprise offrant des systèmes destinés à des artistes, dans un but de production de films d'animation. C'est donc un contexte commercial. L'outil d'animation faciale développé a pour but de permettre d'animer (presque) tous les aspects du visage d'un personnage synthétique. L'animation du visage synthétique peut être contrôlée à partir de capture de mouvements faciaux effectués par un acteur réel. Il devrait ainsi permettre de réaliser l'animation du modèle de la face en temps réel, pendant que l'acteur réel effectue des expressions faciales (pendant qu'il parle), de manière à ce qu'il y ait une synchronisation entre le son et l'image. L'outil d'animation faciale développé est prévu pour fonctionner dans un système d'animation global de tout le corps (visage, corps, bras, jambes, mains), soit le système d'animation de *TAARNA*.

Plusieurs approches sont actuellement proposées pour animer le modèle synthétique de la face. Ces approches incluent l'interpolation d'expressions, la déformation de la face à partir de spécification de paramètres, le contrôle de muscles abstraits et la simulation des mécanismes faciaux à partir de la physique. Chacune de ces approches comporte des avantages spécifiques mais aussi des points faibles qui seront introduits un peu plus loin dans ce mémoire.

Vus les besoins et le contexte de développement et d'évolution du projet, l'approche adoptée est gran-

dement basée sur l'interpolation d'expressions, en essayant de mitiger, voire éliminer ses points faibles. Néanmoins, certains aspects des autres approches sont incorporés via la modélisation initiale de la face. En outre, il y a possibilité d'intégrer des éléments des méthodes à base de muscles abstraits et à base de physique dans le système d'animation.

Le travail effectué permet d'améliorer une approche d'animation de personnages synthétiques existante (synthèse d'expressions faciales par combinaison d'expressions contenues dans une bibliothèque). Ce choix a été influencé par la complexité et la gamme de personnages synthétiques que le système devra supporter, mais aussi par le type d'animation à supporter (exagération d'expressions physiquement impossibles) et le fait que le système d'animation de *TAARNA* utilisait déjà cette approche jusqu'à un certain point. Des modifications effectuées sur la manière d'utiliser l'information pertinente contenue dans la bibliothèque d'expressions faciales font que cette approche devient plus flexible. Compte tenu du contexte d'évolution de l'outil d'animation faciale, c'est-à-dire le système d'animation global de *TAARNA*, il est possible d'enrichir cette approche de certains avantages des autres méthodes d'animation.

Pour ce qui est de la structure du mémoire, une présentation de l'animation faciale en générale est faite pour commencer, suivie des travaux antérieurs effectués dans le domaine. Les détails de l'approche de modélisation d'animation adoptée sont présentés ainsi que la modélisation mathématique du problème de la synthèse d'une expression faciale à partir d'expressions élémentaires contenues dans une bibliothèque. Par la suite, les résultats obtenus et les tests effectués sont présentés. Pour finir, les améliorations possibles à apporter à l'outil d'animation faciale développé sont introduites.

Chapitre 2

Animation faciale

Littéralement, animer signifie amener à la vie. L'action d'animer s'applique à tout changement pouvant avoir un effet visuel. Les variations des caractéristiques telles que la position, la couleur, la structure d'une entité dans le temps sont donc des animations de celle-ci. Une variation, dans le temps, du point de vue d'une scène produit également des effets visuels qui sont considérés comme animation.

De ce fait, la notion de temps est très importante en animation. L'animation est utilisée dans bon nombre de secteurs de l'activité humaine dont l'amusement, l'éducation et l'industrie entre autres.

L'animation conventionnelle est un processus manuel qui consiste à dessiner les positions clé de l'entité à animer. Par la suite, les positions intermédiaires sont interpolées compte tenu de la fréquence à laquelle l'animation finale est désirée (typiquement 24 à 30 images par seconde pour l'oeil humain [3]).

En infographie et plus particulièrement en animation de personnage, le processus de génération d'animation consiste à construire un modèle mathématique de l'entité à animer et à produire des images de l'état ou configuration de ce modèle à des instants donnés. Tout comme en animation conventionnelle, il faut générer les variations de configuration à une certaine fréquence (un certain nombre d'images de configuration du modèle par unité de temps) pour avoir l'impression de continuité de la séquence de celles-ci. Une manière de simplifier la tâche fut d'emprunter l'idée des positions clé de l'animation conventionnelle et de faire l'interpolation du reste des configurations de façon à obtenir la fréquence nécessaire pour avoir la continuité. C'est une approche d'animation qui est très simple à implanter mais qui fait surgir certains problèmes car en animation par ordinateur, c'est le modèle de l'entité, généralement en $3D$, qui est modifié au cours du temps, contrairement à l'image ($2D$) de celle-ci, comme c'est le cas en animation conventionnelle. Le problème de l'interpolation de la forme de l'entité se fait sentir. Par ailleurs, non seulement il faut interpoler dans l'espace (le modèle) mais aussi dans le temps pour assurer une vi-

tesse cohérente de l'animation globale. Pour pallier à ces problèmes, d'autres approches ont été utilisées pour faire l'animation de modèles en infographie.

La modélisation mathématique de certains phénomènes permet de les animer en faisant varier les paramètres du modèle dans le temps. Une procédure se charge de faire varier ces paramètres dans le temps (animation procédurale). Il est ainsi possible de faire toute l'animation du modèle sans avoir à interpoler entre des configurations clé de celui-ci. Cependant, l'efficacité de l'approche est conditionnée par la complexité du modèle et de l'approche de modélisation choisie. Toutefois, elle a le mérite de permettre une simulation du phénomène modélisé. Une façon de contrôler les paramètres du modèle dans le temps est d'aller chercher l'information nécessaire par capture de données sur l'entité réelle dont le modèle a été réalisé.

Il est également possible de spécifier par instruction les configurations désirées pour certains systèmes [3]. L'animation peut ainsi être générée par une liste d'instructions au système d'animation (langage d'animation). Les courbes paramétriques sont également utilisées pour contrôler des paramètres de modèles. En effet, compte tenu des propriétés de celles-ci et de la manière de les contrôler qui est bien connue de la communauté infographique, leur utilisation permet de rendre le contrôle des paramètres de modèles plus lisse et relativement plus facile.

Dans le champ de l'animation par ordinateur, l'animation du corps humain semble plus complexe car les écarts par rapport à la réalité sont plus facilement détectables par un observateur. Toutefois, à l'exception du visage, le corps est constitué de parties plus rigides dont la plupart des mouvements sont associés à ceux du squelette. Pour ce qui est du visage, c'est une partie de l'anatomie qui comporte beaucoup de subtilités et dont les mouvements sont plus souvent des déformations que des déplacements rigides. En fait, le visage transmet plus d'émotions que l'ensemble du reste du corps. Il est donc justifié de considérer son animation de façon particulière.

Tout comme pour l'animation en général, l'animation faciale consiste à produire des variations dans le temps de configurations du modèle de la face. Cela conduit à l'utilisation de certaines méthodes de l'animation telles que la modélisation de l'entité à animer, l'interpolation de positions intermédiaires à partir de positions clé, et la variation dans le temps de paramètres de contrôle des postures du modèle.

La modélisation de la face n'est pas une tâche des plus simples car les muscles en présence sont de diverses sortes et leurs interactions sont très complexes. La modélisation et l'animation de la face intéressent les chercheurs en infographie et les modélisateurs de personnages synthétiques depuis les années 70 du fait des nombreux problèmes rencontrés lorsqu'il était question d'animer le visage de personnages

synthétiques de manière à rendre les expressions les plus réalistes et les plus crédibles possibles, compte tenu de la grande attente des observateurs lorsqu'ils sont en présence d'un visage qui se veut réaliste.

La modélisation de la face doit tenir compte de la possibilité de déformation de celle-ci comme un corps mou, des modifications possibles de la position et de la forme de parties ou de la totalité du visage, de la satisfaction des aspects structurels et fonctionnels dans les mouvements du visage (dynamique des mouvements du visage qui produisent les expressions faciales) et du contrôle de régions particulières de façon indépendante. Pour arriver à des résultats acceptables, le niveau de détail de la modélisation doit être relativement élevé (modélisation de la surface de la peau et des tissus sous-cutanés). En outre, l'animation du visage fait intervenir l'ajustement de plusieurs paramètres dont l'extraction est déjà très difficile en soit.

En plus des différents points énoncés plus haut, il est essentiel en bout de ligne, que la voix soit synchrone à la séquence des expressions faciales générées pour que l'animation soit complète et crédible.

Les travaux de recherche entrepris par les laboratoires d'informatique des universités et ceux des départements de recherche et développement des compagnies ayant affaire à l'animation de personnages dans leurs activités ont conduit à une meilleure connaissance de différents aspects de la face dans le milieu de la recherche informatique et en particulier en infographie. Ces études se sont basées sur des résultats obtenus dans d'autres disciplines de recherche scientifique telles la médecine, la psychologie, la physique et les sciences cognitives.

La modélisation de la face bénéficie actuellement de ces résultats dans la mesure où, partie d'une modélisation uniquement de la surface de la peau, il est maintenant possible de faire des modèles des structures internes de la face de manière à simuler leurs comportements afin de produire des déformations expressives du visage. En outre ces déformations expressives du visage sont en concordance avec les résultats prévus dans d'autres disciplines scientifiques (possibilité de produire une expression émotive en termes de psychologie).

Une présentation sommaire de certaines des études effectuées en infographie en ce qui concerne la face, est faite dans le chapitre qui suit.

Chapitre 3

Travaux antérieurs en animation faciale

La modélisation et l'animation du visage, qu'il soit humain ou non, ont longtemps fasciné les chercheurs en infographie non seulement pour l'omniprésence des visages dans le monde réel, mais aussi pour la génération des déformations de la surface du modèle de manière expressive. En fait le caractère particulier du visage et le fait de son importance dans les media de communication interpersonnelle font qu'il constitue un sujet de recherche dans plusieurs champs d'activité de la recherche scientifique. Ainsi, il se retrouve en télécommunication, en médecine, en science éducative, dans les environnements virtuels et dans les sciences cognitives. Dans chacune de ces disciplines, la modélisation du visage est faite conformément aux besoins qui se présentent.

En ce qui concerne l'infographie et en particulier l'animation de personnages, le facteur déterminant est de comprendre les mécanismes contrôlant les déformations du visage de façon à pouvoir faire une modélisation des plus réalistes possible tout en permettant de produire une variation des propriétés du modèle dans le temps pour produire une animation, et cela à un coût de calcul et de stockage d'information acceptable.

Paul Ekman et Wallace Friesden [9], deux psychologues de la communication non verbale, développèrent en 1977 le système *FACS* (*Facial Action Coding System*), un environnement à base de notation qui détermine un état émotif à partir de la déformation visible du visage. *FACS* définit la notion d'*unité d'action*, qui se compose d'un ou plusieurs muscles contrôlant les distorsions des tissus faciaux. Les muscles du visage peuvent être manipulés individuellement mais ne donnent pas de résultats communicateurs et expressifs visibles. Ekman et Friesden ont ainsi défini 50 unités d'actions indépendantes dont les combinaisons peuvent donner des milliers d'expressions faciales différentes. Cette approche fut à la base de la définition de paramètres permettant de contrôler les déformations des visages modélisés en infographie

[10].

L'approche de Paul Ekman et Wallace Friesden, bien qu'introduisant des considérations de structure musculaire du visage, était orientée vers les déformations de la surface de la peau. En 1981, les investigations de Norman Badler [1] sur les bases structurelles du visage et du caractère élastique des muscles faciaux et de la peau de la partie supérieure du visage donnent un point de départ pour une analyse plus approfondie de la structure musculaire du visage. Cependant, la complexité de la rotation de la mâchoire inférieure rend l'approche peu efficace.

Frederic Parke [10] a été le premier en 1982 à proposer une modélisation du visage tenant compte de la structure particulière de la surface de la peau (courbures complexes). Il définit une paramétrisation de façon à permettre la spécification d'expressions via les paramètres faciaux.

En 1985, Philippe Bergeron et Pierre Lachapelle produisent une animation qui va rentrer pour toujours dans les archives de l'animation faciale par ordinateur: *Tony De Peltrie* [2]. Ils proposent une approche d'animation des expressions faciales d'un personnage synthétique à partir d'interpolation de plusieurs expressions de référence.

Par la suite en 1987, Keith Waters [20] a introduit une modélisation de la face et de sa structure musculaire. Keith Waters et Demetri Terzopoulos [15] proposent en 1991 une approche de modélisation et d'animation qui est une extension de la modélisation initiale des muscles par Waters. Leur approche se base sur les notions de la physique pour modéliser et simuler les mécanismes du visage [15, 7, 8].

Tous ces travaux (dans les universités, mais aussi dans les laboratoires de recherche et développement d'entreprises utilisant l'animation du visage dans l'une ou l'autre de leur activité) ont permis d'une part de définir des approches de modélisation informatique et de les classer selon leurs caractéristiques particulières. Les catégories principales établies à partir de cette classification en fonction de la méthode de contrôle des différentes postures faciales sont les suivantes:

- Bibliothèque d'expressions faciales et interpolation,
- Modélisation paramétrique de la face,
- Modélisation des muscles de la face,
- Modélisation des tissus faciaux à partir de la physique des corps déformables.

Chacune de ces catégories comporte des points forts mais également des inconvénients qui vont être introduits dans les sections qui suivent. Un aspect non négligeable en animation faciale est la façon de

contrôler la génération et l'animation des expressions. Jusqu'à présent les principales approches utilisées pour le contrôle de l'animation sont:

- le contrôle à base de règles,
- le contrôle à base d'analyse,
- le contrôle à base de performance.

Chacune des différentes méthodes de modélisation se prête mieux à l'une ou l'autre des approches de contrôle.

3.1 Méthodes de modélisation

La modélisation informatique des actions du visage est souvent décrite en termes de mouvements de muscles. Toutefois, les modèles (mathématiques) produits ne prétendent pas correspondre nécessairement à la structure et à la mécanique physiologique ou biomécanique de la face.

3.1.1 Bibliothèque d'expressions et interpolation

La modélisation des actes faciaux par bibliothèque d'expressions consiste à digitaliser la face (ou bien la modéliser à l'aide d'outils de *Conception Assistée par Ordinateur*) dans plusieurs postures différentes et à les conserver dans une banque d'expressions. Pour animer le visage digitalisé, le système d'animation fait une interpolation entre deux expressions contenues dans la banque (*inbetween*). La digitalisation de la face est faite par enregistrement des points sur une sculpture de la tête ou sur la tête d'un acteur réel. Cette façon de faire nécessite la redigitalisation (la banque et la face) pour toute nouvelle face à animer. En plus de cela, les expressions accessibles sont limitées à celles disponibles dans la bibliothèque. Le nombre et le stockage des expressions de la bibliothèque demande beaucoup d'espace pour obtenir de bons résultats.

Toutefois, le coût d'obtention d'une configuration intermédiaire de la face est faible dans la mesure où uniquement une interpolation est faite pour obtenir celle-ci.

3.1.2 Modélisation paramétrique

La modélisation paramétrique de la face est basée sur l'utilisation d'un ensemble de paramètres pour contrôler la structure, la forme et la taille du modèle de la face, mais aussi son expression. Le but de ce

type de modélisation est de réduire (minimiser) le nombre de paramètres de contrôle à donner à l'utilisateur. Une indépendance entre les paramètres de contrôle des expressions et les paramètres de contrôle de la structure du modèle de la face est établie. Les paramètres sont fortement liés à une géométrie particulière de la face adoptée dans la modélisation mais dépendent très peu de la dynamique de celle-ci. La détermination de fonctions de contrôle des paramètres est faite par des mesures de distances sur un visage réel (par exemple pour les lèvres). Cependant, il est impossible de paramétrer toutes les facettes de la structure faciale. Des polygones sont donc utilisés pour la modélisation de la surface du visage.

La génération d'une expression se fait par spécification de paramètres. Pour une séquence d'animation, il faut fournir un ensemble de *n-uplets* (langage d'animation) contenant les valeurs de chacun des paramètres de chaque configuration de la face pour chaque instant donné [3].

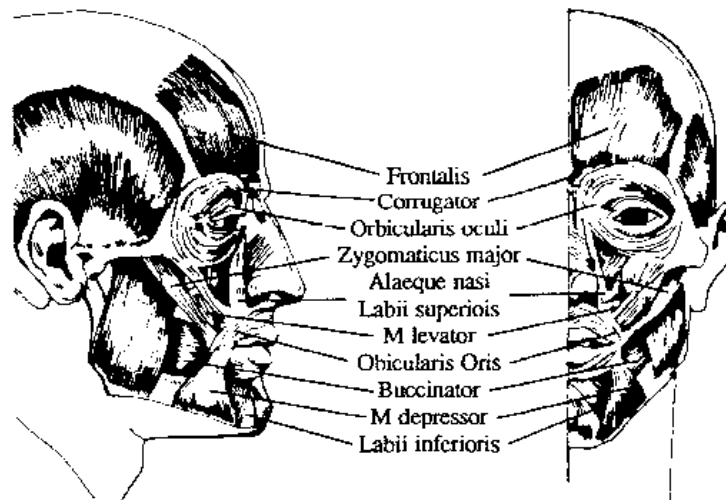
Le nombre de paramètres peut ainsi augmenter très facilement, et la manipulation de ceux-ci n'est pas toujours triviale pour un animateur. En outre, la définition de paramètres faciaux est une tâche très difficile en soit.

3.1.3 Modélisation des muscles de la face

Les modèles à base de muscles ou modèles à muscles abstraits imitent à un niveau élémentaire les actions de muscles faciaux. Les muscles entrant en ligne de compte pour les expressions faciales sont tous sous-cutanés. Ils sont attachés aux os du crâne et à la peau du visage, contrairement à ceux des autres régions du corps (bras, jambe, etc.) qui sont quant à eux attachés aux extrémités à des os. La figure 3.1, tirée de [20], présente un aperçu de la structure musculaire de la face. Pour former les expressions faciales, les muscles sont regroupés en unités d'actions.

Seulement trois types de muscles faciaux sont modélisés [20, 16, 18, 19]: les muscles linéaires- parallèles, les sphincters et les muscles en feuille. La modélisation des muscles est réduite et ne permet de simuler que des caractéristiques primaires de ceux-ci. La paramétrisation est abstraite et ne correspond pas nécessairement à des mécanismes biomécaniques ou neurophysiologiques. Les facteurs déterminants dans la modélisation des muscles sont :

- la proximité des noeuds d'attache du muscle à l'os et à la peau,
- la force de tension du muscle et de la peau,
- la profondeur du tissu musculaire à un noeud et la proximité du noeud,
- la limite d'élasticité du tissu musculaire relaxé et l'interaction intermusculaire.

FIG. 3.1 - *Muscles faciaux (tiré de [20])*

La modélisation d'un muscle est contrainte par les aspects suivants de la structure musculaire abstraite de la face:

- la contraction isotonique du muscle,
- les degrés de liberté des noeuds compte tenu des déplacements accessibles,
- la fixeté du noeud d'attache à l'os du crâne,
- le concept de vecteur muscle [20]¹ est introduit en guise d'abstraction de l'action d'un muscle,
- l'interaction entre différents muscles est modélisée par une interpolation aux limites en supposant un déplacement maximal de noeuds aux points d'attache à la peau et un déplacement nul à l'os.

La modélisation de la face à base de muscles est indépendante d'une géométrie particulière de la face (topologie de la face). Elle est compatible au codage de la face établi par Paul Ekman [9] (*FACS*) et a pour fondements les unités d'action et les travaux de Badler et Summerfield [13]. Le contrôle du modèle se fait via des paramètres des muscles abstraits.

La face est modélisée par un réseau de polygones. Le contrôle des points de la face via la déformation de ce réseau de polygones est fait par l'action du vecteur muscle. En effet, les muscles (vecteur muscle)

1. Un vecteur muscle est décrit comme étant un vecteur avec une zone d'influence circulaire, une direction et un sens allant du point d'attache à l'os vers le point d'attache à la peau; une amplitude variant de zéro au point d'attache à l'os à une valeur maximale au noeud d'attache à la peau.

sont modélisés comme étant des entités attachées en deux noeuds à la structure de la face; un noeud à l'os du crâne et l'autre à un noeud du faciès (points de contrôle de la face). Lorsque le muscle est excité (contraction), il influence les points de la peau situés dans son rayon d'action. Les résultats obtenus avec ce type de modélisation sont impressionnants, compte tenu du type de contrôle qui est adopté.

Toutefois, il subsiste des problèmes quant au contrôle de l'animation. Il est très difficile de contrôler la contraction des muscles (quels muscles actionner avec quels facteurs) pour générer une expression donnée. Quand bien même le modèle serait indépendant de la topologie de face, les déformations résultant de la contraction des muscles peuvent varier de façon notable d'une topologie de face à l'autre.

La simulation du mouvement des muscles sur les os du crâne est difficile. En plus, il n'est pas trivial de simuler certains mouvements subtiles et complexes de muscles faciaux. Par exemple, un muscle fait bouger un autre muscle qui se contracte.

Un dernier point faible de cette approche et non des moindres dans certains contextes de production, c'est qu'il est (presque) impossible de créer des expressions faciales physiquement irréalistes telles que des expressions expressément exagérées ou des expressions faciales de style *cartoon*.

3.1.4 Modélisation des tissus faciaux à partir de la physique des corps déformables

La méthode de modélisation des tissus faciaux à partir de la physique des corps déformables est une extension de celle à base de muscles. En effet, les propriétés des tissus faciaux sont modélisées par-dessus le modèle des muscles. Ici, la forme et les changements dynamiques de la face sont modélisés par des tissus faciaux abstraits. Ces tissus sont considérés comme des couches de réseaux de ressorts élastiques où les unités d'actions sont simulées par des forces (actions des muscles approximées par des fonctions de force). Le mouvement des muscles et leur propagation est intrinsèque au modèle. Plusieurs (quatre) couches de tissus ainsi que les muscles faciaux sont modélisées pour se rapprocher de la réalité et ce, malgré la simplicité du modèle virtuel comparativement à la structure anatomique d'une face réelle. Les cinq couches modélisés sont les suivantes:

- l'épiderme,
- le derme,
- les tissus connecteurs sous-cutanés,
- le faciès,

- les muscles des expressions faciales.

Les quatre premières couches de tissus constituent la peau du visage. Quant à la cinquième, comme l'indique sa dénomination, c'est la couche musculaire sous-jacente impliquée dans la formation des expressions faciales. Le modèle des couches se compose de prismes triangulaires (figure 3.2). Les différentes couches se composent de réseaux de ressorts qui peuvent s'interpénétrer. Les ressorts de couches différentes ont des relations de *contrainte-déformation* différentes. C'est-à-dire que les propriétés des ressorts varient d'une couche à l'autre (constante de raideur - fonction de force) [7]. Les valeurs typiques de constantes de raideur des différentes couches sont présentées au tableau 3.1.

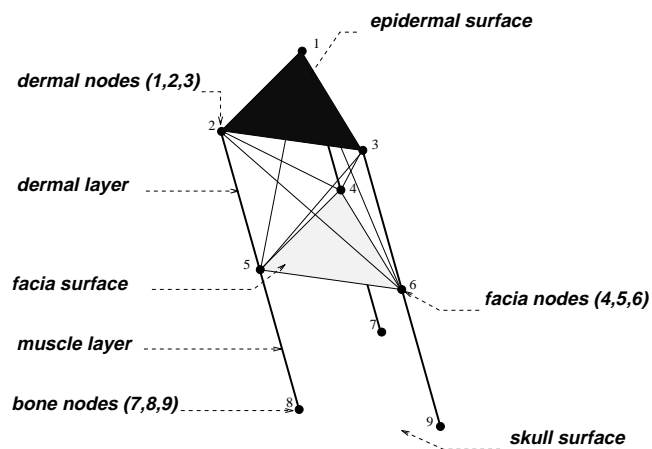


FIG. 3.2 - Couches de tissus faciaux (tiré de [7])

Sur une face réelle, les muscles sous la peau sont rattachés à celle-ci par des tendons élastiques. La modélisation du comportement de ces tendons est faite par une particularisation de l'action des muscles sur les nœuds du faciès se trouvant dans un certain rayon par rapport au vecteur muscle correspondant au muscle en question (le faciès est la couche la plus profonde de la peau). Les nœuds affectés par chaque muscle sont déterminés lors de la construction de la face. Deux aspects supplémentaires de la face sont supportés par cette approche de modélisation. La contrainte de conservation de volume des muscles lors de leurs actions est incorporée via une force additionnelle à un nœud de connexion musculaire. La contrainte de non-pénétration du crâne est quant à elle modélisée par des forces de contraintes calculées à partir de l'estimation des normales au crâne et de la force du réseau de ressorts du faciès au nœud voisin du point du crâne concerné [7].

Le contrôle de la déformation du modèle (génération d'expressions faciales) et de l'animation de celui-ci sont faits par simulation de l'équation discrète du mouvement de Lagrange [7, 15]. Les para-

couches	Épiderme	Derme	Tissus connecteurs	Faciès	Muscles
K	60	30	70	80	10

TAB. 3.1 - Valeurs typiques des constantes de raideur des couches [7]

mètres permettant de simuler l'équation sont obtenus à partir de la modélisation faite précédemment.

3.2 Contrôle de l'animation finale

Le contrôle de l'animation finale constitue une étape essentielle dans le processus d'animation. C'est ce qui permet de vérifier si la modélisation faite est facile à manipuler dans le but de générer une animation du modèle. Les différentes approches de contrôle sont regroupées sous trois méthodes génériques dont le contrôle d'animation à base de règles, à base d'analyse et à base de performance [11].

3.2.1 Contrôle d'animation à base de règles

Basée sur les études effectuées en linguistique et en psychologie, cette approche permet de générer automatiquement des animations faciales par analyse des règles contenues dans un ensemble de règles. Les règles établissent des liens entre les intonations de voix, les émotions et les expressions faciales. L'analyse des règles nécessite plusieurs niveaux de traitement et donc plusieurs effets de synchronisation.

3.2.2 Contrôle d'animation à base d'analyse

Dans le cas de la génération d'animation à base d'analyse, l'information nécessaire à l'animation est extraite à partir de l'analyse d'un vidéo. Les unités d'actions entrant en jeu ainsi que la contraction des muscles sont déduites de cette analyse. L'extraction des paramètres faciaux est très difficile du fait de la complexité de la corrélation entre la déformation de la face et la contraction des muscles faciaux sous-jacents.

3.2.3 Contrôle d'animation à base de performance

Ici, l'information nécessaire est prise sur un acteur réel par capture des mouvements de son visage en y plaçant des marqueurs [21]. Cette approche permet d'avoir la déformation à donner à la surface de peau de la face, sans analyse. Il est possible d'animer en temps réel le modèle de la face à partir de postures

faciales de l'acteur réel si le coût de communication des expressions réelles au modèle synthétique le permet.

Les systèmes d'animation de personnages existant utilisent l'une ou l'autre des approches de modélisation et de contrôle d'animation présentées ci-dessus, lorsqu'il s'agit d'animer la face [17]. Suivant le contexte d'utilisation (production de film, recherche scientifique, éducation, etc.) l'une ou l'autre des approches est préférée aux autres. Toutefois, la méthode de modélisation qui semble la plus proche de la réalité est celle basée sur la physique. Malgré la difficulté de contrôle qui lui est inhérente, cette méthode essaie de simuler les mécanismes de la face pour produire des expressions faciales.

3.3 Le système d'animation de *TAARNA*

Le projet présenté dans ce mémoire a nécessité le développement et l'implantation d'un outil permettant d'animer la majorité des aspects du visage d'un personnage synthétique. Il s'est déroulé en grande partie dans un contexte de production au sein d'une entreprise. L'outil développé fait partie d'un ensemble (système d'animation) visant à permettre l'animation d'objets complexes. Il est donc opportun de présenter sommairement le système d'animation au sein duquel l'outil d'animation faciale a été développé et est appelé à évoluer.

3.3.1 L'animation en général

Le système d'animation de *TAARNA* [14] consiste en une suite d'outils d'animation, de modélisation, d'édition et de rendu, tous distribués et conçus pour l'animation d'objets et personnages complexes [5]. Ces outils communiquent les uns avec les autres via un mécanisme de mémoire partagée (*shared memory*). Ils sont utilisés pour modéliser des objets, pour placer des lumières et leurs paramètres dans une scène, pour spécifier la surface d'objets, modifier les courbes de variation de paramètres et faire le rendu de la scène telle que vue par la caméra. Au centre de la philosophie du système se trouve le concept d'*animateur*. Un *animateur* est un outil (un programme) qui exécute des actions à un instant donné. Lors d'une animation, la caméra en marche (pour filmer la séquence) se charge d'initialiser les *animateurs* dont elle aura besoin au cours du *tournage*. Par la suite elle appellera chacun des *animateurs* initialisés, à chaque unité de temps (à chaque *frame*) pour configurer la scène avant d'en capturer l'image. Le système dispose de plusieurs *animateurs* pour l'animation faciale dont maintenant le nouvel animateur développé dans le cadre de ce projet.

3.3.2 Animation faciale

L'approche de modélisation de l'animation faciale adoptée par *TAARNA* est celle de la bibliothèque d'expressions faciales, avec des extensions et particularités visant à réduire certains des inconvénients et éviter des problèmes causés par les autres approches présentées antérieurement. Elle intègre également la possibilité de modéliser les muscles et les tissus faciaux. Le système dispose d'une bibliothèque standard d'expressions (des modèles de la face avec une expression) plus ou moins élémentaires, sur une face générique qui peut être enrichie par des expressions plus complexes. Pour animer le visage d'un personnage donné, il lui faut une banque d'expressions à lui. La création de cette banque (pour le nouveau personnage) nécessite la modélisation de la face neutre du personnage (*Modeler 3D*, *scanner* ou modèle de muscles pour déplacer des groupes de points en même temps avec cohérence). Un *animateur* du système, *map-expression*, permet de reproduire des expressions d'un modèle de face donné à un autre modèle de face, même si les deux modèles n'ont pas la même topologie. Ainsi, l'artiste n'a pas à refaire une banque d'expressions pour le nouveau personnage. Il lui suffit de choisir les expressions dont il a besoin à partir de la bibliothèque standard, ou de la banque d'un personnage existant et de faire jouer l'*animateur map-expression*.

La construction d'expressions plus complexes est faite par l'*animateur mixer* qui permet de générer une expression faciale qui est la combinaison de plusieurs autres expressions (élémentaires ou non). Les expressions ainsi obtenues peuvent être placées dans une banque.

L'animation des expressions faciales d'un personnage est quant à elle faite par l'*animateur expressor*. Cet *animateur* nécessite que la séquence des expressions faciales clé de l'animation soit créée (à partir de *mixer*) et spécifiée dans un certain ordre, en fonction de la séquence d'animation désirée. Cela fait qu'il est impossible de l'utiliser en temps réel.

Le développement du nouvel *animateur facial live-face* (pour le présent projet) s'est basé sur certaines des idées existantes dans le système d'animation (faciale) de *TAARNA*, tout en essayant de mitiger, voire éliminer certains de ses points faibles. Vu qu'il est appelé à s'intégrer et à évoluer dans le système d'animation de *TAARNA*, l'*animateur facial live-face* se devait également de supporter les concepts en sous-tendant le développement des outils du système. En outre, il doit supporter l'approche d'animation faciale existante dans le système de *TAARNA* et donner à l'artiste (l'utilisateur) la possibilité d'exprimer son talent en lui permettant de contrôler facilement les résultats. Ainsi, l'approche par bibliothèque a été adoptée compte tenu de la diversité de personnages et d'expressions (animaux, cartoon, expressions exagérées) à laquelle les artistes ont affaire, de la qualité, du type d'animation visé et du fait que les artistes

sont habitués à contrôler les résultats d'une animation faciale à partir d'une banque d'expressions.

Chapitre 4

Approche par une banque d'expressions

L'animation d'un visage synthétique consiste à faire prendre différentes expressions à ce visage. L'approche développée dans le présent travail consiste à produire une expression faciale donnée à partir d'un ensemble pré-établi d'expressions constituant une *banque d'expressions*. C'est donc que le résultat (l'expression faciale désirée) correspond à une combinaison des expressions se trouvant dans la banque. La combinaison des expressions de la banque est voulue linéaire; dans ce contexte, il convient que les expressions contenues dans la banque soient le plus indépendantes possible les unes des autres. En fait, on voit à ce que la banque d'expressions constitue un système de référence dans lequel on peut exprimer toute autre expression du visage synthétique donné. Il serait mathématiquement possible de générer à partir de la banque d'expressions initiale, une banque d'expressions orthogonales au sens mathématique (avec la méthode de Gram-Schmidt). Cependant, les approches de résolution numérique utilisées pour la détermination de la combinaison linéaire des expressions (chapitre suivant), tiennent compte de la signification d'une expression en tant qu'ensemble de points du modèle du visage dans une position donnée. Une orthogonalisation de la banque d'expressions ferait qu'une expression n'aurait plus la même signification. Néanmoins, la méthode de Gram-Schmidt peut être utilisée d'une certaine manière pour juger de l'indépendance de la banque d'expressions.

Il est à noter que l'approche proposée se veut automatique mais portant l'empreinte de l'artiste. C'est-à-dire qu'elle devrait permettre de faire l'animation d'un visage synthétique à partir de mouvements faciaux spécifiés par déplacement des points sur ce visage, et cela en conservant l'aspect artistique de la chose (l'artiste retrouve sa signature dans le résultat du système automatique). Les animations de visage nécessitent ainsi peu d'effort et de temps de la part de l'artiste pour une bonne qualité visuelle. Une fois que la banque est créée, très peu de talent est requis pour obtenir des animations de bonne qualité visuelle.

Il faut souligner que le talent de l'artiste fait passer la qualité des animations de bonne à très bonne. Le contrôle des déplacements des points du visage synthétique peut être fait à partir de capture des mouvements du visage d'un acteur réel.

L'intervention de l'artiste se situe au niveau de la modélisation du visage synthétique et de la création de la banque d'expressions initiale. Pour chaque personnage synthétique, on dispose d'une banque d'expressions lui étant propre, compte tenu de la morphologie de son visage. Le système automatique se charge de trouver les coordonnées de l'expression désirée dans la base constituée par la banque d'expressions. Le résultat de la combinaison linéaire est ainsi fonction de la modélisation et des choix faits par l'artiste au niveau de la banque d'expressions, d'où sa signature.

Ce chapitre présente les éléments essentiels à l'animation faciale à partir d'une banque d'expressions. Il discute de l'information pertinente contenue dans une expression et donc dans une banque d'expressions. Il traite également de la capture des expressions faciales d'un acteur réel pour en déduire des expressions similaires pour le visage synthétique.

4.1 La banque d'expressions

La banque d'expressions du personnage synthétique constitue, comme mentionné plus haut, le référentiel dans lequel l'expression désirée est exprimée en terme de coordonnées. L'expression synthétique en elle-même se présente sous la forme d'un modèle 3D du visage synthétique dans une posture donnée. Le modèle du visage synthétique est un ensemble de points à partir desquels on construit des polygones (modèle de type polygonal) ou bien des *patches* (morceaux de surfaces paramétriques - modèle de type *patch*), grâce à des méthodes mathématiques (visualisation d'un nuage de points). Le modèle (visage synthétique) contient une certaine quantité d'informations nécessaires à son identification et à son utilisation. L'utilisation faite du modèle pour les besoins de la synthèse d'expressions faciales ne tient compte que des informations suivantes:

- coordonnées des points du modèle
- liste des primitives du modèle
- lien d'appartenance d'un point à un polygone
- liste des polygones auxquels appartient un point du modèle
- normales aux polygones
- normales aux points du modèle¹
- type de modèle (polygones ou *patches*).

D'autres informations telles que le niveau de détail du modèle et les coordonnées de texture sont également incorporées. Ainsi, une banque d'expressions est en fait un ensemble de modèles du visage synthétique (figure 4.1).

L'approche développée dans ce mémoire utilise les notions de l'algèbre linéaire et d'analyse numérique pour les calculs nécessaires à la composition des expressions. Pour ce faire, la notion de *vecteur expression* est introduite.

Une expression est interprétée comme un *point* en $3m$ dimensions; où m est le nombre de sommets (points 3D) du modèle de visage représentant l'expression en question. La différence entre deux expressions est un *vecteur expression*. Un vecteur expression peut donc s'exprimer comme étant la combinaison

1. La normale dont il est question ici est la normale à la surface du modèle en un point donné.

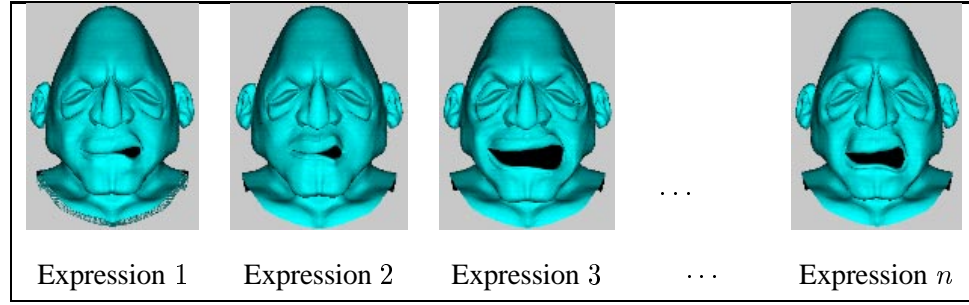


FIG. 4.1 - Exemple de banque d'expressions (modèles)

linéaire d'autres vecteurs expressions. Pour définir un référentiel dans l'espace des expressions, on choisit une expression comme étant l'origine du référentiel; c'est l'*expression neutre* du visage synthétique. Les vecteurs expressions correspondant aux vecteurs de base du référentiel sont obtenus en faisant la différence entre chaque expression de la banque (point dans un espace de dimension $3m$) et l'*expression neutre*.

Soit \mathcal{E} l'ensemble de toutes les expressions possibles d'un personnage synthétique et soit $[\xi_0] \in \mathcal{E}$ l'expression neutre correspondante. On peut définir \mathbf{E} l'ensemble des expressions vectorielles dont les éléments sont des vecteurs expressions obtenus en faisant la différence entre chaque expression de \mathcal{E} et l'expression neutre $[\xi_0]$, origine du référentiel. Mathématiquement, cela s'exprime comme suit:

$$[\xi_i] = \begin{pmatrix} x_{i,1} \\ y_{i,1} \\ z_{i,1} \\ \vdots \\ x_{i,m} \\ y_{i,m} \\ z_{i,m} \end{pmatrix} \quad [\vec{\xi}_i] = \begin{pmatrix} x_{i,1} - x_{0,1} \\ y_{i,1} - y_{0,1} \\ z_{i,1} - z_{0,1} \\ \vdots \\ x_{i,m} - x_{0,m} \\ y_{i,m} - y_{0,m} \\ z_{i,m} - z_{0,m} \end{pmatrix} \quad i = 1, \dots, n$$

$$[\vec{\xi}_0] = [\vec{0}] \quad [\vec{\xi}_i] = [\xi_i] - [\xi_0]$$

L'ensemble \mathbf{E} muni de l'addition interne (+) des vecteurs expressions et de la multiplication externe (\cdot) de vecteur expression par un scalaire appartenant à \mathbf{R} (ensemble des nombres réels), est un espace vectoriel (de dimension $3m$) sur \mathbf{R} et se note $(\mathbf{E}, +, \cdot)$.

Soit $([\vec{\xi}_1], [\vec{\xi}_2], \dots, [\vec{\xi}_n])$ l'ensemble des vecteurs expressions correspondant à une banque d'expressions indépendantes de \mathbf{E} formée de n expressions faciales. Alors tous les vecteurs expressions $[\vec{e}] \in \mathbf{E}$,

tels qu'il existe $(\lambda_1, \dots, \lambda_n) \in \mathbf{R}$ de sorte que $[\vec{e}] = \lambda_1[\vec{\xi}_1] + \lambda_2[\vec{\xi}_2] + \dots + \lambda_n[\vec{\xi}_n]$ constituent un sous espace vectoriel E_n de E .

On peut ainsi exprimer toute expression de E_n dans la base $([\vec{\xi}_1], [\vec{\xi}_2], \dots, [\vec{\xi}_n])$ et par conséquent, toute expression de E_n dans $\mathcal{R}([\xi_0]; [\vec{\xi}_1], [\vec{\xi}_2], \dots, [\vec{\xi}_n])$. La figure 4.2 illustre un référentiel constituée de seulement deux vecteurs expressions $[\vec{\xi}_i]$ et $[\vec{\xi}_j]$.

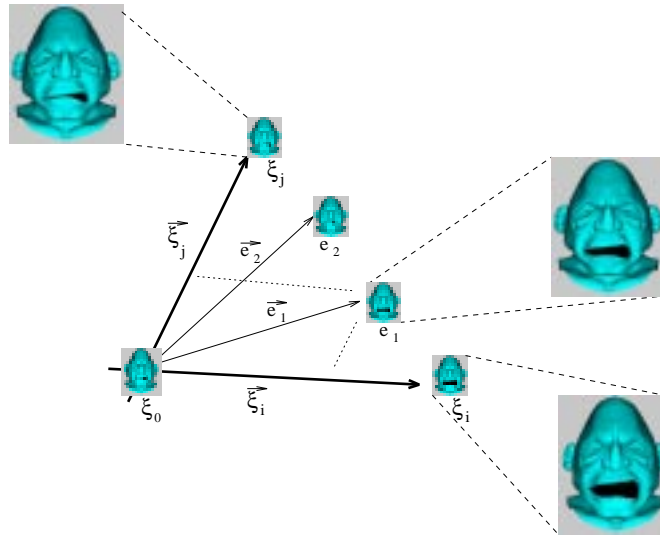


FIG. 4.2 - Expressions vectorielles - Expressions et combinaisons

Le système d'animation faciale de **TAARNA** qui utilise l'approche de la bibliothèque d'expressions de référence, et au sein duquel le projet fut développé, offre la possibilité de faire varier manuellement des potentiomètres contrôlant les valeurs des composantes d'un vecteur $[\lambda]$ de coordonnées d'expressions et par conséquent l'expression faciale finale. Les valeurs clé de $[\lambda]$ sont enregistrées et l'animation continue est générée à partir de celles-ci par interpolation (linéaire ou cubique). Cette façon de faire ne laisse aucune possibilité à une utilisation dans un contexte où les mouvements faciaux capturés sur un acteur réel doivent être utilisés pour animer les expressions faciales d'un personnage synthétique en temps réel. En outre, il est difficile de contrôler localement l'expression faciale résultant des valeurs de $[\lambda]$ spécifiées. Une approche de contrôle de l'expression finale autre que la variation manuelle des valeurs de $[\lambda]$ s'avère donc nécessaire. L'alternative adoptée est de contrôler l'expression finale à partir de manipulation de points de contrôle sur la face, via la banque d'expressions de référence. Il faudra donc déterminer les coordonnées d'expression dans E_n (vecteur $[\lambda]$) automatiquement à partir de la manipulation des points de contrôle de la face. Ainsi il serait possible d'utiliser le système dans un contexte où la manipulation des points de contrôle est faite à partir de capture des mouvements faciaux d'un acteur réel.

Le processus se résume donc ainsi: à partir de la spécification d'une expression faciale désirée par la manipulation des points de contrôle de la face, il faut composer une des expressions du visage du personnage synthétique correspondant à la spécification. Cela revient à trouver les valeurs de λ_i pour une expression $[e]$ donnée, de sorte que l'équation 4.1 soit vérifiée.

$$[e] = \sum_{i=1}^n \lambda_i [\xi_i] \quad (4.1)$$

En donnant plus ou moins d'importance à certains points de contrôle de la face, le niveau de satisfaction de leurs positions dans l'expression de synthèse peut être influencé et pris en compte dans le processus de détermination automatique des coordonnées d'expression. Ainsi la détermination automatique cherchera à satisfaire la position d'un point de contrôle de la face de plus grande importance plutôt que celle d'un autre.

Pour des fins d'efficacité et de production, la spécification de l'expression désirée est faite par la capture du mouvement facial sur un acteur réel.

Dans la suite de ce mémoire et pour des fins de simplification, nous considérons qu'une expression est un vecteur expression et toutes les résolutions de problème sont faites dans l'espace vectoriel des vecteurs expressions $(E, +, \cdot)$. On ne revient à \mathcal{E} que lorsqu'il faut visualiser le résultat en faisant la synthèse des expressions finales (équation 4.2).

$$[e] = [\xi_0] + \sum_{i=1}^n \lambda_i [\xi_i] \quad (4.2)$$

4.2 Capture des postures faciales (*Tracking*)

La séance de capture des postures faciales constitue le point de départ du contrôle des mouvements des points de la face du personnage synthétique. En effet, l'objet de l'animation de personnage à partir de capture de mouvements est de communiquer au personnage synthétique les mouvements d'un sujet réel, mouvements qui ont été capturés pendant la séance de *tracking*. Il est donc opportun à ce moment de décrire l'équipement permettant de faire la capture des expressions faciales de l'acteur réel. Pour des raisons de coûts d'acquisition, de maniabilité et de confort pour l'acteur, plusieurs systèmes ont été testés de manière à choisir le plus adéquat. Nous avons considéré l'enregistrement vidéo de la performance de l'acteur et le port de marqueurs (réfléchissants ou magnétiques) sur le visage de celui-ci. Le choix s'est finalement porté sur le système de capture *FaceTrax*TM de *Adaptive Optics* avec marqueurs faciaux ré-

fléchissant les fréquences lumineuses infrarouge. Tous les systèmes de ce type ont beaucoup de points en commun dans la mesure où ils utilisent le même phénomène physique pour détecter les mouvements du visage, à savoir la réflexion de la lumière. Un aperçu des caractéristiques fondamentales de ces systèmes est présenté en mettant en évidence la philosophie soutendant leur conception.

4.2.1 Système de capture à base de réflexion lumineuse

Les systèmes de capture utilisés sont conçus pour permettre ultimement l'animation en temps réel d'un personnage synthétique à partir d'un sujet. Ils incorporent un dispositif de transfert de données du monde réel (*marqueurs*) à un système graphique qui anime le visage synthétique. Ce dispositif est matérialisé par un lien sériel (*i.e.* RS232) et un protocole de communication permettant de corriger les erreurs mineures pouvant survenir lors de la transmission afin d'assurer l'intégrité des données transmises. Les différents modules des systèmes sont les suivants (vidéo: *Système de capture*):

Un casque muni d'une caméra infrarouge

Le casque est utilisé pour l'acquisition de l'information de position des *marqueurs* en 2D. Les *marqueurs* sont des éléments réfléchissants, placés sur le visage de l'acteur de manière à permettre à la caméra d'acquérir les mouvements de la face. Un microphone est ajouté au casque original de *Adaptive Optics* pour enregistrer simultanément le son (la voix de l'acteur). La figure 4.3 montre une image du casque utilisé pour la capture des mouvements. Les composantes z des déplacements de marqueurs sont estimés lors de la communication des mouvements de ceux-ci au visage du personnage synthétique. Le processus d'estimation est présentée plus loin.

Une unité de traitement

Elle se charge d'initialiser l'ensemble du système, de la calibration des différents modules et du transfert des informations de l'unité casque vers le système graphique 3D.

Une unité d'alimentation

Cette unité alimente la caméra et l'unité de traitement.

Un écran de visualisation des mouvements des marqueurs

Il permet de visualiser les effets des mouvements du visage du sujet via le déplacement de points lumineux sur l'écran.

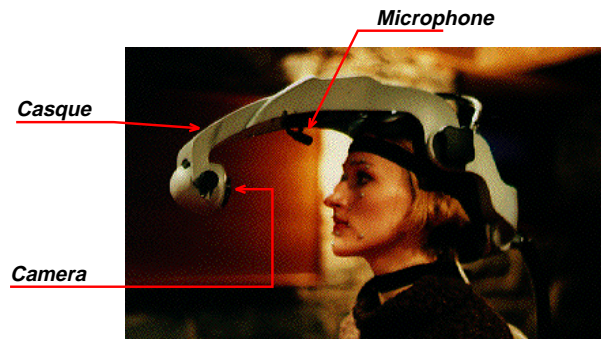


FIG. 4.3 - Actrice réelle avec le casque

Pour assurer un rendu réaliste des mouvements capturés, il faut une correspondance entre le nombre de marqueurs réels et les points de contrôle de l’affichage graphique; en outre ce nombre de points ne doit pas changer pendant une même séance de capture de postures faciales. Un autre aspect important est que la correspondance entre les deux mondes (réel et virtuel) doit tenir compte des différences de morphologie et d’échelles des visages en présence. Pour ce faire, les déplacements du visage du sujet réel subissent des ajustements (angle d’inclinaison du visage, forme, taille, etc.) avant d’être appliqués aux points de la face du personnage synthétique. La transformation des déplacements de marqueurs consiste à appliquer une rotation pour placer les visages dans la même position compte tenu des morphologies, suivie d’une mise à l’échelle faite pour ramener les déplacements des marqueurs réels à l’échelle du visage synthétique.

Pour finaliser la caractérisation des deux visages, les marqueurs et leurs contreparties virtuelles sont répartis de façon à permettre la capture et la communication au visage synthétique des postures faciales, de la façon la plus précise possible.

4.2.2 Répartition des marqueurs

La répartition des marqueurs réels sur la face (figure 4.4) est importante pour la qualité du résultat de l’animation. Une mauvaise répartition des marqueurs réels peut rendre la capture erronée (information insuffisante). Quant aux points de la face du personnage synthétique associés aux marqueurs (marqueurs virtuels), leur répartition influence les expressions synthétiques car ils constituent le lien permettant de contrôler la synthèse des expressions finales à partir des déplacements des marqueurs réels.

De plus, une mauvaise correspondance entre les marqueurs réels et virtuels peut causer la communication erronée des postures faciales du sujet au visage synthétique.

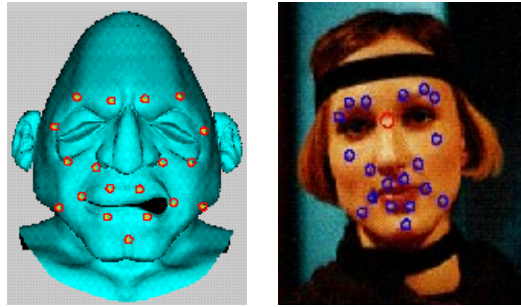
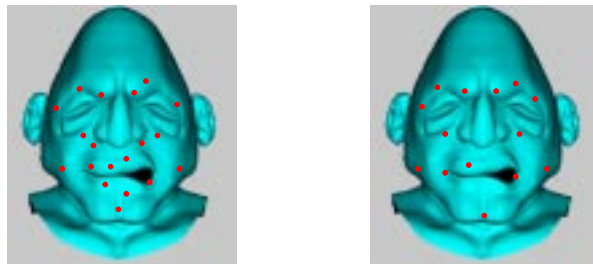


FIG. 4.4 - Répartition des marqueurs virtuels et réels

Pour finir, une mauvaise répartition des marqueurs virtuels sur le visage synthétique se traduit par le déplacement des mauvais points du visage synthétique quand bien même les postures faciales capturées et transmises seraient correctes. La figure 4.5 illustre certains cas de répartition de marqueurs virtuels pouvant altérer l'expression de synthèse finale. Par exemple, dans la première image de la figure 4.5, les marqueurs virtuels des coins de la bouche sont placés sur la lèvre alors que leurs contreparties réelles sont bien aux coins de la bouche. Cela résulte en la génération d'expressions de synthèse qui pourraient ne pas ressembler à l'expression de l'acteur réel au niveau de la forme de la bouche.



Mauvaises correspondances

Pas assez de marqueurs
dans les zones critiques

FIG. 4.5 - Difficulté dans la répartition des marqueurs

4.2.3 Problèmes relatifs à la capture des postures faciales

Compte tenu du principe sur lequel se basent les systèmes de capture, certaines erreurs survenant lors de la capture peuvent être causées par:

- le mélange des marqueurs: au voisinage de la bouche (figure 4.6), deux marqueurs se rapprochant de trop près à l'instant $t + 1$ pourraient être mal identifiés par le système de capture lorsqu'ils se séparent à l'instant d'après (à $t + 2$).
- la perte de marqueurs: lorsqu'un marqueur disparaît du champ de vision de la caméra (il tombe ou sa réflexion est obstruée), le système de capture perd sa trace (figure 4.7).
- les très petits déplacements: un mouvement est produit mais à cause de la sensibilité de la caméra, la position des marqueurs réels est supposée statique.
- les réflexions parasites: lorsqu'une émission ou une réflexion de lumière indésirée tombe dans le champ de vision de la caméra (figure 4.8), cela produit une modification du nombre de marqueurs réels reconnus par le système de capture lors de l'initialisation.
- la taille des marqueurs: lorsque les marqueurs sont trop petits, ils peuvent être interprétés par le système de capture comme du bruit. Il existe une taille minimale pour les marqueurs spécifiée dans le manuel d'utilisation du casque.

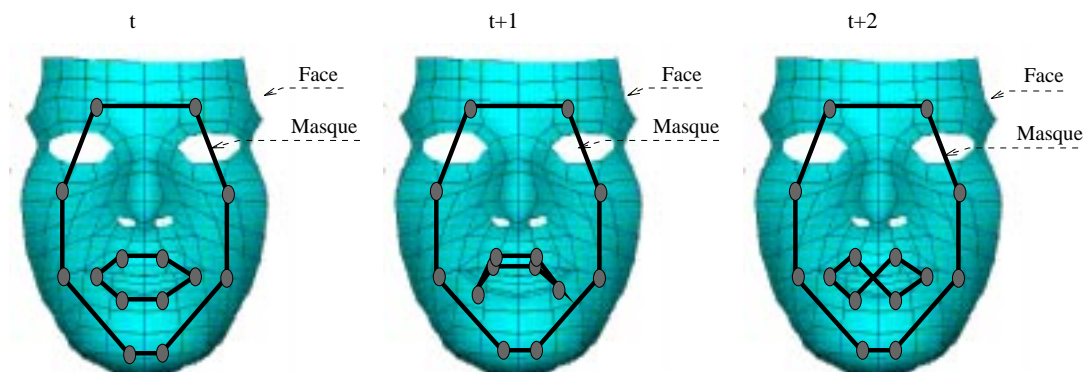


FIG. 4.6 - Mélange de marqueurs

À partir de ces causes d'erreurs probables, certaines zones du visage sont identifiées comme étant plus susceptibles de provoquer des défaillances dans la capture des mouvements faciaux. Cela conditionne donc la répartition des marqueurs réels sur le visage du sujet.

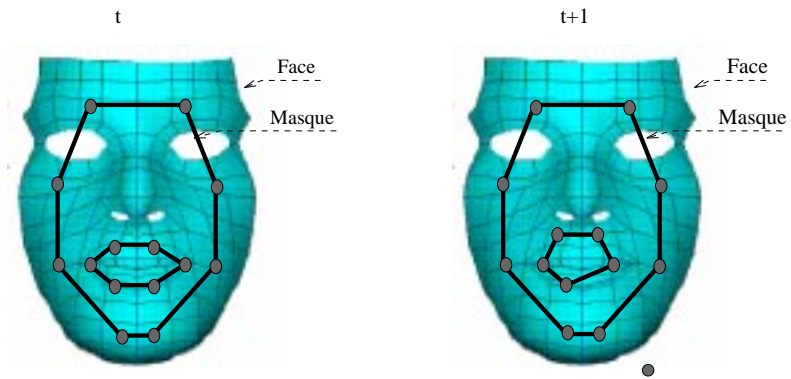


FIG. 4.7 - Perte de marqueurs

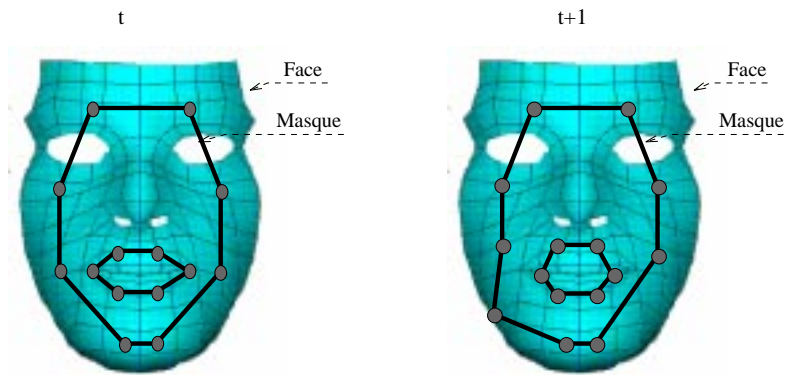


FIG. 4.8 - Réflexion parasite

De par les essais effectués lors des séances de capture, les zones sensibles du visage ont été répertoriées de manière à capturer les mouvements faciaux le plus correctement possible. En raison du nombre limité de marqueurs disponibles (10 à 40 marqueurs), la répartition se fera dans ces zones. Les zones importantes dans la capture des postures faciales sont les suivantes:

- le voisinage de la bouche
- les paupières
- les arcades sourcilières
- les joues
- le menton
- les pommettes

Les trois premières zones sont très importantes dans la répartition des marqueurs tandis que les autres sont d'un niveau d'importance secondaire.

Il est nécessaire pour le dispositif de capture d'avoir un point sur le visage qui ne bouge (presque) pas d'une posture faciale à l'autre. Ce point est utilisé par le système comme référence pour la définition des transformations et les calculs de déplacements. Sur le visage du sujet, un des points qui ne bouge presque pas lors des différentes postures faciales se situe entre les deux yeux, juste au dessus du nez; il est représenté par un marqueur plus gros et donc plus facilement repérable (figure 4.9).

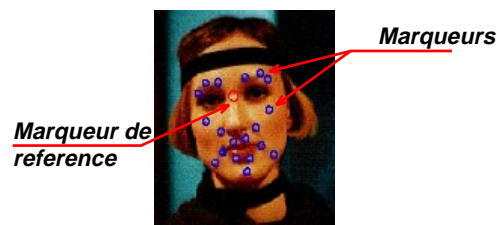


FIG. 4.9 - Marqueur de référence et répartition des marqueurs réels

4.3 Association des marqueurs aux points de contrôle de la face

La communication au personnage synthétique, des postures faciales capturées sur l'acteur réel, n'est possible qu'en établissant un lien entre les deux entités. Ce lien est établi par association des points 3D

du visage synthétique aux marqueurs placés sur le visage de l'acteur réel, et cela en tenant compte de la différence de morphologie possible entre les deux visages. Une explication plus détaillée des liens entre les visages passe par la définition de la terminologie utilisée.

Masque:

À partir de l'enregistrement des positions de marqueurs, la position neutre du visage de l'acteur réel est utilisée pour construire un modèle en connectant les points (associés aux positions des marqueurs) à l'aide d'un outil logiciel (le *modeler*) du système d'animation de TAARNA. Le modèle ainsi obtenu constitue le *masque* qui est le véritable lien entre les deux acteurs en présence. Il comporte k points (nombre de marqueurs sur le visage de l'acteur. $k < m$).

Face:

Le modèle correspondant au visage du personnage synthétique (visage à animer) constitue la *face*. Les points de ce modèle sont à déplacer à partir des postures faciales de l'acteur réel via le *masque*.

La figure 4.10 présente un visage synthétique (*face*) avec son *masque* correspondant. Il est à noter que la *face* comporte un très grand nombre de points (environ 2000 points) tandis que le *masque* n'en a que très peu (entre 15 et 40 points). La composition d'une expression synthétique revient donc à déplacer de façon cohérente un nuage de points à partir de seulement quelques points de contrôle.

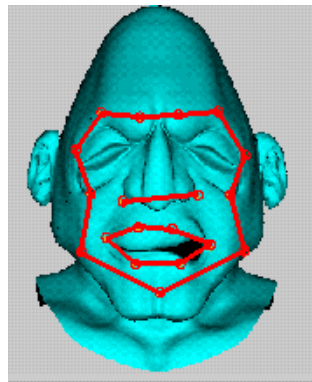


FIG. 4.10 - La face portant le masque

Pour ce faire, le *masque* est porté par la *face* de manière à ce que les deux entités prennent les mêmes postures faciales. Le processus de port du *masque* par la *face* est basé sur l'association des points du visage synthétique aux *marqueurs* placés sur le visage de l'acteur réel.

4.3.1 Modes d'association

Le port du masque par la face peut se faire de diverses manières. Les approches choisies dans le cas présent sont l'association *point à point* et l'association *point à surface* (polygone).

Association point à point

L'association point à point entre le masque et la face consiste à établir une correspondance entre les points. Pour ce faire, les deux *modèles* sont mis en présence et les marqueurs virtuels (les points du masque) sont distribués sur le visage synthétique (les points de la face) de façon à respecter le plus possible la distribution de marqueurs réels du sujet. Chaque marqueur virtuel est associé à *au plus un*² point de la face. Pour établir le lien logique entre les marqueurs virtuels et la face, un algorithme de recherche assez simple est utilisé. Un marqueur virtuel donné est associé au point de la face qui lui est le plus proche et qui n'est pas encore associé à un autre marqueur virtuel (figure 4.11). Cette façon de faire est sujette à l'ordre dans lequel l'ensemble des marqueurs virtuels est parcouru. La répartition des marqueurs virtuels sur le visage synthétique est importante pour éviter les situations où un marqueur virtuel serait associé à un point qui lui est le plus proche mais qui est beaucoup plus proche d'un autre marqueur virtuel. Compte tenu du très grand nombre de points de la face comparativement au nombre de points du masque et du fait que les points du masque sont assez espacés pour permettre un meilleur contrôle, il y a très peu de conflits qui sont alors réglés manuellement.

Pour communiquer le déplacement des marqueurs réels aux points (3D) de la face via les marqueurs virtuels, le processus est direct. Seuls les points de la face qui ont un correspondant dans l'ensemble des marqueurs virtuels sont susceptibles de bouger. Le point de la face se déplace de la même façon que son *marqueur* virtuel. Donc si le marqueur passe de la position T à T' , son homologue sur la face à la position P passera à P' tel que l'équation 4.3 soit satisfaite.

$$P' = P + T\vec{T}' \quad (4.3)$$

2. Si un marqueur virtuel n'est associé à aucun point de la face, alors ses mouvements ne seront pas pris en compte lors de la détermination des coordonnées d'expression dans E_n .

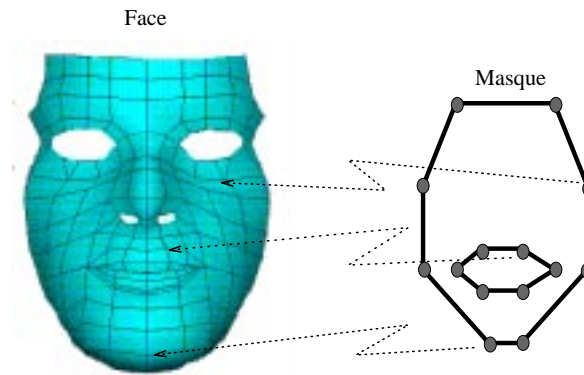


FIG. 4.11 - Association point à points

Zone d'influence d'un point

La posture faciale prise par la face, après déplacement des points associés, se rapproche plus d'une expression faciale à mesure que le nombre de points associés à des marqueurs virtuels augmente. La présente approche d'association des points permet d'augmenter le nombre de points de la face contrôlés par les marqueurs virtuels en introduisant la notion de poids et de zone d'influence d'un marqueur virtuel. Chaque marqueur virtuel dispose d'une zone d'influence. Son activité influence tous les points de la face qui se trouvent dans cette zone suivant un certain poids. La correspondance entre points de la face et marqueurs virtuels se présente comme suit.

Un point donné de la face est influencé par tous les marqueurs virtuels dont les zones d'influence l'englobent. Un marqueur virtuel influence tous les points de la face qui se trouvent dans sa zone d'influence. Ainsi, un point est en correspondance avec plusieurs marqueurs virtuels et un marqueur virtuel agit sur plusieurs points de la face (figure 4.12). Ainsi le lien entre points de la face et marqueurs virtuels n'est plus sujet à l'ordre de parcours de l'ensemble des marqueurs virtuels.

Le déplacement des points de la face se trouvant dans la zone d'influence du masque dépend de l'union des zones d'influence des marqueurs virtuels. Un point de la face sera influencé suivant sa distance par rapport aux marqueurs virtuels qui agissent sur lui. La fonction déterminant l'influence d'un marqueur sur les points de la face situés dans sa zone d'influence se doit de réduire cette influence suivant l'éloignement du point de la face. En outre, elle doit refléter des caractéristiques du marqueur dont elle contrôle l'influence (poids, zone d'influence) et être facile à évaluer. Le choix de la fonction qui permet de déterminer l'influence d'un marqueur virtuel sur un point de la face s'est arrêté sur une gaussienne (figure 4.13) dont la variable est la distance entre point et marqueur. L'écart-type σ est le rayon de la zone

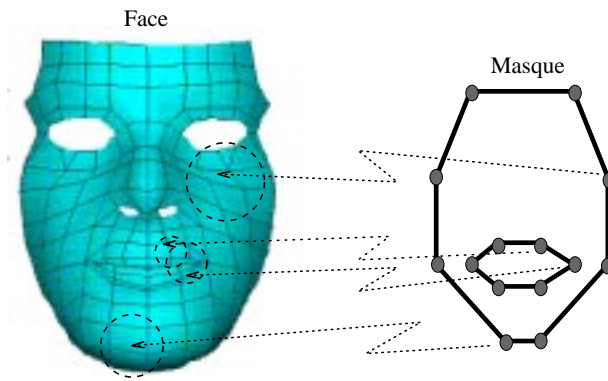


FIG. 4.12 - Association par zone d'influence

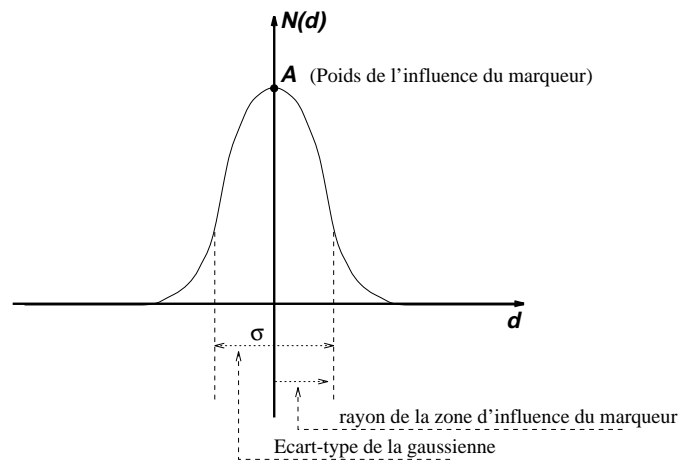


FIG. 4.13 - Fonction de poids des marqueurs

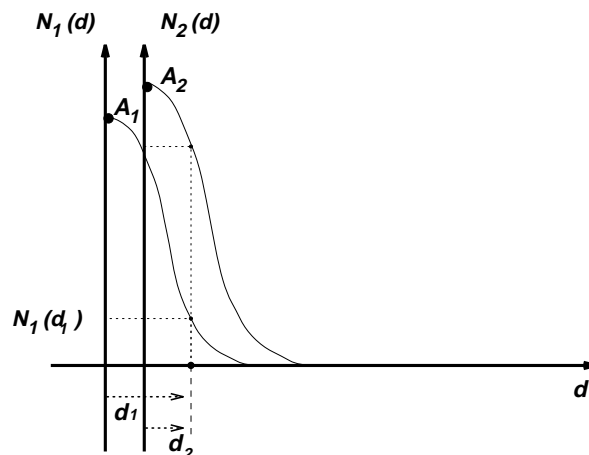


FIG. 4.14 - Influences de plusieurs marqueurs (deux) sur un point

d'influence du marqueur. Chaque marqueur virtuel dispose donc d'une fonction de distribution normale qui régit son influence sur les points se trouvant dans sa zone d'action.

Soit un point P (3D) de la face influencé par k_P ($k_P \leq k$) marqueurs virtuels aux positions (3D) $\{T_1, T_2, \dots, T_{k_P}\}$.

Soient N_i la fonction de distribution d'influence du marqueur virtuel en T_i , d_i la distance de P à T_i et w_i l'influence du marqueur en T_i . Pour éviter des effets de débordements, la contrainte suivante est imposée aux influences des marqueurs virtuels sur le point P , lors du calcul du déplacement:

$$\sum_{i=1}^{k_P} w_i = 1. \quad (4.4)$$

Alors l'influence du marqueur en T_i sur P est donnée par la formule :

$$w_i = \frac{N_i(d_i)}{\sum_{j=1}^{k_P} N_j(d_j)}. \quad (4.5)$$

La fonction de distribution des influences du marqueur en T_i est:

$$N_i(d) = A_i e^{-\left(\frac{d^2}{\sigma_i^2}\right)} \quad (4.6)$$

A_i est la valeur maximale de la fonction de distribution d'influence N_i . C'est le poids du marqueur en T_i (typiquement, $A_i = 1$).

Le calcul effectif du déplacement de P en P' est obtenu par l'équation 4.7, où T_i et T'_i sont respectivement les ancienne et nouvelle positions du marqueur i .

$$P' = P + \sum_{i=1}^{k_P} w_i (T_i \vec{T}'_i) \quad (4.7)$$

Association point à polygone

Une autre approche pour faire porter le masque à la face, et aussi augmenter le nombre de points de la face qui bougent en fonction du masque, consiste à contraindre des points de la face aux polygones du masque. Pour ce faire, le *modèle* correspondant au masque est d'abord triangularisé, c'est-à-dire que tous les polygones le constituant sont subdivisés pour ne conserver que des triangles en guise de polygones. Cette étape est nécessaire pour ramener la méthode de calcul des coordonnées barycentriques au cas simple du triangle. Le port du masque par la face est fait en associant les points de la face au triangle qui leur est le plus proche (figure 4.15). Plusieurs points de la face peuvent être associés à un seul triangle

mais un point donné n'est alors contraint que par *au plus* un triangle. Les points de la face qui ne sont associés à aucun polygone (points libres) sont laissés libres de leurs mouvements.

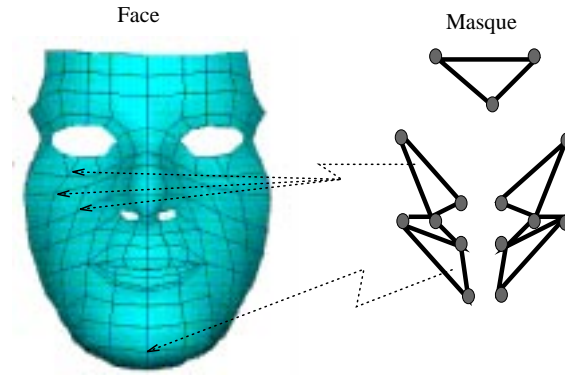


FIG. 4.15 - Association polygone à points

Les postures faciales sont transmises à la face via la déformation des polygones (triangles) du masque. En effet, l'association des points aux polygones se fait en mettant les entités face et masque en présence, dans leurs postures de repos (neutres). Pour chaque point de la face qui est associé à un triangle du masque, ses coordonnées barycentriques relatives à ce triangle sont conservées. Lorsque le masque prend une nouvelle posture, les triangles sont déformés, ce qui a pour effet de modifier la position des points de référence des coordonnées barycentriques. La nouvelle position des points associés est obtenue en replaçant ces derniers en fonction des nouvelles positions des points du triangle associé et en conservant les mêmes coordonnées barycentriques. Ainsi soit P un point de la face associé au triangle (T_1, T_2, T_3) du masque et soient (r, u, v) les coordonnées barycentriques de P relativement à (T_1, T_2, T_3) avec pour origine T_1 , le tout illustré à la figure 4.16 et par les équations 4.11.

$$r = 1 - u - v \quad (4.8)$$

$$P = rT_1 + uT_2 + vT_3 \quad (4.9)$$

$$P = (1 - u - v)T_1 + uT_2 + vT_3 \quad (4.10)$$

$$P = T_1 + u\vec{T_1T_2} + v\vec{T_1T_3} \quad (4.11)$$

Toute position de P est donc calculable à partir de (u, v) et des positions de (T_1, T_2, T_3) .

La séquence vidéo intitulée *Contrôle des expressions synthétiques* présente une comparaison des résultats d'association point à point et point à polygone.

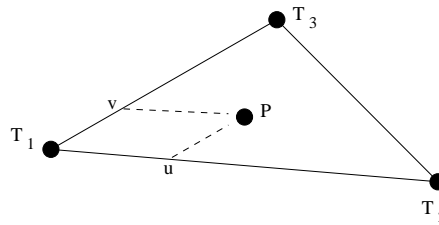


FIG. 4.16 - Coordonnées barycentriques (triangle)

Déplacement 3D

Compte tenu du dispositif de capture des postures faciales disponible, l'information de la position des marqueurs placés sur le visage du sujet est en 2D. La caméra infrarouge ne permet de capturer que les coordonnées X et Y des marqueurs. Pour animer le visage synthétique, la connaissance de la troisième composante Z s'avère d'une utilité non négligeable dans le cas d'expressions dans lesquelles les déplacements prédominants sont en profondeur (en Z) comme pour un sourire en coin. L'approche adoptée ici est d'essayer de déduire la troisième composante des déplacements de marqueurs virtuels, vu qu'après la capture des positions des marqueurs réels en deux dimensions, la composante en Z n'est plus accessible.

L'artiste intervient donc à ce niveau pour juxtaposer les postures neutres de la face et du masque de manière à ce que la répartition des marqueurs virtuels sur la face corresponde à la répartition des marqueurs réels sur le visage du sujet. La position en 3D du masque à la posture neutre est ainsi obtenue de manière à rendre le port du masque par la face le plus cohérent possible. Lors du déplacement des points de la face, les calculs décrits antérieurement sont effectués pour les dimensions X et Y . Pour la dimension Z , la détermination se base sur l'hypothèse que la variation de la normale à un polygone (triangle) de la face varie très peu lorsque celui-ci subit une légère déformation³. Ainsi, pour un point P donné de la face, la nouvelle composante z' de P' est fonction de la normale au polygone \vec{N} et de P en résolvant l'équation 4.12. Le polygone utilisé dans la détermination de z' est choisi en tenant compte du déplacement XY du point P' . Parmi les polygones auxquels P appartient, on choisit celui vers lequel P' se déplace en XY (figure 4.17).

La séquence vidéo intitulée *Composante z* illustre l'impact de l'utilisation ou non de la composante

3. Cette hypothèse ne marche pas pour toutes les régions du visage (lèvres, coins des yeux, coins du nez). Elle est adaptée aux zones de grande surface (front, joues). Toutefois, l'influence de la composante z dans la détermination des coordonnées d'expressions est peu importante comparativement à celle de x et y pris ensemble. Les résultats obtenus avec ou sans estimation de z sont très similaires.

z des déplacements des marqueurs virtuelles.

$$\vec{N} \cdot P\vec{P}' = 0 \quad (4.12)$$

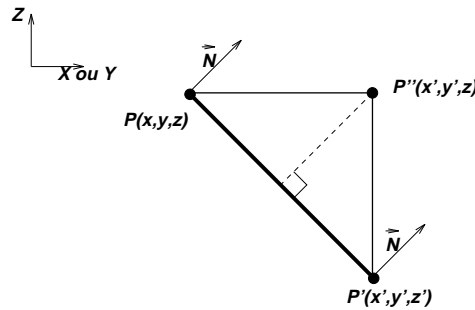


FIG. 4.17 - Calcul de la composante en Z

Les méthodes d'association des marqueurs aux points de la face présentées plus haut (port du masque par la face) permettent toutes la communication des mouvement du masque à la face d'une manière assez précise. Toutefois, les associations points à polygones et celle utilisant une zone d'influence pour les marqueurs présentent un avantage par rapport à l'association point à point. Elles permettent d'avoir plus de points déplacés sur la face, et donc mieux contraindre la détermination automatique des coordonnées d'expression. En outre, l'association point à point peut être faite à partir de celle utilisant une zone d'influence pour chaque marqueur en spécifiant des zones d'influence très petites. Ainsi la méthode d'association adoptée est celle utilisant une zone d'influence pour chaque marqueur.

Quand bien même l'association des marqueurs et les points de la face et la répartition des marqueurs sur le visage de l'acteur seraient idéales, les coordonnées d'expression déterminées sont utilisées pour combiner linéairement les expressions contenues dans la banque. Cela fait que le choix de ces expressions est d'une importance capitale dans l'aspect visuel et esthétique de l'expression finale de synthèse.

4.4 Choix des expressions nécessaires de la banque

La redondance des expressions de la banque est néfaste pour la détermination automatique des coordonnées de l'expression finale. La banque est donc scrutée par l'artiste pour en tirer un ensemble consistant d'expressions, du point de vue mathématique, de manière à permettre une détermination automatique de coordonnées plus efficace. Qualitativement, les suggestions suivantes sont faites pour le choix

des expressions de la banque:

1. Éliminer les expressions de la banque qui peuvent être obtenues par combinaison d'autres expressions de la même banque.
2. Éliminer l'expression neutre de la banque (si elle y est).
3. Dans la banque d'expressions, chaque point de la face correspondant à un marqueur doit bouger au moins une fois dans chaque direction.
4. Les zones où il y a des déplacements de points de la face doivent le plus possible correspondre à des régions où se trouvent des marqueurs.
5. Il est préférable d'avoir dans la banque moins d'expressions que trois fois (chaque point a trois coordonnées X, Y, Z) le nombre de points de la face qui sont influencés par les marqueurs.

Une banque d'expressions tenant compte de ces suggestions est susceptible de donner des meilleurs résultats, vu le nombre de marqueurs réduit comparativement au nombre de points de la face. Bien sûr, ces suggestions ont un caractère qualitatif vis-à-vis de l'artiste utilisant le système d'animation. Toutefois, elles sont d'une importance non négligeable au niveau des calculs nécessaires à la détermination automatique des coordonnées d'expressions. Cette importance du point de vue mathématique sera explicitée dans le chapitre suivant traitant de la modélisation mathématique du problème de détermination des coordonnées d'expressions.

Le postulat de départ qui suppose que partant de mouvements de quelques points sur le visage, il est possible de reconstituer l'expression complète prise par ce dernier guide en tout temps le choix des expressions constituant la banque de référence. En effet, le résultat visuel obtenu est en grande partie fonction de la modélisation qu'aura faite l'artiste, mais aussi de son expérience d'utilisation de l'outil d'animation faciale. Toutefois, il est à noter que la qualité visuelle du résultat augmente de façon remarquable en contraignant plus de points de la face à prendre une position donnée (position déterminée à partir du mouvement des marqueurs). Les suggestions faites ci-haut pour le choix des expressions de la banque ont pour but de permettre une détermination automatique des coordonnées d'expressions qui donnent une expression de synthèse la plus proche possible de celle qu'aura faite l'acteur, lorsque le déplacement des marqueurs est capturé sur la face d'un acteur réel. En plus de ces suggestions l'utilisateur peut être guidé dans son choix par certains critères non quantifiables tels que la beauté d'une expression

(i.e. forme de la bouche plus adéquate). L'artiste peut donner un certain caractère à l'expression finale en choisissant un certain type d'expressions dans la banque de référence.

4.5 Subdivision des expressions

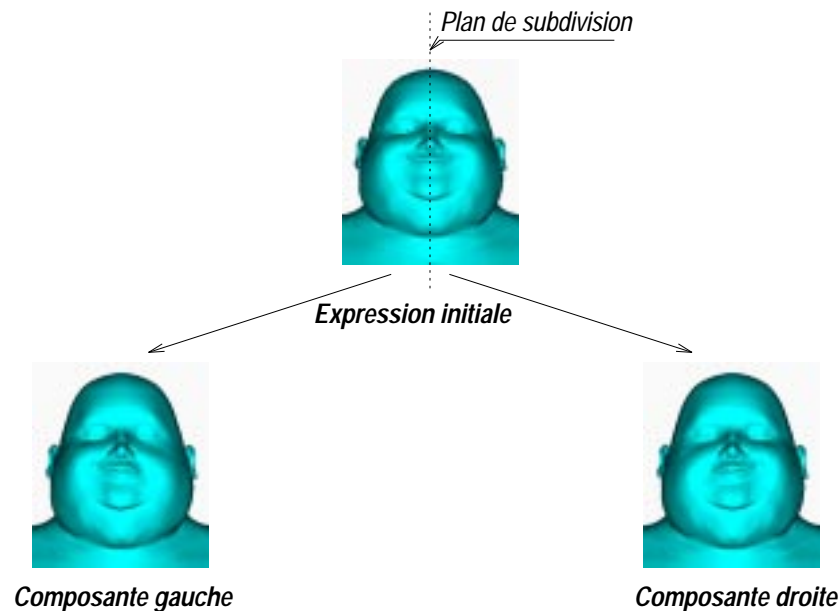
Les suggestions précédentes, prévalant dans le choix des expressions, peuvent être presque toutes vérifiées mathématiquement lors de la détermination automatique des coordonnées d'expression. Cependant, pour le premier critère, vu que les expressions obtenues par combinaison d'autres expressions de la banque sont retouchées par l'artiste, la quantification est plus difficile. Pour outrepasser cette difficulté et utiliser une banque d'expressions plus indépendantes qu'autrement, certaines expressions choisies par l'artiste sont subdivisées en composantes indépendantes de façon dynamique. En outre, cette façon de faire permet d'isoler des régions indépendantes de la face lors de la détermination des coordonnées d'expressions. Un autre point important de la subdivision dynamique des expressions de la banque est qu'il permet de réduire la tâche de modélisation des expressions. En effet, l'artiste n'est plus obligé de modéliser des expressions dissymétriques (plissement de la peau du coin de l'oeil lors d'un clin d'oeil). Il a juste à modéliser des expressions symétriques, ce qui est plus intuitif pour un artiste, et à spécifier que cette expression soit subdivisée lors de la détermination des coordonnées de l'expression finale.

Lors de la synthèse de l'expression finale, la banque d'expressions utilisée est celle générée automatiquement par subdivision (ou non) des expressions spécifiées par l'artiste. La figure 4.18 montre une expression symétrique et ses composantes indépendantes obtenues par subdivision selon un plan vertical passant par le milieu.

La séquence vidéo intitulée *Subdivision des expressions* montre l'influence de la subdivision des expressions dans le résultat final.

4.6 Contraintes sur les coordonnées admissibles des expressions

En plus de pouvoir choisir les expressions entrant en jeu dans la synthèse de l'expression finale et de modéliser les expressions à sa guise, l'artiste (usager) a la possibilité d'imposer des bornes sur les valeurs admissibles des coordonnées d'expressions. Il peut ainsi réduire ou augmenter (ou même annuler) l'influence maximale ou minimale relative d'une expression donnée de la banque sur la synthèse finale. En effet, comme présenté plus haut, un des objectifs de la présente approche est de trouver les coordonnées d'une expression désirée donnée dans un référentiel formé par les expressions contenues dans

FIG. 4.18 - *Subdivision d'expression*

une banque de référence. En pratique, toutes les données contenues dans la banque de référence ne sont pas prises en compte lors de la détermination automatique des coordonnées. Seuls les points de la face qui sont influencés⁴ ($m_i \leq m$) par les marqueurs sont utilisés dans la détermination des coordonnées. Donc l'expression désirée n'est en fait pas une expression, c'est un modèle de la face où les points influencés ont été déplacés conformément aux mouvements des marqueurs. Les coordonnées d'expressions obtenues automatiquement satisfont les positions de ces points influencés lors de la synthèse; les points qui ne sont influencés par aucun marqueur sont positionnés avec ces coordonnées d'expressions sans se préoccuper si leurs positions sont correctes ou non. Les résultats visuels peuvent ainsi être inacceptables dans certains cas. Pour pallier à ce problème, des *bornes* (gardes-fou) sont imposées aux valeurs des coordonnées d'expressions. Ces bornes caractérisent le domaine de définition de la détermination automatique des coordonnées d'expressions. L'influence du choix de ces bornes est illustrée dans un chapitre subséquent traitant des tests et résultats (vidéo: *Bornes sur les valeurs de $[\lambda]$*).

4.7 Le système d'animation faciale

L'animation des expressions faciales du personnage synthétique à partir de la capture des mouvements du visage d'un acteur réel se résume donc ainsi.

4. m_i est le nombre total de points de la face qui sont influencés par les marqueurs.

Après capture de la position neutre des marqueurs du visage de l'acteur, un modèle est généré avec un nombre de points (k points) correspondant au nombre de marqueurs sur le visage de l'acteur; c'est le masque. Les coordonnées des points de ce modèle sont celles des marqueurs. Ce modèle (le masque) est placé en présence de la face du personnage synthétique et les positions des marqueurs virtuels (les points du masque) sont ajustées pour que ceux-ci soient placés dans les mêmes régions que les marqueurs sur le visage de l'acteur. Par la suite, l'utilisateur (l'artiste) spécifie les zones d'influence de chaque marqueur ainsi que son poids et son importance, la banque d'expressions à utiliser et les bornes des valeurs admissibles des coordonnées d'expressions. Le système est alors prêt à fonctionner. Le système d'animation faciale permet d'animer un personnage synthétique en temps réel à partir des mouvements faciaux d'un acteur, pour une trentaine de marqueurs sur la face et un maximum d'environ trente expressions. En pratique, on a rarement besoin de plus de dix-sept expressions pour obtenir une animation de bonne qualité.

Jusqu'à présent, nous avons présenté les notions importantes de la méthode adoptée sans expliquer réellement comment ces étapes sont réalisées.

Une modélisation mathématique du problème de la détermination automatique des coordonnées d'expressions est faite en utilisant des notions bien connues en analyse numérique. Le chapitre suivant décrit cette modélisation mathématique.

Chapitre 5

Modèle mathématique

Dans le chapitre précédent, les concepts de base et les interfaces développés dans l'approche d'animation faciale par banque d'expressions ont été présentés. Cependant, les méthodes utilisées pour déterminer automatiquement et de manière efficace les coordonnées d'expressions et synthétiser l'expression finale, ont été omises. Pour remédier à ce délai, la formulation mathématique du problème est maintenant introduite.

Le problème à résoudre est de composer une expression faciale à partir d'expressions élémentaires constituant la banque d'expressions (base de données). Les données du problème sont la banque de n expressions de référence ($[\vec{\xi}_1], [\vec{\xi}_2], \dots, [\vec{\xi}_n]$) et la posture faciale désirée en terme de vecteur expression ($[\vec{e}]$).

La résolution du problème consiste à déterminer (au mieux) les coordonnées de $[\vec{e}]$ dans $\mathcal{R}([\xi_0]; [\vec{\xi}_1], [\vec{\xi}_2], \dots, [\vec{\xi}_n])$. La combinaison des expressions de la banque de référence pour synthétiser l'expression faciale finale est linéaire comme présenté précédemment. Cela revient donc à trouver $[\lambda] = [\lambda_1, \lambda_2, \dots, \lambda_n]$ tel que $[e] = [\xi_0] + \lambda_1[\vec{\xi}_1] + \dots + \lambda_n[\vec{\xi}_n]$ ($[\vec{e}] = [e] - [\xi_0] = [\xi][\lambda]$).

L'expression vectorielle désirée $[\vec{e}] = [e] - [\xi_0]$ est composée des déplacements (X, Y, Z) des points liés (m_i points influencés par les marqueurs, $k \leq m_i \leq m$) de la face. Supposons pour l'instant que le nombre de points liés est de l'ordre du nombre de marqueurs sur le visage de l'acteur réel, soit quelques dizaines, sur un visage synthétique composé de quelques milliers de points. Dans toutes les résolutions mathématiques qui sont présentées dans ce chapitre, les points considérés (pour la constitution de matrices) sont uniquement les points liés de la face. Les autres points de la face¹ (les points m_f libres) ne rentrent en ligne de compte que lors de la synthèse de l'expression finale. Cela fait que $[\vec{e}]$ est une

1. m_f est le nombre de points de la face qui ne sont influencés par aucun marqueur. $m_f = m - m_i$.

approximation grossière de l'expression finale désirée. Un raffinement de $[\vec{e}]$ pourrait être fait par l'augmentation du nombre de points liés, mais cela introduit un coût additionnel dans la résolution du système d'équations.

Le déplacement des points liés se fait à partir des déplacements des marqueurs de l'acteur réel. Il faut noter que l'acteur et le personnage synthétique ne sont pas nécessairement à la même échelle. La communication des déplacements de marqueurs réels est donc sujette à certains ajustements pour que les déplacements des points liés de la face soient des plus précis possibles. Les ajustements effectués sont différents niveaux d'échelle dont

- une mise à l'échelle automatique entre la face de l'acteur et du personnage synthétique,
- une amplification ou atténuation du mouvement de certains marqueurs,
- une amplification ou atténuation globale des mouvements de tous les marqueurs.

Ainsi le déplacement du point lié est dérivé à partir de celui du marqueur réel auquel ont été appliqués les différents niveaux d'échelle.



Actrice réelle

Acteur synthétique au neutre

FIG. 5.1 - Mise à l'échelle de mouvements de marqueurs

5.1 Données de la banque d'expressions

La résolution du problème de détermination des coordonnées d'une expression donnée dans la base $([\vec{\xi}_1], [\vec{\xi}_2], \dots, [\vec{\xi}_n])$ consiste à trouver les composantes du vecteur $[\lambda]$ tel que $[\vec{\xi}][\lambda] = [\vec{e}]$ avec

- $[\vec{\xi}]$: une matrice contenant les données de la banque d'expressions

- $[\vec{e}]$: un vecteur contenant les données de l'expression désirée
- $[\lambda]$: le vecteur des coordonnées de $[\vec{e}]$ dans $([\vec{\xi}_1], [\vec{\xi}_2], \dots, [\vec{\xi}_n])$.

Les données de la banque d'expressions sont contenues dans $[\vec{\xi}]$ sous forme de colonnes de vecteurs expressions; les composantes de chaque vecteur expression de la banque constituent une colonne de $[\vec{\xi}]$. Si chaque expression de la banque est représentée par m_i points liés, alors chaque colonne de $[\vec{\xi}]$ est un vecteur de dimension $3m_i$ ou $2m_i$ suivant que l'estimation de la composante z des déplacements des points liés est faite ou non.

5.2 Expression faciale désirée

L'expression faciale désirée est, comme mentionné plus haut, la posture faciale prise par l'acteur réel, transmise au personnage synthétique. À partir des liens établis entre les deux ensembles de points (*modèle* des marqueurs de l'acteur réel et *modèle* du visage du personnage synthétique) lors de l'opération d'association, l'expression faciale désirée $[\vec{e}]$ est obtenue en déplaçant les points de la face conformément aux déplacements des marqueurs réels. Suivant le mode d'association, la déformation de la face est faite à partir des formules de déplacement présentées antérieurement (équations 4.3, 4.11, 4.7 du chapitre précédent).

5.3 Mise à l'échelle

Le visage du personnage synthétique et celui de l'acteur réel ne sont généralement pas à la même échelle. Pour communiquer les mouvements du visage de l'acteur réel à la face synthétique, une mise à l'échelle est effectuée sur chaque marqueur (car les marqueurs sont indépendants les uns des autres). Cela donne un premier niveau d'échelle. Les deuxième et troisième niveaux d'échelle sont ceux qui permettent à l'artiste d'amplifier ou d'atténuer le déplacement de certains marqueurs en particulier ou de l'ensemble des marqueurs.

5.3.1 Calcul des facteurs d'échelle des marqueurs

Au début d'une séance de capture de postures faciales sur un acteur, ce dernier essaie de produire les postures maximales des différentes régions de son visage de manière à enregistrer les déplacements

extrêmes de chaque marqueur (*masque*) dans chacune des dimensions (X, Y, Z). Notez que *min* dans ce contexte signifie un déplacement maximal, mais dans la direction négative de l'axe considéré.

- déplacement maximal en X : $\max(x_{masque}) \in [0.0, +\infty)$
- déplacement minimal en X : $\min(x_{masque}) \in (-\infty, 0.0]$
- déplacement maximal en Y : $\max(y_{masque}) \in [0.0, +\infty)$
- déplacement minimal en Y : $\min(y_{masque}) \in (-\infty, 0.0]$
- déplacement maximal en Z : $\max(z_{masque}) \in [0.0, +\infty)$
- déplacement minimal en Z : $\min(z_{masque}) \in (-\infty, 0.0]$

Compte tenu du dispositif de capture utilisé (caméra 2D dans notre situation), les postures faciales sont en deux dimensions et la composante z est donc nulle ($z = 0.0$).

Une inspection de la banque d'expressions de base est effectuée pour déterminer les déplacements extrêmes des points liés de la face (*face*) compte tenu des valeurs extrêmes admissibles des coordonnées d'expressions faciales.

- déplacement maximal en X : $\max(x_{face}) \in [0.0, +\infty)$
- déplacement minimal en X : $\min(x_{face}) \in (-\infty, 0.0]$
- déplacement maximal en Y : $\max(y_{face}) \in [0.0, +\infty)$
- déplacement minimal en Y : $\min(y_{face}) \in (-\infty, 0.0]$
- déplacement maximal en Z : $\max(z_{face}) \in [0.0, +\infty)$
- déplacement minimal en Z : $\min(z_{face}) \in (-\infty, 0.0]$

Ainsi, pour chaque point lié de la *face*, les déplacements extrêmes dans chacune des dimensions sont déterminés.

Le facteur d'échelle de chaque marqueur est calculé pour chaque dimension (X, Y, Z) pour les déplacements positifs (maximum) et négatifs (minimum).

- échelle en X^+ : $\frac{\max(x_{face})}{\max(x_{masque})}$

- échelle en X^- : $\frac{\min(x_{face})}{\min(x_{masque})}$
- échelle en Y^+ : $\frac{\max(y_{face})}{\max(y_{masque})}$
- échelle en Y^- : $\frac{\min(y_{face})}{\min(y_{masque})}$
- échelle en Z^+ : $\frac{\max(z_{face})}{\max(z_{masque})}$
- échelle en Z^- : $\frac{\min(z_{face})}{\min(z_{masque})}$

La figure 5.2 illustre les déplacements extrêmes des marqueurs réels et virtuels par des *boîtes englobantes* alignées sur les axes.

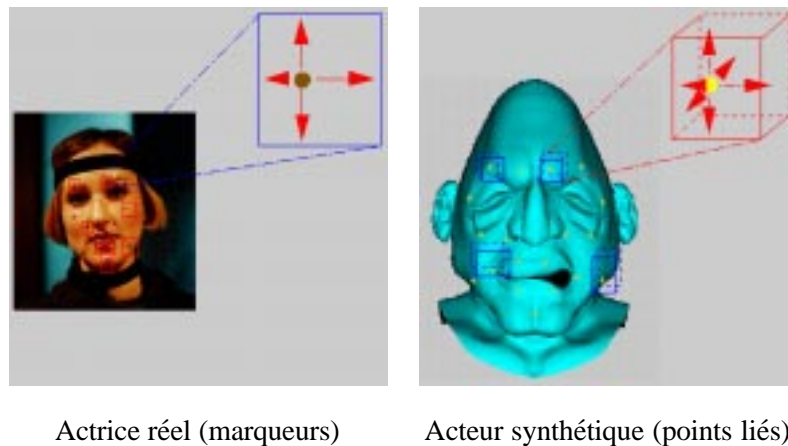


FIG. 5.2 - Boîtes englobantes des marqueurs et des points liés

5.3.2 Déplacement final

En plus du facteur d'échelle déterminé automatiquement pour chaque marqueur, deux niveaux d'échelle permettent de compléter les ajustements faits aux données de capture lors de leur communication au visage synthétique. Il y a d'abord les facteurs d'échelle individuels (contrôle manuel) de chaque marqueur spécifiant l'amplification ou l'atténuation des déplacements de ce marqueur. Puis il y a le facteur d'échelle global des déplacements des marqueurs.

Le déplacement final des points liés est ainsi une concaténation des ajustements automatiques (facteurs d'échelle automatiques) et manuels explicités ci-haut.

Soit à déplacer un point lié de la face $P(x_{face}, y_{face}, z_{face})$ à partir du déplacement d'un marqueur $T(x_{masque}, y_{masque}, z_{masque})$; étant données les conditions suivantes:

$$\text{facteur d'échelle automatique } [auto] = \begin{bmatrix} autoX & 0 & 0 \\ 0 & autoY & 0 \\ 0 & 0 & autoZ \end{bmatrix}$$

$$\text{facteur d'échelle manuel } [man] = \begin{bmatrix} manX & 0 & 0 \\ 0 & manY & 0 \\ 0 & 0 & manZ \end{bmatrix}$$

$$\text{facteur d'échelle global } [global] = \begin{bmatrix} globalX & 0 & 0 \\ 0 & globalY & 0 \\ 0 & 0 & globalZ \end{bmatrix}$$

La nouvelle position P' de P est donc obtenue par:

$$P' = P + [global][man][auto](T\vec{T}') \quad (5.1)$$

Pour le facteur $[auto]$, la mise à l'échelle est fonction du signe de la composante du vecteur $T\vec{T}'$.

5.4 Formulation mathématique du problème

Le problème à résoudre consiste à déterminer les coordonnées de l'expression désirée $[\vec{e}]$ dans un référentiel formé par les expressions de la banque $[\vec{\xi}]$. C'est-à-dire déterminer $[\lambda]$ tel que $[\vec{\xi}][\lambda] = [\vec{e}]$ sujet à $[\lambda_{min}] \leq [\lambda] \leq [\lambda_{max}]$ ($\lambda_{min_i} \leq \lambda_i \leq \lambda_{max_i}$). Pour une banque d'expressions de référence avec n expressions et une face contenant k points liés, les matrices en présence ont les dimensions suivantes:

- $[\lambda]$, $[\lambda_{min}]$ et $[\lambda_{max}]$ sont des vecteurs colonne de dimension n . $[\lambda]$ constitue les poids à donner à chaque expression de $[\vec{\xi}]$ pour que la combinaison linéaire des vecteurs expressions $[\vec{\xi}_i]$ ayant pour coefficient λ_i soit égale à $[\vec{e}]$. $[\lambda_{min}]$ et $[\lambda_{max}]$ sont les bornes des valeurs admissibles de $[\lambda]$.
- $[\vec{\xi}]$ est une matrice de dimensions $2m_i \times n$ (si on ne considère que les déplacements $2D$) ou $3m_i \times n$ (si la composante z est considérée). C'est la représentation de l'information pertinente contenue dans la banque d'expressions.

- $[\vec{e}]$ est un vecteur colonne de dimension $2m_i$ ou $3m_i$ selon composante z . Il représente l'information pertinente de l'expression désirée.

Au niveau de l'expression désirée de la face, les points déplacés à partir des marqueurs sont uniquement les points liés. Cela fait que l'expression désirée utilisée pour les calculs ($[e]$) se réduit à un modèle de la face dans lequel seulement quelques points ont été déplacés. Ce n'est donc pas une expression complète de la face. Une augmentation du nombre de points liés de la face a pour effet de rapprocher $[e]$ d'une expression faciale complète et donc se traduit par un renforcement des contraintes d'obtention d'une expression faciale par combinaison.

Trouver une solution à $[\vec{\xi}][\lambda] = [\vec{e}]$ se ramène donc à résoudre un système de $3m_i$ (ou $2m_i$) équations à n inconnues. Il est important de noter que la matrice $[\vec{\xi}]$ est une matrice quelconque en terme de dimension, des valeurs de ses éléments et de sa configuration. Il n'est donc pas toujours possible de faire une résolution directe ou même précise. L'importance des marqueurs introduits au chapitre précédent se traduit mathématiquement par une pondération des équations de $[\vec{\xi}][\lambda] = [\vec{e}]$ qui correspondent aux positionnements des points de la face associés aux marqueurs, par une valeur quantifiant l'importance de chaque marqueur.

Plusieurs méthodes ont ainsi été investiguées pour résoudre le système d'équations dont la taille peut quelques fois être significative (si tous les points d'une face de 2000 points sont utilisés avec une banque de 30 expressions de référence, $[\vec{\xi}]$ comporte environ 6000 lignes et 30 colonnes). Une bonne approche de résolution doit être robuste à une éventuelle singularité dans $[\vec{\xi}]$, mais également permettre de résoudre un système d'équations aussi bien sur-contraint que sous-contraint. Une présentation sommaire de quelques unes des méthodes de résolution investiguées est faite dans la section suivante.

5.5 Méthodes de résolution

Plusieurs méthodes de résolution de système d'équations de type $[A][x] = [b]$ ont été investiguées. Les sections suivantes présentent sommairement les plus pertinentes d'entre elles tout en soulignant leurs caractéristiques particulières. La méthode de résolution choisie est ensuite présentée plus en détails.

5.5.1 Résolution pseudo-directe

Résolution par calcul de matrice inverse ou pseudo-inverse

La résolution d'un système d'équations peut être faite de manière quasi directe si la matrice associée est bien conditionnée. Par exemple si la matrice est carrée et de rang maximum, c'est-à-dire si elle n'est pas singulière (ne contient que des colonnes et des lignes indépendantes), son inverse peut être calculée et par la suite, la détermination des inconnues se ramène à un simple produit d'une matrice et d'un vecteur. Lorsque la matrice correspondante n'est pas de rang maximum ou n'est pas carrée, c'est-à-dire qu'il y a plus de lignes que de colonnes ou inversement, et que toutes les colonnes ou les lignes ne sont pas indépendantes les unes des autres, il est possible de résoudre le système en calculant la matrice pseudo-inverse et ainsi obtenir la solution de la même manière que précédemment par produit d'une matrice et d'un vecteur.

La pseudo-inverse $[A]^+$ d'une matrice $[A]$ de dimension $m \times n$ de rang maximum se définit comme suit:

- si $m > n$ alors $[A]^+ = ([A]^t[A])^{-1}[A]^t$
- si $n > m$ alors $[A]^+ = [A]^t([A][A]^t)^{-1}$
- si $n = m$ alors $[A]^+ = [A]^{-1}$

La solution du système d'équations est obtenue par: $[x] = [A]^+[b]$.

Résolution par élimination gaussienne

La méthode de résolution par élimination de Gauss a été investiguée afin de voir dans quelles mesures elle pourrait être utilisée vue la facilité de son implantation. Cette méthode consiste à appliquer des transformations linéaires au système d'équations de manière à réduire la matrice correspondant au système en une matrice triangulaire supérieure. La détermination des inconnues se fait ainsi plus facilement, par substitution inverse des variables par des valeurs.

Résolution par factorisation et substitution

Il existe plusieurs méthodes de décomposition matricielle qui permettent par la suite de résoudre le système d'équations par une suite d'étapes de résolutions intermédiaires triviales. Pour cette approche

de résolution les méthodes investiguées sont:

- décomposition LU ,
- décomposition SVD ,
- décomposition QR .

La décomposition de *Cholesky* a aussi été investiguée mais elle n'est applicable que dans le cas très particulier des matrices carrées symétriques et définies positives.

Décomposition LU

La décomposition LU consiste à factoriser la matrice $[A]$ en deux matrices triangulaires ($[A] = [L][U]$). L'une, $[L]$, est triangulaire inférieure et l'autre, $[U]$, triangulaire supérieure. Pour la résolution du système d'équations, une variable intermédiaire $[y] = [U][x]$ est utilisée de sorte que la solution finale $[x]$ s'obtient en deux étapes. Le système d'équations de départ est transformé pour donner:

$$[L][y] = [b],$$

$$[U][x] = [y].$$

À la première étape, le système d'équations $[L][y] = [b]$ est résolu par substitution directe pour obtenir la valeur intermédiaire $[y]$. La solution finale est obtenue à partir de $[U][x] = [y]$ par substitution inverse.

Décomposition SVD

La décomposition en valeurs singulières consiste à déterminer les matrices $[U]$, $[S]$ et $[V]$ de sorte que $[A] = [U][S][V]^t$ avec:

- $[U]$ une matrice orthogonale de dimension $m \times m$,
- $[V]$ une matrice orthogonale de dimension $n \times n$,
- $[S]$ une matrice diagonale de dimension $m \times n$.

Les valeurs de la diagonale de $[S]$ (s_i) sont les valeurs singulières de $[A]$. La solution du système d'équations est donnée par $[x] = [V][diag(1/s_i)][U]^t[b]$.

Décomposition QR

La décomposition QR est une factorisation de la matrice $[A]$ telle que:

- $[A] = [Q][R]$,
- $[Q]$ une matrice orthogonale,
- $[R]$ une matrice triangulaire supérieure.

Pour résoudre le système d'équations de départ, l'orthogonalité de $[Q]$ est exploitée pour effectuer la transformation suivante:

$$[R][x] = [Q]^t[b].$$

Par la suite, les valeurs de $[x]$ sont obtenues par substitution inverse comme dans le cas de la décomposition LU .

5.5.2 Résolution par algorithme de minimisation de fonction

La solution donnée par les méthodes de résolution directe et pseudo-directe est une solution exacte du système d'équations linéaires. Toutefois, il n'est pas toujours possible que le système admette une solution, la matrice correspondant au système d'équations étant quelconque. On ne peut donc pas se fier à ces méthodes dans notre système d'animation faciale où il faut toujours fournir une solution. Pour s'assurer d'avoir toujours une solution, même lorsque le système d'équations n'admet pas de solution exacte, des méthodes de résolution par minimisation de fonction ont été investiguées. Ces méthodes se basent sur l'existence d'une fonction (fonction objective) dont la valeur doit être extrême (minimale ou maximale). Pour la résolution d'un système d'équations linéaires $[A][x] = [b]$, la fonction à minimiser est la norme du vecteur résiduel $[A][x] - [b]$. D'où la résolution du système se ramène à trouver un minimum (local) d'une fonction multi-variables $f([x]) = \|[A][x] - [b]\|$. Les méthodes par excellence qui ont été investiguées sont celle de *Newton* et la méthode du *gradient conjugué* dont les détails sont longuement présentés dans les ouvrages spécialisés [12].

5.5.3 Résolution par moindres carrés linéaires

La résolution par moindres carrés linéaires [6] se base sur le même principe que l'approche précédente qui consiste à minimiser la valeur d'une fonction. Le problème à résoudre peut être défini de la

manière suivante. Étant donnée une matrice $[A]$ de dimension $m \times n$ et de rang $k \leq \min(m, n)$ et étant donné un vecteur $[b]$ de dimension m , trouver un vecteur $[x_0]$ de dimension n de sorte que la norme euclidienne du vecteur $[A][x] - [b]$ soit minimale.

Contrairement aux résolutions par minimisation de fonction, les moindres carrés linéaires n'utilisent pas directement la fonction à minimiser pour trouver la solution du problème. La méthode se base plutôt sur des transformations matricielles orthogonales du système $[A][x] = [b]$ de départ pour résoudre le problème de manière plus efficace.

La résolution par moindres carrés peut être conditionnée pour que la solution satisfasse certaines exigences données. Cette flexibilité fait que la présente approche de résolution se prête mieux au type de problème à résoudre dans le cas de coordonnées d'expressions synthétiques. En effet, les données contenues dans les matrices correspondant au système d'équations à résoudre sont d'une part incertaines compte tenu du fait qu'il pourrait y avoir des erreurs de mesure ou de chiffres significatifs des valeurs numériques, et d'autre part elles sont insuffisantes car le nombre de points liés de la face est très petit comparativement au nombre de points libres. Cela fait que les *expressions de référence* contenues dans la matrice du système d'équations sont des expressions grossières (expressions très incomplètes). En outre les informations de déplacement obtenues par capture de postures faciales sont en deux dimensions (X, Y). L'information de la troisième dimension (Z), lorsqu'elle est utilisée, est incertaine car obtenue par estimation. Lorsqu'elle ne l'est pas, elle fait défaut dans le type de résultats possibles à obtenir pour l'expression finale de synthèse.

Enfin, pour conclure avec l'incertitude et le manque d'information du problème à résoudre, il faut tenir compte du fait que l'utilisation uniquement de l'information pertinente (les points liés) pour résoudre le problème fait que le résultat obtenu garantit la position des points liés dans la synthèse de l'expression finale, mais ne contrôle pas totalement la position des points libres de la face. Ces derniers peuvent se trouver n'importe où dans l'espace $3D$ et ainsi produire une expression de synthèse inacceptable visuellement, vu les niveaux de qualité et de réalisme recherchés dans l'animation finale.

Différents types de conditionnements de la résolution par moindres carrés ont conduit à utiliser les méthodes particulières suivantes:

- NNLS: *Non-Negative Least Squares*,
- LSI: *Least Squares with Inequality*.

NNLS

C'est une approche de résolution du problème des moindres carrés [6] qui assure que les valeurs de solutions obtenues sont positives ou nulles. La spécification du problème des moindres carrés, compte tenu de ce conditionnement est la suivante:

étant donnée une matrice $[A]$ de dimension $m \times n$ et de rang $k \leq \min(m, n)$ et étant donné un vecteur $[b]$ de dimension m , trouver un vecteur $[x_0]$ de dimension n de sorte que la norme euclidienne du vecteur $[A][x] - [b]$ soit minimale avec $[x_0] \geq [0]$ (i.e. minimiser $\|[A][x] - [b]\|$ sujet à $[x] \geq [0]$).

Les coordonnées négatives pour une expression produisent des résultats qui sont souvent inesthétiques visuellement, suivant que les points libres ont des déplacements dans la banque d'expressions de référence qui se prêtent mal à une mise à échelle négative. De ce fait, à moins de nécessité particulière, les valeurs de la solution au problème de moindres carrés sont contraintes à rester positives ou nulles.

LSI

Cette approche de résolution du problème des moindres carrés [6] permet d'augmenter les contraintes de résolution de sorte que le résultat satisfasse un certain nombre d'exigences. La spécification est la suivante:

étant donnée une matrice $[A]$ de dimension $m \times n$ et de rang $k \leq \min(m, n)$ et étant donné un vecteur $[b]$ de dimension m , trouver un vecteur $[x_0]$ de dimension n de sorte que la norme euclidienne du vecteur $[A][x] - [b]$ soit minimale avec $[G][x_0] \geq [h]$ (i.e. minimiser $\|[A][x] - [b]\|$ sujet à $[G][x] \geq [h]$)

La matrice $[G]$ et le vecteur $[h]$ spécifient les exigences supplémentaires que devra satisfaire le résultat. Si $[A]$ est de dimension $m \times n$ alors $[G]$ est une matrice de n colonnes. La matrice $[G]$ et le vecteur $[h]$ sont de dimension respectives $m_2 \times n$ et m_2 . m_2 est donc le nombre de contraintes auxquelles les valeurs de $[x]$ sont sujettes.

Il s'ensuit qu'il est possible de spécifier des bornes supérieures et/ou inférieures quelconques sur les valeurs de $[x]$ ou même de spécifier des contraintes sur la combinaison des composantes de $[x]$ à l'aide de l'approche *LSI*. Ainsi, soit à calculer la solution $[x]$ de $[A][x] = [b]$ tel que pour $i = 1, \dots, n$ les composantes de $[x]$ satisfassent $x_i \geq \beta_i$ et $x_i \leq \alpha_i$ ($m_2 = 2n$). Ces contraintes seront spécifiées dans

les matrices $[G]$ et $[h]$ comme suit:

$$G = \begin{bmatrix} 1.0 & 0.0 & \dots & 0.0 \\ -1.0 & 0.0 & \dots & 0.0 \\ 0.0 & 1.0 & \dots & 0.0 \\ 0.0 & -1.0 & \dots & 0.0 \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ 0.0 & 0.0 & \dots & 1.0 \\ 0.0 & 0.0 & \dots & -1.0 \end{bmatrix} \quad h = \begin{bmatrix} \beta_1 \\ -\alpha_1 \\ \beta_2 \\ -\alpha_2 \\ \cdot \\ \cdot \\ \cdot \\ \beta_n \\ -\alpha_n \end{bmatrix}$$

Le conditionnement du problème des moindres carrés pour satisfaire *NNLS* est donc un cas particulier de *LSI*.

L'approche *LSI* peut également donner un résultat dont la norme est minimale et cela, grâce à l'utilisation des équations normales. Le système d'équations de départ $[A][x] = [b]$ est remplacé par un système d'équations équivalent $[A]^t[A][x] = [A]^t[b]$. La résolution du nouveau système d'équations fournit une solution dont la norme euclidienne est minimale et qui est solution du système d'équations de départ. Ainsi, une résolution par *LSI* de $[A]^t[A][x] = [A]^t[b]$ sujet à $[G][x] \geq [h]$ permet d'obtenir un résultat de norme minimale et satisfaisant les contraintes spécifiées dans $[G]$ et $[h]$. Toutefois, l'utilisation des équations normales sous la forme $[A]^t[A][x] = [A]^t[b]$ n'est possible que si $m \geq n$.

Dans notre cas, il est très rare que le nombre d'expressions contenues dans la banque soit supérieur à trois fois le nombre de points liés de la face. La condition d'utilisation des équations normales est donc satisfaite. Dans les cas rares où cette condition n'est pas satisfaite, l'algorithme *LSI* est utilisé dans sa version originale sans les équations normales.

5.6 Choix des expressions nécessaires de la banque

Dans le chapitre précédent, certaines suggestions ont été introduites relativement au choix des expressions de la banque de référence. Ces suggestions consistaient à suivre les directives suivantes:

1. Éliminer les expressions de la banque qui peuvent être obtenues par combinaison d'autres expressions de la même banque.

2. Éliminer l'expression neutre de la banque (si elle y est).
3. Dans la banque d'expressions, chaque point de la face correspondant à un marqueur doit bouger au moins une fois dans chaque direction.
4. Les zones où il y a des déplacements de points de la face doivent le plus possible correspondre à des régions où se trouvent des marqueurs.
5. Il est préférable d'avoir dans la banque moins d'expressions que trois fois (chaque point a trois coordonnées X, Y, Z) le nombre de points de la face qui sont influencés par les marqueurs.

Les significations mathématiques de ces suggestions sont énumérées ci-dessous:

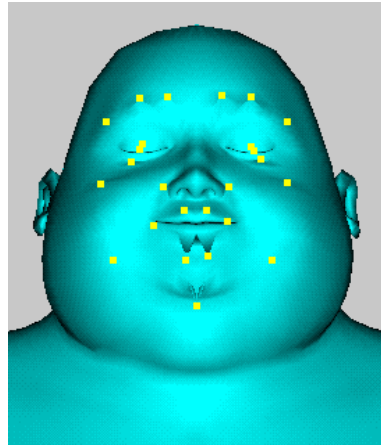
1. Assurer l'indépendance linéaire des colonnes de la matrice $[\vec{\xi}]$.
2. Éliminer une colonne correspondant à $[\vec{0}]$ dans $[\vec{\xi}]$.
3. Éliminer les lignes correspondant à $[\vec{0}]^t$ dans $[\vec{\xi}]$.
4. Éliminer les colonnes correspondant à $[\vec{0}]$ dans $[\vec{\xi}]$.
5. Avoir une matrice $[\vec{\xi}]$ avec plus de lignes que de colonnes (un système d'équations surcontraint) de manière à pouvoir utiliser les équations normales.

Le respect de ces suggestions permet d'obtenir un vecteur $[\lambda]$ plus précis et de faire une résolution plus efficace.

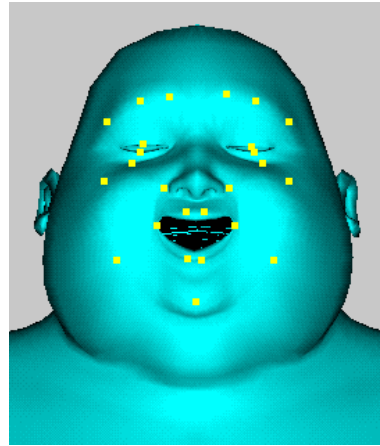
5.7 Utilisation du résultat de la résolution

La résolution de $[\vec{\xi}][\lambda] = [\vec{e}]$ donne le vecteur $[\lambda]$ qui est composé des coordonnées de $[\vec{e}]$ dans le système de coordonnées $([\xi_1], [\xi_2], \dots, [\xi_n])$. L'étape subséquente à la résolution du système d'équations est de voir à l'exactitude tant visuellement que mathématiquement des résultats obtenus. Pour ce faire, les coordonnées contenues dans $[\lambda]$ (solution du système d'équations) sont utilisées pour composer l'expression $[\vec{e}_c] = [\vec{\xi}_c][\lambda]$, mais cette fois-ci, $[\vec{e}_c]$ se compose de tous les points de la face (liés et non liés) et $[\vec{\xi}_c]$ est une matrice dont chaque colonne correspond à une expression complète de la banque de référence. Ainsi, pour résumer l'ensemble du processus, il faut retenir que la résolution du système d'équations fait intervenir un nombre réduit de points de la face et les résultats sont appliqués à l'ensemble des points.

La figure 5.3 présente les positions de points liés fournies au système comme expression désirée et la reconstitution de l'expression à partir des coordonnées calculées.



Expression grossière $[e]$
(seulement points liés déplacés)



Expression complète synthétisée $[e_c]$

FIG. 5.3 - Expressions $[e]$ et $[e_c]$

L'économie de temps de calcul faite lors de la résolution du système d'équations en utilisant juste l'information pertinente serait inutile s'il faut la perdre lors de la recombinaison des expressions de référence (synthèse de l'expression résultat). Pour continuer dans la même direction de réduction du temps de calcul, la synthèse de l'expression résultat ne tient compte que de l'information pertinente contenue dans la banque d'expressions; c'est-à-dire que seuls les points ayant un déplacement non nul sont utilisés pour le calcul de l'expression résultat $[e_c]$.

Les coordonnées d'expressions déterminées à partir de la résolution numérique du système d'équations linéaires résultant de la modélisation mathématique du problème, permettent de générer des expressions de synthèse dont la qualité visuelle peut être qualifiée de bonne. Pour valider la modélisation mathématique, certains tests ont été effectués. Les résultats obtenus sont présentés sous forme de séquences d'animations et sont discutés dans le chapitre qui suit.

Chapitre 6

Résultats

La méthode adoptée pour animer le visage d'un personnage synthétique consiste à composer une expression faciale de ce dernier à partir d'expressions de référence constituant une banque. Ainsi, pour faire prendre à la face du personnage synthétique une expression équivalente à la posture faciale capturée sur le visage d'un acteur réel, un processus de résolution de système d'équations linéaires est effectué afin de déterminer les coordonnées de l'expression équivalente de la face, relativement à un système de coordonnées défini à partir des expressions de référence et de la posture neutre de la face. Par la suite, les coordonnées sont utilisées pour synthétiser l'expression de la face par combinaison linéaire des expressions de référence. Par ailleurs, il faut noter que la détermination des coordonnées est faite uniquement avec les points de la face qui sont liés (associés à des marqueurs) tandis que la synthèse de l'expression finale, utilisant ces coordonnées tient compte de tous les points de la face. Les expressions de référence contenues dans la banque d'expressions (qui constitue le système de coordonnées dans lequel l'expression synthétique est composée) sont modélisées par l'artiste, de là le premier impact de son empreinte sur le résultat final. Les expressions de référence sont des modèles de la face composés de points et de polygones. Ce sont les informations du modèle qui sont pertinentes pour les besoins de la méthode préconisée dans ce mémoire.

6.1 Validation de la méthode utilisée

Pour valider l'approche d'animation adoptée d'une part et le modèle mathématique sous-tendant d'autre part, un certain nombre de tests ont été effectués.

La question à se poser pour la validation de l'approche utilisée est la suivante:

Est-il possible de composer une expression faciale donnée à partir d'une combinaison d'un certain nombre de postures faciales (élémentaires) bien choisies ?

ou de manière inverse:

Est-il possible de décomposer une expression faciale donnée en une combinaison d'un certain nombre de postures faciales (élémentaires) bien choisies ?

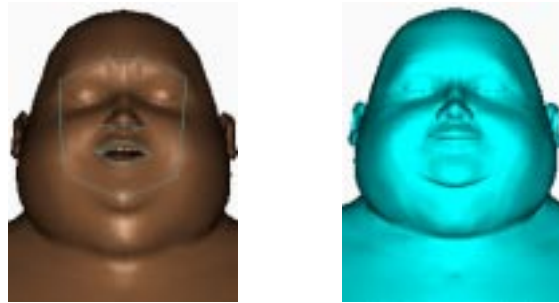
La réponse à l'une ou l'autre de ces questions n'est pas absolument affirmative. Il faut émettre certaines réserves quant à ce qui est faisable en pratique, contrairement au modèle mathématique sous-tendant présenté au chapitre précédent. En effet, la gamme des expressions faciales qu'il est possible d'obtenir à partir d'une banque de référence donnée est conditionnée par la banque d'expressions d'une part et d'autre part par le domaine de définition du système d'équations $[A][x] = [b]$ et l'information pertinente relative à l'expression désirée qu'il est possible d'extraire. Plus l'expression désirée est complète et plus la combinaison de la banque d'expressions à faire pour l'obtenir est précise. La question de la banque d'expressions bien choisie est donc importante parce qu'un déplacement d'un point dans l'expression désirée n'est accessible par combinaison des expressions de la banque que si ce point a un déplacement non nul dans au moins une expression de la banque (condition nécessaire). La prise en compte de cette condition permet de déterminer un premier niveau de consistance dans la banque d'expressions de référence. Ainsi, pour un ensemble donné de marqueurs, pour que toutes les postures des marqueurs soient accessibles par combinaison des expressions de la banque, il faut que chaque point de la face qui est associé à un marqueur ait un déplacement non nul en X et Y dans au moins une des expressions de la banque.

6.1.1 Illustration de la condition nécessaire de consistance

Pour la reconstitution d'une posture faciale donnée (de l'acteur réel ou d'un autre personnage synthétique) sur la face du personnage synthétique de la figure 6.1, deux séries de tests de consistance ont été effectués.

Banque de référence ne respectant pas la condition nécessaire

Soit à reconstituer l'expression faciale de la figure 6.1 à partir de la banque d'expressions de la figure 6.2.



Expression à composer Acteur synthétique au neutre

FIG. 6.1 - *Consistence*

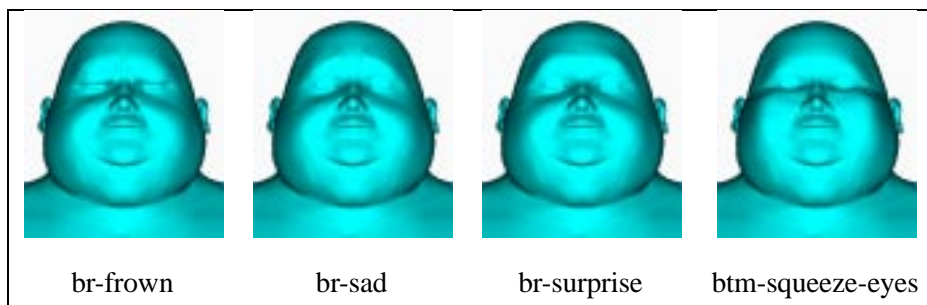


FIG. 6.2 - *Banque d'expressions ne contenant aucun mouvement de la bouche*

Il faut noter que dans cette banque de référence, les points de la face associés aux marqueurs de la bouche n'ont aucun déplacement. Mathématiquement, il est donc impossible de rejoindre les déplacements de ces marqueurs par combinaison linéaire des expressions de la banque.

Le résultat de la résolution donne les coordonnées d'expression suivantes:

$$[\lambda] = [0.595, 1.000, 1.000, 0.111]$$

et par suite la synthèse de l'expression de la figure 6.2. Il est facile de remarquer que la forme de la bouche est différente lorsqu'on passe de l'expression à composer et celle reconstituée à partir de la banque de la figure 6.3. Les déplacements des points de la bouche sont inaccessibles. Pour le haut du visage (front, sourcils, paupières) les deux expressions sont similaires.



Expression à composer



Expression reconstituée
avec banque de la figure 6.2

FIG. 6.3 - *Consistence*

Banque de référence respectant la condition nécessaire

L'utilisation d'une banque d'expressions de référence consistente relativement à la condition nécessaire à l'accessibilité des déplacements de tous les marqueurs permet de corriger l'erreur du test ci-dessus. Cette banque peut être obtenue par ajout des expressions nécessaires (expressions où les points causant l'inconsistance ont des déplacements non nuls) ou remplacement, voire élimination des expressions nuisibles (i.e. expressions pouvant être reconstituées à partir des autres expressions de la banque).

En utilisant la banque de référence de la figure 6.4, la résolution du système d'équations résultant fournit les coordonnées de $[\lambda]$ ci-dessous:

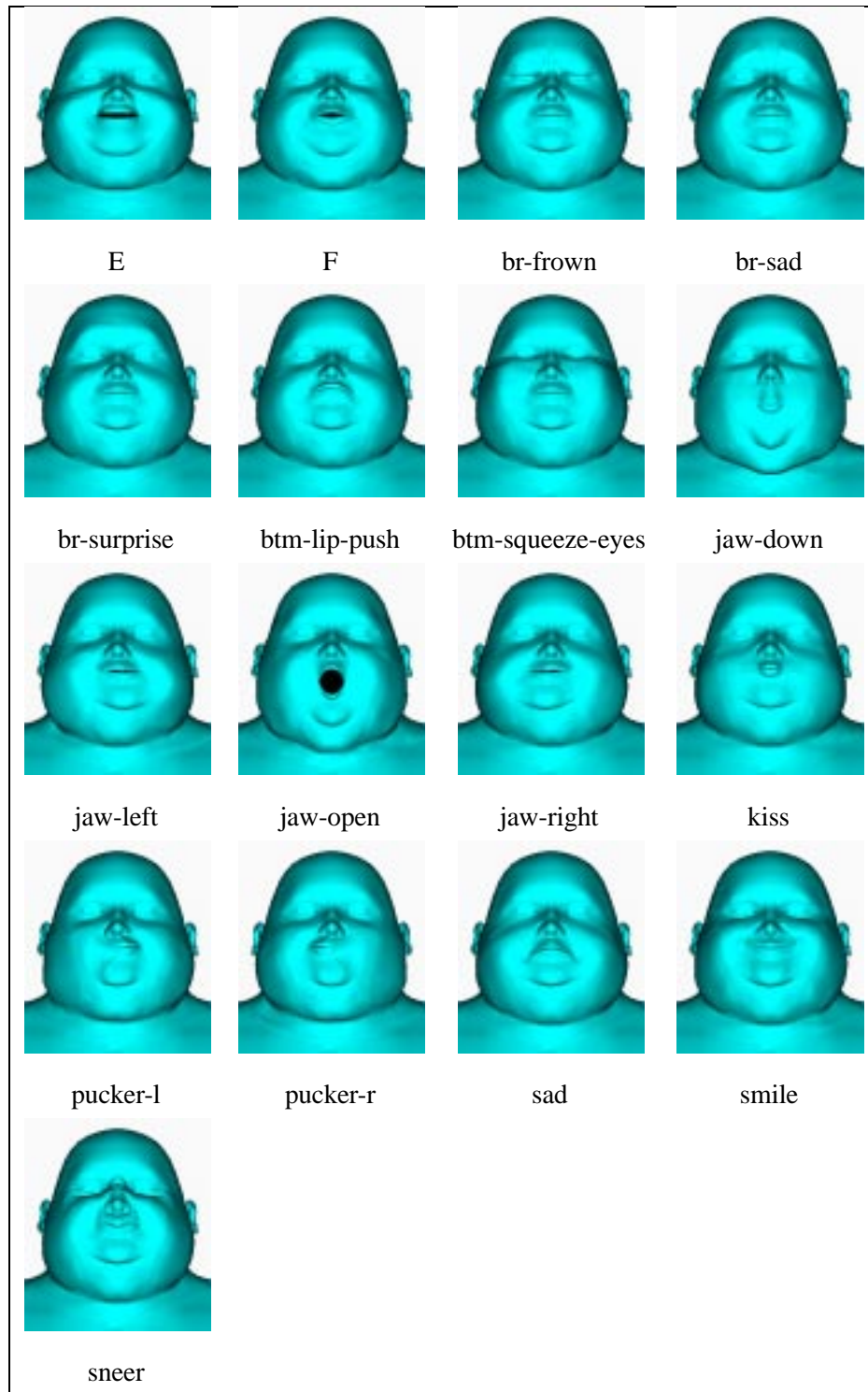


FIG. 6.4 - Banque d'expressions

$$[\lambda] = \begin{bmatrix} 0.107, & 0.452, & 0.586, & 1.000, \\ 1.000, & 9.83 \times 10^{-3}, & 0.00, & 0.00, \\ 0.535, & 0.152, & 0.602, & 0.00, \\ 0.00, & 0.123, & 1.25 \times 10^{-6}, & 5.01 \times 10^{-2}, \\ 1.05 \end{bmatrix}$$

La reconstitution à partir de ces coordonnées est présentée à la figure 6.5.



Expression à composer



Expression composée
avec banque de la figure 6.4

FIG. 6.5 - Expression composée avec banque de la figure 6.4

6.1.2 Illustration de l'influence du domaine de définition du système d'équations

Le domaine de définition du système d'équations peut limiter les expressions qu'il est possible d'obtenir par combinaison. En effet, les valeurs admissibles des coordonnées d'expression contraignent la solution du processus de minimisation qui cherche une solution qui donnerait une expression de synthèse dont la différence avec l'expression désirée est minimale tout en respectant les contraintes du domaine de définition.

Domaine de définition insuffisant

Le présent test consiste à faire la synthèse de l'expression de la figure 6.1 à partir de la banque d'expressions de référence de la figure 6.4 et cela avec différents domaines de définitions pour les valeurs de $[\lambda]$.

Le domaine de définition du système est donné par les bornes sur les valeurs admissibles des coordonnées d'expressions suivantes:

- $0.0 \leq \lambda_1 \leq 0.2$ (E)
- $0.0 \leq \lambda_2 \leq 0.2$ (F)
- $0.0 \leq \lambda_3 \leq 1.0$ (br-frown)
- $0.0 \leq \lambda_4 \leq 1.0$ (br-sad)
- $0.0 \leq \lambda_5 \leq 1.0$ (br-surprise)
- $0.0 \leq \lambda_6 \leq 0.2$ (btm-lip-push)
- $0.0 \leq \lambda_7 \leq 1.0$ (btm-squeeze-eyes)
- $0.0 \leq \lambda_8 \leq 0.2$ (jaw-down)
- $0.0 \leq \lambda_9 \leq 0.2$ (jaw-left)
- $0.0 \leq \lambda_{10} \leq 0.2$ (jaw-open)
- $0.0 \leq \lambda_{11} \leq 0.2$ (jaw-right)
- $0.0 \leq \lambda_{12} \leq 0.2$ (kiss)
- $0.0 \leq \lambda_{13} \leq 1.0$ (pucker-l)
- $0.0 \leq \lambda_{14} \leq 1.0$ (pucker-r)
- $0.0 \leq \lambda_{15} \leq 1.0$ (sad)
- $0.0 \leq \lambda_{16} \leq 1.0$ (smile)
- $0.0 \leq \lambda_{17} \leq 1.0$ (sneer)

La synthèse de l'expression résultante est présentée à la figure 6.6. Les bornes supérieures (0.2) imposées aux coordonnées d'expressions relatives aux expressions de référence produisant l'ouverture de la bouche, contraignent le résultat de la synthèse finale. En effet, la forme de la bouche de l'expression finale est différente de celle de l'expression à composer au niveau de l'amplitude de l'ouverture. Il est possible d'ouvrir la bouche mais l'amplitude de l'ouverture est limitée par les contraintes sur les valeurs des coordonnées d'expressions.

Domaine de définition correct

En jouant sur le domaine de définition (modification des bornes), la résolution du système d'équations permet une synthèse plus consistante de l'expression faciale. Par exemple avec $[0.0] \leq [\lambda] \leq [1.0]$, le résultat de la figure 6.7 semble plus correct.

En définitive, il faut retenir que l'intervention de l'artiste, pour la modélisation des expressions de la banque de référence et pour la spécification du domaine de définition du système d'équations $[A][x] =$

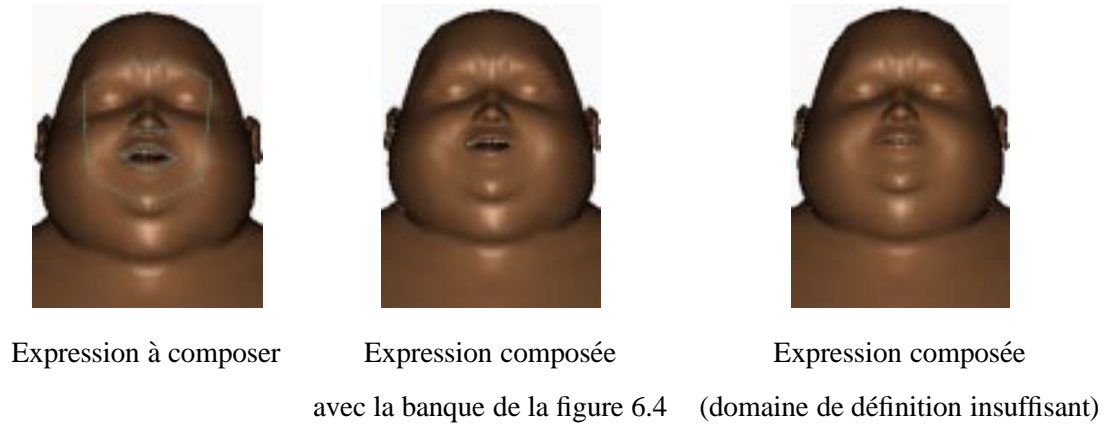


FIG. 6.6 - *Expression composée à partir d'un domaine de définition insuffisant*



FIG. 6.7 - *Expression composée avec $[0.0] \leq [\lambda] \leq [1.0]$*

[*b*] via la spécification des bornes sur les coordonnées d'expressions conditionne ce qu'il est possible d'obtenir en pratique; de là l'empreinte de ce dernier dans l'aspect esthétique du résultat visuel final.

Pour avoir une idée des points introduits ci-haut et des résultats sur l'animation correspondante, le lecteur est invité à visionner les courtes séquences d'animation disponibles aux adresses suivantes (en utilisant *movieplayer*):

<http://www.iro.umontreal.ca/labs/infographie/theses/kouadio/bank-consist.mov>

pour le test de la condition nécessaire,

<http://www.iro.umontreal.ca/labs/infographie/theses/kouadio/df-consist.mov>

pour le test du domaine de définition.

Des séquences d'animation disponibles sur le vidéo du mémoire, intitulées *Bornes sur les valeurs de* [λ] et *Divers* présentent des résultats similaires.

6.1.3 Recomposition d'expression de la banque

La deuxième étape dans le processus de validation de l'approche proposée et du modèle mathématique, est de voir à ce que le système permette de retrouver un résultat des plus précis. Ainsi, les tests suivants ont consisté à spécifier comme expression désirée une expression de la banque de référence. Vingt-cinq points de contrôle (marqueurs virtuels) ont été utilisés pour capturer les déplacements occasionnés par une expression donnée de la banque de référence. Les *animateurs map-expression*, *mixer* et *M-recorder* sont utilisés pour faire la capture des déplacements de l'expression choisie pour les vingt-cinq marqueurs virtuels (masque). Par la suite, ces déplacements sont utilisés comme séquence d'entrée de l'*animateur live-face* pour le calcul des coordonnées d'expression et la reconstitution (synthèse) de l'expression correspondante. Pour ce faire, plusieurs sous ensembles de la banque d'expressions de référence de la figure 6.4 ont été utilisés.

Recomposition de l'expression *F* de la banque

La figure 6.8 présente l'expression de la banque de référence à reconstituer à partir des déplacements occasionnés par celle-ci pour vingt-cinq points de contrôle.



FIG. 6.8 - Expression “F” de la banque de référence de la figure 6.4 ($[\lambda] = [0, 1, 0, 0, \dots, 0]$)

Une expression de référence (*E* (son “i”))

Le résultat ne peut être qu’une multiplication de l’unique expression de la banque par un facteur, comme cela peut s’observer sur la figure 6.9.



Expression à composer Expression reconstituée

FIG. 6.9 - Reconstitution de l’expression “F” de la banque de référence de la figure 6.4 à partir de l’expression “E”

Trois expressions de référence différentes de *F*

Les expressions utilisées pour la reconstitution de celle de la figure 6.8 sont *E* (son “i”), *br-frown*, *br-sad*. Le résultat à la figure 6.10 montre l’expression désirée n’est pas obtenue par reconstitution. Cela signifie qu’il est impossible de reconstituer l’expression désirée, en l’occurrence l’expression *F*, à partir des trois expressions citées ci-dessus.

Quatre expressions de référence différentes de *F*

Pour ce cas-ci, quatre expressions sont utilisées: *E* (son “i”), *br-frown*, *br-sad*, *jaw-down*. Une légère amélioration est observée ici (figure 6.11). Cela est dû au fait que l’expression *jaw-down* produit des



Expression à composer Expression reconstituée

FIG. 6.10 - Reconstitution de l'expression "F" de la banque de référence de la figure 6.4 à partir de trois expressions

déplacements des points intérieurs des lèvres qui sont nécessaires à l'obtention de l'expression désirée F .



Expression à composer Expression reconstituée

FIG. 6.11 - Reconstitution de l'expression "F" de la banque de référence de la figure 6.4 à partir de quatre expressions

Toutes les expressions de référence différentes de F

Est-il possible de reconstituer l'expression F à partir des autres expressions de la banque de référence de la figure 6.4?

Pour répondre à cette question, le présent test a été fait en utilisant toutes les expressions à l'exception de F . Le résultat de la figure 6.12 montre que la reconstitution n'est toujours pas complètement exacte. C'est donc que l'expression F est nécessaire à la banque de référence.



Expression à composer Expression reconstituée

FIG. 6.12 - Reconstitution de l'expression "F" de la banque de référence de la figure 6.4 à partir de toutes les expressions exceptée "F"

Toutes les expressions de référence

La figure 6.13 montre qu'il est possible de retrouver l'expression donnée F de la banque de référence.



Expression à composer Expression reconstituée

FIG. 6.13 - Reconstitution de l'expression "F" de la banque de référence de la figure 6.4 à partir de toutes les expressions

Les valeurs de $[\lambda]$ pour l'expression reconstruite à la figure 6.13 sont données ci-dessous.

$$[\lambda] = \begin{bmatrix} 0.0, & 0.9, & 0.0, & 0.0, \\ -2.11 \times 10^{-2}, & 2.93 \times 10^{-2}, & 0.0, & 0.0, \\ 4.64 \times 10^{-2}, & 9.83 \times 10^{-3}, & 4.80 \times 10^{-2}, & 2.04 \times 10^{-3}, \\ 2.73 \times 10^{-4}, & 0.0, & 0.0, & 0.0, \\ 8.14 \times 10^{-4} \end{bmatrix}$$

Recomposition d'autres expressions de la banque

Les mêmes tests que précédemment ont été effectués pour d'autres expressions de la banque de référence et les mêmes conclusions peuvent être tirées concernant la possibilité ou non de les reconstituer. Les reconstitutions des expressions *E* (son "i") et *kiss* (figure 6.14) sont présentées ci-dessous à titre d'exemple.

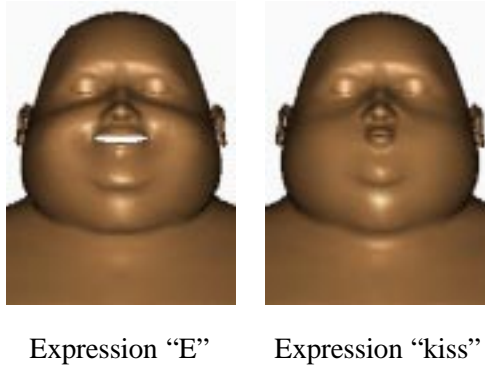


FIG. 6.14 - *Expressions de la banque à reconstituer*

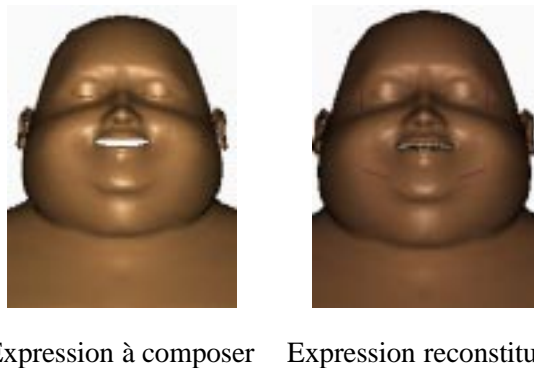


FIG. 6.15 - *Reconstitution de l'expression "E" de la banque de référence de la figure 6.4*

Plusieurs autres tests ont été effectués relativement à la validation de la méthode de résolution utilisée et à la consistance de la banque (i.e. retrouver une expression à partir de ses composantes indépendantes, influence du mouvement d'un marqueur virtuel sur le résultat de la synthèse, etc.) mais ne sont pas présentés ici pour des raisons de limite d'espace.



Expression à composer Expression reconstituée

FIG. 6.16 - Reconstitution de l'expression "kiss" de la banque de référence de la figure 6.4

6.1.4 Correspondance avec le son

Par la suite, la validation subséquente consiste à voir à ce que la séquence de postures faciales faite par un acteur réel concorde avec celles synthétisées par le système pour le personnage synthétique d'une part, mais aussi, que le tout corresponde à la trame sonore enregistrée lors de la capture des postures faciales. L'évaluation de la correspondance des expressions faciales avec le son est faite de manière qualitative, par visionnement de l'animation générée. Les résultats des tests de synchronisation peuvent être observés en visionnant les courtes séquences d'animation disponibles aux adresses suivantes:

<http://www.iro.umontreal.ca/labs/infographie/theses/kouadio/slim-labiales.mov>,

<http://www.iro.umontreal.ca/labs/infographie/theses/kouadio/fats-labiales.mov>,

<http://www.iro.umontreal.ca/labs/infographie/theses/kouadio/slim-phrase.mov>,

<http://www.iro.umontreal.ca/labs/infographie/theses/kouadio/fats-phrase.mov>

ou sur le vidéo:

Phonèmes,

Différents personnages,

Même séquence de capture,

Influence des marqueurs,

Importance des marqueurs.

6.2 Ajout des accessoires

L'approche utilisée qui consiste à synthétiser une expression à partir d'une banque d'expressions élémentaires de référence est poussée plus loin de manière à y ajouter plus de flexibilité. Les coordonnées d'expressions obtenues par résolution du système d'équations sont utilisées pour animer des accessoires de toute sorte qui peuvent être ajoutés à la face du personnage synthétique. Comme exemple d'accessoires, il est possible d'ajouter les dents dans la bouche, la langue, un cigare entre les lèvres, des lunettes, etc. La figure 6.17 présente deux accessoires pouvant être ajouté à la face.

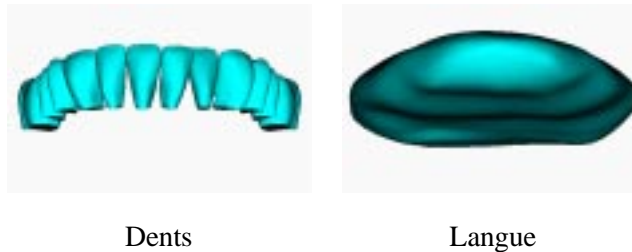


FIG. 6.17 - Exemple d'accessoires



FIG. 6.18 - Accessoires (dents,bouche,langue) ajoutés au visage

Le traitement des accessoires est similaire à celui de la face. En effet, pour chaque accessoire ajouté à la face, il existe une banque des postures de référence de ce dernier, qui permet de définir un système de coordonnées pour l'accessoire en question (figure 6.2). Toutefois, les référentiels des accessoires sont reliés au référentiel de la face car les coordonnées de ceux-ci sont déduites de celles de la face. L'approche adoptée pour établir le lien entre les référentiels des accessoires et celui de la face est voulue la plus flexible possible pour réduire le temps de calcul lors de la synthèse de l'expression finale.

Pour un accessoire donné, chaque posture de référence de la banque associée est reliée à au moins une posture de référence de la face (figure 6.20); mais chaque posture de référence de la face est liée à au

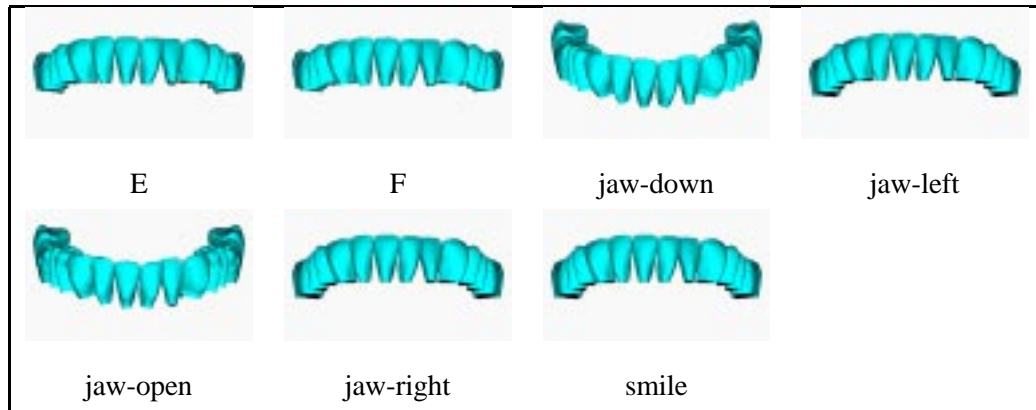


FIG. 6.19 - Banque de postures de référence des dents

plus une posture de référence de l'accessoire en question. La relation définie du référentiel de l'accessoire vers le référentiel de la face est une *fonction surjective*. Ce type de lien a été adopté pour éviter d'avoir dans la banque de référence de l'accessoire des postures qui se répètent car la posture de l'accessoire ne change pas nécessairement pour différentes postures de la face.

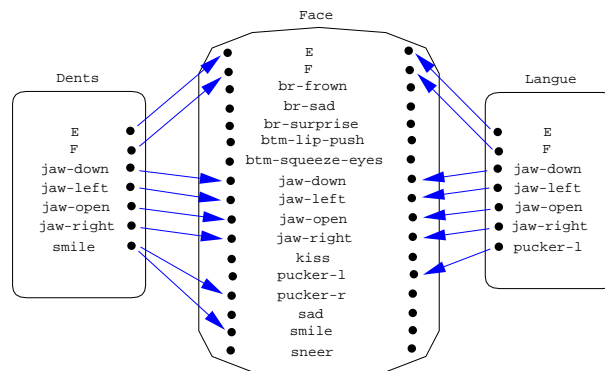


FIG. 6.20 - Définition des liens entre les postures de référence d'accessoires et de la face

Les accessoires n'étant pris en compte que lors de la synthèse finale, il est nécessaire de déterminer les coordonnées de chaque accessoire relativement à son référentiel à partir des coordonnées de la face. La relation (*fonction surjective*) définie entre la banque de référence de la face et celle de chaque accessoire prend toute son importance.

Soient les banques suivantes:

n_ξ références pour la *face*: $([\xi_1], [\xi_2], \dots, [\xi_{n_\xi}])$

n_δ références pour les *dents*: $([\delta_1], [\delta_2], \dots, [\delta_{n_\delta}])$

n_β références pour la *langue*: $([\beta_1], [\beta_2], \dots, [\beta_{n_\beta}])$

Soit $[\lambda_\xi]$ les coordonnées de l'expression faciale obtenues après le processus de résolution du système d'équations. La détermination des coordonnées des dents (accessoire δ) est faite en tenant compte de la *fonction surjective* comme suit:

Soit $g : \{[\xi]\} \rightarrow \{\lambda_\xi\}$ défini par $g([\xi_i]) = \lambda_{\xi_i}$, la fonction qui à chaque expression élémentaire de la face fait correspondre la coordonnée calculée.

Soit $\varphi : \{[\delta]\} \rightarrow \{[\xi]\}$ la fonction surjective définissant le lien entre la banque des dents et celle de la face (i.e. $\varphi([\delta_1]) = \{[\xi_1]\}^1$).

La coordonnée i de l'accessoire δ est déterminée par:

$$\lambda_{\delta_i} = \sum_{\varphi([\delta_i])} g([\xi] : [\xi] \in \varphi([\delta_i])).$$

Par exemple, la coordonnée *smile* de l'accessoire **dent** est obtenue en faisant la somme des coordonnées *pucker-r* et *smile* de la face, tandis que la coordonnée E de l'accessoire **langue** est la même que la coordonnée E de la face. La synthèse de la posture finale de chaque accessoire est faite comme dans le cas de la face à partir de la position neutre. L'équation de synthèse pour les dents sera donc:

$$[\delta] = [\delta_0] + \sum_{i=1}^{n_\delta} \lambda_{\delta_i} [\vec{\delta}_i] \quad (6.1)$$

Les résultats des tests de synthèse des accessoires peuvent être observés à l'adresse

<http://www.iro.umontreal.ca/labs/infographie/theses/kouadio/df-consist.mov>

et sur le vidéo à la séquence intitulée *Filtrage des valeurs de $[\lambda]$: Filtre passe-bas*, qui illustre l'effet du filtrage (présentée au chapitre suivant) mais aussi la synthèse des accessoires.

Le test ultime effectué a consisté à communiquer en temps réel les postures faciales de l'acteur réel au personnage synthétique. Il faut noter que rendu à cette étape, la vitesse de résolution du système d'équations et celle de la synthèse des expressions prennent toute leur importance. La taille de la base de données (nombre de points par expression de référence et nombre d'expressions de la banque) doit être réduite en fonction de la puissance de la machine utilisée et des performances désirées. Pour effectuer le test temps réel sur une **ONYX 200 Mhz** pouvant afficher 10,000 polygones par seconde, le système d'animation

1. $\varphi([\delta_1]) = \{[\xi_1]\}$ signifie que l'image de $[\delta_1]$ par la fonction φ est l'ensemble $\{[\xi_1]\}$ qui comporte un seul élément relativement aux liens de la figure 6.20.

faciale utilise une trentaine de marqueurs (sur le visage de l'acteur), une trentaine de points liés sur la face synthétique et une quinzaine d'expressions de référence.

Les résultats de ce test peuvent être observés sur le vidéo accompagnant le mémoire, à la séquence intitulée *Système de capture*.

Le jugement de la qualité des tests et résultats présentés dans ce chapitre est laissé à la discrétion du lecteur. Cependant, force est de constater que les animations sont de qualité comparable à ce qu'il est possible de générer à partir des systèmes d'animation existants. Le chapitre qui suit donne un aperçu de quelques idées utilisées pour apporter des améliorations aux résultats de la résolution numérique, à la séquence de postures faciales capturées sur l'acteur réel et à la synthèse de l'expression finale.

Chapitre 7

Amélioration des résultats

Les résultats présentés au chapitre précédent sont d'un niveau de réalisme comparable aux meilleurs animations réalisables actuellement. Toutefois, le système laisse encore place à amélioration. En effet, une fois la séquence des postures faciales capturées, il est nécessaire de pouvoir agir sur celle-ci pour corriger les erreurs s'il y en a (de manière à assurer des expressions de synthèse plus précises) ou pour donner un aspect particulier à l'animation du personnage synthétique. En outre, une possibilité d'intervention de l'artiste sur les résultats de la résolution numérique pourrait porter les résultats de l'animation finale au niveau de réalisme qu'il souhaite. D'autres aspects de la face tels que les yeux et les paupières ne sont pas pris en compte dans les calculs introduits jusqu'à présent. Une manière d'influencer la séquence des postures faciales mais également celle des coordonnées d'expressions est le filtrage.

7.1 Filtrage

7.1.1 Filtrage des résultats de capture

Les résultats obtenus à partir des combinaisons linéaires des expressions de la banque de référence sont assez acceptables. Toutefois, il faut avoir souvenance que les déplacements des marqueurs qui sont communiqués au personnage synthétique proviennent de capture de postures faciales sur un acteur réel. Il ne faut donc pas exclure la possibilité de déplacements erronés dus à une erreur de capture ou à la sensibilité du dispositif de capture. Pour pallier à ce problème, un filtrage des résultats de la capture est effectué pour adoucir un peu les variations des déplacements.

Le filtre appliqué à la séquence des déplacements de marqueurs est un filtre passe-bas qui permet ainsi

d'éliminer les variations trop brusques des valeurs des composantes X , Y et Z (hautes fréquences).

La forme du filtre dans le domaine fréquentiel (figure 7.1) est donnée par l'équation 7.1.

$$H(f) = \begin{cases} 1.0 & \text{si } f \in \left[0.0, f_c - \frac{\Delta f}{2}\right] \\ \frac{\sin\left(\frac{(f_c-f)\pi}{\Delta f}\right)+1}{2} & \text{si } f \in \left[f_c - \frac{\Delta f}{2}, f_c + \frac{\Delta f}{2}\right] \\ 0.0 & \text{si } f \in \left[f_c + \frac{\Delta f}{2}, +\infty\right) \end{cases} \quad (7.1)$$

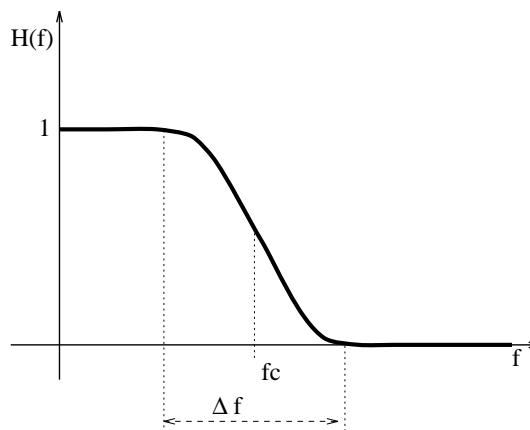


FIG. 7.1 - *Filtre passe-bas (équation 7.1)*

Pour obtenir les valeurs filtrées de la séquence d'entrée, le processus de filtrage suit les étapes ci-dessous pour chacune des dimensions X, Y et Z :

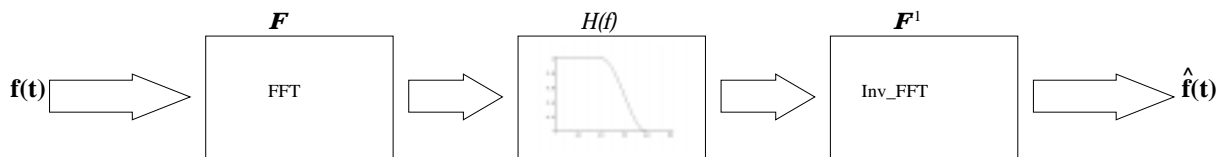
1. la transformée de Fourier de la séquence de déplacements capturés est calculée
2. le résultat de la transformée de Fourier est multipliée par le filtre (filtre dans le domaine fréquentiel ou la transformée de Fourier du filtre)
3. la transformée de Fourier inverse du résultat de la multiplication est calculée pour revenir dans le domaine temporel.

Le synopsis de la figure 7.2 présente les différentes étapes du processus de filtrage.

La courte séquence d'animation

<http://www.iro.umontreal.ca/labs/infographie/theses/kouadio/filter-input.mov>

(ou sur le vidéo *Filtrage de la séquence capturée: Filtre passe-bas*) permet de constater les changements apportés par le filtrage de la capture.

FIG. 7.2 - *Processus de filtrage*

7.1.2 Filtrage des résultats de la minimisation

La résolution du système d'équations, compte tenu de la méthode choisie, est effectuée par itérations successives à partir de valeurs initiales qui sont toutes à zéro. Cela fait que les liens éventuels qui devraient exister entre deux postures faciales successives de l'acteur réel ne sont pas pris en compte. Les coordonnées résultant de la résolution peuvent ainsi refléter des variations indésirables de la séquence de capture. En outre, le système d'équations obtenu par modélisation mathématique du problème ne tient compte que des points liés (le nombre de points liés est de beaucoup inférieur au nombre de points de la face). Le résultat de la synthèse de l'expression finale pourrait ne pas être très esthétique visuellement, vu que les points libres de la face peuvent avoir des déplacements éventuellement nuisibles.

Les coordonnées d'expressions calculées peuvent donc être filtrées pour atténuer l'effet des variations brusques qui pourraient survenir dans la séquence des résultats de calcul.

Filtrage de l'ensemble de coordonnées

Un filtrage similaire à celui de la séquence des résultats de capture (figure 7.2) peut être effectué sur l'ensemble des résultats des résolutions des systèmes d'équations, i.e. sur l'ensemble des $[\lambda]$ calculées à partir de la séquence de capture.

L'animation illustrant l'influence de ce type de filtrage est présentée par la courte séquence

<http://www.iro.umontreal.ca/labs/infographie/theses/kouadio/filter-output.mov>

et sur le vidéo par *Filtrage des valeurs de $[\lambda]$: Filtre passe-bas.*

Il faut noter que plus le filtre (passe-bas) utilisé est large dans le domaine temporel (étroit dans le domaine fréquentiel) plus les expressions sont amenuisées. La séquence d'animation résultante est donc plus monotone.

Le filtrage utilisé pour les animations présentées dans cette section étant un filtrage passe-bas, les résultats sont plus continus (plus lisses). En effet, les hautes fréquences correspondant aux variations

brusques sont éliminées. Cela se traduit par une élimination des petites erreurs de position de marqueurs pour ce qui est de la séquence d'entrée et par une atténuation des changements brusques dans la combinaison des expressions entrant en jeu dans la synthèse à un instant donné.

Toutefois, ce type de filtrage ne peut pas être envisagé dans un contexte où les résultats du système d'équations sont utilisés au fur et à mesure qu'ils sont calculés, ce qui est le cas lorsque l'*animateur* est utilisé en temps réel. Dans cette dernière situation, il faut envisager un filtrage temporel prédictif.

7.2 Traitement des paupières

Les paupières peuvent se déplacer de manière indépendante relativement au reste du visage. L'utilisation des déplacements des marqueurs situés sur les paupières dans le processus de résolution du système d'équations nécessiterait donc l'ajout à la banque des expressions de la face, des expressions particulières où seules les paupières bougent. Cela aurait pour effet d'alourdir la banque d'expressions de la face et donc de réduire l'efficacité (rapidité) de la résolution du système d'équations. En outre, le déplacement des paupières se faisant sur la sphère représentant le globe oculaire, la nécessité de tenir compte de la troisième composante (z) des déplacements des marqueurs de paupières se fait sentir. Une meilleure interpolation de cette dernière devient incontournable.

Cependant, une observation du déplacement des paupières sur un visage permet de constater que celles-ci ont un mouvement qui est pour la plupart du temps vertical pour ne pas dire toujours vertical. Cette caractéristique des paupières est exploitée pour les traiter d'une manière particulière.

L'approche retenue pour les mouvements des paupières est de les traiter en dehors de la résolution numérique. Une banque de positions de référence est ainsi associée aux paupières. Elle contient des positions clé des paupières supérieure et inférieure de chaque oeil auxquelles sont associés des facteurs d'ouverture.

Pour communiquer les déplacements des marqueurs des paupières de l'acteur réel au personnage synthétique, un facteur d'ouverture correspondant au déplacement de chaque marqueur (de paupière) est déterminé. La position à donner à la paupière concernée est alors interpolée dans la banque de référence qui lui est associée compte tenu de son facteur d'ouverture.

Le facteur d'ouverture pour une paupière donnée en fonction de sa banque de référence associée et du déplacement du marqueur est déterminé par le processus suivant.

Une inspection de la banque des positions de la paupière concernée est effectuée, ce qui permet de

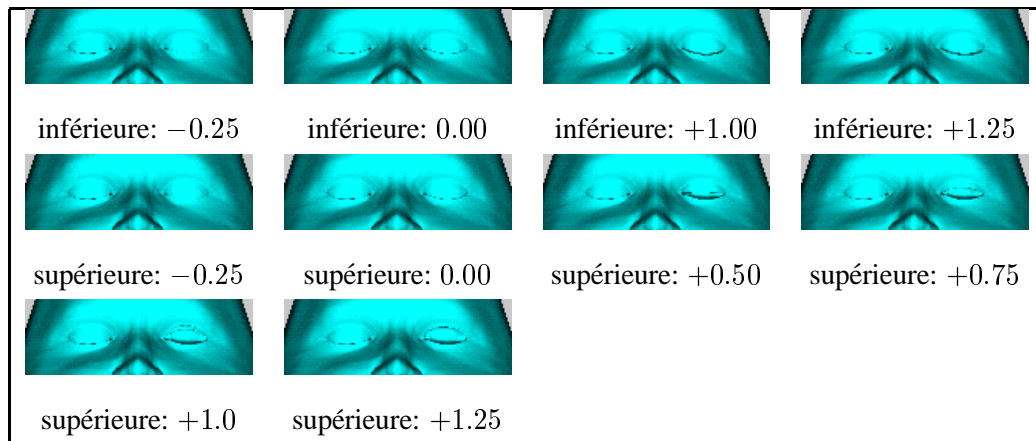


FIG. 7.3 - Exemple de banque de paupières (oeil gauche du point de vue du personnage)

déduire les valeurs maximale et minimale de facteur d'ouverture (f_{max} et f_{min}) ainsi que les déplacements maximal et minimal ($y_{f_{max}}$ et $y_{f_{min}}$) du point de la paupière qui est associé au marqueur placé sur la paupière correspondante de l'acteur réel. Les postures extrêmes effectuées par l'acteur au début de la séance de capture permet d'avoir les déplacements extrêmes des marqueurs de paupière ($y_{m_{max}}$ et $y_{m_{min}}$). À partir de ces valeurs extrêmes, une correspondance est établie entre les variations de chacun de ces paramètres (figures 7.4 et 7.5).

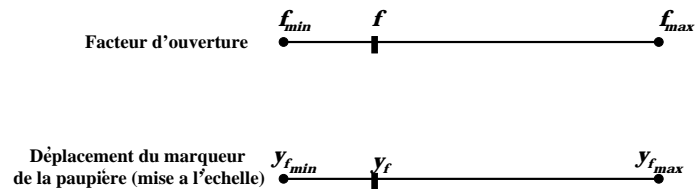


FIG. 7.4 - Correspondance entre variations des paramètres des paupières supérieures

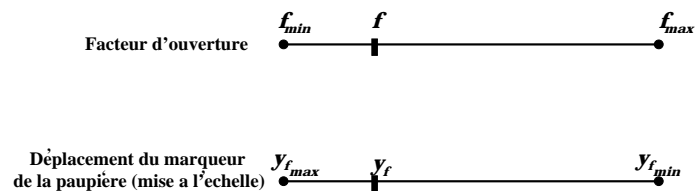


FIG. 7.5 - Correspondance entre variations des paramètres des paupières inférieures

Pour un déplacement de marqueur donné, le déplacement du point de la paupière du personnage associé est calculé via les différents niveaux d'échelles présentés plus haut. Le facteur d'ouverture des pau-

pières supérieure et inférieure correspondant est calculé suivant les équations 7.2 et 7.3 respectivement.

$$f = f_{min} + \frac{(f_{max} - f_{min})(Y - Y_{min})}{Y_{max} - Y_{min}} \quad (7.2)$$

$$f = f_{min} + \frac{(f_{max} - f_{min})(Y - Y_{max})}{Y_{min} - Y_{max}} \quad (7.3)$$

L'animation des paupières ainsi que l'ajout des accessoires ont un apport non négligeable à l'animation de la face. Cela peut s'observer en visionnant côte à côte les animations

<http://www.iro.umontreal.ca/labs/infographie/theses/kouadio/fats-phrase.mov>

et

<http://www.iro.umontreal.ca/labs/infographie/theses/kouadio/df-consist.mov>

ou en visionnant successivement les séquences de vidéo intitulées *Importance des marqueurs* et *Filtrage des valeurs de $[\lambda]$: Filtre passe-bas*. Toutefois, il est possible d'apporter d'autres améliorations au système d'animation faciale. Quelques unes de ces améliorations sont suggérées au chapitre suivant.

Chapitre 8

Améliorations potentielles

L'*animateur* développé dans ce projet permet d'obtenir des animations faciales qui peuvent être jugées bonnes, voir même très bonnes dépendant du talent de l'artiste qui l'utilise. Le lecteur peut en juger par lui-même relativement à la qualité des animations produites. Toutefois, il y a certains aspects qui ouvrent des voies à l'amélioration. Notamment les yeux (et non les paupières), la synchronisation avec la voix, l'interface des actions de l'artiste et la synthèse de l'expression finale.

8.1 Améliorations à court terme

Les améliorations qui suivent peuvent être apportées à l'*animateur* développé sans effort d'implantation majeur.

8.1.1 Mouvement des yeux

L'outil d'animation faciale (*animateur*) tel qu'il existe actuellement ne tient pas compte du mouvement des yeux. Seuls les mouvements de paupières sont pris en compte et incorporés dans le traitement des données de capture. Il serait possible de capturer le mouvement des yeux (mouvements des pupilles) et de l'inclure dans le traitement global de la face, et cela sans un coût supplémentaire important du point de vue de l'implantation. En effet, la caméra servant à capturer les mouvements des marqueurs transmet également une image du visage de l'acteur. Un processus de segmentation (détection de contours et de régions) de cette image pourrait donner la position exacte des pupilles à un instant donné, vu que les marqueurs des paupières sont repérés et que la morphologie d'un visage est assez standard à travers les espèces (les yeux sont au-dessus du nez qui lui est au-dessus de la bouche). De plus seulement une très

petite région de l'image aurait à être traitée. Une autre façon de capturer la position des yeux serait le port par l'acteur de lentilles réfléchissantes. Cela permettrait de capturer les mouvements des yeux en même temps que celui du reste du visage et de les animer par la même occasion avec celui-ci.

Il serait aussi possible d'envisager une utilisation en temps réel de l'animation des yeux avec cette approche de capture. L'inconvénient majeur dans ce cas-ci est dû au fait que les yeux constituent des organes sensibles et que toute manipulation les concernant nécessite beaucoup de soins. Ainsi, à chaque acteur réel, il faudrait des lentilles différentes et leur durée de vie serait assez courte (remplacements fréquents) pour ne pas causer de tords à l'acteur.

8.1.2 Capture de mouvement en 3D

La capture des postures faciales s'effectue actuellement en 2D. Par conséquent la reconstitution des expressions dont les mouvements prédominants sont dans la direction Z (figure 8.1) sont difficiles, même avec l'approximation de la troisième composante. Ce qui est fait pour le moment pour outrepasser cette difficulté est la modélisation d'expressions particulières avec des prédominances de déplacements en Z , mais cela alourdit la banque d'expressions de deux ou trois expressions supplémentaires. En utilisant un système de capture en 3D, les déplacements en Z seraient introduits dans la résolution du système d'équations au même titre que ceux en X et Y . Les positions des points de l'expression de synthèse finale seraient plus précises.



FIG. 8.1 - Expression à prédominance en Z

8.2 Améliorations nécessitant un effort d'implantation

8.2.1 Automatisation

L'artiste utilisant l'*animateur* a un certain niveau de contrôle sur les résultats produits par celui-ci (empreinte de l'artiste). Une question qui revient très souvent lorsqu'il faut concevoir un système est de déterminer quel niveau de contrôle donner à l'utilisateur sans trop alourdir sa tâche. Pour ce qui est de l'*animateur live-face* en question ici, l'utilisateur a le contrôle aux niveaux suivants:

- la banque d'expressions (modélisation et choix des expressions - choix des expressions à subdiviser),
- le déplacement des points de la face (amplification ou atténuation des mouvements des marqueurs réels),
- l'association des marqueurs réels et des points du modèle de la face (port du masque par la face).

Certains de ces contrôles pourraient être automatisés (en partie ou en totalité), non pas pour rendre l'outil nécessairement plus précis, mais pour alléger la tâche de l'artiste d'une part et pour effectuer des choix de paramètres sur une base plus quantitative que qualitative d'autre part. Cela permettrait ainsi une utilisation plus efficace de l'*animateur*.

Génération de la banque d'expressions

La banque d'expressions pourrait être générée automatiquement à partir du modèle de la face, des associations de marqueurs réels et points du modèle de la face, des unités d'actions, en offrant tout de même la possibilité d'effectuer quelques retouches par l'artiste. Cette approche bénéficierait des outils déjà existants dans le système de *TAARNA*, dont l'*animateur map-expression* qui permet de reproduire l'expression d'une face donnée sur une autre de morphologie pouvant être totalement différente.

Valeurs admissibles de coordonnées d'expressions

Le choix des valeurs limites admissibles pour les coordonnées d'expression pourrait également être fait automatiquement. En effet, compte tenu du fait que l'acteur réel, avant de commencer une séance de capture, effectue ces postures extrêmes qui permettent d'aller chercher les déplacements maximum

et minimum de chaque marqueur, il serait possible, par inspection de la banque d'expressions, de déterminer des valeurs maximales et minimales pour les coordonnées de chaque expression de la banque. Il faudrait, pour ce faire, construire des ensembles de valeurs maximales et minimales admissibles de coordonnées d'expression pour chaque marqueur et pour chaque expression de la banque. Par la suite il suffirait de déterminer les valeurs extrêmes globales à partir de ces ensembles.

Un aspect également automatisable du contrôle de l'usager est le choix des expressions de la banque qu'il est judicieux de subdiviser. En effet, une inspection de chaque expression de la banque permettrait de déterminer si elle est judicieusement décomposable en composantes indépendantes ou non. Cela nécessiterait des informations supplémentaires dans la modélisation de la face. Les unités d'actions pourraient être utilisées pour la détermination de la décomposabilité d'une expression en composantes indépendantes.

8.2.2 Coordonnées d'expression et expression de synthèse

Couche de tissus

L'expression de synthèse finale est reconstruite à partir uniquement des points de la surface de la face. Une amélioration de ce processus consisterait à utiliser une couche de tissus en-dessous de la surface, qui serait déformée au lieu de la surface du visage. Par la suite la surface serait relaxée par-dessus la couche de tissus. En outre, le système d'animation de *TAARNA* permet d'ajouter des tissus et des muscles à un modèle et de les animer grâce à l'*animateur layers-dynamics*. Les tissus seraient ainsi animés en fonction des coordonnées d'expression et des déplacements de marqueurs.

Expressions antagonistes

Le processus de détermination des coordonnées d'expression pourrait être modifié de manière à ce que le résultat de la résolution du système d'équations ne comporte pas d'inconsistance relativement aux composantes d'expressions antagonistes¹. En effet, il serait possible de spécifier des relations entre les expressions deux à deux, et tenir compte de ces relations lors de la résolution du système d'équations de sorte que deux expressions élémentaires antagonistes n'entrent pas dans la synthèse d'une même expression finale.

1. Deux expressions telles que l'une déplace la mâchoire inférieure à gauche et l'autre la déplace vers la droite, sont considérées comme étant antagonistes

Poids des expressions

Certaines expressions de la banque pourraient avoir plus d'importance que d'autres lors de la détermination de ces coordonnées. Cela permettrait à l'artiste de faire prédominer ou non une expression particulière dans le résultat.

Résolution indépendante par région

La résolution du système pourrait également être faite par morceaux. Les équations dans chaque morceau représenteraient une zone indépendante du visage. Cela aura pour effet de donner une précision locale très efficace pour les positions de points de l'expression de synthèse. Cependant, il pourrait se poser le problème de continuité d'une zone indépendante à l'autre. La discontinuité peut être amoindrie en incluant certains points en bordure des régions dans plusieurs régions à la fois et en factorisant la contribution de chaque région dans le déplacement de ces points.

8.2.3 Déplacement hiérarchique des points liés

Les points liés à déplacer dans la face pour spécifier l'expression désirée (pour la résolution du système d'équations) sont définis initialement lors de l'association des marqueurs. Le niveau de détail de l'expression désirée est donc établi une fois pour toute. La possibilité de pouvoir spécifier plusieurs niveaux de détails pour l'expression désirée serait un autre moyen de contrôler les résultats des calculs. Le premier niveau de déplacement grossier serait celui des marqueurs virtuels (les points correspondant aux marqueurs réels). Par la suite, suivant le niveau, plus de points de la face seraient déplacés pour mieux contraindre la résolution du système d'équations. Les déplacements des points pour différents niveaux seraient similaires à ceux effectués dans [4] pour la détection de collisions.

8.2.4 Synchronisation du son

Lorsque le système n'est pas utilisé en temps réel, l'information sonore capturée pendant la séance d'enregistrement des postures faciales n'est utilisée qu'à la fin du processus, lors de la génération de l'animation finale. Les erreurs survenues du côté synchronisation des enregistrements ne sont découvertes qu'à ce moment-là. Le seul moyen de les corriger est en jouant sur les enregistrements de mouvements spatiaux de la face, ce qui est très fastidieux pour l'artiste.

Il serait possible d'analyser la trame sonore pendant (ou après) la résolution numérique du système

d'équations ou pendant la synthèse de l'expression finale et d'ajuster les coordonnées d'expressions en fonction du son à l'instant en question (par exemple ajustement des contours de la bouche).

Chapitre 9

Conclusion

L'outil d'animation faciale développé dans le présent projet est en évolution. L'objectif visé dans le développement de l'*animateur* facial était de permettre l'animation de différents aspects du visage, à partir de spécifications de mouvements faciaux pouvant être capturés sur un acteur réel. Les expressions faciales du personnage synthétique devaient pouvoir être synthétisées en temps réel au fur et à mesure que les mouvements faciaux sont capturés. Par ailleurs, la face du personnage synthétique peut être quelconque (humain, animal ou *cartoon*) et l'animation de celle-ci devait être réaliste et répondre aux attentes d'un observateur extérieur. Cela a guidé le choix de l'approche d'animation utilisée qui est basée sur l'existence d'une bibliothèque d'expressions de référence.

Les spécifications que doit rencontrer un système d'animation faciale telles que présentées antérieurement, ont grandement influencé la conception et le développement qu'a nécessités ce projet.

Les méthodes utilisées pour arriver à satisfaire les objectifs fixés sont basées sur des notions mathématiques. Ainsi, la détermination des poids à donner à chaque expression de référence entrant en compte dans la synthèse de l'expression finale est faite en utilisant les notions de l'algèbre linéaire (espace vectoriel, dépendance linéaire de vecteurs, résolution d'équations linéaires) et de l'analyse numérique (moindres carrés, décomposition de matrices).

Pour valider l'approche adoptée et les modifications apportées à la méthode de base, des tests ont été effectués de manière à vérifier la robustesse et l'exactitude de la modélisation mathématique du problème. En outre, le traitement dynamique de la bibliothèque d'expressions ajoute plus de flexibilité à la méthode; les expressions accessibles par la face ne sont plus limitées à la matrice statique des expressions de la bibliothèque. Le système d'animation au sein duquel l'*animateur* est appelé à évoluer, a permis de faciliter la conduite de ces tests et la réalisation du projet dans son ensemble.

Rendu à ce stade dans l'évolution du projet de maîtrise, il est opportun de se demander si les buts fixés ont été atteints. Fort des tests effectués et des résultats obtenus d'une part, et de la qualité des animations produites par des artistes avec l'*animateur* développé d'autre part, la réponse à la question de l'atteinte des objectifs fixés est affirmative. En effet, l'*animateur* en question est utilisé pour l'animation des expressions faciales dans le film d'animation "*The Boxer*".

Toutefois, il y a encore place à améliorations, pour incorporer l'animation de la totalité des aspects de la face. Par exemple, la prise en compte des mouvements de pupilles ou d'accessoires dont les positions peuvent être indépendantes de celle de la face (boucle d'oreille), ou des effets faciaux impossibles à capturer sur l'acteur réel avec le système de capture utilisé (mouvement de la langue hors de la bouche). Il serait également possible de tenir compte de régions voisines de la face telles que les mouvements au niveau du cou. Les modifications apportées à l'approche d'animation des expressions faciales synthétiques à partir d'une bibliothèque d'expressions permettent de l'utiliser dans un contexte d'animation en temps réel.

Bibliographie

- [1] Norman I. Badler. Animating facial expression. In *Proceedings Computer Graphics*, pages 245–252. ACM SIGGRAPH, 1981.
- [2] Philippe Bergeron and Pierre Lachapelle. Controlling facial expressions and body movements in the computer-generated animated short “tony de peltrie”. In *SIGGRAPH '85 Advanced Computer Animation seminar notes*. July 1985.
- [3] James D. Foley, Andries van Dam, Steven K. Feiner, and John F. Hughes. *Computer Graphics, Principles and Practice, Second Edition*. Addison-Wesley, Reading, Massachusetts, 1990. Overview of research to date.
- [4] Jason Harrison and David Forsey. A kinematic model for collision response. In *Proceedings of Eurographics Workshop on Animation and Simulation '94*, 1994.
- [5] Kaveh Kardan. Tools for the animation of facial expressions using computer graphics. Society of Motion Picture and Television Engineers, Inc., 1992.
- [6] Charles L. Lawson and Richard J. Hanson. *Solving Least Square Problems*. Prentice-Hall, 1974.
- [7] Yuencheng Lee, Demetri Terzopoulos, and Keith Waters. Constructing physics-based facial models of individuals. In *Proceedings of Graphics Interface '93*, pages 1–8, Toronto, Ontario, Canada, May 1993. Canadian Information Processing Society.
- [8] Yuencheng Lee, Demetri Terzopoulos, and Keith Waters. Realistic face modeling for animation. In Robert Cook, editor, *SIGGRAPH 95 Conference Proceedings*, Annual Conference Series, pages 55–62. ACM SIGGRAPH, Addison Wesley, August 1995. held in Los Angeles, California, 06-11 August 1995.

- [9] Ekman P and Friesden W. Manual for the facial action coding system. In *Consulting Psychologist 1977 Press*, 1977.
- [10] F. I. Parke. Parameterized models for facial animation. *IEEE Computer Graphics and Applications*, 2:61–68, November 1982.
- [11] Catherine Pelachaud, Norman I. Badler, and Marie-Luce Viaud. Final report to nsf of the standards for facial animation workshop. 1994.
- [12] William H. Press, Saul A. Teukolsky, William T. Vetterling, and Brian P. Flannery. *Numerical Recipes in C: The Art of Scientific Computing (2nd ed.)*. Cambridge University Press, Cambridge, 1992. ISBN 0-521-43108-5.
- [13] Summerfield Quentin. In *Analysis, Synthesis and Perception of Visible Articulatory Movements*. Academic Press Inc, 1984.
- [14] Taarna Studio. *Taarna System Tutorial*. 1991.
- [15] D. Terzopoulos and K. Waters. Techniques for realistic facial modeling and animation. In Springer-Verlag, editor, *Computer Animation*, pages 59–74, 1991.
- [16] Carol Leon-Yun Wang. Automating facial gestures and synthesized speech in human character animation. In *Proceedings of the 1991 Western Computer Graphics Symposium*, pages 39–40, April 1991.
- [17] Carol Leon-Yun Wang. Current trends in facial animation or langwidere: not just another witch. In *Proceedings of the 1992 Western Computer Graphics Symposium*, pages 103–107, April 1992.
- [18] Carol Leon-Yun Wang. Talking heads:developing pull. In *Proceedings of the fifth Annual Western Computer Graphics Symposium*, pages 114–119, March 1993.
- [19] Carol L.Y. Wang and David R. Forsey. Langwidere:a new facial animation system. In *Proceedings of computer animation'94*, pages 59–68, May 1994.
- [20] Keith Waters. A muscle model for animating three-dimensional facial expression. In Maureen C. Stone, editor, *Computer Graphics (SIGGRAPH '87 Proceedings)*, volume 21, pages 17–24, July 1987.

- [21] Lance Williams. Performance-driven facial animation. In Forest Baskett, editor, *Computer Graphics (SIGGRAPH '90 Proceedings)*, volume 24, pages 235–242, August 1990.