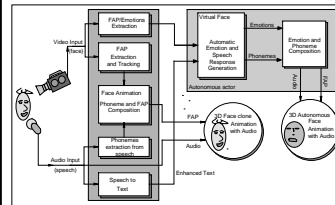


Animation faciale

Animation et Multimedia

Vue d'ensemble ... exemple

➤ Communication entre un visage réel et virtuel



Quelle est la (les) technologie(s)?

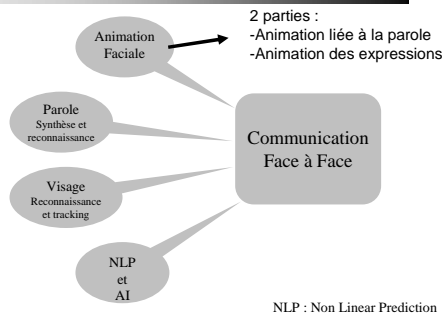
Quels sont les progrès?

Que manque-t-il?

Quel est le futur?

Animation et Multimedia

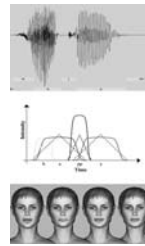
La communication face to face :



NLP : Non Linear Prediction

Animation et Multimedia

Animation de la parole : Hiérarchie



Etapes	Technologie	Méthodes
Temporisation des phonèmes pour la parole (synthétique ou réelle)	Reconnaissance des phonèmes	Manuel, semi-automatique ou automatique
Transition des phonèmes	Coarticulation	Règles de base, automatique
Génération des visèmes et animation	Définition des visèmes Synchronisation avec le son	Automatique

Animation et Multimedia

4 techniques d'animation

1. Posture clef
Keyframing
2. Synthèse vocale animée
Text-to-speech (text to visual)
3. Synchronisation lèvres/audio + expression
Semi-automatic processing
4. Capture de mouvement
MoCap

Animation et Multimedia

1. Posture clef

➤ Animation créée par interpolation de posture ou d'expressions existantes

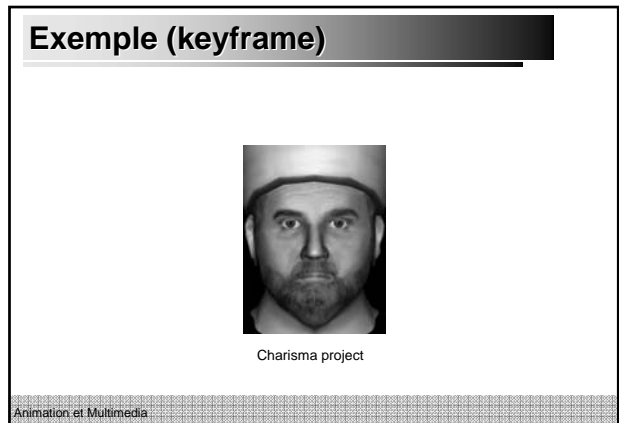
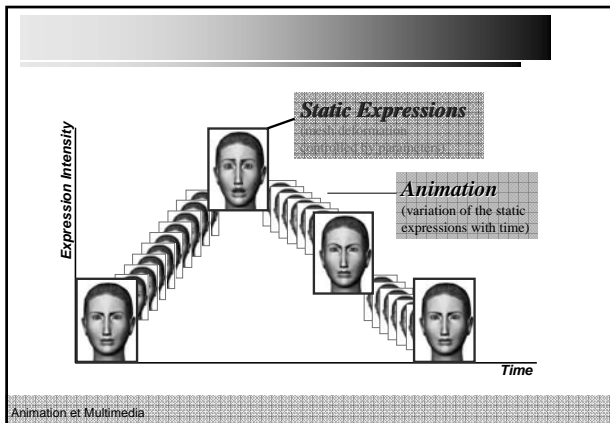
➤ AVANTAGES

- Confère un haut degré de contrôle sur l'animation
- Interpolation des paramètres ... calculs simples

➤ INCONVENIENTS

- Construction de l'animation coûteuse en temps
- Difficile de designer des animations réalistes
- Définir toutes les expressions avant !!

Animation et Multimedia



2. Text-to-speech

➤ L'animation et la voix sont construite automatiquement à partir des phonèmes et des visèmes

Text → Phonème + timing → Visème → Param. animation

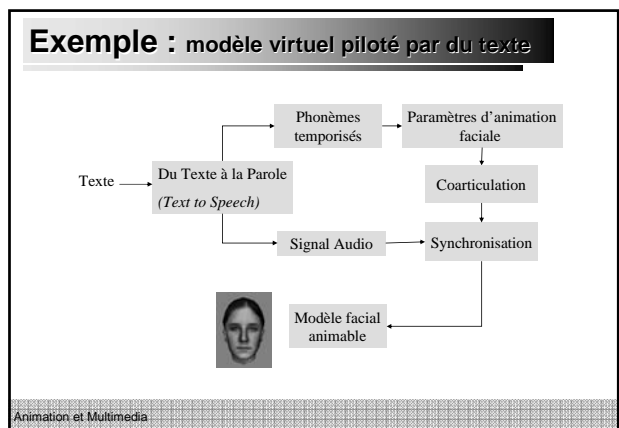
➤ **AVANTAGES**

- Très rapide
- L'animation peut être calculée en temps-réel

➤ **INCONVENIENTS**

- Voix synthétique
- Seulement animation des lèvres

Animation et Multimedia



Coarticulation de la parole

➤ Phénomène observé durant la parole

- Quand l'animation faciale, correspondant à un segment de phonème ou visème, est influencée par les segments environnant

un segment V1-C-V2 ou V1 n'est pas prononcé et V2 est prononcé

Look-Ahead Model
Transition vers le début de V2 dès que V1 est fini

Time-Looked Model
La transition vers V2 commence à un intervalle de temps fixe avant que V2 ne commence

Hybrid Model
La transition vers V2 a lieu en 2 phases

M. M. Cohen, D.W. Massaro, "Modelling coarticulation in synthetic visual speech", in N. M. Thalmann and D. Thalmann, Models and techniques in Computer Animation, Springer-Verlag, 1993, pp. 139-156.

Animation et Multimedia

Modèle d'articulation

➤ Chaque segment de la parole (visème) augmente et diminue avec le temps

➤ Les visèmes adjacents ont des fonctions de recouvrement qui se mélangent au court du temps

➤ Chaque visème peut avoir une fonction différente de variation pour chaque articulation

A. Löfqvist, "Speech as audible gestures", in Speech Production and Speech Modeling, Kluwer Academic Publishers, 299-322

Animation et Multimedia

Coarticulation : définition (for talking head)

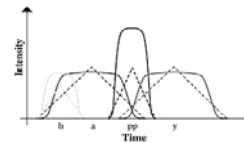
Pelachaud (1991)
« Look ahead model based on deformability of phonemes.
Also considered muscle contraction times »

Cohen & Massaro (1992)
« Non-linear dominance and blending functions designed for each phoneme »

Animation et Multimedia

Coarticulation : résumé

- Définir un poids (dominance) et recouvrement en accord avec les groupes de phonèmes



Animation et Multimedia

3. Semi-automatic processing

- Synchronise l'animation à la voix
... puis ajout des émotions

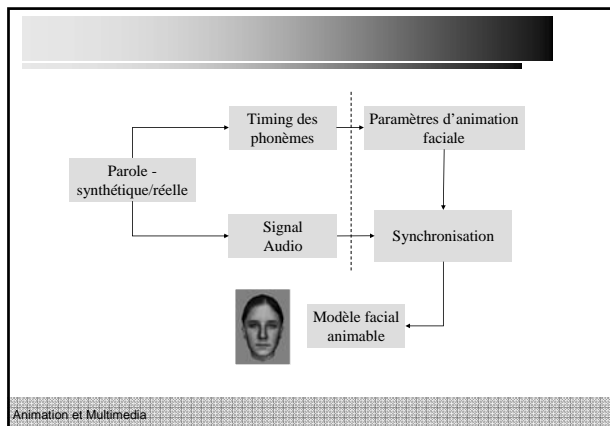
➤ AVANTAGES

- Rapide (quelques minutes pour créer une animation)
- Animation facile à générer
- Large éventail de possibilités

➤ INCONVENIENTS

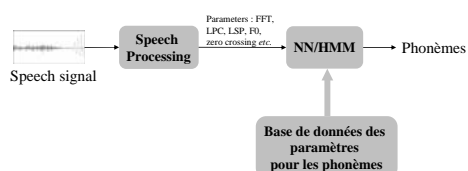
- Pas temps-réel

Animation et Multimedia



Animation et Multimedia

Analyse acoustique

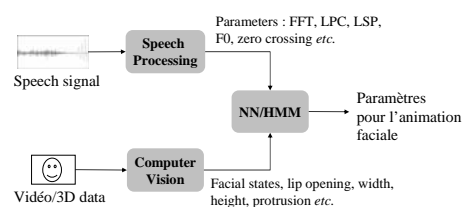


HMM : Hidden Markov Model

NN : Neural Network

Animation et Multimedia

Analyse acoustique + visuelle



Animation et Multimedia

Comparaison

Analyse Acoustique	Analyse Acoustique et Visuelle
Les phonèmes résultant sont utilisables avec n'importe quel système d'animation faciale	Les paramètres de sortie (état du visage/lèvres, ...) sont attachés à un système d'animation particulier
L'animation faciale résultante n'est pas affectée par la base de données d'entraînement	L'animation faciale résultante est étroitement affectée par la base de données d'entraînement
Facilité dans la collecte des données (seulement la parole)	Les données capturées sont synchronisées (parole et vidéo)
Seuls les mouvements des lèvres et de la bouche peuvent être générés	La technique peut être employée pour la synthèse d'autres mouvements faciaux (sourcils, ...)
Un modèle de coarticulation peut être appliqué sur les phonèmes du résultat	La coarticulation fait partie de l'analyse
Grande dépendance de la langue	Dépend moins de la langue

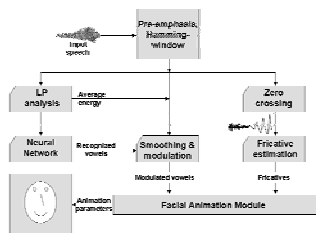
Animation et Multimedia

Les défis ...

- Indépendance de la langue et de l'utilisateur
- Indépendance du modèle faciale utilisé pour l'animation
- Requiert un entraînement minimale de le part de l'utilisateur du système
- Outils et algorithmes simples à implémenter

Animation et Multimedia

Tête parlante animée par la voix



Animation et Multimedia

Analyse de la parole

Extraction des paramètres

Choix de l'analyse Linear Prediction (LP)

« LP derived reflection coefficients are directly related to vocal tract shape » [Wakita]

Les phonèmes peuvent être caractérisés par leur forme

Limitations

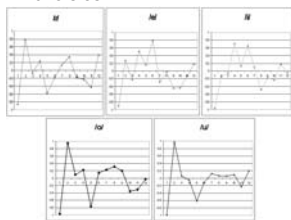
Marche bien pour les voyelles, /a/, /e/, /i/, /o/, /u/.

Pour les consonnes ?

Animation et Multimedia

Utilisation d'un réseau de neurones – Neural Network - NN

- Répartition typique des coefficients pour les cinq voyelles choisies



Propagation sur les 3 couches précédentes

12 points d'entrée

10 points cachés

5 points de sortie

Cinq voyelles utilisées

/a/, /e/, /i/, /o/, /u/

12 speakers masculins

5 féminins

Animation et Multimedia

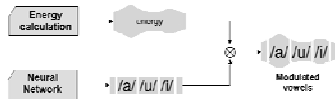
Résultats de l'entraînement du NN

		Reconnu				
		/a/	/e/	/i/	/o/	/u/
Prévu	/a/	241	2	15	11	0
	/e/	0	177	89	0	5
	/i/	0	3	301	0	2
	/o/	10	0	0	224	36
	/u/	4	12	0	88	143

Animation et Multimedia

Analyse de l'énergie

- Transition voyelle – voyelle
- Demi voyelle
- consonnes



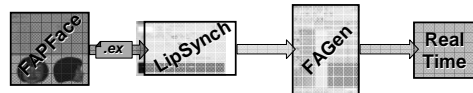
Les paramètres correspondant aux voyelles sont modulés

Animation et Multimedia

3.B Design d'expression

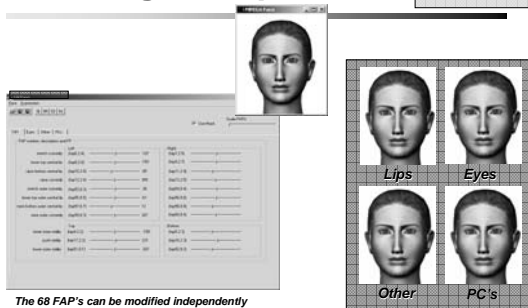
FAPFace

- software to design Static expressions :
-> Phonemes and emotions design
- Output Format : .ex files



Animation et Multimedia

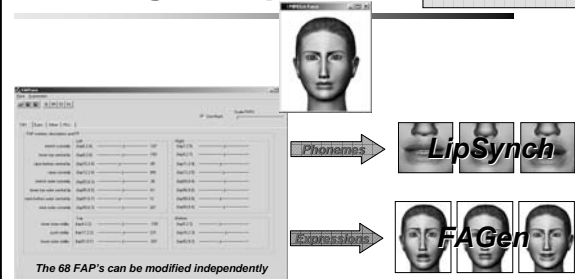
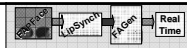
IIIa. Design d'expression



The 68 FAP's can be modified independently

Animation et Multimedia

IIIa. Design d'expression



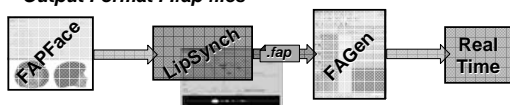
The 68 FAP's can be modified independently

Animation et Multimedia

IIIb. Speech Synchronisation

LipSynch

- software to build Speech animation files using real audio recording
- can be used in any language by just building new phonemes
- Output Format : .fap files

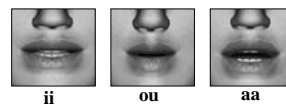


Animation et Multimedia

IIIb. Speech Synchronisation



Phoneme : Smallest speech element



Viseme : Visual representation of a phoneme



Animation et Multimedia

IIIb. Speech Synchronisation

The screenshot shows the FAPFace software interface. At the top, there are three small windows showing different facial expressions. Below them is a 'Phonemes' track with a grid of phonemes. A 'Sound' track is at the bottom. A 'FAPFace' window on the right shows a 3D face model. A 'FAGen' window is also visible, showing a 3D face model.

Animation et Multimedia

IIIc. Emotion Editor

FAGen

- software to build facial expression and emotions
- 4 different layers: speech, emotions, head movements and eyes movements
- Output Format : .fap files

```

graph LR
    FAPFace[FAPFace] --> LipSynch[LipSynch]
    LipSynch --> FAGen[FAGen]
    FAGen -- ".fap" --> RealTime[Real Time]
  
```

Animation et Multimedia

IIIc. Emotion Editor

The screenshot shows the FAPFace software interface. On the left, there is a 'LipSynch' window. The main interface has 'speech' and 'Sound' tracks. A 'FAPFace' window on the right shows a 3D face model.

Animation et Multimedia

IIIc. Emotion Editor

The screenshot shows the FAPFace software interface. On the left, there is a 'LipSynch' window. The main interface has an 'emotion' track. A 'FAPFace' window on the right shows a 3D face model.

Animation et Multimedia

IIIc. Emotion Editor

The screenshot shows the FAPFace software interface. On the left, there is a 'LipSynch' window. The main interface has a 'Head Movements' track. A 'FAPFace' window on the right shows a 3D face model.

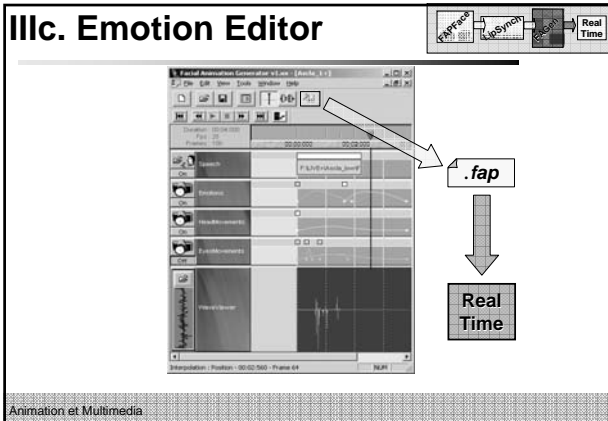
Animation et Multimedia

IIIc. Emotion Editor

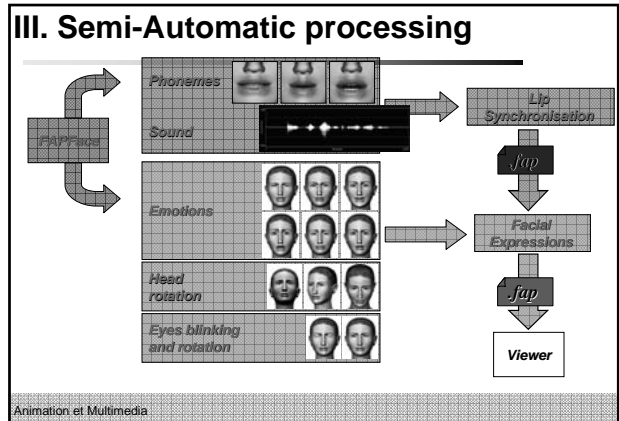
The screenshot shows the FAPFace software interface. On the left, there is a 'LipSynch' window. The main interface has an 'Eyes Movements' track. A 'FAPFace' window on the right shows a 3D face model.

Animation et Multimedia

IIIc. Emotion Editor



III. Semi-Automatic processing



4. Capture de mouvement

- Animation construite par système de capture optique (VICON)

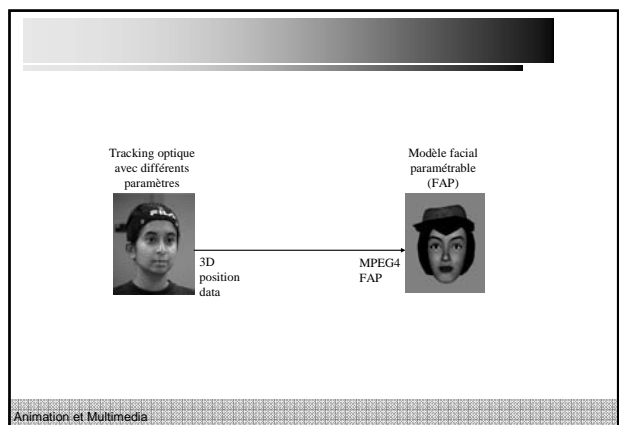
➤ AVANTAGES

- Très réaliste
- La voix, l'animation et les émotions sont enregistrées simultanément

➤ INCONVENIENTS

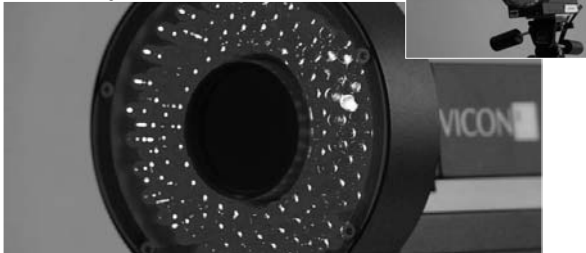
- Coûts des équipements

Animation et Multimedia



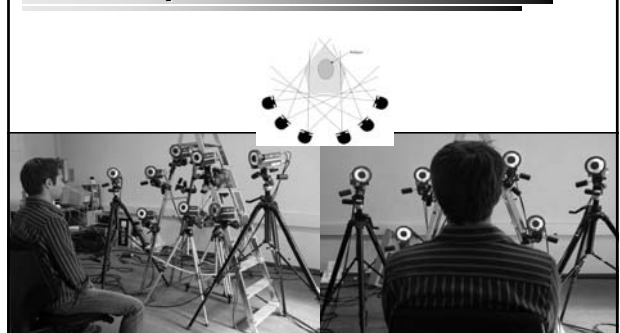
VICON Cameras

8 Infra-Red High-Resolution Cameras



Animation et Multimedia

Camera position



Animation et Multimedia

Markers

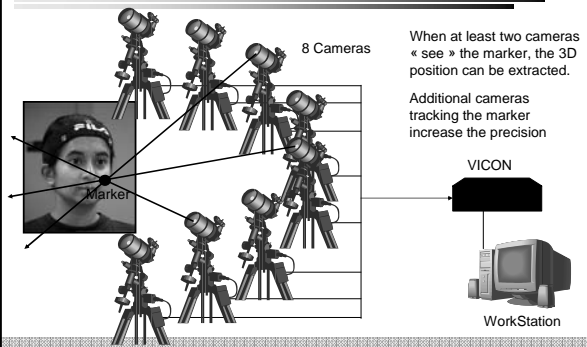
- 30 facial markers (3mm hemisphere)



- 3 markers for head spatial position
- 6 markers for eyebrows
- 4 markers for eyes
- 3 markers for nose
- 4 markers for cheeks
- 8 markers for mouth
- 2 markers for chin

Animation et Multimedia

VICON System



Animation et Multimedia

How to animate the virtual face?

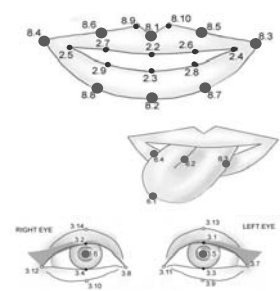
The challenge is to deform the virtual face in a realistic way, just using the spatial position of 30 points ...



Animation et Multimedia

Transformation in MPEG-4 format

- MPEG-4 FAP that are not tracked are interpolated from FAP that are tracked
- Some FAP cannot be tracked for obvious reasons (manual work)



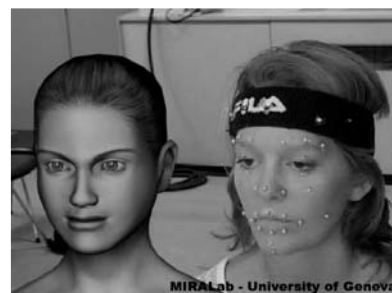
Animation et Multimedia

From VICON to FAP



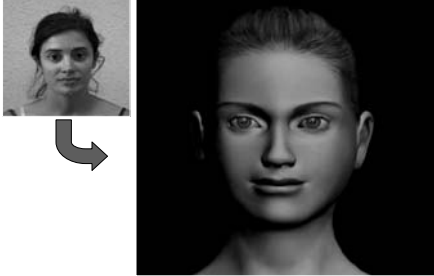
Animation et Multimedia

Results



Animation et Multimedia

Results



Animation et Multimedia

Results

Result on a virtual face



Animation et Multimedia

Exemple : Animation pour Half Life 2 ☺

➤ Exactement la même approche



Animation et Mul

■ Faceposer allows you to direct each actor in every scene.