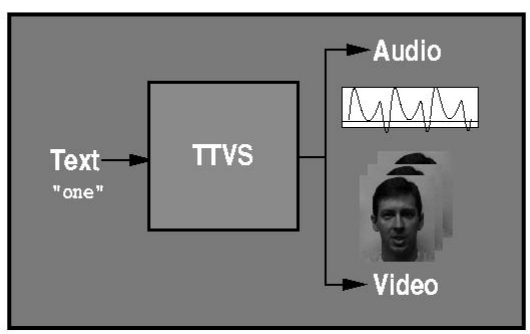
MikeTalk : A Photorealistic Talking Facial Display

Tony Ezzat
Tomaso Poggio

IAP, January 2000

TTVS Overview

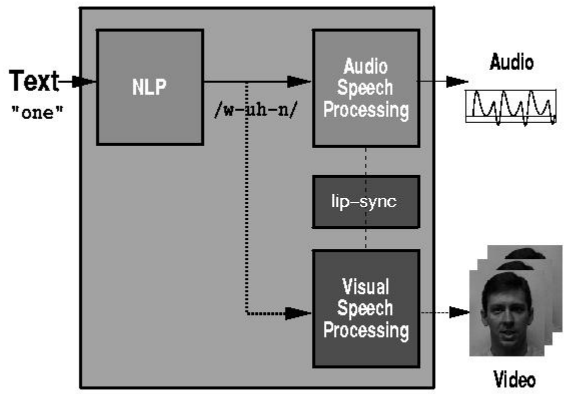
- **Input: Text**
- **Output: photorealistic talking face uttering text**



IAP, January 2000

TTVS Modules

CBCl MIT



IAP, January 2000

Phoneme Corpus

CBCl MIT

- **Step 1:**
  - **collect a visual corpus from a subject.**
  - **corpus contains 44 words**
  - **one word for each American English phoneme.**

monophthongs	
ii	bead
i	bid
e	bed
a	bad
o	body
aa	father
uh	bud
oo	baud
u	book
uu	boot
@	about
@@	bird

diphthongs	
ou	boat
ei	bait
au	bout
ai	bide
oi	boyd
e@	there
i@	near
u@	moor

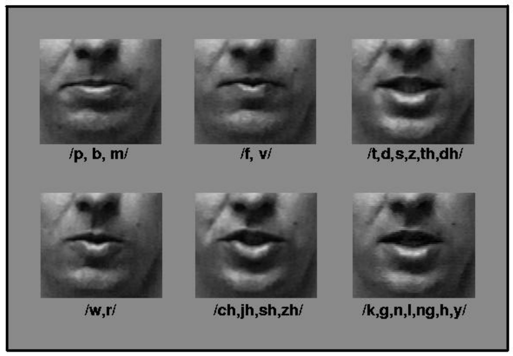
consonants	
r	ride
l	light
w	wide
y	yacht
m	might
n	night
ng	song
b	bite
d	dog
g	get
p	pet
t	tea
k	key
v	veal
dh	then
z	zeal
zh	garage
f	feel
th	thin
s	seal
sh	shore
h	head
jh	jeep
ch	chore

IAP, January 2000

6 Consonantal Visemes

CBCl MIT

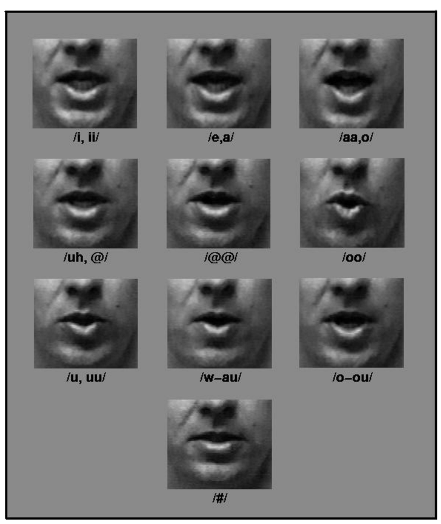
- **Step 2**
  - **extract one image per phoneme: <u>viseme</u>**
  - **group visemes together by visual similarity**



IAP, January 2000

9 Vocalic Visemes (+ 1 SilenceViseme)

CBCl MIT

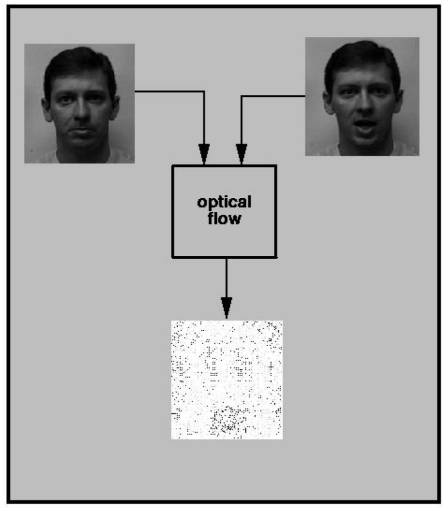


IAP, January 2000

Optical Flow

CBCl MIT

- **To interpolate between two visemes, optical flow is first computed**
- **A 2D motion vector field is produced**

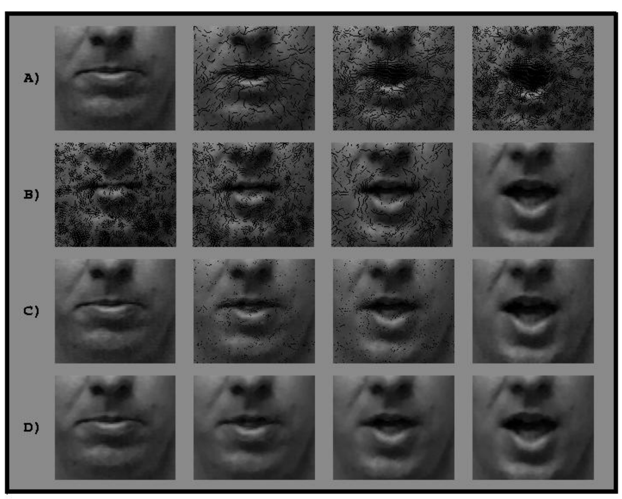


IAP, January 2000

Morphing

CBCl MIT

- **Forward warping A to B**
- **Forward warping B to A**
- **Blending**
- **Holefilling**

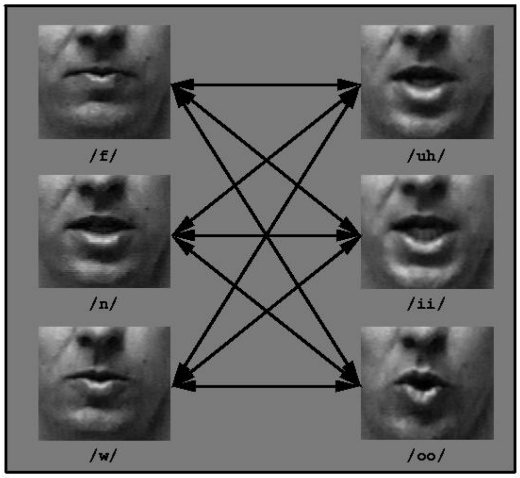


IAP, January 200(

Synthesis Database

CBCl MIT

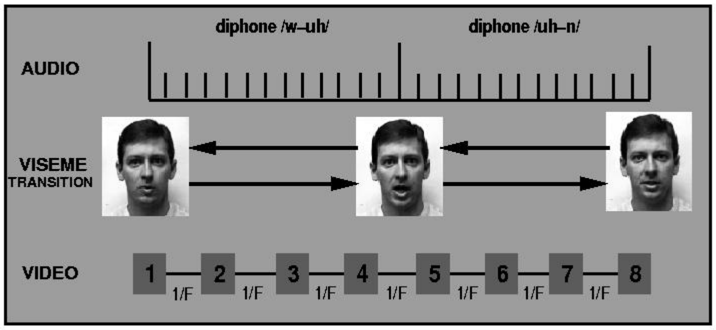
- 16 visemes total
- 256 optical flow vectors total, from every viseme to every other viseme



IAP, January 2000

Concatenation and Lip Sync

CBCl MIT



- Load the correct viseme transitions
- Concatenate viseme transitions
- Sample the viseme transitions using audio durations

IAP, January 2000

Further Work

CBCl MIT

- Learning of dynamics/coarticulation for smoother mouth motion
- Visual Prosody: eye movements, eye brows, head movements
- Viewpoint changes

IAP, January 2000

Applications

CBCl MIT

- Special Effects/Digital Actors
- Desktop agents
- Automated news/web page presenters

IAP, January 2000