



Multimodal Speaker Identity Conversion



- *"The show must go on"* -



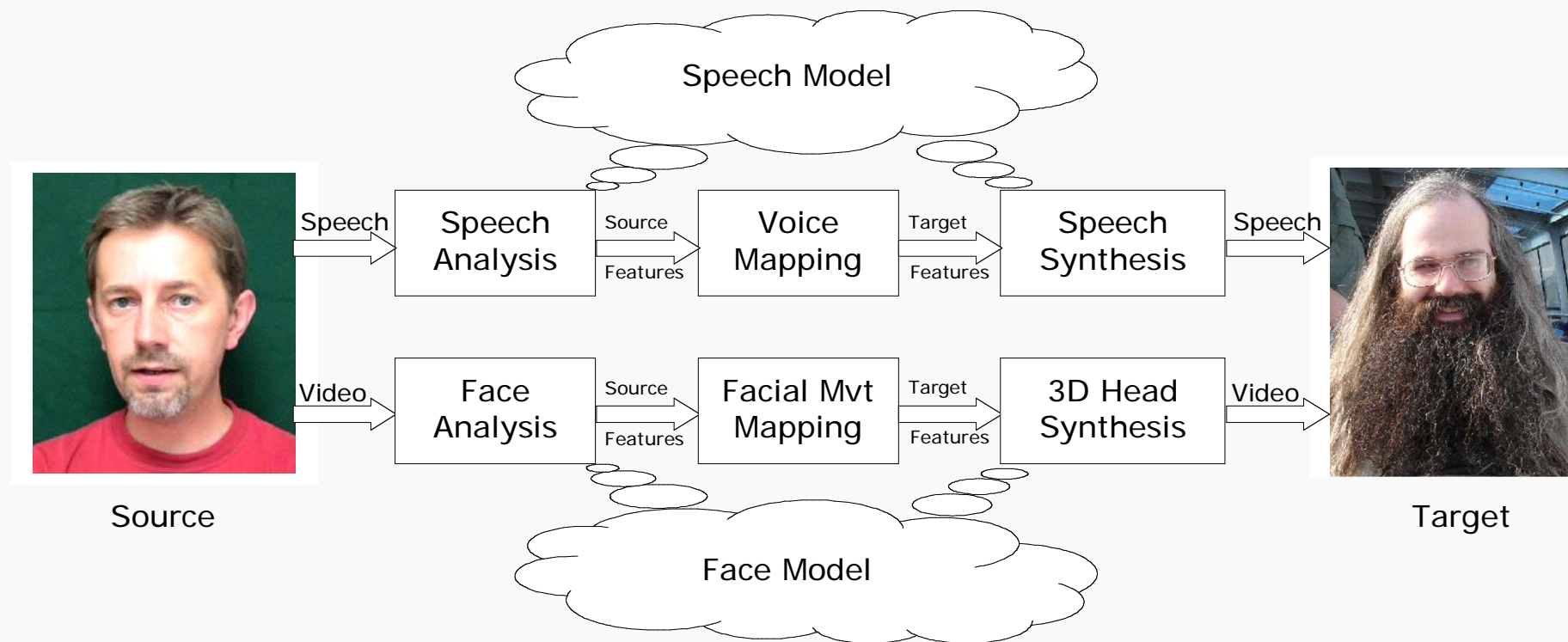
L. Arslan T. Dutoit Z. Inanoglu M. Jottrand M. Markaki I. Panzic M. Saraçlar K. Stankovic Y. Stylianaou A. Zara



General view of the problem

eNTERFACE '07

The SIMILAR NoE
Summer Workshop
on Multimodal Interfaces



Quite a challenging project...

Contents

eNTERFACE '07

The SIMILAR NoE
Summer Workshop
on Multimodal Interfaces



§ Voice conversion

- ú eNTERFACE06

- ú eNTERFACE07

§ Face Conversion

- ú eNTERAFCE06

- ú eNTERFACE07

§ DEMO

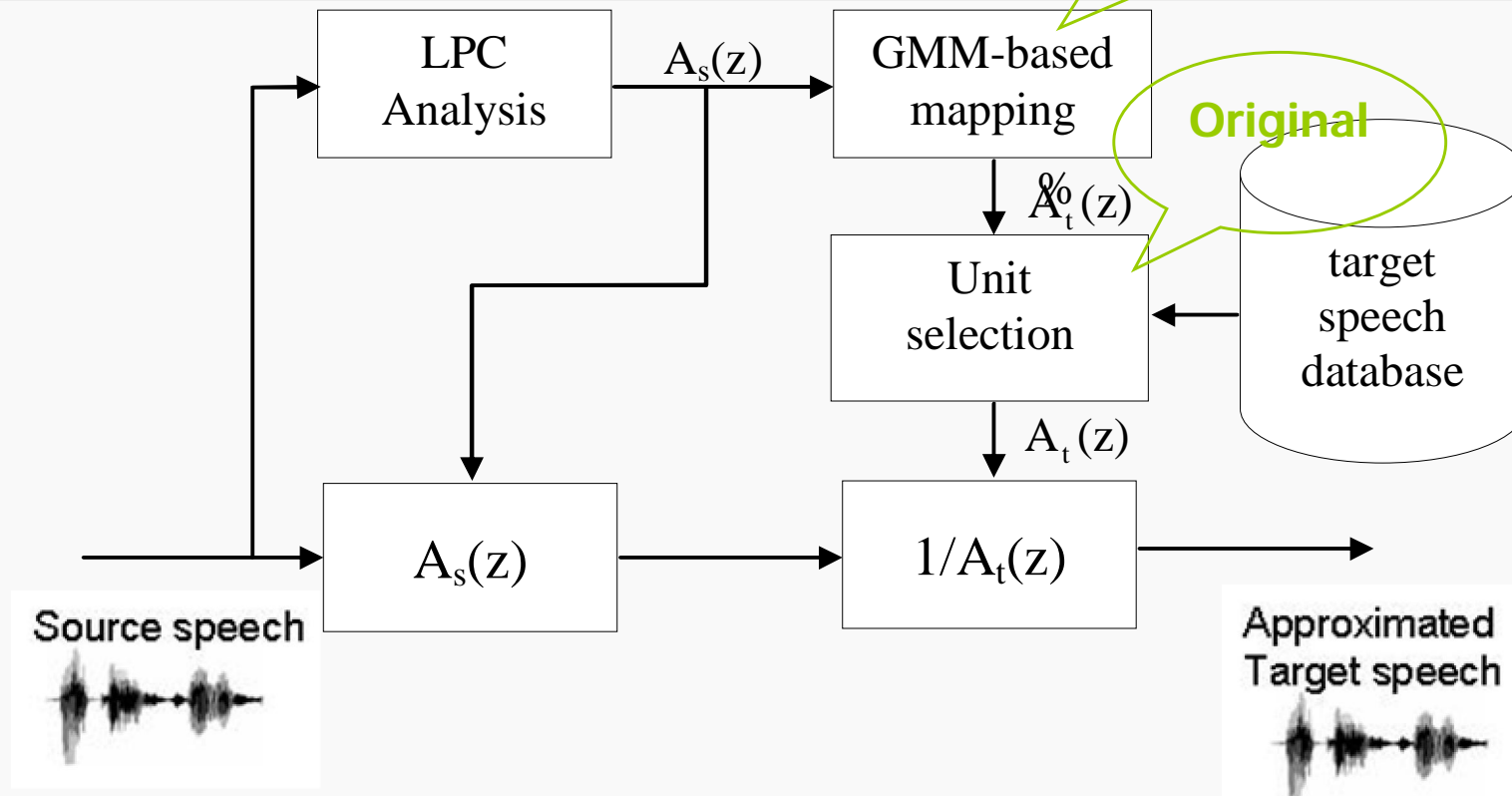
eNTERFACE'06: Voice conversion

eNTERFACE '07

The SIMILAR NoE
Summer Workshop
on Multimodal Interfaces

Stylianou
98

Original



Good results on the spectral envelopes,
but excitation bears speaker info too!

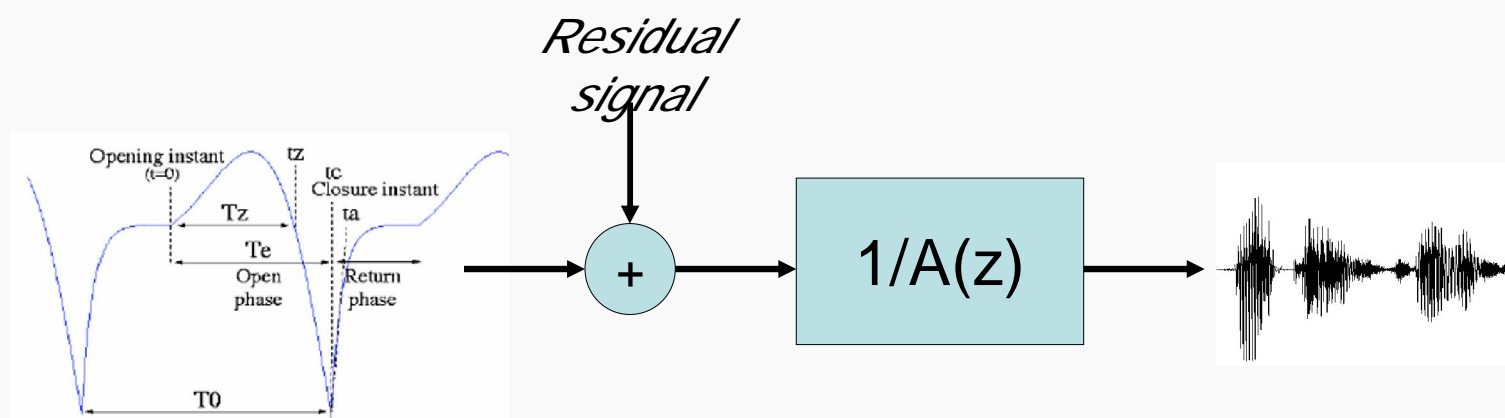
eNTERFACE'07:

Voice conversion (ARX-LF)

eNTERFACE'07

The SIMILAR NoE
Summer Workshop
on Multimodal Interfaces

§ Try using a new Source-Tract separation algorithm:
ARX-LF (“AutoRegressive eXogeneous Lijlencrants-Fant
model”: Vincent, Rossec, Chonavel 2007)



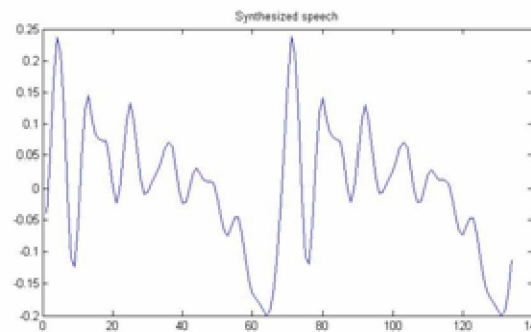
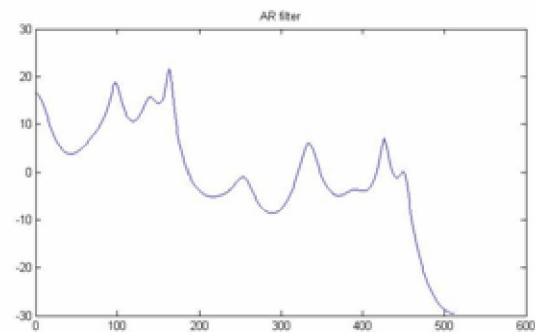
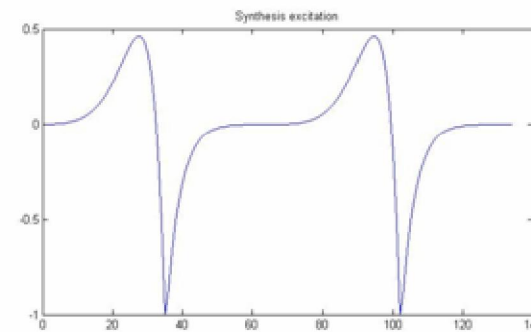
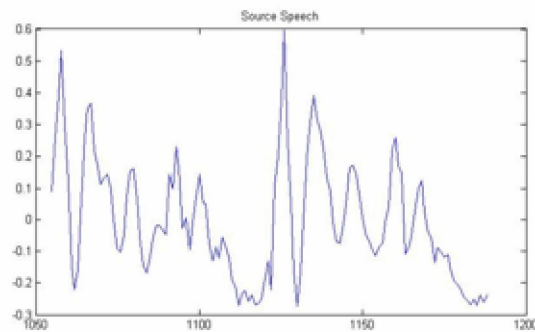
eNTERFACE'07: Voice conversion results (ARX-LF)

eNTERFACE '07

The SIMILAR NoE
Summer Workshop
on Multimodal Interfaces



§ Copy synthesis works fine (but sensitive to GCI!)



eNTERFACE'07:


Voice conversion results (ARX-LF)

eNTERFACE '07

The SIMILAR NoE
Summer Workshop
on Multimodal Interfaces



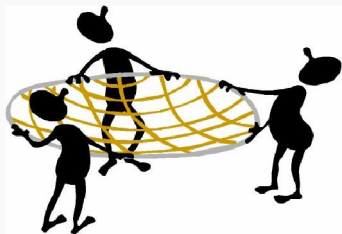
- § Parameter transplantation (target → source) gives **bad audio results**
- § Transplanting source (LF) parameters does not change speaker ID much

Source	Target	Transplantation
		
		

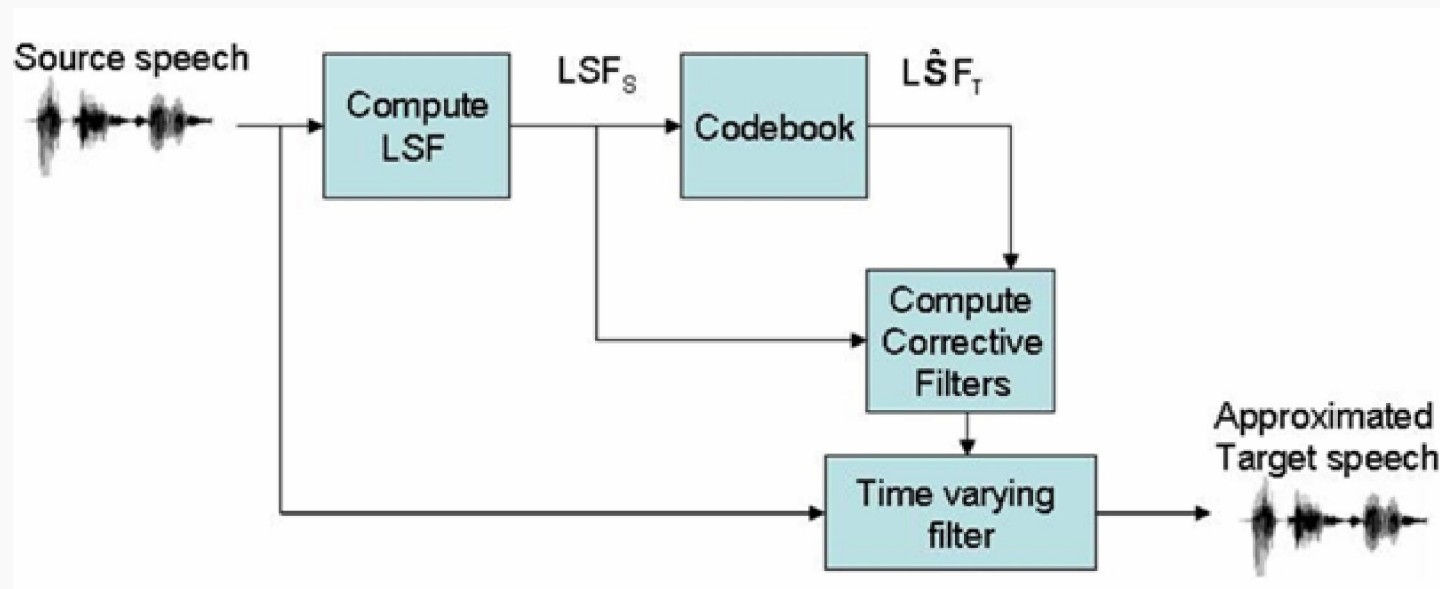
eNTERFACE'07: Voice conversion (STASC)

eNTERFACE '07

The SIMILAR NoE
Summer Workshop
on Multimodal Interfaces



§ Study L. Arslan's STASC ("Speaker Transformation Algorithm using Segmental Codebooks") conversion method



eNTERFACE'07:

Voice conversion results (STASC)







eNTERFACE'07

The SIMILAR NoE
Summer Workshop
on Multimodal Interfaces



§ Similar to last year's audio results:

- ú Incomplete ID conversion
- ú **But** smoother converted speech
- ú **And** using a less computationally expensive algorithm

Source	Target	Conversion
		
		

eNTERFACE'06: Face Analysis

eNTERFACE '07

The SIMILAR NoE
Summer Workshop
on Multimodal Interfaces



- § Tracking the face *region*
 - § Extracting facial *features* such as lips, nose, etc.
 - § Tracking these features throughout the video
- è Active Appearance Models
- § Animating a 3D Xface model (No face conversion achieved)



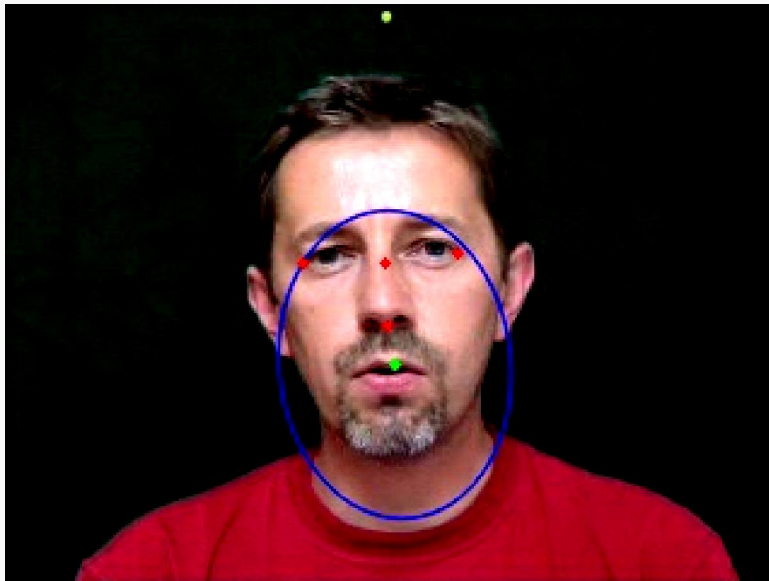
eNTERFACE'07: Face Analysis

eNTERFACE'07

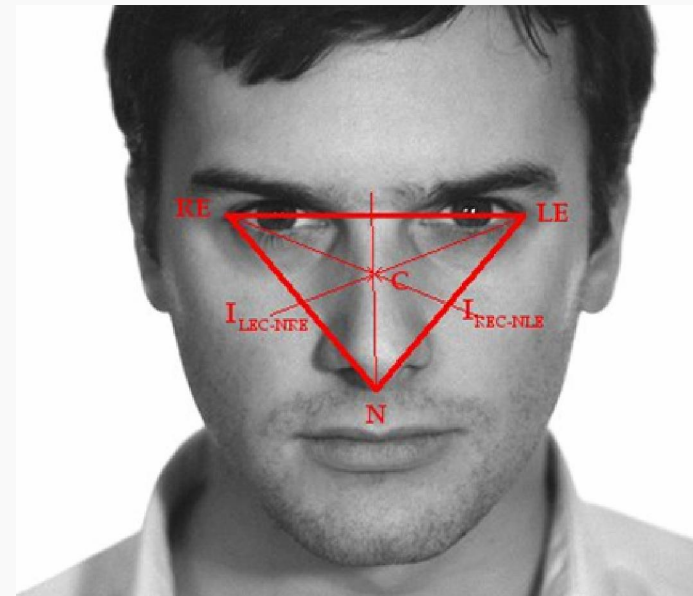
The SIMILAR NoE
Summer Workshop
on Multimodal Interfaces

Improving the global head tracking

§ Previously:
Ellipsis-based



§ Currently:
(Lips+nose)-based



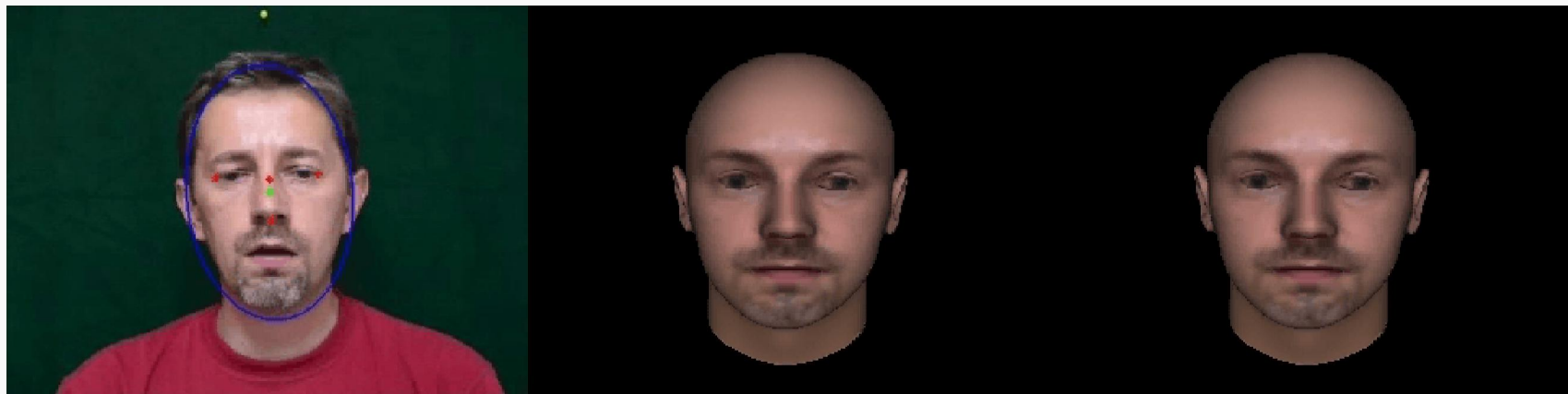
eNTERFACE'07: Face Analysis Results

eNTERFACE'07

The SIMILAR NoE
Summer Workshop
on Multimodal Interfaces



Improving the global head tracking
(using an XFACE model, VISAGE doe not
render global head movements)



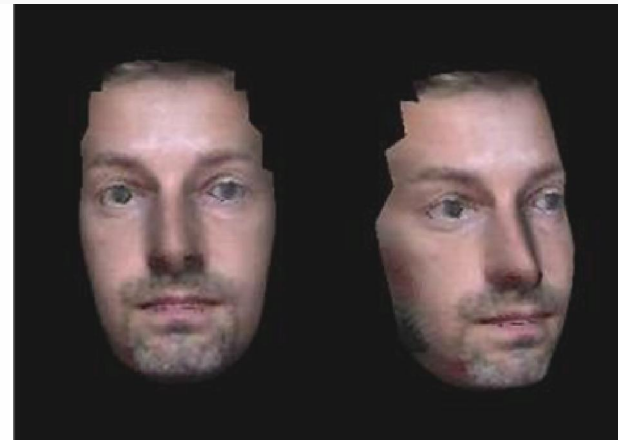
eNTERFACE'06: Face Modeling

eNTERFACE '07

The SIMILAR NoE
Summer Workshop
on Multimodal Interfaces



§ Building photo-realistic models using PHOTO-FIT



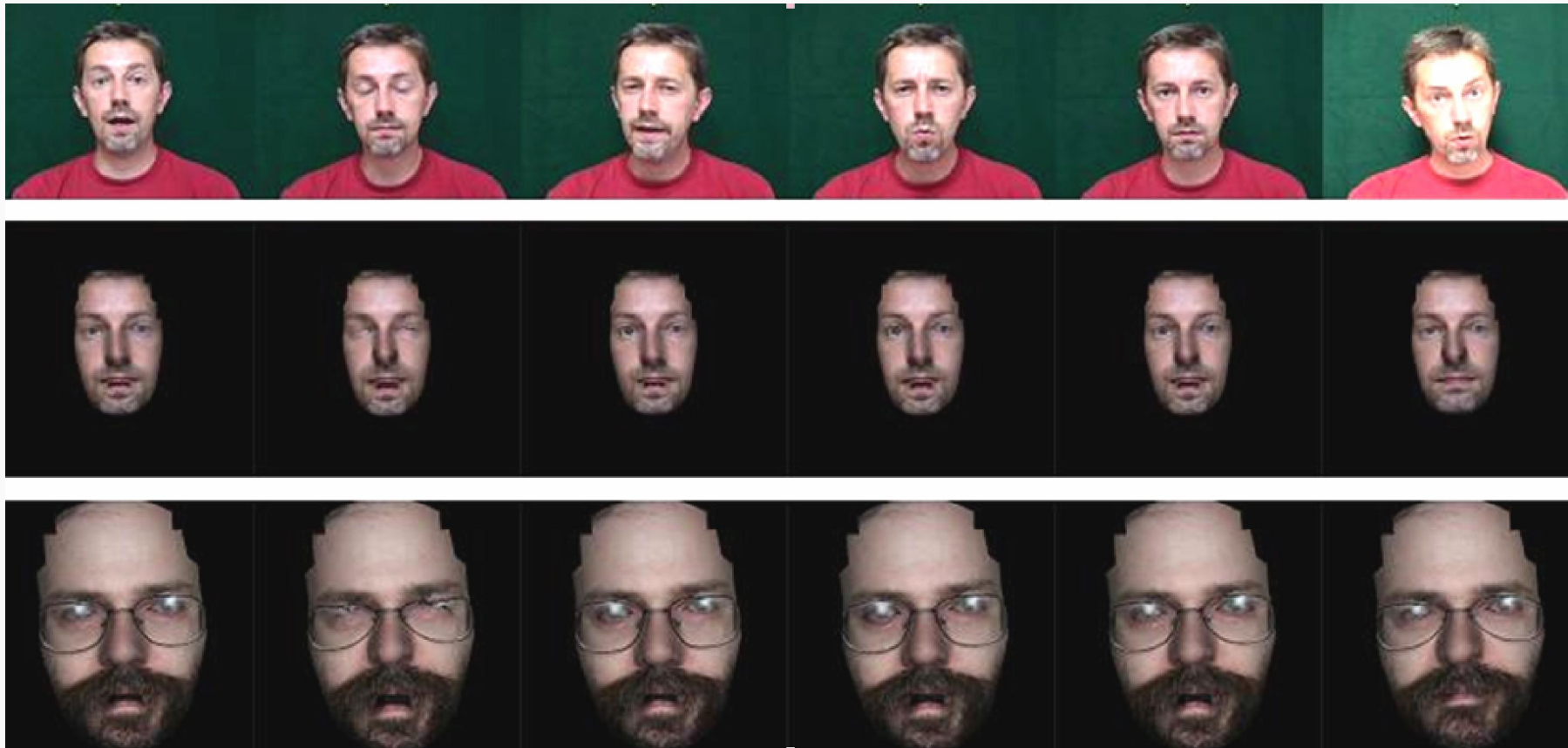
eNTERFACE'06: Face Synthesis

eNTERFACE '07

The SIMILAR NoE
Summer Workshop
on Multimodal Interfaces



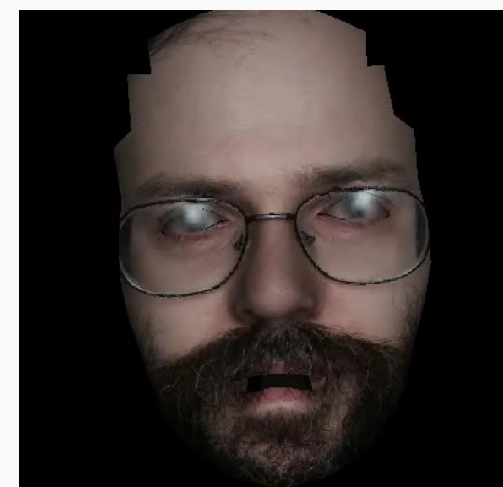
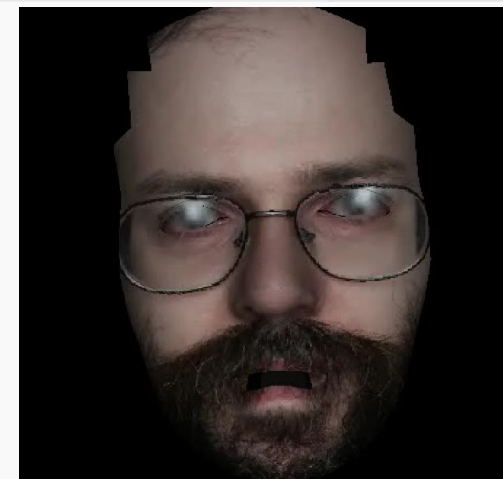
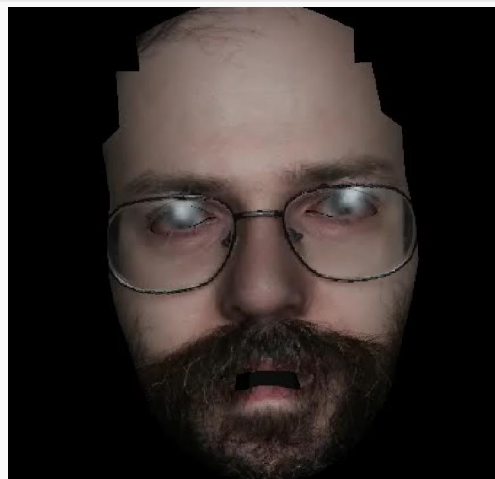
§ Using the source FAPs to drive the target model



DEMO TIME

eNTERFACE '07

The SIMILAR NoE
Summer Workshop
on Multimodal Interfaces



Artefacts J

eNTERFACE '07

The SIMILAR NoE
Summer Workshop
on Multimodal Interfaces

