

Voice transformation and speech synthesis for video games

Snorre Farner¹, Axel Roebel,
Christophe Veaux¹, Gregory Beller¹ ,
Xavier Rodet¹,
and Laurent Ach²

¹ www.ircam.fr

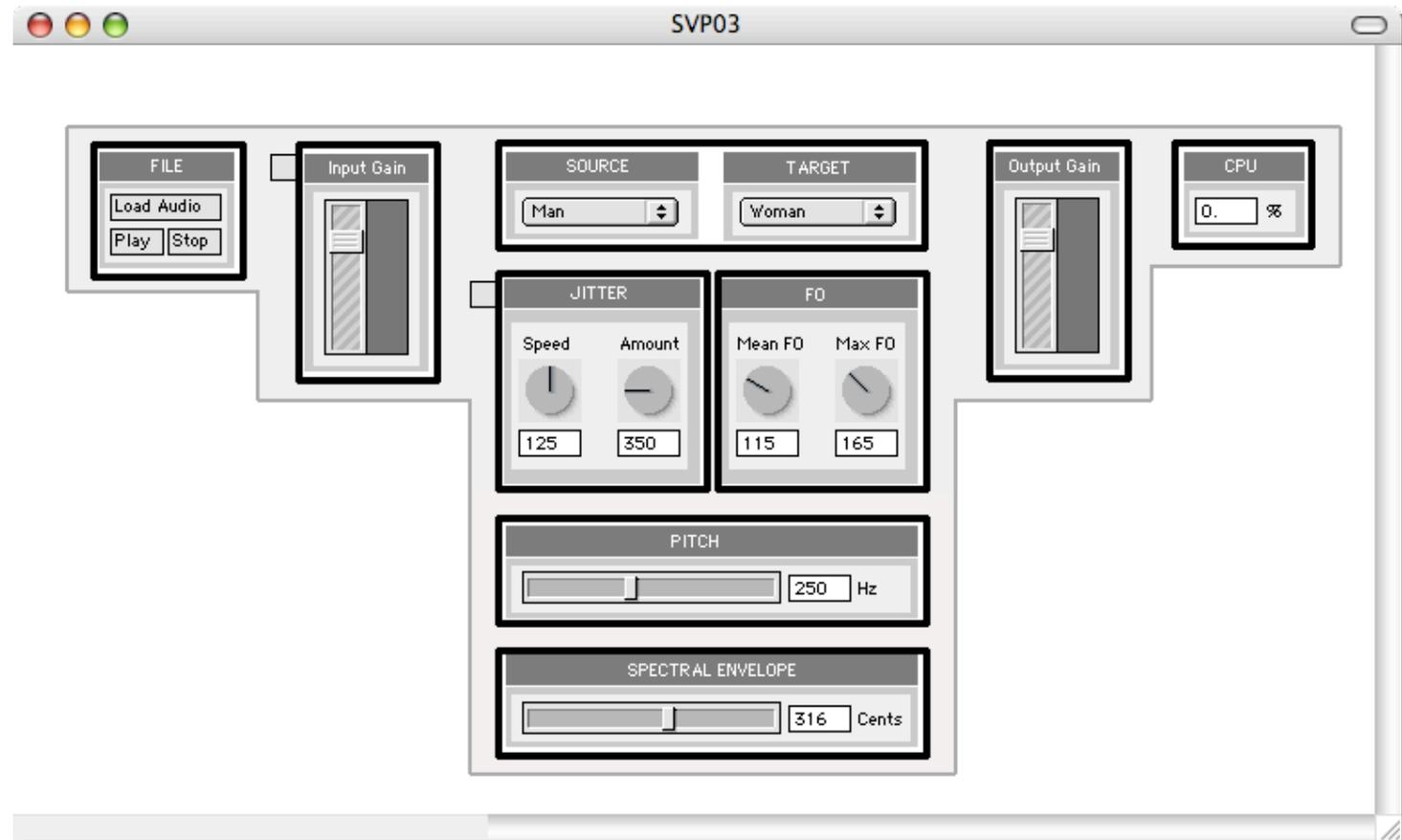


² www.cantoche.com



www.parisgdc.com

Demo: Real-time transformation



Overview

- Introduction
- Advanced voice transformation
- Expressivity transformation
- Text-to-speech synthesis
- Avatar production
- Demo: speaking avatars

Introduction

- Application of speech in games:
 - narrators and NPCs in video games
 - players' communication in multiplayer role-playing games
 - expressive voice in multimédia: the ANR-*Vivos* project
- Non-entertainment games:
 - educational games
 - e-learning
 - “serious games”

Current use of speech in games

- prerecorded speech (narrator, NPCs)
- player's speech (VoIP)
- basic sound effects on the voice

- Limitations:
 - utterances must be predetermined
 - recording of several actors may be necessary

Artistic research at IRCAM

- Our objectives: **artistic applications**
 - music, multimedia, films, dubbing, cartoon characters, etc.
- Requirements:
 - very high **sound quality**
 - very high degree of **naturalness**
 - **automatic** solution
 - **user control**

Speech tools

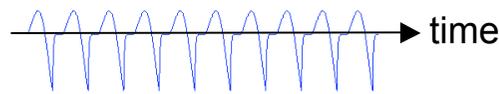
- We present a set of tools to:
 - transform the voice of **one actor** into several different voices
 - **design the voice** of a playing character based on the player's voice
 - **modify** speech to express **emotions**
 - produce **arbitrary sentences** by text-to-speech synthesis
 - create a **visual avatar** (Cantoche)
 - transform in **real time**

Library of voice transformation “voiceTrans”

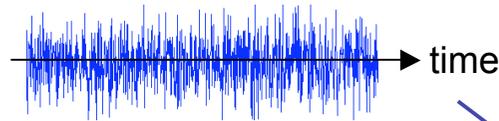
- Transformation of **type**:
 - sex, age, animal voice, fictional voice,...
- Transformation of **voice quality**:
 - whispering, breathy, hoarse,...
 - dark/bright, nasal, strong/weak,...
 - relaxed/tense, creaky
- Transformation of **speech style**:
 - trembling, singing, stuttering,...
 - lively, dull, eager, lazy, drunk,...

The voice

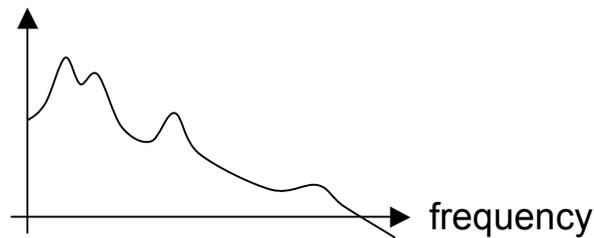
- Pulsation of vocal folds



- Turbulence in constrictions



- Vocal tract resonance



- Speech signal

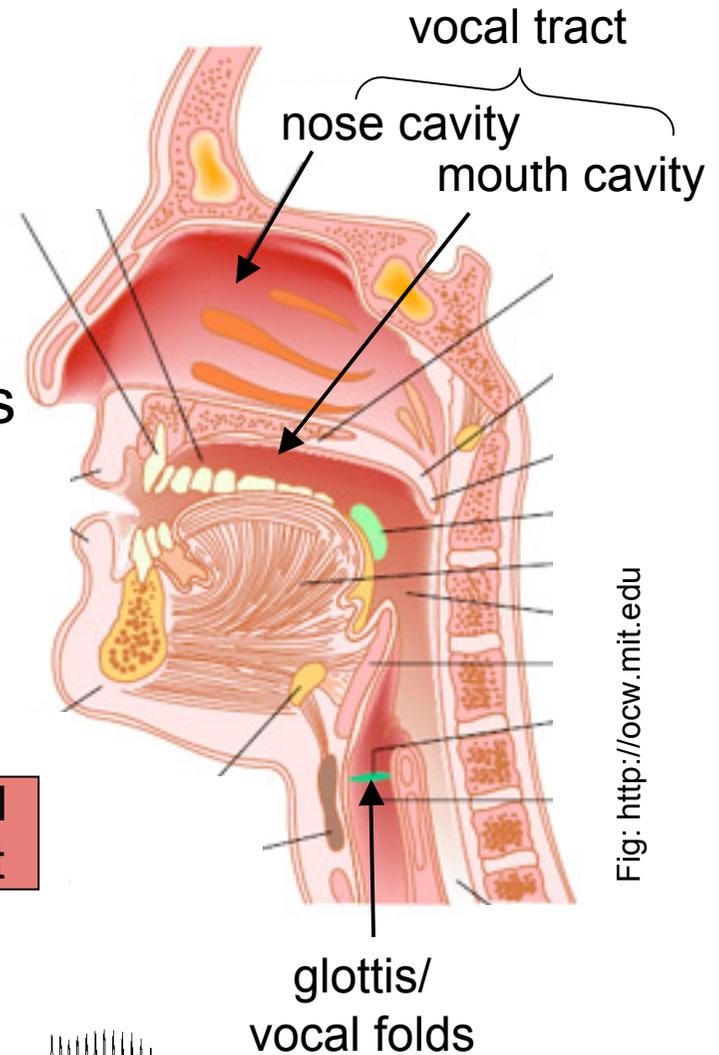


Fig: <http://ocw.mit.edu>

Signal transformation

- Modification of

- pitch



- vocal tract



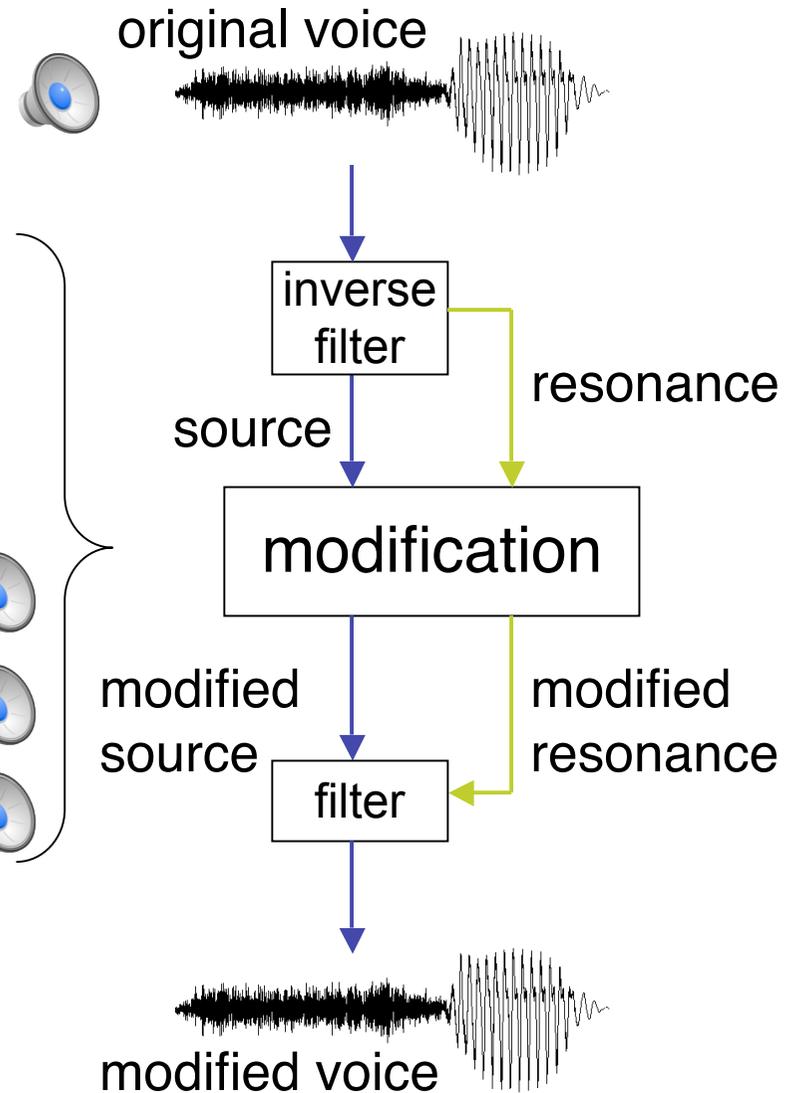
- voiced contents



- noise contents



- glottal source



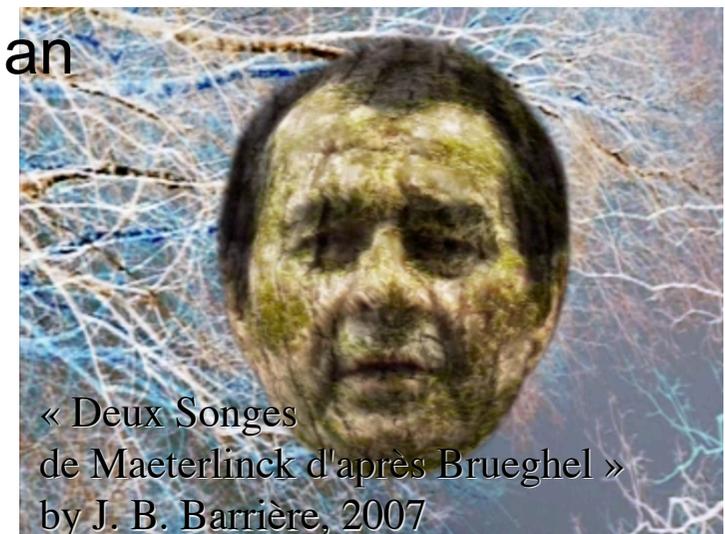
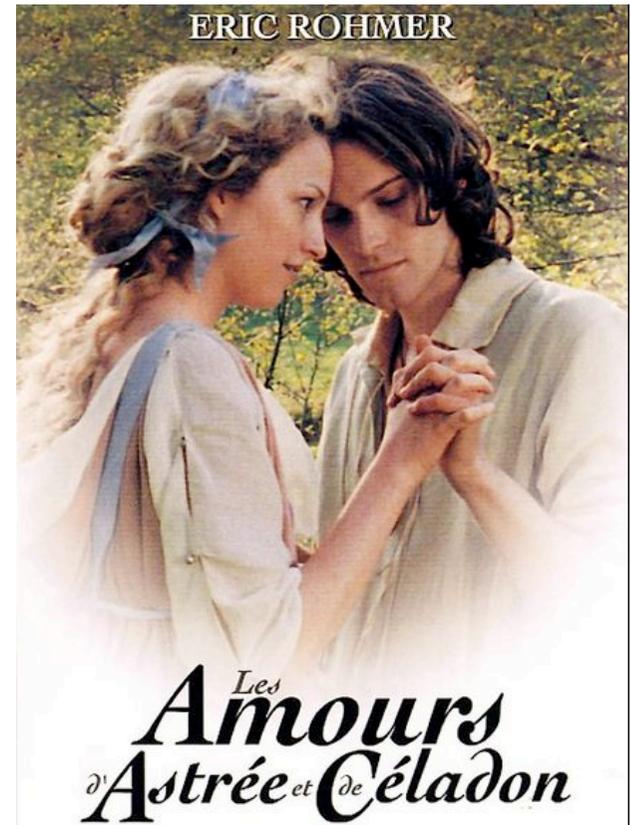
* Sound examples also available at

<http://recherche.ircam.fr/anasyn/farner/pub/GDC08>

www.parisgdc.com

Transformation of sex and age

- Disguising man to woman:
 - ...also the voice:  → 
 - Céladon  → Alexie 
- One actor to 12 persons:
 -  →  5th Blind (woman)
 -  →  Oldest Blind Woman
 -  →  Oldest Blind Man
 -  →  3rd Blind (man)
- Monologue → dialog
 



Other voice transformations

- original 
- breathy 
- whispering 
- creaky  (irregular vocal-fold movement)
- softer voice  (glottal source)
- trembling 
- dull  and eager  speech
- drunk 

Text-to-speech synthesis

- Construction of **database**:
 - **Recording** of actor(s)
 - **Segmentation** and classification
- **Text analysis**
 - ⇒ syntax ⇒ phone sequence
- **Prosody management** (duration, intensity, pitch)
 - from model ⇒ target prosody, or
 - naturally by selection by phonologic position
- **Selection** of speech units
- **Concatenation** and possibly **modification**

Examples of synthesis

- “C’est un soldat(,) à cheveux gris” 
- “Mon chien...” 
- Monologue:  → dialog: 

Training expressivity

- Basic emotions:
 - neutral
 - happiness
 - fear
 - sadness
 - anger
- Acoustic attributes:
 - pitch
 - speech rate - duration
 - force - intensity
 - articulation degree
 - phonation - voice quality
- Introvert ↔ extrovert
- Different intensity levels
- Intentions and attitudes:
 - surprise, disgust, discretion, excitation, confusion

Transformation of expressivity

- Construction of expressivity database
- Training of expressivity models
- Two complementary approaches):
 1. expressivity criterion in unit-selection stage
 2. transformation of synthetic or natural speech
 - analysis and segmentation of speech
 - transformation of prosody and timbre

Preliminary examples

introvert extrovert

- neutral 
- happiness 
- fear  
- sadness  _
- anger 
- negative surprise  _



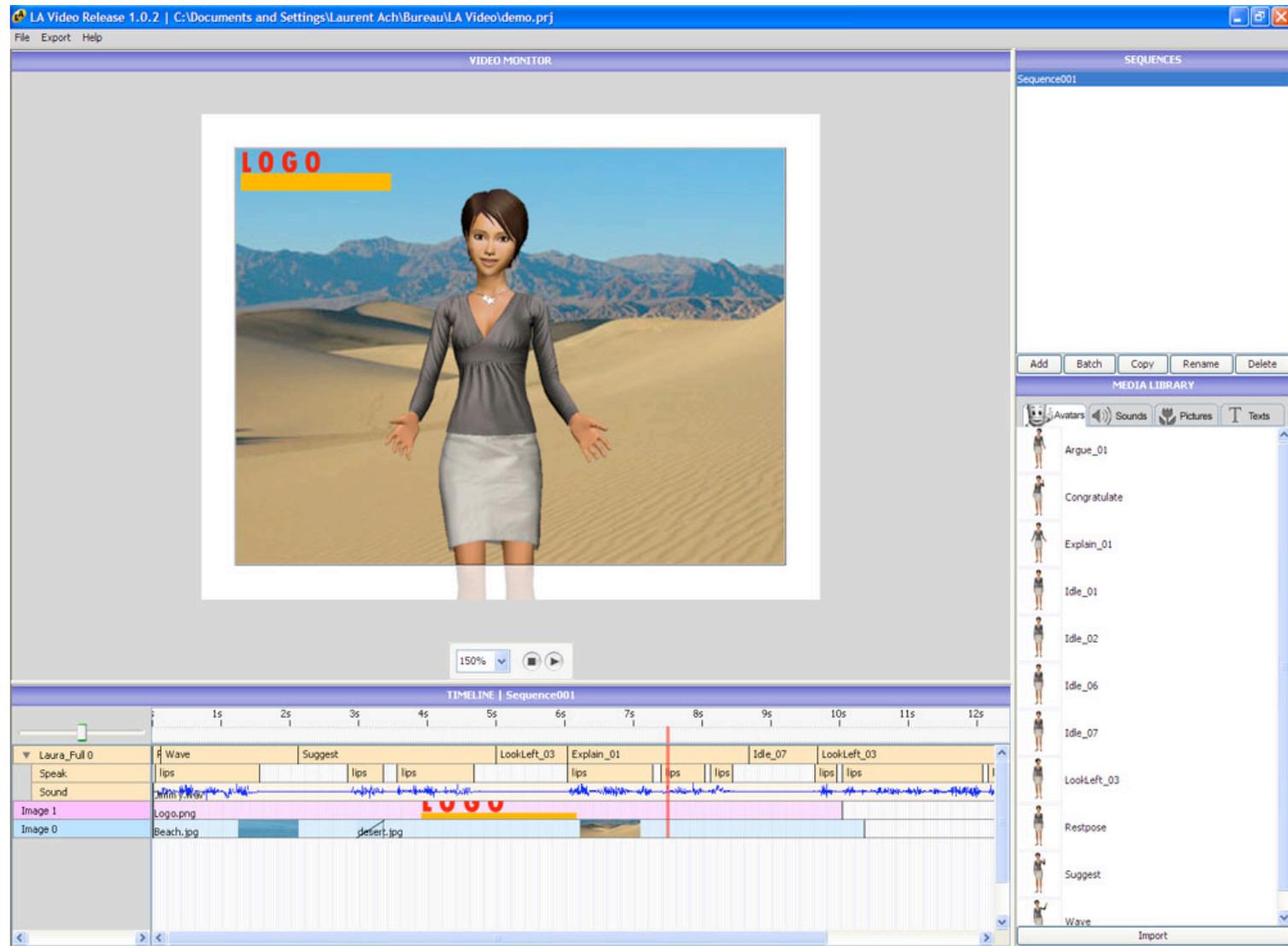
Living Actor™ Avatars

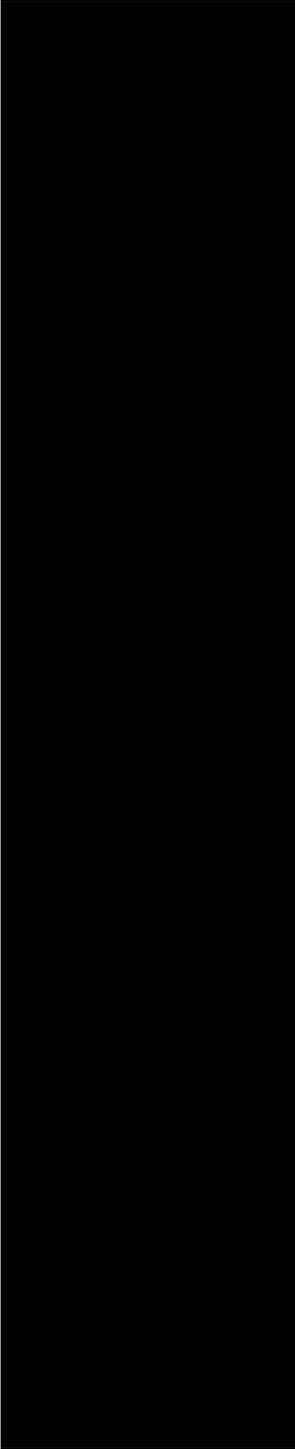
- behavior depending on avatar personality
- gestures and expressions from voice analysis
- mixing avatar animations, audio and images data
- Speaking Avatars
 - emotion detection in voice
 - multimodal correlations
 - voice transformation





Living Actor™ – Creation





Demo: Speaking Avatars

One actor → 4 characters