RAMCESS & HandSketch
A Framework for Realtime Control of Expressive Singing Synthesis

Interspeech 2007 - Synthesis of Singing Challenge

Nicolas d’Alessandro, Thierry Dutoit
Faculté Polytechnique de Mons (Belgium)
Table of Contents

- Introduction
- Context of the Research
- Voice Quality Control Library
- RAMCESS 1.x to 2.x
- HandSketch Bi-Manual Controller
- Synthesizer Song
- Comments about the Results
Content of the Research

Production

Perception

Interaction

Instrument making

Voice

Quality

Control

VQCLib

RAMCESS

HandSketch
Voice Quality Control Library

jointly maximize expressive & realtime abilities of the glottal source + vocal tract filter synthesis scheme

\[ n_g(t) = \frac{1 + e^{at}(a_{\alpha m} \sin(\frac{\pi t}{\alpha m}) - \cos(\frac{\pi t}{\alpha m}))}{1 + e^{a\alpha_m}} \]

\[ H_{TL}(z) = \frac{b_{TL}}{1 - a_{TL} \cdot z^{-1}} \]
Voice Quality Control Library

jointly maximize expressive & realtime abilities of the glottal source + vocal tract filter synthesis scheme

Glottal source, lattice filter and conversion functions are all available as Max/MSP objects
http://vqclib.blogsite.org

\[ /a/ = [x^a_1, x^a_2, \ldots, x^a_N] \]

\[ \Delta_{LAR} = [\Delta l_1, \Delta l_2, \ldots, \Delta l_N] \]

LSF Preset Interpolation

LSF to LAR

LAR to Reflection
RAMCESS 1.x to 2.x

a unified dimensional representation of voice quality variations and realtime mappings to achieve it

RAMCESS 1.x to 2.x

A unified dimensional representation of voice quality variations and realtime mappings to achieve it

Frequency | Effort | Pressure | Registers | Contents
---|---|---|---|---

Dimensional Mapping

Intra-dimensional Relations

\[ O_q = f(F_0, P, E, M_i) \]

Phonetograms coupling effects effort/pressure

The need to process a database: from pure synthesis to voice transformation!
RAMCESS 1.x to 2.x

extraction of all glottal source parameters in a coarticulated voice database is not trivial!


HandSketch Bi-Manual Controller

fretless angular control of pitch
radial control of “presfort”
pressure control of energy
finger position control of multiphones browsing

Synthesizer Song

Max/MSP programming, full realtime control, dba @ 16k
Comments about the Results

- Intuitive management of prosodic contour
- Lattice browsing in coarticulated contents
- Large spectrum of possible VQs (with only 1 dba) (sounds)
- Realtime local playing with the “presfort” axis is really expressive (e.g. crescendos+vibrato)
- Encouraging sound quality (now: use residuals, go to 44k,...)
- Algorithms known not to be robust in the analysis of coarticulated corpus, < 2% wrong frames so +/- automatic
- Ready for implementing things like NUU with larger corpus