

web [www.leoalmanac.org](http://www.leoalmanac.org)

email [info@leoalmanac.org](mailto:info@leoalmanac.org)

phone +90 216 483 9292

address Sabancı University, Orhanlı - Tuzla, 34956 Istanbul, Turkey

Lea

## **LEONARDO ABSTRACT SERVICES (LABS) 2008-2009 REALTIME AND ACCURATE MUSICAL CONTROL OF EXPRESSION IN SPEECH SYNTHESIS**

**By Nicolas d'Alessandro**

In the early days of speech synthesis research, understanding voice production has attracted the attention of scientists with the goal of producing intelligible speech. Later, the need to produce more natural voices led researchers to use prerecorded voice databases, containing speech units, reassembled by a concatenation algorithm. With the outgrowth of computer capacities, the length of units increased, going from diphones to non-uniform units, in the so-called unit selection framework, using a strategy referred to as "take the best, modify the least".

Today the new challenge in voice synthesis is the production of expressive speech or singing. The mainstream solution to this problem is based on the "there is no data like more data" paradigm: emotion-specific databases are recorded and emotion-specific units are segmented.

In this thesis, we propose to restart the expressive speech synthesis problem, from its original voice production grounds. We also assume that expressivity of a voice synthesis system rather relies on its interactive properties than strictly on the coverage of the recorded database.

To reach our goals, we develop the RAMCESS (Realtime and Accurate Musical Control of Expression in Speech Synthesis) software system, an analysis/resynthesis pipeline which aims at providing interactive and realtime access to the voice production mechanism. More precisely, this system makes it possible to browse a connected speech database, and to dynamically modify the value of several glottal source parameters.

In order to achieve these voice transformations, a connected speech database is recorded, and the RAMCESS analysis algorithm is applied. RAMCESS analysis relies on the estimation of glottal waveforms and vocal tract impulse responses from the prerecorded voice samples. We cascade two promising glottal flow analysis algorithms, ZZT and ARX-LF, as a way of reinforcing the whole analysis process.

Then the RAMCESS synthesis engine computes the convolution of previously estimated glottal source and vocal tract components, within a realtime pitch-synchronous overlap-add architecture (PSOLA). A new model for producing the glottal flow signal is proposed. This model, called SELF (Spectrally Enhanced Liljencrants-Fant), is a modified LF

web [www.leoalmanac.org](http://www.leoalmanac.org)  
email [info@leoalmanac.org](mailto:info@leoalmanac.org)  
phone +90 216 483 9292  
address Sabancı University, Orhanlı - Tuzla, 34956 Istanbul, Turkey

Lea

model, which covers a larger palette of phonation types and solving some problems encountered in realtime interaction.

Variations in the glottal flow behavior are perceived as modifications of voice quality along several dimensions, such as tenseness or vocal effort. In the RAMCESS synthesis engine, glottal flow parameters are modified through several dimensional mappings, in order to give access to the perceptual dimensions of a voice quality control space.

The expressive interaction with the voice material is done through a new digital musical instrument, called the HandSketch: a tablet-based controller, played vertically, with extra FSR sensors. In this work, we describe how this controller is connected to voice quality dimensions, and we also discuss the long term practice of this instrument.

Compared to the usual prototyping of multimodal interactive systems, and more particularly digital musical instruments, the work on RAMCESS and HandSketch has been structured quite differently. Indeed our prototyping process is rather inspired by the traditional instrument making and based on embodiment. This luthery-inspired methodology leads us to propose the Analysis-by-Performance (AbP) paradigm, a methodology for approaching signal analysis problems. The main idea is that if signal is not observable, it can be imitated with an appropriate digital instrument and a highly skilled practice. Then the signal can be studied by analyzing the imitative gestures.

Degree: Applied Sciences

Year: 2009

Pages: 214

University: University of Mons

Supervisor: Thierry Dutoit

Semail: [thierry.dutoit@umons.ac.be](mailto:thierry.dutoit@umons.ac.be)

Supervisor 2:

Language: English

Dept: Signal Processing

Copyright: Nicolas d'Alessandro © University of Mons 2009

Lang\_author: English, French

Url: <http://nicolasdalessandro.net/phd/phd-print.pdf>

Email: [nicolas@dalessandro.be](mailto:nicolas@dalessandro.be)

Keywords: Realtime, Voice Synthesis, Tablet, Wacom, HandSketch, Performance, hci

web [www.leoalmanac.org](http://www.leoalmanac.org)

email [info@leoalmanac.org](mailto:info@leoalmanac.org)

phone +90 216 483 9292

address Sabancı University, Orhanlı - Tuzla, 34956 Istanbul, Turkey



Lea

## LEONARDO LABS

Leonardo LABS is annually peer reviewed by an international panel of reviewers. This year the panel selected the following abstracts to be published on Leonardo Electronic Almanac (LEA) in order to bring attention to the outstanding work of these authors by the international community interested in the intersection of art, science and technology.

The Peer Review Panel for 2009-2011 are: Yiannis Colakides, CoDirector, New Media (NeMe), Limassol, Cyprus, Christo Doherty, Chair, Digital Arts, University of the Witwatersrand, Johannesburg, South Africa, Ryszard W. Kluszczynski, PhD., Chair, Department of Media and Audiovisual Culture, University of Lodz, Poland, Tom Lesser, Director, Center for Integrated Media, California Institution of the Arts, Valencia, California, United States, Rachel Mayeri, Assistant Professor of Media Studies, Harvey Mudd College, Claremont, California, United States, Andrea Polli, Director, Interdisciplinary Film and Digital Media (IFDM) and Mesa Del Sol Chair of Digital Media, University of New Mexico, Albuquerque, New Mexico, United States, Sheila Pinkel, Professor of Art, Pomona College, Claremont, California, United States.