Is this guitar talking or what!?

Principal Investigator

Nicolas d'Alessandro

Candidate Participants

Maria Astrinaki, Loïc Reboursière, Thierry Dutoit
numediart – Institute of New Media Art Technology
University of Mons, Belgium

Abstract

In this project project we want to blend recent research done at numediart institute on speech / singing synthesis and detection of guitar playing techniques. It seemed interesting to explore the control of speech and singing synthesis from an already-expressive gesture such as the one of an instrumental technique, in our case the guitar playing. After having worked on developing and improving both parts, i.e. speech and singing synthesis algorithms and detection of guitar playing techniques, we now want to explore and propose several mappings between the guitar seen as a controller and the speech / singing synthesizer, regarding to a performance context. Porting the synthesis on mobile platforms, such as iOS or Android, will also be investigated, in order to propose portable, on-instrument, easy-to-use solutions.

Objectives

In this project we aim at developing a new framework for performative speech and singing synthesis, i.e. a realtime synthesis system where voice is directly produced by gestural control, with no more reference to textual input. We want to address both phonetical and prosodical issues, with applications in speech and singing synthesis. The target of this new system is to extend the context in which performative speech / singing synthesis is produced and explore a new relationship : controlled by an electric guitar. Indeed, we have known for a long time that managing all the parameters of speech production is tough for a single performer. However, we want to explore the idea of using instrumental gestures, i.e guitar playing techniques, since the subtlety and richness of the instrumental technique is a good start to have a refined control of the synthesized speech / singing. We want to see how intelligibility, naturalness and even speaker identity can be addressed as a guitar performance, involving the player and the audience.

Objective 1 – Interactive Control of Voice Production

Voice synthesis can be split in various types of typical issues to be solved: articulation and coarticulation of phonemes, speech timing management, intonation modelling, voice quality dimensions, etc. Most of these tasks refer to a significantly different representation of data and the development of appropriate human-computer interaction (HCI) models has not been widely studied for these tasks. We want to develop new interaction paradigms for voice synthesis based on an actual musical instrument.

Objective 2 – Second Release for the MAGE Platform

Most of current voice synthesis architectures are designed like a giant script which aims at writing down a waveform onto the hard drive. Eventually if we consider realtime and interactive voice synthesizers, we find out that their structure is quite monolithic and difficult to break down. MAGE, a platform for reactive HMM-based speech and singing synthesis, brought a first solution to reactively design voice synthesis with various components and controls being detachable on heterogeneous platforms – computers or mobile devices – while maintain sound quality and low latency. However, the currently provided controls are rather limited. We want to provide various reactive context control modules embedded in the platform, easily accessible to the user / developer / performer. Additionally we would like to explore the idea of reactive interpolation control between different speaking styles and voices over the currently synthesized voice. These integrated new control parameters can result to the release of a more complete, stable and flexible version of the MAGE platform. There is a linguistic and sociological interest in questioning and validating several properties of voice (at various levels: intelligibility, naturalness and identity) when this voice is produced by a performer.

Objective 3 – Guitar Independent Algorithms for Playing Techniques to Control a Voice Synthesizer

In the Guitar As Controller project, algorithms to detect playing techniques detection have been developed. A nylon string guitar has been used to create the database on which the algorithms have been created and tested. Other types of guitar / string need to be tested in order to make the algorithms guitar independent or to provide preset settings for each types of guitar.

Background

Speech and Singing Synthesis

Speech is a performance, a realtime, dynamic gestural phenomenon that involves vocal organs, face and body and transmits messages with both information and emotions. Even though vocal behavior and expression is a very complex mechanism, it is the richest and most ubiquitous modality of communication used by human beings, a highly interactive and social process. People have always been fascinated for artificial speech and singing production, and in the early years of voice synthesis, intraoral vocal gestures were being imitated by manual manipulation of physical objects, such as the von Kempelen's machine [1] or the Voder [2]. The simple models used at that time had quite poor intelligibility and naturalness but

these systems provided a rich and embodied interaction paradigm, like a musical instrument. When voice synthesis was implemented on computers, Text-To-Speech (TTS) became the main trend, where static text was converted into an intelligible and natural waveform, implicitly defining the keyboard input as the interaction model. From a human-computer interaction (HCI) perspective, the textual input is more of a historical heritage, based on similarities between first desktop computers and typewriters. The last twenty years brought many innovations to replace this textual input at the graphical level (e.g. mouse and icons) but voice synthesis has rarely been considered with more interactive input than text.

Existing Systems

Nowadays, with the recent emerging technologies in sensors and computational abilities, we have a broad and complete framework to achieve high-quality, expressive and ubiquitous speech and singing synthesis, with interactive manipulation possibilities. As a consequence, the idea of *performative voice synthesis* [3] has been reinvestigated within the digital instrument making context. Performative voice synthesis is the production of digital voice directly from realtime gestures, without any reference to textual input, or musical score in the case of singing. We can cite the Voicer [4] as a pioneer investigation. Our teams are investigating this research axis for the last ten years and are have been developing systems such as :

- the RAMCESS synthesizer : a concatenative voice synthesizer with realtime control of the phonetic stream (several vowels and consonants), the intonation and the voice quality [5];
- the HandSketch digital instrument : a tablet-based singing instrument giving a refined control on the voice quality parameters, allowing to play melodic lines with expressive quality [6];
- the MAGE platform : a platform for HMM-based speed and singing synthesis that provides both reactive contextual and prosodic control, allowing fast and easy prototyping [7, 8].

Human – Computer Interaction

The realtime control of voice properties starts to be more and more studied from the human-computer interaction (HCI) point of view. We find several research exploring the control and imitation of voice intonation with hand gestures both for speech [9] and singing [10]. The idea of replacing the text keyboard by a phoneme-based typing interface has been investigated, but initiators of this project are the firsts to approach artificial speech and singing control using and electrical guitar.

Augmented Guitar / Guitar as Controller

Guitar has maintained a close relationship with technological innovation throughout its history, from acoustic to electric and now to virtual [11]. Beyond the guitar, this virtuality has affected the way we play instruments in general. The appearance of the MIDI norm was a main event in the establishment of a communication between computers and traditional instruments. With an increased merging of the two, instruments have become more and more interfaces to control computational sound processing or (multi-)media elements. The example of a MIDI keyboard controlling a sampler is one of the many examples that can be found for several years in electronic music. The term "augmented instrument" is generally used to refer to a traditional (acoustic) instrument with added sonic possibilities. The augmentation can be physical like John Cage's sonic research on prepared pianos, but nowadays the term has acquired a more computational meaning: the use of digital audio to enhance the sonic possibilities of a given instrument as well as the use of sensors and/or signal analysis algorithms to give extra and expressive controls to the player [12]. Several elements on guitar playing techniques can be found in the literature: in [13] and [14] focus has been put on measuring or estimating the point where the string has been plucked, i.e the plucking point. In [15], left-hand fingering of a guitar player has been analyzed and characterized. Regarding the mapping side, several papers emphasize the use of an augmented guitar in an artistic context [16], [17], [18] or [19]. In most of these examples added sensors or analysis algorithms are mapped to control synthetic parameters. Recently, during the Numediart Guitar As Controller [20] project we have developed algorithms to detect the most common guitar playing techniques. Most of these algorithms are showing good results and are usable realtime so that the expressively of the instrumental is not lost during the analysis process.

Technical Description

In this project we aim at bringing several performers together and make them collaborating on heterogeneous devices – computers, cellphones, touch screens – and controllers, e.g. guitars, in order to arouse the possible parameters used for artificial voice production. This part of the proposal describes the different technologies that are envisioned and gives the main research and development axes that will be followed in order to build the new system. We also provide a more technical description of the environments and devices that will be used. Finally we also describe the project management strategies that will be deployed in the team.

Research and Development Axes

1. **Interactive Speech / Singing Synthesis (WP 1)**

   The first aspect of this research is to work on the recently introduced reactive HMM-based speech synthesis platform, called MAGE [7,8]. MAGE is a new C/C++ software platform for reactive speech synthesis. In this case, "reactive" means that both phonetic content and prosody of synthetic speech can be controlled by the user in realtime. It is using the HMM-based speech synthesis approach [21], and is more precisely based on the HTS system [22]. MAGE is a partial redesign of HTS, allowing the user to alter ongoing parameter trajectories of the speech production model. Combined also with a reactive natural processor module (RNLP) [23] enables the control of the targeted sentence itself. In summary, along with the enabled contextual control, here is a list of speech production parameters that can be altered on-the-fly:

   - pitch curve: for the ongoing label, the pitch information can be replaced by an arbitrary value or deviated from its HMM-generated value by a given ratio;
   - speed and duration: speech timing can be altered in various ways; label duration can be overwritten or deviated, scaled or shifted from its HMM-generated value; the speech speed can also be modified;
   - vocal tract length: the spectral envelopes corresponding to each label can be transformed so that it corresponds to a change of the speaker's vocal tract length.

   In spite of the very good results that MAGE has to present in terms of reactivity and quality of the final targeted output there are still some issues that need to be addressed. In fact, MAGE seems to have memory leak issues that need to be solved before embedding the platform into mobile devices. The synthesizer is very stable in environments with sufficient available memory, such as laptops, but not in mobile devices where memory is limited (WP 6).

   Instead of having only context and prosody control, we also aim at enriching the already existing control parameters provided from the first beta version of the platform. On the one hand we would like to embed the RNLP module into the platform and on the other hand we would like to enable reactive interpolation properties between various speaking styles or voices. Improving the memory management of the platform and adding new control parameters of the artificial voice, that can be used with the existing once or separately could lead to a new release of the platform.

2. **Detection Algorithms for Guitar Playing Techniques (WP 2)**

   The second aspect of the research is to enhance the recently developed playing techniques detection algorithms for the guitar. The different algorithms have been developed as Matlab files as well a C++ VST audio plug in. The enhancement will be of two types:

   - ameliorate algorithms results
   - generalize it to different type of guitar, i.e. electric guitar

   The last thing we want to test is the portability of these algorithms on mobile devices in terms of performance and quality (WP 6).

3. Mappings (WP 3)

One other significant aspect of the this project is the performance context in which we want to integrate the blending of the guitar controller and the voice synthesizer. Developing different mappings between the playing techniques of the controller and the available voice control parameters of the synthesizer will be explored. First, we would like to start with a natural sounding voice within a given context that will be transformed into sound for various performance cases. In other words, taking an existing phonetic context and placing it into a sound and musical context. Following an inverse process, we would like to begin by gradually building phonetic contexts, upon which we will apply different prosodic controls. These different mappings will be used to conduct case studies (WP 4).

Prototyping Cycle

As in any HCI application development, the team work flow is made of iterations between various phases, including research & development (described above) but also updating the case study (WP 4) and validating in a real performance context (WP 5). Updating the case study will consist in constantly refining the different scenario that are developed. The current starting points are:

1 - Improving the existing MAGE synthesizer and testing
the playing techniques algorithms with an electric guitar;
2 - blending of the guitar controller and the MAGE synthesizer in performance paradigms;
3 - building different mappings between the valuable components of the two systems;
4 - importing this final application also in mobile devices, rather than only personal computers.

Along with the iterative development of the platform, these starting points might evolve. We want to maintain a relevant amount of work in discussing these user experience scenario. within the team. The other aspect is evaluation of the system in a real performance context. As we are talking about a performing system, we will involve other participants of the workshop in creating small performing groups and/or attending the performance and we will collect various kinds of user feedback.

Software Environments

The development of the software will be achieved with openFrameworks. The openFrameworks project aims at providing an open-source creative environment for aggregating various software components, initially for computer graphics (OpenGL, image management, etc). One significant advantage of openFrameworks is to be cross-platform (Linux, Windows, iOS). Moreover the open and free add-on format attracted a lot of third party developers to wrap a large amount of new functionalities: OpenCV, XML, Open Sound Control, network protocols, etc. We aim at taking advantage of these existing modules for building real-case prototypes during the workshop. We also aim at contributing to this international open-source community with the various advances of the project (WP 6): voice synthesis, meaningful parameter mapping, GUIs, etc.

Facilities and Equipment

The team will essentially work with available devices brought by the participating lab: laptops, guitars, several iPods, iPhones and iPads. As performing the devices is an important part of the work, we might need a separate room for running the audio performances without disturbing the other groups. We would also require usual team work facilities, such as a projector, a screen, a whiteboard, meeting space, etc.

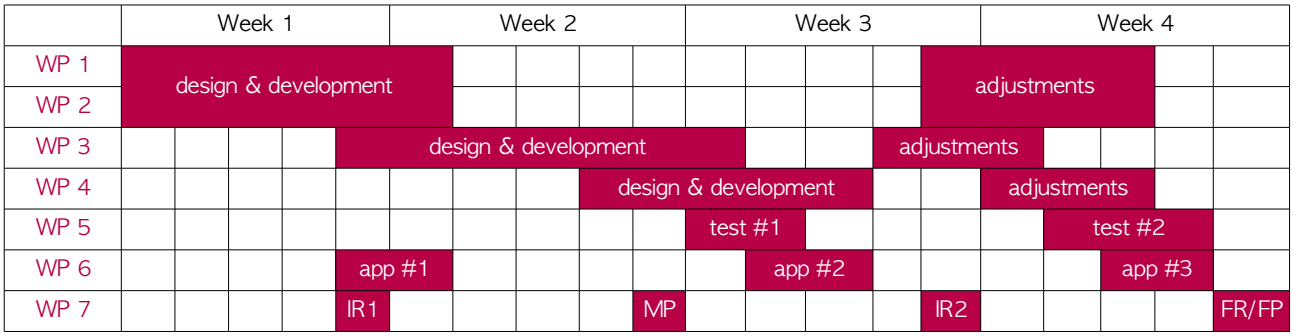
Project Management

The whole project will be supervised by Nicolas d'Alessandro, from the University of Mons. He should stay on the site of the workshop for the whole period. Based on the subscribed participants, sub-teams will be gathered around the specific work packages of the project. As it has been done during previous eNTERFACE workshops (2005-2011), we consider important that all participants can go for interdisciplinary interest and discussions, especially regarding the performing side of the project. Thus we will foster that everybody can perform the prototypes.

Schedule

In this part we gather the various work packages (WP N) that have been highlighted in the technical description and set them down on a one-month schedule. We also added one typical work package which concerns the reporting task (preparing slides, writing intermediate and final reports).

- WP 1 – interactive voice synthesis : improving and enriching the MAGE synthesizer, providing access to various voice parameter controls;
- WP 2 – guitar gesture recognition : testing if the playing techniques detection algorithms are guitar independent;
- WP 3 – mapping : developing various mappings between the guitar controller and the MAGE synthesizer;
- WP 4 – human-computer interaction models : prototyping a first paradigm based on user case studies for different mappings;
- WP 5 – performance / assessment : testing the prototypes, as a performer, as an audience;
- WP 6 – application integration : wrapping and sharing technologies as an application for both computer and mobile devices (iOS and android);
- WP 7 – reporting: preparing the intermediate and final documents (slides, reports).



IR1 : internal report #1;
MP: mid-term presentation;
IR2: internal report #2;
FR/FP: final report/presentation.

Benefits

In this part we describe what are the main deliverables and benefits that the team will provide at the end of the workshop :

1. A new major release of the MAGE synthesizer;
2. Assessment of the guitar technique detector to be guitar-independent;
3. Various mappings between guitar playing and speech / singing production properties;
4. Some platform-specific code for using the mappings and sound synthesis;
5. A scientific report will be provided at the end of the workshop.

Team Profile

Leader: Nicolas d'Alessandro

Staff proposed by the leader: Maria Astrinaki, Loïc Reboursière, Thierry Dutoit

Researcher profiles needed: voice analysis/synthesis, realtime audio software architecture, distributed architecture, human-computer interaction with some major interests in digital instrument making, linguistics, digital art performers, C/C++ development, iOS device development.

References

〔1〕 W. von Kempelen, Mechanismus der menschlichen Sprache nebst Beschreibung einer sprechenden Maschine, 1791.

〔2〕 H. Dudley. "*The Carrier Nature of Speech*," in Bell System Tech., 19:495–515, 1940.

〔3〕 N. d'Alessandro, B. Prichard, J. Wang and S. Fels, *"Interactve Manipulaton of Speech and Singing on Mobile Distributed Platorms,"* Proceedings of CHI 2011, Vancouver, May 2011.

〔4〕 L. Kessous and D. Arfb, *"Bi-Manuality in Alternate Musical Instruments,"* in Proceedings of New Interfaces for Musical Expression, pp. 140–145, 2003.

〔5〕 N. d'Alessandro, O. Babacan, B. Bozkurt, T. Dubuisson, A. Holzapfel, L. Kessous, A. Moinet, and M. Vlieghe, *"RAMCESS 2.x Framework - Expressive Voice Analysis for Realtme and Accurate Synthesis of Singing,"* Journal of Multmodal User Interfaces, vol. 2, no. 2, pp. 133–144, 2008.

〔6〕 N. d'Alessandro and T. Dutoit, *"Handsketch Bi-Manual Controller: Investgaton on Expressive Control Issues of an Augmented Tablet,"* in Proceedings of New Interfaces for Musical Expression, pages 78–81, 2007.

〔7〕 M. Astrinaki, O. Babacan, N. d'Alessandro, T. Dutoit, "*sHTS: A Streaming Architecture for Statistical Parametric Speech Synthesis"*, First International Workshop on Performative Speech and Singing Synthesis (p3s-2011), March 14-15, Vancouver, BC, Canada, 2011.

〔8〕 http://www.numediart.org/demos/mage_phts

〔9〕 C. d'Alessandro, N. d'Alessandro, J. Simko, S. Le Beux, F. Cetn, H. Pirker, *"The Speech Conductor: Gestural Control of Speech Synthesis,"* eNTERFACE'05 Summer Workshop on Multmodal Interfaces, Belgium, 2005.

〔10〕 N. d'Alessandro, C. Ooge, T. Dutoit, S. Fels, *"Analysis-by-Performance: Gesturally-Controlled Voice Synthesis as an Input for Modelling of Vibrato in Singing,"* Proceedings of the Internatonal Computer Music Conference, Huddersfeld, 2011.

〔11〕 G. Carfoot, *"Acoustic, Electric and Virtual noise: The Cultural Identity of the Guitar,"* Leonardo Music Journal, 16:35–39, 2006.

〔12〕 E. R. Miranda and M. Wanderley, New Digital Musical Instruments: Control and Interaction Beyond the Keyboard, *W. A-R Editions, Middleton, Computer Music and Digital Audio Series*, volume 21, 2006.

〔13〕 C. Traube and P. Depalle, *"Extraction of the Excitation Point Location on a String Using Weighted Least-Square Estimation of Comb Filter Delay,"* in Proceedings of the Conference on Digital Audio Effects (DAFx), 2003.

〔14〕 H. Pettinen and V. Välimäki, *"A Time-Domain Approach to Estimating the Plucking Point of Guitar Tones Obtained with an Under-Saddle Pickup,"* in Applied Acoustics, volume 65, pages 1207–1220, 2004.

〔15〕 E. Guaus and J. L. Arcos, *"Analyzing Left Hand Fingering in Guitar Playing,"* in Proceedings of SMC, 2010.

〔16〕 O. Lähdeoja, Une approche de l'instrument augmenté: La guitare électrique, *PhD Thesis, Ecole Doctorale Esthétique, Sciences et Technologies des Arts*, Paris 8, 2010.

〔17〕 M. Puckette, *"Patch for Guitar"*, in Proceedings of the Annual Pure Data Convention, 2007: http://crca.ucsd.edu/~msp/lac/.

〔18〕 L. Reboursière, C. Frisson, O. Lähdeoja, J. A. Mills, C. Picard, and T. Todoroff, *"Multimodal Guitar: A Toolbox for Augmented Guitar Performances,"* in Proceedings of New Interfaces for Musical Expression, 2010.

〔19〕 R. Graham, *"A Live Performance System in Pure Data: Pitch Contour as Figurative Gesture,"* in Proceedings of the Annual Pure Data Convention, 2001.

〔20〕 Loïc Reboursière, Otso Lähdeoja, Ricardo Chesini Bose, Thomas Drugman, Stéphane Dupont, Cécile Picard-Limpens, Nicolas Riche, *"Guitar As Controller,"* in QPSR of the numediart research program, 2011.

〔21〕 K. Tokuda, T. Kobayashi, and S. Imai, *"Speech Parameter Generation from HMM Using Dynamic Features,"* in Proc. of the IEEE International Conference on Audio, Speech and Signal Processing, pages 660–663, 1995.

〔22〕 H. Zen, K. Tokuda, and A. W. Black, *"Statistical Parametric Speech Synthesis,"* in Speech Communication, 51:1039–1064, 2009.

〔23〕 T. Dutoit, An Introduction to Text-to-Speech Synthesis, *Klumer Academic Publishers*, 1997.