Vokinesis : syllabic control points for performative singing synthesis

Samuel Delalez
LIMSI, CNRS, Université Paris-Saclay, B508,
rue John von Neumann, Campus Universitaire,
F-91405 Orsay, France
delalez@limsi.fr

Christophe d'Alessandro
Sorbonne Universités, UPMC Univ Paris 06,
CNRS, UMR 7190, Institut Jean Le Rond
d'Alembert, 4 place Jussieu, F-75005, Paris,
France
cda@limsi.fr

ABSTRACT

Performative control of voice is the process of real-time speech synthesis or modification by the means of hands or feet gestures. Vokinesis, a system for real-time rhythm and pitch modification and control of singing is presented. Pitch and vocal effort are controlled by a stylus on a graphic tablet. The concept of Syllabic Control Points (SCP) is introduced for timing and rhythm control. A chain of phonetic syllables have two types of temporal phases : the steady phases, which correspond to the vocalic nuclei, and the transient phases, which correspond to the attacks and/or codas. Thus, syllabic rhythm control methods need transient and steady phases control points, corresponding to the ancient concept of the arsis and thesis is prosodic theory. SCP allow for accurate control of articulation, using hand or feet. In the *Tap mode*, SCP are triggered by pressing and releasing a control button. In the *Fader mode*, continuous variation of the SCP sequencing rate is controlled with expression pedals. Vokinesis has been tested successfully in musical performances, using both syllabic rhythm control modes. This system opens new musical possibilities, and can be extended to other types of sounds beyond voice.

Author Keywords

singing synthesis, new musical instrument, performative syllable re-sequencing

ACM Classification

H.5.5 [Information Interfaces and Presentation] Sound and Music Computing—Signal analysis, synthesis, and processing, H.5.2 [Information Interfaces and Presentation] User Interfaces.

1. INTRODUCTION

The human voice is probably the most expressive and widely used musical instruments. A unique feature of the vocal instrument is its full embodiment: contrary to all the other musical instruments, all the control commands are internal. Although co-verbal gestures of the upper or lower limbs often happen in expressive singing performances, they have only limited effects on vocal production, and can be controlled or neutralized for stage direction purposes. Performative voice synthesis, real-time voice synthesis control (using hands or feet) is a new research paradigm for voice features analysis by synthesis [4], which raises many fundamental questions on the control gestures, control parameters, and finally prosodic representation. It is necessary to design new interfaces and new methods for synthetic singing, because common interfaces like keyboards, sliders, buttons etc, are not suited for it. Several performative singing synthesis systems have been proposed recently. They are using continuous bi-manual control surfaces, like graphic tablets [1, 5, 9] or continuous keyboard-like surfaces (Roli, Linnstrument, Soundplane, etc. according to the Multidimentional Polyphonic Expression (MPE) protocol). In these systems, the musician is playing mainly with vocalic or noisy vocal sounds, but with somewhat limited articulation capabilities. The real-time control of these vocal synthesizers is limited to voice quality, vowel quality and melodic dimensions. Real time external control of vocal articulation appeared very difficult, and is almost impossible because of the number and velocity of voice articulation organs: hands or feet are not able to reproduce the coordinated motions of tongue, lips and jaws for consonant production.

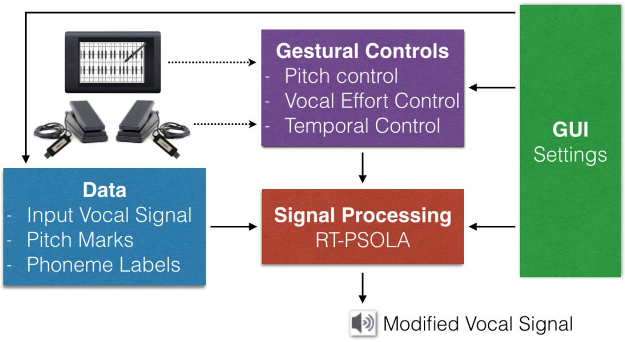

Figure 1: Vokinesis - System Overview

Natural voice modification allows for very natural sounding synthesis, and has been used with success in offline voice synthesizers such as Vocaloid [11]. In this research the question of articulated singing control is addressed with a different perspective, based on real-time control and modification of pre-recorded voice samples. Intonation and vocal effort are controlled using a graphic tablet, like in other voice instruments. Consonantal timing and rhythm are controlled thanks to syllabic sized chunks manipulation, using various methods, like tapping or continuous expression pedal motions. An overview of the system Vokinesis is displayed in Figure 1. A pre-recorded and labeled (pitch marks and phoneme labels) voice signal is modified with the RT-PSOLA algorithm [12], according to the musician's

NIME'17, May 15-19, 2017, Aalborg University Copenhagen, Denmark.
.

hands and feet gestures (see the accompanying videos for a quick introduction and overview of the possibilities of this system). In a first section real-time control of the various singing and voice parameters is presented, with some emphasis on rhythmic organisation of voice production. The Vokinesis system is described in Section 3. A discussion on application, assessment and future work is presented in Section 4.

2. CONTROLLING VOICE PARAMETERS

New control strategies are needed for performative singing synthesis, because internal voice controls must be externalized. The approach chosen here is to control some aspects of the voice signal (i.e. a phenomenological approach), in contrast with physical modelling (aiming at controlling physical aspects of voice production). This avoids the difficulty induced by a complete articulation control like in older performative speech synthesis systems [6, 8]. Another choice is to work directly with voice samples, and not with a terminal-analog voice synthesizer (like e.g in Cantor Digitalis). Methods for real-time control of some speech parameters are therefore needed. The parameters under control in Vokinesis are pitch, vocal effort and timing (for timing control, the player aims at time-instants in the original signal noted $t_o(t)$).

In summary, when playing Vokinesis, the synthesis signal corresponds to pitch, vocal effort and time scaling of an original signal, according to the pitch, vocal effort and target time-instants controlled by the player's gestures.

2.1 Pitch and vocal effort control

Pitch and vocal effort are controlled by the motion and pressure of a stylus on a graphic tablet. This very effective and intuitive melodic control method has been used in previous performative synthesis systems [5, 9], showing expressive and accurate control of intonation [3, 4, 7]. A printed mask representing the pitch scale (keyboard, guitar fretting, Indian raga mode) is attached to the graphic tablet for visual reference. Pitch is controlled by hand gestures similar to writing gestures. Note that accuracy can be enhanced by a sophisticated pitch tuning algorithm especially developed for stylus control [15]. Vocal effort (i.e. a combination of intensity and voice spectral tilt) is controlled by the pressure on the stylus.

2.2 Principles of timing control

Two main scales can be considered for voice rhythm [13] : intonative (or melodic) rhythm and syllabic rhythm. Rhythm at the intonational level has already been explored in synthesizers using a graphic tablet and only vocalic (or sustained) sounds [5, 9]. When articulation is also considered, intonational rhythm is built on the underlying syllabic rhythm. In the field of rhythm perception, it is widely acknowledged that the P-center (Perceptual Center) is able to define the syllable rhythmic event (or rhythmic beat), and is situated near the vowel onset [13, 16, 17]. Thus, controlling syllabic rhythm induces controlling the moment of occurrence of the P-centers, with one event for each syllable.

The most significant contribution of Vokinesis is a method for accurate control and sequencing of syllabic articulation timing. A new method is designed for accurate and intuitive time-domain manipulation of any prerecorded voice segments, at a phonemic level of detail. The aim is to be able to control fine articulation timing, but also to keep an excellent naturalness and sound quality. The controlled time scale is of the order of magnitude of a syllable, i.e. a minimum of about 80-100 ms. This time scale can be controlled e.g. by finger taping. Using continuous controllers (instead of tapping), it is even possible to decompose this time scale and to control articulation timing (a minimum of about 10 ms).

Table 1: Example of a word (1), along with its phonetic transcription (2), and its split into syllables (3) and arsis and thesis (4) (arsis are inside brackets)

(1)	Manual
(2)	_ m æ n j u ə l _
(3)	[_ m æ n] [j u] [ə l _]
(4)	[_ m] æ [n j] u [] ə [l _]

2.3 Syllabic Control Points

Syllabic rhythm depends on the syllable structure. The syllable is often described with three parts: the attack, the vocalic nucleus, and the coda. The attack and the coda correspond to one or more consonants, and the nucleus to the vowel. A syllable always contains a vocalic nucleus, but the attack and the coda are not necessarily present. For the purpose of rhythm production, this definition of the syllable, as a one-to-three phased unit, appeared not well suited

In an actual voice utterance, syllables are chained, and the attacks and codas of successive syllables correspond to the open and closure motions of the vocal apparatus, when the vowels correspond to the open positions. These cycles of opening and closing can be exploited for rhythmic control. Then the concepts of ”arsis” and ”thesis” (derived from Greek prosody) are very useful for our purpose. ”Thesis” represents the stable part of the segment, in our case the vowel or nucleus, and arsis represents the transient part between nuclei. The coda of one syllable and the attack of the next one (if they exist) are grouped to form the arsis. If there are no coda and no attack, the arsis still exists and corresponds to a short transition between two vowels. As an example, Table 1 shows the syllables and the arsis and thesis splits of the word ”manual”. This word is made of three syllables. It contains three thesis, but four arsis. Controlling syllabic rhythm induces controlling these seven time points.

We define the *Syllabic Control Points* (SCP) as temporal marking points for rhythm production. An example of SCP for the sentence ”My name is” (/majnejmɪz/) is displayed in Figure 2, top panel. *Vocalic Points* (Pv) are the SCP that corresponds to the vocalic nuclei or thesis, and *Transient Points* (Pt) those that correspond to the transient phases or arsis. These points define a target temporal location for each phase: when a vocalic phase is triggered (see sections 2.4 and 2.5), the target time-instant aims at the corresponding Pv until the next transient phase is triggered. Once this transient phase is triggered, the target time-instant evolves from the current Pv to the next Pt, and the synthesis signal will loop around this Pt until the next vocalic phase is triggered, and so on. On Figure 2, seven SCP are plotted on the spectrogram for the displayed sentence. Controlling the timing of these points allows for accurate rhythmic control while preserving the correct articulation.

2.4 Point rhythm control: Tap mode

The *Tap mode* is demonstrated in the second part of the attached video. In this mode, arsis and thesis are controlled by tapping a control button, as shown on Figure 2. Pressing the control button triggers a vocalic phase, while releasing it triggers a transient phase. At the beginning, the first Pt is selected (i.e. $t_o(t) = Pt(1)$). When the control button is pressed, the target time-instant evolves from the first Pt to the first Pv. Once this Pv is reached ($t_o(t) = Pv(1)$),

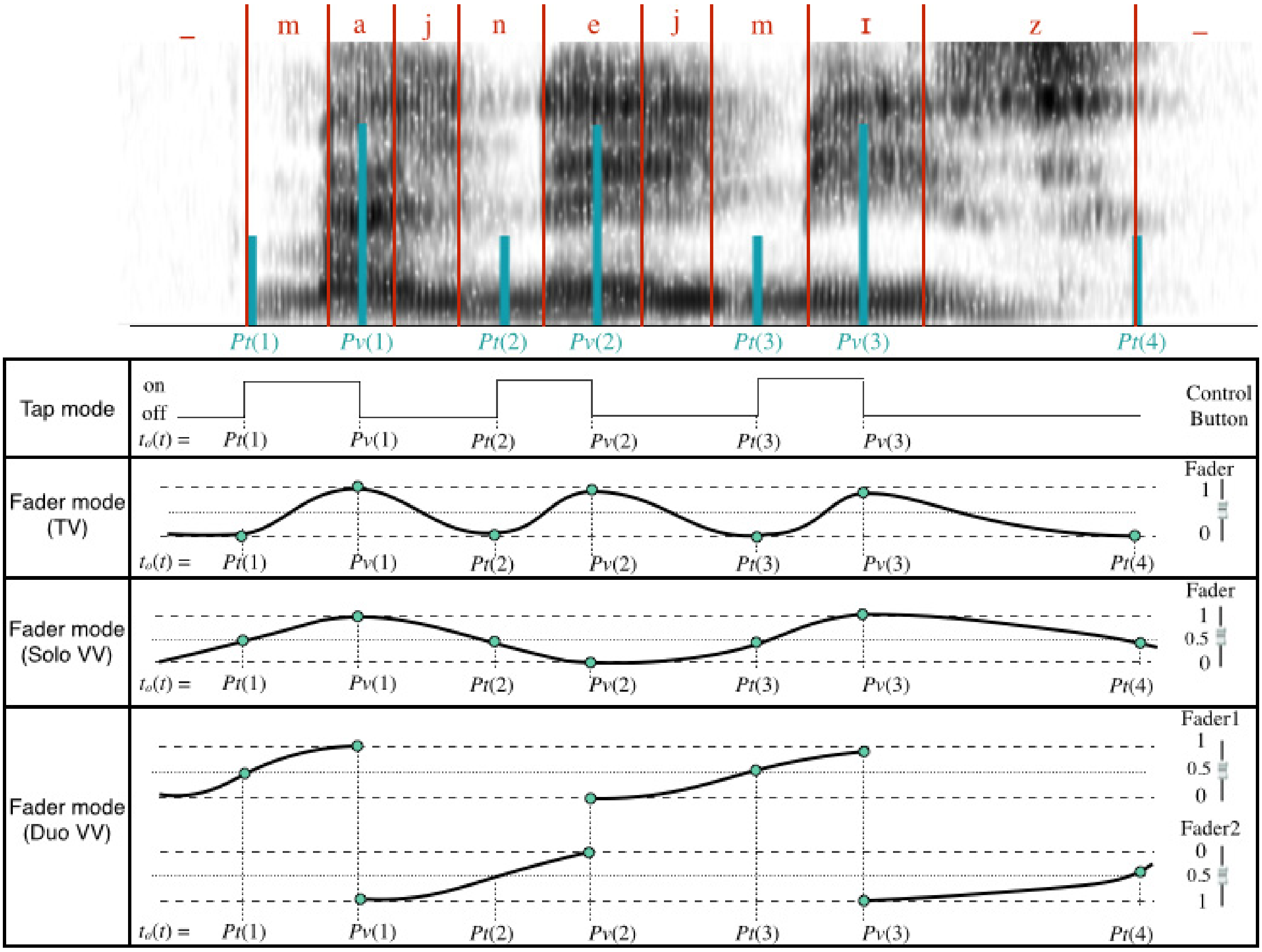

Figure 2: Syllabic Control Points and rhythm control for the sentence ”*my name is*”. Top panel: phonetic labels, spectrogram and syllable control points. Bottom panel: modes of rhythmic control (from top to bottom: tap mode, transient-vocalic mode, solo and duo vowel-transient-vowel modes).

the synthesis signal loops around the corresponding original sample until the control button is released. Then, the target time-instant evolves from the current Pv ($Pv(1)$) to the next Pt ($Pt(2)$). If the control button is pressed again, the target time-instant evolves from the current Pt ($Pt(2)$) to the next Pv ($Pv(2)$), and so on until the end of the original signal is reached. One syllable is pronounced by a release-pressure-release sequence. Note that samples are played at a predefined rate. This rate can be set independently for vowels, consonants and silences, but it can not be controlled in real time. Then in principle shortening of a recorded utterance can degrade the sound quality, as part of the signal may be truncated.

2.5 Continuous rhythm control: Fader mode

The *Fader mode* (demonstrated in the third part of the attached video) allows for more accurate rhythm control than the tap mode. Unlike in the *Tap mode*, the playback rate during transient phases can be varied continuously using a *fader* controller. Several types of faders have been tested. As the hands are busy with melodic control, expression pedals seem well suited. In addition, using hands and feet for independent pitch and rhythm control seems easier than controlling pitch with one hand and rhythm with the other.

Figure 2 shows the three fader modes available. In each mode, a fader has two extreme positions: position 0 and position 1. Moving a fader from an extreme position to another moves the target time-instant forward in the original signal.

In the first mode, Transient-Vocalic (TV) mode, moving the fader from position 0 to 1 performs a TV transition, moving it from 1 to 0 performs a VT transition. This fader mode is equivalent to the tap mode, but with controlled velocity for each phase. In the second mode, the solo Vowel-Transient-Vowel (VTV) mode, moving the fader from position 0 to 1 or from position 1 to 0 performs a VTV transition: one goes from a syllable nucleus to the next syllable nucleus with a single one-way displacement. In the third mode, the Duo VTV mode, two faders are used instead of one (e.g. two expression pedals). In Duo VTV mode, the faders are only active during their bottom to top displacement, and have to be used successively: moving Fader 1 from bottom to top performs the first VTV transition, then moving Fader 2 from bottom to top performs the second VTV transition, and so on.

It is important to note that even if the fader does not reach an extreme position, a change in direction in the second half of its way triggers the next transition.

Preserving transient phases is of uttermost importance for intelligibility. For this, the target time-instant reading rate is limited by a *maximum transient velocity* during transient parts, and a *catch-up velocity* accelerates the target time-

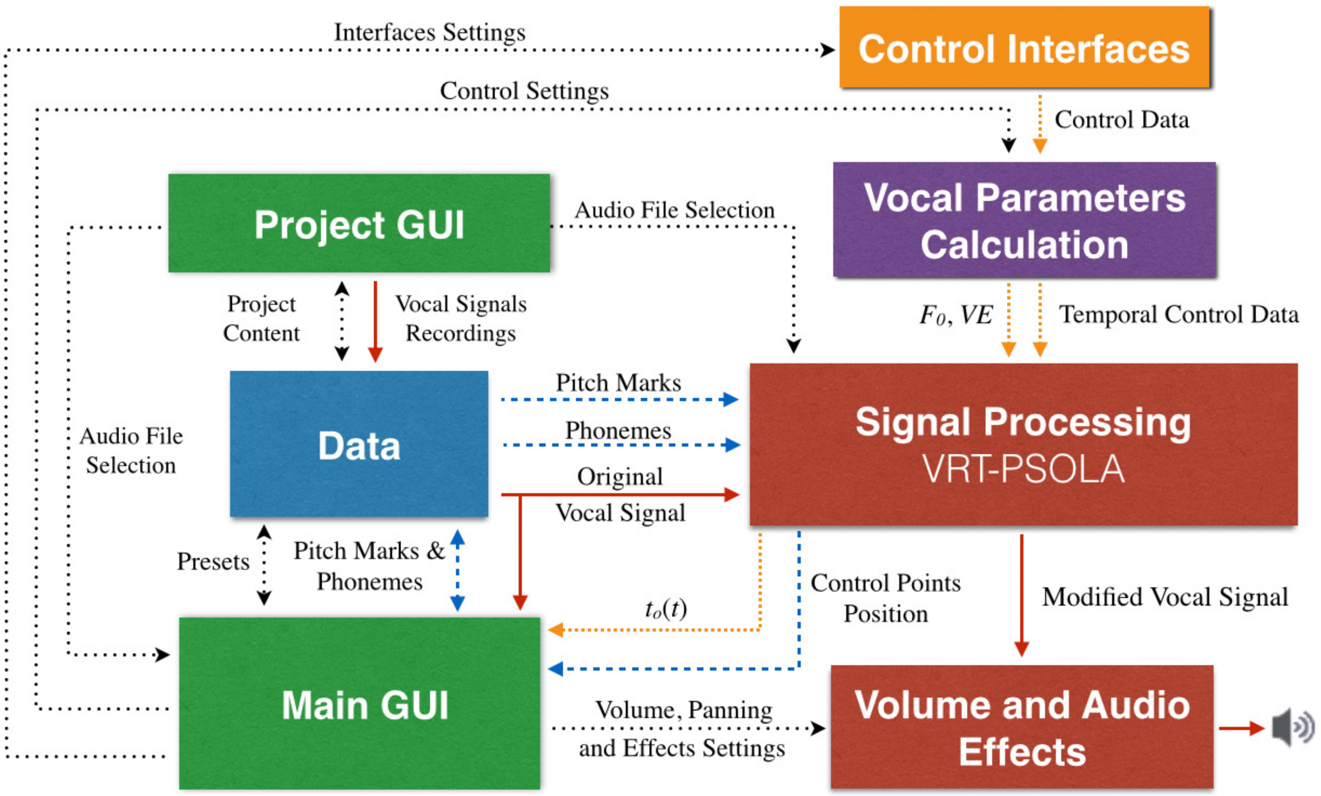

Figure 3: Vokinesis : System Architecture. Red (plain) arrows correspond to audio signals. Blue (dashed) arrows correspond to analysis data. Black (large-dotted) arrows correspond to configuration data. Orange (small-dotted) arrows correspond to real-time controlled data. F_0: Fundamental Frequency, VE: Vocal Effort, $t_o(t)$: target time-instant.

instant rate to catch up the fader position during steady vocalic parts. Both *maximum transient velocity* and *catch-up velocity* can be adjusted in the settings.

2.6 Polyphony

Vokinesis is not limited to a monophonic voice. As some graphic tablets also provide a *touch* function, fingers positions can be detected and exploited for a second voice. The pitch of the second voice is controlled by detecting the finger motion on the same surface as the stylus. Vocal effort is set for both voices by the pressure applied on the stylus, and then it is the same for both voices. An example of polyphony is displayed in the accompanying video.

3. VOKINESIS : SYSTEM PRESENTATION

Performative voice modification is carried out in three steps. The first step consists in preparing the input signals that will be modified and the corresponding analysis data. The next step consists in configuring the control interfaces and the playing modes, and the final step consists in re-synthesizing an original speech file according to its analysis data and to the musician's gestures. This section goes through these steps to present Vokinesis and its architecture, based on Figure 3. Then, an explanation of the the signal processing algorithms is provided.

3.1 Signals and data preparation

Vokinesis is a project oriented software: when the program is launched, a project folder needs to be loaded (or created) from the *Project GUI*. In this GUI, original speech files can be added to the project, and their names will be displayed in a table. A project folder contains a *ProjectTable.txt* file which stores the loaded speech files paths. It also contains a *data* folder, which stores the analysis data for each audio file of the project.

When an audio file is loaded in the project (or recorded), several analysis label tiers are needed: pitch period marks, phoneme boundaries and syllabic control points. Various methods for pitch period detection have been proposed. These labels can be obtained automatically with reasonable accuracy using phonetic software tools (e.g. Praat [2]). Phoneme boundaries can be placed automatically thanks to automatic speech alignment tools (e.g. easyalign [10]). Pitch marks and phoneme labels are stored in the *data* folder. SCP are determined from the phoneme boundaries: the Pv are placed in the center of the vowels boundaries, and the Pt are placed in the center of the last consonant of an arsis, or at the the center of a transition between two vowels.

A loaded file can be selected from the *Project GUI* in order to modify it. Its waveform, spectrogram, pitch marks and phoneme labels can be displayed in the *Main GUI*. The automatic pitch periods detection or phoneme alignment algorithms cited above can sometimes lead to inexact results. In this case the mistakes can be corrected by hand in the *Main GUI*.

3.2 Configurations

Although this paper focuses on using the system with graphic tablets for pitch and vocal effort control, and with expression pedals for syllabic re-sequencing, any MIDI keyboard or controller can be configured from the *Main GUI* and used to control any vocal parameter. In Figure 3, the *Interfaces Settings* allow to choose a control interface for such or

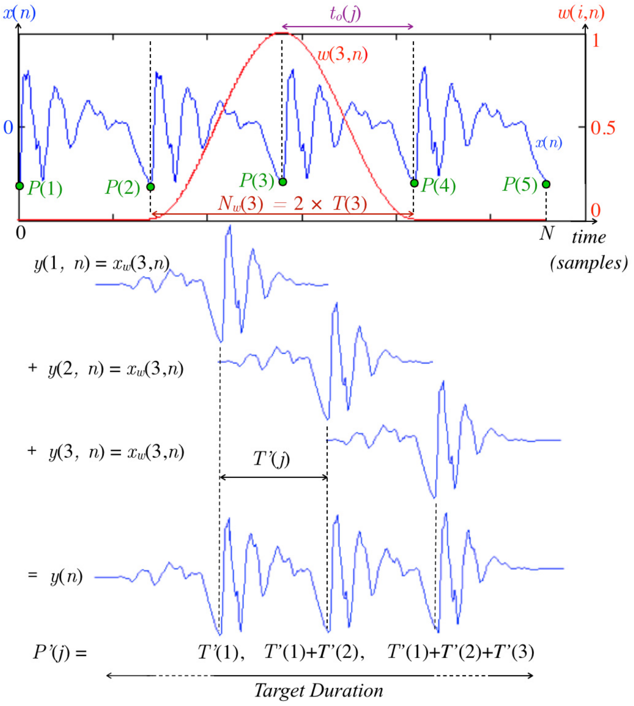

Figure 4: Principle of the VRT-PSOLA algorithm for voiced sounds. Pitch periods of the input signal are extracted, duplicated or withdrawn, and re-sequenced for obtaining a synthesis signal with desired pitch and duration.

such vocal parameter, and the *Control Settings* define how to use the selected interfaces (i.e. pitch range, duration control mode).

A set of configurations can be saved as a preset for each audio file that the project contains. Preset files are saved in the *data* folder: when a file is selected from the *Project GUI*, the corresponding configuration preset is loaded.

3.3 Signal control

The musician's gestures are captured and sent to the *Vocal Parameters Calculation* part. The control data is converted into vocal parameters according to the control settings made in the *Main GUI*. The *Signal Processing* part modifies the original vocal signal selected in the *Project GUI* according to the controlled vocal parameters, the phoneme labels and the pitch marks. The target time-instant $t_o(t)$ is calculated according to the temporal control data and is sent to the *Main GUI* for display. Original vocal effort is attenuated according to the selected controller value (e.g. the stylus pressure): if the stylus pressure is maximal, original vocal effort will not be attenuated.

Audio effects such as delay, reverb and equalization can be configured from the *Main GUI* and applied to the synthesis signals, which are then sent to the audio output.

3.4 Signal re-synthesis

Vokinesis synthesis engine is designed for real-time scaling of pitch, time and vocal effort. The input values for pitch, vocal effort and timing modification are provided by the player's gestures. The synthesis voice signal with corresponding pitch, vocal effort and target time-instant is computed by the Vokinesis Real-Time Pitch Synchronous Overlap-Add (VRT-PSOLA) algorithm. This method is based on re-sequencing of pitch periods on the original signal for re-synthesizing them with target pitch, vocal effort and timing. The original PSOLA algorithm [14] has been designed for speech synthesis. In the case of singing synthesis specific improvements are needed, because there are much larger variations in intonation and durations in singing than in speech. The real-time implementation of this algorithm, RT-PSOLA [12] has been used. VRT-PSOLA does not only allow to lengthen an original signal, but it also allow to hold any part of the original signal for an infinite duration.

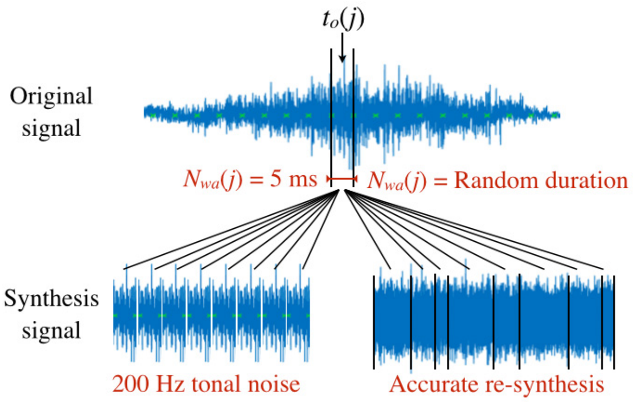

Figure 5: Principle of the VRT-PSOLA algorithm for unvoiced sounds. Random durations are used to avoid tonal noise in time scaled unvoiced sounds.

Figure 4 shows an example of the modification of a voiced (i.e. periodic) input signal $x(n)$ with the VRT-PSOLA algorithm. $P(i)$ corresponds to the i^{th} original pitch mark, and $P'(j)$ to the j^{th} synthesis pitch mark. $T(i)$ and $T'(j)$ correspond to the i^{th} original period duration and to the j^{th} synthesis period duration, respectively. The parameters that are controlled by the player are the target time-instant $t_o(j)$ and the synthesis period $T'(j)$. If $t_o(j)$ is between $P(i)$ and $P(i+1)$, the selected original short term signal will be $x(i,n)$, with $x(i,n) = x(P(i-1) : P(i+1))$. Each synthesis short term signal $y(j,n)$ is calculated for each synthesis period $P'(j)$ according to equation (1):

$$y(j,n) = w(j,n) \times ((1-\alpha) \times x(i,n) + \alpha \times x(i+1,n)) \quad (1)$$

where $w(i,n)$ is a hanning window with a size of $N_w(i) = 2 \times T(i)$ and $\alpha(j)$ is a weighting interpolation factor defined by equation (2):

$$\alpha(j) = \frac{t_o(j) - P(i)}{P(i+1) - P(i)} \quad (2)$$

Each short term signal is then added to the output signal $y(n)$ with a temporal spacing defined by $T'(j)$. In the figure, the player holds the double period around $P(3)$ for the *Target Duration*, with a synthesis period $T'(j)$ close to the original period $T(3)$.

As shown on Figure 5, repeating unvoiced (aperiodic) parts of an input signal in regular intervals leads to undesirable tonal noise. This has been improved by using random durations for each repetition of an unvoiced part.

The last improvement concerns vocal effort modification: original spectral envelope can be modified by the spectral tilt filter proposed in [9], as well as original intensity.

4. DISCUSSION AND CONCLUSION

4.1 Summary

Vokinesis is a performative singing synthesizer. To the best of our knowledge it is the only system that is able to sing articulated speech with full control on timing, pitch and vocal effort. Other performative systems are limited to a reduced set of speech sounds. The key features of the system are the introduction of phonetically informed Syllabic Control Points for accurate sequencing of voice samples, together with sophisticated methods for pitch and vocal effort scaling. Voice data, i.e. sound samples enriched with pitch marks, phoneme labels and SCP must be prepared in advance for the performance. The advantage is that any type of voice (or even labelled sound samples!) can be played with Vokinesis.

4.2 Assessment

Some features of Vokinesis have been tested in earlier studies. Pitch control using a stylus on a graphic tablet seems both easy to learn and accurate. In a prosody mimicry task subjects were asked to reproduce the intonation of original sentences with a graphic tablet and with their own voice [4]. The results showed that subjects were able to reproduce the intonation as well with the tablet as with their own voice. In the case of melodic accuracy in a musical context, it has been shown that subjects were as accurate (if not better) with the graphic tablet as with their own voice [3].

The pitch and time scaling algorithms in Vokinesis have been used in a speech synthesis experiment [7]. The aim was to compare the expressive quality of pitch contours obtained by gestural control on a graphic tablet vs computed with statistical models. It appeared that chironomic (i.e. hand controlled) stylisation of intonation was significantly more expressive than statistical modeling. In this experiment no SCP are used, time scaling is only controlled by varying the sample playback rate.

Vokinesis's *Tap mode* and *Fader mode* with two expression pedals have been recently demonstrated in public live performances. This showed that the system can be used successfully as a musical instrument, with unique capabilities.

4.3 Demonstration video

A brief demonstration of Vokinesis is presented in the accompanying video. The video begins with Karaoke example: a song is played with Vokinesis (with another melody) along with recorded music. In a second part, the tap mode is demonstrated. The space bar of the computer is used as a control button, and the graphic tablet for pitch and vocal effort. In the third part, the duo fader mode is demonstrated, focusing on the two expression pedals and (bare) feet.

4.4 Future work

Our current and future research projects are along 3 main lines. For assessing the rhythmic accuracy and precision achieved with Vokinesis, formal testing of the SCP concept in phonetic and musical tasks is in progress. Procedure for automatic SCP labeling, and sensitivity to the SCP positions are under study. Depending on the musical project, various organizations of the voice data can be studied and tested: bag of syllables, full song, selected phonemes etc. Vocal effort modification can be improved, and other voice variation capabilities can be added (e.g. whispering, tenseness and roughness of the voice, smiling voice etc.). Finally, deeper exploration of the creative applications of this kind of system will be adressed in composition and improvisation projects. The concept of SCP can be extended, beyond voice sounds, to any sound samples enriched with labels that can be used as SCP.

5. ACKNOWLEDGEMENTS

This work was supported by the Agence Nationale de la Recherche ANR, through the ChaNTeR project (ANR-13-CORD-0011, 2014-2017). Many thanks to René Durand (IJLRA) for shooting and editing the demonstration video.

6. REFERENCES

[1] M. Astrinaki. *Peformative statistical parametric speech synthesis applied to interactive designs.* PhD thesis, University of Mons, 2014.

[2] P. Boersma et al. Praat, a system for doing phonetics by computer. *Glot international*, 5(9/10):341–345, 2002.

[3] C. d'Alessandro, L. Feugere, S. Le Beux, O. Perrotin, and A. Rilliard. Drawing melodies: Evaluation of chironomic singing synthesis. *JASA*, 135(6):3601–3612, 2014.

[4] C. d'Alessandro, A. Rilliard, and S. Le Beux. Chironomic stylization of intonation. *JASA*, 129(3):1594–1604, 2011.

[5] N. D'Alessandro and T. Dutoit. Handsketch bi-manual controller: investigation on expressive control issues of an augmented tablet. In *Proceedings of the 7th international conference on New interfaces for musical expression*, pages 78–81. ACM, 2007.

[6] H. Dudley, R. Riesz, and S. Watkins. A synthetic speaker. *Journal of the Franklin Institute*, 227(6):739–764, 1939.

[7] M. Evrard, S. Delalez, C. d'Alessandro, and A. Rilliard. Comparison of chironomic stylization versus statistical modeling of prosody for expressive speech synthesis. In *INTERSPEECH*, 2015.

[8] S. S. Fels and G. E. Hinton. Glove-talk: A neural network interface between a data-glove and a speech synthesizer. *Neural Networks, IEEE Transactions on*, 4(1):2–8, 1993.

[9] L. Feugère, C. d'Alessandro, B. Doval, and O. Perrotin. Cantor digitalis: chironomic parametric synthesis of singing. *EURASIP Journal on Audio, Speech, and Music*, 2017.

[10] J.-P. Goldman. Easyalign: an automatic phonetic alignment tool under praat. 2011.

[11] H. Kenmochi and H. Ohshita. Vocaloid-commercial singing synthesizer based on sample concatenation. In *INTERSPEECH*, volume 2007, pages 4009–4010, 2007.

[12] S. Le Beux, B. Doval, and C. d'Alessandro. Issues and solutions related to real-time td-psola implementation. In *Audio Engineering Society Convention 128*. Audio Engineering Society, 2010.

[13] P. Mairano. *Rhythm typology: acoustic and perceptive studies.* PhD thesis, Università di Torino, 2011.

[14] E. Moulines and J. Laroche. Non-parametric techniques for pitch-scale and time-scale modification of speech. *Speech communication*, 16(2):175–205, 1995.

[15] O. Perrotin and C. d'Alessandro. Adaptive mapping for improved pitch accuracy on touch user interfaces. In *Proc. NIME 2013 Daejeon+ Seoul, Korea Republic*, pages 186–189, 2013.

[16] S. K. Scott. *Perceptual centers in speech - An acoustic analysis.* PhD thesis, University College London (University of London), 1993.

[17] P. Wagner. *The rhythm of language and speech: Constraining factors, models, metrics and applications.* PhD thesis, Habilitationsschrift, University of Bonn, 2008.