Chorus Digitalis: polyphonic gestural singing

Lionel Feugère, Sylvain Le Beux, Christophe d'Alessandro

Chorus Digitalis: polyphonic gestural singing

Lionel Feugère
LIMSI-CNRS
BP 133 - F91403
Orsay, France
lionel.feugere@limsi.fr

Sylvain Le Beux
LIMSI-CNRS
BP 133 - F91403
Orsay, France
sylvain.le.beux@limsi.fr

Christophe d'Alessandro
LIMSI-CNRS
BP 133 - F91403
Orsay, France
cda@limsi.fr

Abstract

Chorus Digitalis is a choir of gesture controlled digital singers. Chorus Digitalis is based on Cantor Digitalis, a gesture controlled singing voice synthesizer, and the Méta-Mallette, an environment designed for collective electronic music and video performances. Cantor Digitalis is an improved formant synthesizer, using the RT-CALM voice source model and source-filter interaction mechanisms. Chorus Digitalis is the result of the integration of voice synthesis in the Méta-Mallette environment. Each virtual voice is controlled by both a graphic tablet and a joystick. Polyphonic singing performances of Chorus Digitalis with four players will be given at the conference. The Méta-Mallette and Cantor Digitalis are implemented using Max/MSP.

Keywords: Singing Synthesis, Gestural Control, virtual choir, electronic music

1. Introduction

Chorus Digitalis is a choir of virtual singers, based of Cantor digitalis, a virtual singer controlled by a WACOM tablet and a joystick. The choir can integrate several singers on a same computer. The virtual singer has been improved, with individualized voices for different vocal types (baritone, tenor, alto, soprano), and source-filter interactions (the formant and fundamental frequency are tuned before synthesis). Although several virtual singers, including gesture controlled virtual singers, have been reported since many years, virtual choral experiment can scarcely found. Playing virtual voices in a choir is a very interesting and rewarding musical experience, reported in this workshop.

The paper is organized as follows. In section 2, Cantor Digitalis, is presented. It is an improved real-time singing synthesizer, including a sophisticated source component, specific features for different voice types, and mechanisms for dealing with source-filter interactions. Section 3 presents the integration of several singers in a same environment for collective electronic music performance, the Méta-Mallette. Section 4 reports on experiments on virtual choral singing.

2. Cantor Digitalis

Cantor Digitalis is a singing voice gestural synthesizer based on a source/filter model, including source/filter interactions.

2.1. Singer voices individualization

The characterization of a speaker is influenced by multiple factors, including the shape of his vocal tract and his glottal source, the cultural way of controlling his vocal system, his organ health, etc. In our synthesizer, the source parameter modification are mapped to higher vocal dimensions (i.e. tension, breathiness, roughness, vocal effort) in such way that it reproduces the behaviour of a natural glottis. The cultural way of controlling the glottis should ideally be found in the way the synthesizer player controls the interface.

We extracted the formant parameters of six speakers for the vowels /a, i, u/ from a LIMSI-CNRS database, where speakers were asked to produce vowels with different vocal efforts. For each of these six speakers, we arbitrarily associated a central value for breathiness (pink noise modulated by the glottal air flow wave) and tenseness (related to source parameters such as open quotient O_q and asymmetry coefficient α_m), in order to emphasize more each speaker identity. We also created five other speakers with the formants values of singers from the Ircam AudioSculpt software (soprano, alto, countertenor, tenor, bass) with /a, i, o/ vowels.

Most of the speaker formant frequencies, bandwidths and amplitudes were further tuned so as to give the most coherence between the three vowels of each speaker and between the speakers themselves (concerning intensity level), and so that it fits with our source model (good sound quality, not too strong resonances).

2.2. Source: RT-CALM Model

The real-time version of CALM source model of the glottal air flow wave [1] provides the source of our model.

To avoid discontinuities for high pitched notes, we use a high sampling rate of $8*44100Hz$. As it is well known from Fourier theory, the main side effect of oversampling is aliasing. Figure 2 shows a broader look at the version upsampled

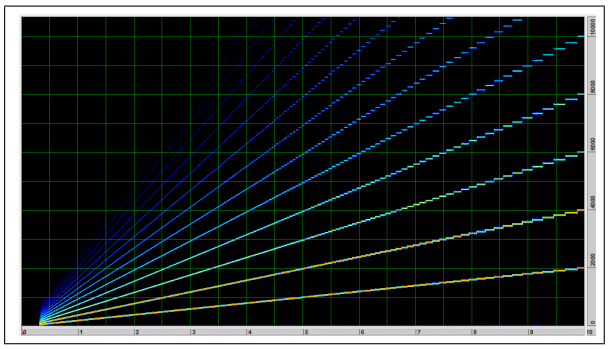

Figure 1. Spectrogram of glottal flow derivative for a sweep from 50 Hz to 2 kHz with an oversampling factor of 2.

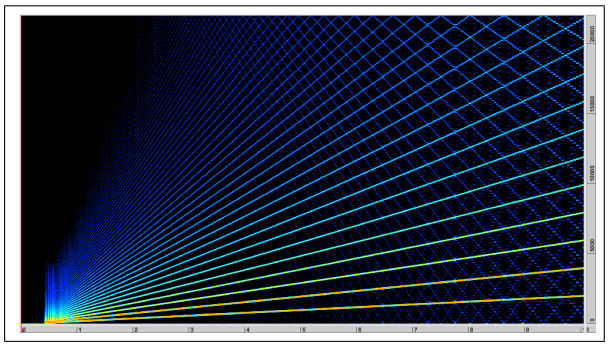

Figure 2. Visualization of aliasing with a background level fixed at -40 dB (oversampling x8)

by 8 (up to 20 kHz instead of 10 kHz) to highlight the effect of aliasing. Thus, in order not to lose in one side what we gain on the other, it is required to achieve band-limiting before downsampling to a proper sampling rate acceptable by the audio device (most likely 44.1 kHz). We use a spline transition band filter with 20 coefficients, from the filter design tools by Vaasko & Välimäki [12] in Matlab.

The passage from mechanism M_1 to mechanism M_2 (mechanisms as defined by Roubeau et al. [11]) in the former version of the synthesizer presented a strong perceived sound discontinuity. Lyrical singers are trained to learn how to dissimulate this transition. And those who master *voix mixte* (pitch range located between the two mechanisms) are able to produce a M_1 timbre while using M_2 mechanism and a M_2 timbre while using M_1 mechanism when they are singing in the overlap zone of M_1 and M_2 (Castellengo etal. [5]). In our synthesizer, in order to avoid modelling this timbre shift, which aim is to smooth the transition, we fixed the mechanism for the different pitch ranges that we provide. Three pitch ranges can be selected: G#2-G4 and G#3-G4 for which we associated mechanism M_1 and G#3-G5 for which we associated mechanism M_2. Thus, no mechanism transition is possible inside each pitch range.

Up to now, our synthesizer features a unique voice range profile for each mechanism based on experiments. Here, as the total pitch range is larger than in natural voice (the synthesizer features 4 octaves for each voice range profile), it induces that the spectral tilt becomes very high when going to highest pitches. Consequently, the signal pressure level reaches a very low value at low pitch, which is incompatible while using several voices with different pitch ranges at the same time: the low pitch range voice tends to be inaudible compared to higher pitch range (with higher vocal effort). A proper setting would require having a single voice range profile for each singer, associated with his own formants analysis data. It means that recordings of voice range profiles are needed for each register, data that we don't have at our disposal for the moment. In the meantime, in order to tackle this issue, we decided to reduce the spectral tilt maximum value allowing to achieve a voice signal loud enough in low pitch range.

2.3. Filter: four parallel resonant filters

The filter part of our source-filter model represents what happens in the vocal tract above the glottis, that is to say how the glottal waveform is transformed in the vocal tract from the pharynx to the leaps. The temporal evolution of the positions of the articulators (tongue, jaw, lips, uvula) creates cavities whose global shape corresponds to particular resonance frequencies. To reproduce the behaviour of the articulators, resonant filters are commonly used in formant synthesizers, characterized by their respective central frequency, amplitude and bandwidth.

The formants of each vowel correspond to the peaks of its spectral envelope. Thus, a given vowel is reproduced by setting the frequencies of the filters to its formants values. In this synthesizer, we use four formant filters.

2.4. Source Filter interaction

As any complex cavity, the vocal tract presents resonances located at several characteristic frequencies. This matter of fact can be used, for instance, to produce a louder voice: these cavities may be adjusted to increase the overall acoustic power by boosting particular frequency ranges (like for the so called "singer formant"). Conversely, it also happens that for a same vowel, when changing the pitch, we slightly moves our articulators in order to keep the pitch or its first order harmonics close to a formant resonance enabling a higher efficiency in terms of ratio of the voice acoustic level over vocal effort (i.e. vocal strength).

In our synthesizer, our aim is to get a voice intensity level rather constant when moving the pitch solely, and to avoid saturation in all cases. On an computational point of view, it is easier to smoothly modify the formants values when getting in resonance with source harmonics, rather than to arrange the formants frequencies constantly while the fundamental frequency moves. As a matter of fact, if we take into account the four formants frequencies together with the fundamental frequency of the glottis and its many harmonics, the resonances often occur.

An efficient way to avoid too frequent resonances is to take into account the first six harmonics of the fundamental

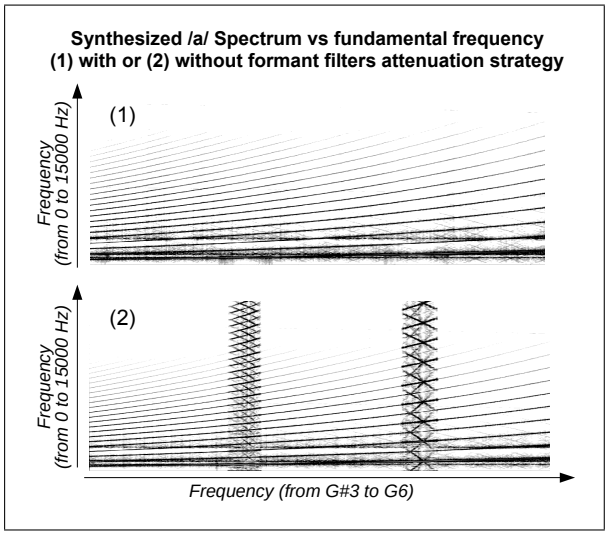

Figure 3. Source-Formant interaction for avoiding "harmonic whistling".

frequency as references to modify the formants amplitudes. In natural voice, one may suppose that the modification of articulator locations causes both a frequency and an amplitude shift of the formant resonances. Here, we choose to act only upon its amplitude, so as to not alter vowels too strongly. Thus, when F_0 or one of its first six harmonics approaches the frequency of one of the first three formants, then the amplitude of this particular formant is decreased gradually to reach its maximum reduction when exactly matching with F_0 or one of its harmonic values. The two parameters – the amplitude of reduction and the frequency interval where the decreasing occurs – depends on the formant index and continuously changes with F_0. These values were chosen empirically for our voice synthesizer, so as to hear the minimum amplitude shift resulting from the combination of the resonance and the applied correction. The difference with and without the attenuation is illustrated on Figure 3.

Many interdependent parameters are present in our implementation, and as the formant amplitude attenuation strategy is applied in the same manner whichever vowel or speaker is considered, it is not possible to remove them all without creating some artefacts. For instance, for some vowel or speaker configurations, the reduction of the formants can be perceived. However, our main goal is to achieve the most natural voice has possible, and soft resonances also happens with natural voices.

3. Chorus Digitalis

3.1. The Méta-Malette

The Méta-Mallette is a music software developed by Puce Muse. It targets the general public by proposing various computer music instruments (transformation, manipulation of samples, synthesis, ...) to be used in orchestra by the

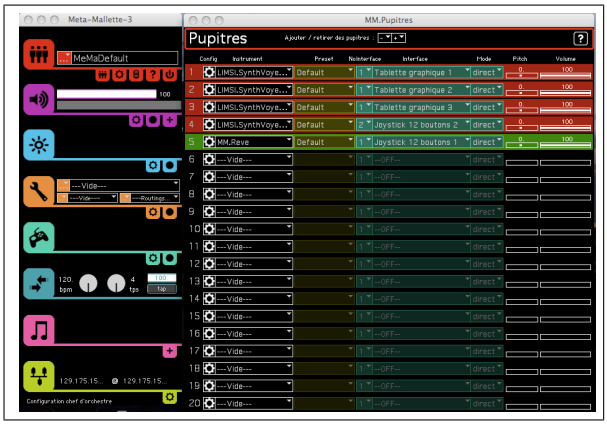

Figure 4. Screen shot of the Méta-Mallette, an environment for implementing electronic orchestra.

help of only one computer for all the orchestra and several USB interfaces. Each instrument has a dedicated number of Audio & video I/Os that can be linked each one another. The instruments are available from Puce Muse or can be developed by outside people, and shared in a library under different licences.

The present work is part of the European project OrJo, led by Puce Muse, around development of the Méta-Malette (see the screen shot of the software 5). Our role in this project is to create voice instruments to be added in the Méta-Mallette to be played together or with other instruments from Méta-Mallette. We saw the opportunity to use the Méta-Mallette environment to test our voice synthesis instrument in polyphony.

3.2. Use of the Méta-Mallette to play a virtual choir: the Chorus Digitalis

Each of the chorus musicians controls one voice synthesizer, with a particular singer characterization.

The graphic tablet is featured with a keyboard layer with visual and sensitive marks (see Figure 5). The pitch can be modified continuously on X axis (increasing from left to right) and a semi-tone scale is indicated visually by a keyboard sketch in relief thanks to relief printing at each notes. It enables users to get precise notes on a given musical scale and thus to play in polyphony in an easier way.

The pitch range can easily be changed (G#2-G4, G#3-G5, or G#4-G6) by pressing a button situated on the grip of the stylus. With the X and Y axes of the joystick, the non preferred hand controls tenseness, shimmer, jitter and breathiness. The speaker can be changed by pressing the joystick trigger, and three other buttons are used to choose the desired vowel.

In the Chorus Digitalis, a current limitation in the USB data flow does not allow us to use only one laptop. A time lag occurs if we have more than three voice synthesizers together.

Figure 5. Keyboard for controlling the virtual singer.

3.3. Others control mappings

In the Méta-Malette, others control mappings and control interfaces are available: voice quality with a joystick, diphonic singing with two graphic tablet.

4. Experiments in polyphonic Singing

The Chorus Digitalis quartet, composed of 4 musicians, 4 graphic tablets and 4 joysticks, has been recently formed (see Figure 6). The choir is able to play polyphonic choral music (e.g. Bach chorals or Renaissance polyphonic music), with a limited amount of training. The system suits well to musical games or other types of improvisation.

Each of the voice used is mapped to a different speaker to differentiate each one another. Besides, each of us plays in a different vocal register. Three use mechanism M_1 (bass, tenor and alto registers) and one other uses mechanism M_2 (soprano registers).

Extension of the choir to more than 4 voices is planned. Perceptual and performance experiments are planed, e.g. F_0 accuracy measurements while mimicking a given natural voice, learning to play the instrument for subjects with different musical backgrounds.

5. Acknowledgments

Chorus Digitalis is developed in the framework of the OrJO project funded by FEDER and the Région Ile-de-France, in collaboration with Puce Muse, UPMC and 3Dlized. The OrJo participants are Puce Muse (Serge De Laubier, Guillaume Evrard, Guillaume Bertrand), the UPMC (Hugues Genevois, Boris Doval, Vincent Goudard), and 3DLIZED (Philippe Gerard, Adel Keita).

Figure 6. The Chorus digitalis performing.

References

[1] D'Alessandro, N., d'Alessandro, C., Le Beux, S., and Doval, B., "Real-time calm synthesizer: new approaches in hands-controlled voice synthesis", Proceedings of the 6th International Conference on New Interfaces for Musical Expression (NIME'06), p. 266-271, June, 2006

[2] C. d'Alessandro, N. D'Alessandro, S. Le Beux, J. Simko, F. Cetin, H. Pirker "The Speech Conductor: Gestural Control of Speech Synthesis" , Proc. of eNTERFACE 2005 Workshop, Mons, Belgium

[3] N. D'Alessandro, P. Woodruff, Y. Fabre, T. Dutoit, S. Le Beux, B. Doval, C. d'Alessandro , "Realtime and accurate musical control of expression in singing synthesis", in *Journal on Multimodal User Interfaces*, 2007, Vol. 1, No. 1.

[4] N. D'Alessandro, B. Doval, S. Le Beux, P. Woodruff, Y. Fabre RAMCESS: Realtime and Accurate Musical Control of Expression in Singing Synthesis , Proc. of eNTERFACE 2006 Workshop, Dubrovnik, Croatia.

[5] M. Castellengo, B. Chuberre, N. Henrich, "Is voix mixte, the vocal technique used to smooth the transition across the two main laryngeal mechanisms, an independent mechanism?", in *International Symposium on Musical Acoustics*, 2004

[6] S. De Laubier, V. Goudard. "Puce Muse - La Méta-Mallette," *Journée d'Informatique Musicale*, 2008.

[7] S. Le Beux, "Contrôle gestuel de la prosodie et de la qualité vocale", Thèse de doctorat de l'Université Paris-Sud XI Orsay, France, Décembre 2009

[8] R. F. Orlikoff. "Vowel Amplitude Variation Associated With The Heart Cycle", in *Journal of Acoustic Society of America (JASA)*, Vol. 88, pp. 2091, 1990

[9] Puckette, M., and Zicarelli, D. Max/MSP. Cycling 74/IRCAM, version 5.1, 1990-2010.

[10] Juan G. Roederer, "The Physics and Psychophysisc of Music", 4th Edition, 2008.

[11] B. Roubeau, N. Henrich, M. Castellengo, "Laryngeal vibratory mechanisms: The notion of vocal register revisited", in *Journal of Voice*, 2009, 23(4):425–438.

[12] Vesa Välimäki, Timo I. Laakso, "Principles of Fractional Delay Filters" IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP00), Istanbul, Turkey, 59 June 2000.