Cantor Digitalis: chironomic parametric synthesis of singing

Lionel Feugère, Christophe d'Alessandro, Boris Doval, Olivier Perrotin

Feugère *et al. EURASIP Journal on Audio, Speech, and Music Processing* (2017) 2017:2
DOI 10.1186/s13636-016-0098-5

EURASIP Journal on Audio, Speech, and Music Processing

RESEARCH Open Access

Cantor Digitalis: chironomic parametric synthesis of singing

Lionel Feugère[1], Christophe d'Alessandro[1,2]*, Boris Doval[2] and Olivier Perrotin[1]

Abstract

Cantor Digitalis is a performative singing synthesizer that is composed of two main parts: a chironomic control interface and a parametric voice synthesizer. The control interface is based on a pen/touch graphic tablet equipped with a template representing vocalic and melodic spaces. Hand and pen positions, pen pressure, and a graphical user interface are assigned to specific vocal controls. This interface allows for real-time accurate control over high-level singing synthesis parameters. The sound generation system is based on a parametric synthesizer that features a spectral voice source model, a vocal tract model consisting of parallel filters for vocalic formants and cascaded with anti-resonance for the spectral effect of hypo-pharynx cavities, and rules for parameter settings and source/filter dependencies between fundamental frequency, vocal effort, and formants. Because Cantor Digitalis is a parametric system, every aspect of voice quality can be controlled (e.g., vocal tract size, aperiodicities in the voice source, vowels, and so forth). It offers several presets for different voice types. Cantor Digitalis has been played on stage in several public concerts, and it has also been proven to be useful as a tool for voice pedagogy. The aim of this article is to provide a comprehensive technical overview of Cantor Digitalis.

Keywords: Digital musical instrument, Voice synthesis, Gestural control, Singing voice

1 Introduction

Cantor Digitalis is a singing instrument, i.e., a performative singing synthesis system. It allows for expressive musical control of high-quality vocal sounds. Expressive musical control is provided by an effective human-computer interface that captures the player's gestures and converts them into synthesis control parameters [1, 2]. High-quality vocal sounds are produced by the synthesis engine, which features a specially designed formant synthesizer and an elaborate set of singing rules [3–5].

Cantor Digitalis is a musical instrument, and it is regularly played on stage by *Chorus Digitalis*[1], the choir of Cantor Digitalis. The expressiveness and sound quality of this innovative musical instrument have been recognized, as it was awarded the first prize of the 2015 International Margaret Guthman Musical Instrument Competition (Georgia Institute of Technology)[2]. Cantor Digitalis is distributed as a free software, accompanied by a detailed documentation. However, the scientific basis and technical details underlying Cantor Digitalis have never been published and discussed. The aim of the present work is to provide a comprehensive technical description of Cantor Digitalis, including the interface, the formant synthesizer, and the singing synthesis rules.

The sound synthesis components of Cantor Digitalis are in the tradition of formant synthesis. Apart from tape-based music using recorded voices and vocoders, synthetic voices first appeared in contemporary music pieces thanks to the "Chant" program [6]. "Chant" was based on a formant voice synthesizer and synthesis by rules, i.e., a parametric model of voice production[3]. Other research groups also proposed rule-based formant synthesizers [7, 8]. The main advantage of parametric synthesis is its flexibility and economy in terms of memory and computational load. The next generation of voice synthesis systems was based on recording, concatenation, and modification of real voice samples[4] or statistical parametric synthesis [9]. A formant synthesizer is preferred for Cantor Digitalis because flexibility and real time are the main issues for performative singing synthesis.

*Correspondence: cda@limsi.fr
[1]LIMSI, CNRS, Université Paris-Saclay, Bât 508, rue John von Neumann, Campus Universitaire, F-91405, Orsay Cedex, France
[2]Sorbonne Universités, UPMC Univ Paris 06, CNRS, UMR 7190, Institut Jean Le Rond d'Alembert, 4 place Jussieu, F-75005, Paris, France

Singing instruments have been proposed by different research groups [3, 10–13]. The graphic tablet has been proposed for approximately a decade [10, 14, 15] for controlling intonation and voice source variation. This interface appeared as a very effective choice. It has been extensively tested for intonation control in speech and singing synthesis [16, 17]. Additionally, this interface allows much expressiveness [18] because it takes advantage of the accuracy and precision acquired through writing/drawing gestures. This is the interface chosen for Cantor Digitalis.

This article describes the three main original components of Cantor Digitalis: the interface, the synthesis engine, and the rules for converting the input of the former into parameters for the latter. The next section presents the general architecture of Cantor Digitalis and the main issues and choices related to chironomic control of the singing voice. Section 3 presents the parametric formant synthesizer. Section 4 describes the synthesis rules, i.e., the transformation of parameters issued by the chironomic interface into synthesizer parameters. Section 5 discusses the obtained results, illustrated by audio-visual files, and proposes some directions for future work.

2 Chironomic control of the singing voice

The Cantor Digitalis architecture is illustrated in Fig. 2. It is composed of three layers: the interface, the synthesis/mapping rules, and the parametric synthesizer. This architecture follows the path of music production, from the player to sound. Initially, the musician plans to produce a given musical phrase, with a given vowel, given dynamics, given voice quality, and so forth. The planned musical task is then expressed through hand gestures related to the interface, i.e., through motions of a stylus and fingers on the graphic tablet (after selection of the voice type and other presets). The interface captures high-level parameters that are perceptually relevant to the player, such as the vowel quality or pitch. These high-level parameters are then converted into low-level synthesis parameters through a layer of synthesis/mapping rules. Low-level synthesis parameters drive the parametric voice synthesizer for sound sample production. The resulting sound is played back; listened to by the musician, who reacts accordingly; and the perception-action loop for performative singing synthesis is closed. Before addressing the control method itself, the control parameters must be identified.

2.1 Singing voice parameters

Cantor Digitalis is restricted to vocalic sounds (the case of consonants being considerably more difficult for real-time high-quality musical control [5, 19]). The corresponding parameters are pitch, voice force (or vocal effort), voice quality, and vowel label. The main perceived dimensions of voice quality are [20] voice tension (lax/tense voice), noise (aspiration noise in the voice resulting in breathiness and structural aperiodicities such as vocal jitter or shimmer resulting in roughness or hoarseness), and vocal tract size (or larynx height). All high-level parameters are listed below:

- Pitch P corresponds to the perceived melodic dimension of voice sounds. It is often the most important musical dimension.
- Vocal effort E corresponds to the dynamics, i.e., perceived force of vocal sounds. It is also an essential musical dimension.
- Vowel height H defines the openness or closeness of the vowel and corresponds to the vertical axis in the vocalic triangle, a classical two-dimensional representation for vowels. This dimension is related to the vertical position of the tongue, which depends on the aperture of the jaw.
- Vowel backness V defines the front-back position of the vowel and corresponds to the horizontal axis of the vocalic triangle. It is related to the position of the tongue relative to the teeth and the back of the mouth.
- The noise dimension in vocal sounds can be decomposed into two components. The first component is roughness or hoarseness R due to structural aperiodicities, i.e., random pitch period or amplitude perturbations. It defines the hoarse or rough quality of the voice.
- Breathiness B is the second noise dimension. Leakage at the glottis produces aspiration or breath noise in the voice. The extreme case of breathiness is whispering, with no fold vibration, resulting in unvoiced vowels.
- Tenseness T defines the tense/lax quality of the voice, i.e., the degree of adduction/abduction of the vocal folds.
- Vocal tract size S defines the apparent vocal tract size of the singer. Vocal tract size is singer dependent, but it also varies according to the larynx position or lips rounding for the same individual.
- Pitch range is singer dependent. A typical singer range is approximately 2 octaves. For simplicity, a unique pitch range size of 3 octaves is implemented. To play either low (e.g., bass) or high (e.g., soprano) voices, a pitch offset parameter P_0 is introduced.
- Laryngeal vibration mechanism M defines the vibration mode of the vocal folds used by the singer. Only chest and falsetto mechanisms are used, corresponding to $M = 1$ and $M = 2$, respectively.

All these dimensions are expressed in normalized units (between 0 and 1), except P_0 and M, and are summarized in Table 1. For most musical performances, only vowel label, pitch, and vocal effort are controlled with the

Table 1 High-level parameter control

Parameter	Voice dimension	Control
	Chironomic control	
P	Pitch	stylus x
E	Vocal effort	stylus pressure
H	Vowel height	finger y
V	Vowel backness	finger x
	Graphical user interface	
R	Roughness	preset, GUI
T	Tension	preset, GUI
B	Breathiness	preset, GUI
S	Vocal tract size	preset, GUI
P_0	Pitch offset	preset, GUI
M	Voice mechanism	preset, GUI

The corresponding synthesis parameters are indicated, along with their control types

help of chironomy. The other dimensions are controlled using a graphical user interface (GUI) on a computer. Note that other shares of parameter controls between the GUI and chironomy are possible; for instance, modulation of breathiness, voice tension, or vocal tract length can be assigned to the stylus (see Table 2 for musical examples).

2.2 Chironomic control: an augmented graphic-touch tablet

Following previous experiments, a Wacom Intuos 5M-touch tablet has been chosen as the interface, allowing for bi-manual chironomic control: the tablet detects the position of the pen pressure over the 2D plan, as well as the finger position over the surface. This interface is preferred for two main reasons. On the one hand, it is reactive, with no noticeable latency. The time resolution of the Wacom Intuos 5M tablets is 5 ms with a pen and 20 ms with a finger. This resolution proved short enough to provide the player with the feeling of direct causality between gesture and sound, similar to that for an acoustical instrument. On the other hand, this interface provides a fine spatial resolution, avoiding noticeable quantization effects in parameter variation. Wacom Intuos 5M tablets have a spatial resolution of 5080 lines per inch (0.005 mm) with a pen tip diameter of approximately 0.25 mm and 2048 levels of pressure. In addition, this interface allows for accurate, reproducible, and intuitive gestures. The pen tablet takes advantage of our writing ability, developed since childhood. The touch technology also takes advantage of the widespread habit of finger gestures on phones and computer tablets.

For increased intonation accuracy, the tablet is equipped with visual references. A printed template is superimposed on the active zone of the tablet with pitch and vowel targets. The top of Fig. 1 summarizes the *interface* part of Cantor Digitalis. Melodic accuracy and precision with this interface are comparable or better than those obtained by singers [17].

2.3 Voice source control

Melody and dynamics are the most important musical features. They are associated with the tablet's pen handled by the preferred hand to guarantee the best possible accuracy for intonation and dynamics. Pitch *P* is controlled by the X-stylus position, in a left-right organization similar to a keyboard. For accurate pitch targeting, the template attached on the tablet is carefully calibrated. It can represent a keyboard, a guitar fingerboard, or any specific melodic arrangement (e.g., the notes of a given mode or a raga), depending on the musical purpose. Examples of a "keyboard" template, with black and white keys, and a template for an indian modal scale (raga Yaman), are presented in Fig. 1.

For the flat and continuous tablet surface, a template representing the melodic scale is needed. The exact pitch (for each note of the melodic scale) corresponds to the printed key center line because whereas in a traditional keyboard the same pitch is associated to the whole key width, the pitch varies continuously according to the pen position on the tablet. The thick vertical lines correspond to the pitches of the chromatic keys, while the thin vertical lines correspond to the diatonic keys. Note that a dynamic intonation correction algorithm is available. This option can help less experienced users or for virtuoso passages [21].

Vocal effort *E* is the second main voice source parameter. It controls musical dynamics and does not need as much precision as pitch. The pen pressure has been chosen because of its analogy with vocal effort: a harder pressure corresponds to a higher vocal effort and a louder sound. Additionally, voice sound production occurs when the air flow of the lungs oversteps a phonation threshold. This threshold represents the sub-glottal pressure required to start vocal fold vibration. Thus, we introduced a vocal effort threshold E_{thr} under which no voiced sound is produced. A linear mapping between the stylus tip pressure and the vocal effort parameter appeared convenient.

2.4 Control of the vocalic space

Vowel label control is assigned to the non-preferred hand. The vocalic space is represented by a two-dimensional vocalic triangle or trapezium [22]. The two axes match the opening degree of the jaw (open-close vowel axis) and the position of the tongue in the mouth (antero-posterior vowel axis), respectively. The Wacom Intuos 5M-Touch tablet allows for the use of fingers at the same time as the stylus. The two dimensions of the vocalic space *H* and *V*

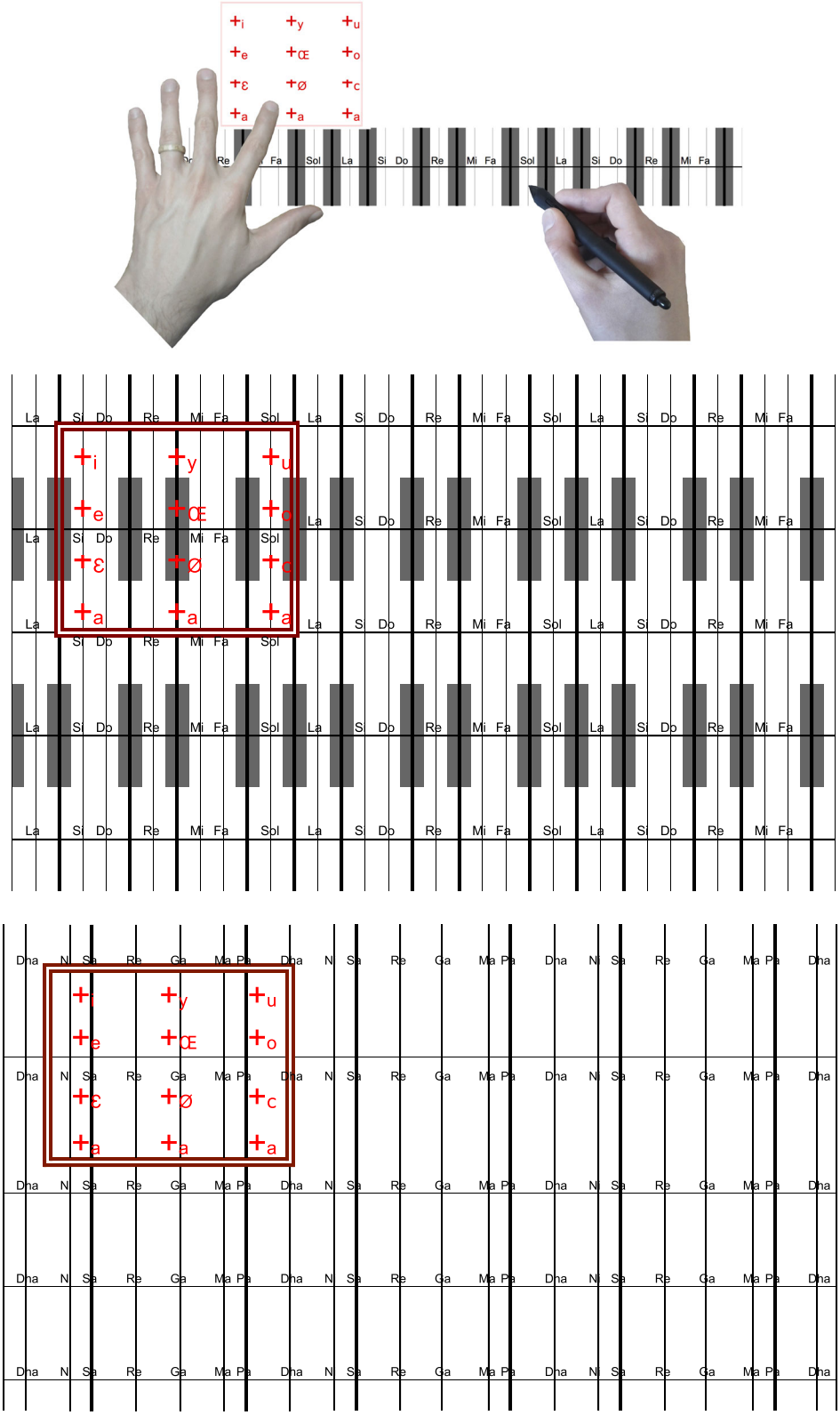

Fig. 1 Bi-manual control of the vocalic color (chosen in the 2D rectangle space with the finger position of the non-preferred hand) and the pitch (controlled with the preferred hand by the pen position along the X-dimension). *Middle panel*: full view of the tablet chromatic pattern. *Bottom panel*: full view of the tablet raga-Yaman pattern

Table 2 Examples of voice types, parameter variations, and parameter dependency rules, including public performance sequences (with corresponding sounds and videos)

	P_0	M	S	B	R	T	Additional file number and type
Parameter variations							
Changing vocal tract size	44	1	[0,1]	0.15	0.06	0.5	16 (sound)
Changing tension	44	1	0.29	0.15	0.06	[0,1]	17 (sound)
Changing breathiness	44	1	0.29	[0,1]	0.06	0.5	18 (sound)
Changing roughness	44	1	0.29	0.15	[0,1]	0.5	19 (sound)
Changing pitch	44	1	0.29	0.15	0.06	0.5	13 (video)
Changing effort	44	1	0.29	0.15	0.06	0.5	14 (video)
Changing vowels	44	1	0.29	0.15	0.06	0.5	15 (video)
Voice types							
Bass	32	1	0.21	0.2	0.06	0.5	20 (sound)
Tenor	44	1	0.29	0.15	0.06	0.5	21 (sound)
Alto	44	1	0.32	0.1	0.06	0.5	22 (sound)
Noisy Alto	44	1	0.33	0.3	0.06	0.5	23 (sound)
Soprano	56	2	0.35	0.1	0.06	0.5	24 (sound)
Noisy Soprano	56	2	0.41	0.3	0.06	0.5	25 (sound)
Bulgarian Soprano	56	1	0.53	0.1	0.06	0.66	26 (sound)
Baby	68	2	0.59	0.1	0.06	0.	27 (sound)
Gull	44		0.29	1	0.06		28 (sound)
Lion	8	1	0	0.7	0.2	0.5	29 (sound)
Didgeridoo	8	1	0	0.6	0	0.	30 (sound)
DesertBreeze	68	1	0	0.9	0.2	1.	31 (sound)
Whispering	56	1	0.35	0.6	0	0.8	32 (sound)
Woodbells	56	1	1	0	0.1	0.8	34 (sound)
Wind	56	1	0	1	0		33 (sound)
Live concert extracts							
Raga1	32	1	0.29	0.15	0.06	0.5	35 (video)
Raga2	32	1	0.29	0.15	0.06	0.5	36 (video)
Cold song	44	1	0.29	0.15	0.06	0.5	37 (video)
The lion sleeps tonight	m.c.	1	m.c.	m.c	0.06	0.5	38 (video)
Bulgarian song	m.c.	1	m.c.	0.1	0.06	0.66	39 (video)
Laugh	m.c.	m.c.	m.c.	m.c	0.06	0.5	40 (video)
Dependency rules							
Long-term perturbations (OFF/ON)	44	1	0.29	0.15	0.06	0.5	1, 2 (sounds)
Phonation threshold (OFF/ON)	44	1	0.29	0.5	0.06	0.5	5, 6 (sounds)
Laryngeal mechanism ($M = 1, M = 2$)	44	1, 2	0.32	0.15	0.06	0.5	3, 4 (sounds)
First formant tuning to E (OFF/ON)	44	1	0.29	0.15	0.06	0.5	7, 8 (sounds)
Formant frequency tuning to f_0 (OFF/ON)	56	2	0.35	0.15	0.06	0.5	9, 10 (sounds)
Formant amplitude attenuation (OFF/ON)	56	2	0.35	0.15	0.06	0.5	11, 12 (sounds)

"m.c." stands for "many configurations"

can be controlled by the two-dimensional positions of a finger, y and x, respectively. French vowels are represented in a specific area in the left-top corner of the tablet (see Fig. 1 and Table 3).

2.5 Control of the voice quality

Voice quality dimensions are controlled using a GUI on the computer screen. Each parameter (roughness R, tension T, breathiness B, vocal tract size S, and laryngeal

Table 3 Base vocalic formant center frequencies, bandwidths, and amplitudes

			Formant values of the generic voice (Hz, dB, Hz)														
			F_{i_G}					A_{i_G}					B_{i_G}				
Vowel	V	H	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
/i/	1	0	215	1900	2630	3170	3710	10	18	20	30	40	−10	−10	−8	−4	−15
/e/	1	1/3	410	2000	2570	2980	3900	10	15	20	30	40	−1	−3	−2	−2	−5
/ɜ/	1	2/3	590	1700	2540	2800	3900	10	15	30	50	40	0	−4	−5	−12	−24
/y/	1/2	0	250	1750	2160	3060	3900	10	10	20	30	40	−12	−9	−14	−11	−11
/œ/	1/2	1/3	350	1350	2250	3170	3900	10	10	20	30	40	−6	−3	−8	−8	−10
/ø/	1/2	2/3	620	1300	2520	3310	3900	10	10	20	30	40	−3	−3	−3	−7	−14
/u/	0	0	290	750	2300	3080	3900	10	10	20	30	40	−6	−8	−13	−8	−9
/o/	0	1/3	440	750	2160	2860	3900	10	12	20	30	40	−6	−1	−10	−6	−28
/ɔ/	0	2/3	610	950	2510	2830	3900	10	12	20	30	40	−3	0	−12	−15	−20
/a/	–	1	700	1200	2500	2800	3600	13	13	40	60	40	0	0	−5	−7	−24

The sixth formant is defined by $F_{6_G} = 2F_{4_G}$, $A_{6_G} = -15$ dB, and $B_{6_G} = 150$ Hz irrespective of the vowel

vibratory mechanism M) corresponds to a slider. Because only three octaves can be represented on the tablet, the pitch range is selected among seven possibilities. The voice dimensions used in Cantor Digitalis along with their control types are presented in Table 1.

3 Parametric formant synthesizer

3.1 Formant synthesizer architecture

The sound of Cantor Digitalis is computed by a formant synthesizer [23, 24], based on the linear model of speech production [25]. The main advantages of formant synthesis are its low computational cost, allowing for real-time processing, and the parametric representation of the vocal sounds, allowing for full control of voice type and voice quality. A new parallel/series formant synthesizer has been designed, and its general architecture is shown in Fig. 2 (bottom). According to the source-filter theory of speech production, the vocal sound $\mathcal{S}$ in the spectral domain is the product of a glottal flow derivative model $\mathcal{G}'$ and a vocal tract model $\mathcal{V}$. The glottal flow derivative model $\mathcal{G}'$ is composed of two elements: periodic pulses weighted by a factor A_g, filtered by the glottal formant response GF and the spectral tilt response ST, and a Gaussian white-noise $\mathcal{N}$, filtered by a bandpass filter NS and pondered by a factor A_n and the harmonic part GF × ST. The vocal tract model $\mathcal{V}$ is the sum of resonant filter responses $\mathsf{R_i}$ pondered by an anti-resonance filter response BQ.

$$\begin{aligned}\mathcal{S}(f) &= \mathcal{G}'(f)\,\mathcal{V}(f)\\ &= \left(\sum_n \delta(f - nf_0)\,\mathsf{GF}(f)\,\mathsf{ST}(f)\right.\\ &\quad \left. + A_n\left[\sum_n \delta(f - nf_0)\,\mathsf{GF}(f)\,\mathsf{ST}(f)\right] \otimes \left[\mathcal{N}(f)\,\mathsf{NS}(f)\right]\right)\\ &\quad \times \mathsf{BQ}(f)\sum_{i=1}^{5}\mathsf{R_i}(f)\end{aligned} \tag{1}$$

Note that the glottal source derivative is used for the source. Assuming that the lip radiation component of the speech production model can be modeled as a derivation and that the source-filter model is linear, the radiation component can be included directly in the source component. Figure 2 and each term of Eq. 1 are explained in detail in the following sections.

3.2 Voice source model

A parametric model of the glottal flow derivative, equivalent to the LF model [26], is used for the voiced source. The model is described in the spectral domain, according to previous results [27]. The spectral approach is well suited to real-time implementation because of a low computational load. The perceptive parameters of voice quality are genuinely linked to spectral descriptions, such as spectral richness or harmonic amplitudes.

3.2.1 A spectral model

The first version of Cantor Digitalis used the causal anticausal linear voice source model (CALM) [13, 28]. In the current version of Cantor Digitalis, a simpler version is used, computing only the magnitude spectrum of the glottal flow derivative (not the phase spectrum). The magnitude spectrum is a combination of a spectral peak, the glottal formant, and dynamic slope variation for high frequencies.

The model is described by five parameters: fundamental frequency f_0, glottal formant frequency F_g and bandwidth B_g, maximum excitation (glottal flow derivative negative peak) A_g, and high frequency attenuations T_{l_1} and T_{l_2} (or spectral tilt).

3.2.2 Glottal formant

For each voicing period, the glottal flow derivative is computed with two cascaded linear filters: the glottal formant

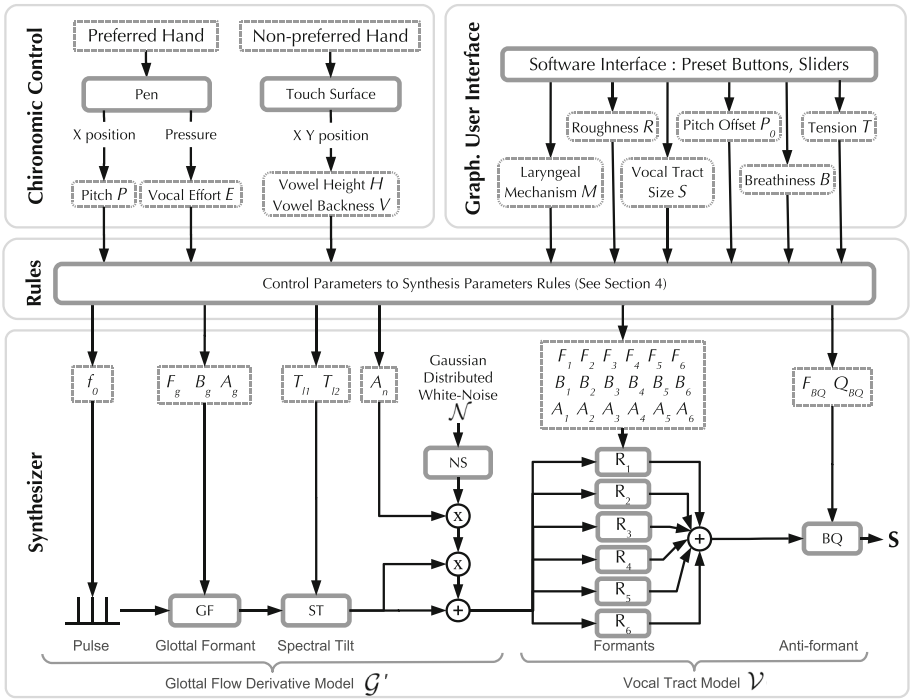

Fig. 2 The three main components of Cantor Digitalis: interface, rules, and synthesizer (see text for details)

filter represents the main source-related spectral peak, and the spectral tilt filter represents high-frequency voice variation. The glottal formant is computed in time domain as the impulse response of the first linear filter GF (GF box in Fig. 2).

The transfer function of the glottal formant is computed using a 2-pole 1-zero digital resonant filter in series with a 1-zero derivation filter standing for the lip radiation component (T_s being the sampling period, fixed at 1/96000 sec) [28] [5]:

$$\mathsf{GF}(z) = -\frac{A_g z^{-1}\left(1 - z^{-1}\right)}{1 - 2e^{-\pi B_g T_s}\cos(2\pi F_g T_s)z^{-1} + e^{-2\pi B_g T_s} z^{-2}} \tag{2}$$

3.2.3 Spectral tilt

A 2-pole 2-zero low-pass filter accounts for the spectral slope in high frequencies (ST box in Fig. 2). Its transfer function reads as (derived from [28])

$$\mathsf{ST}(z) = \mathsf{ST}_1(z) \times \mathsf{ST}_2(z) \tag{3}$$

where

$$\mathsf{ST}_i(z) = \frac{1 - (\nu_i - \sqrt{\nu_i^2 - 1})}{1 - (\nu_i - \sqrt{\nu_i^2 - 1})z^{-1}}, i = 1, 2 \tag{4}$$

$$\nu_i = 1 - \frac{cos(2\pi 3000 T_s) - 1}{10^{T_{l_i}/10} - 1} \tag{5}$$

with T_{l_i} $(i = 1, 2)$ corresponding to attenuation in dB at 3000 Hz.

3.2.4 Unvoiced source component

The unvoiced source component is computed using a Gaussian white noise $\mathcal{N}$ filtered by a wide band-pass second-order filter NS to simulate the effect of flow turbulence at the glottis. Turbulent noise sources can be modeled by a high-pass filter with a small spectral tilt in high frequencies [29]. We chose a second-order Butterworth filter with 1000 and 6000 Hz as cutoff frequencies. This noise, with amplitude A_n, is then modulated by the glottal flow derivative for mixed (noisy voiced) voice source qualities and added to it.

3.3 Vocal tract model

The vocal tract in Cantor Digitalis is computed with the help of a cascade/parallel formant synthesizer. This hybrid structure allows for fine adjustment of the voice spectrum and then fine control of the vowel quality, vocal tract size, and singer individuality.

3.3.1 Vocal tract resonances

The parallel components of the vocal tract filter are composed of six band-pass filters, each corresponding to one formant. Each formant is a 2-pole 2-zero digital resonator filter $\mathsf{R_i}$ with transfer function (formant central frequency F_i, formant bandwidth B_i, and gain A_i, $i \in [1, 6]$) [30]:

$$\mathsf{R_i}(z) = \frac{A_i\left(1 - e^{-\pi B_i T_s}\right)\left(1 - e^{-\pi B_i T_s} z^{-2}\right)}{1 - 2e^{-\pi B_i T_s}\cos(2\pi F_i T_s)z^{-1} + e^{-2\pi B_i T_s} z^{-2}} \tag{6}$$

The first three formants contribute to the vowel identification. The remaining three formants contribute to voice timbre. The "singing formant," described in the analysis of lyric voices, can be produced by grouping the third, fourth, and fifth formants [31].

3.3.2 Hypo-pharynx anti-resonances

In the spectrum of natural voice, one can observe anti-resonances at approximately 2.5–3.5 kHz and 4–5 kHz for vowels. The presence of the hypo-pharynx, composed of the laryngeal cavity and the bilateral piriform sinuses in the lower part of the vocal tract, appears to be primarily responsible for creating the anti-resonances. These spectral valleys are a clue for speaker identification but do not vary much between vowels of the same person [32, 33].

In Cantor Digitalis, an anti-formant is disposed in cascade after the parallel formant filters. A second-order fixed anti-resonance (BQ box in Fig. 2) is computed by a *notch filter*, a bi-quadratic second-order filter. Its transfer function is (quality factor Q_{BQ}, anti-resonance frequency F_{BQ}) [30]

$$\mathsf{BQ}(z) = \frac{1 + \beta_{BQ} z^{-1} + z^{-2}}{1 + \alpha_{BQ} + \beta_{BQ} z^{-1} + (1 - \alpha_{BQ}) z^{-2}} \tag{7}$$

where

$$\alpha_{BQ} = \frac{\sin(2\pi F_{BQ} T_s)}{2Q_{BQ}} \tag{8}$$

$$\beta_{BQ} = -2\cos(2\pi F_{BQ} T_s) \tag{9}$$

Finally, all the parameters of the synthesizer are summarized in Table 4.

Table 4 Low-level synthesis parameters

Parameter	System	Description	Unit
f_0	Voice source GF	Voice fundamental frequency	Hz
F_g	Voice source GF	Glottal formant center frequency	Hz
B_g	Voice source GF	Glottal formant bandwidth	Hz
A_g	Voice source GF	Voice source amplitude	1
T_{l_1}, T_{l_2}	Voice source ST	Voice source spectral tilt	dB
A_n	Noise source NS	Aspiration noise amplitude	1
F_1–F_6	Vocal tract R_1-R_6	Formant center frequency	Hz
B_1–B_6	Vocal tract R_1-R_6	Formant bandwidth	Hz
A_1–A_6	Vocal tract R_1-R_6	Formant amplitude	dB
F_{BQ}	Vocal tract BQ	Anti-formant center frequency	Hz
Q_{BQ}	Vocal tract BQ	Anti-formant quality factor	1

4 Voice dimensions to parameter mapping

In this section, the mapping between voice dimensions and synthesis parameters is detailed. Recall that voice dimensions are managed by the actions of the player on the chironomic interface and the GUI. Both interfaces can be used simultaneously and in real time. The chironomic interface is preferred for fast musical actions, such as playing notes of a melody, and the GUI is preferred for slower action, such as creating one's own voice character. The interplay between voice dimensions is rather intricate for some parameters in Cantor Digitalis. This is because as much knowledge as possible from the singing voice analysis literature has been incorporated in the mapping procedures, including formant tuning, vocal effort modeling, periodicity perturbations, voice mechanism modeling, and voice type settings.

4.1 Fundamental frequency

The fundamental frequency f_0 is mainly driven by the pitch voice dimension P defined by the stylus as well as by the pitch offset P_0 and several perturbations.

4.1.1 Pitch control

Pitch perception is very accurate, with the threshold for absolute pitch discrimination being in the case of synthetic vowels of approximately 5 to 9 cents (vowels with $f_0 = 80$ or 120 Hz) [34]. Considering the dimensions and resolution of the tablet, mapping approximately 3 octaves (35 semitones) of pitch to the X-axis corresponds to a pitch resolution of 0.08 cents for the smallest spatial step on the tablet (0.005 mm). Practically, the stylus tip width of approximately 0.25 mm corresponds to ±4 cents; this allows for an accuracy under the different limens for pitch perception.

In addition to the tablet X dimension, pitch is computed according to the pitch range of a given voice because only 35 semitones (ST) are represented on the tablet. The pitch offset parameter P_0 defines the pitch of the leftmost note on the tablet in semitones. $P_0 = 69$ corresponds to a fundamental frequency of 440 Hz. The absolute pitch linked to the control P_{abs} is (in semi-tones)

$$P_{\text{abs}} = P_0 + 35P \tag{10}$$

4.1.2 Jitter

Jitter, i.e., random perturbation of f_0, is useful for obtaining a hoarse voice quality. It is computed as a percentage of f_0 (in Hz) and is controlled by the roughness R voice dimension. In normal voices, jitter is generally less than 1%. However, in pathological voices, jitter can be as large as 5% (i.e., almost a tone) [35]. A maximum of 30% jitter is set here to go beyond the human limit. Jitter is computed with the help of a centered random Gaussian noise generator $\mathcal{N}_{\mathcal{R}}$ with unity variance.

4.1.3 Long-term f_0 perturbations

In addition to additive and structural noises, slow and small amplitude random perturbations of the source contribute to a more lively quality of the sound. These perturbations are due to the heartbeat and muscular instabilities [35–37].

Of course, in the case of expert singers, minimal perturbation is expected. Nevertheless, a small amount of perturbations still remain in the most experienced singer and may add naturalness to the synthetic voice. The perturbation due to the heartbeat can be identified in the sound pressure level and f_0 curves of natural voices. f_0 (in Hz) fluctuates up to 1% during a heartbeat cycle, and the voice amplitude fluctuates between 3 to 14%. Both depend on the mean vocal effort [38]. The additive amplitude and f_0 perturbation terms p_{heart} are modeled on a cardiac cycle as follows (with $\beta = 0.001$ ms^{-1} damping coefficient, A_{heart} amplitude depending on vocal effort, and f_c heartbeat frequency typically set to 1 Hz):

$$p_{\text{heart}} = A_{\text{heart}} e^{-\beta t} \begin{cases} cos\left(8\pi f_c t - \frac{\pi}{2}\right) & \text{for } t \in \left[0; \frac{1}{4f_c}\right] \\ cos\left(4\pi f_c t + \frac{\pi}{2}\right) & \text{for } t \in \left[\frac{1}{4f_c}; \frac{1}{f_c}\right] \end{cases} \tag{11}$$

When applied to f_0, A_{heart} is set to 0.15 semitone for low vocal effort ($E = E_{thr}$), 0.01 semitone for high vocal effort (E = 1) and is logarithmically interpolated for other values of E.

Other perturbations can be added. Under a f_0 variation of 5 Hz, they are not sufficiently slow to be considered as a controlled intonation fluctuation but sufficiently slow to be perceived as a pitch variation [37]. We added a pink noise to the pitch, whose amplitude is empirically limited to 0.2 semitone for low vocal effort ($E = E_{\text{thr}}$), to 0.01 semitone for high vocal effort ($E = 1$), low-passed at the cut-frequency of 5 Hz, and independent from f_0 (named p_{slow}). The filter output is reset every two cardiac cycles to avoid a deviation that is too high for a singing context.

In summary, f_0 is computed as follows (in Hz):

$$f_0 = 440 \cdot 2^{(P_0 + 35P + p_{\text{heart}} + p_{\text{slow}} - 69)/12} (1 + 0.3R\mathcal{N}_{\mathcal{R}}) \tag{12}$$

4.2 Voice source

The voice source parameters (amplitude of noise A_n, amplitude of source A_g, formant center frequency F_g, bandwidth B_g, and spectral tilts T_{l_1} and T_{l_2}) are computed as functions of the voice parameters P, T, E, B, and R.

4.2.1 Long-term voice amplitude perturbations

Similar to f_0, long-term perturbations also affect the sound level, from 3 to 14% deviation on amplitude voice signal [38]. However, for the perturbations to modify all the variables related to vocal effort ($A_n, A_g, F_g, B_g, T_{l_1}$, and T_{l_2}), they are applied at the output of control parameter E and not directly on A_g:

$$E_p = E + p_{\text{heart}} + p_{\text{slow}} \tag{13}$$

The heart perturbation p_{heart} has the same form as that for the f_0 perturbation (Eq. 11) with A_{heart} set to 0.1 for low vocal effort ($E = E_{\text{thr}}$), 0.02 for high vocal effort (E = 1) and is logarithmically interpolated for other values of E. The mathematical expression p_{slow} of long-term perturbations over E is chosen identically as for the one over f_0 (see above).

p_{slow} is empirically limited to 0.08 for low vocal effort ($E = E_{\text{thr}}$), and to 0.015 for high vocal effort ($E = 1$) [sound examples in Additional files 1 and 2][6].

4.2.2 Glottal formant central frequency and bandwidth

The glottal formant has a major influence on the relative amplitudes of the first harmonics. Its characteristics (center frequency and bandwidth) depend on the shape of the glottal flow derivative model, given by the glottal formant frequency F_g and bandwidth B_g. In CALM, the latter can be defined as a function of the open quotient O_q and asymmetry coefficient α_m [28]:

$$F_g = \frac{f_0}{2O_q} \tag{14}$$

$$B_g = \frac{f_0}{O_q \tan(\pi(1 - \alpha_m))} \tag{15}$$

O_q and α_m are expressed from the tension T and the perturbed vocal effort E_p (defined in Section 4.2.1). Furthermore, to distinguish between chest and falsetto registers, respectively produced in laryngeal vibratory mechanisms $M = 1$ and $M = 2$, two expressions of O_q and α_m are given:

$$O_q = \begin{cases} 10^{-2(1-O_{q_0})T} & \text{if } T \leq 0.5 \\ 10^{2O_{q_0}(1-T)-1} & \text{if } T > 0.5 \end{cases} \tag{16}$$

where $O_{q_0} = 0.903 - 0.426E_p$ for $M = 1$ and $O_{q_0} = 0.978 - 0.279E_p$ for $M = 2$, and

$$\alpha_m = \begin{cases} 0.5 + 2(\alpha_{m_0} - 0.5)T & \text{if } T \leq 0.5 \\ 0.9 - 2(0.9 - \alpha_{m_0})(1 - T) & \text{if } T > 0.5 \end{cases} \tag{17}$$

where $\alpha_{m_0} = 0.66$ for mechanism $M = 1$ and $\alpha_{m_0} = 0.55$ for mechanism $M = 2$. Figures 3 and 4 show the evolution of O_q and α_m as functions of T and E_p for each laryngeal vibratory mechanism. α_m is limited to 0.51 such that B_g does not reach 0 in the computer program.

The constants are chosen such that the standard ranges for O_q (for $T = 0.5$) are $[0.3, 0.8]$ for $M = 1$ and $[0.5, 0.95]$ for $M = 2$ [39], and 0.66 for α_m for $M = 1$ and 0.55 for $M = 2$. When the tenseness T decreases to 0, O_q increases to 1 and α_m decreases to 0.5. When the tenseness T increases to 1, O_q decreases to 0.1 and α_m increases to 0.9, and the values are set here to go beyond the human

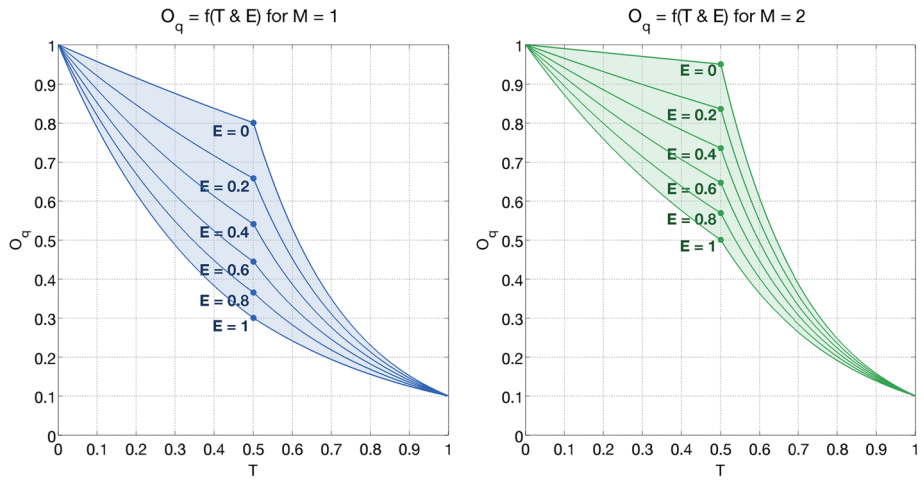

Fig. 3 Evolution of open quotient O_q according to vocal effort E and tension T for the two mechanisms

limit. The exponential function in Eq. 16 is deduced from Henrich et al. [40], who shows that the perception of O_q variations is proportional to O_q.

From these expressions and Eq. 12, one can compute the center frequency F_g and the bandwidth B_g as a function of P, P_0, M, T, and R, with Eqs. 14–15. The glottal formant center frequency F_g is generally situated below the first vocal tract formant F_1 and in the area of f_0. Moreover, it is proportional to pitch and depends on tenseness and vocal effort. As a simple rule, an increase in vocal effort and/or tenseness results in an increase of the glottal formant center frequency. Note that the effect of tenseness is larger on F_g (and B_g) and that it changes only the glottal formant center frequency and bandwidth, while E_p also changes the spectral tilt (see below).

As O_q varies between approximately 0.1 and 1, the variation of the glottal formant center frequency is between approximately $F_g \simeq 0.5f_0$ for the laxest voice quality and $F_g \simeq 5f_0$ for a very tense voice (see Fig. 5).

The glottal formant bandwidth B_g influences the first harmonic amplitudes relative to the higher harmonics. As α_m varies between approximately 0.51 and 0.9, the variation of the glottal formant bandwidth is between approximately $B_g \simeq 0.03f_0$ for the laxest voice quality and $B_g \simeq 31f_0$ for a very tense voice (see Fig. 5).

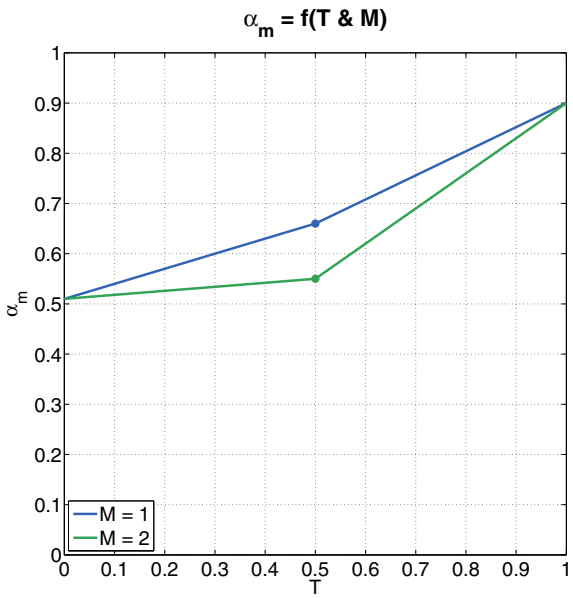

Fig. 4 Evolution of asymmetry coefficient α_m according to tension T for the two mechanisms

4.2.3 *Voice spectral tilt*

Pressure on the stylus controls the vocal effort, the main influence of which is to change the voice spectral tilt. The spectral tilt is inversely proportional to the pressure (i.e., a strong pressure corresponds to a low spectral tilt or a boost in high frequency). Spectral tilt is controlled by a series of two low-pass filters driven by two parameters T_{l_1} and T_{l_2} (see Eqs. 3 to 5). The spectral tilt parameter T_{l_1} varies in mechanism $M = 1$ (resp. $M = 2$) between 6 dB (resp. 9 dB) for a minimum tilt and maximum effort and 27 dB (resp. 45 dB) for a minimum effort and maximum tilt, whereas the spectral tilt parameter T_{l_2} varies in mechanism $M = 1$ (resp. $M = 2$) between 0 dB (resp. 1.5 dB) for a minimum tilt and maximum effort and 11 dB (resp. 20 dB) for a minimum effort and maximum tilt.

$$T_{l_1} = \begin{cases} 27 - 21E_p \text{ dB} & \text{for } M = 1 \\ 45 - 36E_p \text{ dB} & \text{for } M = 2 \end{cases} \tag{18}$$

$$T_{l_2} = \begin{cases} 11 - 11E_p \text{ dB} & \text{for } M = 1 \\ 20 - 18.5E_p \text{ dB} & \text{for } M = 2 \end{cases} \tag{19}$$

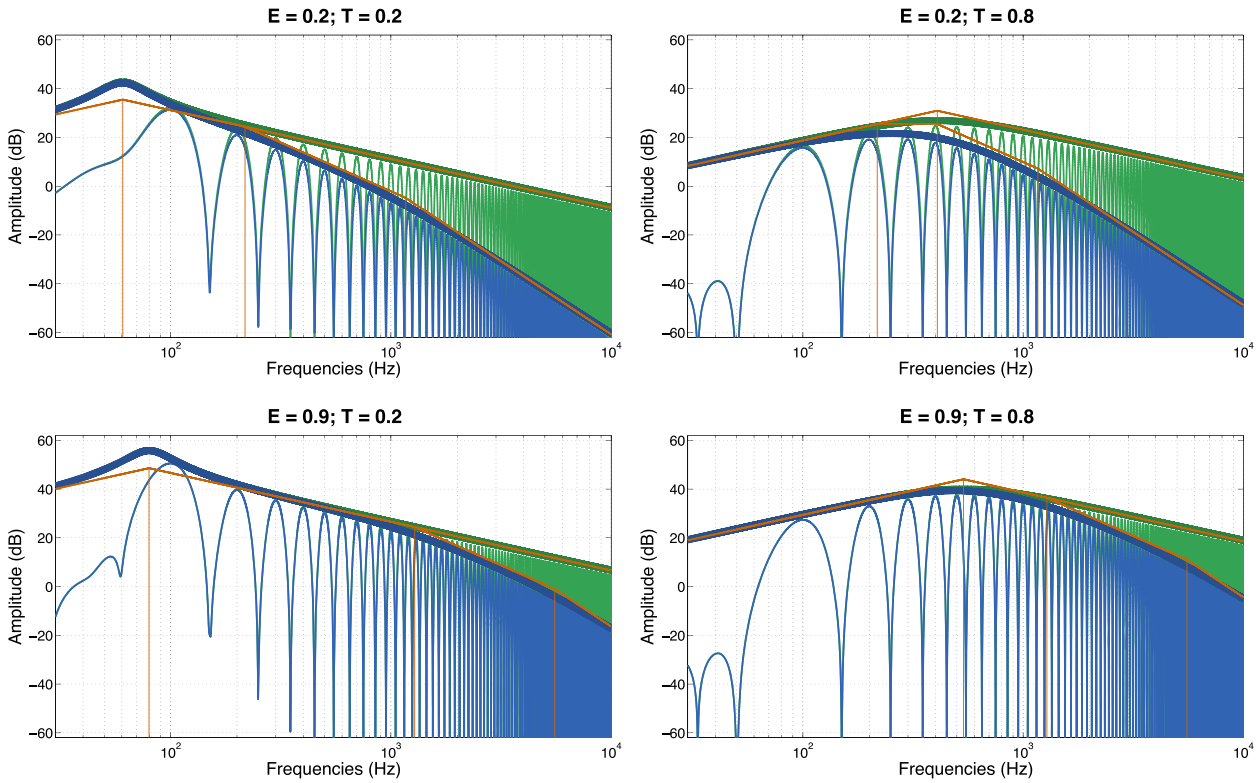

Fig. 5 Evolution of glottal flow derivative spectrum according to E and T. In *green* are the spectra of the glottal formant only. The glottal formant filter's frequency response is indicated with a *thick green line* as the envelope of the glottal formant's spectrum. In *blue* is the overall glottal spectrum including glottal formant and spectral tilt. The overall frequency response is represented with a *thick blue line* as the envelope of the overall glottal spectrum. *Top row*: low vocal effort ($E = 0.2$); *bottom row*: high vocal effort ($E = 0.9$); *left column*: low tension ($T = 0.2$); *right column*: high tension ($T = 0.8$). These spectra have been calculated for $M = 1$ and $f_0 = 100$ Hz

The sum of T_{l_1} and T_{l_2} corresponds to the attenuation (dB) of the glottal flow derivative at 3000 Hz.

Figure 5 displays the effect of vocal effort E and tension T on the glottal flow derivative spectrum with and without spectral tilt for 4 pairs of (E, T) values corresponding to low/high tensions and low/high vocal effort [sound examples in Additional files 3 and 4][7].

4.2.4 Voicing amplitude and shimmer

Voice sound production occurs when the air flow of the lungs exceeds a phonation threshold. This threshold represents the sub-glottal pressure required to start vocal fold vibration. Thus, the sound level for a vowel cannot be arbitrary small: there is a minimum amplitude step in vocal effort between silence and phonation with an hysteresis effect between starting and ending of phonation. The phonation threshold is set as $E = E_{\text{thr}} = 0.2$ [sound examples in Additional files 5 and 6][8].

Shimmer, i.e., random perturbation of A_g, is found in hoarse voice quality. A small amount of perturbations can be incorporated in voice amplitude A_g computation. Shimmer is computed as a percentage of A_g and controlled by the roughness R voice dimension. Although typical values are approximately 2.3% for a normal voice [41], for simulation of very rough voices, a maximum of 100% shimmer on A_g is allowed in the system. Shimmer is computed with the help of a centered random Gaussian noise generator $\mathcal{N}_{\mathcal{R}}$ with unity variance.

One can show that changing O_q, hence, changing F_g and B_g, in Eq. 2 has an effect on A_g. Then, A_g must be normalized by O_q. Additionally, a correlation between E and A_g is introduced. This is because in natural voice, the sound pressure level (SPL) depends on E. The chosen reference for sound level as a function of vocal effort comes from [42], extrapolated to sung voice: SPL $\simeq 39E + 60$(dB), i.e., approximately 40 dB between low and high vocal efforts. The parameter $C_{Ag} = 0.2$ in Eq. 20 represents the signal amplitude at the phonation threshold value. It can be modified from the GUI.

Finally, A_g can be computed as follows (where p_{hon} is a binary function equal to 1 if phonation is present or 0 if there is no phonation):

$$A_g = \begin{cases} 0 & \text{if } E_p \leq E_{\text{thr}} - 0.05 p_{\text{hon}} \\ \left((1 - C_{Ag})\frac{E_p - E_{\text{thr}}}{1 - E_{\text{thr}}} + C_{Ag}\right)(1 + R\mathcal{N}_{\mathcal{R}})/O_q & \text{if } E_p > E_{\text{thr}} - 0.05 p_{\text{hon}} \end{cases} \quad (20)$$

Note that although aspiration noise is modulated by A_g, it is not equal to 0 for $E_p \leq E_{\text{thr}} - 0.05p_{\text{hon}}$. Indeed, aspiration noise is expected to be produced below the phonation threshold of vocal effort (for $0 < E_p < E_{\text{thr}} - 0.05p_{\text{hon}}$). Then, the aspiration noise modulation is set to $\left((1 - C_{Ag})\frac{E_p - E_{\text{thr}}}{1 - E_{\text{thr}}} + C_{Ag}\right)(1 + R\mathcal{N}_{\mathcal{R}})/O_q$ regardless of E_p. For simplicity, it is not mentioned in Fig. 2 or Eq. 1.

4.2.5 Noise amplitude

The breathiness dimension B directly controls the A_n parameter, i.e., noise amplitude or the amount of aspiration or breath noise in the voice source. Voicing can be switched off by a voiced-unvoiced command. This allows for breathy vowels without any periodic component. The relation between B and A_n is directly given by:

$$A_n = \begin{cases} B & \text{if voicing is on} \\ 1.5E_pB & \text{if voicing is off} \end{cases} \quad (21)$$

When voicing is on, the factor 1 is empirically set to have a maximum signal-to-noise ratio of approximately −12 dB for standard voices. When voicing is off, a dependency on E_p is added to enable control over the loudness of the signal.

4.3 Vocal tract formants

4.3.1 Generic formant values

Almost all the chironomic or GUI control parameters have an effect on vocal tract formants: vowel, voice quality dimensions, pitch and vocal effort. The different voices are computed using generic formant center frequencies F_{i_G}, −3-dB-bandwidth B_{i_G}, and amplitude A_{i_G} ($i \in [1, 6]$, G stands for "generic") reported in Table 3 for $H, V = \{0, 0.5, 1\}$. These values have been measured for a tenor voice singing at a comfortable pitch and vocal effort level. Note that formant values measured for other singers can also be used and can be easily edited using the GUI.

The ten chosen vowels (/i,y,u,e,ø,o,ɜ,œ,ɔ,a/) are sufficient for computing the entire vocalic space. The other vowels (i.e., $H, V \neq \{0, 0.5, 1\}$) are computed using a 2-D interpolation between the four closest canonical vowels in the space formed by the vowel height H and vowel backness V dimensions. These values are defined for any $H, V \in [0, 1]$. H and V are controlled by the finger position of the non-preferred hand on a vocalic triangle printed at the top-left corner of the graphic tablet.

4.3.2 Vocal tract length

The vocal tract length is an important factor that influences vocal identity. Male vocal tracts are on average longer than female vocal tracts because of anatomical differences. The longer the vocal tract is, the lower its formant frequencies. Voices corresponding to different vocal tract sizes are created by multiplying the formant central frequencies by the same factor. The vocal tract size parameter S is mapped to a vocal tract scale factor α_S ranging from 0.5 to 2.2 with the linear equation:

$$\alpha_S = 1.7S + 0.5 \quad (22)$$

4.3.3 Larynx position adaptation to f_0

A modification of approximately 10% of the formant positions is noticeable between f_0 = 200 Hz and f_0 = 1000 Hz [43]. This is achieved by multiplying the central frequency of all formants by a factor K, depending on f_0, with $K(f_0 = 200\text{ Hz}) = 1$ and $K(f_0 = 1000\text{ Hz}) = 1.1$ (which is equivalent to modifying the length of the vocal tract):

$$K = 1.25 \cdot 10^{-4} f_0 + 0.975 \quad (23)$$

Vowel height H and vowel backness V are used to find the closest vowel of the generic voice. Then, the formant center frequencies are obtained by the generic formant values F_{i_G} ($i \in [1, 6]$) from Table 3, scaled by the vocal tract size scale factor α_S and larynx position factor K:

$$F_i = K\alpha_S F_{i_G}(V, H) \text{for } i \in [1, 6] \quad (24)$$

4.3.4 First formant tuning

The main control parameter for F_1 is vowel height H. In the vocalic triangle, F_1 represents the vertical dimension. However, vocal backness V also has a slight influence. In addition to Eq. 24, the first formant center frequency depends on other parameters: f_0 and vocal effort E.

In speech, increased vocal effort results in a higher first formant frequency F_1. F_1 increases at a rate of approximately 3.5 Hz/dB on average for French oral and isolated vowels [44]. Extrapolating this result for singing voice, a rule for automatic F_1 and vocal effort dependency is implemented. For our generic tenor voice at $f_0 = 200$ Hz, the sound level varies by approximately 40 dB between $E = 1$ (maximum value) and $E = E_{\text{thr}}$ (phonation threshold). The generic F_{1_G} for this voice corresponds to a medium vocal effort $\left(E = \frac{1 - E_{\text{thr}}}{2}\right)$. Then, for a 3.5 Hz/dB increase, the dependency rule between F_1 and E must satisfy $F_1\left(E = \frac{1 - E_{\text{thr}}}{2}\right) = K\alpha_S F_{1_G}$, and $F_1(E = 1) - F_1(E = E_{\text{thr}}) = 40 \times 3.5$ Hz. This corresponds to the term "$\frac{140}{1 - E_{\text{thr}}}E - 70$ Hz" in Eq. 25. Note that as in natural voices, the vowel identity tends to disappear for high pitch, with all vowels becoming close to each other [sound example in Additional files 7 and 8][9].

Singers can adapt their two first vocalic formants as a function of f_0 and its harmonics to exploit the vocal tract resonances as much as possible. The effect is to increase the sound intensity [43, 45]. Soprano singing /ɑ,o,u,ɛ/ vowels with a low vocal effort tend to adjust their first formant with the first harmonic f_0 [10]. The first formant is tuned to the first source harmonic $F_1 = f_0 + 50$ Hz above a pitch threshold. Of course, for very high f_0, formant tuning is no longer possible because the fundamental frequency

is well above the possible first formant frequency. In summary, F_1 is computed according to the following equation:

$$F_1 = \max\left(f_0 + 50\text{ Hz}, K\alpha_S F1_G(V,H) + \frac{140}{1-E_{\text{thr}}} - 70\text{ Hz}\right) \quad (25)$$

4.3.5 Second formant tuning

The main control parameter for F_2 is vowel backness V, which is the horizontal dimension in the vocalic triangle. The vocal tract length factor α_S modifies the formant frequency proportionally.

For high pitched voices, $2f_0$ and F_2 can come close together. In this case, there is some evidence of vocal tract resonances tuning as a function of f_0. Soprano singing /ɑ,o,u,ɛ/ vowels with a low vocal effort tend to adjust their second formant to the second harmonic $2f_0$ [10]. The second formant is tuned to the second source harmonics $F_2 = 2f_0 + 50$ Hz above a pitch threshold. For very high f_0, formant tuning is no longer possible because the second harmonic is well above the second formant frequency. In summary, the second formant center frequency F_2 is computed as a function of vowel backness V, vowel height H, vocal tract scale factors α_S and K, and f_0 [sound examples in Additional files 9 and 10][11]:

$$F_2 = \max\left(2f_0 + 50\text{ Hz}, K\alpha_S F2_G(V,H)\right) \quad (26)$$

4.3.6 Formant bandwidths

Formant bandwidths for any vowel are obtained from generic values B_{i_G} (given in Table 3 for canonical vowels), interpolated using vowel height H and vowel backness V.

4.3.7 Formant amplitudes

As for center frequencies and bandwidths, formant amplitudes A_i ($i \in [1,6]$) are obtained by interpolation of the values in Table 3 using vowel height H and vowel backness V.

These values must be corrected depending on f_0. In parallel formant synthesis, the coincidence of f_0 or its harmonic with formant center frequencies is likely to produce artifacts. A sharp resonant filter with a narrow bandwidth is likely to amplify source harmonics too much when multiples of the fundamental frequency f_0 match with the formant frequency F_i ($i \in [1,6]$). In natural voice, this effect is occasionally searched for (e.g., in diphonic singing). To correct possible outstanding harmonics, the first three resonant filter amplitudes A_i ($i \in [1,3]$) are decreased automatically and progressively when the closest kth harmonic ($k \in [0,7]$) of f_0 is becoming closer to the central frequency F_i of the resonant filter i [sound examples in Additional files 11 and 12][12]:

$$\begin{aligned} &\text{if } |(k+1)f_0 - F_i| < \Delta F_i\text{:} \\ &\quad A_i = A_{i_G} - \left(1 - \tfrac{|(k+1)f_0 - F_i|}{\Delta F_i}\right)\text{Att}_{\max_i} \\ &\text{else: } A_i = A_{i_G} \end{aligned} \quad (27)$$

ΔF_i is the frequency interval around the formant central frequency where the attenuation is applied, and it is a linear function of f_0. Its values typically range from 15 to 100 Hz for f_0 from 50 to 1500 Hz. $\text{Att}_{\max_i}$ is the attenuation amplitude at the formant central frequency F_i and is a linear function of f_0. Its values typically range from 10 to 25 dB for f_0 from 50 to 1500 Hz. All these values have been set empirically. For higher order harmonics, no correction is needed because artifacts are not perceived.

4.3.8 Anti-formants

A quality factor of 2.5 and a central frequency of 4700 Hz are used for the generic voice. The piriform sinus shape appears to be person dependent [32]. As the vocal tract size is likely to change the piriform sinus size, the central frequency of the piriform sinus anti-resonance is also multiplied by the vocal tract size scale factor α_S.

5 Results and discussion

In this section, the evaluation of Cantor Digitalis is presented. Following objective evaluation for melodic accuracy and precision, sound quality and musical use are demonstrated with the help of didactic videos, live performance videos, and audio demonstrations for typical voices built with the synthesizer. Applications and the software distribution are presented before the conclusions and perspectives.

5.1 Evaluation of melodic accuracy and precision

Assessment of melodic precision and accuracy in singing using Cantor Digitalis compared to natural singing has recently been reported in a companion paper [17]. The reader is referred to this publication for details on the evaluation; only the main results are summarized here.

Melodic accuracy and precision were measured for a group of 20 subjects using a methodology developed for singing assessment [46]. The task of the subjects was to sing ascending and descending intervals and short melodies as well as possible. Three singing conditions were tested: chironomy (Cantor Digitalis), mute chironomy (Cantor Digitalis, but without audio feedback), and singing (i.e., the subjects' own natural voice). The mute chironomy condition was used for studying the role played by the different (audio, visual and motor) modalities involved when playing Cantor Digitalis.

All the subjects showed comparable proficiency in natural and Cantor Digitalis singing, with some performing significantly better in chironomic singing. Note that for a majority of the subjects, this test was the first contact with Cantor Digitalis. Thus, trained players are likely to

obtain even better results. However, professionally trained singers would most likely also outperform chironomic singers.

Surprisingly, for chironomic conditions, the subjects performed equally well with or without audio feedback: both conditions do not show any significant difference. This result was further investigated in a complementary study [47], showing a generally high visuo-motor ability among subjects and the dominance of vision on audition in targeting visual and audio targets: the subjects rely considerably on visuo-motor skills for playing Cantor Digitalis. This situation is somewhat similar to keyboard playing, where the musician can play with a comparable precision on a mute keyboard.

Note that in the current version of the software, an intonation correction algorithm is also available [21].

5.2 Playing with Cantor Digitalis

Using the 2D tablet surface is preferred for expressive melodic control (in principle, only a 1D parameter). An example of an X-Y trace in time for a simple melody is presented in Fig. 6. Pitch vibrato corresponds to the circles around the notes, while pitch transitions correspond to the larger curves linking the notes. An example of virtuoso melodic gestures is provided in an additional video file [see Additional file 13].

Gestures for vocal efforts and voice quality variations are also intuitively produced by the player. A video example shows two musical sentences with low and high vocal efforts [see Additional file 14]. Gestures for playing vowels and semi-vowels are shown in a third video example [see Additional file 15]. Spectrograms of vocalic variations for whispered speech are shown in Fig. 7. It is also possible to play with the GUI in real time. An example of changing vocal tract size is provided in an additional sound file [see Additional file 16].

As a parametric synthesizer, Cantor Digitalis is not limited to a specific voice. On the contrary, all voice types or other sounds close to the vocal model can be designed. The vocal individuality of a singer results in a specific combination of formants, pitch range, and voice qualities.

The base formant values, measured for a tenor voice, are extrapolated to produce voices with a different mean vocal tract size. A factor smaller than one increases the vocal tract size, such as for a bass singer, whereas a factor greater than one decreases the vocal tract size, such as for soprano or female alto voices. Baby voices are built with a very short vocal tract size, whereas giant voices are built with a very large vocal tract.

Voice source parameters must also be adjusted to create different voices: laryngeal vibratory mechanism, vocal tension, hoarseness, and breathiness. This is demonstrated in additional sound files with dynamic parameter modification on an ascending and descending pitch scale [sound examples in Additional files 17, 18, and 19].

Cantor Digitalis offers voice presets for different vocal types, such as the western classical vocal quartet (bass, tenor, alto and soprano [sound examples in Additional files 20, 21, 22, 23, 24, and 25]) or folk Bulgarian soprano [see Additional file 26]. "Baby" (short vocal tract, very high pitch [sound examples in Additional files 27 and 28]) or "giant" (long vocal tract, low pitch, [sound examples in Additional files 29 and 30]) voices are obtained by pushing some parameters beyond their natural boundaries. Vocal sounds can be turned in wind-like (tense voiced vowels [sound example in Additional file 31] or unvoiced vowels [sound example in Additional files 32 and 33]) or bell-like (very low pitch, vocal tract impulse responses [sound example in Additional file 34]) sounds. All the parameters can be varied independently to build a new voice type. Other formant parameters can also be used for the generic voice.

The parameter values of the voices used for the sound and video examples are presented in Table 2. Figure 7 presents spectrograms of lyric bass voice, Bulgarian soprano voice, whispered voice, and bell-like vocal impulses.

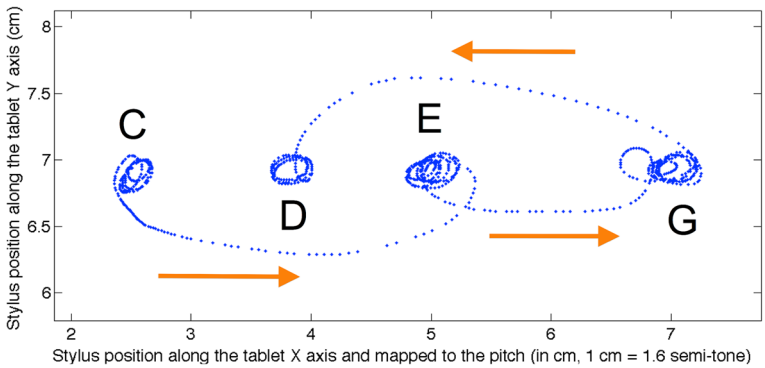

Fig. 6 Trace on the tablet of the melody CEGD played with vibrato. The *red arrows* indicate time

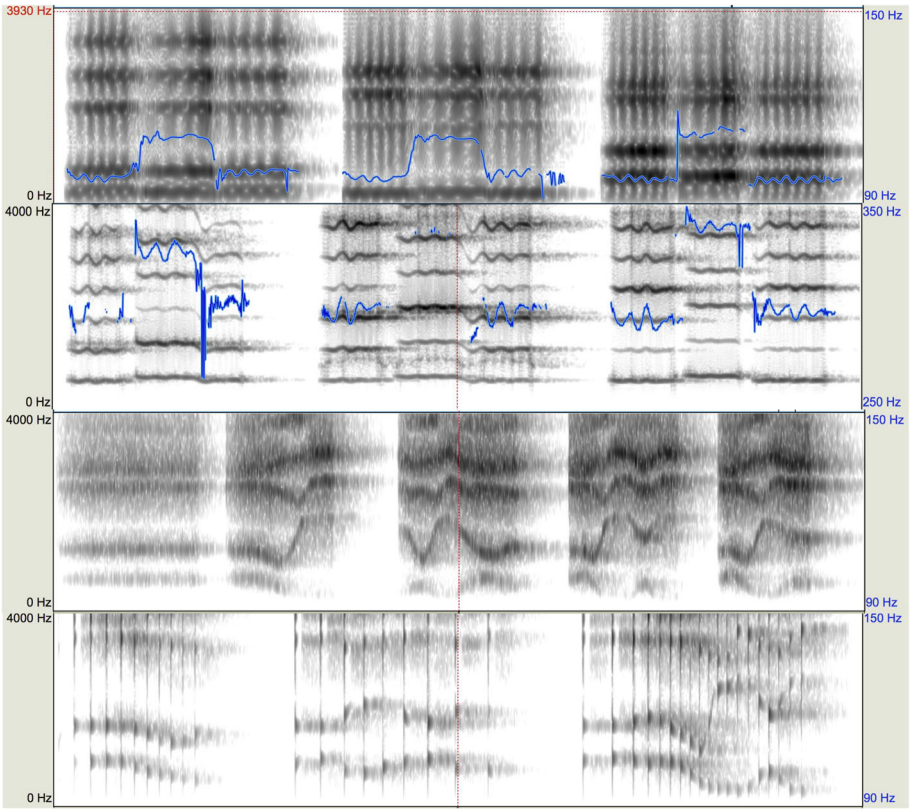

Fig. 7 Spectrograms of different voice types, with pitch (*blue thin line*). *Top*: bass voice [sound example in Additional file 20]. *Second*: Bulgarian soprano voice [sound example in Additional file 26]. *Third*: whispered voice [sound example in Additional file 28]. *Bottom*: bell-like vocal impulses [sound example in Additional file 34]

5.3 Chorus Digitalis and voice factory

The effectiveness of Cantor Digitalis as a musical instrument has been demonstrated during several successful concerts by the *Chorus Digitalis*[13], a choir of Cantor Digitalis. Each musician plays one Cantor Digitalis on her/his laptop with a dedicated loudspeaker for each voice, located just behind each player. Concert video excerpts are associated with this paper (North Indian vocal style [see Additional files 35 and 36]; Opera vocal style [see Additional file 37]; modern vocal style [see Additional file 38]; Bulgarian vocal style [see Additional file 39]; and performing laughs [see Additional file 40]).

Another application of Cantor Digitalis is the Voice Factory software [4], included in Cantor Digitalis. Thanks to this educational tool, the main concepts in the field of voice production can be manipulated and heard in real time. Voice source parameters, formants, and source/filter dependencies can be listened to separately or in combination. However, the most important feature is dynamic control through user gestures during the construction and deconstruction of the voice model, providing an interactive and instructive audio-visual tool. This tool has been used in various contexts: science festivals, classes in several universities, and elementary schools.

5.4 Software implementation and distribution

Cantor Digitalis is implemented in Max[14] [15]. It is distributed under an open-source CeCILL license (a GPL-like license designed by CNRS). Interested readers are able to find all the details of the implementation directly in the software documentation and patches at the following addresses: http://cantordigitalis.limsi.fr or https://github.com/CantorDigitalis/.

The code sources are given in Max 6. It is composed of a main Max patch calling Max *abstractions*. The main Max patch follows the source-filter structure with several sub-patches addressing the rules and parameter mappings. Table 5 presents the list of sub-patches and their references to the corresponding sections of this article. External open-source codes are used, particularly the *s2m.wacom* and *s2m.wacomtouch* Max objects[16] (CeCILL license) allowing to receive the tablet data in Max.

Although the continuous surface appears very adapted for Cantor Digitalis, it is possible to plug in any MIDI interface. MIDI piano keyboards with pedal and wheel

Table 5 List of Max patches, with short description and reference to the corresponding sections in the text

Max patch	Description	Text section
Control	Receive and normalize data from the tablet or MIDI interfaces	2.3, 2.4, 5.4
Voice factory	GUI to set voice quality parameters	2.5
	GUI to construct/deconstruct the vocal model	5.3
GlottisMapping_HL	Compute high-level parameters for the source model	2.1
	from the interface data (E, E_{thr}, p_{hon}, B, T, R)	
	Compute pitch (P_0, P, f_0)	4.1.1
heartPerturbations	Compute heart perturbations for f_0 and E	4.1.3, 4.2.1
otherPerturbations	Compute slow perturbations for f_0 and E	4.1.3, 4.2.1
GlottisMapping_LL	Compute low-level parameters for the source model	4.2.4, 4.2.2, 4.2.3
	(A_g, F_g, B_g, T_{l_1}, T_{l_2})	
VowelMapping	Compute high-level parameters for the vocal tract model	2.1
	from interface data (H, V, S)	
	Compute vocal tract scale factor α_S	4.3.2
VowelRules	Compute generic formant values and interpolate	4.3.1, 4.3.6, 4.3.7
	(F_{i_G}, B_{i_G}, A_{i_G}, F_{BQ}, Q_{BQ})	
	Apply vocal tract length on formants and anti-formants	4.3.3, 4.3.8
SourceFilterDependencies	Compute larynx position adaptation to f_0	4.3.3
	Compute first and second formant tuning	4.3.4, 4.3.5
	Compute formant amplitude attenuation	4.3.7
Glottis	Compute jitter and shimmer	4.1.2, 4.2.4
	Compute amplitude of noise and noise source (A_n, NS)	4.2.5, 3.2.4
	Compute the glottal flow derivative model $\boldsymbol{\mathcal{G}}'$ (GF, ST)	3.2.2, 3.2.3
VocalTract	Compute the vocal tract model $\boldsymbol{\mathcal{V}}$ (R_i, BQ)	3.3.1, 3.3.2

controls have been tested and allow music to be played that requires fast phrases, which is more difficult with the pen tablet.

The robustness of the software implementation has been practically assessed by an important number of downloads to date. The code has already been ported by developers outside our research group to other musical interfaces, such as the Haken Continuum[17] [48], the Madrona Labs Soundplane[18], and the ROLI Seaboard [49][19].

6 Conclusions

Cantor Digitalis is a successful chironomic parametric singing synthesis system. This article aims at presenting the scientific and technical design of this system. As described in the present article, Cantor Digitalis is limited to vocalic synthesis. However, consonant synthesis by rules in the same framework has also been developed [5, 19]. A bi-tablet version of Cantor Digitalis, the Digitartic system, has been demonstrated, with a limited number of consonants (French consonants except /ʀ,l/). Adding consonants on a single tablet proved difficult because too many parameters have to be controlled by the player (pitch, vocalic space, place and manner of articulation, articulation phase, and intensity on attacks and vowels). As the resulting sound quality is generally inferior to that of vowels, one can consider that the question of articulation for consonant for future real-time singing instruments is still open.

Another important question is the automatic learning of specific voices. Statistical parametric learning, such as in modern text-to-speech technology, or other machine learning techniques could be used for incorporating specific voice characters with Cantor Digitalis.

Endnotes

[1] http://cantordigitalis.limsi.fr/chorusdigitalis_en.php

[2] http://guthman.gatech.edu/pastcompetitions

[3] A performative version of "Chant" had been proposed very early [50]

[4] Along this line, the Vocaloid system [51] witnessed phenomenal popular success.

[5] the correct $b1$ coefficient expression is given in [13]

[6] Additional files 1 and 2 are audio examples without and with the perturbations, for a medium vocal effort $E = 0.5$.

[7] Additional audio files 3 (laryngeal mechanism $M = 1$) and 4 (laryngeal mechanism $M = 2$) illustrate the audio effects of Eqs. 16, 17, 18, and 19, with vocal effort $E = 0.8$, tension $T = 0.5$, and fundamental frequency $f_0 = 280$ Hz, Alto voice type.

[8] Additional audio files 5 and 6 are examples of crescendi and decrescendi, without and with phonation threshold. Vocal effort increases from 0 to 0.4 and then decreases from 0.4 to 0, with breathy voice ($B = 0.5$). Note that breath noise remains for $0 < E < E_{thr}$.

[9] Additional audio files 7 and 8 are crescendi without and with formant tuning. Parameter E increases linearly from 0 to 1.

[10] This effect was already in the CHANT program [6]

[11] Additional audio files 9 and 10 are without and with formant tuning, Eqs. 26, and 25, and larynx position adaptation, Eq. 24).

[12] Additional audio files 7 and 8 are glissandi without and with formant amplitude attenuation. Two whistling resonances are attenuated with the rule (beginning and middle of the sound)

[13] http://cantordigitalis.limsi.fr/chorusdigitalis_en.php

[14] http://cycling74.com/products/max/

[15] Max works under OS X and Windows, and *s2m.wacom* works with Max only on OSX 10.6 or later. Then, Cantor Digitalis can be used with all its features on Mac OS X and Windows, except for the graphic tablet control under Windows. On Windows, the current possible controls are the following: MIDI interface like piano keyboard with wheels and pedal; mouse and computer keyboard; and any other control from Max messages. Max Standalones compiled for Mac OS X and Windows are also provided.

[16] http://metason.cnrs-mrs.fr/Resultats/MaxMSP/index.html

[17] See https://youtu.be/R2XRfhu95Dc

[18] See https://youtu.be/oVQMHX4bQuo

[19] See https://youtu.be/mC4pmokMwRo

Additional files

Additional file 1: Long-term perturbation (OFF). See Table 2 for details. (MP3 191 KB)

Additional file 2: Long-term perturbation (ON). See Table 2 for details. (MP3 193 KB)

Additional file 3: Laryngeal mechanism (M1). See Table 2 for details. (MP3 7140 KB)

Additional file 4: Laryngeal mechanism (M2). See Table 2 for details. (MP3 55 KB)

Additional file 5: Phonation threshold (OFF). See Table 2 for details. (MP3 111 KB)

Additional file 6: Phonation threshold (ON). See Table 2 for details. (MP3 115 KB)

Additional file 7: First formant tuning to E (OFF). See Table 2 for details. (MP3 117 KB)

Additional file 8: First formant tuning to E (ON). See Table 2 for details. (MP3 119 KB)

Additional file 9: Formant frequency tuning to f_0 (OFF). See Table 2 for details. (MP3 127 KB)

Additional file 10: Formant frequency tuning to f_0 (ON). See Table 2 for details. (MP3 137 KB)

Additional file 11: Formant amplitude attenuation (OFF). See Table 2 for details. (MP3 105 KB)

Additional file 12: Formant amplitude attenuation (ON). See Table 2 for details. (MP3 101 KB)

Additional file 13: Changing pitch. See Table 2 for details. (MP4 2220 KB)

Additional file 14: Changing effort. See Table 2 for details. (MP4 1830 KB)

Additional file 15: Changing vowels. See Table 2 for details. (MP4 2890 KB)

Additional file 16: Changing vocal tract size. See Table 2 for details. (MP3 633 KB)

Additional file 17: Changing tension. See Table 2 for details. (MP3 298 KB)

Additional file 18: Changing breathiness. See Table 2 for details. (MP3 337 KB)

Additional file 19: Changing roughness. See Table 2 for details. (MP3 355 KB)

Additional file 20: Bass. See Table 2 for details. (MP3 328 KB)

Additional file 21: Tenor. See Table 2 for details. (MP3 332 KB)

Additional file 22: Alto. See Table 2 for details. (MP3 314 KB)

Additional file 23: Noisy Alto. See Table 2 for details. (MP3 288 KB)

Additional file 24: Soprano. See Table 2 for details. (MP3 308 KB)

Additional file 25: Noisy Soprano. See Table 2 for details. (MP3 295 KB)

Additional file 26: Bulgarian Soprano. See Table 2 for details. (MP3 274 KB)

Additional file 27: Child. See Table 2 for details. (MP3 299 KB)

Additional file 28: Gull. See Table 2 for details. (MP3 198 KB)

Additional file 29: Lion. See Table 2 for details. (MP3 97 KB)

Additional file 30: Didgeridoo. See Table 2 for details. (MP3 290 KB)

Additional file 31: DesertBreeze. See Table 2 for details. (MP3 360 KB)

Additional file 32: Whispering. See Table 2 for details. (MP3 340 KB)

Additional file 33: Wind. See Table 2 for details. (MP3 364 KB)

Additional file 34: Woodbells. See Table 2 for details. (MP3 236 KB)

Additional file 35: Raga 1. See Table 2 for details. (MP4 5470 KB)

Additional file 36: Raga 2. See Table 2 for details. (MP4 7320 KB)

Additional file 37: Cold Song. See Table 2 for details. (MP4 1410 KB)

Additional file 38: The Lion sleeps tonight. See Table 2 for details. (MP4 2850 KB)

Additional file 39: Bulgarian song. See Table 2 for details. (MP4 6030 KB)

Additional file 40: Laugh. See Table 2 for details. (MP4 1250 KB)

Acknowledgements
This work was supported by the Agence Nationale de la Recherche ANR, through the ChaNTeR project (ANR-13-CORD-0011, 2014–2017), the 2PIM-MI3 project (ANR/06RIAM02, 2007–2009), and the Conseil régional d'Ile de France through the ORJO project ORJO (OTP 31226, 2009–2013).

Authors' contributions
All the authors, LF, CdA, OP, and BD, participated to the design, implementation, and dissemination of Cantor Digitalis. LF and OP developed the software and documentation, with contributions of CdA and BD. LF and

CdA prepared the manuscript, with contributions of OP and BD. CdA initiated and supervised the project, including coordination and obtaining funding. All authors read and approved the final manuscript.

Competing interests
The authors declare that they have no competing interests.

Received: 1 September 2016 Accepted: 6 December 2016
Published online: 23 January 2017

References

1. ER Miranda, MM Wanderley, *New digital musical instruments: control and interaction beyond the keyboard*. A-R Editions, (Middleton, WI, USA, 2006), pp. 1–18
2. PR Cook, in *Proceedings of the 5th Conference on New Interfaces for Musical Expression (NIME'05)*. Real-time performance controllers for synthesized singing, (Vancouver, BC, Canada, 2005)
3. S Le Beux, L Feugère, C d'Alessandro, in *12th Annual Conference of the International Speech Communication Association (INTERSPEECH 2011)*, ed. by P of the conference ISSN: 1990-9772. Chorus Digitalis : experiment in chironomic choir singing, (Firenze, Italy, 2011), pp. 2005–2008
4. L Feugère, C d'Alessandro, B Doval, in *Intelligent Technologies for Interactive Entertainment, 5th International ICST Conference, INTETAIN 2013*. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, ed. by M Mancas, N d'Alessandro, X Siebert, B Gosselin, C Valderrama, and T Dutoit. Performative voice synthesis for edutainment in acoustic phonetics and singing: a case study using the "Cantor Digitalis", vol. 124 (Springer, Mons, Belgium, 2013), pp. 169–178
5. L Feugère, C d'Alessandro, Gestural control of voice synthesis. the Cantor Digitalis and digitartic instruments. Traitement Du Signal. **32**(4), 417–442 (2015). doi:10.3166/TS.32.417-442
6. X Rodet, Y Potard, J-B Barrière, The CHANT project: from the synthesis of the singing voice to synthesis in general. Comput. Music J. **8**(3), 15–31 (1984)
7. G Berndtsson, The KTH rule system for singing synthesis. STL-QPSR. **36**(1), 1–22 (1995)
8. PR Cook, SPASM, a real-time vocal tract physical model controller; and singer, the companion software synthesis system. Comput. Music J. **17**(1), 30–44 (1993)
9. M Umbert, J Bonada, M Goto, T Nakano, J Sundberg, Expression control in singing voice synthesis: features, approaches, evaluation, and challenges. IEEE Signal Process. Mag. **32**(55–73) (2015)
10. MM Wanderley, J-P Viollet, F Isart, X Rodet, in *Proc. of the 2000 International Computer Music Conference (ICMC2000)*. On the choice of transducer technologies for specific musical functions, (Berlin, 2000), pp. 244–247
11. PR Cook, CN Leider, in *Proceedings of the 2000 International Computer Music Conference (ICMC2000)*. SqueezeVox: a new controller for vocal synthesis models, (Berlin, 2000)
12. L Kessous, Contrôles gestuels bi-manuels de processus sonores. PhD thesis. Université de Paris VIII (2004)
13. N d'Alessandro, P Woodruff, Y Fabre, T Dutoit, S Le Beux, B Doval, C d'Alessandro, Real time and accurate musical control of expression in singing synthesis. J. Multimodal User Interfaces. **1**(1), 31–39 (2007)
14. L Kessous, in *Proceedings of the International Conference on New Interfaces for Musical Expression (NIME'02)*. Bi-manual mapping experimentation, with angular fundamental frequency control and sound color navigation, (Dublin, 2002), pp. 113–114
15. M Zbyszynski, M Wright, A Momeni, D Cullen, in *Proceedings of the 7th Conference on New Interfaces for Musical Expression (NIME'07)*. Ten years of tablet musical interfaces at cnmat, (New York, USA, 2007), pp. 100–105
16. C d'Alessandro, A Rilliard, S Le Beux, Chironomic stylization of intonation. J. Acoust. Soc. Am. **129**(3), 1594–1604 (2011)
17. C d'Alessandro, L Feugère, S Le Beux, O Perrotin, A Rilliard, Drawing melodies: Evaluation of chironomic singing synthesis. J. Acoust. Soc. Am. **135**(6), 3601–3612 (2014). doi:10.1121/1.4875718
18. N d'Alessandro, O Babacan, B Bozkurt, T Dubuisson, A Holzapfel, L Kessous, A Moinet, M Vlieghe, Ramcess 2.x framework—expressive voice analysis for realtime and accurate synthesis of singing. J Multimodal User Interfaces. **2**(2), 133–144 (2008)
19. L Feugère, C d'Alessandro, in *Proceedings of the 13th Conference on New Interfaces for Musical Expression (NIME'13)*. Digitartic: bi-manual gestural control of articulation in performative singing synthesis, (Daejeon, Korea Republic, 2013), pp. 331–336
20. J Laver, *The phonetic description of voice quality. New edition edn*. (Cambridge University Press, Cambridge, 2009)
21. O Perrotin, C d'Alessandro, Target acquisition vs. expressive motion: dynamic pitch warping for intonation correction. ACM Trans. Computer-Human Interact. **23**(3), 17:1–17:21 (2016)
22. GE Peterson, HL Barney, Control methods used in a study of vowels. J. Acoust. Soc. Am. **24**(2), 175–184 (1952)
23. DH Klatt, Software for a cascade/parallel formant synthesizer. J. Acoust. Soc. Am. **67**(3), 971–995 (1980)
24. JN Holmes, Formant synthesizers: cascade or parallel? Speech Commun. **2**, 251–273 (1983)
25. G Fant, *Acoustic Theory of Speech Production*. (Mouton, The Hague, 1960)
26. G Fant, J Liljencrants, Q Lin, A four-parameter model of glottal flow. STL-QPSR. **55**(2), 1–13 (1985)
27. B Doval, C d'Alessandro, N Henrich, The spectrum of glottal flow models. Acta Acoustica. **92**, 1026–1046 (2006)
28. B Doval, C d'Alessandro, N Henrich, in *Proceedings of Voqual'03 : Voice Quality : Functions, Analysis and Synthesis*, ed. by ISCA. The voice source as a causal/anticausal linear filter, (Geneva, Switzerland, 2003)
29. KN Stevens, HM Hanson, in *Vocal fold physiology: voice quality control*, ed. by O Fujimara, M Hirano. Classification of glottal vibration from acoustic measurements (Singular, San Diego, 1995), pp. 147–170
30. R Bristow-Johnson, Cookbook formulae for audio EQ biquad filter coefficients. http://www.musicdsp.org/files/Audio-EQ-Cookbook.txt. Accessed 21 Dec 2016
31. J Sundberg, Level and center frequency of the singer's formant. J. Voice. **15**(2), 176–186 (2001)
32. T Kitamura, K Honda, H Takemoto, Individual variation of the hypopharyngeal cavities and its acoustic effects. Acoust. Sci. Tech. **26**(1), 16–26 (2005)
33. J Dang, K Honda, Acoustic characteristics of the piriform fossa in models and humans. J. Acoust. Soc. Am. **101**(1), 456–465 (1997)
34. JL Flanagan, MG Saslow, Pitch discrimination for synthetic vowels. J. Acoust. Soc. Am. **30**(5), 435–442 (1958)
35. J Kreiman, B Gabelman, BR Gerratt, Perception of vocal tremor. J. Speech Lang. Hear. Res. **46**, 203–214 (2003)
36. RF Orlikoff, RJ Baken, Fundamental frequency modulation of the human voice by the heartbeat: preliminary results and possible mechanisms. J. Acoust. Soc. Am. **85**, 888–893 (1989)
37. S Ternström, Choir acoustics: an overview of scientific research published to date. Int. J. Res. Choral Singing. **1**(1), 3–12 (2003)
38. RF Orlikoff, Vowel amplitude variation associated with the heart cycle. J. Acoust. Soc. Am. **88**(5), 2091–2098 (1990)
39. N Henrich, C d'Alessandro, B Doval, M Castellengo, Glottal open quotient in singing: measurements and correlation with laryngeal mechanisms, vocal intensity, and fundamental frequency. J. Acoust. Soc. Am. **117**(5), 1417–1430 (2005)
40. N Henrich, G Sundin, D Ambroise, C d'Alessandro, M Castellengo, B Doval, Just noticeable differences of open quotient and asymmetry coefficient in singing voice. J. Voice. **17**(4), 481–494 (2003)
41. ACND Felippe, MHMM Grillo, TA-SH Grechi, Standardization of acoustic measures for normal voice patterns. Rev. Bras. Otorrinolaringol. **72**(5), 659–664 (2006)
42. H Traunmüller, A Ericksson, Acoustic effect of variation in vocal effort by men, women and children. J. Acoust. Soc. Am. **107**(6), 3438–3451 (2000)
43. E Joliveau, J Smith, J Wolfe, Vocal tract resonances in singing: the soprano voice. J. Acoust. Soc. Am. **116**(4), 2434–2439 (2004)
44. J-S Liénard, M-G Di Benedetto, Effect of vocal effort on spectral properties of vowels. J. Acoust. Soc. Am. **106**(1), 411–422 (1999)
45. N Henrich, J Smith, J Wolfe, Vocal tract resonances in singing: strategies used by sopranos, altos, tenors, and baritones. J. Acoust. Soc. Am. **129**(2), 1024–1035 (2011)
46. PQ Pfordresher, S Brown, KM Meier, M Belyk, M Liotti, Imprecise singing is widespread. J. Acoust. Soc. Am. **128**(4), 2182–2190 (2010)
47. O Perrotin, C d'Alessandro, Seeing, listening, drawing: interferences between sensorimotor modalities in the use of a tablet musical interface. ACM Trans. Appl. Percept. **14**(2), 10:1–10:19 (2016)
48. L Haken, E Tellman, P Wolfe, An indiscrete music keyboard. Comput. Music J. **22**(1), 30–48 (1998)

49. R Lamb, AN Robertson, in *Proceedings of the International Conference on New Interfaces for Musical Expression (NIME). NIME '11*. Seabord: a new piano keyboard-related interface combining discrete and continuous control, (Oslo, Norway, 2011), pp. 503–506
50. F Déchelle, C d'Alessandro, X Rodet, in *Proc. of the 1984 International Computer Music Conference (ICMC1984)*. Synthèse temps-réel sur microprocesseur TMS 320, (Paris, 1984), p. 15
51. H Kenmochi, H Oshita, in *Proc. Interspeech' 2007*. Vocaloid—commercial singing synthesizer based on sample concatenation (Antwerp, 2007)