

A Singing Toolkit: Gestural Control of Voice Synthesis, Voice Samples and Live Voice.

D. H. Molina Villota, C. D'Alessandro, G. Locqueville, and T. Lucas *

Institut Jean Le Rond d'Alembert
Equipe Lutheries-Acoustique-Musique
Sorbonne Université - Centre National de la Recherche Scientifique
Paris, France
daniel.molina.villota@sorbonne-universite.fr

Abstract. The Singing Toolkit demo presents three approaches to real-time gestural control of voice : control of vocal synthesis using the Cantor Digitalis instruments; syllabic re-sequencing and modification of pre-recorded vocal tracks with the Voks instrument; control of real-time vocal performances, using DAFx and inertial devices. These three approaches exemplify the potential of gesture-based control to enhance vocal performances, expand the creative possibilities in vocal music production, and open up new avenues for expressive control and artistic exploration.

Keywords: Gestural Control of Voice, IMU, Theremin, Voks, Chironomic, Gesture, Cantor Digitalis

1 Introduction

The Singing Toolkit demonstrates our recent work in three directions for real-time gesture control and modification of voice signals. The first instrument, Cantor Digitalis, is a formant synthesizer using bimanual (chironomic) gestures for melodic and formantic control with the help of graphic tablet. The second instrument, Voks, allows for syllabic resequencing using tapping gestures and chironomic control of intonation and voice quality. The third approach is real-time voice transformation through gesture-controlled vocal effects using the IMU RiOT-Bitalino inertial measurement units (an Ircam and Bitalino joint project).

* This Research is funded by ANR National Research Agency: Analysis and Transformation of Singing Style ANR-19-CE38-0001 & Gepeto: GEsture and PEdagogy of inTONation ANR-19-CE28-0018



This work is licensed under a Creative Commons Attribution 4.0 International License (CC BY 4.0).

2 Cantor Digitalis : Chironomic control of synthesized voice

Cantor Digitalis ¹ is a vowel and semi-vowel singing instrument controlled by chironomic gestures [2]. It translates manual gestures into formant synthesis parameters based on the linear model speech production [1], allowing musicians to control the pitch, vocal effort, and vowel of a synthetic voice in real time. The primary gesture interface used for controlling Cantor Digitalis is the Wacom graphic tablet. Writing or drawing gestures by the preferred hand are controlling pitch and vocal effort, while the other hand control the vowel space using a 2D (2 formants) surface, as shown in Figure 1a.

The pen's low latency (5 ms) makes sound produced by Cantor Digitalis seem to exhibit a direct causality similar to that of acoustic instruments. A visual cue is also printed on the tablet to enhance usability. The graphic tablet has proven effective for controlling voice intonation and singing with Cantor Digitalis. Cantor Digitalis can also be controlled with other continuous interfaces, e.g. the Roli Seaboard RISE Multi-dimensional Polyphonic Expression interface (MPE) [6]. In this case, pitch is controlled using a chromatic keyboard, and vocal effort is controlled by pressure on the touch surface. MPE allows for continuous transitions between notes and pressure levels. Cantor Digitalis [7] [3] won the first prize in the Margaret Guthman Musical Instrument Competition (2015). Cantor Digitalis is limited to vowels or vocalic sounds, to the exclusion of most consonants.

3 Voks: Syllabic sequencing of a prerecorded voice

The Voks singing instrument [4] makes it possible to control any voice utterance, including consonants. As it appeared impossible to control each individual articulatory parameter in real time, the syllable is chosen as rhythmic control unit. In practice, the user first loads a sample recording of the desired text being uttered, together with a syllabic annotation of said recording. The loaded sample needs not have any particular rhythm or melody. Then, during the performance, the system resequences the loaded sample, with a rhythm, pitch and vocal quality controlled in real time by the performer's manual gestures.

Syllabic sequencing: Syllabic rhythm control is performed using a cyclic tapping gesture. Several interfaces can capture such gesture data, including buttons, keys, pads, and pressure sensors. Upon tapping/pressing or releasing one's finger on the interface, a one-time signal is sent to the system, triggering advancement of a virtual playhead to the next frame timestamp.

Other gestures: In addition to rhythm sequencing, other parameters are to be controlled by the performer: pitch, vocal effort, vocal tract stretching factor. Some of those parameters are common to Cantor Digitalis, although they are not implemented in the same way — in Cantor, synthesis parameters are controlled directly, whereas in Voks, a prerecorded sample is modified in real time based on control values.

Following Cantor Digitalis, the graphic tablet and MPE interfaces are used to control pitch and vocal effort in Voks. In addition, the theremin has been used as a control

¹ <https://github.com/CantorDigitalis>

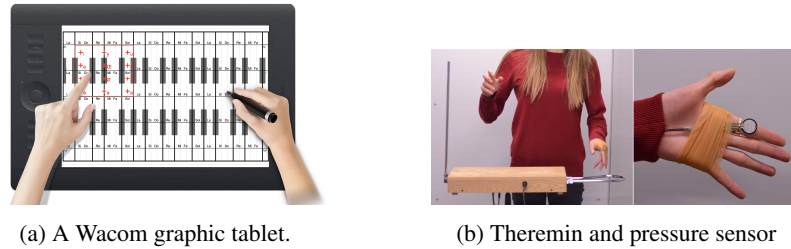


Fig. 1: Two interfaces that can be used for gestural control of vocal synthesis. (a) The Wacom tablet has been used with Cantor Digitalis (pen and finger) and Voks (pen only) (b) The theremin and pressure sensor have been used to control Voks.

interface, with one antenna controlling pitch and the other controlling vocal effort, and an added pressure sensor placed in between the thumb and index of the performer for rhythm control. T-Voks (i.e. Voks played by a Theremin and a rhythm control button) won second place in the 2022 Guthman musical instrument competition.

4 Gesture Control of Digital Audio Effects with IMU

The third tool in the Singing Workshop is interactive real-time gestural control of digital audio effects (DAFx) for voice. The the BITalino R-IoT (abbreviated as R-IoT)[5] is chosen because of its lightness and powerfullness. It is a 9-axis digital IMU sensor (LSM9DS1) that provides absolute orientation in space with low latency over the OSC protocol. The data flow follows the structure indicated in Figure 2. First, R-IoT data is carried to the computer by a router through wifi. Then, data from R-IoT (orientation, quaternions, and acceleration) is received in MAX using the dedicated Bitalino object and Mubu package (by IRCAM). For each DAFx, a selection of parameters, mapping, limit conditions, and appropriate scaling must be made. The data is then sent from Max to the TouchOSC object in Ableton Live using the OSC protocol. There, another mapping is performed to assign those OSC values to different controls in the effects used.

Now we will describe briefly some effects that have been implemented. We have mapped hand rotation to panning: visually, the performer can make an opening gesture, which allows capturing an appropriate range of orientation values for the axis of rotation. Body limitations help define the scaling limits in MAX so that the movement adequately covers the maximum, minimum, and center of stereo panning. Figure 3 a) illustrates this gesture simply. The second effect is an overdrive effect. Within the specific musical piece for which it has been developed, this effect involves distortion applied to all vocal tracks, which gradually increases towards the end of the song. The backward movement of the hand, as shown in Figure 3 b), relates to the incremental distortion by tilting the arm. Finally, another performer triggers a delay effect momentarily using the same gesture. In this case, the sudden movement launches the delay effect based on the speed of the motion, making the control of the delay much more efficient than with a traditional knob. This movement can be seen in Figure 3 c).

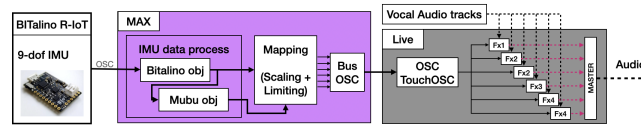


Fig. 2: Flow diagram for Interactive Vocal DAFx with R-devices using MAX and Ableton LIVE.

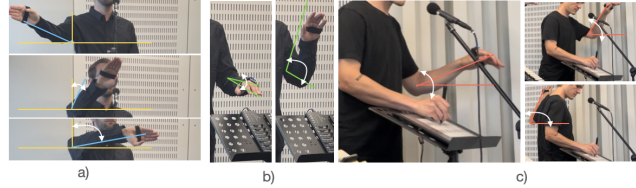


Fig. 3: Schema for the configuration of a) Panning, b) distortion, c) delay using the R-IoT devices.

5 The Demo

The Singing Workshop the demo consists of a room with the three devices set up, each with its corresponding interfaces and computers. Additionally, there will be a poster and three assessors who will explain how the three devices work using musical pieces as examples, within there are also included some tracks of the Chorus Digitalis project, including Cantor Digitalis, Voks and real voices.

References

1. L. Feugère and C. d'Alessandro. Contrôle gestuel de la synthèse vocale. les instruments cantor digitalis et digitartic. *Traitement du Signal*, 32:417–442, 12 2015.
2. L. Feugère, C. d'Alessandro, B. Doval, and O. Perrotin. Cantor digitalis: chironomic parametric synthesis of singing. *EURASIP Journal on Audio, Speech, and Music Processing*, 2017.
3. L. Feugère, S. Le Beux, and C. d'Alessandro. Chorus digitalis: polyphonic gestural singing. 03 2011.
4. G. Locqueville, C. d'Alessandro, S. Delalez, B. Doval, and X. Xiao. Voks: Digital instruments for chironomic control of voice samples. *Speech Communication*, 125:97–113, 2020.
5. Plux Wireless Biosignals and IRCAM. *User Manual R-IoT BitAlino*. <https://www.bitalino.com/storage/uploads/media/manual-riot-v12.pdf>.
6. Roli. Zimphony: The seaboard rise with cantor digitalis, 2016. <https://youtu.be/mC4pmokMwRo>.
7. S. S.Le Beux, L. Feugère, and C. d'Alessandro. Chorus digitalis: Experiments in chironomic choir singing. pages 2005–2008, 08 2011.