



HAL
open science

Borrowed Voices

Christophe d'Alessandro, Xiao Xiao, Grégoire Locqueville, Boris Doval

► **To cite this version:**

Christophe d'Alessandro, Xiao Xiao, Grégoire Locqueville, Boris Doval. Borrowed Voices. International Conference on New Interfaces for Musical Expression NIME'19, Jun 2019, Porto Alegre, Brazil. pp.2.2-2.4. hal-02197066

HAL Id: hal-02197066

<https://hal.science/hal-02197066>

Submitted on 30 Jul 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Borrowed Voices

CHRISTOPHE D'ALESSANDRO, LAM, Institut Jean le Rond d'Alembert, Paris

XIAO XIAO MIT Media Lab, LAM, Institut Jean le Rond d'Alembert, Paris

GRÉGOIRE LOCQUEVILLE, LAM, Institut Jean le Rond d'Alembert, Paris

BORIS DOVAL, LAM, Institut Jean le Rond d'Alembert, Paris

1. PROJECT DESCRIPTION

Borrowed voices is a performance featuring performative voice synthesis, with two types of instruments: C-Voks and T-Voks. The voices are played *a cappella* in a double choir of natural and synthetic voices.

Performative singing synthesis is a new paradigm in the already long history of artificial voices. The singing voice is played like an instrument, allowing singing with the *borrowed voice* of another. The relationship of embodiment between the singer's gestures and the vocal sound produced is broken. A voice is singing, with realism, expressivity and musicality, but it is not the musician's own voice, and a vocal apparatus does not control it.

The project focuses on *control gestures*: the music explores vocal sounds produced by the vocal apparatus (the basic sound material), and "played" by the natural voice, by free-hand Theremin-controlled gestures, and by writing gestures on a graphic tablet. The same (types of) sounds but different gestures give different musical "instruments" and expressive possibilities.

Another interesting aspect is the distance between synthetic voices and the player, the voice being at the same time embodied (by the player gestures playing the instrument with her/his body) and externalized (because the instrument is not her/his own voice): *two different voices sung/played by the same person*.



Fig. 1. T-Voks (Theremin + button, right) & C-Voks (stylus/tablet + button/surface, left)

2. TECHNICAL NOTES

Borrowed voices is part of a long-term research project on performative voice synthesis initiated at the *Speech Conductor workshop* [1]. Previous systems *Ramcess* [2], *Cantor Digitalis* [3], and *Vokinesis* [4][5], explored various synthesis engines (formant synthesis, concatenative synthesis) and control

Licensed under a Creative Commons Attribution 4.0 International License (CC BY 4.0). Copyright remains with the author(s).

interfaces (graphic tablet, MIDI and MPE keyboards, pedals, Puce-Muse MétaTouche, Touché, Expressive-E Touché, Theremin ...).

This research has been accompanied by several musical performances and concerts featuring the *Chorus Digitalis*, an experimental musical choir of synthetic voices. Note that performative voice synthesis systems have been developed in other research groups: the *Voicer* [6] or *Handsketch* [7], or for musical projects like *Luna Park* [8].

T-Voks (presented in an accompanying paper [9] is a Theremin-controlled voice synthesizer. This will be the first public appearance of this new instrument, with its impressive visual and sound presence. For T-Voks the Theremin's frequency antenna modifies the output pitch of the target utterance while the amplitude antenna controls not only volume as usual but also voice quality and vocal effort. An additional pressure sensor attached to the player's volume-control hand handles syllabic sequencing. Metrical control is needed for accurate syllabic timing control. A pair of control points defines each syllable, considered as the basic rhythmic frame. The "*arsis*" corresponds to the constriction (weak beat of the syllable) and the "*thesis*" corresponds to the vocalic nucleus (strong beat of the syllable). The left hand uses a force sensitive resistor (FSR) button for biphasic sequencing of rhythmic units (see Figure 1, left).

C-Voks is a voice synthesizer controlled by a pen on a graphic tablet, using drawing gestures and a new version of the Vokinesis system [6][7]. Different modes of timing control are available: speech rate, scrubbing and metrical control. The speech rate mode corresponds to direct control of the signal playback speed. Scrubbing corresponds to direct control of the playback time position. Metrical control is the same mode as used in T-Voks. The non-preferred hand or feet using a button or pedals perform biphasic sequencing of rhythmic units. Vocal effort or voice quality is controlled by pressure on the tablet and settings.

The sound engines for C-Voks and T-Voks are similar: pitch, time scales, vocal effort, voice quality are modified according to these gestural controls with the help of a real-time high-quality vocoder (WORLD).

This gives an interesting confrontation of free hand gestures of the Theremin and calligraphic gestures for singing or speaking.

3. PROGRAM NOTES

The piece played at NIME'19, entitled "*Borrowed Voices*" is especially composed to explore the various possibilities of vocal instruments played in double a choir of natural and artificial voices. It is a collective creation of the *Chorus Digitalis*, featuring Xiao Xiao (T-Voks, voice), Grégoire Locqueville (C-Voks, voice), Christophe d'Alessandro (C-Voks, voice), Boris Doval (C-Voks, voice). The players' natural voice join occasionally the synthetic vocal quartet in a double choir formation (8 voices). New musical possibilities opened by performative singing synthesis are explored in prepared improvisation and compositions. The sound and gestural material offered by this musical formation allow for :

- **Voice deconstruction** sounding like “computer music” or “electroacoustic voice.” Parametric representation and modeling of the voice allows for extreme variations. Specific features of the voice can be emphasized (formants, pitch, voice quality, vocal tract size, roughness), and a rich sonic material based on the voice can be worked out in real time.
- **Voice imitation** on the contrary privileges the proximity between natural and synthetic voice. How close to a natural voice can a synthetic voice be? In some situation, a realistic voice is desirable. It is at the (possibly interesting) risk of an “uncanny valley” effect.
- **Voice extension** in between deconstruction and imitation, the augmented voice is a realistic-sounding voice with augmented (naturally impossible) features: for instance, a voice with a very large register, a male/female voice, a very slow, very rapid pronunciation, small and large vocal tracts. Another aspect of voice augmentation is the specific vocal gestures allowed by the control interfaces: here the Theremin and graphic tablet.

Borrowed Voices is composed of four clearly contrasted movements rolled into a single unbroken piece. Like previous works of the *Chorus Digitalis* the music is polystylistic: it makes use of multiple styles and techniques, and various languages, i.e. pieces of *Borrowed Styles*. The four movements of *Borrowed Voices* are braid together as follows. The introduction, or **Movement 1**, is an 8-voice motet written by Christophe d’Alessandro. This movements are based on English textual material, a 15th century poem by Julian of Norwich. The motet is designed as a contrapuntal vocal game, based on the strong rhythmic and verse structure of the text, and on the synthetic/natural dialectics both at the choir level (4 + 4) and at the individual level (double voice). After a short transition comes **Movement 2**. It is a solo French song: La vie en rose (Edith Piaf, Louis Guglielmi, arranged by Boris Doval), sung by T-Voks, with a 3-voices accompaniment. The solo voice is sung in an impersonation paradigm, with the typical vocal expressiveness and musicality of the French “chanson réaliste”. The vocal arrangement (3 C-Voks) by Boris Doval for the accompaniment explores freely the sounds and gestures of the synthetic choir. **Movement 3** is a Chinese poem recitation, performed by Xiao Xiao on T-Voks. It shows the musicality of Chinese speech recited using Theremin gestures. The hand motions are drawing in space a gestural equivalent of Chinese tones and vocal expression. **Movement 4** concludes the piece in a prepared improvisation, featuring T-Voks, 3 C-Voks, and natural voices of the 4 musicians. The improvisation is prepared in the sense that the texts (and moods) of Movements 1, 2 and 3 are reused (English, French and Chinese) by the 4 singers in this conclusive part.

4. MEDIA LINK(S)

- Video: <https://youtu.be/XxIz6MnT9HM>
- Video: <https://www.dropbox.com/s/4xyvyv0e5jvy2b3/TVoks%20NIME19%20submission.mov?dl=0>

ACKNOWLEDGMENTS

Part of this work has been done in the framework of the SMAC (FEDER IF0011085) project. Xiao Xiao was partially supported by the MIT-France program during her stay at Sorbonne Université.

REFERENCES

- [1] C. d'Alessandro, N. D'Alessandro, S. Le Beux, J. Simko, F. Çetin et Hannes Pirker, (2005), « The Speech Conductor: Gestural Control of Speech Synthesis » proc. eINTERFACE 2005, Mons, Belgium, pp. 52-61
- [2] N. D'Alessandro, B. Doval, C. d'Alessandro, S. Le Beux, P. Woodruff, Y. Fabre, T. Dutoit « RAMCESS: Realtime and Accurate Musical Control of Expression in Singing Synthesis », Journal on Multimodal User Interfaces, Vol. 1, No. 1, March 2007, p 31-39.
- [3] L. Feugère, C. d'Alessandro, B. Doval, and O. Perrotin. Cantor digitalis: chironomic parametric synthesis of singing. EURASIP Journal on Audio, Speech, and Music, 2017.
- [4] S. Delalez and C. d'Alessandro. Adjusting the frame: Biphasic performative control of speech rhythm. In Proc. INTERSPEECH 2017, Stockholm, Sweden, pages 864-868, 2017.
- [5] S. Delalez and C. d'Alessandro. Vokinesis : syllabic control points for performative singing synthesis. In Proc. NIME 2017 Aalborg University Copenhagen, Denmark, pages 198{203, 2017.
- [6] L. Kessous. Gestural control of singing voice, a musical instrument. Proceedings of Sound and Music Computing, 2004.B.A. R. Jensenius and M. J. Lyons, *A NIME Reader*. Cham: Springer International Publishing, 2017.
- [7] N. D'Alessandro and T. Dutoit. Handsketch bi-manual controller: investigation on expressive control issues of an augmented tablet. In Proceedings of the 7th international conference on New interfaces for musical expression, pages 78{81. ACM, 2007.
- [8] G. Beller, « Gestural Control of Real-Time Speech Synthesis in Luna Park », Proc. Sound and Music Computing, Padova, 2011
- [9] X. Xiao, G. Locqueville, C. d'Alessandro, B. Doval. T-Voks: Singing and Speaking with the Theremin. In Proc. NIME 2019 (this issue)