Real-Time Performance Controllers for Synthesized Singing

Perry R. Cook
Princeton Computer Science (also Music)
35 Olden St. Princeton, NJ 08540
(1)609-258-4951
prc@cs.princeton.edu

ABSTRACT

A wide variety of singing synthesis models and methods exist, but there are remarkably few real-time controllers for these models. This paper describes a variety of devices developed over the last few years for controlling singing synthesis models implemented in the Synthesis Toolkit in C++ (STK), Max/MSP, and ChucK. All of the controllers share some common features, such as air-pressure sensing for breathing and/or loudness control, means to control pitch, and methods for selecting and blending phonemes, diphones, and words. However, the form factors, sensors, mappings, and algorithms vary greatly between the different controllers.

Keywords

Singing synthesis, real-time singing synthesis control.

1. INTRODUCTION

As might be expected, attempts to create controllers for computer voice models consistently point up similar sets of problems. From a technical standpoint, the sheer number of parameters that need to be controlled in an expressive voice model present daunting issues of sensors, bandwidth, systems, and mappings. From a musical, linguistic, and perceptual standpoint, nearly all humans possess a voice, and have years of experience “playing” it (not necessarily musically, but still expressively). Further, humans closely attend to the voices of others, so we are extremely critical of synthesized voices. This paper describes a number of devices to address some of these problems, providing the ability to control a variety of singing synthesis models in real time for musical performance.

2. SQUEEZEVOXEN

There once were two singer/engineers (Perry Cook and Colby Leider), both interested in vocal synthesis, who had been foraging used music stores and Ebay for old accordions. Thus the SqueezeVox project [1] was born in 2000 with hopes of creating meaningful and expressive (or at least fun) controllers for computer voice models. The project recognized that to successfully control a vocal model, independent controls are needed for pitch, breathing, and articulation (spectral features). The accordion is an instrument with components that map somewhat naturally to these requirements. Melody pitch is controlled with the right hand keyboard, “breathing” is provided naturally in the bellows mechanism (though modern accordions “sing” the same when breathing in or out), and the left hand provides an array of buttons (from 10 to 120 depending on instrument type and size). The SqueezeVox project exploited these features of the accordion to control a variety of voice models (formant models, acoustic tube models, FOF synthesizers, orchestras of formant chanting monks, etc.). Bart (Colby’s first SqueezeVox), Lisa (Perry’s, shown in Figure 1), Maggie (Perry’s concertina, shown in Figure 2) and Santa’s Little Helper (Colby’s toy accordion) make up the complete fleet of SqueezeVoxen.

Lisa features a traditional synthesizer keyboard, with a linear FSR located next to the keyboard to control fine pitch bend or vibrato. When used with a formant model of Tuvan overtone singing, the piano key selects the base pitch, and the FSR controls the moving overtone. On the left side, 64 buttons map to phonemes, diphones, words, and phrases, depending on programming, and four bend sensors map to formant positions in a resonant filter model, or articulator positions (jaw drop, tongue tip, tongue hump, and velum opening) in an acoustic tube model [2]. A small 2D “trackpad” also allows for single-finger adjustment of vocal tract resonances in the famous Peterson/Barney [3] vowel space. The bellows vent button, which normally allows rapid inspiration or expiration of air from the bellows without sounding tones, signals the software to create breathing sounds with no phonation. Two speakers mounted inside Lisa allow sound to be projected from the instrument itself, an important design consideration when crafting new musical instruments [4]. Lisa and Bart made their performance debut in Princeton’s “Beyond the 88: A Festival of New Music for Alternative Keyboard Instruments” in 2001.

Similar to Lisa, concertina Maggie’s left side has buttons (32 plus “bank-switch”) and bend sensors (Fig. 2 upper). However, Maggie differs much from her larger SqueezeVox siblings. The right hand provides pitch control via four buttons functioning as brass instrument transposition valves, and a thumb slider (Fig. 2 lower) selecting the overtone/octave. Maggie has no internal speakers, and made her performance debut at Seattle NIME 2001 in the Experience Music Project Museum JBL Theater, performing “7 Minutes from Tibet,” controlling multiple models of Tibetan chant, Tuvan overtone singing, and banded-waveguide Tibetan prayer bowls.

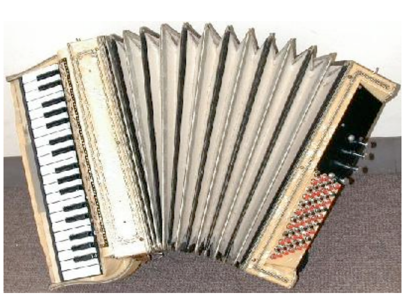

Figure 1. SqueezeVox Lisa.

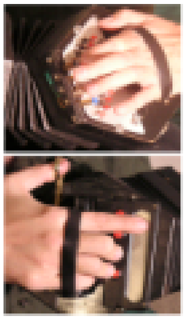

Figure 2. Maggie.

Nime’05, May 26-28, 2005, Vancouver, BC, Canada.

3. THE COWE

So the Zen Master says to the hot dog vendor, "Make me one with everything." (Thanks to Michael Gurevich for this). Originally intended in 2003 as a percussion and sound effects controller [5] [6] that included every type of common sensor (one with everything), the "Controller, One With Everything" (COWE, Figure 3) quickly found use as an interface for vocal models. The COWE has a breath-pressure sensor, linear FSR, and thumb slider, but also borrows from SqueezeVox Bart by adding a two-axis accelerometer tilt sensor for controlling vowel space. Touching the linear FSR overrides the vowel-space tilt control and activates pre-stored linear phoneme/ word sequences, allowing the user to naturally and smoothly sing text. Eight push buttons are divided into four buttons acting as brass-valve pitch control, and four for programmatic control. Four rotary knobs provide additional controls. Two buttons on the underside switch voiced/unvoiced synthesis, and overtone singing, with the overtone controlled by tilt.

Continuous blowing into the breath-sensor became tiring, so a large stuffed toy cow was called into service, reinforcing the COWE name. The cow was re-stuffed with an exercise ball, which when inflated acts as an air reservoir. Inserting the breath pressure hose into the ball's inflating nozzle turns the COWE into a bagpipe-like interface, allowing the user to squeeze the cow under the arm to control breath pressure (Figure 4). The COWE has performed in a variety of venues, including the Princeton Listening in the Sound Kitchen Festival in 2003, and the Cornell Music Festival in 2004.

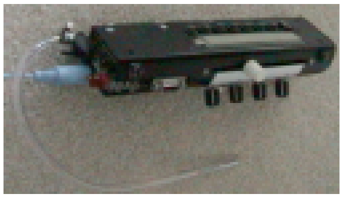

Figure 3. The COWE controller.

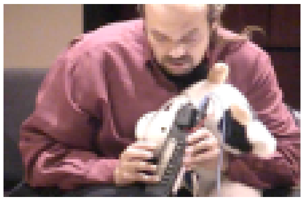

Figure 4. COWE with cow.

4. THE VOMID

Maggie and the COWE proved to be conveniently portable devices, certainly when compared to the larger Lisa and Bart, but the brass-instrument pitch control metaphor proved difficult to maneuver. The piano keyboard was sorely missed when trying to play rapid passages, even though the author is a brass player, because in actual brass playing, lip tension controls the overtone, not a thumb slider. Beginning in 2004, lessons learned from the SqueezeVoxen and the COWE were brought forward and improved in the Voice-Oriented Melodica Interface Device (VOMID) (Figure 5). Based on the melodica, which is a small handheld reed keyboard instrument blown by the player, the VOMID is built on a highly modified (nearly entirely gutted and rebuilt) Korg MicroKontrol device.

The VOMID is designed to be suspended by a neck strap on the chest, thus played somewhat like an accordion (Figure 6). Thanks to Korg, the VOMID sports a 37 note keyboard, 16 programmable touch-sensitive buttons, a joystick, eight rotary pots, and eight slide pots (all programmable). Custom additions to the base controls include a breath pressure sensor, sensitive to both blowing and sucking, mapped to phonation (singing) when blown, and breathing sounds when sucked. A linear FSR is located along side the top two octaves of the keyboard, and is mapped to continuous pitch control, directly related to the discrete pitches of the keyboard. If a key is held down, the FSR behaves as a pitch modifier (bend, vibrato), but if no key is held down, the FSR directly controls pitch over the two octave range. In this way, the "best of both worlds" of accurate discrete pitch (which a real singer does not enjoy), and smooth continuous pitch (which singers do), are available. There is an additional linear FSR next to the bottom octave of the keyboard, for arbitrary programmatic control and special effects. A rotary pot is located on the top plate so as to be easily visible to the player, along with four LEDs (different colors) to display status. Finally, there is a three-axis accelerometer inside, sensitive to leaning and shaking.

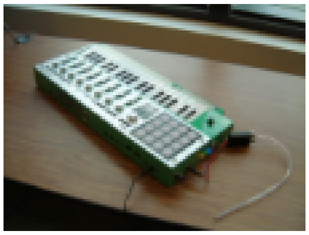

Figure 5 The VOMID

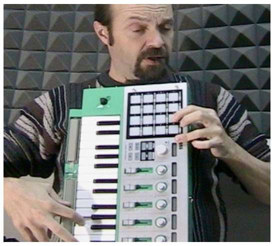

Figure 6 VOMID in performance.

5. SOFTWARE

All devices send MIDI to programs written in a variety of languages and systems. These include the Synthesis ToolKit in C++ (STK) [7], Max/MSP [8], and ChucK [9]. All programs run on Windows and Mac OS X platforms. STK and ChucK programs run additionally under LINUX. Vocal models include source-filter formant models, acoustic-tube models, FM, FOFs, and concatenative PCM (for overview see [10]).

6. REFERENCES

[1] Cook, P. and Leider, C. SqueezeVox: A New Controller for Vocal Synthesis Models," Proceedings of the ICMC (International Computer Music Conference), Berlin, 2000.

[2] P. Cook, "SPASM: a Real-Time Vocal Tract Physical Model Editor/Controller and Singer: the Companion Software Synthesis System," CMJ (Computer Music Journal), 17: 1, pp 30-44, 1992.

[3] Peterson, B. and Barney, H. "Control Methods Used In a Study of the Vowels," Journal of the Acoustical Society of America, 24, 1952.

[4] Cook, P. "Remutualizing the Musical Instrument, Co-Design of Synthesis Algorithms and Controllers," Journal of New Music Research, March 2005.

[5] P. Cook, "Physically Informed Sonic Modeling (PhISM): Synthesis of Percussive Sounds," CMJ 21:3, 1997.

[6] P. Cook, "Modeling Bill's Gait: Analysis and Parametric Synthesis of Walking Sounds," Proc. Audio Engr. Society 22 Conference on Virtual, Synthetic and Entertainment Audio, Helsinki, Finland, June 2002.

[7] G. Scavone and P. Cook, "Synthesis Toolkit in C++ (STK)," Audio Anecdotes, Volume 2, K. Greenebaum and R. Barzel Eds., A.K. Peters Press, 2004.

[8] http://www.cycling74.com/products/maxmsp.html

[9] G. Wang and P. Cook, "ChucK: A Concurrent, On-the-fly, Audio Programming Language," Proc. ICMC, Oct. 2003.

[10] P. Cook, "Singing Voice Synthesis History, Current Work, and Future Directions," CMJ 20:2, 1996.