The Vocal Augmentation and Manipulation Prosthesis (VAMP): A Conducting-Based Gestural Controller for Vocal Performance

Elena Jessop
MIT Media Lab, Opera of the Future
20 Ames Street, E15-445
Cambridge, MA 02139
ejessop@media.mit.edu

Abstract

This paper describes The Vocal Augmentation and Manipulation Prosthesis (VAMP) a gesture-based wearable controller for live-time vocal performance. This controller allows a singer to capture and manipulate single notes that he or she sings, using a gestural vocabulary developed from that of choral conducting. By drawing from a familiar gestural vocabulary, this controller and the associated mappings can be more intuitive and expressive for both performer and audience.

Keywords: musical expressivity, vocal performance, gestural control, conducting.

1. Introduction

The Vocal Augmentation and Manipulation Prosthesis, or VAMP, is a glove-shaped musical controller that is worn by a vocalist. Through simple gestures with his or her gloved arm, the performer is able to capture particular notes and manipulate them to harmonize with himself or herself. This gestural controller differs from previous systems both in its intended users and in the conceptual basis for its gestural vocabulary. This controller was created for vocal performers in order to let the performer serve simultaneously as the conductor and the performer of a piece of solo vocal music, extending his or her voice purely through free gesture without touching buttons, dials, or a computer. In keeping with the use of this controller for vocal performance, the mappings of gesture to sound manipulation are inspired by the gestural vocabulary of choral conducting.

This instrument was originally inspired by the author's work on Tod Machover's upcoming opera, *Death and the Powers*. In this opera, the character of Nicholas has a robotic arm that must also serve as an engaging musical instrument. Such an instrument must be constrained to the physical form of an arm and limited by the unknown instrumental experience of an opera singer. One way to incorporate the performer's significant vocal abilities was to create a controller in the shape of an arm that allowed the performer to manipulate his or her own voice. Thus, much of the audience's focus remains on the sound of the performer's voice, a key component in an opera production. It is also necessary for gestural mappings to be intuitive and clear for an audience that may not have significant experience with electronic music.

Additionally, it is necessary for this controller and associated software to calculate and produce all vocal effects in real time. There are no pre-recorded samples triggered by the controller; all samples are recorded in real time and manipulated in real time

2. Background

2.1 Gestural Control of the Voice

Numerous wearable music controllers that capture gestures through a variety of sensors have been created for enhancing vocal performance. One well-developed gestural instrument is Michel Waisvisz's "The Hands," which incorporates small keyboards on the player's hands, pressure sensors manipulated by the player's thumbs, and sensors to detect the tilt of the hands and the distance between them [1]. Waisvisz has used this instrument to manipulate a variety of parameters to change the sound of his voice and other sonic sources.

Another such instrument is Laetitia Sonami's "Lady's Glove," developed by Sonami and Bert Bongers [2]. This glove utilizes flex sensors on each finger, a Hall Effect sensor on the thumb and magnets on the other four fingers, switches on top of the fingers, and ultrasonic receivers. Data from these sensors is used to control sound, lighting, and even motors, usually for a vocal performance [3].

Another gestural controller that has been occasionally used for vocal performance is the Bodycoder System created by [4]. In early forms, this system employed resistive sensors on knee and elbow joints and keypad-like switches in a glove. Switches triggered pre-recorded samples and selected particular audio and visual patches. In the authors' more recent work with the Bodycoder

System in vocal performances such as "The Suicided Voice" and "Etch," the glove switches trigger particular MSP patches and video events, and all sound manipulation is performed live [5].

The French singer Emilie Simon performs with an arm-mounted controller that allows her to sample and manipulate her voice and the sound of other accompanying instruments. Similarly, Donna Hewitt performs with the eMic[6], a standalone microphone equipped with controls that allow a performer to filter and process his or her voice live. A major difference between VAMP and these previous vocal controllers is that VAMP takes advantage of a pre-existing and fairly intuitive gestural vocabulary and is controlled solely by gesture.

2.2 Glove-Based Controllers

The glove form has been also been used in multiple other musical contexts, including a glove-based music player [7] that allows the wearer to select and play music from a music library, various data-gloves used for additional expression in performance (e.g. [8]) or conducting (e.g. [9]), and a glove that allows the wearer to control sounds in 3D space [10].

2.3 Conducting Systems

There has been significant previous work on capturing the expressive movement vocabulary of a conductor for digital music performance, using on-the-body sensors and/or visual processing techniques. One notable example in this category is Teresa Marrin's "Conductor's Jacket" [11]. This system used EMG sensors on the conductor's biceps and triceps, along with a chest strap that collected physiological data such as heart rate and galvanic skin response. Marrin's extended work with the system, as described in [12], showed that the muscular tension of the arms provided the most data about dynamic intensity from pianissimo to fortissimo. However, most conducting systems, including [7, 12] are designed to interpret conducting gestures in real time for control over a pre-recorded or pre-scored piece of music. VAMP allows the performer conductor-like control over notes generated and recorded instantaneously.

3. System

The base of VAMP is a soft, stretchable fabric glove which extends to the performer's shoulder. This glove is made in the shape and size of a given performer's arm, in order to obtain the most sensitive data about that performer's movement. The current version of the glove has been made of thick, stretchable velvet and sewn by hand to fit the author. By using fabric stretched and form-fitted to the arm, the glove can stay in place without using a potentially uncomfortable elastic band around the upper arm.

3.1 Gestural Sensors

A series of sensors on the glove measure various aspects of the performer's gestural behavior. Two 4.5" flex sensors are sewn onto the glove, one over the elbow joint and one over the wrist joint. When the sensors are used as variable resistors in a voltage divider construction, voltage measurements correlate to the amount of strain. The flex sensor at the elbow measures only the amount of unidirectional bend in the elbow, while the sensor at the wrist can detect the wrist bending either forward or backward from center (though these directions are not differentiated in the output). Stitches at both ends and over the middle of each sensor keep the sensors secure to the glove and limited to bending with the associated joints.

Second, the glove is outfitted with an accelerometer attached to the top of the forearm. This accelerometer is aligned to detect acceleration along the axis that a conductor moves his or her arm when s/he conducts a downbeat. Finally, there is a small 1 lb. pressure sensor attached to the index finger of the glove. This sensor is approximately the size of a fingertip, with a thin, non-sensitive flexible extension that is sewn down the middle of the palm.

Figure 1. Wearing VAMP

3.2 Software System

The data from all the sensors on the glove is collected using an Arduino-compatible Funnel I/O (attached to the upper arm of the glove), and sent wirelessly over a serial connection using Xbee to a Macbook Pro running a Java applet. This Java program utilizes the Processing API and Processing's Arduino libraries to enable communication with the Funnel I/O. In the Java program, the sensor information is collected, analyzed, and mapped, and the desired sound modifications are calculated. Instructions for the desired modifications are then sent to a Max/MSP patch running on the same computer, using [13]'s `MaxLink` libraries for Java. The performer sings into a microphone, sending audio data that is amplified and modified in the Max patch. This allows all of the audio input, processing, and output to be done through Max 5.0, while the sensor input and calculations are carried out using Java and Processing.

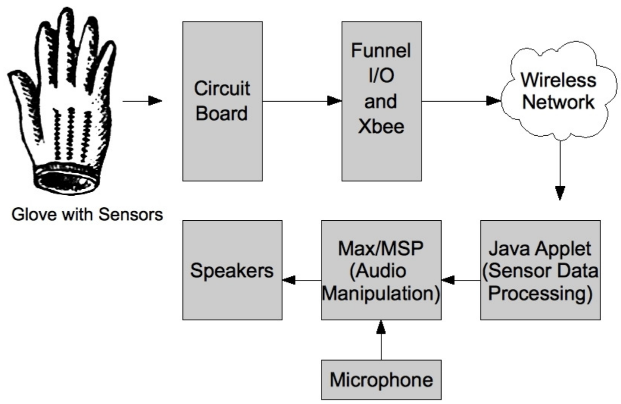

Figure 2. System Diagram

4. Mappings

The mappings between the performer's gesture and the sound modifications were inspired by the movement vocabulary of choral conducting. The specific conducting actions used as the basis for this controller's gestural vocabulary included setting a tempo, controlling amplitude, and adding vocal parts. This vocabulary was also extended with more controller-specific (though still intuitive) actions, such as physically grabbing and releasing individual notes. All these mappings, described in the following sections, are computed in real time.

4.1 Capturing Notes

When the performer closes his or her thumb and forefinger, putting pressure on the glove's pressure sensor, the audio signal that is currently coming into the Max Patch is captured and "frozen." For instance, when the performer sings a note and touches his or her thumb and forefinger together, the current note is held and extended, regardless of other notes the performer sings, until the performer "releases" the note by separating his or her fingers. The pressure from the sensor is regarded as a binary input: pressure above a given level represents a held note, and pressure below that level represents a released note.

The implementation of the "frozen" note processing uses the Max `pfft~` subpatch `solofreeze.pfft` designed by Jean-François Charles [14]. This subpatch uses Jitter matrices to do spectral processing on a Fast Fourier Transform of the audio signal, which allows not only for the necessary computation to be done in real time, but also for a richer sound quality by repeating multiple frames blended together in a stochastic process.

4.2 Following a Beat

One of the primary tasks of a choral conductor's gestures is to set a tempo for a given choral work and have the performers follow that tempo. VAMP provides the ability to pulse a sustained note to a beat pattern indicated by the performer's gesture. Using the accelerometer data from the movement of the performer's forearm, the software constantly examines data patterns over time and locates peaks in the data, which represent downbeats. When two consecutive peaks are detected less than two seconds apart, the length of time between those peaks is set as the beat length (the current tempo), and the program goes into "beating mode." All peaks detected at approximately one beat length apart afterwards trigger amplitude modifications of the sustained note; the amplitude is set to the current high level at each detected downbeat, then fades out after half the calculated beat length. This makes the sound pulse in time with the performer's downbeats.

While the system detects that this "beating" is occurring, it recalculates the beat length with every downbeat and allows the performer a little flexibility in the exact timing of beats. This allows the performer to adjust the tempo and still have the system respond correctly to each downbeat. When the system does not see a beat when expected, it waits for half a second before turning off the "beating mode" and restoring the amplitude of the sound to the previous high level.

4.3 Crescendos and Decrescendos

Additionally, this system allows the performer control over the amplitude of the note s/he is sustaining through gestures indicating crescendos and decrescendos. For a crescendo, the performer extends her arm and reaches out her hand; for a decrescendo, the performer pulls back her hand to near her body. Analysis of the sensor data from the glove indicates that these gestures are primarily characterized by the degree to which the arm is bent at the elbow. Thus, the amount of bend detected by the sensor on the elbow is mapped to the amplitude of the sustained pitch. The range of amplitude of this effect was empirically determined to allow the performer the greatest expressivity in volume without disappearance or significant distortion of the sustained sound.

4.4 Chorusing

In keeping with the choral style explored in this controller, the final effect that the performer can control through this system is the addition of another sustained note in harmony with the one that the performer is holding. The fundamental frequency of a held note is calculated with the `fiddle` external for Max, developed by Miller Puckette [15]. Given this fundamental frequency, any harmony n semitones above the fundamental can be calculated in 12-tone equal temperament using the equation

$$F_{harmony} = F_{fundamental} \times \sqrt[12]{2}^{\,n} \qquad (1)$$

This harmonic frequency is calculated in Max from the fundamental frequencies of any "captured" note. Then, by subtracting the fundamental frequency from the harmonic frequency, we can determine the amount by which the sustained signal needs to be shifted by Max's `freqshift` object. By the performer raising his or her wrist, s/he can bring in and adjust the amplitude of this harmony note.

5. Applications and Future Work

Early performances with VAMP have received positive feedback both on the intuitive nature of the gestural language of the controller and on the specificity and clarity of the resulting performance. Audience members have found the use of the controller to be "expressive and immediate," with a "clear correlation between gesture and sound." The author has found it easy and intuitive to add layers and expression to her vocal performance by using VAMP.

Future extensions of VAMP may include additional features inspired by other choral conducting techniques. For instance, through the use of position sensors or image processing to determine the location and direction of the performer's arm, it would be possible to let the performer give cues to other sections of the room associated with specific vocal parts and hear those vocal parts chorus in harmony with the performer's currently held note. Additional gestures may be added to manipulate the voice by further developing and extending the sense of tangibility of the held notes.

Additionally, other sensors could be added to give a range of performance possibilities. For instance, it would extend the versatility of the device to incorporate pressure sensors on all fingertips, then use different sets of mappings depending on which fingers are touching.

VAMP will also be developed further for use in Machover's upcoming opera *Death and the Powers*. For this performance, it may be useful to further extend the layering effects possible using this system, perhaps allowing the performer to capture and manipulate multiple notes at a time. The system will need to be made quite sturdy to withstand the stresses of rehearsal and performance. Additionally, incorporating this system into *Death and the Powers* will require work with the opera singer playing the role of Nicholas, creating mappings that will take advantage of his particular vocal technique, produce the musical effects desired by the composer, and still retain the intuitive sensibility of a conductor's gesture.

6. Acknowledgments

Thanks to Joe Paradiso and the members of MAS.837, Principles of Electronic Music Interfaces. Thanks also to Tod Machover, Alex McDowell, and Peter Torpey. This research is supported through the Opera of the Future group at the MIT Media Laboratory.

References

[1] A. J. Bongers. "Tactual Display of Sound Properties in Electronic Musical Instruments." *Displays*, 1998, pp 129-133.

[2] B. Bongers. "Physical Interfaces in the Electronic Arts: Interaction Theory and Interfacing Techniques for Real-Time Performance," in *Trends in Gestural Control in Music*, M. M. Wanderley and M. Battier, Eds. Paris: IRCAM, 2000.

[3] L. Sonami. "Lady's Glove," [Web Site], accessed 12/2008. Available: http://www.sonami.net/lady_glove2.htm.

[4] M. Bromwich and J. Wilson. "'Bodycoder': A Sensor Suit and Vocal Performance Mechanism for Real-Time Performance," in *Proc. of the 1998 International Computer Music Conf.*, pp 292-295.

[5] M. Bokowiec and J. Wilson-Bokowiec. "The Suicided Voice," "Etch," in *Proc. of the 2008 Conf. on New Interfaces for Musical Expression*.

[6] "clatterbox>>eMic" [Web Site]. Available: http://www.clatterbox.net.au/instruments/emic.

[7] K. Hayafuchi and K. Suzuki. "MusicGlove: A Wearable Musical Controller for Massive Media Library," in *Proc. of the 2008 Conf. on New Interfaces for Musical Expression*.

[8] M. Marshall. "the_fm_gloves," [Web Site]. Available: http://www.marktmarshall.com/projects/the_fm_gloves.

[9] T. Machover. "Hyperinstruments: A Progress Report, 1987-1991," MIT Media Laboratory, 1992.

[10] J. Schacher. "Gesture Control of Sounds in 3D Space," in *Proc. of the 2007 Conf. on New Interfaces for Musical Expression,* pp. 358-362.

[11] T. Marrin and R. Picard. "The 'Conductor's Jacket': A Device for Recording Expressive Musical Gestures," in *Proc. of the 1998 International Computer Music Conf.*, pp 215-219.

[12] T. Marrin. "Inside the Conductor's Jacket: Analysis, Interpretation, and Musical Synthesis of Expressive Gesture." Ph.D. Dissertation, Massachusetts Institute of Technology, 2000.

[13] Jesse Kris. "jklabs:: maxlink," [Web Site], 2008. Available: http://jklabs.net/maxlink/.

[14] J. Charles. "A Tutorial on Spectral Sound Processing Using Max/MSP and Jitter," in *Computer Music Journal*, Fall 2008, pp. 87-102.

[15] Available: http://crca.ucsd.edu/~tapel/software.html.