Surface Electromyography for Direct Vocal Control

Courtney N. Reed
Centre for Digital Music
Queen Mary University of London
London, UK
c.n.reed@qmul.ac.uk

Andrew P. McPherson
Centre for Digital Music
Queen Mary University of London
London, UK
a.mcpherson@qmul.ac.uk

ABSTRACT

This paper introduces a new method for direct control using the voice via measurement of vocal muscular activation with surface electromyography (sEMG). Digital musical interfaces based on the voice have typically used indirect control, in which features extracted from audio signals control the parameters of sound generation, for example in audio to MIDI controllers. By contrast, focusing on the musculature of the singing voice allows direct muscular control, or alternatively, combined direct and indirect control in an augmented vocal instrument. In this way we aim to both preserve the intimate relationship a vocalist has with their instrument and key timbral and stylistic characteristics of the voice while expanding its sonic capabilities. This paper discusses other digital instruments which effectively utilise a combination of indirect and direct control as well as a history of controllers involving the voice. Subsequently, a new method of direct control from physiological aspects of singing through sEMG and its capabilities are discussed. Future developments of the system are further outlined along with usage in performance studies, interactive live vocal performance, and educational and practice tools.

Author Keywords

voice, singing, sEMG, biosignals, direct control

CCS Concepts

•**Hardware** → *Sensor devices and platforms;* •**Applied computing** → Sound and music computing; •**Human-centered computing** → **Gestural input;**

1. INTRODUCTION

In order to digitally expand the capabilities of the voice while preserving the emotional communication and technique in singing as a musical craft, the inherent difficulty of interfacing with the voice must be tackled: how do you design for something which you can neither see nor touch? Up to now, the voice controllers have revolved around more overt audio analysis and feature extraction. This paper discusses how refocusing design around the physiological interactions which drive vocal technique, independent from any audio produced, can provide a means of direct control.

The 2003 NIME paper by Michael J. Lyons well summarises the core issues with direct control and physiological interaction in voice controllers: "Current ways of interacting with computers neglect most of physiology of human-human interaction and are surely unsuitable for most forms of communication, especially expressive forms such as music" [29]. Emotional expression and communication in singing is hypothesised to utilise existing neural pathways from verbal communication for encoding and interpreting emotion in speech [22, 23], making this interaction especially critical in vocal music. Current vocal interfaces can be broken down into two main categories: controllers which use audio characteristics of the voice to control another instruments, or model-based vocal synthesis controllers which use other forms of interaction, mainly the hands; thus, there is a present gap in the utilisation of vocal technique and clear need to center the vocalist in the design of voice controllers.

In this paper, we present sEMG as a practical and verifiable way of measuring vocal musculature. We thus provide a source of direct control and a sense of tangibility to the voice. This paper begins by defining direct and indirect control in musical interfaces. Next, some instrument augmentations that effectively combine these control methods as well as existing voice controllers are discussed. Related work with sEMG sensing is presented before a method of this sensing is introduced and discussed for direct voice control. This paper thus presents a way to fill the gap for voice control to provide new design and research opportunities, further including vocalists and the voice, an instrument we all play, in digital music.

2. DIRECT AND INDIRECT CONTROL

The distinction between direct and indirect control is well-defined in HCI: coined by Shneiderman, *direct manipulation* describes user actions which are rapid and mimic real-life interactions with objects in an incremental and easily reversible way [48]. An example would be a touch screen, which allows users to directly "touch" objects to open them. Indirect manipulation instead involves an intermediary stage where some translation must occur between the user and machine. Instead of touching an application, a user could indirectly open them at the command line.

In musical interfaces, we thus define this intermediary translation stage as feature extraction. Indirect control would be audio-signal-driven sound synthesis [46] where analysis and parametrisation of sound drives interaction. Truly indirect control-based interfaces would include instruments such as the MIDI guitar, which uses audio signal to generate symbolic MIDI data [53] or Max Mathews's electronic violin through which filter parameters are controlled by audio amplitude [32]. Direct control would not involve this parametrisation. A direct control comparison to the MIDI guitar would be the K-Bow, a violin bow controller which

NIME'20, July 21-25, 2020, Royal Birmingham Conservatoire, Birmingham City University, Birmingham, United Kingdom.

generates MIDI data from bow position, acceleration, pressure, and grip [33]. Features for control could also include sensor measurements and raw audio for excitation of digital synthesis, for instance piezo sensing for string plucking [17] and resonance modeling in the Caress instruments [36]. As seen in cases like this, it is important to note that the use of audio signal does not imply indirect control; the distinction is in the presence of the translation as a result of audio analysis and feature extraction for control parameters.

Therefore, the key affordance of direct over indirect control is that it does not rely on audio production to provide interaction. Ancilliary gestures [13], which do not produce sound, are important in emotional communication [9, 1] as well as group dynamics and synchronisation [11]. Such aspects of performance could be used for direct control. Additionally, most audio analysis, especially spectral analysis, introduces some latency into a system. An unpredictable connection between user and interface can also result from imperfect audio analysis; pitch tracking algorithms are not always accurate and can behave in strange ways. Finally, audio-based indirect control implies that there is an acoustic sound which any digital synthesis must compete with.

2.1 Combining Direct and Indirect Control

Many digital instrument designers have successfully balanced both types of control in a single interface. This is particularly present in augmented instrument design, where the common design goal is to allow the musician to use their existing technique on an otherwise traditional instrument for digital synthesis aspects, sonic or otherwise.

Such augmented instruments include the Overtone Fiddle [41], svampolin [42], and other related hybrid violin controllers [44, 45]. The augmentation of the violin involves multi-modal tracking of several elements of performance, including upbow and downbow detection with electrodynamic pickups on the bridge and pitch tracking via left hand finger placement on pressure sensors embedded into the fingerboard. The key control element here, pitch data, is based on a fusion of this sensor data as well as parameters extracted from the audio signal. Some augmentations were intended for use in teaching, meaning that "both pedagogically and motivationally, players need to feel like they are playing a real violin;" thus, focus was placed on low-latency response and coupling acoustic sound analysis and gestural-based controls to reinforce sensorimotor mappings in learning [42, 43].

This work on the violin was partly inspired by the ESitar, an augmented sitar which also uses a variety of sensors for detecting gestures such as hand position, fret placement, and thumb pressure in addition to audio analysis [25]. This combination of control helps to determine gestures which may otherwise be undetectable from an audio-only standpoint, such as the performer bending a string for pitch variance. The ESitar also provides coordinated visual representation based on direct gestural control, which is useful in teaching contexts for reinforcement of finger placement as well as in creative performance.

This blend of controls can also be found in new instrument design, such as the Bellyhorn[1]; Verdonk describes how visible excitation methods can reinforce connections between synthesis elements to preserve human interaction and expression [52]. Using vocal audio features and direct pressure from body, the bellyhornist can influence the drone produced inside the instrument. Singing loudly or putting one's head further into the horn define the volume of the drone, while lifting the horn influences pitch. The player can also lie on the instrument and change its shape to influence the sound. This combination of control allows the user to create sound through exploration.

3. VOCAL INTERFACES

A look through NIME conference proceedings back to 2001 reveals the voice is a relatively uncommon focus, with fewer than 20 papers devoted to control aspects of the singing voice. The controllers which do exist can be divided roughly into two categories: those using features of the voice (most are indirectly extracted from audio) to control aspects of other-instrument synthesis, and those using non-vocal direct control to manipulate digital vocal synthesis.

3.1 The Voice as a Controller

There are many instances of vocal audio signal features being used to indirectly control synthesis for other instruments. Vowel detection with the Wahwactor allows for control of guitar filtering (a wah-wah pedal) in the guitarist uttering "wah-wah," as a way to reduce the learning demands of using a foot pedal [28]; a similar example is the synthesis of bass guitar using volume, pitch, and timbre extracted from the voice [20]. Other devices have been developed commercially, notably the Vocoder[2] and TalkBox[3] for changing instrumental filtering using vocal formants and mouth shape derived from the audio signal. Newer digital controllers include imitone[4], a voice-to-MIDI controller comparable to the MIDI guitar, and the OVox plug-in by Waves[5] which uses vocal features to control filtering and modulation. Audio signal has also been translated into tactile physical vibration in musical installations [19].

Articulatory aspects of the voice have also been the subject of a few direct control-driven interfaces. There have been several mouth or vocal tract interfaces developed using facial and mouth tracking, such as the Mouthesizer [29], mapped to a variety of sound synthesis parameters [5, 47] or as MIDI controllers [40]. Ultrasound has been used in the case of the Tongue'n'Groove [54] to use tongue contour and motion for controlling other digital instruments as well as a vocal model, although the system was not used for gestural recognition of vocal technique, "but rather to explore how to leverage the fine motor control skills developed by the tongue for expressive music control." Outside of a musical context, ultrasound has also been used to detect speech formants for direct control [26].

3.2 Controllers for Vocal Synthesis

Direct control is more prominent in controllers for vocal synthesis; however, the majority of this direct control relies on gestures unrelated to the voice or vocal performance, such as hand movements [10, 55, 56], manipulation of vocal tracts made of soft materials [57], or browser-based [51] and stylus/tablet control interfaces [6, 12] to change vocal models. Digital vocal processing in computer-based audio plug-ins is also popular; for instance, the Dehumaniser[6] provides modulators, scrubbing, spectral shifting, and a variety of filtering to create artificial monster voices or modify existing audio. Some instruments such as the SqueezeVoxen, COWE, and VOMID [3] incorporate direct controls for voice synthesis which are similar to actual singing, including air pressure sensing for breath control and mouthpieces for phoneme

---

[1] dianneverdonk.com/bellyhorn_pulseyarn

[2] en.wikipedia.org/wiki/Vocoder

[3] en.wikipedia.org/wiki/Talk_box

[4] imitone.com

[5] waves.com/plugins/ovox-vocal-resynthesis

[6] krotosaudio.com/dehumaniser

measurement, through amalgams of other instruments such as accordions and keyboards.

Although removed from organic voice production, some groups such as the Cantor Digitalis[7] team have been able to turn this control of voice synthesis into an art form in its own right. Artists like Pamela Z[8] use custom MIDI controllers and a variety of processing software to process their own voices in real-time, combining an extensive knowledge of traditional vocal techniques with digital capabilities.

3.3 Direct Vocal Control

Among this variety of vocal interfaces, we have identified a clear gap: control of synthesis using direct vocal control. We are interested in utilising the well-developed sensorimotor techniques of vocalists, particularly for audio-independent control. Additionally, we aim to provide this direct control in a format which is affordable enough to be used widely in design for the voice. While physiological measurement tools such as ultrasound are effective for articulatory control, the equipment required is both too cumbersome to be used in performance contexts and currently too expensive for many instrument makers, especially those working outside of academia.

We therefore propose a method of direct physiological vocal sensing using surface electromyography (sEMG) as a minimally disruptive and cost-effective way to bridge this gap and provide a method for voice controller design that can be used by the wider music community.

4. DIRECT CONTROL WITH sEMG

Electromyography (EMG) is the process of measuring electrical neuron activation of the muscles. In the case of this paper and others related to musical interfaces, these electrical signals are measured with sEMG across the skin in a minimally invasive way using surface electrodes. Raw EMG signals usually exist between 0 and 10 mV peak-to-peak and lie between 0-500 Hz, although the usable signal (i.e. that above power-line interference) is mainly between 50-150Hz [2, 4]. These signals can be useful in exploring the vocal mechanism without examining vocal audio signal, thus addressing the control gap.

4.1 sEMG in Practice

sEMG has appeared frequently at NIME in gestural studies. Work by the Embodied Audiovisual Interaction (EAVI) Group at Goldsmiths University of London incorporates sEMG measurements for control in performance [49, 50] and for studying performance gesture and playing techniques [8]. This work has been used in a variety of musical contexts, for instance sensing of the position of the arms and head and gestural controlled pitch mapping. Currently, the group is in development of a low-cost dedicated board for sEMG human-computer interaction in music and instrument making [7], as many sEMG sensing devices are limited to DIY, as done in this paper, or medical-grade [and therefore highly expensive] equipment.

The MYO armbands (Thalmic Labs, now North), which utilise a combination of sEMG and rotational sensing, have also been featured in musical performance [39], composition [31], and studies of musical gesture and control [21, 30, 39] by researchers at RITMO at the University of Oslo. Despite some limitations in terms of gesture classification, users who applied the bands for musical control were found to quickly learn how to adapt their movements for sound production and modification [39].

Another recent study by Kapur et al. involved the design of a wearable interface, AlterEgo, for subvocal (unvoiced) speech recognition [24]. The design captured low-level neuromuscular activity on the face and jaw articulator muscles used in speech production while users silently spoke to themselves. The signal was trained in a recognition model to classify subvocal movements corresponding to specific words. sEMG thus proves to be an effective tool for gestural analysis and classification, even in unvoiced speech; we thus believe exploration of this technology would address audio-independence in direct control.

5. SYSTEM DESIGN

The following section provides a method for sEMG in measuring aspects of vocal performance for direct control. We first discuss sEMG signal acquisition, filtering, and integration with the Bela platform [34], before demonstrating how sEMG can be used to directly measure both vocalised and subvocalised singing.

5.1 Sensing

The controller consists of three 10 mm reusable gold-plated silver cup electrodes (Medimaxtech, New Malden, UK) each with a 120 cm wired connection; the end- and mid-muscles electrodes are placed across the muscle being sensed, while a reference electrode is placed on nearby non-muscular tissue, typically a bony or cartilaginous part of the body. Processing involves two stages, beginning with an analogue preamplification circuit (Figure 1) to acquire electrode signal. The circuit is powered by two 9 V batteries, allowing for external powering and portability, as well as noise reduction from grid power sources.

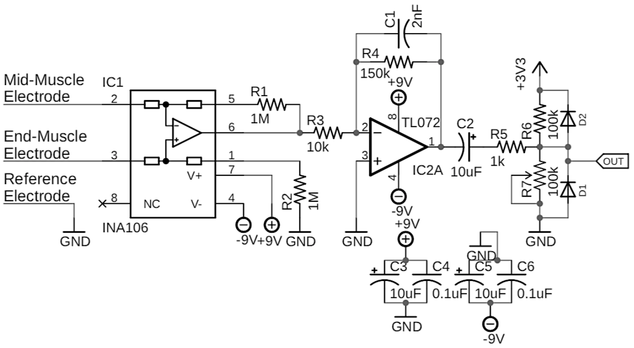

Figure 1: sEMG signal acquisition and preamplifier schematic using three electrodes

The signal acquisition stage of this design is based partially on the open-source EMG Circuit v7.1 (Advancer Technologies[9]). A differential amplifier IC is used to amplify any small voltage difference between the two muscle electrodes. These areas, which otherwise have equal electrical potential, will differ as the muscle is activated and contracts. Differential amplification also reduces noisiness through common mode rejection. The gain of this stage is set to 110. The signal is then passed to an inverting amplifier and a first-order low-pass filter with a 530.5 Hz cutoff to restrict the signal to an appropriate range for sEMG. The sEMG sensor is used in tandem with the Bela board [35], an open-source embedded computing platform which allows for ultra-low latency for signal processing. A variable voltage divider using external power from Bela allows for gain control and

---

[7] cantordigitalis.limsi.fr/

[8] pamelaz.com/

[9] advancertechnologies.com/p/muscle-sensor-emg-circuit-kit-bronze.html

prepares the raw signal with DC offset voltage for use with the platform.

5.2 Usage

An example of the potential use of this controller for direct control with vocal musculature can be seen in the movement of the omohyoideus muscle when singing descending pitches. The omohyoid is an extrinsic laryngeal muscle which lowers the larynx (Figure 2); the main function of this muscle is thus to generate lower fundamental frequencies [16]. This muscle passes beneath the sternocleidomastoideus, one of the neck muscles, which is typically why vocalists are trained to keep the chin down, thus relaxing the neck and keeping pressure off the omohyoideus and other surrounding muscles.

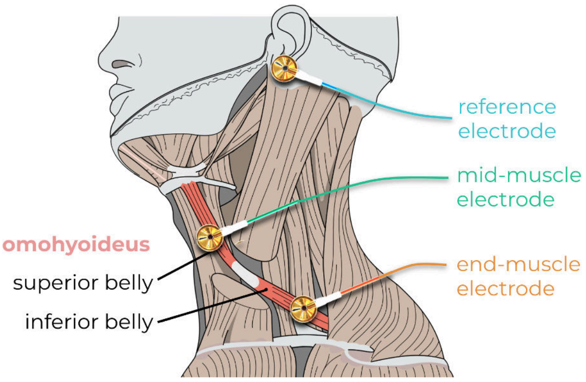

Figure 2: Placement of the three electrodes for sensing activation of the omohyoid.[11]

In a short self-study, a mezzo-soprano with conservatory-level voice training and 7+ years of professional performance experience (and also the first author of this paper), performed a short chromatic exercise in the lower-register of their voice. The mid-muscle electrode was placed on the upper portion of the right omohyoid (the superior belly) in the middle of the neck adjacent to the thyroid cartilage, while the end-muscle electrode is placed at the inferior belly close to the scapula (Figure 2). The reference electrode was placed on the right earlobe. The electrodes were secured with Ten20 conductive paste (Weaver and Company). The vocalist sang chromatics descending from G3 (typical mezzo range extends to F3); a breath was taken before the first two notes but not before the third to observe any contrasts in sEMG as a result of breathing. Additionally, the vocalist indicated the start and end of note events through pressing and releasing a button connected to Bela. The button presses were timestamped for synchronisation and confirmation of voltages were observed in the GUI during the singing exercise.

The neuron activation and contraction of the omohyoid in singing the first three semitones of this downward chromatic sequence can be clearly observed (Figure 3). Markers indicate points where different actions occurred in the signal recording; the different pitches are noted above. sEMG signal is not continuous, but rather the sum of discrete neuron impulses [50] which can be seen in the voltage spikes during this reading. The inhalation taken before G3 is sung is first visible; with each successive downward movement, the amplitude of the signal voltage increases—this is perhaps due to the greater downward laryngeal movement needed to achieve lower pitches at the bottom end of the voice range. The two inhalations are also visible in this case, as a slight lowering of the larynx also occurs during deep breathing where more space is created in the vocal tract [16]. This type of breathing is a core facet of vocal pedagogy and allows the vocalist to shape the vocal tract for rounded, warm tones and provide airflow support for vocal fold vibration.

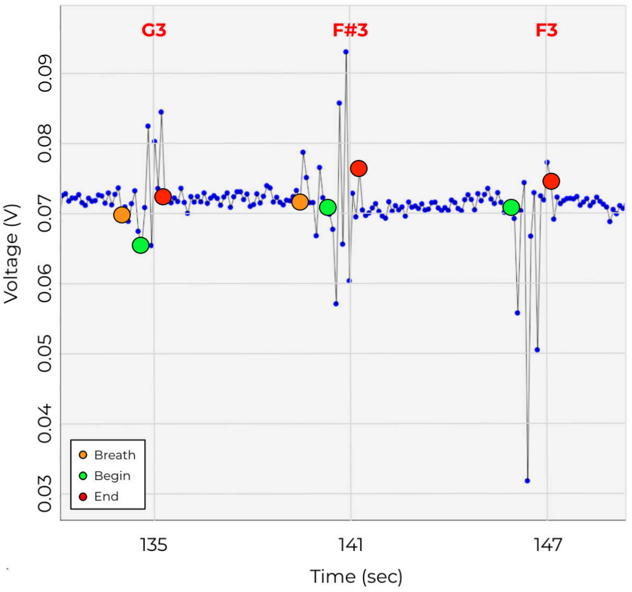

Figure 3: Muscular activation during singing.

The same exercise was repeated in a mental rehearsal to determine the presence of subvocalisation of the same muscular activation (Figure 4). Imagining and executing an activity will result in similar neural activation; in this case, mental rehearsal of a vocal exercise will excite the parts of the brain necessary to perform that exercise [15, 27], resulting in low-level activation which can be detected by electrodes, as done with AlterEgo [24]. Breathing was repeated following the same pattern as done in the previous vocalised trial.

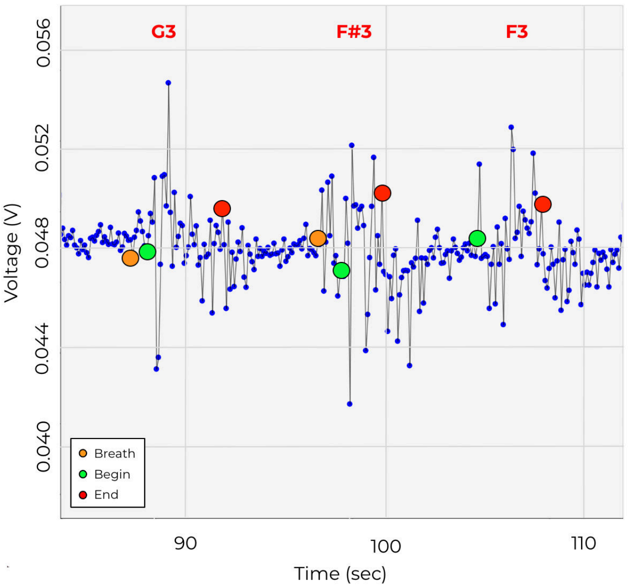

Figure 4: Muscular activation during subvocalisation.

---

[11] electrode images: Pulse Medical; muscular diagram: Olek Remesz, Wikimedia Commons

Although it is clear that the subvocal signal has smaller amplitude and lies more closely in the range of electrical noise in the system, the same markers can still be seen. The difference between the notes and greater downward motion of the larynx is less visible, perhaps a muscular distinction which is lost in mental rehearsal. However, despite not actually producing sound, the gesture required for lowering the larynx can still be observed; thus, we see how sEMG measurements of vocal musculature can provide an audio-independent method of direct control.

6. DISCUSSION

The potential for sEMG as a method for direct sensing of the voice is very promising, as observed in the previous examples. We here have verified that sEMG can be employed to observe musical vocal gesture even while no audio is produced. Thus we provide another outlet for visualising vocal technique in the laryngeal muscles which has not been done previously. Compared to conventional audio analysis, sEMG provides a more introspective look at what the vocalist is doing or intending to do in their performance and the beginnings of a gestural vocabulary, much like we see in other instruments. In areas where audio analysis of voice may struggle in accuracy or ambiguity, such as in pitch recognition, sEMG data can provide means of augmented support, as done with instruments such as the ESitar. The presence of sEMG signal during mental rehearsal and subvocalised singing on its own provides a basis for a wealth of studies regarding musical imagery and rehearsal and learning practices used by vocalists.

6.1 Performance and Education

Direct control using vocal musculature through sEMG and its mapping to digital synthesis techniques provides many opportunities for creative composition and live performance. We are especially interested in the use of such systems in educational contexts, similar to the augmented instruments discussed earlier in this paper. We believe this representation of muscular activation can provide visual reinforcement elements for students, which have been beneficial in addressing common hurdles in instrument learning in other sensor-oriented practice tools such as the 3D Augmented Mirror for violin bowing and performance gesture [38], visualisation of lip embouchure on the flute [18], and the Elbow Piano for piano touch [14].

This use of sEMG also provides a path for new research into a critical area of voice pedagogy: the vocal mechanism, which is obscured in many regards, being within the larynx. Voice teachers do not have the ability to observe or actively adjust laryngeal technique in ways that can be done with other instruments (such as a violin teacher moving a student's hand placement while they hold their bow). Providing a way for teachers and students to measure muscular activity and observe changes in their technique can help to bridge this gap. We hope that sEMG in this way also provides a new direction for vocal physiology research, which otherwise involves invasive medical procedures [37].

6.2 Future Development

Future editions of the system will see a focus on real-time digital filtering of sEMG data and mapping to on-board synthesis using Bela. Additional electrodes will be added to incorporate grouped movements of multiple muscles for different vocal techniques; further, we aim to build a classification system for these grouped movements for robust vocal gesture recognition.

7. CONCLUSION

This paper provides a practical verified system which can be used to study sEMG control with the voice. We provide a method for direct control using the voice which operates independently from audio signal analysis. We find sEMG an appropriate and minimally invasive way to measure and visualise vocal musculature activation in both vocal and subvocal contexts. Through this direct control, sEMG can be applied for new vocal interfaces and augmentations and provide new directions for research in vocal gesture and voice education.

8. ACKNOWLEDGMENTS

Courtney N. Reed is funded by an Electronic Engineering and Computer Science Studentship from Queen Mary University of London. Andrew P. McPherson is funded by EPSRC grant EP/N005112/1 (Design for Virtuosity).

The authors would like to thank Joseph T. Colonel for his thoughtful input on circuit design.

9. REFERENCES

[1] B. Buck, J. MacRitchie, and N. J. Bailey. The Interpretive Shaping of Embodied Musical Structure in Piano Performance. *Empirical Musicology Review*, 8(2), 2013.

[2] R. Chowdhury, M. Reaz, M. Ali, A. Bakar, K. Chellappan, and T. Chang. Surface Electromyography Signal Processing and Classification Techniques. *Sensors*, 13(9), 2013.

[3] P. R. Cook. Real-Time Performance Controllers for Synthesized Singing. In *Proc. NIME*, 2005.

[4] C. J. De Luca. *Surface Electromyography: Detection and Recording.* DelSys Inc., 2002.

[5] G. C. de Silva, T. Smyth, and M. J. Lyons. A Novel Face-tracking Mouth Controller and its Application to Interacting with Bioacoustic Models. In *Proc. NIME*, 2004.

[6] S. Delalez and C. D'Alessandro. Vokinesis : syllabic control points for performative singing synthesis. In *Proc. NIME*, 2017.

[7] B. Di Donato, A. Tanaka, M. Zbyszynski, and M. Klang. EAVI EMG board. In *Proc. NIME*, 2019.

[8] M. Donnarumma, B. Caramiaux, and A. Tanaka. Combining EMG and MMG sensing for musical practice. In *Proc. NIME*, 2013.

[9] M. Doğantan-Dack. In the Beginning Was Gesture. In *New Perspectives on Music and Gesture.* Ashgate, 2011.

[10] N. D'Alessandro, C. D'Alessandro, S. Le Beux, and B. Doval. Real-time CALM Synthesizer New Approaches in Hands-Controlled Voice Synthesis. In *Proc. NIME*, 2006.

[11] T. Eerola, K. Jakubowski, N. Moran, P. E. Keller, and M. Clayton. Shared periodic performer movements coordinate interactions in duo improvisations. *Royal Society Open Science*, 5(2), 2018.

[12] L. Feugère and C. D'Alessandro. Digitartic: bi-manual gestural control of articulation in performative singing synthesis. In *Proc. NIME*, 2013.

[13] R. I. Godøy and M. Leman. *Musical Gestures: Sound, Movement, and Meaning.* Routledge, 2010.

[14] A. Hadjakos, E. Aitenbichler, and M. Mühlhäuser. The Elbow Piano: Sonification of Piano Playing Movements. In *Proc. NIME*, 2008.

[15] A. R. Halpern and R. J. Zatorre. When that tune runs through your head: A PET investigation of

auditory imagery for familiar melodies. *Cerebral Cortex*, 9, 1999.

[16] W. J. Hardcastle. *Physiology of Speech Production: An Introduction for Speech Scientists*. Academic Press Inc., 1976.

[17] J. Harrison, R. H. Jack, F. Morreale, and A. P. McPherson. When is a Guitar not a Guitar? Cultural Form, Input Modality and Expertise. In *Proc. NIME*, 2018.

[18] F. Heller, I. M. C. Ruiz, and J. Borchers. An Augmented Flute for Beginners. In *Proc. NIME*, 2017.

[19] C. J. Holbrow, E. Jessop, and R. Kleinberger. Vocal Vibrations: A Multisensory Experience of the Voice. In *Proc. NIME*, 2014.

[20] J. Janer. Voice-controlled plucked bass guitar through two synthesis techniques. In *Proc. NIME*, 2005.

[21] A. R. Jensenius, V. G. Sanchez, A. Zelechowska, and K. A. V. Bjerkestrand. Exploring the Myo Controller for Sonic Microinteraction. In *Proc. NIME*, 2017.

[22] P. N. Juslin and P. Laukka. Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychological Bulletin*, 129(5), 2003.

[23] P. N. Juslin and D. Västfjäll. Emotional responses to music: The need to consider underlying mechanisms. *Behavioral and Brain Sciences*, 31(5), 2008.

[24] A. Kapur, S. Kapur, and P. Maes. Alterego: A Personalized Wearable Silent Speech Interface. In *Proc. IUI*, 2018.

[25] A. Kapur, A. J. Lazier, P. Davidson, R. S. Wilson, and P. R. Cook. The Electronic Sitar Controller. In *Proc. NIME*, 2014.

[26] N. Kimura, M. Kono, and J. Rekimoto. Sottovoce: An Ultrasound Imaging-Based Silent Speech Interaction Using Deep Neural Networks. In *Proc. CHI*, 2019.

[27] B. Kleber, N. Birbaumer, and T. a. L. M. Veit, R. adn Trevorrow. Overt and imagined singing of an Italian aria. *NeuroImage*, 36(3), 2007.

[28] A. Loscos and T. Aussenac. The Wahwactor: A Voice Controlled Wah-Wah Pedal. In *Proc. NIME*, 2005.

[29] M. J. Lyons, M. Hähnel, and N. Tetsutani. Designing Playing, and Performing with a Vision-Based Mouth Interface. In *Proc. NIME*, 2003.

[30] C. P. Martin, A. R. Jensenius, K. A. V. Bjerkestrand, and V. Johnson. Stillness Under Tension: Performance for Myo armbands and Bela embedded computers. In *MusicLab Vol.1: Biophysical Music*, 2017.

[31] C. P. Martin, A. R. Jensenius, and J. Torresen. Composing an Ensemble Standstill work for Myo and Bela. In *Proc. NIME*, 2018.

[32] M. Mathews. Electronic Violin: A Research Tool. *Journal of the Violin Society of America*, 8(1), 1984.

[33] K. A. McMillan. Stage-Worthy Sensor Bows for Stringed Instruments. In *Proc. NIME*, 2008.

[34] A. McPherson. Bela: An embedded platform for low-latency feedback control of sound. *Journal of the Acoustical Society of America*, 141(3618), 2017.

[35] A. McPherson and V. Zappi. An Environment for Submillisecond-Latency Audio and Sensor Processing on BeagleBone Black. In *Proc. AES*, 2015.

[36] A. Momeni. Caress: An Enactive Electro-acoustic Percussive Instrument for Caressing Sound. In *Proc. NIME*, 2015.

[37] A. Nacci, G. Baracca, S. O. Romeo, M. D. Cavaliere, and M. A. Barillari. Endoscopic and Phoniatric Evaluation in Singing Students. *Journal of Voice*, 33(2), 2017.

[38] K. Ng. *Interactive Multimedia for Technology-Enhanced Learning with Multimodal Feedback*. Springer, Berlin.

[39] K. Nymoen, M. R. Haugen, and A. R. Jensenius. MuMYO — Evaluating and Exploring the MYO Armband for Musical Interaction. In *Proc. NIME*, 2015.

[40] N. Orio. A gesture interface controlled by the oral cavity. In *Proc. ICMC*, 1997.

[41] D. Overholt. The Overtone Fiddle: an actuated acoustic instrument. In *Proc. NIME*, 2011.

[42] L. S. Pardue, K. Buys, M. Edinger, D. Overholt, and A. P. McPherson. Separating sound from source: sonic transformation of the violin through electrodynamic pickups and acoustic actuation. In *Proc. NIME*, 2019.

[43] L. S. Pardue, C. Hart, and A. P. McPherson. A Low-Cost Real-Time Tracking System for Violin. *Journal of New Music Research*, 44(4), 2015.

[44] L. S. Pardue and A. P. McPherson. Near-field optical reflectance sensing for violin bow tracking. In *Proc. NIME*, 2013.

[45] L. S. Pardue, D. Nian, C. Hart, and A. P. McPherson. Low-Latency Audio Pitch Tracking: a Multi-Modal Sensor-Assisted Approach. In *Proc. NIME*, 2014.

[46] C. Pöpel and R. B. Dannenberg. Audio Signal Driven Sound Synthesis. In *Proc. ICMC*, 2005.

[47] C. Pöpel, J. Feitsch, M. Strobel, and C. Geiger. Design and Evaluation of a Gesture Controlled Singing Voice Installation. In *Proc. NIME*, 2014.

[48] B. Shneiderman and P. Maes. Direct Manipulation vs. Interface Agents. *Interactions*, 4(6), 1997.

[49] A. Tanaka and R. B. Knapp. Multimodal Interaction in Music Using the Electromyogram and Relative Position Sensing. In *A NIME Reader: Fifteen Years of New Interfaces for Musical Expression*. Springer, 2017.

[50] A. Tanaka and M. Ortiz. Gestural Musical Performance with Physiological Sensors, Focusing on the Electromyogram. In *The Routledge Companion to Embodied Music Interaction*. Routledge, 2017.

[51] N. Thapen. Pink Trombone, https://dood.al/pinktrombone, 2017.

[52] D. Verdonk. Visible Excitation Methods: Energy and Expressiveness in Electronic Music Performance. In *Proc. NIME*, 2015.

[53] J. A. Verner. Midi Guitar Synthesis: Yesterday, Today and Tomorrow. *Recording Magazine*, 8(9), 1995.

[54] F. Vogt, G. McCaig, M. A. Ali, and S. Fels. Tongue'n'groove: An ultrasound based music controller. In *Proc. NIME*, 2002.

[55] X. Xiao, G. Locqueville, C. d'Alessandro, and B. Doval. T-Voks: the Singing and Speaking Theremin. In *Proc. NIME*, 2019.

[56] T. Yonezawa, N. Suzuki, K. Mase, and K. Kogure. Handysinger: Expressive Singing Voice Morphing using Personified Hand-puppet Interface. In *Proc. NIME*, 2005.

[57] F. Yoshimura and J. Kazuhiro. A "voice" instrument based on vocal tract models by using soft material for a 3D printer and an electrolarynx. In *Proc. NIME*, 2019.