THE APPLICATION OF A NOVEL VOICE-DRIVEN MIDI CONTROLLER IN MUSIC EDUCATION AND TRAINING

Christos Chousidis
Southampton Solent University, United Kingdom

Laurentiu Lipan
Polytechnic University of Bucharest, Romania

Abstract

Music technology is an integral part of music education and training today. A series of applications are developed to assist musicians to record their performance to write music score, to analyze rhythmic and melodic patterns and evaluate their progress. However, the human singing voice which is the dominant means of musical expression it lacks this feature. The system presented in this paper implements an efficient method to convert Electroglottographic signal into MIDI messages. The paper describes the characteristics the operation and the limitation of this novel system and examines its potential application in music education and training.

Introduction

The evolution of computing systems along with the advent of music synthesizers during the last decades, created a new field of technology that of Music Technology. This evolution, gave new potentials for music and music education and made the music creation accessible to all. As a consequence, a significant need for encoding the musical performance appeared. Thus, several MIDI controllers that replicate piano keyboards, guitars, wind and brass instruments and percussions have been designed. However, the human singing voice, which is the dominant means of musical expression, lacks this feature. This is because unlike the other music performance actions, which deliver mainly mechanical information, singing voice provides performance information which can only be obtained through its acoustic impact, which is the produced audio sound. In this paper a novel method of extracting singing performance information is presented (Garcia, 1854). The paper describes and analyzes the characteristics and the abilities of a voice-driven MIDI controller based on the electroglottographic signal.

The electroglottographic signal is a relatively simple signal comparing to its related audio. It gives information for the closure time of the vocal folds by measuring the variation of their electrical impedance and it contains all the necessary information to describe singing. The system presented here is based on the authors' previous research and implementation of an efficient way to convert electroglottographic signals into MIDI messages. In this work we describe the characteristics, the setup, and the operating principle of the system, discuss its limitations and examine its potential application in music education and training.

The remainder of this paper is organized as follows: in section two and three, the fundamentals of the Electroglottography and the MIDI protocol are presented respectively. In section four, the operation of the proposed EGG-driven controller is described. In section five, an evaluation of the proposed system is performed. In section six, the potential application of the system in music training is discussed. Finally, in section seven, the conclusions of this work are presented.

The Electroglottographic Signal

Electroglottography (EGG) is a non-invasive method for describing and analyzing vocal folds' operation. An EGG device, known also as laryngograph, produces a signal that is directly related to the closure time of the vocal folds by measuring the electrical impedance variations (due to abduction and adduction) of vocal folds. This is achieved by means of a small A/C electric current applied by two electrodes, externally on the neck on both sides of the thyroid cartilage. The acquired signal gives an accurate description of the vibration period of vocal folds, especially because it operates on the source of phonation, this way avoiding the interferences related to vocal tract and airborne noise (Howard, 1995). The electrodes of the device apply a high frequency (300 kHz to 5 MHz) and low voltage (approximately 0.5 Volt) signal, which flows across the larynx and the nearby tissues (Henrich, Roubeau, & Castellengo, 2003). The produced waveform is a representation of the glottal movement. Figure 1 shows an acoustic waveform (vowel /a/ phonation that corresponds to note A2, ~110 Hz), along with its respective EGG signal.

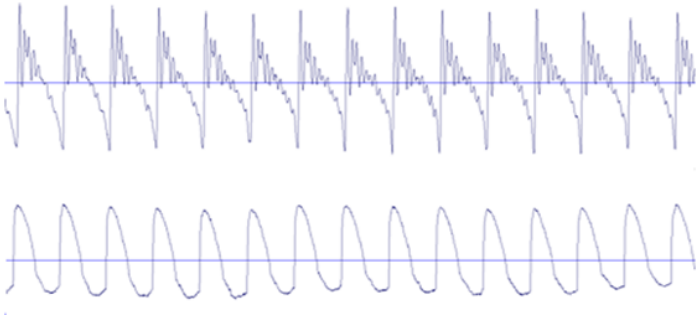

Figure 1: Microphone signal (vowel /a/ phonation, top) and its corresponding EGG signal (bottom).

Figure 2 shows a single period of the EGG signal and its interpretation when it comes to the glottal cycle. The graph describes the open and close periods of vocal folds and shows the contacting and de-contacting event.

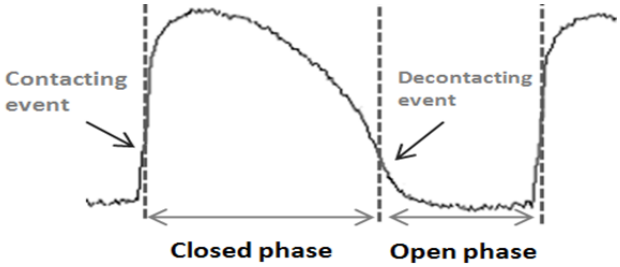

Figure 2: The EGG signal.

The MIDI Protocol

The musical Instrument Digital Interface (MIDI) standard is one of the biggest and most long-lasting innovations in the history of music technology. It was officially introduced in 1983, and it was designed to regulate the communication between different types of music devices such as synthesizer, sampler, drum machine, computer with sequencer and etc. MIDI is a serial asynchronous digital command protocol that encodes actions taken during musical performances into digital messages. MIDI messages are able to encode parameters that describe not only the start and the end of a musical note's execution, the pitch of the note and its variation, but also a plethora of control actions

that are used to differentiate the musical expression in various instruments such as vibrato, breath control, sustain and dumper pedals, after-touch and many others (Chousidis, Rigakis, Hadjinicolaou, & Antonidakis, 2010). The success of the protocol is due to the fact that it has almost 120 different control messages that are able to describe with clarity all kinds of musical expressions. An important feature is also that the protocol provides a fundamental system of networking based on 16 channels. This allows the formation of simple star or daisy chain networks where several independent MIDI devices are able to operate and exchange information over the same infrastructure.

An in-depth analysis of all types of messages and features included in the protocol is beyond the scope of this paper. In the system analyzed here the basic *note-on* and *note-off* messages are used to encode the initiation and the termination of a singing tone. In addition, the *volume control* message is used to encode the variations of loudness and the *pitch-bend* message to encode the pitch variation during singing.

The *note-on* message is generated whenever a note is played in a MIDI controller. The message consists of 3 bytes. The first byte is called the *status byte,* and it contains information about the type and the channel of the message. The second and third byte are called *data bytes,* and they contain information regarding the note number that is played and the initial volume of the note called velocity, respectively. The termination of a note played can be achieved in two ways in MIDI. The first way is the transmission of a *note-off* message for the specific note. This message has the same format like *note-on* with a difference only in the code part of its status byte. The alternative way is the transmission of a *note-on* message with the value zero in the velocity data byte. The system described here uses the second method. The *pitch-bend* message is used to modify the pitch of the notes played on a given MIDI channel. The *pitch-bend* message uses two *data bytes* to describe a 14-bit of information that provides a range of 16384 values. The two bytes are used here to provide the necessary resolution required for a smooth transition within the pitch values. The *volume control* message is a *control-change* message that is generated from a MIDI controller every time a variation of the initial velocity value of a specific note takes place. This message is used for notes with long duration and consequently applies in singing. It consists of one status and two data bytes. The first data byte describes the type of the control message (volume control in this case), while the second data byte describes the value of the controller. Figure 3 provides a detailed description of the above described MIDI messages.

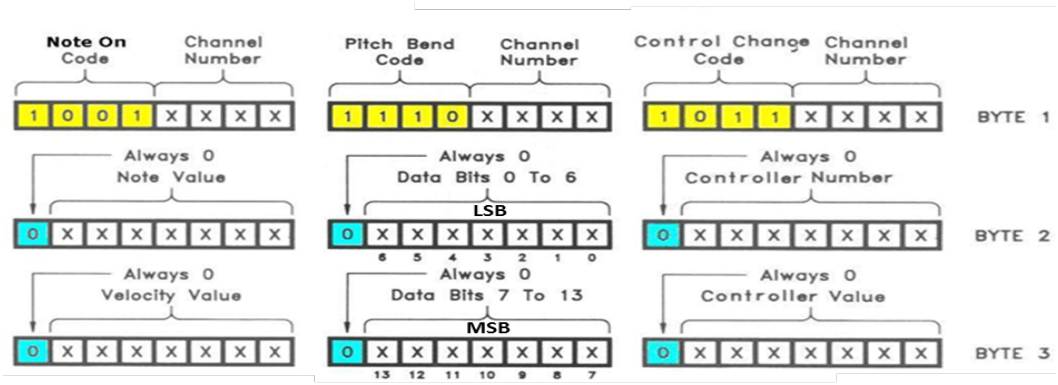

Figure 3: The format of note-on, pitch-bent and control-change MIDI messages.

Operation of the EGG-Driven MIDI Controller

The first task of a voice-driven MIDI controller is to generate a *note-on* message at the beginning of a phonation. This message must contain the pitch and the loudness of the note. Then, during the duration of the note, the controller must identify any possible variation of the pitch or loudness and generate the appropriate *pitch-bend* and *volume control* messages respectively (Kehrakos, Chousidis, & Kouzoupis, 2016). Therefore, a voice driven MIDI controller must be able to perform two main operations, which are to constantly identify the pitch and the loudness of the phonation.

Pitch Extraction

The extraction of the pitch from a complex audio signal is a popular research topic. The main task in this process is to identify the value of the fundamental frequency f_0. This is, however, a relatively complex task especially when the source signal has a rich harmonic content such as singing voice. The most efficient methods for pitch extraction, when it comes to complexity are the Time-Domain methods. Time-Domain methods use a basic approach to the problem of f_0 estimation by usually looking at the acoustic pressure waveform and attempting to detect f_0 from that waveform. These methods are easier to be implemented and require less computational time. However, they perform better when they are applied to simple signals, which have a relatively low harmonic content. For that reason, the EGG signal is used in this implementation instead of the acoustic signal. A relatively efficient time domain method of extracting the f_0 of a signal is the method based on autocorrelation. The autocorrelation function is the correlation of a waveform with itself and it is defined in equation (1).

$$R_{xx}(t_1 t_2) = E\{X(t_1)X(t_2)\} \tag{1}$$

For a harmonic signal in the form of $X(t)=A\cdot\sin(\omega t+\theta)$, equation (1) can be expressed as in (2).

$$R_{xx}(t_1 t_2) = E\{A^2\}\frac{1}{2}\cos(\omega[t_1 - t_2]) \tag{2}$$

From (2) it is shown that the autocorrelation function is also a periodic function, and it is a measure of similarity as a function of time lag *(t1-t2)*. When this method is used, the EGG signal provides accurate information for the fundamental frequency of the signal. Figure 4 shows the advantage of using the EGG signal over the acoustic signal for pitch estimation using autocorrelation.

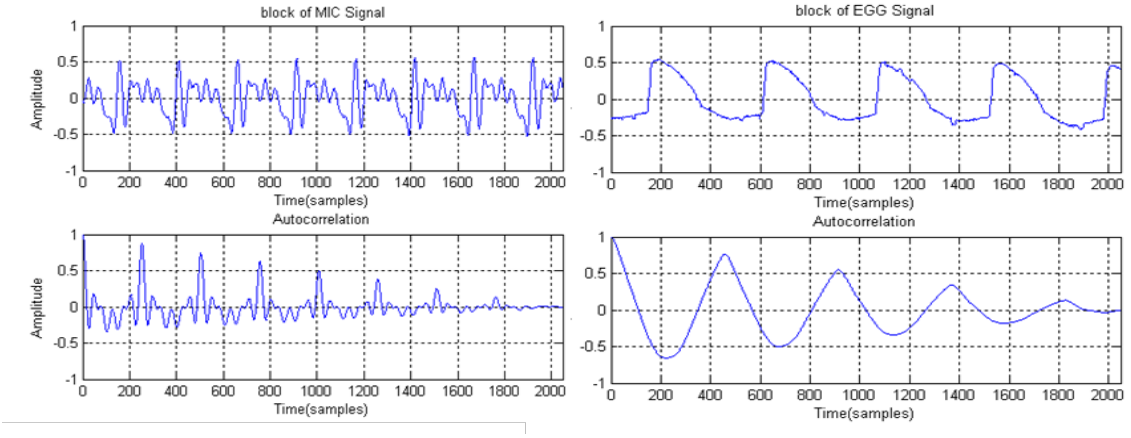

Figure 4: The autocorrelation function in EGG and audio signal.

Main Operation

The main algorithm that describes the system's operation it is shown in the block diagram in Figure 5.

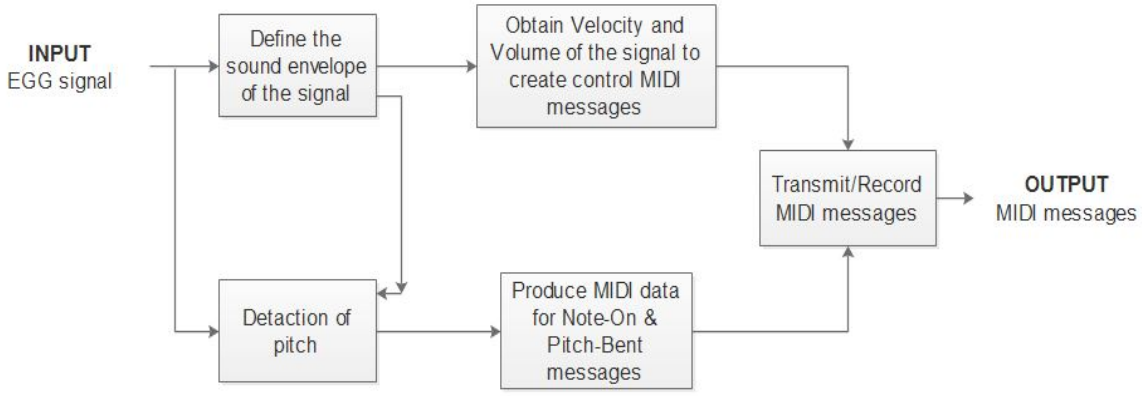

Figure 5: The autocorrelation function in EGG and audio signal.

When the RMS value of the EGG signal passes a predefined threshold, the attack time and steady state part of the waveform are calculated. That is considered as the beginning of the sound production. As long as the signal remains in steady-state, the envelope identification algorithm acts as a trigger for the pitch identification algorithm. During the attack time interval, but also during idle, the pitch detection algorithm creates no signal. Thus, unnecessary calculations are avoided. The frequency estimation takes place for each predefined frame using autocorrelation. With the calculation of the sound envelope, we actually get information for both the attack and volume of the signal. The algorithm scales and assigns integer values in the range from 0 to 127 and then appends these digitized values to the MIDI messages. The calculation of the fundamental frequency provides the information for the identification of the musical note produced by the singer. However, the identified frequency as it is expected, will not always match an exact musical note. Rounding this value to the closest note will cause confusion to the performer who in most cases, gets a real-time feedback of the sound he or she produces through MIDI. In the system analyzed in this paper this problem is addressed by using the *pitch-bend* MIDI messages. The algorithm generates a *note-on* message only the first time a note is triggered. After that, all pitch variations are conveyed using *pitch-bend* messages to correct the pitch of the initially identified musical note. This feature is a significant novelty of the proposed system over other similar implementations.

System Evaluation

The proposed system was tested using numerous recorded EGG files collected from both male and female professional singers. The resulting MIDI messages were recorded into MIDI files in order to be able to compare them with the EGG source signal. In this section we examine the accuracy of the system when it comes to the identifications of the fundamental frequency f_0 and loudness and also its ability to generate the appropriate *note-on*, *pitch-bend* and *control-change* messages. Figure 6 shows the EGG signal of the note E3, singing the vowels /a/, /e/, /i/, /ou/ and the calculated RMS levels.

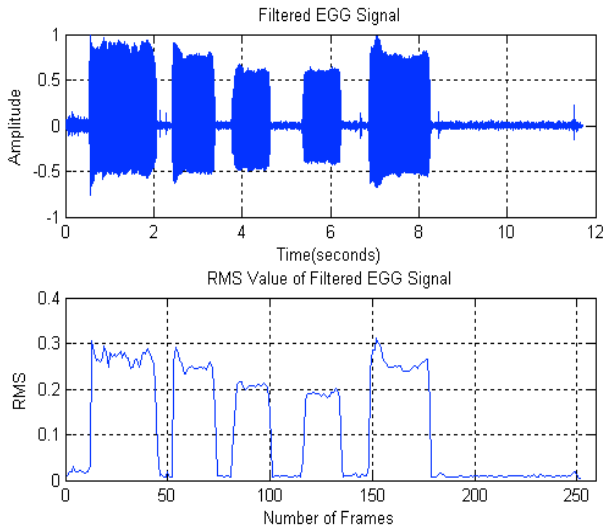

Figure 6: A typical (~8 s duration) EGG signal along with its extracted level values.

Figure 7 shows the calculation of the fundamental frequency and the generation of note-on messages by calculating the corresponding note number and velocity values.

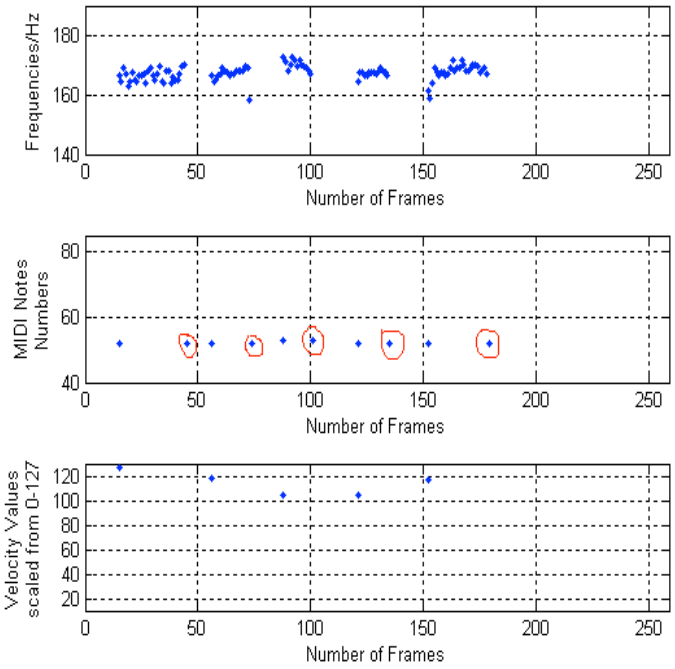

Figure 7: Pitch extraction and note-number and velocity calculation.

Figure 8 presents the *pitch-bend* generation process. It is shown here that a great number of *pitch-bend* messages follow after each *note-on* message in order to capture the frequency variations during singing. The *note-number* messages marked in red have the same value with those preceding them, as they are acting as *note-off* messages, (Note-On with velocity zero). These messages are generated by the system when the phonation ends.

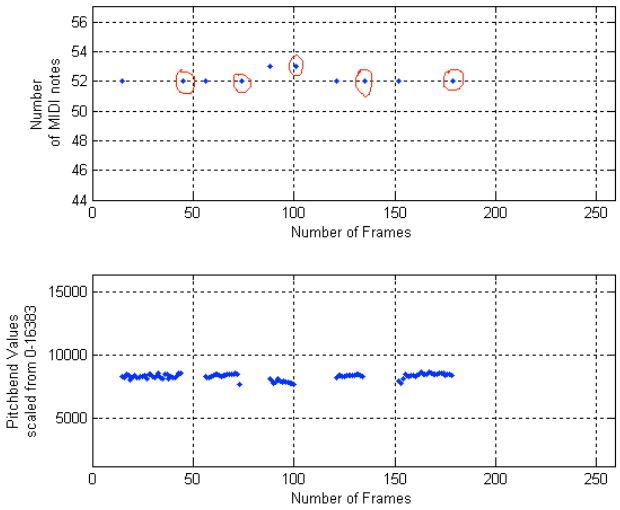

Figure 8: Generation of Pitch-Bend messages.

Application of the System

The human voice is the oldest and most expressive musical instrument. Developing vocal skills requires time and practice. Singing training is a much more complex process than learning a regular music instrument. This is mainly because the sound production system (i.e., the instrument) is part of the singer's body and its operation is not fully controllable.

Singing training focuses in a number of issues that we are not usually facing in the regular musical training. Some of those are, the proper posture of the body, proper breathing techniques, appropriate control of the phonation system, enhancement of the pitch and loudness range etc. Singing training also requires significant self-study and practicing, probably much more than the regular instruments training. Education and training of conventional acoustic and electronic instruments have benefited from the evolution of music technology. Keyboard, guitar, percussion, and wind instrument students can monitor their performance using music software, record and repeat their exercises, identify and correct their mistakes and generally monitor their progress. The EGG-driven MIDI controller presented in this paper, gives these advantages to the singers. The system can be used with any music production software such as Pro Tools, Cubase, Logic etc. without the need of additional microphones and amplification systems. Singers can get a real-time visual and audio feedback on their pitch, vibrato and loudness (Hoppe, Sadakata, & Desain, 2006). They can also record and reproduce their performance and make corrections. They can also use the controller to write scores in order to capture and share their musical ideas. The controller can be used in combination with the multitrack functions of the music software to practice vocals and singing in parallel melodic lines.

Singing teachers also can benefit from the system presented here. By visualizing of their examples, they can make explanation easier, they can provide a more efficient feedback and highlight vocal issues.

In addition to the above the EGG-driven MIDI controller presented in this work can be a very powerful live performance tool for singers. With appropriate training, it can be used to drive synthesizers, control automation systems, effect processors and even

DMX lighting networks simultaneously with singing as it is an independent source of information which is not affected by the noise level of its environment.

Conclusions

In this paper, the characteristics the operation principles and the potential applications of an EGG-driven MIDI are presented and analyzed. The system uses electroglottographic signal to identify the singing information and generates appropriate MIDI messages that describe the initiation and the termination of a phonation and also the variation in pitch and loudness. The novelty of the system is that monitors continuously the pitch of the produced note during the phonation and not only at onset and constantly tunes the corresponding MIDI note using pitch-bent messages. This allows the system to have a plethora of implementations in music education and training but it also be used efficiently in live performance.

References

Chousidis, C., Rigakis, I., Hadjinicolaou, M., & Antonidakis, E. (2010). A MIDI to DMX512 interfacing protocol implemented using microcontroller. *International Journal of Electronic Engineering Research, 2*(5), 713–724

Garcia, M. (1854). Observations on the human voice. *Proceedings of the Royal Society of London, 7,* 399-410.

Henrich, N., Roubeau, B., & Castellengo, M. (2003, April). *On the use of electroglottography for characterisation of the laryngeal mechanisms.* Paper presented at the Stockholm Music Acoustics Conference Stockholm, Sweden.

Hoppe, D., Sadakata, M., & Desain, P. (2006). Development of real-time visual feedback assistance in singing training: A review. *Journal of Computer Assisted Learning, 22*(4), 308-316.

Howard, D. M. (1995). Variation of electrolaryngographically derived closed quotient for trained and untrained adult female singers. *Journal of Voice, 9*(2), 163-172.

Kehrakos, K., Chousidis, C., & Kouzoupis, S. (2016, May). *A reliable singing voice-driven MIDI controller using electroglottographic signal*. Paper presented at the 140th Convention of Audio Engineering Society, Paris.

Authors Details

Christos Chousidis
christos.chousidis@solent.ac.uk

Laurentiu Lipan
laurentiulipan@gmail.com