The Birl: An Electronic Wind Instrument Based on an Artificial Neural Network Parameter Mapping Structure

Jeff Snyder
Princeton University
Princeton, NJ
josnyder@princeton.edu

Danny Ryan
Princeton University
Princeton, NJ
danny@dannyryan.io

ABSTRACT

The Birl is an electronic wind instrument developed by the authors, which uses artificial neural nets for the mapping of fingering systems and embouchure position. The design features of the instrument are described, and the selected machine-learning mapping strategy is discussed.

Keywords

Wind, controller, instrument, machine learning, neural net, capacitive sensing, embouchure, fingering, breath, physical modeling

1. INTRODUCTION

Although the keyboard paradigm has dominated electronic music interfaces since the 1960s, researchers have designed many alternative musical interfaces. Researchers have explored wind controllers, developing several experimental designs in an effort to harness the expressive potential of wind players for electronic music. The development of the Birl seeks to add to that exploration through the creation of a new electronic wind instrument that uses artificial neural networks (ANNs) for musical parameter mapping.

2. GOALS

The primary goal of this research project was to design an electronic wind instrument that is more expressive that that of currently available commercial instruments. In particular, we were interested in developing an instrument that presents the performer with a larger space of pitch possibilities by enabling half holing and alternate fingerings. We were also interested in increasing timbre control through embouchure sensing and in creating an instrument with an attractive and classic look and feel.

3. RELATED WORK

3.1 Commercial Wind Controllers

There are several commercially available wind controllers on the market. The most well known are the Akai EWI and EVI series[1], and the Yamaha WX series[2]. There are a handful of boutique controllers, such as the Morrison Digital Trumpet[3], the Synthaphone[4], and the Eigenharp[5].

[1]http://www.akaipro.com/category/ewi-series
[2]http://usa.yamaha.com/products/music-production/midi-controllers/wx5/

NIME'14, June 30 – July 03, 2014, Goldsmiths, University of London, UK.

3.2 Experimental Wind Controllers

In addition to the established commercial instruments, the NIME research community has also been working on wind-style control of electronic music for decades. Several important papers have pushed the field in new directions. Perry Cook's *HIRN* controller[3] is the most directly related to our project. Cook focuses on increasing the amount of sensing data that is collected from the controller and foreshadows our work in the last paragraph by suggesting machine-learning techniques as a possible solution to the difficulties of mapping multidimensional sensor data to synthesis parameters. Gary Scavone's *PIPE*[11] is also related, as well as the wealth of other wind control research he has generated over the years[13]. Duncan Menzies's work on the *P-bROCK* digital bagpipes[7] explores similar issues of low latency wind control, in developing a teaching and practice tool. Tomás Henriques impressive *Double Slide Controller*[5] builds on the field of trombone-inspired controllers that was paved by Nic Collins's *Trombone Propelled Electronics*[1] and Perry Cook's *TBone*[2]. More distantly related, but important in terms of sensing strategies, are instruments such as Palacio-Quintin's *Hyperflute*[9] or Ystad and Vioner's *Virtually Real Flute*[15] that fit existing wind instruments with sensors to increase control possibilities.

3.3 Other Related Research

Outside of the wind controller paradigm, there are several areas of research that are significantly connected to this work. In particular, this project was informed by research on machine learning in HCI problems and musical instrument design. The Wekinator software[4], developed by Rebecca Fiebrink, has been invaluable to the rapid prototyping of the Birl. Other related research focuses on parameter mapping strategies in general[6] or sensing techniques[10].

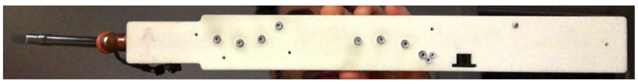

Figure 1: Top view of Birl prototype

4. THE INSTRUMENT

4.1 Fingering System

Commercial electronic wind instruments have switch-like keys, and the pitch is determined by matching a digital

[3]http://www.digitaltrumpet.com.au/
[4]http://www.synthophone.info/
[5]http://www.eigenlabs.com/

word created by scanning these keys to values in a lookup table of known key combinations. If a match is found, the proper pitch value is output. The first version of the Birl, used in the piece *Concerning the Nature of Things (2009)* [14], was designed this way. With this simple mapping, if the player performs a key combination that does not exist in the lookup table, a problem arises — the instrument must either maintain the current pitch or cut off the sound, neither of which make sense musically. Many experimental designs in the NIME community have explored wind controllers with continuous sensing of the fingering keys[3, 9, 11, 15]. However, with this increased depth of input data there are mapping problems. Assuming that the designer is interested in using continuous fingering sensors in a traditional manner (i.e. to determine the pitch of the synthesis output), how does one choose a pitch to output when the input space contails all possible key (2^N) combinations? A simple lookup model is not possible. One option is to use numerical models of tone hole acoustics to determine the effective length of the tube. Scavone and Cook used this approach[12], and it is particularly intriguing. We intend to explore it in the future.

In the prototype, we evaluated an alternative approach based on machine learning techniques. Using Fiebrink's Wekinator[4] as a rapid prototyping platform, we experimented training ANNs with input from the fingering sensors. After training an ANN on a set of known input values (i.e. fingering example 1 = pitch D), the network allows the performer to interpolate between these values in an intuitive fashion. Therefore, every possible finger combination results in a pitch output. Even if some pitches are strange or unexpected, the behavior is something like an acoustic instrument, where alternate fingerings may produce slightly out-of-tune pitches or in-between tones. This permits both the use of half-holing techniques for bending notes and of the discovery of serendipitous "extended techniques" — microtones without the need for exhaustive scale design.

The fingering sensors on the current prototype are aluminum standoffs connected to plated holes in the PCB with screws. They are scanned sequentially using Cypress Semiconductor's CSD capacitive sensing library. A single scan of the 12 finger sensors takes about 5 ms with the current settings, achieving 12-bit resolution and acceptable noise immunity. A new prototype currently in fabrication includes optical reflectance sensing, which has been shown by Menzies[8] to be effective in sensing open-hole wind instrument fingering on his electronic bagpipes and may offer a reduction of the scan time by half.

4.2 Breath Control

Commercial experimental electronic wind instruments typically measure breath pressure the same way. Scavone's PIPE[11] is an exception in which the player to maintains a static pressure rather than a continuous air steam. Because we are interested in harnessing the skill of trained musicians, we implemented the well-accepted continuous airstream measurement system. A Freescale MPXV5004GP gauge pressure sensor is connected to the mouthpiece via flexible silicone tubing. The current synthesis system uses the standard mapping of breath pressure to amplitude.

4.3 Embouchure Sensing

Most commercial electronic wind instruments enhance the mouthpiece with an additional "bite" sensor, which measures something akin to the performer "biting the reed". Some experimental instruments add other sensor systems, such as the myoelectric sensors in Cook's HIRN controller[3] and the ultrasonic distance sensors in Palacio-Quintin's Hyperflute[9]. Our machine learning approach opens up more possibilities for embouchure sensing and is an exciting area in the development of this instrument. The current prototype uses three copper pieces near the lips as capacitive proximity sensors. A user trains a neural net on several different embouchures (i.e. tightened lips, dropped jaw, bottom lip drawn back) and then maps these embouchure positions to synthesis parameters. This allows for the detection of various embouchure positions without explicit mappings of those particular positions, adding a desirable layer of flexibility. We are currently investigating the use of the Swept Frequency Capacitive Sensing (SFCS) technique developed by Munehiko Sato et al[10], which could theoretically detect many different embouchure positions through the use of a single metal mouthpiece or a mouthpiece with a conductive coating.

4.4 The Brain

There are three microcontrollers (MCU) on the birl prototype. A Cypress Programmable System On Chip (PSOC) 8-bit MCU runs at 24MHz and handles the fingering sensor input. Another PSOC is dedicated to the proximity sensing in the mouthpiece. The primary brain, an Atmel AVR32 UC3A 32-bit microcontroller, runs at 66MHz. The brain retrieves the fingering and mouthpiece sensor data from the PSOC MCUs and the breath sensor data from an AD7680 analog-to-digital converter. The brain processes this data and communicates with a computer via UDP network packets in OSC format. The Birl transmits communication over a standard Ethernet CAT5 cable. Once neural nets for the fingering and embouchure sensors is trained, the nets are exported as an array of weights and loaded into the primary brain MCU. The neural net can be used in real-time by the instrument without connection to a computer. The AVR32 can handle some light audio DSP and includes a 16-bit digital-to-analog converter, so we have designed a differential audio amplifier and a 3.5mm audio jack onto the brain PCB. Onboard synthesis has not yet been tested. In the current prototyping stage, a multimedia computer receiving control data over OSC handles the audio synthesis.

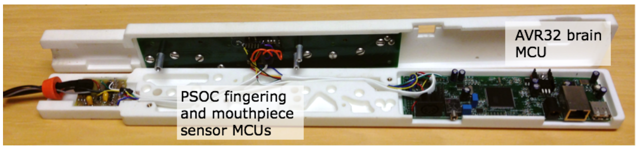

Figure 2: Internal view of Birl prototype

4.5 Case Design

The current Birl prototype case is 3D printed in nylon. It is a clamshell layout, designed for milling in wood on a 3-axis CNC mill. When the structure is finalized, we intend to mill the case out of a hardwood such as maple or walnut. The current mouthpiece is a repurposed vacuum cleaner attachment for cleaning computer keyboards. This is attached to the main body of the instrument with a piece of copper plumbing tubing at 135°, which positions the mouthpiece toward the mouth and puts the hands in a comfortable position. The unusually long and thin mouthpiece was chosen to discourage visual comparison to a soprano saxophone, which the commercially available Akai and Yamaha controllers resemble. In the updated prototype currently in fabrication, the mouthpiece is 3D printed in nylon and designed with better integration of the capacitive sensors. We intend for the finished case to have a classic feel, subtly recalling Renaissance or Medieval wind instruments.

4.6 Artificial Neural Net

Based on tests with the Wekinator system, we determined that a single-layer neural network was sufficient to achieve desirable parameter mappings. We tested the prototype with all sensors as input values and with the following outputs: a floating-point pitch output, an amplitude, and six variable timbre parameters. The ANN output calculation takes around 1ms on the current 32-bit brain microcontroller — fast enough to be incorporated into the firmware loop without slowing down the sensor scan rate.

It is important to note that the Birl does not perform any time-based gesture following or prediction. The machine learning is strictly used to provide an instantaneous output state based on the current sensor input.

5. EVALUATION

5.1 Overview of Evaluation with Expert Users

No formal user studies have been conducted. However, the instrument has been tested informally with four expert saxophonists. The primary goal of these informal user tests was to test the hypothesis that ANNs are an effective tool for handling parameter mapping on a wind instrument with continuous sensing. The secondary goal was to gather feedback about the instrument design, so that we can improve the instrument's usability in subsequent prototypes.

During testing, OSC data from the birl was sent to a laptop computer running OS X. On the computer, the training and running of ANNs were handled by the Wekinator software, which then sent parameter data to a Max/MSP patch for audio synthesis. A small buffer size of 64 samples was used to keep audible latency to a minimum. Two synthesis methods were tested — a simple triangle wave with variable FM and a physical model patch, the "blotar", from PeRColate[6] by Dan Trueman and Luke Dubois.

Figure 3: Experts testing Birl protoype

5.2 Evaluation Procedure

Each user first built training examples correlating fingerings with the pitches of the chromatic scale, usually about 30 examples per note. Then each user trained three embouchure positions and correlated them with particular synthesis parameters by describing to us the timbre that each position should map to. The input data was a set of 17 floating-point features — the 13 fingering keys, the pressure sensor, and the three capacitive sensors on the mouthpiece. The output parameters were a single floating point MIDI value for pitch, an amplitude value, and three timbre parameters. The neural nets were trained on this data, and then the user experimented with the instrument for a few hours.

5.3 Feedback Regarding Machine Learning

All of the performers found the ANN to be useful and to perform surprisingly well. The transition between two trained fingerings often displayed a linear pitch change, "smooth bends". For instance, when moving from a trained C to a trained D, a user opens the lowest finger hole sensor. If the user opens the capacitance fingering hole gradually, the rise from C to D is noticeably smooth. Many (though not all) transitions between trained pitches exhibit this characteristic. Each of the performers found the ability for smooth bends between pitches to be both interesting and musical.

When trained on a chromatic scale, two-fold cross validation showed that the neural net had an RMS error of zero. Thus all explicit mappings perform as expected.

Performers agreed that untrained fingering combinations produced interesting output, but we expect tone-hole modeling would likely produce output that is more intuitive. Random fingering combinations produce various "in between" and out-of-tune pitches, which could be interesting for an experimental musician because they are repeatable within a neural net configuration.

5.4 Feedback Regarding Specific Details

5.4.1 Fingering System

All of the users reported that the finger holes felt too small, presumably because they were all saxophonists used to larger keys. We expect this to be different for players used to recorder or bagpipes, but it was significant enough for us to reconsider the design of the keys. We are considering alternatives, such as countersinking the case around the holes or using larger surfaces for the sensors.

The users liked the "open holed" style capacitive sensors for the finger keys, but they found the continuous sensing and the lack of haptic feedback frustrating for the thumb keys. They found the thumb keys harder to avoid touching by accident. The users also wanted the left hand thumb key mapped to octave. To address this feedback, the next revision utilizes mechanical momentary pushbuttons instead of capacitive sensors for the thumb keys.

5.4.2 Pinky Keys

Much like the C# and C keys on a flute, there are three holes for the right pinky (smallest finger). All users declined to use the pinky keys, because they were too small and close together. The next prototype will only have two right hand pinky keys that are larger and farther apart.

5.4.3 Case Shape

The shape of the case was successful. All users commented that they liked how it positioned their hands and mouth. We recommended a technique of placing the base of the instrument on the left leg while seated. This posture was found to be comfortable for all users. Use of the instrument with a strap while standing has yet to be evaluated.

5.4.4 Speed of Response

All users commented that the speed of the response was excellent and that they could perceive no delay whatsoever. The brain microcontroller was sending breath data to the computer every 2 ms and fingering/embouchure data to the computer every 5 ms. In fact, the extended technique of flutter tonguing turned out to be possible because the rate of data capture and transfer was fast enough to capture that speed of breath pressure variation. The 2 ms breath sensor data transfer rate could theoretically capture the 500

[6] http://music.columbia.edu/percolate/

Hz maximum of the mechanical frequency response of the pressure sensor according to the data sheet, which suggests the possibility of "simultaneous singing and playing" multiphonics. This has not yet been tested.

5.4.5 Embouchure Sensing

All users were excited by the possibilities of the trained embouchure sensing. Tests verified that the technique shows significant promise although the reliability of the embouchure sensing on the current prototype suffered from several problems. The capacitive sensors on the mouthpiece gave rather noisy signals, probably due to the free-hanging wires connecting the sensors. The motion of these wires could alter the fields. The sensors were attached to the mouthpiece with wire ties so they would slowly drift in position. The trained system only worked for a few minutes before the sensor outputs would change too much to classify correctly. These issues will be addressed by integrating the sensors into the mouthpiece itself and by experimenting with SFCS.

5.4.6 Synthesis

Of the synthesis methods tested, simple FM and physical modeling, the users enjoyed experimenting with both. The physical model allowed for some very interesting possibilities with embouchure sensor training. The users especially liked getting the "blotar" model to squeal by adjusting their embouchure. While other synthesis methods (i.e. granular, concantenative) are possible, we found that there was much to be explored in the two tested synthesis types. Additionally, they are both algorithmically efficient such that it will be possible to implement them in real-time on the AVR32. Thus we are currently focusing on these two methods.

6. FUTURE WORK

Circuit boards for the next prototype are currently in fabrication, and an adjusted case design is being developed to be CNC milled in wood. Features to be explored in this upcoming hardware prototype include the evaluation of SFCS for embouchure capture, countersunk case holes for easier tactile fingering feedback, and integration of an LCD screen and controls for adjustment of parameters and loading of presets directly on the device. The substitution of IR reflectance sensing for the finger holes is also being evaluated.

The current training system was unwieldy for the testers because the Wekinator platform is designed for flexible and broad applications. We are working developing our own training GUI in OpenFrameworks that is customized for an end user to train fingering, breath, and embouchure classification. When this interface is completed, we will work on transferring the generated ANN to the Birl without firmware reprogramming. Instead, we intend to store and load training presets in the instrument itself via onboard EEPROM. The goal is to require a computer only for training, after which the instrument may be unplugged and run from a 12V Li-ion battery, calculating the interpolated output parameters from the neural net and generating the audio synthesis onboard the 32-bit microcontroller.

After these next steps, we will conduct a more formal user study. Five examples of the instrument will be given to professional performing musicians for extended periods of time, soliciting formal feedback from them. An evaluation with non-expert players is not planned, since the goals of the instrument do not include ease of use by non-experts.

7. CONCLUSION

The Birl, an experimental electronic wind instrument, has demonstrated that machine-learning techniques are applicable to parameters of wind controller design such as fingering systems and embouchure detection. While there is much room for improvement on the instrument, the experiments so far are promising and encouraging for further exploration.

8. ACKNOWLEDGEMENTS

Additional authors: Michael Mulshine (Princeton University) and Yicheng Sun (Princeton University).

9. REFERENCES

[1] N. Collins. Low brass: The evolution of trombone propelled electronics. *Leonardo Music Journal*, 1(1):41–44, 1991.

[2] P. R. Cook. Tbone: An interactive waveguide brass instrument synthesis workbench for the next machine. In *Proceedings of the ICMC '91*, pages 297–297, Montreal, Canada, October 1991.

[3] P. R. Cook. A meta-wind-instrument physical model, and a meta-controller for real time performance control. In *Proceedings of the ICMC '92*, pages 273–276, San Jose State University, USA, October 1992.

[4] R. A. Fiebrink. Real-time human interaction with supervised learning algorithms for music composition and performance. *PhD thesis*, Princeton University, Princeton, NJ 2011.

[5] T. Henriques. Double slide controller. In *Proceedings of NIME 2009*, pages 260–261, Pittsburgh, PA, USA, 2009.

[6] A. Hunt, M. Wanderley, and R. Kirk. Towards a model for instrumental mapping in expert musical interaction. In *Proceedings of the ICMC '00*, pages 209–212, Berlin, Germany, 2000.

[7] D. W. Menzies and A. P. McPherson. An electronic bagpipe chanter for automatic recognition of highland piping ornamentation. In *Proceedings of NIME 2012*, Ann Arbor, MI, USA, 2012.

[8] D. W. Menzies and A. P. McPherson. A digital bagpipe chanter system to assist in one-to-one piping tuition. In *Proceedings of the 2013 Stockholm Music Acoustics Conference/Sound and Music Computing Conference (SMAC/SMC)*, Stockholm, Sweden, July 2013.

[9] C. Palacio-Quintin. The hyper-flute. In *Proceedings of NIME 2003*, pages 206–207, Montreal, Canada, 2003.

[10] M. Sato, I. Poupyrev, and C. Harrison. Touché: enhancing touch interaction on humans, screens, liquids, and everyday objects. In *Proceedings of CHI'12*, pages 483–492, Austin, TX, USA, 2012.

[11] G. Scavone. The pipe: explorations with breath control. In *Proceedings of NIME 2003*, pages 15–18, Montreal, Canada, 2003.

[12] G. Scavone and P. R. Cook. Real-time computer modeling of woodwind instruments. In *Proceedings of the 1998 International Symposium of Musical Acoustics*, pages 197–202, Leavenworth, WA, USA, 1998.

[13] G. Scavone and A. da Silva. Frequency content of breath pressure and implications for use in control. In *Proceedings of NIME 2005*, pages 93–96, Vancouver, Canada, 2005.

[14] J. Snyder. Exploration of an adaptable just intonation system. *PhD dissertation*, Columbia University, New York, NY 2010.

[15] S. Ystad and T. Voinier. A virtually real flute. *Computer Music Journal*, 25(2):13–24, Summer 2001.