MICROPHONE-BASED ELECTRONIC WIND INSTRUMENT BY FEATURE EXTRACTION FROM BREATH SIGNALS

Francesco Bigoni
Aalborg University – Copenhagen
fbigon17@student.aau.dk

Pelle Juul Christensen
Aalborg University – Copenhagen
pjch17@student.aau.dk

Rasmus Eklund
Aalborg University – Copenhagen
reklun13@student.aau.dk

Javier Molina Garcia
Aalborg University – Copenhagen
jmolin17@student.aau.dk

Cumhur Erkut
Aalborg University – Copenhagen
cer@create.aau.dk

ABSTRACT

An *electronic wind instrument* is an analog or digital electronic instrument actuated by blowing onto an electronic sensor. Through the history of electronic wind instruments, the refinement of the physical interfaces has not been followed by major innovations regarding breath and embouchure detection: the industry is still largely relying on pressure sensors for measuring air flow. We argue that many sound production techniques for acoustic wind instruments depend on breath *quality* in addition to breath quantity, and that most of the commercially available electronic options do not account for this. A series of breath signal measurements indicated that an electret microphone flush-mounted in a plastic tube is a suitable sensor for feature extraction of the player's breath. Therefore, we propose the implementation of an electronic wind instrument, which captures the envelope and frequency content of the breath and detects whether the signal is voiced or unvoiced. These features are mapped to the parameters of an FM synthesizer, with an external MIDI keyboard providing pitch control. A simple evaluation shows that the proposed implementation is able to capture the intended signal features, and that these translate well into the character of the output signal. A short performance was recorded to demonstrate that our instrument is potentially suitable for live applications.

1. INTRODUCTION

During the late 1970s, breath-controlled electronic instruments started to bloom along with digital keyboards. Yamaha, Roland and most notably Akai (with their *EWI*, the electronic wind instrument *par excellence*), proved themselves to be the leaders in the electronic wind instrument market. Their products try to emulate acoustic wind instruments (woodwind and brass), according to the different types of interface and playing techniques. They enable multiple types of output, ranging from physical models of acoustic wind instruments to abstract synthesizers.

Although the interfaces for electronic wind instruments went through many years of development and refinement, they are still disconnected from their acoustic counterparts when it comes to sound production techniques. The development of a "good" saxophone sound requires a large degree of control over embouchure, reed pressure, air flow and vocal tract configuration; extended techniques such as growling (singing into the horn while blowing), flutter-tonguing (vibrating the false vocal folds while blowing) and multiphonics (playing several tones at the same time) require an even higher degree of awareness regarding the position of the articulators in the vocal tract. However, most electronic wind instruments focus on the mere detection of the airflow by employing a pressure sensor inside the mouthpiece, adding modulation capabilities (e.g. vibrato and pitch bend) by a supplementary bite pressure sensor. It has been shown that the vocal tract configuration of the player has an influence on timbre which goes well beyond capturing bite pressure and air flow [1], suggesting that several modes of sonic interaction through breath are still largely unexplored.

In this paper, we investigate the literature regarding sound production techniques on acoustic wind instruments. Using this knowledge as a background, we build a simple interface that employs a single microphone to capture the breath signal, which is then processed by an audio plug-in that extracts basic audio features in real time and maps them to the parameters of an FM synthesizer.

2. RELATED WORK

2.1 Commercial wind instruments

The Akai *EWI* (Electronic Wind Instrument) is arguably the most popular commercial wind instrument. It has been played by prominent jazz/fusion instrumentalists such as Michael Brecker and Bob Mintzer, but also by more experimental players such as Marshall Allen. The EWI features a clarinet-like form factor, a mouthpiece bite sensor, a breath pressure sensor, an internal sound module, and MIDI capabilities for controlling external synthesizers [2]. The EWI was first invented by Nyle Steiner, who earlier released a trumpet-like version called

EVI (Electronic Valve Instrument). Both were later licensed to Akai.

Yamaha introduced the *WX series* with the launch of the *WX7* wind instrument in 1987 [3] [4]. Like the EWI, the WX series (now discontinued) included a mouthpiece bite sensor for pitch bend. Another major company, Roland, has developed the *Aerophone*, which mounts an actual plastic reed on the mouthpiece and a bite sensor analogous to the ones employed on the EWI and WX series [5]. The plastic reed aims to create a resistance feel analogous to that of an actual reed instrument.

Wind controllers without a dedicated synthesizer module seem to constitute an even smaller niche market. Examples include *TE Control USB MIDI Breath Controller* [6] and *Hornberg Research HB1 Breath Station* [7], which employ similar sensor technologies as the aforementioned electronic wind instruments, with added features such as breathe-in detection (Hornberg) and two-axis accelerometer (TE Control). The early example of Bill Bernardi's *LYRICON* by Computone [8], on the other hand, employs a membrane and photo cell to convert the acoustic signal from a saxophone-like instrument (featuring an actual reed, mouthpiece and keys) to send a control voltage to an external synthesizer module — an interface design that, to our knowledge, has not been replicated since.

2.2 Academic research on wind instruments

Although commercial electronic wind instruments are largely standardized, experimental alternatives can be found in the world of academia.

HIRN by Perry R. Cook exhibits an innovative take on an electronic wind instrument, where breath pressure is complemented by an array of other features: bite tension, lip tension (through a myoelectric sensor), pitch when singing or buzzing into the mouthpiece, fingering (through buttons) and spatial position, as well as three rotation controls – one on the mouthpiece and two on the body – plus a linear slide control for the right hand. This configuration, which provides multiple dimensions of expression, is mapped to the parameters of a physical model which combines reed excitation (e.g. clarinet), jet excitation (e.g. flute) and lip excitation (e.g. trombone), allowing a continuous variation between these timbral qualities , through manipulation of the interface [9].

The Pipe by Gary P. Scavone employs a pressure sensor, for measuring the static pressure in a closed, "flow-free" pressure chamber; force-sensitive resistors, for emulating the continuous nature of tone holes in acoustic woodwind instruments; an accelerometer, for measuring the angle at which the instrument is held. Scavone points out that breath is an "intuitive" excitation signal for driving physical models of sustained sounds and argues that the addition of a miniature microphone in the mouthpiece would allow to retrieve information on the player's vocal tract [10].

More recently, *The Birl* by Jeff Snyder and Danny Ryan [11] utilized a pressure sensor for tracking the airflow and capacitive sensors for tracking tone hole fingering and lip position, allowing *artificial neural networks* (ANNs) to create a mapping from tone hole and embouchure positions to the pitch and timbre of a synthesis algorithm. The ANNs were trained by expert performers to generate the appropriate mappings. The authors noted that ANN-produced mapping improved pitch linearity and continuity for shifting tone-hole positions.

The aforementioned examples show several ways of mapping multidimensional features such as fingering and lip position. However, none of them (except, to some extent, HIRN) seems to relate inherently to sound production techniques that are pertinent to acoustic wind instruments. On the other hand, an enduring lineage of influential saxophone teachers such as Sigurd Rascher, Joe Allard, and Allard's student Dave Liebman, has been teaching how to place the articulators in the vocal tract in order to develop a good and personal saxophone sound [12]. Extended techniques such as multiphonics or growling require an extraordinarily high degree of awareness of the vocal tract. In the following, we outline two studies that are supporting Rascher and Allard's approach.

Scavone et al. [1] measured the influence of vocal tract configuration on different saxophone blowing techniques by analysing the transfer function between upstream and downstream acoustic pressures. The terms *upstream* and *downstream* refer to the air columns on the upper teeth and reed side of the saxophone mouthpiece respectively. Since the placement of microphones into the mouthpiece was difficult due to the small dimensions and the high sound pressures reached, the authors employed pressure transducers instead. Their tests covered both conventional and extended registers, pitch bending, multiphonics, as well as bugling — playing an overtone series without a change in fingering, by varying the vocal tract configuration. The authors demonstrated that vocal tract formation has a fundamental influence on timbre for a plain sound in the conventional register, but also for the outer registers and the extended techniques, which are unplayable without the right placement of the articulators. One trial showed that even when a player is asked to pitch-bend by varying their lip pressure exclusively, some degree of vocal tract manipulation would still be detected [1].

In another study, Scavone et al. measured the frequency content of breath pressure signals applied to wind instruments [13]. Their measurements showed that breath signals contain frequencies up to 10 kHz, with a significant amount of energy in the first 1 kHz. The authors tested several different pressure sensors, arguing that even though some sensors have an upper frequency limit of up to 2.5 kHz, MIDI wind instruments cannot be expected to support more than 500 Hz of breath pressure bandwidth, due to the constraints of MIDI specifications. Furthermore, Yamaha WX11 and WX5 are found to have an input sampling rate of 320 Hz and 280 Hz respectively, which make them unsuitable for tracking extended techniques that require breath pressure modulation at audio rate, e.g. flutter-tonguing (i.e. vibrating the false vocal folds while blowing) and growling. On the whole, these studies indicate that the mappings chosen in most commercial

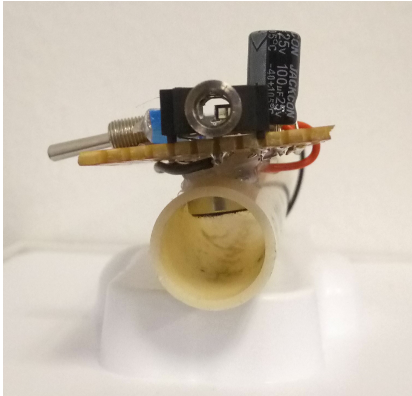


Figure 1: Wind instrument prototype (front view) consisting of an omnidirectional electret microphone (diameter = 0.8 cm); a 3 V button cell battery feeding the microphone with a supply voltage through a simple circuit; a plastic tube (length = 10 cm, diameter = 1.5 cm) with a round hole at 1 cm from the tail end, for flush-mounting the microphone.

breath controllers are disputable when taking the acoustics of wind instruments into account.

3. DESIGN

All the modern commercial wind instruments we have encountered employ a pressure sensor for detecting the air flow, and this technology does not seem to have been updated throughout the years. In the previous section, we have shown that these common pressure sensors are not suitable for capturing the nuances of breath signals [13]. Scavone et al. mention the disconnect felt by many woodwind performers while performing on the currently available electronic wind instruments, and stress the need for a more versatile instrument, able to achieve a full range of sensory input [13]. We argue that the quality of the breath signal has a significant impact on the sound production techniques of acoustic wind instruments, and that this fact is not reflected by their electronic counterparts.

Therefore, we set to design a *microphone-based* wind instrument, which allows us to extract currently unused features of breath signals and map them to the parameters of a synthesizer. For the scope of this paper, we had the following requirements: 1) accurate detection of the breath envelope, allowing for a playing experience which is both smooth and responsive; 2) analysis of the breath signal, which makes it possible to indirectly extract information about the vocal tract manipulation; and 3) discrimination between vocalized and unvocalized breath, allowing the player to sing into the mouthpiece to achieve special effects.

3.1 Breath signal measurements

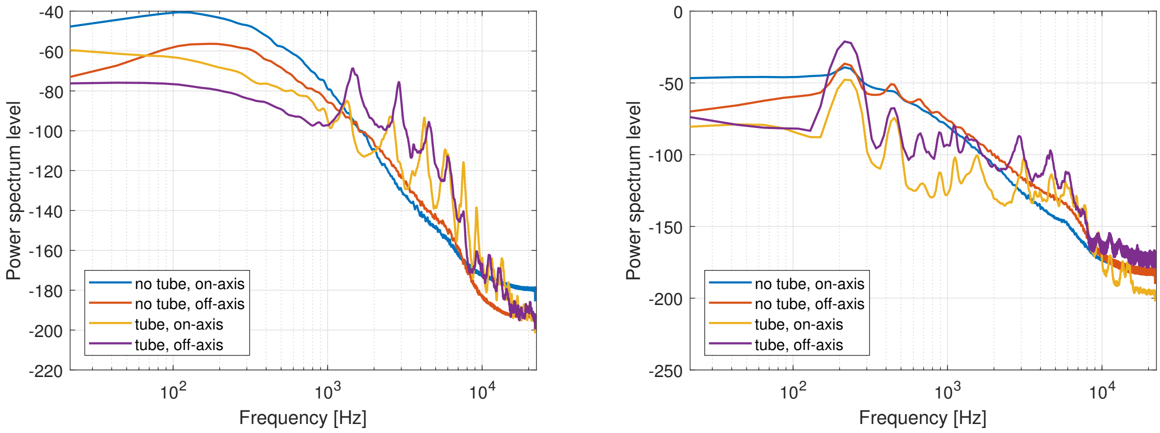
We assume that useful features describing the quality of the player's blow could be extracted from the breath signal by using a simple miniature microphone as an interface. Therefore, a series of preliminary tests were carried out, with the purpose of pinpointing different features of breath and understanding the potential issues of such a setup. The making of a solid and playable interface goes beyond the scope of this paper. However, it is worth investigating if adding a mouthpiece (in our case, a short plastic tube) and varying the incidence angle of the player's blow into the microphone gives different results than e.g. blowing directly into a mobile phone headset microphone. For this purpose, a simple prototype consisting of a cheap electret microphone flush-mounted into a plastic tube through a hole on the side wall was realized as shown in Figure 1. The microphone output was recorded while a player was blowing into it with and without the aid of the plastic tube.

Three different types of actions were performed: 1) blowing into the microphone without any conscious vocal tract formation; 2) blowing into the microphone while singing a pitched ”u” sound; 3) blowing into the microphone while varying the vocal tract configuration, i.e. while forming the vowels ”u”, ”a”, ”e” and ”i” (unpitched) at a tempo of 120 bpm, where each vowel gets two beats.

Each action was recorded with the following setups: a) on-axis blow into the microphone in free air ($\theta = 0°$); b) off-axis blow into the microphone in free air ($\theta = 90°$); c) on-axis blow into the microphone, placed at the tail end of the tube ($\theta = 0°$, d_{tube} = 1.5 cm, l_{tube} = 10 cm); d) off-axis blow into the microphone, flush-mounted on the side wall of the tube at 1 cm from the tail end ($\theta = 90°$, d_{tube} = 1.5 cm, l_{tube} = 10 cm), where θ is the incidence angle, d_{tube} is the tube diameter and l_{tube} is the tube length.

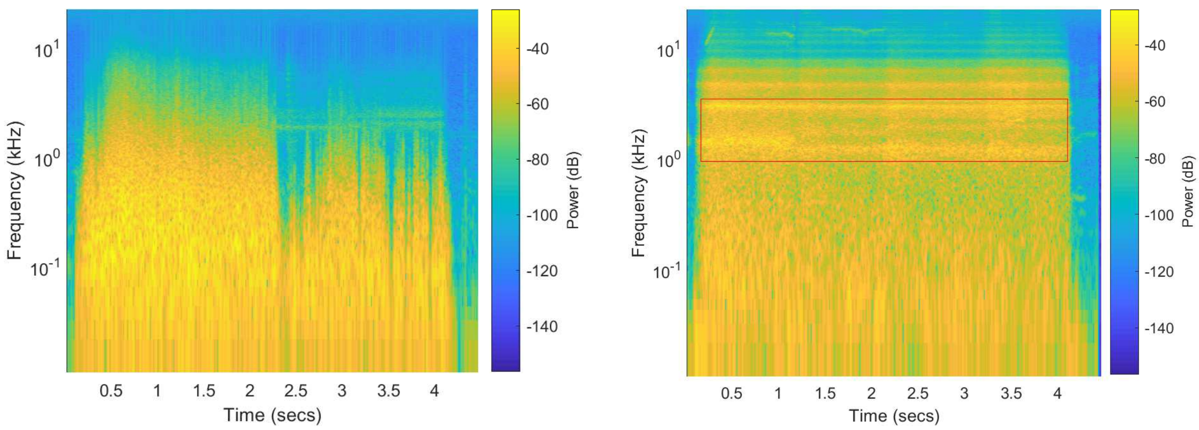
We expected that on-axis incidence would produce a strong DC component in the signal, resulting in a loss of sensitivity, as the microphone membrane is pushed towards the back electrode by the air stream. By flush-mounting the microphone on the side wall, the DC component should be attenuated and the microphone sensitivity should be preserved.

Figures 2a and 2b show the averaged power spectra of a normal blow (i.e. no conscious vocal tract formation) and a voiced blow (i.e. singing a pitched ”u” sound while blowing) respectively. It can be noted that switching from on-axis to off-axis incidence provides a 40 dB attenuation at DC in all cases. The power spectra of normal blows in free air are smooth (with most energy at low frequencies) since a normal blow is essentially a noise signal. When the tube is added, harmonic content appears above 1 kHz (both on-axis and off-axis) due to standing waves in the tube. This is an unwanted effect, as these resonances are not contained in the actual breath signal, but effectively added by the tube. However, the advantages of such a construction become evident by looking at the power spectra of voiced blows in Figure 2b: voiced blows are basically undetectable with a free air mount, as the harmonic content is barely visible (and only for off-axis incidence). When using the tube, the fundamental frequency and its harmonics (produced by the vibrating vocal folds of the player) are well visible as low-frequency harmonic content (and, in our case, well separated from the



(a) (b)

Figure 2: Power spectra of normal blows (a) and voiced blows (b) with/without the tube for on-axis/off-axis incidence.



(a) (b)

Figure 3: Spectrograms of varied vocal tract configurations for off-axis incidence, without the tube (a), and for off-axis incidence, with the tube (b). The red box highlights the region (around 2-4 kHz) where formants are emerging behind the standing-wave pattern.

standing wave pattern in the tube). Moreover, the initial test has shown the practical advantages of including a mouthpiece in the prototype. According to the performer's feedback, when blowing through a tube it was easier to make sure that the stream of air hit the microphone evenly at any point in time; without the tube (and with pursed lips) it was easy to mistakenly direct the stream of air away from the microphone; when the microphone was installed at the tube tail end (for on-axis incidence) it blocked the air stream, making it harder to control the vocal tract configuration.

We can also assume that adding a simple mouthpiece approximates the impedance change at the mouthpiece that is experienced when dealing with acoustic wind instruments under normal playing conditions [13], thereby offering a more realistic "resistance" feel to the performer.

Figure 3 shows spectrograms of varying vocal tract configurations for off-axis incidence, in free air and with the tube respectively. When blowing into the microphone in free air, the acoustic energy is mainly located at low frequencies and there is very little trace of formants. On the other hand, a couple of formants (boxed in red in Figure 3b) are emerging when the tube is added, between the first and the second tube resonance and between the second and the third resonance respectively (with the lowest formant clearly moving while the performer alters the position of the articulators). Even though these formants are drowning into the standing wave pattern (due to the unwanted effect of the tube) it is fair to assume that manipulating the vocal tract configuration would have a measurable impact on the frequency distribution of the breath signal, at least at loud dynamics.

On the whole, flush-mounting the microphone on the side wall of a small plastic tube offers the following advantages: reduced DC component and, thereby, preserved microphone sensitivity; enhanced experience

Purpose	Required function
Event detection, time course, dynamics	Envelope detection
Timbre, dynamics, vocal tract configuration	Spectral centroid
Discriminate between voiced and unvoiced sounds	Harmonicity ratio

Table 1: Desired functionality vs required implementation blocks.

for the performer: correct impedance shift, ease of focusing the air stream, ease of vocalizing and varying the vocal tract formation; ease of detection of voiced vs. unvoiced sounds; richer high-frequency content, with partly disclosed formants during vocal tract formation. This has the disadvantage of added resonances due to standing waves in the tube. Therefore, the prototype shown on Figure 1 was chosen for the current implementation.

3.2 Feature extraction

To achieve the requirements presented in Section 3, the implementation blocks outlined in Table 1 are needed.

The envelope of the input signal is detected by employing a leaky integrator, which is an envelope follower defined by the equation [14]:

$$e\,[n] = (1 - b\,[n])|x\,[n]\,| + b\,[n]\,e\,[n-1]\,, \tag{1}$$

where $|x\,[n]\,|$ is the rectified input signal and $e\,[n]$ is the resulting envelope. This is equivalent to implementing a weighted running average, i.e. a first-order IIR lowpass filter. However, the leaky integrator differs from a linear integrator in that its coefficient $b\,[n]$ depends on whether the signal is on a rising or a falling edge. $b\,[n]$ is in fact defined as

$$b\,[n] = \begin{cases} b_r, & \text{if } |x\,[n]\,| > e\,[n-1] \\ b_f, & \text{if } |x\,[n]\,| \leq e\,[n-1]\,, \end{cases} \tag{2}$$

where b_r is the rising-edge (or attack) coefficient and b_f is the falling-edge (or decay) coefficient. By manipulating these coefficients, it is possible to adjust the integration time of the filter on both edges, varying between faster (but more affected by overshoot and ripple) and slower (but smoother) attacks/decays.

By calculating the spectral centroid, which can be thought of as the center of mass of the spectrum of the signal, information regarding the desired sound brightness (aiming for a perceptual mapping) can be extracted. Even though the centroid is unsuitable for inferring a complete description of the vocal tract configuration, we argue that it offers at least an indication regarding the manipulation of the mouth cavity, since forming different vowels while blowing into the tube cause a change in brightness of the resulting sound. For instance, an *"i"* vocal tract configuration will result in a greater high-frequency content than an *"a"*.

The spectral centroid (C) is calculated as the weighted mean of the frequencies of the spectrum [15]:

$$C = \frac{\sum_{i=0}^{N-1} f_i \cdot |F_i|}{\sum_{i=0}^{N-1} |F_i|}, \tag{3}$$

where N is the number of frequency bins, F_i is the magnitude of the i'th frequency bin of the DFT of the input signal and f_i is the center frequency of bin i.

For voiced blows, we assume that the frequency content of the signal will be more harmonic, rather than noisy. The signal should show a fundamental frequency and a number of harmonics with amplitudes well above the noise floor. The harmonicity of a signal can be discriminated in a simple way by calculating the harmonicity ratio (HR), which is the ratio between the energy of the highest peaks of the DFT of the input signal and the total energy of the signal:

$$HR = \frac{\sum_{i=0}^{K-1} \text{DESC}(|F|^2)_i}{\sum_{i=0}^{N-1} |F_i|^2}, \tag{4}$$

where K is the number of peaks to evaluate and DESC(X) is the function that sorts X in descending order. The harmonicity ratio is used in a number of contexts: among them, the audio section of the MPEG-7 standard (in that case, calculated using the autocorrelation inside each audio frame, with analogous results) [16].

4. IMPLEMENTATION

The microphone signal received from the construction presented in Section 3 is fed into the feature extraction algorithms of Section 3.2. These features should be mapped to the parameters of a sound synthesis algorithm.

The choices of mapping and synthesis are crucial to both the aesthetics and interaction of our instrument. We chose a monophonic FM synthesizer for several reasons: a) it is straightforward to implement, b) it offers a rich palette of sounds, c) it has been used with other electronic wind instruments [11], d) it offers a perceptual mapping of the spectral centroid to the FM modulation index, which controls the bandwidth of the synthesized signal [17].

Table 2 offers an overview on the chosen mapping. The envelope is mapped directly to the amplitude of the synthesized signal. The spectral centroid is scaled by a user-defined constant and mapped to the modulation index.

Once a vocalized blow pushes the harmonicity ratio above a user-defined hysteretic threshold, the envelope is scaled by an irrational number and summed to the base (rational) FM carrier-modulator frequency ratio. This causes a shift from a harmonic to an inharmonic spectrum, reminiscent of the ring modulation-like growl produced by vocalizing into a brass or woodwind instrument.

As the current mapping is lacking pitch control, due to the fact that it our prototype is not a full interface — no keys, buttons or tone holes are present — a small MIDI keyboard (Korg nanoKEY 2) has been added to the setup, providing MIDI pitch control. The described feature extractors, sound synthesis algorithm and mapping were implemented as two DAW audio plugins in *JUCE* [18], supporting both VST (in our case, 2.0) and AudioUnits from the same codebase.

Feature	Type of mapping	Synthesis parameter
Envelope (input signal)	Linear (direct)	Carrier envelope and FM ratio offset envelope for voiced sounds
Spectral centroid	Linear (scaled)	Modulation index (scaled and smoothed)
Harmonicity ratio	Threshold	FM ratio offset (gated on voiced sounds)

Table 2: Features-to-synthesis mapping.

Our application was split into two parts for handling MIDI and audio separately. The MIDI plug-in sends the MIDI messages from the external keyboard as OSC packets, which are then received by the audio processing plug-in. The audio processing plug-in exposes a GUI which makes all the user-defined constants available to the user: centroid scaling and smoothing factors, harmonicity threshold, envelope detection coefficients (b_r and b_f), base FM ratio, input gain and dry/wet mix. To prevent clicks and artifacts during MIDI control or automation, all of the available GUI parameters were smoothed. Finally, we included a brief tail-out period to avoid clicks at NOTE OFF messages.

5. EVALUATION

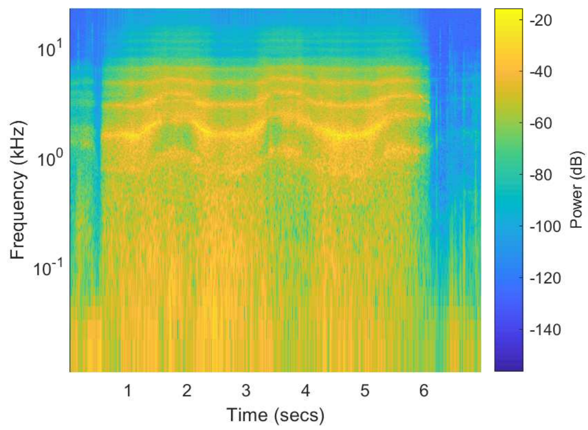
As a general performance test, the plug-ins were loaded both in the commercial DAW Reaper, as well as MATLAB's Audio Test Bench. Although the code was not optimized for all sample rates and block sizes, the plug-ins seem to perform well in Reaper at a sample rate of 44100 Hz and with an block size of 512 samples. Neither remarkable latency, clicks or dropouts were observed. The Audio Test Bench showed no buffer underruns or overruns.

To evaluate the breath signal feature mapping, four scenarios were recorded. The first three focus on the variation of spectral centroid, envelope and harmonicity ratio as isolated features. The last is a short musical example exploring the possibilities of the instrument in a performance scenario, in which several techniques, such as flutter tonguing, changing vocal tract configuration, and vocalising while blowing, were used. The input and output files for all scenarios are available online on SoundCloud [1] .

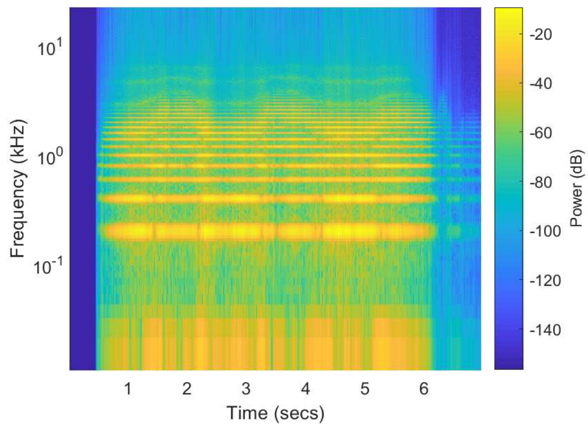
Figure 4 shows the outcome of the centroid scenario. At least two formants are clearly gliding up and down in the input signal plot (Figure 4a). Accordingly, the output signal (Figure 4b, recorded at 20% dry / 80% wet signal) shows that the modulation index follows the centroid, increasing the sideband energy at the expected times.

The envelope detection scenario is shown in Figure 5. The rising-edge and falling-edge coefficients of the leaky integrator in Eq. (2) were respectively set to 0.995 and 0.9 during the recording. These values were chosen to

[1] `https://soundcloud.com/francescobigoni/sets/electronic-wind-controller/s-cRdgk`



(a)



(b)

Figure 4: Spectral centroid scenario. The performer alternated between open (*"a"*) and closed (*"e"*) vowels, while blowing normally. The harmonic content of the output (4a) follows the input signal (4b).

provide an acceptable trade-off between smoothness and responsiveness. The output signal (Figure 5b) follows the envelope of the input signal at all dynamics, including the short transients.

Finally, the harmonicity ratio scenario is shown in Figure 6. In this case, the output signal (6b) is smoothly alternating between harmonic and inharmonic spectral content.

For the performance scenario the performer (author FB) noted a fair degree of expressiveness, given the early stage of the interface, which is not yet ready for a proper live setting. At a later stage of development, a more through evaluation should be conducted, to assess how our instrument fares in a performance situation. Given the fact that feature extraction and mapping worked as expected, and despite the fact that there is room for improvement, we find it fair to conclude that our wind instrument is fulfilling the implementation requirements we set forth.

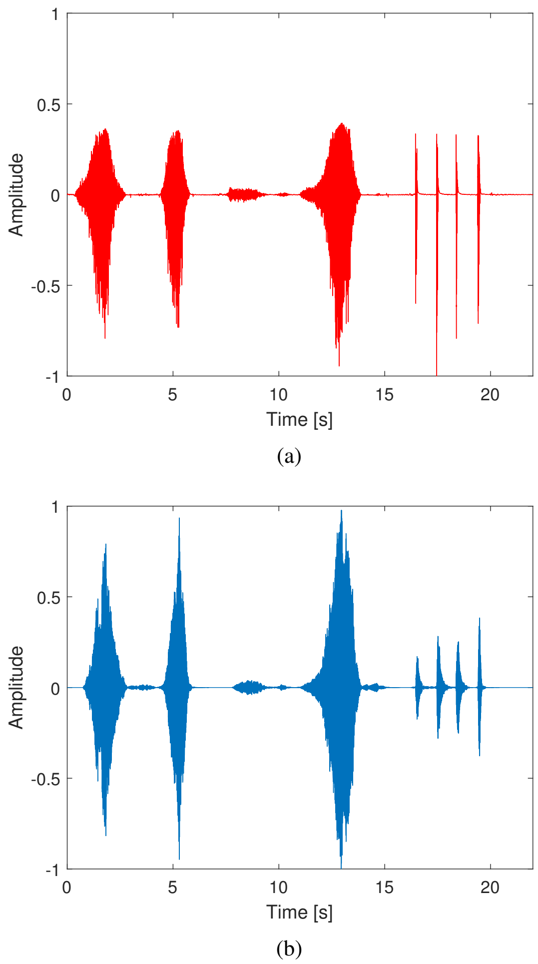


Figure 5: Envelope scenario. The performer sought to utilize different dynamics and durations, such as long/short, soft/loud, and abrupt/longer events, with no conscious vocal tract formation. The input signal is shown in 5a, the output signal in 5b.

6. FUTURE WORK

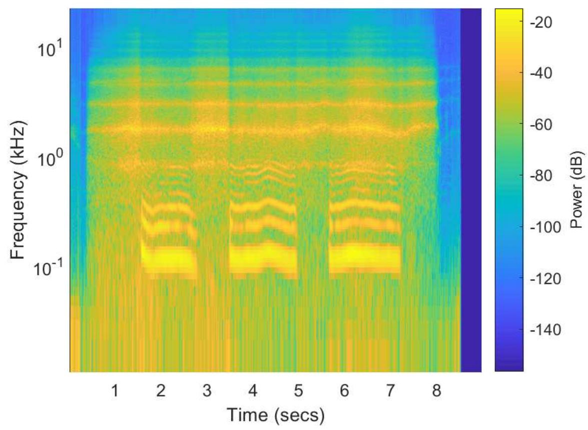
Although our wind instrument fulfills the implementation requirements, there is room for improvement and further development. In the following, the two main development areas are outlined.

Interface: Our prototype, made using a small plastic tube as a mouthpiece, constitutes a basic interface. However, the embodied aspects of interaction with the wind instrument should be investigated. At the present time we are employing a single omnidirectional microphone, but other kind of sensors could be added. For instance, signal detection for feature extraction could be split between a pressure sensor for envelope detection and a microphone for frequency analysis. This would also improve noise immunity: our setup is prone to noise and feedback in proximity of other sound sources. Pitch control could be obtained by using keys, tone holes or continuous sliders. For less traditional modes of interaction, an accelerometer could be employed to track the movement of the controller during the performance, either for adding multiple dimensions to sound synthesis (in an analogous way to HIRN [9]): studies of clarinet performances show a correlation between instrument movement and performance character [19]. Ideally, the instrument should become an embedded system, with its built-in microcontroller.

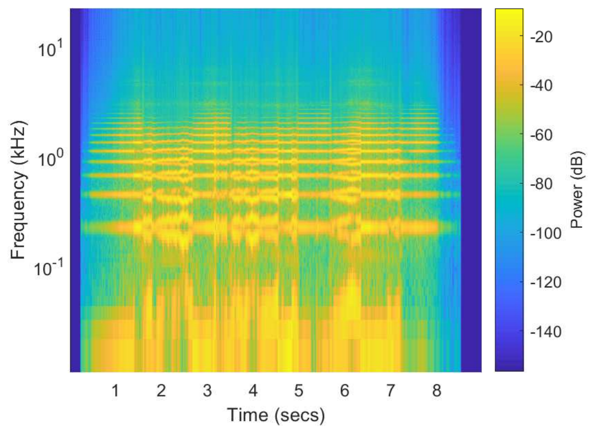
Audio signal processing: we argue that tube resonances are not a major issue with respect to the current mapping, even though they establish a (small) error source for the SC and HR calculations. However, since these resonances are invariant (being only dependent on the tube length) they could be easily removed by implementing a linear predictive coding algorithm. Envelope, spectral centroid and harmonicity ratio seem to capture some of the nuances of a wind instrument performance; however, other audio descriptors (e.g. spectral envelope and fundamental frequency) could be investigated for improved control. As an alternative to (or in combination with) feature extraction, machine learning algorithms could provide a relevant way of classifying breath quality. Finally, the application of our method to other audio engines, possibly with contrasting types of mapping, would provide a more extensive evaluation framework.

7. CONCLUSION

A microphone-based wind instrument has been implemented, with the purpose of extracting basic audio features of breath and mapping them to the parameters of an FM synthesizer. Through research on the state of the art of electronic wind instruments, it was shown that there is room for enhancing expressiveness by exploring the frequency content of breath signals. We argue that capturing the full bandwidth of the breath signal provide information regarding the player's vocal tract configuration, establishing a direct link between the sound production techniques that are inherent to acoustic wind instruments and our implementation. A series of measurements of breath signals in multiple configurations supports these assumptions. A simple prototype, consisting of an omnidirectional electret microphone flush-mounted in the side wall of a small plastic tube, supplied the best conditions to record the breath signal and extract its features in real time. We found that changing the vocal tract configuration (by forming different vowels while blowing) has an effect on the spectral centroid, which moves up when closing the vowel and down when opening it. Since the 90° mounting reduces the strong DC component of the signal, this setup allows to discriminate between voiced and unvoiced sounds in a relatively easy way. Based on the aforementioned findings, an audio plug-in was implemented that captures breath signals in real time and maps audio features are to the parameters of an FM synthesizer. The GUI allows the player to tweak feature extraction and synthesis parameters during the performance. A simple technical evaluation, showing envelope, centroid and harmonicity ratio extraction in a separate manner, proves that the proposed implementation carries out the desired functions with no remarkable latency, clicks or artifacts at the output.



(a)



(b)

Figure 6: Harmonicity ratio scenario: the performer alternates between unvoiced and voices blows, shifting the harmonicity ratio. The input signal is shown in 6a, the output signal in 6b.

8. REFERENCES

[1] G. Scavone, A. Lefebvre, and A. Silva, "Measurement of vocal-tract influence during saxophone performance," in *The Journal of the Acoustical Society of America*, vol. 123, May 2008, pp. 2391–2400.

[2] "The EWI Story by Joel Peskin," accessed: 4/12/2017. [Online]. Available: http://www.patchmanmusic.com/articles/JoelPeskinEWIStory.html

[3] "Yamaha WX5," accessed: 18/12/2017. [Online]. Available: https://www.emusician.com/gear/yamaha-wx5

[4] "Yamaha WX7 – Good Design Gold Award," accessed: 18/12/2017. [Online]. Available: https://www.yamaha.com/en/about/design/synapses/id_006/

[5] Roland Corporation, "Aerophone AE-10 — Digital Wind Instrument," accessed: 5/10/2017. [Online]. Available: https://www.roland.com/us/products/aerophone_ae-10/

[6] TEControl, "USB MIDI Breath Controller," http://tecontrol.se/products/usb-midi-breath-controller, accessed: 5/10/2017.

[7] Hornberg Research, "Hornberg Research Products," https://www.hornberg-research.de/index.php/en/breath-controller-products-en, accessed: 5/10/2017.

[8] "Lyricon," accessed: 5/10/2017. [Online]. Available: http://www.jorritdijkstra.com/lyricon/

[9] P. Cook, "A meta-wind-instrument physical model, and a meta-controller for real time performance control," in *Proceedings of the ICMC '92*. San Jose State University, USA, Oct 1992, pp. 273–276.

[10] G. P. Scavone, "The Pipe: explorations with breath control," in *Proceedings of the 2003 conference on New interfaces for musical expression*. National University of Singapore, 2003, pp. 15–18.

[11] J. Snyder, "The Birl: Adventures in the development of an electronic wind instrument," in *Musical Instruments in the 21st Century: Identities, Configurations, Practices*, T. Bovermann, A. de Campo, H. Egermann, S.-I. Hardjowirogo, and S. Weinzierl, Eds. Springer, 2017, pp. 181–205.

[12] D. Liebman, *Developing a Personal Saxophone Sound*, 2nd ed. Dorn Publications, 1994.

[13] G. Scavone and A. da Silva, "Frequency content of breath pressure and implications for use in control," in *Proceedings of the 2005 Conference on New Interfaces for Musical Expression*. National University of Singapore, 2005, pp. 93–96.

[14] L. Peltola, C. Erkut, P. R. Cook, and V. Valimaki, "Synthesis of Hand Clapping Sounds," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 15, no. 3, pp. 1021–1029, Mar 2007.

[15] A. Farnell, *Designing Sound*. The MIT Press, 2010.

[16] J. Luque, D. F. Larios, E. Personal, J. Barbancho, and C. Len, "Evaluation of MPEG-7-Based Audio Descriptors for Animal Voice Recognition over Wireless Acoustic Sensor Networks," *Sensors (Basel, Switzerland)*, vol. 16, no. 5, May 2016.

[17] J. M. Chowning, "The synthesis of complex audio spectra by means of frequency modulation," *Computer Music Journal*, vol. 1.2, p. 46, Apr 1977.

[18] "JUCE," accessed: 14/12/2017. [Online]. Available: https://juce.com/

[19] E. C. Teixeira, M. A. Loureiro, M. M. Wanderley, and H. C. Yehia, "Motion analysis of clarinet performers," in *Journal of New Music Research*, vol. 44, Apr 2015, pp. 97–111.