EXPLORATIONS IN CONVOLUTIONAL SYNTHESIS

Tamara Smyth, Andrew R. Elmore
School of Computing Science, Simon Fraser University
tamaras@cs.sfu.ca, aelmore@sfu.ca

ABSTRACT

In this work we further explore a previously proposed technique, used in the context of physical modeling synthesis, whereby a waveguide structure is replaced by a low-latency convolution operation with an impulse response. By doing so, there is no longer the constraint that successive arrivals be uniformly spaced, nor need they decay exponentially as they must in a waveguide structure. The structure of an impulse response corresponding to an acoustic tube is discussed, with possible synthesis parameters identified. Suggestions are made for departing from a physically-constrained structure, looking in particular at impulse responses that are mathematically-based and/or that correspond to hybrid or multi-phonic instruments by interleaving two or more impulse responses. The result is an exploration of virtual musical instruments that are either based on physical instruments, completely imagined, or somewhere in between.

1 INTRODUCTION

It is common practice for contemporary musicians to explore playing techniques by producing sounds on their instruments that may be unusual, surprising or perhaps completely atypical of the instrument. In the case of wind instruments, extended playing techniques such as side and flutter tonguing, false fingering, split tones, circular breathing, are typically practiced by very accomplished musicians as who have highly proficient and virtuosic control over their airflow and embouchure. By employing unusual playing techniques, the player achieves a sound that seemingly extends or redefines the instrument itself.

In addition to extending playing techniques, it has become increasingly common for modern musicians to *effectively* modify the instrument itself, both offline and during real-time performance. Some notable examples include recent performances by clarinetist Francois Houle creating a polyphonic effect by blowing into two different instruments simultaneously, extending the clarinet by directing the radiated sound onto the strings and soundboard of an open grand piano, or removing parts of the clarinet bore during performance. Modern trumpet players, such as slide trumpet player Steve Bernstein, often use a technique where the microphone is *swallowed* by the trumpet, effectively altering the horn's frequency response by removing the radiation transfer function of the flare, producing a low-pass sound similar to a rubber mallet dragging along the taut skin of a drum. Many musicians will also explore custom built instruments, with slight modification to the shape of the bore, the bell or the mouthpiece.

Some such instrument extensions can be awkward, costly or impractical, and limiting if the musician doesn't want the change to be permanent or damage the instrument. In any event, no matter what physical modifications are made, the instrument will always be constrained by the acoustics governing the system, which in the case of wind instruments, will yield a periodic impulse response with uniformly spaced arrivals (or echos) corresponding to the length of the instrument, with each consecutive arrival decaying over time due to reflection and propagation losses. Changing or removing components of the instrument will alter the frequency response of these losses, but they will always have a low-pass characteristic and an exponential decay in amplitude, and the impulse response will always be periodic. Convolutional synthesis would allow for greater possibilities related to the practice of instrument extension since, unlike the impulse responses corresponding to a waveguide synthesis model (which aims to model a physical structure), the employed impulse responses are not necessarily physically constrained.

In this work we further explore this previously proposed technique of convolutional synthesis [5], and propose a solution whereby the performer may enhance or replace their instrument all together with a synthesized parametric impulse response. That is, if the signal generated by the reed may be estimated from the signal recorded at the bell during a real-time performance of the instrument [6], it may be convolved directly with a new parametric impulse response corresponding to a new instrument, either physics-based, completely imagined, or perhaps somewhere in between. Alternatively, rather than using the estimated reed pulse directly, it may be used to extract key playing parameters that can then be remapped to the control parameters of a synthesized source, perhaps a parametric pulse train or a more rigorous physics-based model of a generalized-reed [7], which may then be convolved with a new parametric impulse response.

SMC 2009, July 23-25, Porto, Portugal

In both cases, using a low-latency convolution operation [1] as described in Section 4, will allow for a level of real-time interactive control comparable to a waveguide model.

The term *Convolutional Synthesis* borrows from the term *Convolutional Reverb*, a technique used for simulating reverberation of a physical or virtual space by using its characteristic impulse response. In this work, measurements of acoustic tubes, and musical instrument bores, are used as a base from which a parametric impulse response may be synthesized, corresponding to possible wind instrument bores, either existing, or imagined.

2 PARAMETRIC IMPULSE RESPONSES

2.1 Synthesizing one-dimensional propagating waves

A physics-based model of a wind instrument usually consists of a source (the reed pulse) and a filter (the bore, bell and possibly the mouthpiece). Depending on the reed configuration, the flow from the reed may have a strong or weak coupling to the connected acoustic tube, that is, the resonance of the bore and bell may have a varying degree of influence over the resonance or the reed pulse. The oscillating reed provides an excitation to the bore by modulating the air flowing through an aperture with a time-varying cross-sectional area, providing a pressure input. This input is effectively convolved with the instrument's impulse response (as measured at the reed end) to produce the pressure at the base of the bore (or mouthpiece).

Pressure waves travelling along the instrument bore are frequently modeled using the digital waveguide [3] structure seen in Figure 1. This enables use of appropriate waveguide elements to account for propagation losses, $\lambda(\omega)$, reflection losses at the tube ends $R_{cl}(\omega)$ and $R_{op}(\omega)$, and a change of tube length corresponding to the delay z^{-L}.

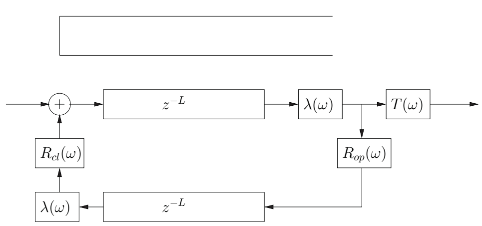

Figure 1. A waveguide model of a cylinder closed at one end with reflection $R_{cl}(\omega)$ and open at the other with reflection $R_{op}(\omega)$.

The impulse corresponding to the waveguide structure in Figure 1, as driven and measured at the closed end, is fairly straightforward to synthesize by following the signal flow of the propagating wave in response to an applied impulse: if an impulse enters the tube at the (not perfectly) closed end, the pulse would be measured at that location at that point in time. After the impulse travels round-trip to the bell and back, a second arrival, A_2, would be measured $2L$ samples later at the bore base, and would consist of the initial pulse filtered according to

$$A_2 = \lambda^2(\omega)R_{op}(\omega)(1 + R_{cl}(\omega)), \tag{1}$$

where the term $(1+R_{cl}(\omega))$ corresponds to the need to sum the left and reflected right going traveling waves to obtain the actual pressure as recorded at the *closed* end. The arrival A_2 would also travel down to the bell and back, and produce a third arrival, consisting of A_2 with additional filtering according to

$$A_3 = \lambda^4(\omega)R_{op}^2(\omega)R_{cl}(\omega)(1 + R_{cl}(\omega)). \tag{2}$$

Every subsequent arrival would be the result of the previous arrival having been subjected to another round of wall and reflection losses.

Figure 2 shows the impulse response corresponding to the closed-open tube and waveguide structure shown in Figure 1. The impulse response has a very clear periodic structure, with uniformly spaced sequence of arrivals corresponding to the length of the tube. The losses in the system have a low-pass characteristic, causing an overall exponential decay, with each arrival being increasingly smeared in time.

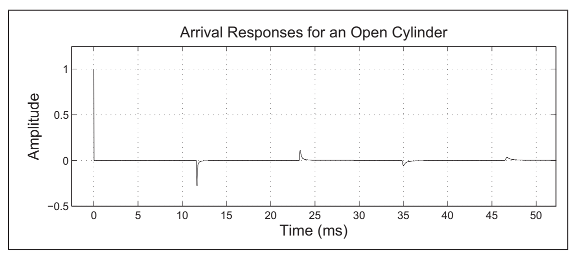

Figure 2. A synthesized impulse response corresponding to a cylindrical acoustic tube closed at one end and open at the other.

This impulse response may be synthesized by creating a sequence of arrivals, each one containing the appropriate filtering corresponding to (1) and (2), up to the number of desired arrivals. One parameter that may be made available to the user in this case is the spacing between the arrivals, which may be stretched or contracted (which would correspond to a change in delay z^L in in Fig. **??**) depending on the desired pitch. Alternate filters for R_{op} (theoretical, measured, or imagined) may be swapped in to the impulse response during performance, corresponding to a real-time change of the bell shape. Similarly, the transfer function R_{cl} may be modified to simulate a change in reed: a lip reed for example, would have more significant loss than would a woodwind reed.

2.2 Other Possible Impulse Responses

Regardless of the filter transfer functions or the delay length L, the impulse response corresponding to Figure 1 will always have uniformly spaced arrivals, and if stable, an overall amplitude envelope that decays over time. Yet when synthesizing the impulse response directly, that is, by not using a waveguide model, there is no need to adhere to these physical constraints. The period between pulse arrivals need not be uniform but may have some other structure, perhaps randomly distributed, or according to some mathematical model such as the golden ratio as in Figure 5. It may also be possible to create hybrid or multi-phonic instruments by interleaving two separate impulse responses (as in Figure 3). Also, the sequences need not decay exponentially; they may grow and then suddenly drop, with rates specified parametrically if so desired.

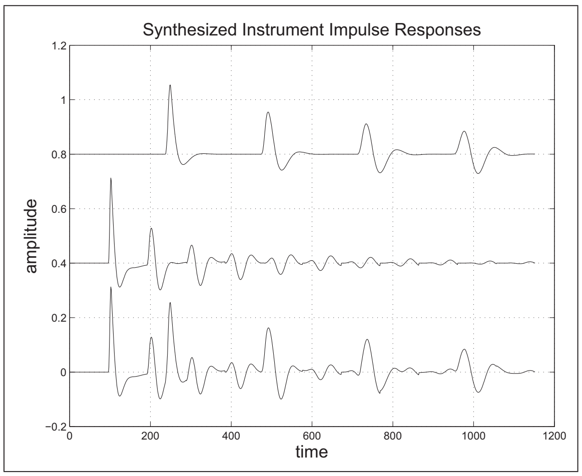

Figure 3. An impulse response is synthesized by interleaving two separate impulse response (top and middle) to produce a multi-phonic impulse response (bottom).

It is also possible to use an entire impulse response as measured from an instrument, such as the one shown for a trumpet in Figure 4. Though periodicity is less visible here, the impulse response is still composed of a sequence of arrivals, each having a transfer function corresponding to any losses in the system, most notably the wall losses and the reflection at the bell. This impulse response could be used as is, or by changing identified parameters much like was done in [2], but instead of replacing leading zeros of a measured reflection function with a delay line, we propose simply making the number of zeros in the impulse response variable, and use it to create a periodic impulse response. Another possibility is to convolve this measured impulse response with one that has been synthesized, creating a hybrid instrument.

3 CONVOLUTIONAL SYNTHESIS

Once a parametric impulse response is in place, the system needs an excitation mechanism, that is, a signal to be convolved with the impulse response. Here we consider three possibilities: a parametric pulse train, a physics-based model of a reed, and a signal corresponding to a reed pulse, estimated be inverse filtering during a live performance.

3.1 A parametric pulse train

In this case, results are explored using a sequence of bell-curve shaped functions (such as gaussian), the width, shape (symmetry) and periodicity of which are controllable parameters. Since it may be desirable to have the source strongly coupled with the acoustic filter, the periodicity of the pulse train may be set automatically to a value corresponding to the fundamental frequency, or a harmonic, of the the parametric impulse response.

A graphical user interface, as shown in Figure 5, was developed to allow for modification of both source (pulse train) and impulse response parameters, and to perform the convolutional synthesis.

3.2 A physics-based reed model

In this case, we explore the use of a physics-based model of a generalized reed as a source. The reed model depends on a bore model for obtaining the downstream (bore) pressure, which is used for establishing the pressure difference across the reed, a requirement for oscillation.

The problem that may be encountered here is that certain reed configurations will not oscillate under all bore conditions (this is an expected limitation of a physics-based model). Rather than limit the producible sounds by limiting the possible impulse responses, we propose the use of two impulse responses (as seen in Figure 6), whereby one impulse response, $h(t)$, is convolved with the bore input pressure (the product of the airflow from the reed U and the bore impedance Z_0) to produce the bore pressure required to achieve reed oscillation, and the other impulse response, $g(t)$, is convolved with the input pressure to obtain the produced sound. It should be noted that if keeping a physical structure is desirable, $g(t)$ may be inferred from $h(t)$, with the assumption that a transmission is always amplitude complementary to its corresponding reflection (that is, it has transfer function $G(z) = 1 + H(z)$ for pressure waves [4]).

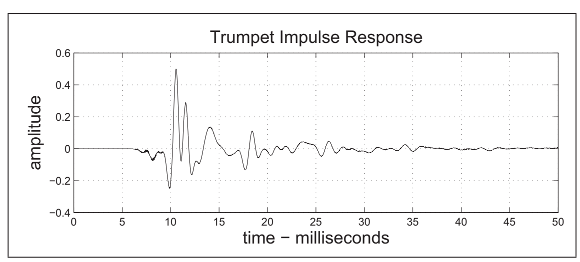

Figure 4. An impulse response from a trumpet.

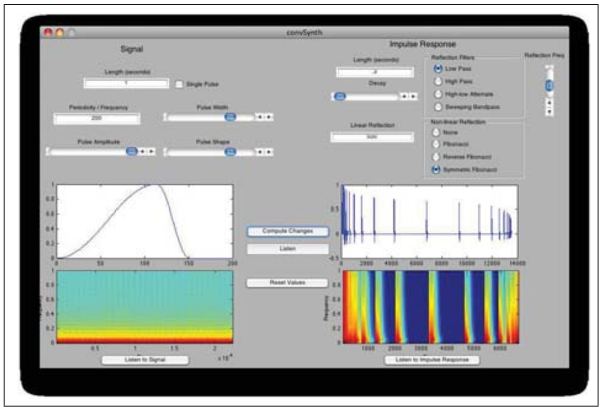

Figure 5. A graphical user interface allows for parametric synthesis of both input source and impulse response. For the impulse response, the user may specify the spacing of the arrivals, the filters of which they are composed, and the individual arrival and overall amplitude envelopes. For the source, the user may specify periodicity, shape and width of pulses.

It may also be possible however, to use completely unrelated impulse responses.

3.3 An estimated reed pulse

In this case, the bore impulse response is not only used for synthesis, but for analysis and estimation of the reed pulse, during an actual performance. If the reed pulse may be estimated using an inverse filter corresponding to the correct impulse response (which will change during performance), the reed pulse may be directly convolved with an impulse response similar or entirely different from the one corresponding to the instrument being played. Alternatively, features corresponding to different playing techniques may be extracted from the estimated reed pulse, and then remapped to control parameters of the pulse train or reed model described above. This application is the ultimate aim of the convolutional synthesis technique and relies on successfully estimating the reed pulse, a research goal that has obtained several results [6].

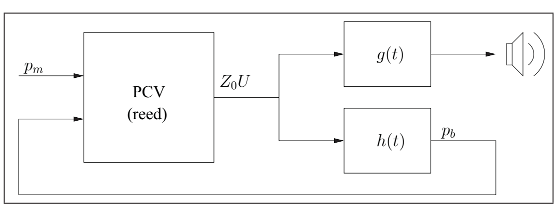

Figure 6. A signal flow diagram of convolutional waveguide synthesis in the context of a reed instrument. The bore pressure p_b is obtained by convolving the bore input pressure Z_0U with the impulse response $h(t)$, and the model output is obtained by a convolution with the impulse response $g(t)$.

4 LOW-LATENCY CONVOLUTION

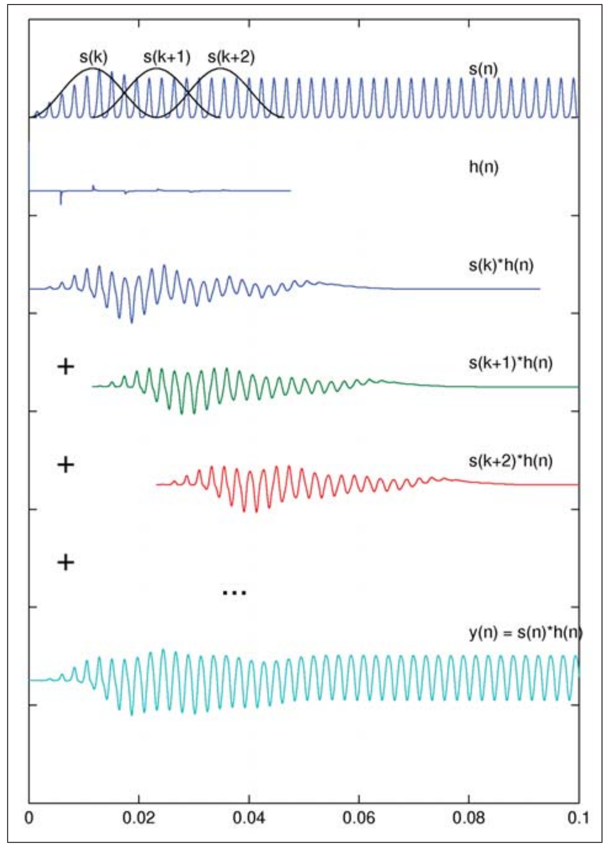

Figure 7. Fast convolution with latency

Though the waveguide structure, as seen in Figure 1, could have been used to filter all three source/excitation examples described in Section 3, we propose the use of a low-latency convolution, as described in [1], with a complete impulse response (i.e. not just a reflection function corresponding to a single round-trip down the bore to the bell and back).

If latency is not a concern, a simple implementation of fast convolution as illustrated in Figure 7 may be used, whereby a blocksize of samples $s(k)$ is convolved with the entire impulse response $h(n)$ and stored. One hopsize later, the next blocksize of samples $s(k+1)$ is also convolved with $h(n)$ and added to the previously stored result. This process continues until the end of the input signal $s(n)$ is reached. The problem with this implementation is that there is a latency corresponding to the length of the impulse response. That is, if parametric changes are made to the impulse response, the effects of one or more previous impulse response state(s) would continue to be heard for a time, in the worst case, corresponding to the sum of the lengths of the impulse response and blocksize (see the lengths of the component signals $s(k)*h(n)$, $s(k+1)*h(n)$ etc. in Figure 7 as compared to counterpart components in the low-latency convolution illustrate in Figure 8).

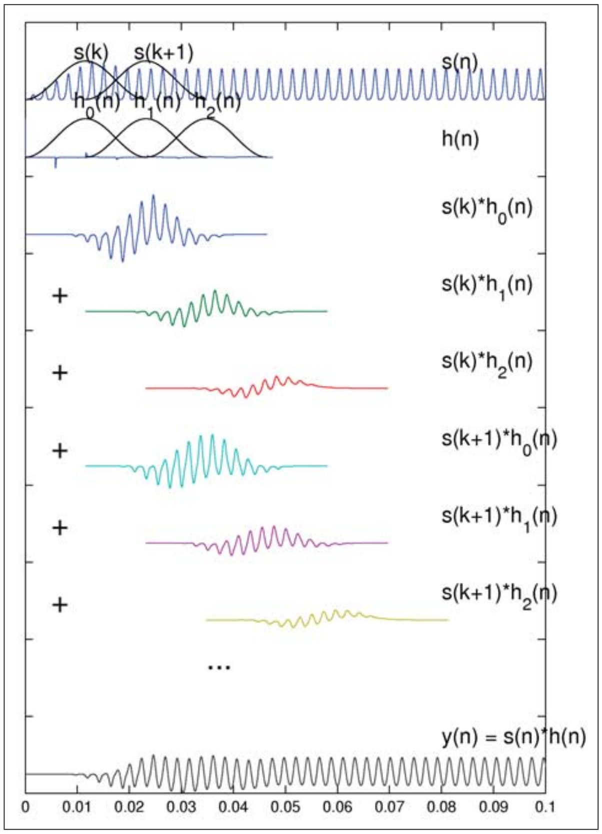

Figure 8. Fast convolution with low-latency.

In a low-latency convolution operation as described by [1] however, *both* signal $s(n)$ and impulse response $h(n)$ are windowed into a sequence of blocks (as shown in Figure 8). The convolution operation of a signal frame with an impulse response frame need only occur at the time it is needed for continued sound output. Because the result of a previous lengthy convolution is not stored, the user may modify the impulse response every blocksize of samples, where the blocksize is significantly shorter than the entire length of the impulse response. If the impulse response is parametric, the user's modifications may be heard in real-time, yielding the same level of interactive control as the waveguide model.

5 CONCLUSION

At the expense of computational complexity, the low-latency convolution affords the user some additional advantages to the waveguide model. For example, it is possible to build impulse responses less typical of a physical system, which would be difficult to obtain within a waveguide structure. There is no longer the physical constraint of having impulse response arrival uniformly spaced. Rather, they may be spaced according to some other structure, or interleaved in patterns that create multiple tones, creating multi-phonic or hybrid instruments. The impulse responses are also not constrained by stability in that arrivals must not decay.

This work is ultimately intended to be used in real-time performance, affording modern wind instrument players the ability to extend their instruments in addition to extending their playing technique. The proposed convolutional techniques permits sound exploration that is no longer limited by the physical structure of the instrument being played.

6 REFERENCES

[1] W. G. Gardner, "Efficient convolution without input-output delay," *Journal of the Audio Engineering Society*, vol. 43, no. 3, pp. 127–136, March 1995.

[2] X. Rodet and C. Vergez, "Phyical models of trumpet-like instruments: Detailed behavior and model improvements," in *Proceedings of ICMC 1996*. Clear Water Bay, Hong-Kong: International Computer Music Conference, August 1996.

[3] J. O. Smith, "A basic introduction to digital waveguide synthesis," http://ccrma.stanford.edu/˜jos/swgt/, February 2006.

[4] ——, "Physical audio signal processing for virtual musical instruments and audio effects," http://ccrma.stanford.edu/˜jos/pasp/, December 2008.

[5] T. Smyth and J. Abel, "Convolutional synthesis of wind instruments," in *Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA'07)*, New Paltz, New York, October 2007.

[6] ——, "Estimating the reed pulse from clarinet recordings," in *Proceedings of ICMC 2009*, Montreal, Canada, August 2009, forthcoming.

[7] T. Smyth, J. Abel, and J. O. Smith, "A generalized parametric reed model for virtual musical instruments," in *Proceedings of ICMC 2005*. Barcelona, Spain: International Computer Music Conference, September 2005, pp. 347–350.