

A Physical Modeling Based Music Representation Employing Measured Acoustic Properties of Musical Instruments and Inferred Performance Parameters

Xiaoxiao Dong, Mark Sterling, Mark F. Bocko
University of Rochester

(Dated: March 26, 2008)

We present an empirical approach for the physical modeling synthesis of monophonic musical instrument sounds and demonstrate the new method for a B^b clarinet. The physical model incorporates measured acoustic input impedances of the instrument air column for all notes in the instrument compass. The control parameters, which include the fingerings, the blowing pressure and the instrumentalist's embouchure clamping force on the clarinet mouthpiece, are inferred from audio recordings of real clarinet music with a computer algorithm. The control parameter time-histories form a highly compact yet expressive representation of the musical sound. We find that updating the control parameters 10 to 20 times per second enables highly accurate reproduction of the original musical sound at an aggregate data rate of only a few hundred bytes per second. To quantitatively assess the performance of the physical model a distance metric that emphasizes the timbral aspects of musical sound was defined. The metric provides a quantitative method to gauge the accuracy of the parameter extraction routines, to test the effect of minor alterations to the model, and to compare the empirical physical modeling approach with other forms of musical sound synthesis. Furthermore, the empirical physical modeling framework is sufficiently general that it may be extended to include other wind, stringed and perhaps even percussion musical instruments.

PACS numbers:

I. INTRODUCTION

In the work of Schumacher¹ *et al* the basic methods of physical modeling musical sound synthesis for a number of instrument families, including woodwinds, were laid out. This work led to the MSW (McIntyre, Schumacher and Woodhouse) model²⁻⁴, which is a minimal block diagram description of sound generation incorporating nonlinearities, feedback, and impulse responses. The advent of inexpensive computing power has enabled this line of research⁵, with most of the attention being devoted to digital waveguide models⁶. A more recent precedent for this paper may be found in Gazengel⁷. That work, however, predates the true ubiquity of powerful personal computers that exists today. For example, we found that the reflection function method—which was extensively employed in early work on Clarinet physical modeling, primarily as a means to introduce a computational savings—was not crucial to our physical modeling implementation. The instrument air column as a linear system will be discussed directly in terms of time-domain impulse responses. Similarly, we were able to choose a relatively high sample rate of 44.1 kHz for simulations. Furthermore, we address many issues related to implementing a physical modeling *synthesizer*, such as the practical issue of handling note transitions during a simulation.

A large body of literature exists on the measurement of the acoustic properties of wind instruments⁸, with the B^b clarinet being a popular system to study. Various methods have been developed which make it possible to obtain high quality measurements of the acoustic impedance of an instrument's air column with relatively

simple instrumentation⁹. By combining these two lines of research i.e., physical modeling algorithms for sound synthesis and accurate measurements of the acoustic properties of musical instruments, we have developed an empirically based form of instrument physical modeling and sound synthesis.

The musical expression of the performer may be captured and incorporated into an empirical physical model in the form of the time-histories of the instrument model control parameters. These may be measured directly, although it may be difficult to do so in a manner that does not hinder the musician, or they may be inferred from audio recordings of a musical performance by various parameter estimation techniques.

This paper outlines the construction of an empirical physical model sound analysis and re-synthesis system. We illustrate the method in the case of monophonic clarinet music but it should be possible to generalize the method to other members in the wind instrument family as well as stringed and perhaps even percussion instruments. There are two aspects of this work: the design and implementation of the physical model itself and obtaining a suitable set of acoustic measurements on a model instrument. The model described here provides a method for encoding a specific musical performance in the language of the control parameters, which enables a natural representation of the more subtle features in a musical performance such as vibrato and variations of timbre. The virtues of this representation are extreme compactness and musical realism¹⁰. The availability of freely alterable physical control parameters to the end user also gives the computer music performer new degrees of freedom in musical expression, however these are

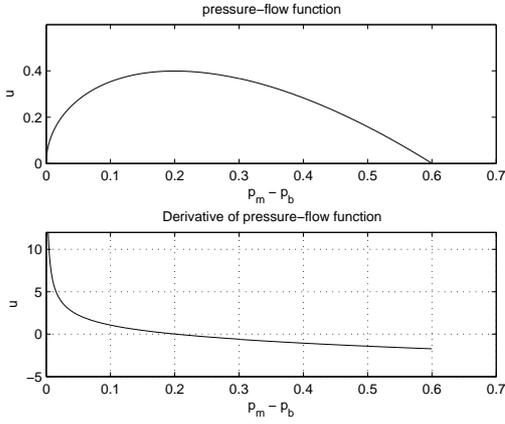


FIG. 1. A typical Pressure Flow curve and its derivative for single reed valve. The lower curve is the slope of the flow characteristic and is helpful in understanding the behavior of the system when it is linearized about an operating point.

expressed in terms of the musical gestures that a musician would employ in performance so it is interesting to contemplate if it may be possible to define a “language” of elemental musical gestures from which an expressive musical performance may be synthesized.

II. EMPIRICAL PHYSICAL MODEL

A wind musical instrument may be described by the following equations.

$$p_b = \int_0^\infty h(t - \tau)u(\tau)d\tau \quad (1)$$

where p_b is the pressure at the input of the instrument bore, $u(t)$ is the time dependent volume flow velocity and $h(t)$ is the impulse response function corresponding to the input acoustic impedance of the instrument bore. The pressure difference across the reed valve is then

$$\Delta p = p_m - p_b \quad (2)$$

where p_m is the blowing pressure supplied by the performer. The resulting volume flow velocity through the reed valve¹¹ is then given by

$$u = g(\Delta p) = U_M \frac{3\sqrt{3}}{2} \left(1 - \frac{\Delta p}{p_{\text{ext}}}\right) \left(\frac{\Delta p}{p_{\text{ext}}}\right)^{1/2} \quad (3)$$

In this model a nonlinear system (the reed valve) is cascaded with a linear system (the instrument bore) and the output of the linear system is provided as feedback to the nonlinear element along with an external input representing the blowing pressure of the player. In the case of a clarinet, the reed-mouthpiece assembly is the nonlinear element while the instrument air column corresponds to

the linear portion of the model. A typical example of the nonlinear function g is shown in Fig. 1.

The bore of the clarinet may be represented by a lumped impedance. This is possible because the clarinet bore is assumed to behave linearly in all typical operating regimes and also because plane wave propagation is assumed, that is, wave propagation is considered to take place in the lowest mode, which enables a one-dimensional representation of the clarinet. The response of the bore at a point just inside the mouthpiece of the instrument may be characterized as a filter, or in the time domain as the impulse response function. Since the impulse response of the bore is different for each possible fingering the filter must be treated as time-dependent over a relatively long time scale, i.e. from one note to the next.

The continuous time equations may be converted to their discrete time version as follows. First we let L denote the length (in samples) of the bore impulse response. For the simulations presented below, $L = 7000$ with a sampling rate $f_s = 44100$ Hz. The filter coefficients for different notes are stored in column vectors h_i with the subscript $i \in (0, 46)$ denoting the bore responses for each of the notes ($i = 0$ is a zero valued filter which can be used during silent passages or rests). Although many alternate fingerings are available we limited this set to 46 “standard” fingerings found in any elementary method book for the clarinet. A state vector U_n , containing the past values of u is created.

$$U_n = \begin{bmatrix} u[n] \\ u[n-1] \\ \vdots \\ u[n-L+1] \end{bmatrix} \quad (4)$$

The bore pressure, at any sample, is then equal to an inner product.

$$p_b[n] = h_i^T U_{n-1} \quad (5)$$

$$u[n] = g(p_m - h_i^T U_{n-1}) \quad (6)$$

The nonlinear nature of the reed makes its treatment more complex and several approaches are available. In the simplest form the reed may be considered as static or memory-less. That is, the reed acts as a pressure controlled flow valve that relates the instantaneous value of the volume flow velocity to the pressure difference across the reed valve. In a more detailed approach, the dynamical properties of the reed may be included by introducing the second order equations of motion for the displacement of the reed tip as a function of the forces acting on the reed. In our simulations, the reed dynamics are included according to the prescription¹² where the second order system of the reed is approximated by an IIR filter.

$$x[n] = b_1 p_b[n-1] + a_1 x[n-1] + a_2 x[n-2] \quad (7)$$

The filter coefficients depend upon the resonant frequency and mechanical quality factor of the reed. The volume flow can be written as the following function of the reed displacement¹².

$$u = \Theta(1 - p_m + x) \times (1 - p_m + x) \times \text{sgn}(\Delta p) \sqrt{|\Delta p|} \quad (8)$$

A. Acoustic Measurements

In the empirical physical modeling approach acoustical measurements of actual instruments become a part of the synthesis engine. This will enable the development of synthesis engines that are based upon measurements of the most outstanding examples of specific musical instruments. Furthermore, researchers will be able to explore the differences in the generated sounds and the playing characteristics of various examples of the same type of instrument. It also may be possible to employ empirical physical modeling to compute the acoustical properties of “theoretical instruments” from dimensional specifications and then to synthesize the sound of the instrument. This would enable a purely computational approach to the design of musical instruments and allow efficient trials of a great many design variations of standard instruments or even for entirely new types of instruments to be explored.

A number of review papers^{9,13} have been written on the subject of the acoustic input impedance of wind musical instruments. Methods for measurement of acoustic impedance include the capillary method^{8,14}, the reflection function method, and the piezoelectric disk method⁹. They differ principally in the number of microphones that are present in the system and the type of transducer. Electret condenser microphones are readily available and provide an accurate method of measuring pressure. The problem often is one of creating a high impedance acoustic source so that the impedance of the sample may be determined simply from a measurement of pressure at the input of the instrument.

The piezoelectric disk method⁹ was chosen as the measurement technique to furnish data for our empirical physical model. The impedance head consists of a piezoelectric disk rigidly fastened to one end of a cylindrical cavity. The cavity also contains a hole where a small microphone can be mounted. The acoustic impedance can then be determined from a measurement of pressure. The piezo-disk is modeled as a piston at one end of the air column, with the piston behaving as a second order system. The analysis⁹ assumes that the disk is excited in its lowest flexural mode, where the displacement is maximum at the center of the disk and is axially symmetric. The natural resonant frequency of this disk mode determines the maximum frequency at which acoustical measurements can be made without errors caused by the transducer or overly complex calibration procedures. We employed Mouser Part Number AB2040B and AB2065B piezoelectric benders which cost less than \$ 1 US each. The lowest resonant frequencies of these elements are 4 kHz and 6.5 kHz respectively and in both cases we found

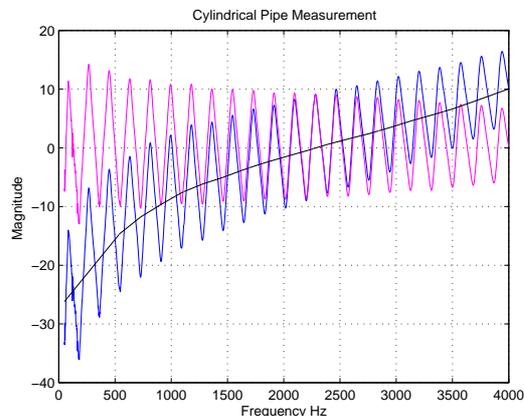


FIG. 2. Impedance reference reading taken at end of a cylindrical PVC tube (blue). The same data is shown with its midline (black) subtracted (magenta). In log magnitude, the midline occurs at the characteristic impedance of the pipe, which is known to be a constant.

that the calibration procedure resulted in quality acoustic impedance curves.

The audio measurements were taken on a plastic Selmer B^b Clarinet employing the following equipment: a Countryman Isomax B6 lavalier microphone with its output going into a Mackie mixer Model 1402, various lengths of 0.5 inch diameter PVC tubing, a Samson audio amplifier to drive the piezoelectric bender and a Stanford Research Systems SR 780 spectrum analyzer. A swept sine signal was generated by the analyzer and amplified to drive the piezoelectric bender element. The pressure at the input of the clarinet was measured in response to the excitation. For each of 46 standard fingerings of the clarinet we measured the acoustic impedance from 50 Hz to 4 kHz with a resolution of 2047 points covering this frequency span, which gives the following frequency resolution.

$$\Delta f = \frac{4000 - 50}{2047} \text{Hz} = 1.9 \text{Hz} \quad (9)$$

Starting with the measured acoustic impedance data it is possible to zero-pad and interpolate to create impedance curves with a higher effective sampling rate to be used in the empirical synthesis model, but of course these are limited by the frequency range and resolution of the original measurements. Gazengel⁷ discusses the various consequences and pitfalls of interpolating such measured impedances, especially as they apply to time domain representations.

A calibration is performed by connecting the measurement system to a known acoustic impedance. This was provided by a 3-ft length of cylindrical PVC pipe with an inner diameter of 0.5 in. Figure 2 shows the calibration data. The midline of the curve (plotted as log magnitude) represents the characteristic impedance Z_0 . We plot an example of the magnitude adjustment in Fig. 4. A similar procedure can be employed for the phase of the impedance data as well. However unlike the case

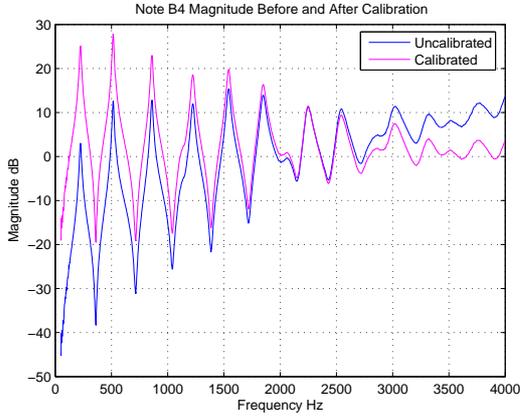


FIG. 3. Calibrated and uncalibrated impedance magnitude for note B4.

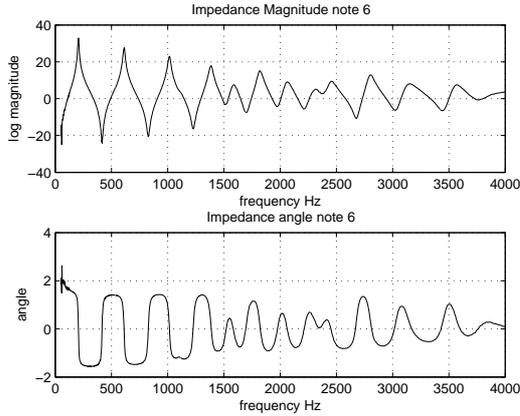


FIG. 4. Example of calibrated impedance magnitude and phase. The peaks of the upper curve occur at resonances of the air column. At very low frequencies, the measurements are more corrupted by noise (due to the lower response of the piezo disk) and this is reflected in the measurements, especially in the low-frequency phase response.

for the magnitude, the phase calibration was computed separately for each different note of the clarinet. In the course of the experiments a linear trend in the unwrapped phases of all the impedance measurements was noted. The best fit to a straight line was found by a least squares fit and subtracted from the measured phase for each note. The following equations describe this procedure. Assume the unwrapped phase is stored in a 2047×1 column vector called a . We define a 2047×2 matrix B .

$$B = \begin{bmatrix} 1 & 1 \\ 2 & 1 \\ \vdots & \vdots \\ 2046 & 1 \\ 2047 & 1 \end{bmatrix} \quad (10)$$

The parameters of the linear model are computed with a standard least squares calculation¹⁵.

$$m = (B^T B)^{-1} B^T a \quad (11)$$

To apply the calibration the line resulting from the least squares fit is subtracted.

$$a_{\text{cal}} = a - Bm \quad (12)$$

To find the impulse response of the air column for any note the inverse Fourier transform of the acoustic impedance is computed. An example is shown in Fig. 6. Below, we describe the algorithm for computing the discrete impulse responses.

Our impedance measurements are interpolated and zero-padded after the calibration procedure so that simulations may be carried out at various sampling rates. In fact, some interpolation and zero-padding must be performed even if we set the sampling rate at 8000 Hz (Nyquist frequency of 4000 Hz). This is because the measurements start at 50 Hz and without filling in the data below 50 Hz there would be a mismatch between the radian frequencies of the DFT and those implied by the actual measurement frequencies.

After calibration the impedance curves are stored in a set of row vectors. Let Z denote any one of these vectors. The procedure for finding the discrete impulse response is the same for each fingering of the instrument. In addition to the given measurement, the spectrum analyzer returns a vector containing the analysis frequencies. Call this vector M . With the data stored in the vectors M and Z we perform a linear interpolation of the impedance curves and for values outside of M we set the interpolated value of Z to zero. The frequencies at which the vector Z is interpolated are stored in a row vector Q and the interpolated curve itself is stored in a vector Z_q . In the programs that calculated the clarinet bore impulse responses Q was a 1×4001 vector consisting of evenly spaced values between 0 and 4000 Hz. The following vector H can be formed from the entries of Z_q .

$$H = \begin{bmatrix} Z_q(1) \\ Z_q(2) \\ \vdots \\ Z_q(4001) \\ Z_q^*(4001) \\ \vdots \\ Z_q^*(2) \end{bmatrix} \quad (13)$$

This vector possesses conjugate mirror symmetry corresponding to a purely real signal in the time domain. The inverse discrete Fourier transform of this vector is the impulse response of the air column $h(n)$. The impulse response $h(n)$ can be re-sampled to any effective sampling rate. We found it necessary to run the clarinet simulations at sampling rates higher than 8000 Hz, which is not surprising. Even though the fundamental frequency of all of the clarinet notes (E3 through C[#]7, written) is below the Nyquist rate at an 8000 Hz sampling rate for the highest note (C[#]7 written which sounds

Computation of Clarinet Impulse Responses from Piezo-Disk Impedance Head Measurements

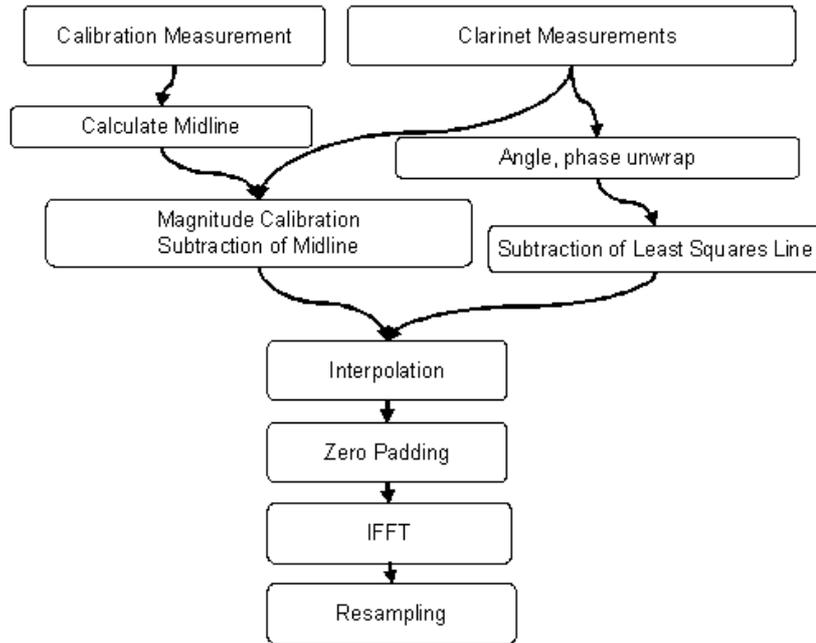


FIG. 5. Flowchart describing the various stages of the algorithm which returns the impulse responses for the empirical physical model.

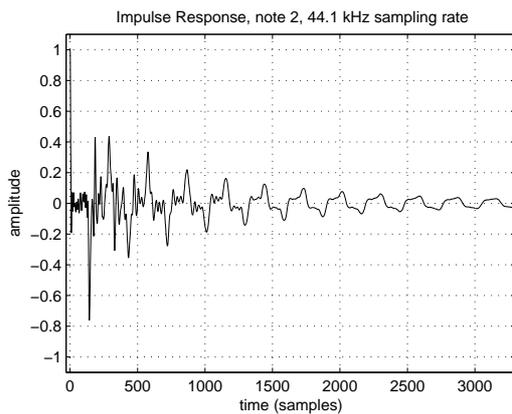


FIG. 6. Example of impulse response response computed for note 25. The impulse response is normalized so that the value at sample 0 equals 1.

at a fundamental of 1975.5 Hz) in the highest register only one upper harmonic lies below 4000 Hz. This is problematic since the tone quality of musical sound is contained in the overtone structure, however such high notes are not often employed. To extend the impedance measurements above 4000 Hz is possible but would require a redesign of the measurement apparatus. It also is possible that due to the nonlinearity of the clarinet model the simulation may require a greater bandwidth than the Nyquist frequency.

Following calibration of the impedance data it is neces-

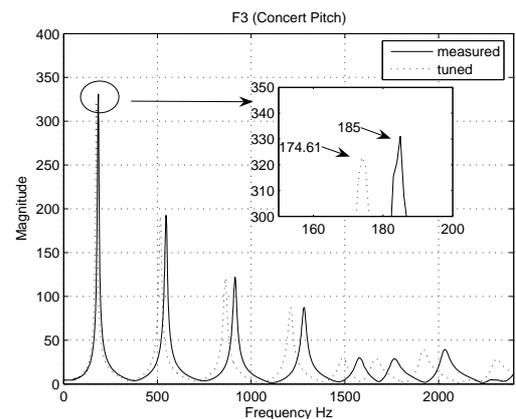


FIG. 7. Tuning of impedance curve to concert pitch. The tuning is accomplished by a linear scaling the frequency axis so that the resonance peaks are shifted more at higher frequencies than at lower frequencies. Frequency scaling, as opposed to a frequency shift, preserves harmonic ratios (even though the resonances of wind instruments are only approximately harmonic).

sary to scale the frequency axes of the impedance curves to produce diatonically correct synthesized clarinet tones. As discussed by Benade and Ibbi⁹ the piezo-disk method of measuring acoustic impedance can result in shifts of the impedance maxima from their values in normal playing conditions. The reason for this is that the volume of the impedance head is not precisely the same as that

of the clarinet mouthpiece. Furthermore the volume of air in the player’s oral cavity may produce an additional shift. To account for these measurement errors a scaling transformation is applied to the frequency axis for each measured impedance curve. The scaling factor is chosen so that the frequency of the fundamental resonant peak in the impedance spectrum is equal to the equally tempered scale frequency for the note, for A440 tuning.

B. Vocal Tract Effect

The effect of the vocal tract of the player is an important aspect in the functioning of the instrument^{16–18}. The volume of the oral cavity and vocal tract, over which the player has great influence, add to the reactance of the bore. This represents an additional aspect of the connection between instrument and player in addition to the blowing pressure and embouchure force and determines which note the instrument produces under many circumstances, especially for notes for which the register hole is opened and the conditions for oscillation¹⁰ may be met at more than one frequency. These conditions are

$$\omega < \omega_r \quad (14)$$

$$X_v + X_p > 0 \quad (15)$$

where ω is the operating frequency, ω_r is the resonance frequency of the reed and X_v, X_p represent the reactance of the instrument and oral cavity at the operating frequency ω . The second condition simply states that the reactance of the coupled system of oral cavity and instrument should be positive (in general, the reactance of the oral cavity is less than zero). These conditions arise from a detailed analysis of the mouthpiece valve¹⁹. The model is capable of producing oscillations at frequencies at which these conditions are met and subtle variations of the parameters under the control of the musician favors one of the modes thereby forcing the instrument break into oscillation in that mode. It is reported that the reactance of the mouth cavity is generally less than zero, contributing to the overall damping in the system¹⁰. From the point of view of the player, then, the goal should be to arrange the damping to achieve the desired pitch.

Measurements of vocal tract impedance may be found in the literature^{16,18}. We do not attempt to include a full parametric description of the vocal tract in the physical model, rather we employ the simplification of representing the vocal tract as a second order system with resonant frequency equal to the frequency of the desired pitch. Adding the reactance of the player’s vocal tract to the reactance of the clarinet bore ensures that the model will sound at the correct pitch by making the total reactance $X_v + X_p$ a global maximum at the frequency of the desired note. The effect of the vocal tract is assumed fixed for each note and thus is not included as one of the time-dependent control parameters. However in a more refined model time dependent vocal tract effects could be

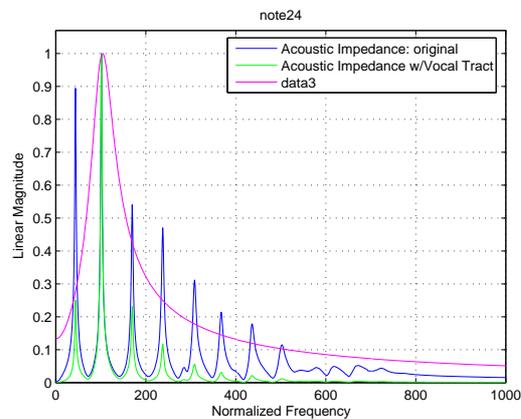


FIG. 8. Example (shown here in linear magnitude) of altering the impedance curve to account for the effect of vocal tract.

included. For example the well-known opening clarinet glissando in George Gershwin’s ”Rhapsody in Blue” is affected with the aid of such vocal tract manipulations.

The complete sound synthesis algorithm proceeds as follows. First, the calibrated and tuned impedance curves are created. For each such curve we identify the corresponding pitch and construct a second order response with a resonant frequency at the frequency of the desired pitch and an independent damping. We use this device to increase the impedance maxima at the desired pitch and to reduce the others. An example of this is shown in Fig. 8.

C. Radiation

The final step in the synthesis algorithm is to model the sound radiation from the instrument. In the clarinet, sound is radiated both from the bell and from the tone hole lattice. For notes in the upper register, most of the sound is radiated from the tone holes and the bell can be removed without significantly altering the perceived tone quality. The clarinet displays a complex directional sound radiation pattern²⁰. One approach is to measure the transfer function between the inside of the mouthpiece and an external microphone for each of the 46 fingerings of the clarinet similar to the acoustic impedance measurements. A simpler option is to treat the system as an unflanged cylindrical pipe and compute the radiation transfer function. Although the frequency-dependent sound radiation pattern is highly complex^{21,22} the transfer characteristic in the forward direction is a high pass filter with a 6 dB per octave roll-off below the pass-band. The cutoff frequency is equal to the frequency at which the product of the wavenumber k and pipe radius a is equal to 2.

$$ka = 2 \quad (16)$$

We employ this simplified approach for the time-being.

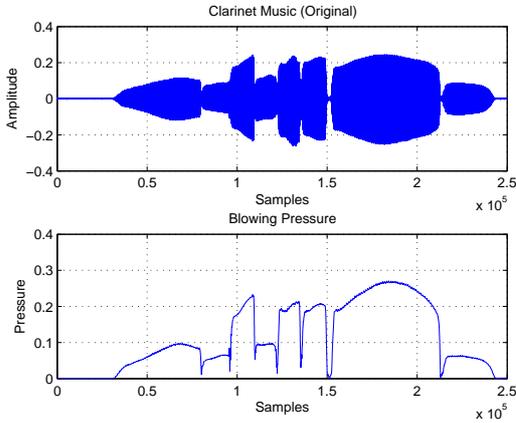


FIG. 9. Envelope calculation from an audio recording of solo clarinet music.

III. CONTROL PARAMETERS

The control parameters of the physical model map onto the gestural elements of a musical performance such as the sequence in which the keys are pressed in wind instruments, the length of the stopped string in stringed instruments, the embouchure clamping force on the mouthpiece of a woodwind, or the force on the strings and velocity of the bow in a stringed instrument. In the clarinet the control parameters may be divided into two groups, those associated with the linear portion of the model and those associated with the nonlinear elements in the model. The time record of the clarinet fingerings is associated with the linear part of the model and determines which instrument impulse response function is selected at any given time. The blowing pressure and embouchure force both directly affect the nonlinear portion of the clarinet model.

To create musical sounds with the model described above it is necessary to specify the time histories of the various control parameters. Identifying combinations of the parameters that produce musically pleasing results is a challenge in physical modeling synthesis so in the following section we present a method for extracting appropriate control parameter histories from a recording of an actual instrument. These parameter histories, which when provided to the model enable a re-synthesis of the original signal, serve as a highly compact representation of the sound. In the following section we describe our method of parameter estimation for a recorded sound.

A. Envelope Detection, Blowing Pressure

Much of the expressiveness of a musical recording is present in the envelope of the sound wave, for example note “shapes”, i.e., the attacks and releases, dynamics and amplitude modulation (vibrato) all are visible in the waveform envelope. Therefore the control parameter extraction algorithm begins with a calculation of the envelope. This is accomplished simply by low-pass filtering the absolute value of the waveform.

$$\text{env } p = h_{lp} * |p| \quad (17)$$

We have found that setting the blowing pressure in the model equal to an appropriately scaled value the envelope leads to satisfactory results in most cases. A detailed analysis of the clarinet, and in particular of the conditions under which steady oscillation is maintained, lends validity to the range of blowing pressures that are arrived at by this simple method. As discussed above, Fletcher provided a general theory of oscillations¹⁹ in brass and woodwind instruments. When considering a reed-mouthpiece coupled to a specific acoustic system, such as a pipe, there are additional conditions for oscillation¹⁰. With a linearized model of the reed the oscillation thresholds are determined by a pair of requirements: resonance (i.e. in-phase pressure and velocity)²³ and negative damping of the reed. These requirements lead to

$$\text{Im } Y_r = -\text{Im } Y_p \quad (18)$$

$$-\text{Re } Y_r > \text{Re } Y_p \quad (19)$$

where Y_r is the admittance of the reed and Y_p is the admittance of the air column. If the reed nonlinearity is considered to be a lumped admittance connected in parallel with the admittance of the bore and also possibly with the admittance of the player’s vocal tract, then, for the quasistatic case, the admittance of the reed is purely real and is equal to the slope of the flow curve at the operating point.

Further, when the coupling of the generator to a resonator is considered, two additional conditions involving the reed and air column admittances arise. The admittances of the reed and of the air column are effectively connected in parallel with each other. Therefore, the total admittance is their sum. For a parallel resonator resonance occurs when the effort and velocity variables (voltage and current for example) are in phase. This means that the imaginary part of the total admittance should be equal to zero. Since resonance is assumed, a condition is sought such that the corresponding pole is in the right half plane. This occurs when the real part of the total admittance is negative. Since the real parts of the admittances sum, this means that the reed conductance should be negative and larger in magnitude than the real part of Y_p .

The conditions given above define the range of values of the parameters for which the clarinet will oscillate. Thus, we can also employ them to furnish a set of operating regions for the control parameters of the physical model.

Clearly, the blowing pressure p_m should be large enough to set the operating point of the reed on the negative sloping region of the pressure-flow characteristic. This occurs for values of the pressure difference larger than $P_{\text{ext}}/3$. The conductance of the reed is found by evaluating the derivative pressure flow characteristic at the operating point.

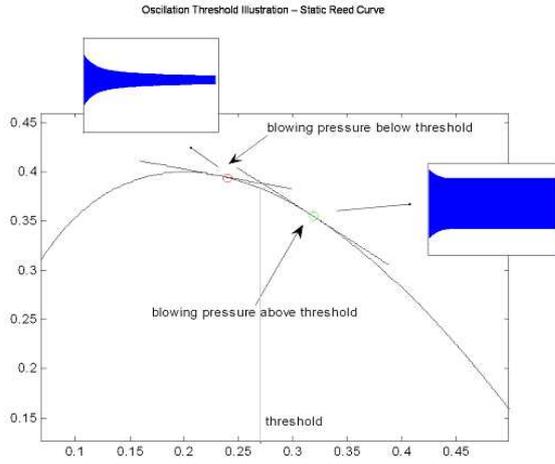


FIG. 10. By running the model with a constant blowing pressure and a fixed impulse response we can demonstrate the threshold for which the model will maintain a steady tone. The threshold computed from theory is represented by a vertical line.

In general, Z_p consists of a finite number of resonance peaks in approximately harmonic ratios. The inequality will be most strongly satisfied at these admittance minima or impedance maxima. On the other hand, the equality of the susceptances shifts the oscillation frequency away from the exact resonance whenever the reed admittance is not simply a conductance.

B. Linear Parameters

The linear control parameters, that is, the control parameters pertaining to the instrument air column, are a crucial ingredient in re-synthesis. Using a note recognition algorithm^{24–26} we process the recorded audio to determine the onset and cut-off times of the different note events and their pitches. To account for small variations in tuning the pitches are "rounded" to the closest concert pitch in standard A440 tuning and integer values between 1 and 46 are assigned. Thus, for example, 1 corresponds to a written E3, sounding as D3 on a piano keyboard. A vector containing the times of the onsets and releases (in samples at 44.1 kHz) and their corresponding note numbers fully specifies the control of the linear part of the clarinet model. The total length of time L that our simulation will run is known in advance. The onsets and pitches eventually are condensed into a single $1 \times L$ vector, which we call id , with values in $(0, 46)$ -where a value of zero is used during musical rests or silences to make the bore impulse response zero. At any sample, the bore impulse response enters the model as part of an inner product or linear functional. The vector id simply selects which impulse response vector to place in the inner product, the other entry being the state vector U .

$$p_b = (h_{id(n)}, U) \quad (20)$$

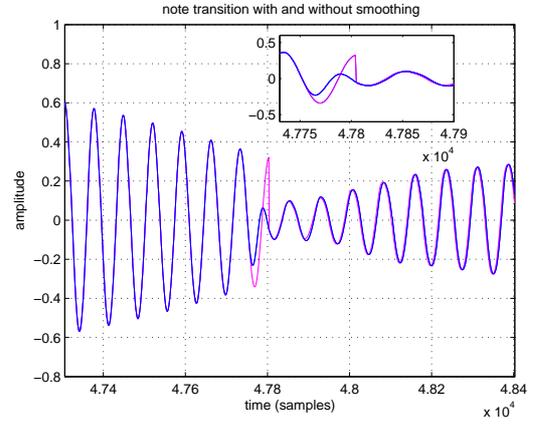


FIG. 11. Effect of smoothing on a note transition. An abrupt transition results in a discontinuity which becomes audible as a click.

The vector h is, in general, not fixed, but changes at each new note. In an actual clarinet, the impulse response is determined by how the player holds their fingers. Even the most dexterous clarinet player would not be able to change the fingering in the space of one sample at a sampling rate of 44.1 kHz. Not only are there transients in the system due to the finite amount of time taken by the system to build up an oscillation, but there also are transients caused by the fact that between two notes, as the player is shifting their fingers, the impulse response of the bore is not any one of the 46 impulse responses for fixed fingerings but some intermediate state between two of them.

The empirical measurement of the time-varying nature of the bore impulse response as discussed above would be a significantly more involved task than is possible with our current experimental setup. One can imagine a machine with false fingers that would fit over the clarinet and be able to hold fingerings with partially covered tone holes, and partially depressed keys. Impedance measurements could then be taken in much the same manner as they have been. Alternatively one could simply interpolate between the impulse responses of the beginning and ending notes, which we employed.

Although the method outlined above would be more desirable in the sense of being physically motivated, the actual implementation would add a great deal of complexity. Still, it was found that some approximation of this behavior was needed. Abrupt changes often resulted in click like distortions in the synthesized sounds. To remedy this, a region around each onset is defined where the impulse response in the simulation is a linear combination of the impulse responses for the next and previous notes. Call these impulse responses h_1 and h_2 , for example. Then the following equation determines the impulse response, where the constants c are chosen so that there is a continuous transition.

$$h = c_1 h_1 + c_2 h_2 \quad (21)$$

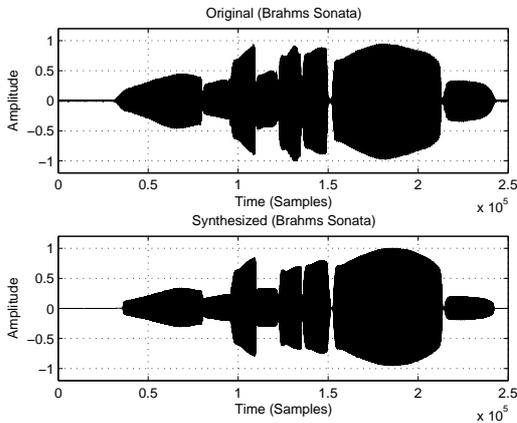


FIG. 12. Original and synthesized clarinet music shown in the time domain.

The coefficients in the equation are calculated according to cosine smoothing. Taking samples x from the beginning of the smoothing region, which is a window of J samples

$$c_1 = \frac{1}{2} \left(1 - \cos \left(\frac{\pi x}{J} \right) \right) \quad (22)$$

$$c_2 = \frac{1}{2} \left(1 - \cos \left(\frac{\pi(J-x)}{J} \right) \right) \quad (23)$$

IV. SYNTHESIS RESULTS EVALUATION

Synthesis examples from our empirical physical model are presented in this section. Control parameters were computed as outlined in the preceding section. The source recordings were of a professional clarinetist playing in anechoic conditions. A plot of the time domain waveforms of the original source recording and the synthesized output of the physical model are given in Fig. 12. Five notes are represented in these waveforms. Time-frequency distributions of the same five notes are given in Fig. 13.

Evaluation of the clarinet empirical model requires a metric that quantifies the error between the synthesized sound and the original sound. The exact pitches of the synthesized music rarely match those in the original recording. This mismatch can thwart attempts to employ simple metrics such as the mean squared error, that is, small errors in pitch may alter the inner product significantly but this difference may be imperceptible to a listener. A more useful metric is one that takes account of perceptual differences of musically important features. The determination of which features are musically important, however, can be largely subjective and we found no qualitative metric that could eclipse listening tests.

Extensive listening tests were conducted with trained musicians and although we did not endeavor to conduct a quantitative preference study at this point the results

were deemed highly satisfactory in general, especially in the sustained portions of notes. One feature that was noted by many trained listeners was the lack of crispness of the attacks, which in retrospect was expected since “tonguing” was not included explicitly in the model. Clearly the onset transient is much different for a note that is tongued, i.e. striking the reed with the tip of the tongue at the beginning of the note, than by simply blowing into the clarinet. To include tonguing will require a specification of the initial conditions (displacement and velocity) for the reed at the beginning of each note.

V. CONCLUSION

In summary, we have created a system for synthesizing clarinet music which combines a description of the physics of the instrument, experimental measurements of the instrument acoustic properties and a parameter estimation routine that enables determination of the parameters that when provided to the model accurately reproduces a recorded musical passage. This system is capable of recreating a performance of a piece of music while maintaining, to a high degree, both the timbre of the musical instrument and the emotional sweep of the player. This capability depends upon the accuracy of the empirical physical model and the time-histories of the control parameters.

This work may be extended to encompass a large variety of instruments besides the clarinet. The impedance head design, and hence the experimental portion of our work, may be employed in measuring the acoustics of other instruments in the woodwind family such as the saxophone, oboe and bassoon and also the brass instruments. The primary difference in developing a full empirical physical model for these instruments is in the characterization of the nonlinear active component in an MSW model. For the saxophone, the reed-mouthpiece assembly is virtually the same as that of the clarinet. The oboe and bassoon, however, employ double reeds and the pressure versus flow relationship is somewhat different than in the clarinet²⁷. The experimental portion of an empirical physical model need not be restricted to an acoustical impedance measurement, or of the properties of a linear resonant structure. As alluded to above, among the many features of an instrument which may be tested, we can include instrument radiation patterns as well.

Another possible extension of this work is to focus on deploying the model in applications. An end user likely places a premium on ease of use and speed. While the implementation presented above is an accurate realization of the MSW equations, we do not claim that it is the most computationally efficient form possible. The time-varying and nonlinear nature of the reed makes fast computation difficult. FFT methods for computing filter outputs do not apply. The ways in which researchers have sought to create computationally fast empirical models has been either in the digital waveguide modeling context or by introducing simplified models of the bore impedance. However as computing power increases there is less of a premium on computational efficiency.

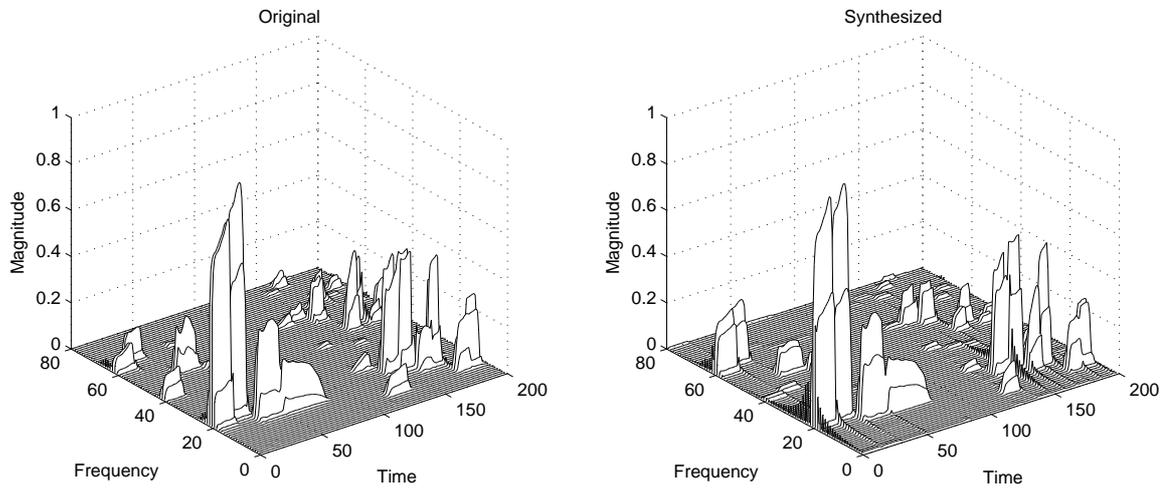


FIG. 13. These waterfall plots show the spectra of a sequence of the same two notes as played on a real instrument (left) and resynthesized by our empirical physical modeling algorithms (right).

Outside of applications requiring a real-time or near real-time computation we see potential in the community of people, amateur and professional electronic musicians, who work with the growing numbers of audio editing software packages. Thankfully, there are already a few standard systems for incorporating modularity into computer audio tools, such as the VST standard developed by Steinberg. VST plugins can be created with a freely available C++ class. Once compiled, they take the form of a DLL which certain audio applications can access.

- ¹ R. T. Schumacher, “Ab initio calculations of the oscillations of a clarinet”, *Acustica* **48**, 71–85 (1981).
- ² M. E. McIntyre, R. T. Schumacher, and J. Woodhouse, “On the oscillations of musical instruments”, *J. Acoust. Soc. Am.* **74**, 462–470 (1983).
- ³ S. Adachi, “Principles of sound production in wind instruments”, *Acoust. Sci. and Tech.* **25**, 400–405 (2004).
- ⁴ S. D. Sommerfeldt and W. J. Strong, “Simulation of a player-clarinet system”, *J. Acoust. Soc. Am.* **83**, 1908–1918 (1988).
- ⁵ V. Valimaki, R. Rabenstein, D. Rocchesso, X. Serra, and J. O. Smith, “Signal processing for sound synthesis: computer-generated sounds for all”, *IEEE Signal Processing Magazine* **24**, 8–10 (2007).
- ⁶ J. O. Smith, “Physical modeling using digital waveguides”, *Computer Music Journal* **16**, 74–87 (1992).
- ⁷ B. Gazengel, J. Gilbert, and N. Amir, “Time domain simulation of single reed wind instruments. from the measured input impedance to the synthesis signal. where are the traps?”, *Acta Acust.* **3**, 445–472 (1995).
- ⁸ J. Backus, “Input impedance curves for the reed woodwind instruments”, *J. Acoust. Soc. Am.* **56**, 1266–1279 (1974).
- ⁹ A. H. Benade and M. I. Ibsi, “Survey of impedance methods and a new piezo-disk-driven impedance head for air columns”, *J. Acoust. Soc. Am.* **81**, 1152–1167 (1987).
- ¹⁰ N. H. Fletcher and T. D. Rossing, *The Physics of Musical Instruments*, second edition (Springer) (2000).
- ¹¹ J. Dalmont, J. Gilbert, and S. Ollivier, “Nonlinear char-

- acteristics of single-reed instruments: Quasistatic volume flow and reed opening measurements”, *J. Acoust. Soc. Am.* **114**, 2253–2262 (2003).
- ¹² P. Guillemain, J. Kergomard, and T. Voinier, “Real-time synthesis of clarinet-like instruments using digital impedance models”, *J. Acoust. Soc. Am.* **118**, 483–494 (2005).
- ¹³ J. P. Dalmont, “Acoustic impedance measurement, part i: A review”, *Journal of Sound and Vibration* **243**, 427–439 (2001).
- ¹⁴ J. Kergomard and R. Causse, “Measurement of acoustic impedance using a capillary: An attempt to achieve optimization”, *J. Acoust. Soc. Am.* **79**, 1129–1140 (1986).
- ¹⁵ D. C. Lay, *Linear Algebra and Its Applications*, second edition (Addison-Wesley) (2000).
- ¹⁶ J. Backus, “The effect of the player’s vocal tract on woodwind instrument tone”, *J. Acoust. Soc. Am.* **78**, 17–20 (1985).
- ¹⁷ A. H. Benade, “Interactions between the player’s windway and the air column of a musical instrument”, *Cleveland Clinic Quarterly* **53**, 27–32 (1986).
- ¹⁸ C. Fritz and J. Wolfe, “How do clarinet players adjust the resonances of their vocal tracts for different playing effects?”, *J. Acoust. Soc. Am.* **118**, 3306–3315 (2005).
- ¹⁹ N. H. Fletcher, “Autonomous vibration of simple pressure-controlled valves in gas flows”, *J. Acoust. Soc. Am.* **93**, 2172–2180 (1993).
- ²⁰ F. Otondo and J. H. Rindel, “Directional representation of a clarinet in a room”, in *Proceedings of the Nordic-Baltic Acoustical Meeting*, Kgs. Lyngby, Denmark, 113–116 (2002).
- ²¹ H. Levine and J. Schwinger, “On the radiation of sound from an unflanged circular pipe”, *Physical Review* **73**, 383–406 (1947).
- ²² G. P. Scavone, “Modeling wind instrument sound radiation using digital waveguides”, in *1999 International Computer Music Conference, Beijing, China* (1999).
- ²³ W. H. Hayt and J. E. Kemmerly, *Engineering Circuit Analysis* (McGraw-Hill, Inc.) (1993).
- ²⁴ G. Velicic, E. L. Titlebaum, and M. F. Bocko, “Musical note segmentation employing combined time and frequency analysis”, in *ICASSP 2004 Tokyo, Japan (In-*

ternational Conference on Acoustics, Speech, and Signal Processing), IV:277–280 (2004).

²⁵ A. Klapuri, A. Eronen, J. Seppänen, and T. Virtanen, “Automatic transcription of music”, in *Proc. Symposium on Stochastic Modeling of Music, 22th of October, Ghent, Belgium*. (2001).

²⁶ C. Chafe and D. Jaffe, “Source separation and note iden-

tification in polyphonic music”, in *ICASSP 1986 (International Conference on Acoustics, Speech, and Signal Processing)* (1986).

²⁷ A. Almeida, C. Vergez, and R. Causse, “Quasistatic nonlinear characteristics of double reed instruments”, *J. Acoust. Soc. Am.* **121**, 536–546 (2007).