A Formant Filtered Physical Model for Wind Instruments

Axel Nackaerts, *Student Member, IEEE*, Bart De Moor, *Senior Member, IEEE*, and Rudy Lauwereins, *Senior Member, IEEE*

***Abstract*—In this paper, we report on our research concerning the calibration of physical models for sound synthesis. We combine waveguide physical modeling synthesis with formant filtering, by dividing the nonlinear description of the reed mechanism into a nonlinear part and an input-dependent linear filter. We elaborate on the calibration of the model and assess its performance by comparing it to a single-reed, cylindrical bore instrument, the clarinet.**

***Index Terms*—Digital filters, music, nonlinear systems, parameter estimation, sound synthesis.**

I. Introduction

In this paper, the focus is on the calibration of wind instrument models. The models presented here are a combination of simplified physical models for wind instruments and the formant filtering concept. These hybrid models are relatively easy to calibrate to recordings of a real instrument. We will mainly look at cylindrical bore, single-reed instruments, but the same concept can be used with minor changes for conical bores and other excitation mechanisms.

After an introduction to single-reed woodwind acoustics and the calibration problem of physical models for this class of instruments, we present our model in Section II. The calibration of the model is addressed in Section III and the performance is evaluated in Section IV.

The paper starts with an introduction to single-reed woodwind acoustics in Section I-A, followed by a description of physical modeling in Section I-B and of the calibration problem in Section I-C.

A. Single-Reed Woodwind Acoustics

Generally speaking, there are three types of instruments: those which generate waves in strings (piano, guitar, violin family, etc.), those which generate waves in air (clarinet, trumpet, organ, flutes etc.) and the percussion instruments. The second group can be classified by the excitation mechanism used: single reed (clarinet, saxophone, etc.), double reed (oboe, bassoon, etc.), air jet (flutes, organ, etc.), or the player's lips (trumpet, trombone, etc.) [1]. This paper discusses the single-reed woodwind instruments, with the clarinet as a typical example. The clarinet consists of three major parts as shown on Fig. 1. The mouthpiece/reed assembly (Fig. 2) is the excitation mechanism. The cane reed combined with the shape of the mouthpiece act as a pressure-controlled valve. The bore with the toneholes is a resonator. Depending on the configuration of the instrument, certain wave propagation modes are allowed. The bell is an impedance matching device between the surrounding air and the instrument bore.

The pressure difference between the oral cavity of the player (p_{oc}) and the pressure inside the mouthpiece induces a movement of the reed. The air flow into the mouthpiece is determined by this pressure difference and the tip opening by aerodynamic laws. This injection creates a pressure wave that propagates into the bore of the instrument. The bore acts as a filter on the traveling wave. When the wave reaches the bell, part of its energy is radiated outwards and part is reflected. The reflected wave travels back to the mouthpiece where it interacts again with the reed. This procedure repeats itself and creates a continuous pressure oscillation. This is a simplified view on the principles, detailed analysis can be found in [1] or [2].

The player can select different reeds and mouthpieces with different properties. During performance, the player controls the oral cavity pressure and the displacement of the reed at rest (y_0), as well as the configuration of the toneholes (Fig. 3). In a lesser extent, the player is able to modify the damping of the reed movement.

B. Physical Models

Physical models are based on the acoustical analysis of a real instrument. The modeling involves the decomposition of the instrument in several components as described in previously and subsequently generating a mathematical description of these components. The combination of the descriptions yields a complete model or virtual instrument that can be used for real-time performance. Fig. 4 shows a typical model of a wind instrument (after [3]). This approach has the advantage that we can build libraries of components and create new or physically not realizable instruments as we wish. Real instruments can be modeled by combining components to copy reality.

An important class of physical models are the digital waveguide models [4], [5]. These are essentially discrete-time

Manuscript received June 8, 2000; revised September 26, 2002. This work was supported by the Flemish Government: Research Council KUL (GOA Mefisto-666); FWO: (G.0240.99, G.0407.02, G.0197.02, G.0141.03, G.0491.03, G.0120.03, research communities ICCoS, ANMMM, Ph.D., and postdoctoral grants); AWI: Bil. Int. Research Program; IWT (Ph.D. grants, Soft4s); EU: Eureka 2063-Impact, Eureka 2419-FLiTE, CAGE, ERNSI; Federal State DWTC (IUAP IV-02, IUAP V-22), PODO-II (CP/40); Industrial contract research (IPCOS, Electrabel, Elia, Data4s, LMS, VIB). The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Bryan George.

A. Nackaerts is with the I.W.T. (Flemish Institute for Scientific and Technological Research in Industry), ESAT, B-3001 Leuven, Belgium (e-mail: Axel.Nackaerts@esat.kuleuven.ac.be).

B. De Moor and R. Lauwereins are with Katholieke Universiteit Leuven, Leuven, Belgium.

Digital Object Identifier 10.1109/TSA.2002.807351

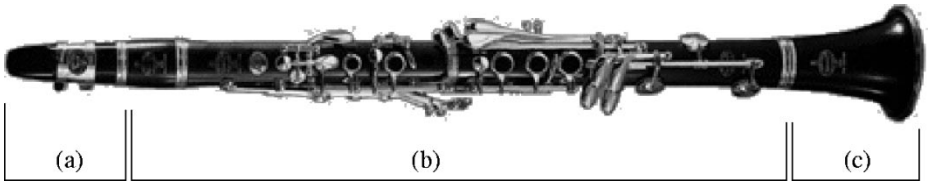

Fig. 1. The clarinet consists of three parts: (a) the mouthpiece/reed assembly, (b) the instrument bore with toneholes, and (c) the bell.

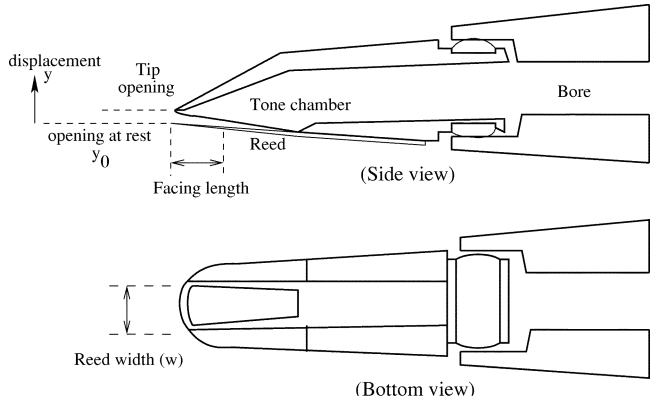

Fig. 2. The clarinet mouthpiece and reed assembly (side view and bottom view). The reed opening is determined by the pressure difference inside and outside the mouthpiece. The opening at rest and the facing length are chosen by the player by moving the player's lower lip. The reed width is fixed.

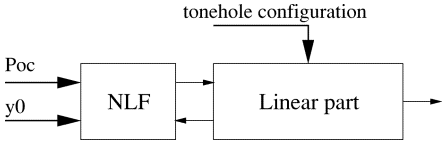

Fig. 3. The three main parameters controlled by the player during performance are the oral cavity pressure p_{oc}, the tip opening at rest (y_0) and the configuration of the toneholes (the fingering pattern). The tonehole configuration is implemented by building a different model for each possible pattern. Both p_{oc} and y_0 are direct inputs to the nonlinear function (NLF).

models resulting from the generic solution of the wave equations describing the system, often combined with models of lumped elements. This computationally interesting scheme is widely used.

C. The Calibration Problem

A roadblock to the general acceptance of physical models as "consumer" synthesis tools is the difficulty involved in creating and tweaking the parameters of those models. Mastering the control of all the model parameters to achieve one particular sound is difficult to learn and requires a good understanding of both the physical structure of the instrument and of the model used. These concepts are not user-friendly and merely tweaking the knobs will rarely produce a pleasing result, in contrast with subtractive or FM synthesis. The issue of calibration of physical models becomes more and more important as it opens the road to parametric sound compression in, e.g., MPEG-4 Structured Audio.

II. Physical Model

In this section we combine the traditional physical modeling view with formant filtering. The choice of the model determines the range of sound that it will be able to reproduce. We need a model that is flexible enough but also fairly easy to calibrate. The full acoustic model has too many parameters for our purposes, so we start with a very simple model.

A. Simple Physical Model

A simple and generic physical model has been studied by McIntyre, Schumacher, and Woodhouse (MSW) [6]. In the MSW model, tone production can be divided into two main parts: a nonlinear excitation and a linear resonator (see Fig. 5). For a given instrument, MSW synthesis models the objects and actions as a compact set of equations. The most complicated and instrument-specific equations describe the excitation, while the linear resonator primarily consists of a delay line and a linear filter. To obtain accurate results with this approach, a different model is needed for each note. The effective bore length (and thus the time delay) changes for different notes, as does the tonehole configuration and thus the internal reflection function. Notes played in a different register by using register holes or a different embouchure on the real instrument can also be modeled with a different bore length.

B. Physical Models With Formant Filtering

The linear part of the model is described first. The nonlinear part is split into the nonlinearity itself and a formant filter. Several one-note models are then combined into a multi-note model.

1) The Linear Part: The linear part of the model represents the instrument bore, the toneholes, and the bell. As we are only modeling one note, it is possible to lump all filtering effects together in one filter, and all time-delay into one delay line.

The time delay originates mainly in the length of the bore. Its length can be calculated with

$$L = \frac{F_s}{\chi f_0} \tag{1}$$

where F_s the sampling frequency, χ the integer proportionality factor depending on the bore type (conical or cylindrical) and f_0 the fundamental frequency of the note being modeled. It is clear that L seldom has an integer value. The most useful methods for achieving fractional delay lengths include Lagrange [7] and all-pass interpolation [8]. We used a first order Lagrange (linear) interpolator because of its simplicity. The coefficients of the general Lagrange interpolator are given by

$$h(n) = \prod_{k=0,k\neq n}^{N} \frac{D-k}{n-k} \tag{2}$$

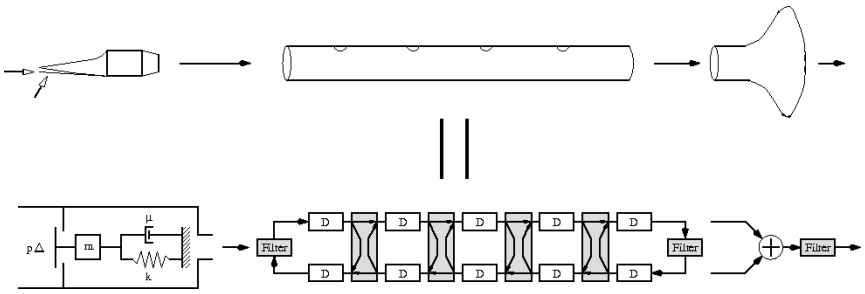

Fig. 4. A typical physical model for a wind instrument. From left to right one sees the nonlinear excitation mechanism, the resonating cavity or instrument bore, and the bell. The various parts are translated to delay lines and filters.

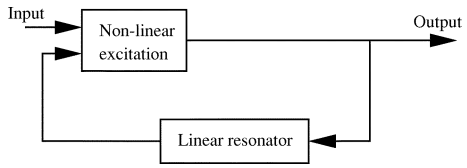

Fig. 5. McIntyre, Schumacher, and Woodhouse model simplifies the full model of Fig. 4 by dividing the instrument in a linear and a nonlinear part.

where D is the fractional delay and N the order of the filter, $n = 0, 1, \ldots N$. The first order filter ($N = 1$) is then calculated as

$$h(n) = [h(0)\ h(1)] = [-(D-1)\ D]. \tag{3}$$

Due to the physical structure of the bore, the toneholes, the bell, dispersion, and losses, the frequencies of the partials might be inharmonic. A very similar situation exists when modeling string instruments: the stiffness of the string introduces inharmonic behavior for the partials. We can easily correct these small frequency differences by using an allpass filter [9]. The allpass filter introduces a frequency-dependent phase delay, leading to slightly different frequencies for the higher partials. The bore length yielding a nth partial of frequency ν_r is

$$L = \chi \times n \times \frac{c}{\nu_r} \tag{4}$$

where c is the speed of sound in air. The frequency dependent bore length difference between the recording (with frequencies ν_r) and the synthetic sound ν_s can now be calculated as

$$\Delta L = \chi \times n \times \left(\frac{c}{\nu_r} - \frac{c}{\nu_s}\right). \tag{5}$$

The phase $\theta(\nu)$ of the all-pass filter now becomes

$$\theta(\nu) = 2\pi \frac{\Delta L}{c} \nu_s. \tag{6}$$

Several techniques exist for the design of all-pass filters [10]. In this model, we used a simple first order all-pass filter of the form

$$\mathcal{H}_{ap}(z) = \frac{a + z^{-1}}{1 + az^{-1}} \tag{7}$$

where a is a real coefficient. Note that this filter is not able to correct all the partials individually. The slight inharmonicity of the partials is almost inaudible.

2) The Nonlinear Part: Several nonlinear models exist for the single-reed mechanism [5], [11]. In this work, we used the Pressure-Dependent Reflection Coefficient method [12] in combination with a damped mass-spring system (see Fig. 6) [13]. This combination allows the use of natural parameters (reed stiffness, resonant frequency, reed tip/mouthpiece distance) and is inherently passive, thus improving the numerical stability. The reflection coefficient is calculated as

$$\mathcal{R} = \frac{Z_r - Z_b}{Z_r + Z_b} \tag{8}$$

where Z_r is the acoustic impedance of the reed/mouthpiece structure and Z_b represents the acoustic impedance of the instrument bore. Z_r is determined by

$$Z_r = \frac{p_\Delta}{u_r} \tag{9}$$

where u_r is the volume flow through the tip opening, and the pressure difference $p_\Delta = p_{oc} - p^-$ between the oral cavity and the tone chamber results from the feedback and input. The model is free of delay-free loops as there is no implicit feedback with this definition of p_Δ, so no special computational problems occur [14]. Note that in reality, the pressure difference is given by $p_\Delta = p_{oc} - (p^+ + p^-)$. Implementing this requires either the introduction of a delay, the analytical solution of the equations or an iterative numerical solution. Both choices for p_Δ generate quite similar output with appropriate choice of p_{oc}. The volume flow u_r is calculated as

$$u_r = w(y_0 - y) sgn(p_\Delta) \sqrt{\frac{2|p_\Delta|}{\rho}} \tag{10}$$

(w is the width of the reed, y_0 the tip opening at rest, y is the reed displacement relative to y_0, with a positive value for upward (closing) displacement, and ρ the density of air). This follows from conservation of energy and Bernouilli's equation. Further refinements of this formula were proposed in [15]. The tip opening was found as the displacement of the reed modeled as a damped mass-spring system with transfer function

$$\mathcal{H}(s) = \frac{A}{s^2 + 2\zeta\omega_n s + \omega_n^2} \tag{11}$$

in which ζ is the damping coefficient and ω_n the natural resonance frequency of the reed. The factor A is chosen such that the displacement y is equal to the rest opening y_0 when the dc input

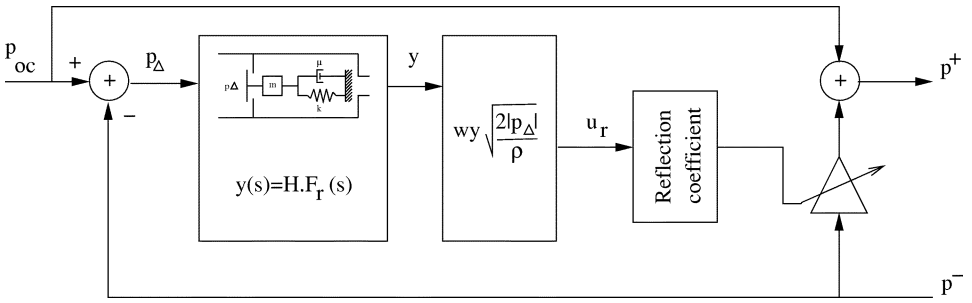

Fig. 6. The nonlinear excitation mechanism. The input and reflected wave determine the pressure difference p_Δ which drives a second-order damped mass-spring system, yielding the new reed opening y. With this opening, the volume flow u_r is calculated. The acoustic impedance of the reed/mouthpiece Z_r can now be calculated with u_r and p_Δ. Knowledge of Z_r leads to the reflection coefficient $\mathcal{R}$.

pressure equals the cut-off pressure, closing the tip opening. The displacement is then

$$y(s) = \mathcal{H}F_r(s) \tag{12}$$

where $F_r = S_r p_\Delta$ represents the force on the reed with surface S_r. The actual implementation uses a discretization $H(z)$ obtained by applying the backward Euler transform on (11). Fig. 6 represents the complete nonlinear part. A more accurate model, including displacement-dependent damping as a parameter has also been implemented [16]. Other models can be found in [14] and [17].

We can interpret the linear part of the model as the part that determines the allowed frequencies, and the nonlinear part as the distributor of the incoming and reflected energy at these frequencies. Of course, there is some interaction between the two parts through the feedback. In the normal playing range of the instrument, the reed behavior can be approximated by a memoryless function [6], [15], [18]. By doing this, the whole nonlinearity becomes memoryless and will not introduce any phase change in the loop. As the frequencies of the partials are defined by the total time delay of the loop (true delay and phase changes), they are independent of the nonlinearity itself. In this light, a fortissimo-played (but not overblown) instrument can be seen as a fully excited system, in which the amplitude of every allowed frequency has reached its maximum. It also means that when played on a softer level the amplitudes of some of the allowed frequencies will be lower, but the frequencies themselves won't change. Fig. 7 shows the spectra of a clarinet, recorded at three playing levels. We see that for low-amplitude input, only the fundamental frequency and the first harmonic are present with significant amplitudes. Louder playing levels introduce more harmonics. Physically, this is caused by the nonlinearity of the reed at higher displacements. For (theoretical) amplitudes larger than the opening at rest of the reed, hard clipping (*beating*) occurs.

The idea is now to split the nonlinearity into two parts: a nonlinear function, with maximum p_{oc} input (but still a variable y_0) and a linear filter, dependent on p_{oc}. The linear filter will adjust the amplitudes of the partials for all possible input levels. By doing this, the player loses direct control over the blowing pressure in the reed model (the player now controls the filter) but still controls the reed opening.

To achieve this division, a formant filter is used, as shown in Fig. 8. The first step in the formant filter design is to

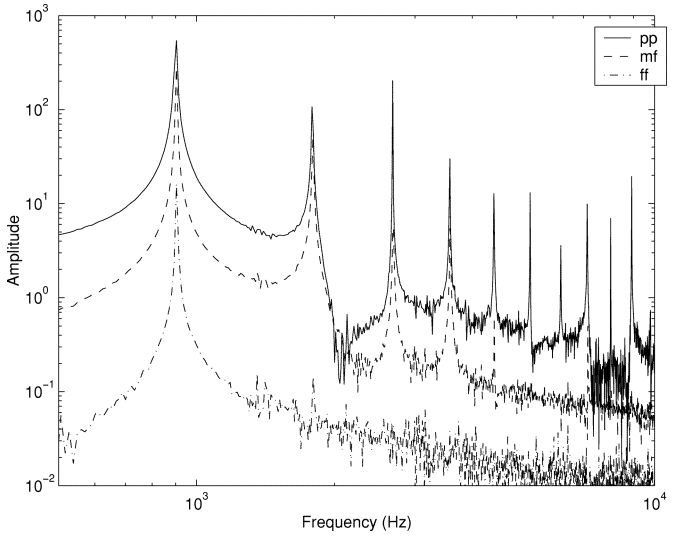

Fig. 7. The amplitude spectrum of a real clarinet played at three input levels (*pianissimo*, *mezzoforte*, and *fortissimo*). We see that for low-amplitude input, only the fundamental frequency and the first harmonic have significant amplitudes. Louder playing levels introduce more harmonics. Physically, this is caused by the nonlinearity of the reed at higher displacements. For (theoretical) amplitudes larger than the opening at rest of the reed, hard clipping (*beating*) occurs.

determine the spectral envelopes of the maximum-input model $(\mathcal{H}_{\text{maximum-input envelope}})$ and of a recording $(\mathcal{H}_{\text{desired envelope}})$. The formant filter is given by

$$\mathcal{H}_{\text{filter}} = \frac{\mathcal{H}_{\text{desired envelope}}}{\mathcal{H}_{\text{maximum-input envelope}}}. \tag{13}$$

Note that the poles of this filter might introduce unacceptable gains, so some precautions might be necessary: adding a high-pass filter to decrease the dc gain, adding a lowpass filter to decrease the high-frequency gain, and damping the resonant peaks. The order of this filter and the methodology used to calculate the envelopes will determine the final precision of the filter. Ideally, this filter should be determined at all possible inputs. In practice, the filter will only be determined for a few input pressures. By using interpolation, one is able to calculate intermediate approximations.

Measurements however showed that there is a positive correlation between playing frequency and blowing pressure with mechanical blowing. This was not always the case when the instrument was played by a professional performer [19]. This effect is attributed to the nonlinear dependence of the hydrody-

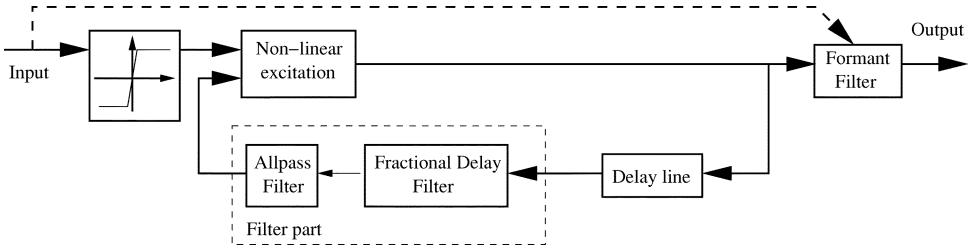

Fig. 8. The dynamic model has been combined with a formant filter. The physical model now acts as an oscillator with maximum input. The linear part has been split in a delay line and a filter part including a fractional delay filter and the allpass filter. The output goes through the input-dependent formant Filter.

namic force on the reed and by the elastic collision of the reed on the lay at high pressures [20]. The same also occurs if the lip pressure is changed. We can include this effect by slightly altering the length of the delay line. Again, this is not included in the nonlinear part of the model but in the characteristics of the resonator.

One aspect not yet present in the model is the aerodynamic noise, resulting from the turbulent air flow in the mouthpiece. This is modeled as white noise added to the reflected wave, modulated by the volume flow through the reed channel [21], [22]. The noise level can be estimated by examining the noise subspace [23].

3) Multinote Models: When playing a note sequence on a real instrument one effectively changes the acoustic configuration. During the transition of one note to another, the real instrument goes through a whole range of intermediate states. The output of the presented model should match the output of the real instrument during these states as closely as possible. Several possible multinote models have been studied [24]. Fig. 9 shows the chosen multi-note model. During the note transition, two calibrated formant filtered models coexist with joined nonlinear function (NLF). The parameters a and b determine the relative importance of the two resonators during the transition. They can be seen as the transmission and reflection parameters of the toneholes, lumped together at the output of the delay lines. This approximation is fairly accurate with adequate choice of the parameters a and b. Note that this is not a tonehole model but an approximation of the complete instrument during the transition. Many other multi-note models exist and can be found in [5], [25]–[27].

III. Calibration

The calibration of the physical model is quite important, as inappropriate parameters might render the model totally unplayable. Previous work has been done on the estimation of the nonlinear function and the resonator [28]–[30]. These methods try to measure the characteristics of the instrument (nonlinearity and resonator properties) directly by applying a known input. Our goal is to calibrate a model without specialized equipment. To calibrate the presented model, only a recording of a note played over the complete dynamic range is needed, for each note to be modeled. Following the design flow of Fig. 10, we will now calibrate the formant filtered model. The first part of the calibration process is the estimation of the parameters of the reed model. It is known that the clarinet operates in the stiffness-dominated region of the mass-spring system. Normal

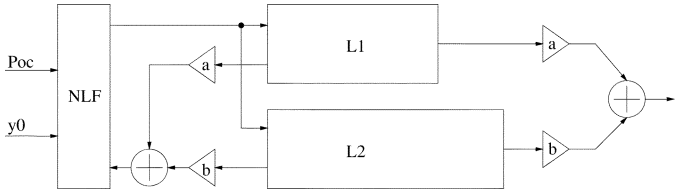

Fig. 9. Multinote model using two resonators and one nonlinear function during the note transition. The parameters a and b determine the relative importance of the resonators.

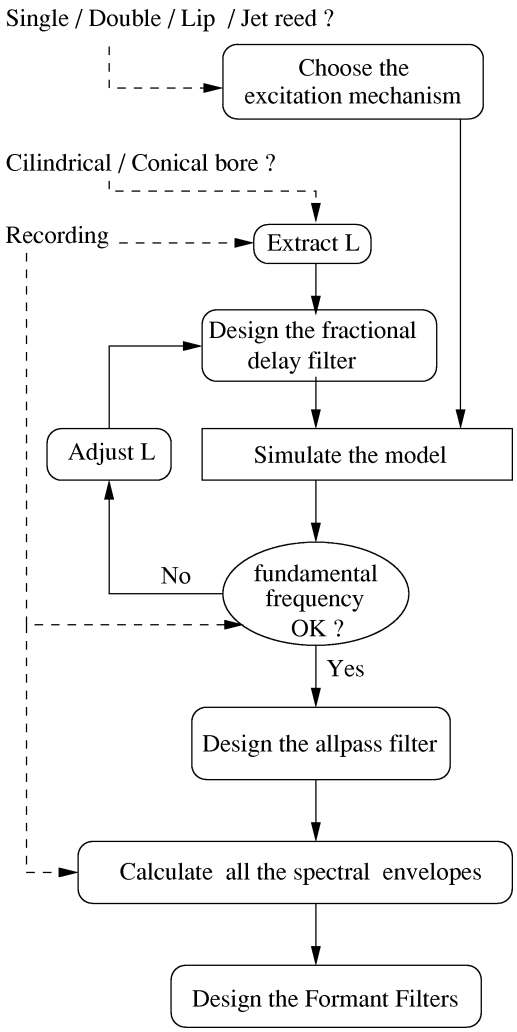

Fig. 10. The design flow for calibrating the formant filtered physical models. One starts by building a general model using an appropriate nonlinear characteristic and using rule-of-thumb parameters. The procedure then designs the linear part such that the simulation matches the recording.

playing frequencies for a clarinet are under 1 kHz so we expect the resonant frequency to be in the range 2–3 kHz. The damping coefficient ranges from 0.05 to 0.35 with little effect on the frequency domain behavior. With a lower damping coefficient for the reed, the model tends to "lock in" and oscillate at the reed's resonant frequency. Detailed measurements of different reeds have confirmed these parameter ranges [16]. These parameters do influence the time-domain (transient) response but have little effect on the steady-state tone. In the normal playing range, they affect the amplitude of the higher harmonics. The main contribution of these parameters is in the player interaction: they determine the oral cavity pressure and rest opening needed to generate a tone of a certain amplitude. This effect is captured in the parameterization of the formant filter.

It is fairly easy to design the linear part, by calculating the Short-Time Fourier Transform of the recording and extracting the frequencies of the partials. Using these values with (1), (2), and (5) gives the delay length, fractional delay and phase correcting filter.

The last step is the design of the formant filter. As stated earlier, the order of the filter determines the precision of the result. We start with a second-order approximation of the envelope and extend it to Nth order using LPC.

The second-order spectral envelope can be described by

$$\mathcal{H}_{\text{env}} = \frac{A_{\text{env}}\omega_{\text{env}}^2}{s^2 + 2\xi_{\text{env}}\omega_{\text{env}}s + \omega_{\text{env}}^2} \tag{14}$$

in which A_{env} denotes the amplitude, ξ_{env} the damping coefficient $(0 < \xi_{\text{env}} \leq 1)$ and ω_{env} the resonant frequency in the absence of damping [31]. The lower limit for the parameter $\xi_{\text{env}}(\xi_{\text{env}} > 0)$ ensures that a stable filter is obtained. Without the upper limit $\xi_{\text{env}} \leq 1$, the poles of system would split to two different, real poles. The coefficients of this filter can be found by minimizing the weighted least squares problem (WLS) with cost function

$$F = \sum W|\mathcal{S}|(|\mathcal{S}| - |\mathcal{H}_{\text{env}}|)^2. \tag{15}$$

Here, $\mathcal{S}$ is the spectrum of the signal as determined by the Short-Time Fourier Transform. W is a scaling factor, chosen large when $|\mathcal{S}| > |\mathcal{H}_{\text{env}}|$ and small otherwise. The goal of this cost function is to minimize $(|\mathcal{S}| - |\mathcal{H}_{\text{env}}|)^2$, but weighted such that the dominant spectral peaks are accurately followed. This is done by an extra multiplication with $|\mathcal{S}|$: strong peaks have more impact that small peaks or the noise floor. The extra weighting with W is needed to ensure that the spectral envelope is "dressed" on top of those peaks. Standard LS fits the envelope through the peaks in the spectrum: giving more weight to the spectrum above the fitted envelope pulls the fit upwards. The WLS problem is solved with a Gauss-Newton algorithm. Fig. 11 shows two such spectral envelopes. Plotting the extracted coefficients (see Fig. 12) as a function of time for a soft-to-loud sequence reveals that the spectral envelopes go from highly-resonant (with resonance peak around the fundamental frequency), to fully damped ($\xi = 1$, with -3dB frequency around three times the fundamental frequency). The resulting formant filter according to (13) is

$$\mathcal{H}_{\text{filter}} = \frac{A_r\omega_r^2 s^2 + 2\xi_s\omega_s s + \omega_s^2}{A_s\omega_s^2 s^2 + 2\xi_r\omega_r s + \omega_r^2} \tag{16}$$

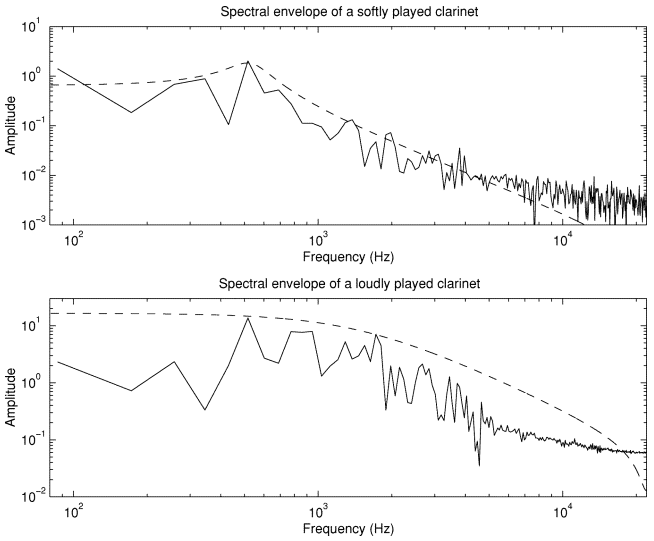

Fig. 11. The second-order envelope spectrum of equation (14) (dashed line) fitted to the spectrum of a softly and loudly played clarinet (full line).

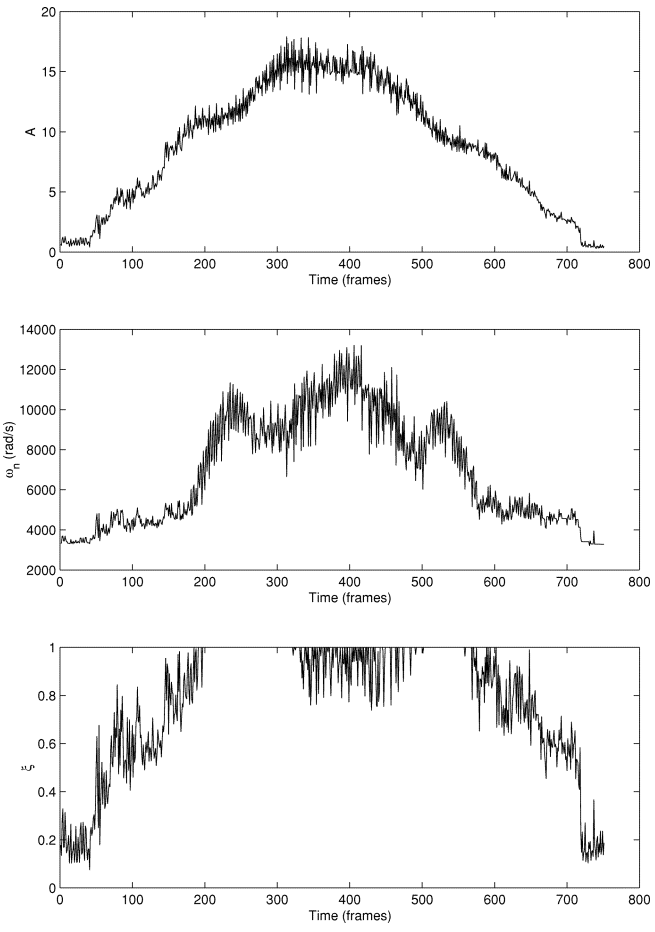

Fig. 12. Evolution of the envelope parameters for a *pp* to *f f* to *pp* recording. Shown are the extracted amplitude (A_{env}), resonant frequency $(\omega_{\text{n,env}})$ and damping coefficient (ξ_{env}). Note that ξ_{env} clips to its maximal value (1) for the highest amplitudes. The frames are 1024 samples or 23.2 ms long (44.1kHz sampling rate), the complete signal has 17.4 s duration.

in which the subscripts r and s denote the parameters estimated from, respectively, the recording and the synthetic output.

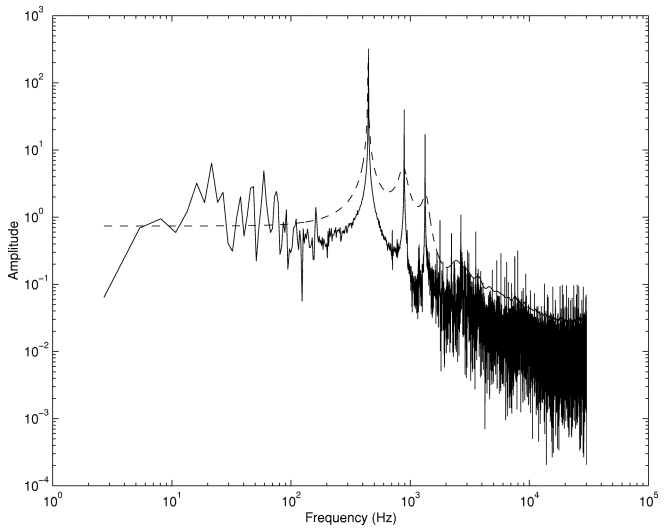

Fig. 13. The spectrum envelope (dashed line) as determined with the LPC method for a pp played clarinet (full line). The sampling frequency is 44.1 kHz.

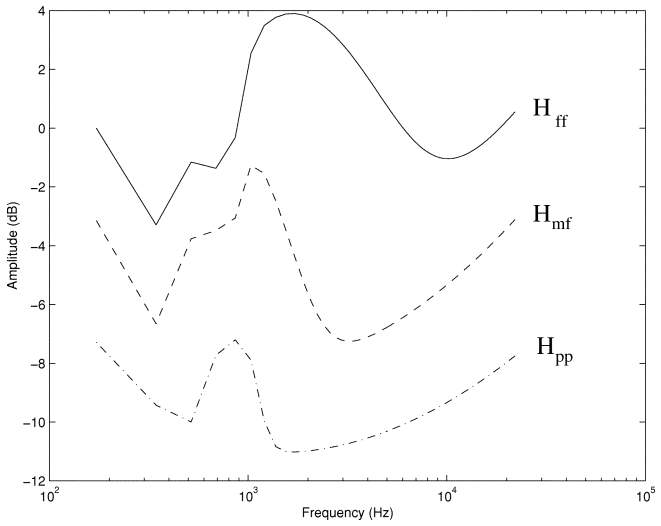

Fig. 14. Three extracted Formant Filters for three amplitudes determined with the LPC method.

Extension to N-th order is possible using a technique like Linear Predictive Coding. LPC fits an allpole model with transfer function

$$\mathcal{H}_{lpc}(z) = \frac{G}{1 + \sum_{k=1}^{p} a_k z^{-k}} \tag{17}$$

where G is a gain factor. LPC does not exactly yield the spectral envelope, but a best fit according to some cost function. The standard LP method uses least squares as the error measure. Linear prediction has two important spectral matching properties [32]: globally, the matching process should perform uniformly over the whole frequency range, irrespective of the general shaping of the spectrum and locally, the minimization of the error measure results in a model spectrum $\hat{\mathcal{S}}$ that is a good estimate of the spectral envelope of the signal spectrum $\mathcal{S}$. Using the standard LP error measure has the disadvantage of the cancellation of errors: the contributions to the error when $\mathcal{S} > \hat{\mathcal{S}}$ cancel those when $\mathcal{S} < \hat{\mathcal{S}}$. The cost function (15) solves this, but has to be minimized using a nonlinear optimization routine. When allowing enough poles, the resulting filter $\mathcal{H}_{\text{lpc}}$ still follows the envelope closely when using standard LPC. Note that if too many poles are selected, the result will be the spectrum itself and not the envelope. If the LP order is kept far below the number of partials, this is not an issue. We choose to use this filter as the description of the envelope ($\mathcal{E}_{\text{lpc}}$). An example of an envelope determined with the LPC method can be seen in Fig. 13.

In the low-order case, it is fairly easy to parameterize the filter for the complete input range. For the Nth order LPC case, the spectrum envelope was extracted for only three playing levels ($\mathcal{E}_{pp}, \mathcal{E}_{mf}$), $\mathcal{E}_{ff}$ and for the model output ($\mathcal{E}_{synth}$)). The formant filter frequency response ($\mathcal{H}_{ff,mf,pp}$) was then calculated as

$$\mathcal{H}_{ff,mf,pp}(z) = \frac{\mathcal{E}_{ff,mf,pp}(z)}{\mathcal{E}_{synth}(z)}. \tag{18}$$

These can be found in Fig. 14. The filters representing all other levels were obtained by interpolating between these three levels depending on the input level. This has been validated by comparing the interpolated filters with the filters extracted of a soft-to-loud sequence. Only a slight difference could be heard when the mf point was well chosen. The formant filters are thus

$$\mathcal{H} = a_1\mathcal{H}_{pp} + a_2\mathcal{H}_{mf} + a_3\mathcal{H}_{ff} \tag{19}$$

where a_1, a_2 and a_3 determine the interpolation point. The interpolation parameters are simply determined by the local amplitude A_{signal} and the amplitude of the pp, mf and ff signals (A_{pp}, A_{mf}, A_{ff}) as

$$a_1 = 1 - q \quad a_2 = q \quad a_3 = 0 \tag{20}$$

where $q = (A_{\text{signal}} - A_{pp})/(A_{mf} - A_{pp})$ in the case $A_{\text{signal}} \leq A_{mf}$, and

$$a_1 = 0 \quad a_2 = 1 - q \quad a_3 = q \tag{21}$$

where $q = (A_{\text{signal}} - A_{mf})/(A_{ff} - A_{mf})$ when $A_{\text{signal}} > A_{mf}$. Alternatively, very long FIR formant filters can be obtained by directly calculating the Wiener filter that equalizes the model output to match the recorded output.

IV. Performance

The models were implemented as MatLab software. The calibration data are 16-bit 44.1kHz recordings of a clarinet, made in a dry recording room, using an AKG C4000B microphone. The musician played several chromatic scales with different loudness, as well as soft-to-loud sequences. These recordings were segmented per note by hand-selecting fragments with steady loudness. The segments were used by the calibration algorithm.

The derived models have a very good spectral match with the recordings. Informal listening tests indicated that it is difficult to discern real and synthetic sound for steady state tones. Dynamic changes also sound natural if the input is derived from real data.

As an objective test, we calculated the weighted relative power spectral error (RPSE) defined as

$$\text{RPSE} = \frac{\sum_{\text{bins}} W|\mathcal{S}^2_{\text{reference}} - \mathcal{S}^2_{\text{synthetic}}|}{\sum_{\text{bins}} \mathcal{S}^2_{\text{reference}}} \tag{22}$$

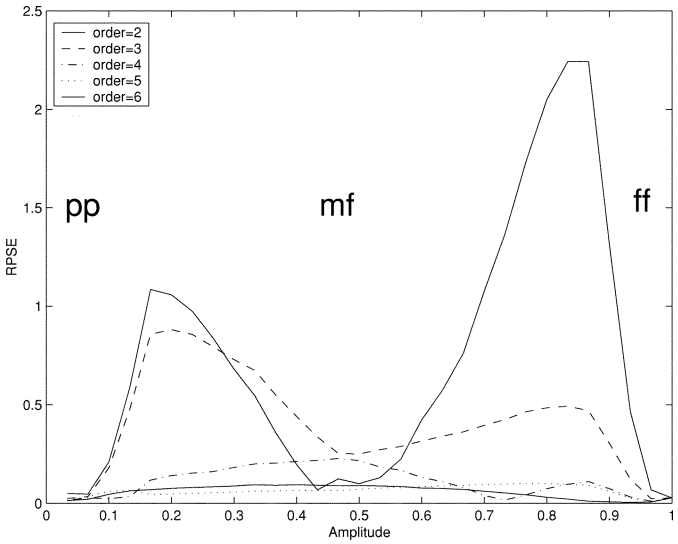

Fig. 15. This figure shows the Relative Power Spectral Error for the formant filtered model. The formant filters are IIR filters calculated using LPC.

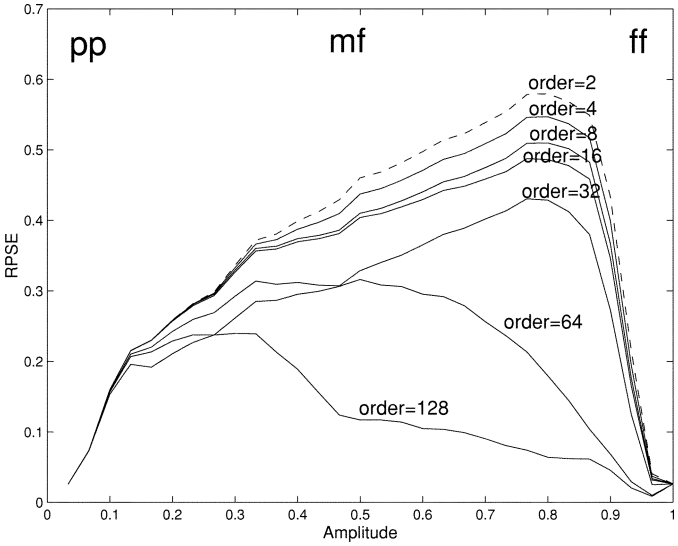

Fig. 16. This figure shows the Relative Power Spectral Error for the formant filtered model. The formant filters are FIR Wiener filters with *order* taps. Note that the scale of the Y axis differs from the scale of Fig. 15.

in which $\mathcal{S}$ is the amplitude of the spectrum and W a weighting factor. The weighting factor W was chosen flat up to 8 kHz and declining for higher frequencies. One could use a perceptual auditory model for a more accurate representation of the subjective sound quality [33]. Figs. 15 and 16 show the RPSE for increasing filter order, using the LPC-derived IIR formant filters and the Wiener FIR filters. As can be seen on the figures, it is clear that higher-order filters provide a lower RPSE. The amplitudes at which $\mathcal{H}_{pp}$, $\mathcal{H}_{mp}$ and $\mathcal{H}_{ff}$ were determined are indicated.

The transient behavior at onset of a note mostly depends on the chosen nonlinear excitation mechanism. Small adjustments of the parameters of the nonlinear part can alter the attack time and overblowing sensitivity. Fig. 17 shows the spectrogram of a tone going from pp to ff. Several audio examples can be found at ftp://ftp.esat.kuleuven.ac.be/pub/sista/nackaerts/examples/formant/. The calculation load is somewhat higher than

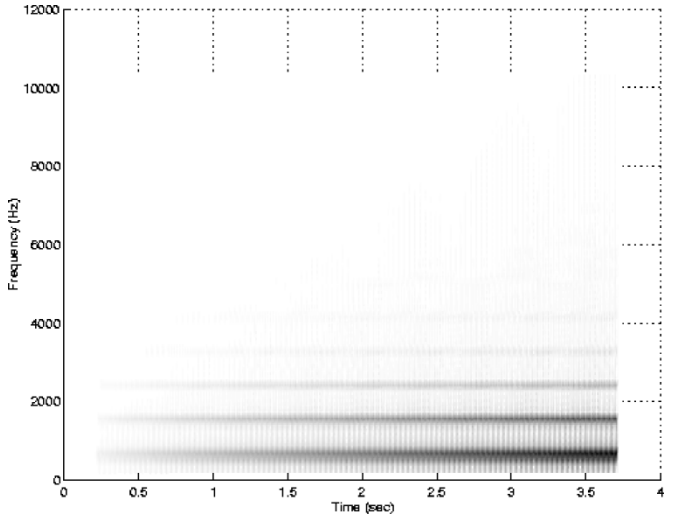

Fig. 17. Spectrogram of the formant filtered physical model of a clarinet, from pp to ff.

conventional models and depends largely on the order of the formant filter. In the case of a second order filter, the impact is very low. An MPEG-4 Structured Orchestra Language [34] implementation of the model showed that realtime performance is easily attained. For instance, the model with second order formant filter runs seven times faster than realtime on a Pentium 1 GHz machine, using the sfront} [35] SAOL-to-C compiler and GNU gcc.

V. Conclusion

A combination of physical modeling synthesis and formant filtering with improved calibration properties has been elaborated for wind instruments. These models can be calibrated solely based on a recording of the output of a real instrument. Simulation of the model proved that the steady-state output matches the recorded tone closely.

References

[1] A. H. Benade, *Fundamentals of Musical Acoustics*, 2nd ed. New York: Dover, 1990.

[2] N. H. Fletcher and T. D. Rossing, *The Physics of Musical Instruments*. New York: Springer-Verlag, 1991.

[3] J. O. Smith III, "Physical modeling using digital waveguides," *Comput. Music J.*, vol. 16, no. 4, pp. 74–91, 1992.

[4] Discrete-Time Modeling of Acoustic Systems With Applications to Sound Synthesis of Musical Instruments——, (1996, Apr.). *http://www-ccrma.stanford.edu/~jos/waveguide/* [Online]

[5] G. P. Scavone, "An Acoustic Analysis of Single-Reed Woodwind Instruments With an Emphasis on Design and Performance Issues and Digital Waveguide Techniques," Ph.D. thesis, CCRMA at Stanford University, Mar. 1997.

[6] M. E. McIntyre, R. T. Schumacher, and J. Woodhouse, "On the oscillations of musical instruments," *J. Acoust. Soc. Amer.*, vol. 74, no. 5, Nov. 1983.

[7] T. L. Laakso, V. Välimäki, M. Karjalainen, and U. Laine, "Real-time implementation techniques for a continuously variable digital delay in modeling musical instruments," in *Proc. 1992 Int. Computer Music Conf.*, 1992, pp. 14–18.

[8] J. O. Smith III, "Techniques for Digital Filter Design and System Identification With Application to the Violin," Ph.D. dissertation, Elect. Eng. Dept., Stanford Univ., Stanford, CA, 1983.

[9] D. A. Jaffe and J. O. Smith III, "Extensions of the karplus-strong plucked string algorithm," *Comput. Music J.*, vol. 7, no. 2, pp. 56–69, 1995.

[10] T. Q. Nguyen, T. I. Laakso, and R. D. Koilpillai, "Eigenfilter approach for the design of allpass filters approximating a given phase response," *IEEE Trans. Signal Processing*, vol. 42, pp. 2257–2263, Sept. 1994.
[11] D. H. Keefe, "Wind-instrument reflection function measurement in the time domain," *J. Acoust. Soc. Amer.*, vol. 99, no. 4, pp. 2370–2381, 1996.
[12] J. O. Smith III, "Efficient simulation of the reed-bore and bow-string mechanisms," in *Proc. 1986 Int. Computer Music Conf.*, Den Haag, The Netherlands, 1986, pp. 275–280.
[13] W. E. Worman, "Self-Sustained Nonlinear Oscillations of Medium Amplitude in Clarinet-Like Systems," Ph.D. dissertation, Case Western Reserve Univ., 1971.
[14] G. Borin, G. De Poli, and D. Rocchesso, "Elimination of delay-free loops in discrete-time models of nonlinear acoustic systems," *IEEE Trans. Speech Audio Processing*, vol. 8, pp. 597–605, Sept. 2000.
[15] J. Backus, "Small vibration theory of a clarinet," *J. Acoust. Soc. Amer.*, vol. 35, no. 3, pp. 591–599, 1963.
[16] W. Ermens, "Single-Reed Wind Instruments: The Non-Linear Oscillator," M.S. dissertation, Dept. of Electrical Engineering, Katholieke Universiteit, Leuven, Belgium, 1999.
[17] D. H. Keefe, "Physical modeling of wind instruments," *Comput. Music J.*, vol. 16, no. 4, pp. 57–73, 1992.
[18] C. J. Nederveen, *Acoustical Aspects of Woodwind Instruments*. Amsterdam, The Netherlands: Frits Knuf, 1969.
[19] N. Bak and P. Domler, "The relation between blowing pressure and blowing frequency in clarinet playing," *Acustica*, vol. 63, p. 238, 1987.
[20] A. Hirschberg, R. W .A. van de Laar, J. P. Marrou-Maurières, A. P. J. Wijnands, H. J. Dane, S. G. Kruijswijk, and A. J. M. Houtsma, "A quasistationary model of air flow in the reed channel of single-reed woodwind instruments," *Acustica*, vol. 70, pp. 146–154, 1990.
[21] D. A. Jaffe and J. O. Smith III, "Performance expression in commuted waveguide synthesis of bowed strings," in *Proc. 1995 Int. Computer Music Conf.*, 1995, pp. 343–346.
[22] G. P. Scavone, "Modeling and control of performance expression in digital waveguide models of woodwind instruments," in *Proc. 1996 Int. Computer Music Conf.*, 1996, pp. 224–227.
[23] Y. Ephraim and H. L. Van Trees, "A signal subspace approach for speech enhancement," *IEEE Trans. Speech Audio Processing*, vol. 3, pp. 251–266, July 1995.
[24] A. Nackaerts, B. De Moor, and R. Lauwereins, "Note transitions in physical models of a wind instrument," in *Acoustics and Music, Biology and Chemistry, Business and Economics*, N. Mastorakis, Ed. Singapore: World Scientific, 2000, Mathematics and Computers in Modern Science, pp. 59–63.
[25] T. L. Laakso, V. Välimäki, M. Karjalainen, and U. Laine, "Real-time implementation techniques for a continiously variable digital delay in modeling musical instruments," in *Proc. 1992 Int. Computer Music Conf.*, 1992, pp. 14–18.
[26] G. P. Scavone and P. R. Cook, "Real-time computer modeling of woodwind instruments," in *Proc. 1998 Int. Symp. Musical Acoustics*, Leavenworth, WA, 1998.
[27] M. Van Walstijn and G. Scavone, "The wave digital tonehole model," in *Proc. 2000 Int. Computer Music Conf.*, Berlin, Germany, 2000.
[28] P. R. Cook, "Non-linear periodic prediction for on-line identification of oscillator characteristics in woodwind instruments," in *Proc. 1991 Int. Computer Music Conf.*, 1991, pp. 157–160.
[29] G. Scavone and P. R. Cook, "Combined linear and nonlinear prediction in calibrating models of musical instruments to recordings," in *Proc. 1994 Int. Computer Music Conf.*, 1994, pp. 433–434.
[30] V. Välimäki, M. Karjalainen, and J. Huopaniemi, "Measurement, estimation, and modeling of wind instruments using DSP techniques," in *Proc. Int. Congr. Acoustics (I CA '95)*, Trondheim, Norway, June 26–30, 1995, pp. 513–516.
[31] A. Nackaerts, B. De Moor, and R. Lauwereins. Dynamic timbre matching for sound synthesis using time-varying parametric filters. presented at *Proc. 43rd IEEE Midwest Symp. Circuits and Systems (MWSCAS 2000, CDROM)* [Online]
[32] J. Makhoul, "Linear prediction: A tutorial review," *Proc. IEEE*, vol. 63, pp. 561–580, 1975.
[33] B. Paillard, P. Mabilleau, S. Morisette, and J. Soumagne, "Perceval: Perceptual evaluation of the quality of audio signals," *J. Audio Eng. Soc.*, vol. 40, pp. 21–31, 1992.
[34] International Standards Organization, International Standard ISO 14 496 (MPEG-4),Part 3 (audio), usbpart 5 (Structured Audio), CH:ISO, Geneva, Switzerland, 1999.
[35] J. Lazzaro and J. Wawrzynek, "Compiling MPEG-4 structured audio intoC," in *Proc. 2nd IEEE MPEG-4 Workshop and Exhibition (WEMP)*, San Jose, CA, June 18–20, 2001.

Axel Nackaerts (S'02) was born in Leuven, Belgium, on September 21, 1975. He received the degree of burgerlijk elektrotechnisch ingenieur from the Katholieke Universiteit Leuven in 1998. Since 1998, he has been pursuing the Ph.D. degree as a Research Assistant with the I.W.T. (Flemish Institute for Scientific and Technological Research in Industry) in the SCD Research Group of the Department Elektrotechniek, Faculty of Engineering, K.U. Leuven.

His principal areas of interest are the design and calibration of physical models of musical instruments and their implementation with the MPEG-4 Structured Audio Orchestra Language.

Bart De Moor (M'86–SM'93) was born on July 12, 1960, and received his doctoral degree in applied sciences in 1988 from the the Katholieke Universiteit Leuven, Leuven, Belgium.

He is Full Professor at the Department of Electrical Engineering, Katholieke Universiteit Leuven, in the research group SCD. He was a Visiting Research Associate (1988–1989) with the Departments of Computer Science and Electrical Engineering, Stanford University, Stanford, CA. His research interests include numerical linear algebra, system identification, control theory and datamining. He has more than 200 papers in international journals and conference proceedings (see http://www.esat.kuleuven.ac.be/sista-cosic-docarch for a list), is the co-author of several books on system identification and neural nets. In 1991–1992, he was the Chief of Staff of the Belgian Federal Minister of Science, Wivina Demeester-DeMeyer, and later on the staff of the Belgian Prime Minister, Wilfried Martens. From 1994 to 1999, he was the main Advisor on Science and Technology policy of the Flemish Minister-President, Luc Van den Brande.

Dr. De Moor has received several national and international awards for his work. He is a member of several boards of administrators of (inter)national scientific, cultural and commercial organizations. He is also one of the founders of ISMC NV and Data4S NV.

Rudy Lauwereins (M'98–SM'01) received the M.Eng. and Ph.D. degrees in electrical engineering from the Katholieke Universiteit Leuven, Belgium, in 1983 and 1989, respectively.

In 1991, he was a Visiting Researcher at the IBM Almaden Research Center, San Jose, CA, on a postdoctoral NFWO research fellowship. In 1998, he was advisor to the team that won the Northern European section of the Texas Instruments DSP Solutions Challenge. He became a Professor at the Katholieke Universiteit Leuven, Belgium, in 1993 and Vice-President of IMEC in 2001. His main research interests are in computer architectures, implementation of interactive multimedia applications on dynamically reconfigurable platforms, and wireless communication. In these fields, he has authored and co-authored more than 250 publications in international journals and conference proceedings.

Dr. Lauwereins was voted by students as the Best Teacher of the engineering faculty's Master's curriculum in 2000-2001 and received a nomination for the same prize in 2001-2002.