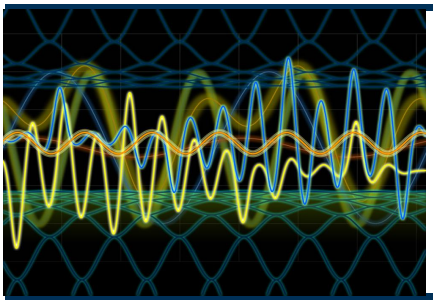
Volume 49 http://acousticalsociety.org/



Fourth Vienna Talk on Music Acoustics

University of Music and Performing Arts
Vienna, Austria
11-14 September 2022

Musical Acoustics: Paper Session 9

A method for tracking the frequencies and amplitudes of partials in transient musical signals

Péter Rucz
Faculty of Electrical Engineering and Informatics, Department of Networked Systems and Services, Budapest University of Technology and Economics: Budapesti Muszaki es Gazdasagtudomanyi Egyetem, Budapest, H-1117, HUNGARY; rucz@hit.bme.hu

Péter Tamás Nagy and András Szabó
Faculty of Mechanical Engineering, Department of Hydrodynamic Systems, Budapest University of Technology and Economics: Budapesti Muszaki es Gazdasagtudomanyi Egyetem, Budapest, H-1111, HUNGARY; pnagy@hds.bme.hu; aszabo@hds.bme.hu

Pitch detection and tracking is a classical problem of signal processing. Over the years, many approaches have been elaborated that can be applied in various fields, including the analysis of musical or speech signals. In this contribution, a method is introduced for extracting the frequencies and amplitudes of partials in transient musical signals, intended to be used for post-processing recorded audio samples. A band-pass filtered and a complex base-band signal are created, which are processed block by block. In each block, the frequency of each partial is assumed to be constant, while the amplitude varies exponentially with a constant growth rate in time. These parameters are found by means of a least-squares fit performed on the complex, base-band signal. Then, the signal block is reconstructed by fitting a harmonic function onto the corresponding block of the band-pass filtered signal. The performance of the proposed technique is compared with that of an other method applicable for pitch detection and tracking. The above signal processing strategy is applied for analyzing sound samples of an artificially blown tenor recorder. Systematic frequency changes are found during the attack stage, which are interpreted by analyzing the jet-drive model of sound production by the recorder.

Published by the Acoustical Society of America

 https://doi.org/10.1121/2.0001680

1. INTRODUCTION

Pitch detection and tracking are classic problems of signal processing. Many approaches were developed in the last few decades, aiming at various fields of applications, such as music or speech processing, among others. In this contribution an approach is proposed for identifying and tracking the changes of the frequencies and magnitudes of partials in transient musical samples. The objective of this paper is to introduce this method and demonstrate its application by analyzing recorded sound samples.

The proposed technique processes the signal in a block-by-block manner and makes use of complex modulation and least-squares fits for estimating the parameters of the signal. In its present form, the proposed approach is an offline method, and real-time processing is out of scope. In this paper, the approach is applied for analyzing the attack transients of a tenor recorder, sounded with different fingerings and blowing pressures. The method is shown to be capable of reliably tracking the rapid variations of the frequencies and amplitudes of the most significant partials during the onset of the sound.

The paper is structured as follows. Section 2 introduces the proposed method as well as an alternative strategy based on the extended Kalman filter. The attack transients of a tenor recorder are analyzed in Section 3 and the temporal growths and the frequencies of the partials are examined in detail. Explanations to the observed changes of the frequencies during the attack transient are sought in Section 4 by using a phase lock condition derived for a simplified jet-drive model of the recorder in the complex frequency domain. Finally, Section 5 concludes the paper by a brief summary.

2. METHODOLOGY

A. SIGNAL MODEL

The attack transients of air jet instruments are characterized by the following properties. In the onset of the sound a transition from the edge tone to the "pipe" tone is observed. First, the aeroacoustic sources of the oscillating air jet dominate the spectrum, while the vibrations of the air column enclosed in the resonator develop slower. In the attack, multiple acoustical modes can be excited at once, while a single mode becomes dominant finally in the steady state. These two phenomena lead to the observation that the partials develop with different envelopes in time and their frequencies are also expected to change during the transient. The attack transients of flute-like instruments have already been analyzed in various studies;[1–3] however, to the best knowledge of the authors, the frequency variations of the partials were not yet examined in detail.

The analyzed signal $p(t)$ is assumed to be composed of N partials, whose amplitude, frequency and phase shift is expected to vary slowly in time:

$$p(t) \approx \sum_{n=1}^{N} \tilde{P}_n(t) \cos\left(2\pi \tilde{f}_n(t) t + \tilde{\phi}_n(t)\right) + \nu(t), \tag{1}$$

where $\tilde{P}$, $\tilde{f}$, and $\tilde{\phi}$ are the slowly varying amplitude, frequency, and phase, respectively, with the subscript n referring to the n-th partial, and $\nu(t)$ represents an additive noise that is not correlated to any of the partials. By slowly varying it is supposed that the relative change of the respective quantity is small in one period.

While the signal model was introduced here for air reed instruments, various other kinds of musical instruments (e.g. strings, woodwinds, or brass instruments) produce similar signals. Thus, the method may be applied for analyzing sound samples of a large variety of instruments.

B. THE PROPOSED APPROACH

The proposed method estimates the parameters $\tilde{p}$, $\tilde{f}$, and $\tilde{\phi}$ of the signal in three different paths, as shown in Figure 1. The input signal $p(t)$ is sampled at discrete time instances $p[k] = p(kT_\mathrm{s})$, with T_s

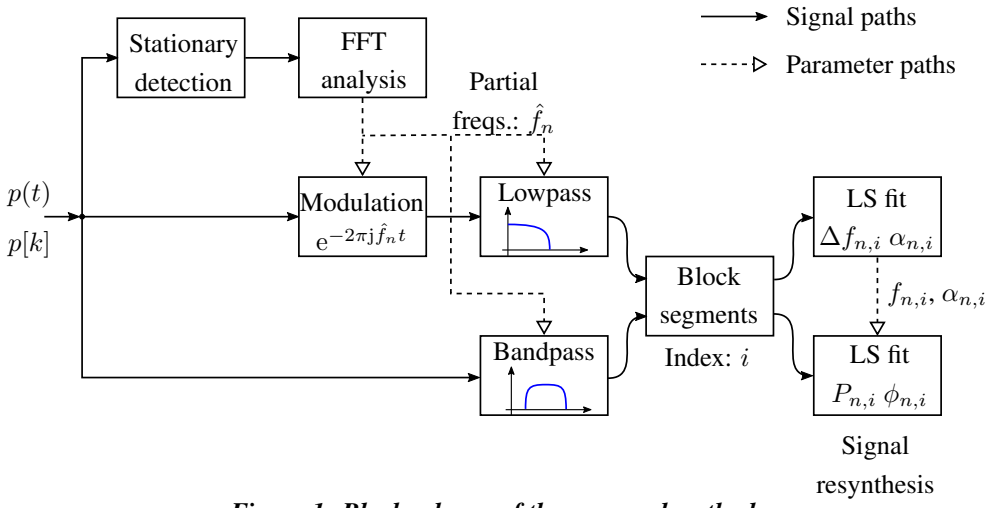


Figure 1: Block scheme of the proposed method

denoting the sampling interval. First, the stationary section of the signal is detected and a frequency analysis is performed using the fast Fourier transform. From this step, an estimate on the frequencies of the partials $\hat{f}_n$ ($n = 1, 2 \dots N$) is attained. It is noted, that this step can also be skipped if an a priori estimate on the frequencies of the partials is available.

The other two paths process the signal partial-by-partial. In the base-band branch (middle branch in Figure 1), a complex modulation is performed, using the modulation signal $\exp(-2\pi \mathrm{j} \hat{f}_n t)$. Then, low-pass filtering is applied and a slowly varying base-band signal is attained. In the band-pass branch (bottom branch in Figure 1), a band-pass filter with a center frequency of $\hat{f}_n$ is applied. Both the base-band and the band-pass filtered signals are segmented into possibly overlapping blocks. Then, in each block i, the frequency $f_{n,i} = \hat{f}_n + \Delta f_{n,i}$ and the temporal growth rate $\alpha_{n,i}$ of the signal are assumed to be constant and the parameters $\Delta f_{n,i}$ and $\alpha_{n,i}$ are determined by means of a least-squares fit procedure, which is discussed in detail in the sequel. Once these two parameters are attained, the output of the band-pass filter is utilized for fitting the amplitude $P_{n,i}$ and phase shift $\phi_{n,i}$. Performing the fits on successive blocks of the signal allows for the reconstruction (re-synthesis) of the original signal from the partials.

By means of the modulation by $\exp(-2\pi \mathrm{j} \hat{f}_n t)$, a complex valued signal is attained, whose spectrum is shifted by $-\hat{f}_n$ on the frequency axis due to the modulation property of the Fourier transform:

$$p_{\mathrm{mod},n} = p(t) \exp(-2\pi \mathrm{j} \hat{f}_n t). \tag{2}$$

The slowly varying frequency $\tilde{f}_n$ will be shifted close to $f \approx 0$ if the estimate $\hat{f}_n$ approximates f_n well.

Applying the low-pass filter on the modulated signal filters out other partials, and a slowly varying complex exponential signal is found on the output of the filter:

$$p_{\mathrm{lpf},n}(t) = \tilde{P}_n(t) \exp\left(2\mathrm{j}\pi \Delta f_n(t) t + \tilde{\phi}_n(t)\right) + \nu_n(t), \tag{3}$$

where $\Delta f_n(t) = \tilde{f}_n(t) - \hat{f}_n$ is the slowly varying frequency difference and $\nu_n(t)$ is the residual signal of the n-th low-pass filter that contains the noise let through by the filter and residuals from other partials not being completely suppressed.

The effect of the complex modulation and low-pass filtering on a periodic signal is depicted in Figure 2. As an example, a signal with a frequency of $113.5\,\mathrm{Hz}$ was taken, and the modulation frequency was set to $105\,\mathrm{Hz}$. The low-pass filter is a fourth order Butterworth filter with a cutoff frequency of $25\,\mathrm{Hz}$. The

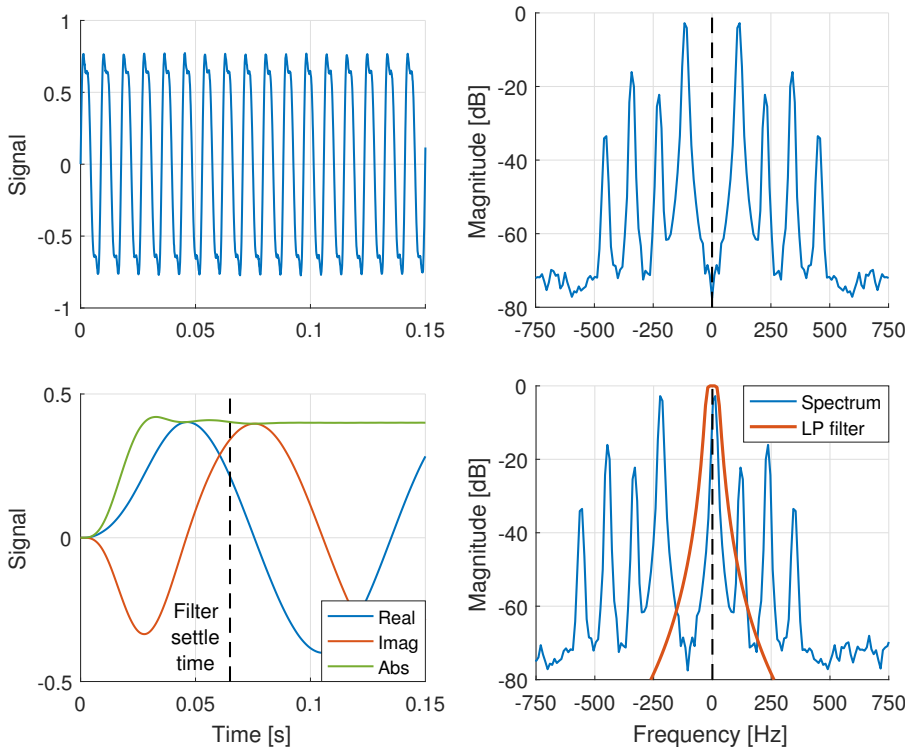


Figure 2: Illustration of complex modulation and filtering on a periodic signal. Left column: time domain signals. Right column: spectra. Top row: original signal. Bottom row: modulated signal and the effect of low-pass filtering.

diagrams in the first row of Figure 2 show the original signal in the time and frequency domains. The shift of the spectrum and the effect of the low-pass filter is illustrated in the frequency domain in the bottom right diagram. The effect in the time domain is visible in the bottom left, where the settle time of the filter is also marked. It should be noted that the filter shifts the phase of the signal. However, in an offline application this shift is avoided by applying the filter twice: first on the original signal and for the second time on the time-reversed signal. Thus, the square of the magnitude characteristics and a zero phase shift is attained. Such a filtering procedure is available in the `filtfilt` function of Matlab,[4] for example.

If a short block of the signal is taken, it can be assumed that the frequency difference Δf_n and the phase shift ϕ_n are constant in the block. Furthermore, it can also be supposed that the amplitude $\tilde{P}_n$ varies exponentially in the block. With these assumptions, the ratio of two subsequent samples of the output of the low-pass filter is written omitting the effect of the noise as:

$$\gamma_n[k] = \frac{p_{\mathrm{lpf,n}}[k+1]}{p_{\mathrm{lpf,n}}[k]} \approx \exp(\alpha_n T_\mathrm{s}) \exp(2\mathrm{j}\pi\Delta f_n T_\mathrm{s}) = \gamma_n, \tag{4}$$

where α_n is the temporal growth rate of the n-th partial, and γ_n is independent of k because of assuming a constant frequency and temporal growth rate within one block.

The value of γ_n is found by solving the overdetermined set of equations, written for the pairs of successive samples in a signal block consisting of L samples, by means of a least-squares fit:

$$\begin{bmatrix} p_{\mathrm{lpf},n}[0] \\ p_{\mathrm{lpf},n}[1] \\ \vdots \\ p_{\mathrm{lpf,n}}[L-1] \end{bmatrix} \gamma_n = \begin{bmatrix} p_{\mathrm{lpf},n}[1] \\ p_{\mathrm{lpf},n}[2] \\ \vdots \\ p_{\mathrm{lpf,n}}[L] \end{bmatrix}. \tag{5}$$

Evaluating the relative error of the fit by substituting the resulting γ_n back into (5) gives a useful indicator on the goodness of the fit and the correctness of assuming a single complex exponential in the filtered signal block. Repeating the same procedure for each block i, gives the block-wise estimates $\gamma_{n,i}$, from which the frequency differences $\Delta f_{n,i}$ and the temporal growth rates $\alpha_{n,i}$ are found as

$$\Delta f_{n,i} = \frac{\angle\gamma_{n,i}}{2\pi T_\mathrm{s}}, \qquad \alpha_{n,i} = \frac{\log|\gamma_{n,i}|}{T_\mathrm{s}}. \tag{6}$$

While the base-band signal already contains the phase shift and the amplitude of the n-th partial, we propose to evaluate the amplitude and the phase shift from the band-filtered signal. This approach provides somewhat more flexibility, as a filter with different characteristics can be chosen in the low-pass and band-pass branches. However, the advantages of using different filters is yet to be explored. The output of the band-pass filter is assumed to take the form for the n-th partial in the i-th block

$$p_{n,i}(t) = A_{n,i}\exp(\alpha_{n,i}t)\cos\left[2\pi(\hat{f}_n + \Delta f_{n,i})t\right] + B_{n,i}\exp(\alpha_{n,i}t)\sin\left[2\pi(\hat{f}_n + \Delta f_{n,i})t\right]. \tag{7}$$

The coefficients $A_{n,i}$ and $B_{n,i}$ are found by applying a least-squares fit anew. The amplitude $P_{n,i} = \sqrt{A_{n,i}^2 + B_{n,i}^2}$ and phase $\phi_{n,i} = \tan^{-1}(B_{n,i}/A_{n,i})$ are also found immediately.

C. EXTENDED COMPLEX KALMAN FILTER

As an alternative approach, a Kalman filter-based frequency tracking method was also examined. The Kalman filter is a state estimator that predicts the state vector $\mathbf{x}$ based on the vector of observations $\mathbf{y}$ and the presumed dynamic behavior of the system. While the Kalman filter is optimal for linear systems, its extensions (such as the extended and unscented Kalman filters, or particle filters) make it applicable in case of nonlinear systems as well. The discrete-time equations for a nonlinear dynamic system are written as

$$\mathbf{x}[k+1] = \boldsymbol{f}(\mathbf{x}[k]) + \mathbf{w}[k] \tag{8}$$

$$\mathbf{y}[k] = \boldsymbol{h}(\mathbf{x}[k]) + \mathbf{v}[k] \tag{9}$$

The operation denoted by $\boldsymbol{f}$ produces the next state from the current one, while $\boldsymbol{h}$ gives the observation $\mathbf{y}$ from the state vector. The vectors $\mathbf{w}$ and $\mathbf{v}$ are random samples representing the additive noise. It is noted that it was assumed here that the system has no inputs. The two equations (8) and (9) are often referred to as the state equation and the observation equation, respectively.

The block scheme of the Kalman filter is depicted in Figure 3. Two kinds of superscripts are introduced in the block diagram: the superscript $-$ refers to an estimated value, while values with a superscript $+$ denote variables that are already updated based on the actual observations.

In a recent paper by Das *et al.*,[5] an extended complex Kalman filter (ECKF) was used for tracking the fundamental frequency of a musical signal. Here, the latter approach was implemented and utilized to be able to compare the results of the proposed method to an alternative strategy. For the ECKF, we suppose that the observations are samples of a quasi-harmonic signal:

$$y[k] = \tilde{a}\cos(\tilde{\omega}kT_\mathrm{s} + \tilde{\phi}) + v[k], \tag{10}$$

with the slowly varying amplitude $\tilde{a}$, angular frequency $\tilde{\omega}$, and phase $\tilde{\phi}$. The variance of the noise v is denoted by σ_v^2. Following the approach published by Dash *et al.*,[6] the state space vector $\mathbf{x}$ is chosen as

$$\mathbf{x}[k] = \begin{bmatrix}\beta[k]\\ u[k]\\ u[k]^*\end{bmatrix} = \begin{bmatrix}\exp(\mathrm{j}\omega T_\mathrm{s})\\ a\exp(\mathrm{j}\omega kT_\mathrm{s} + \mathrm{j}\phi)\\ a\exp(-\mathrm{j}\omega kT_\mathrm{s} - \mathrm{j}\phi)\end{bmatrix}, \tag{11}$$

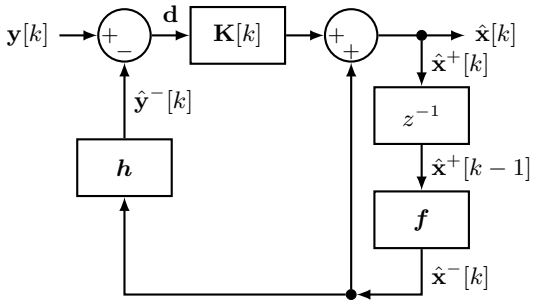


Figure 3: Block structure of the Kalman filter

where the superscript $*$ denotes complex conjugate.

This choice leads to the discretized system:

$$\mathbf{x}[k+1] = \boldsymbol{f}(\mathbf{x}[k]) = \mathbf{F}[k] \cdot \mathbf{x}[k] = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \beta[k] & 0 \\ 0 & 0 & \beta[k]^* \end{bmatrix} \cdot \begin{bmatrix} \beta[k] \\ u[k] \\ u[k]^* \end{bmatrix} \tag{12}$$

$$\mathbf{y}[k] = \boldsymbol{h}(\mathbf{x}[k]) = \mathbf{H} \cdot \mathbf{x}[k] = [0 \quad 0.5 \quad 0.5] \cdot \mathbf{x}[k] \tag{13}$$

The estimated amplitude $\hat{a}$, frequency $\hat{f}$, and phase $\hat{\phi}$ are found at each time step k as:

$$\begin{aligned} \hat{a}[k] &= \sqrt{\hat{u}[k]\hat{u}[k]^*} \\ \hat{\phi}[k] &= \frac{1}{2j} \ln\left(\frac{\hat{u}[k]}{\hat{u}[k]^* \hat{\beta}[k]^{2k}}\right) \\ \hat{f}[k] &= \frac{\ln \hat{\beta}[k]}{2\pi \mathrm{j} T_\mathrm{s}} \end{aligned} \tag{14}$$

The Kalman filter estimates the state vector $\hat{\mathbf{x}}$ for each sample of the signal. First, a predicted state $\hat{\mathbf{x}}^-$ is generated using (12). Then, a prediction of the output vector $\hat{\mathbf{y}}^-$ is computed using (13). The estimated state $\hat{\mathbf{x}}^+$ is a result of updating the prediction $\hat{\mathbf{x}}^-$ using the innovation $\mathbf{d} = \mathbf{y} - \hat{\mathbf{y}}^-$, and the feedback matrix $\mathbf{K}$. The latter feedback matrix is computed making use of the estimated covariance matrix of the predicted state vector and the assumed variance of the observation vector $\mathbf{y}$. Finally, the estimated covariance matrix corresponding to the state vector $\mathbf{x}^+$ is also updated.

The prediction of the covariance matrix of $\mathbf{x}^-$ involves two terms. First, the Jacobian of $\partial \boldsymbol{f}/\partial \boldsymbol{x}$ is utilized for propagating the covariance of $\mathbf{x}^+$. Second, an additive term, called the covariance matrix of the process noise, denoted by $\mathbf{Q}$, is added to the estimated variance. Essentially, the Kalman filter performs a weighted averaging between the state inferred from the observation and the predicted state, where the weights are inversely proportional to the variances. Therefore, by increasing the assumed process noise $\mathbf{Q}$ the filter will follow the observation (i.e., the signal y in our case) more closely, which results in a faster settling of the filter, but comes at a cost of the filter becoming more sensitive to the measurement noise.

Das *et al.*[5] proposed an adaptive choice for the process noise: $\mathbf{Q}[k] = \mathbf{I}\sigma_w^2[k]$, where

$$\log_{10}\left(\sigma_w^2[k]\right) = -c + \left|y[k] - \hat{y}^-[k]\right|, \tag{15}$$

and c is a positive scalar. This approach is also adopted here, with a slight improvent of introducing a diagonal pre-factor for the process noise

$$\mathbf{Q}[k] = \mathrm{diag}\ \langle \mathbf{q}_d \rangle\, \sigma_w^2[k], \tag{16}$$

where $\mathbf{q}_d$ is a vector with three positive elements and $\mathrm{diag}\,\langle\cdot\rangle$ represents a diagonal matrix constructed from the elements inside the brackets. This modification is reasonable as the elements of the estimated state vector $\hat{\mathbf{x}}$ contain different parts of information on the three estimated parameters. The first element $\hat{\beta}$ represents only the estimated frequency, while the other two elements are related to the frequency, phase, and amplitude at once. For example, if the amplitude of the signal is expected to vary while its frequency is constant, the first element in $\mathbf{q}_d$ can be remarkably smaller than the other two elements.

It is worth mentioning that in the paper by Das *et al.*[5] the Kalman filter is supplemented by other features, including onset detection and note change detection; however, these options are not used here. In the following examples, when the proposed method and the ECKF are compared, the input of the Kalman filter is always a band-filtered signal in which only one partial of the sound signal is dominant.

3. ANALYSIS OF THE ATTACK TRANSIENTS OF A TENOR RECORDER

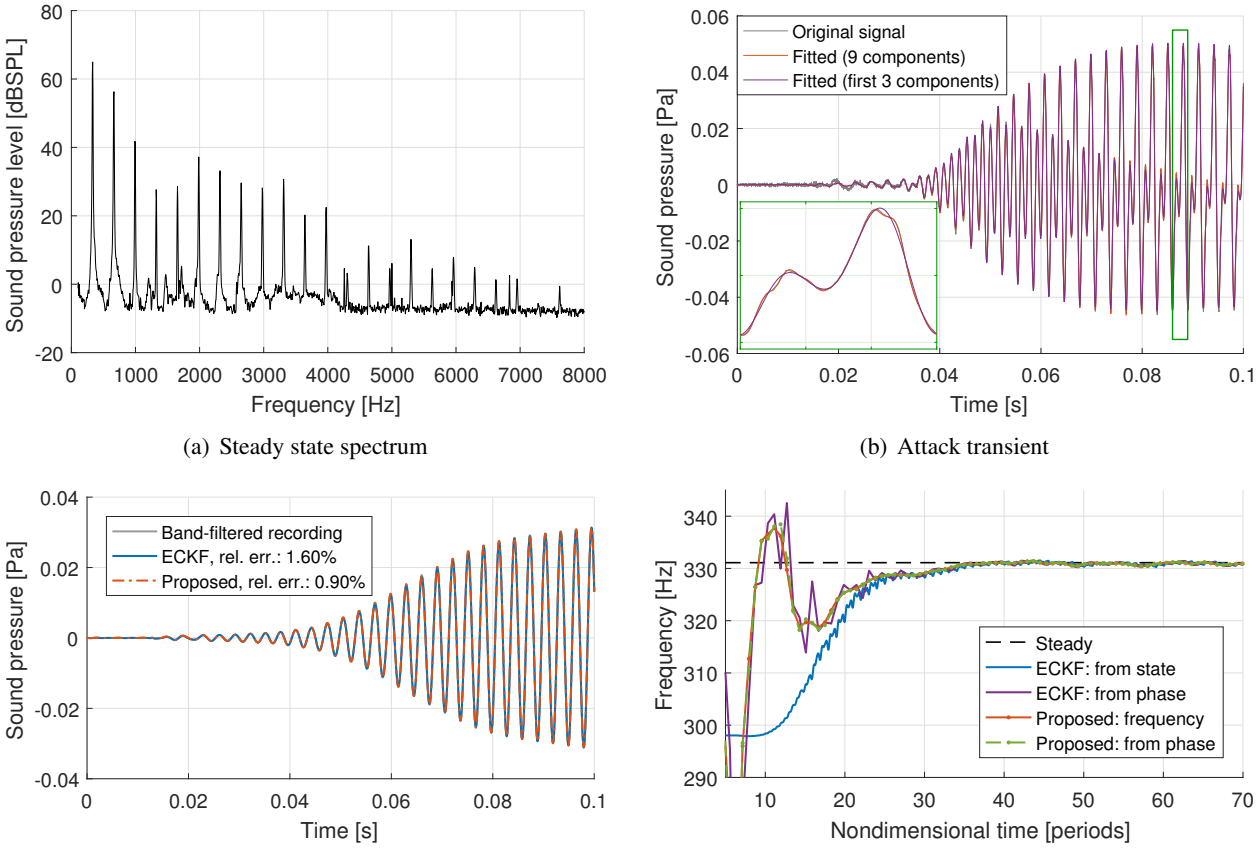
Sound recordings of a Yamaha YRT-304B II tenor recorder were made at the pipe organ laboratory of the Fraunhofer Institute for Building Physics in Stuttgart, Germany. The recorder was placed into the research pipe organ, on top of one of the windchests designed to suite experimental organ pipes. An airtight sealing between the mouthpiece of the recorder and the windchest of the pipe organ was provided. A valve installed in the windchest was applied for closing or opening the flow into the mouthpiece of the recorder. The pressure inside the windchest could be controlled by adjusting the frequency of the fan of the pipe organ's wind system in the range of 0 to $1\,250\,\mathrm{Pa}$. A Honeywell 163PC01D36 pressure sensor was utilized for measuring the pressure inside the windchest.

The sound radiated by the recorder was measured using a 1/2" GRAS 46AF condenser microphone with a 26TK type preamplifier. The microphone was located in front of the mouth of the recorder at a distance of $150 \pm 5\,\mathrm{mm}$. A Brüel & Kjær Nexus type 2690-OS2 conditioning amplifier was used for amplifying the microphone signal. Both the windchest pressure and amplified microphone signals were sampled and recorded synchronously by a Data Translation DT9836 16-bit data acquisition card running at a sampling frequency of $40\,\mathrm{kHz}$ ($T_\mathrm{s} = 25\,\mu\mathrm{s}$).

Various recordings of different fingerings were taken using the above measurement setup. In the transient recordings, the frequency of the fan and thus the windchest pressure was kept constant. The recorder was sounded three times with each sound lasting for ≈ 5 seconds with ≈ 3 seconds pause between them with opening and closing the valve in the windchest manually. By analyzing the signal of the pressure sensor located inside the windchest showed that the manual opening of the valve gave reproducible measurements. This pressure signal was also used as the trigger signal of the transient frequency analysis.

First, the analysis of the attack transient of the $\mathrm{E4}$ note (fundamental frequency $\approx 330\,\mathrm{Hz}$) is discussed. The windchest pressure was set to $120\,\mathrm{Pa}$, and the recorder was sounded three times. Figure 4 displays the results of the analysis of one of the three sounds. The steady state spectrum is depicted in Fig. 4(a). Naturally, harmonic partials dominate in the steady state spectrum, with the first three partials (fundamental, octave, and fifth) having the highest levels. Diagram 4(b) displays the reconstruction of the attack signal using the proposed method. The analysis parameters were set as follows. Both in the base-band and the band-pass branch Butterworth filters (orders 5 and 4, respectively) were used with a bandwidth of $b = \hat{f}_1/4$, where $\hat{f}_1 \approx 332\,\mathrm{Hz}$ resulting from the FFT analysis of the steady state part. The block size was 128 samples with an overlapping of 50%, which results in a temporal resolution of frequency and temporal growth rate values of $1.6\,\mathrm{ms}$. As it is visible, a very good reconstruction is already attained with using three components, while adding more partials further reduces the difference compared to the original signal. The development of the second partial is also worth observing here: in the early stage of the attack this partial grows very quickly but then its amplitude diminishes and the fundamental component becomes clearly dominant.

The tracking of the development of the fundamental component both by the proposed method and the ECKF is shown in Fig. 4(c). For this comparison, the input signal of the ECKF is chosen as the band-



(a) Steady state spectrum

(b) Attack transient

(c) Reconstruction of the fundamental component

(d) Transient frequency analysis

Figure 4: Analysis of the $\mathrm{E}4$ ***note of the tenor recorder. Blowing presure:*** $120\,\mathrm{Pa}$**.**

pass filtered signal corresponding to the fundamental component. As described in Section 2.3, the tracking behavior of the Kalman filter depends on the assumptions made on the process and measurement noises. Here, the coefficient in (15) was set to $c = 3$ and the diagonal pre-factor $\mathbf{q}_d = [1 \quad 10 \quad 10]$ was used. This choice led to a filter that follows the changes of the parameters of the input signal quite rapidly, while it allows the estimated frequency, phase, and amplitude to fluctuate a bit from sample to sample. In general, increasing the coefficient c leads to a slower tracking with smaller variance of the estimated state. As it is seen, both methods capture the temporal changes of this partial quite well. The relative L_2 errors of the two methods were also evaluated and the proposed method was found to give a slightly smaller error than the ECKF. The almost perfect reconstruction of the partials was also made audible by subtracting the re-synthesized signal from the original recording. In this case, only the background noise is heard first, and when the recorder starts to sound, an increased amount of flow noise is heard, which results from the microphone being located near to the mouth of the recorder.

Diagram 4(d) displays the change of the frequency of the fundamental component in the attack transient evaluated using both methods. Interestingly, while both approaches reconstruct the signal quite accurately, the predicted frequencies differ remarkably for the first 20–25 periods of the sound. (Compare the curves: “ECKF: from state” and “Proposed: frequency”.) The latter discrepancy can be explained by an inherent ambiguity that arises when the frequency and phase of the signal has to be estimated at the same time: the change of the running phase of the signal may be attributed to a change of its frequency as well as a variation of the phase shift (phase modulation). As the ECKF has to estimate both quantities $\hat{\phi}$ and $\hat{f}$ from a single

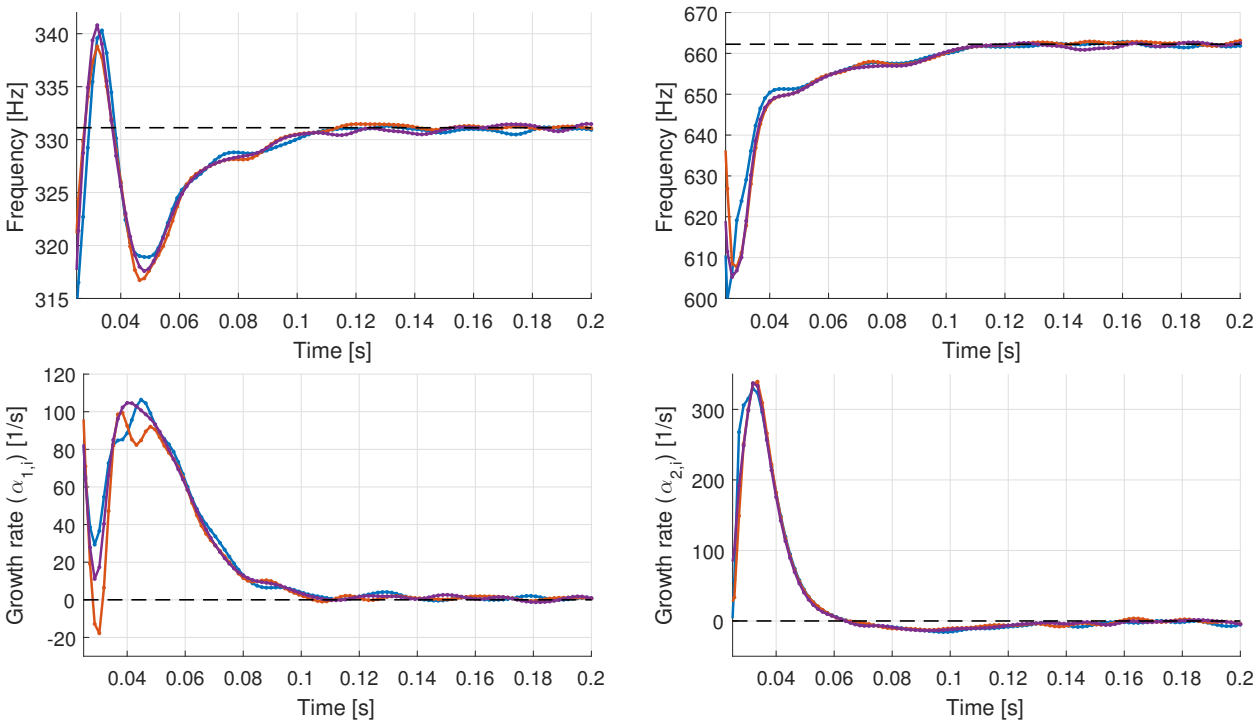
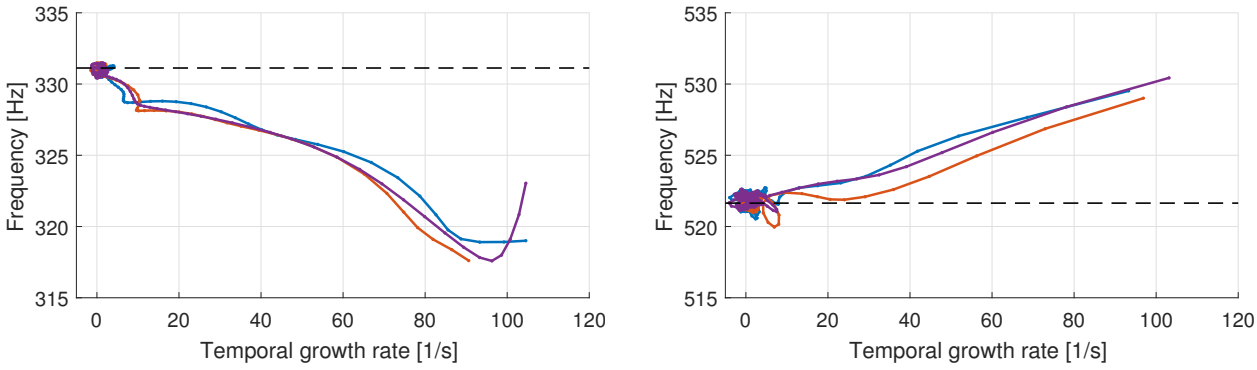


Figure 5: Frequency and temporal growth rate of the first two partials in the attack transient of the E4 ***note. The three colors denote three subsequent recordings with the same conditions. Top row: frequency. Bottom row: temporal growth rate. Left column: fundamental. Right column: second partial.***

sample, the deviation of the estimated and the observed sample will cause a change of both the phase and the frequency estimation. To highlight this ambiguity, two additional curves are also plotted in Fig. 4(d) ("ECKF: from phase" and "Proposed:from phase"), where the frequency of the signal is calculated from the change of the running phase in one signal block. As it is visible, when the frequency is extracted from the running phase, the frequency estimation of the ECKF is much closer to that of the proposed method, yet the variance of the estimation is greater. It is also worth noting that the initial frequency estimation of the ECKF also stems from the spectral analysis of one block of the signal (256 samples in our case). As in the attack transient, the frequency of the signal varies, the initial frequency guess is sensitive to the timing of the first analysis window. For the proposed method, the phase–frequency ambiguity is mitigated by approximating both quantities from a block of samples instead of a single observation. When the frequency is extracted from the phase shifts of subsequent time windows, the result matches almost perfectly with the estimated frequency (see the curves "Proposed: frequency" and "Proposed: from phase").

Next, the frequency and the temporal growth rate of the first two partials (fundamental and octave) of the same E4 note are analyzed in more detail using the proposed method. The results are depicted in Figure 5 with the three curves in each of the diagrams showing the analysis of the three subsequent sounds taken using the same setup. While it is a known feature of wind instruments that subsequent attack transients exhibit some variations,[7] the results of the evaluation were found to be well reproducible.

Looking at the frequencies of the first two partials, it is visible that both vary significantly in the first few tens of milliseconds of the attack. It is also interesting to observe that the two partials are not harmonic in the early stage of the attack ($t < 0.06\,\mathrm{s}$ in the figure). This observation is explained by the first two, slightly inharmonic acoustical modes of the resonator being simultaneously excited by the air jet in the early stage of the attack. After an initial "overshoot", the frequency of the fundamental component converges



(a) E4, 120 Pa (b) C5, 155 Pa

Figure 6: Frequency of the fundamental component as a function of the temporal growth rate during the attack.

to its steady state value from below, while in case of the second partial no overshoot was observed. By comparing the temporal growth rates, the quicker development of the second partial is clearly observed, its maximal temporal growth rate is about three times as high as that of the fundamental. It is also visible that α_2 becomes negative for $0.65 < t < 0.15$, which is in correspondence with the results of Fig. 4(b), where the second partial visibly diminishes after its initial growth.

It is instructive to display the frequency of the fundamental component as a function of the temporal growth rate. Figure 6 compares the notes E4 and C5. Here, the three colors also denote three repetitions of the same note. As it is visible, two different behaviors are found: for E4 the frequency gradually increases as the temporal growth rate decreases during the attack stage, while for the C5 note, the frequency of the fundamental decreases with decreasing the temporal growth. Notes of the diatonic scale from C4 to G5 were analyzed, and it was found that notes clearly form two groups regarding this behavior. Notes from C4 to G4 behave similar to E4, and notes from A4 to G5 exhibit a behavior similar to that of C5. In the next section an attempt is made to explain this interesting observation at least qualitatively by a simplified physical model of recorder-like instruments.

4. INTERPRETATION BY A TRANSIENT JET-DRIVE MODEL

A simplified model of sound generation in a recorder-like instrument is based on the McIntyre – Woodhouse – Schumacher model,[8] that represents the excitation mechanism as a nonlinear system and the acoustical resonator as a linear one. There are different possibilities for modeling the excitation of the resonator by an oscillating air jet, jet-drive and discrete vortex models[9] being two of the alternatives. Here, a state-of-the-art lumped jet-drive model[10] is considered.

The jet-drive model, depicted schematically in Figure 7 assumes that the jet emerging from the windway of the instrument takes up a Bickley profile while it propagates to the upper lip (labium). The jet is sensitive to cross-stream perturbations of the velocity field, and its most sensitive region is located near the exit of the windway. The acoustical particle velocity which results from the resonator reflecting the pressure waves generated by the jet,[11] causes a perturbation of the velocity field of the jet. The perturbation propagates with an effective velocity of $\approx 0.4u_{\mathrm{jet}}$, with u_{jet} denoting the maximum velocity of the jet at the windway exit. At the same time, the perturbation waves are amplified by the jet, with the displacement becoming η at the edge. The time delay of the perturbation from the windway exit to the labium is denoted by τ. The oscillation of the air jet results in an oscillating volume flow in and out of the resonator. Assuming incompressibility in the small region of the mouth opening results finally in a dipole pressure source Δp driving the resonator.

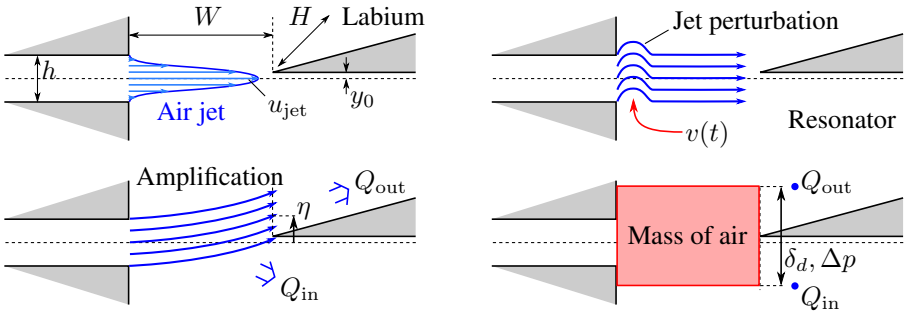


Figure 7: Illustration of the lumped jet-drive model.

The resonator is characterized by its frequency dependent input admittance $Y_{\mathrm{in}}(\omega)$, which is the ratio of the particle velocity response v and the driving pressure Δp. The input admittance of the tenor recorder was computed in the frequency domain using a one-dimensional acoustical model that takes closed and open toneholes, as well as wall and radiation losses into account. Following Terrien *et al.*,[10] the input admittance is decomposed into M modes, each mode represented by its eigen angular frequency ω_m, damping factor ξ_m, and magnitude F_m ($m = 1, 2, \ldots M$):

$$Y_{\mathrm{in}}(\omega) \approx \frac{a_0}{\mathrm{j}\omega b_0 + c_0} + \sum_{m=1}^{M} \frac{\mathrm{j}\omega F_m}{\omega_m^2 + 2\mathrm{j}\omega\omega_m\xi_m - \omega^2}. \tag{17}$$

In a recent paper by the authors,[12] the jet–drive model was extended into the complex frequency domain, allowing for the examination of transient effects, where the magnitude of the signals is allowed to vary exponentially in time. The Laplace variable is introduced as $s = \alpha + \mathrm{j}\omega$ and a time dependence of $\exp(st)$ is assumed. Both the stability analysis of the jet and the model of the resonator was extended into the complex frequency domain, and different conditions for self-sustained oscillations were investigated. Here, for the sake of simplicity, only the phase lock condition is examined, written for the n-th aerodynamic mode of the air jet as:

$$-2n\pi = \angle Y_{\mathrm{in}}(f) - 2\pi f\tau(f) - \frac{3\pi}{2} \qquad n = 1, 2, \ldots \tag{18}$$

The phase lock condition states that for self-sustained oscillations, the phase delay of the resonator and that of the perturbations traveling with the jet should give exactly n whole cycles. The constant of $-3\pi/2$ incorporates that v points outward of the resonator and contains a conversion from displacement to velocity.

The phase lock condition (18) is solved for the frequency f at different blowing pressures. The graphical representation of the solution is illustrated in Figure 8. In the left diagram the poles of the resonator are shown over the complex plane. As the resonator is a passive system, all its poles are located on the left half-plane. The three dashed-lines illustrate three different temporal growth rates that are also shown by the same colors on the right hand side of the figure. As it is visible, as the temporal growth rate increases, the frequency axis moves away from the poles resulting in less steep peaks of the magnitude diagram and the phase jumps near the poles becoming less abrupt.

The graphical solution of the phase lock condition is shown for two different blowing pressures on the bottom right phase diagram of Figure 8. For the sake of simplicity, the frequency and temporal growth rate dependence of the phase velocity of the perturbations were not taken into account here. As it is visible, including the temporal growth in the model results in a change of frequency. The markers ▲ / ▼ denote the two cases where the oscillation frequency increases / decreases with increasing the temporal growth rate. The two inlays on the diagram zoom on the intersection points for the two cases.

While the lumped jet–drive model operates with many simplifications, and its extension into the complex frequency domain is non-trivial for the jet oscillations, it still can qualitatively explain the observed changes

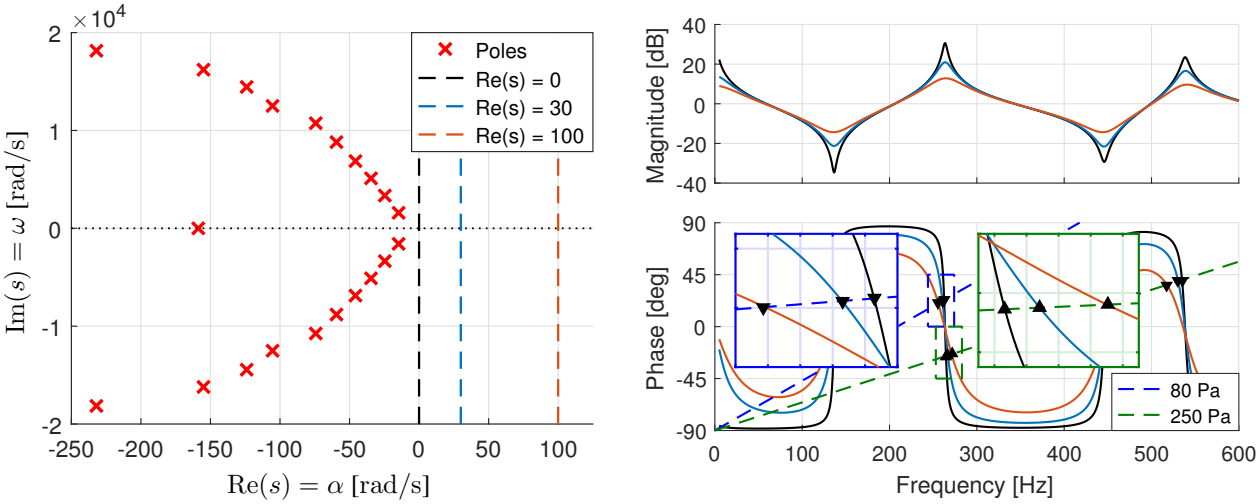


Figure 8: Phase lock condition in the complex frequency domain. Left: poles of the input admittance. Right: Bode diagram with different temporal growth rates $\alpha = \mathrm{Re}(s)$***. The bottom right diagram illustrates the change of frequencies resulting from the change of the temporal growth rate for two different blowing pressures.***

of the fundamental frequency during the attack transient. A more detailed and quantitative analysis requires further investigations. The experimental validation of the complex frequency model of the jet is also a challenging task.

5. SUMMARY AND OUTLOOK

This paper introduced a method for the detailed analysis of the frequencies and amplitudes of partials in transient musical signals. Relying on the assumption that the signal is composed of partials whose frequency, amplitude, and phase vary slowly in time, a block-processing approach was proposed, which uses complex modulation and least squares fits to estimate the time-varying parameters of the signals.

The proposed method was used for the analysis of the recordings made on a tenor recorder with different fingerings. The attack transients of the sounds were analyzed, and the results of the proposed signal processing method were compared to an alternative strategy that uses the extended complex Kalman filter. Both methods were found to be capable of tracking the changes of the amplitude and estimate the frequency of the signal. For the ECKF, the inherent ambiguity of the phase and frequency estimation from a single observation renders the transient frequency tracking challenging.

Very interestingly, the frequency analysis of the recorder sounds revealed that notes clearly form two groups: from C4 to G4 the frequency of the fundamental component gradually increases during the attack (the frequency increases as the temporal growth rate decreases), while for A4 to G5, the opposite behavior was observed. The explanation of these unexpected results were sought for using a lumped jet-drive model of the recorder, which has been extended into the complex frequency domain. It was found that the latter model is capable of at least qualitatively explaining the changes of the frequency during the attack transient.

The proposed method seems to be promising for analyzing transients of musical signals. The authors plan to apply the approach on sound samples of various wind and string instruments. Different options regarding the filter choices of the base-band and band-pass stages are also to be explored.

ACKNOWLEDGMENTS

This work has been supported by the Hungarian National Research, Development and Innovation Office under contract No. K 143436.

REFERENCES

[1] M. Castellengo. Acoustical analysis of initial transients of flute-like instruments. *Acustica-Acta Acustica*, 85:387–400, 1999.

[2] A. Miklós and J. Angster. Properties of the sound of flue organ pipes. *Acustica-Acta Acustica*, 86(4):611–622, 2000.

[3] A. Ernoult and B. Fabre. Temporal characterization of experimental recorder attack transients. *Journal of the Acoustical Society of America*, 141(1):383–394, 2017.

[4] Mathworks Inc. `filtfilt`: Zero-phase digital filtering. https://mathworks.com/help/signal/ref/filtfilt.html Accessed: 10/31/2022.

[5] O. Das, J. O. Smith, and C. Chafe. Improved real-time monophonic pitch tracking with the extended complex Kalman filter. *Journal of the Audio Engineering Society*, 68(1/2):78–86, 2020.

[6] P. K. Dash, R. K. Jena, G. Panda, and A. Routray. An extended complex Kalman filter for frequency measurement of distorted signals. *IEEE Transactions on Instrumentation and Measurement*, 49(4):746–753, 2000.

[7] J. Angster, P. Rucz, and A. Miklós. Acoustics of organ pipes and future trends in the research. *Acoustics Today*, 13(1):12–20, 2017.

[8] M. E. McIntyre, R. T. Schumacher, and J. Woodhouse. On the oscillations of musical instruments. *Journal of the Acoustical Society of America*, 74(5):1325–1345, 1983.

[9] R. Auvray, A. Ernoult, B. Fabre, and P.-Y. Lagrée. Time-domain simulation of flute-like instruments: Comparison of jet-drive and discrete-vortex models. *Journal of the Acoustical Society of America*, 136(1):389–400, 2014.

[10] S. Terrien, C. Vergez, B. Fabre, and D. Barton. Calculation of the steady-state osciallations of a flute model using the orthogonal collocation method. *Acustica-Acta Acustica*, 100:690–704, 2014.

[11] P. T. Nagy, A. Szabó, and Gy. Paál. A feedback model of the edge tone, using the adjoint Orr-Sommerfeld equation. *Journal of Fluid Mechanics*, 915:A13, 2021.

[12] P. T. Nagy, P. Rucz, and A. Szabó. Examination of jet growth and jet-drive in the recorder by means of linearized numerical and lumped models. *Journal of Sound and Vibration*, 527:116857, 2022.