A NEW ESTIMATION TECHNIQUE FOR DETERMINING THE CONTROL PARAMETERS OF A PHYSICAL MODEL OF A TRUMPET

Wim D'haes [a,b], *Xavier Rodet* [a]

[a] IRCAM – Centre George Pompidou – 1, place Igor-Stravinsky · 75004 Paris · France
[b] Visionlab – University of Antwerp (UA) – Groenenborgerlaan 171 · 2020 Antwerp · Belgium

ABSTRACT

A new estimation technique is proposed which computes the control parameters of a physical model of a trumpet in order to simulate a recording of a real instrument. First, the physical constraints of the instrument and the prior knowledge about how a player controls a trumpet are described. This is taken into account during the design of the data set and guarantees that these constraints are respected. Then, an estimation procedure minimizes two perceptual similarity criteria in function of the control parameters. The first criterium expresses the difference of the spectral envelopes and the second one the difference in fundamental frequency. An optimization technique is proposed that yields an optimal solution for the fundamental frequency, and a conditional suboptimal solution for the spectral envelope. A robust implementation of the technique was developed for which it is shown that the estimated parameters are unique and that the optimization does not suffer from local minima.

1. PHYSICAL CONSTRAINTS AND PRIOR KNOWLEDGE

In the dissertation of Christophe Vergez a physical model is developed that computes the sound produced by a trumpet from time varying control parameters of a player [14]. Recent research focusses on the automatic determination of control parameters in order to simulate a recorded sound [5, 8]. Also for other physical models, control parameter estimation is actively researched [9, 12, 13].

The control parameters $\bar{P}$ of this physical model, consist of the pressure in the mouth P_M, the lip frequency P_L, the tube length P_T and damping factor of the lips P_D. In previous work, the conditions were derived for which the maximal resonance for this non linear system with delayed feedback was obtained [6]. There, it was shown that the resonance frequency f_τ of the tube controlled with a the parameter P_T could be computed using

$$f_\tau = \frac{F_s}{\lfloor \frac{F_s}{2P_T} \rfloor + \lambda_0} \tag{1}$$

with F_s being the sampling frequency and λ_0 a constant. By keeping all parameters constant and varying the the lip frequency it was observed that a maximal resonance was obtained at multiples of f_τ with a value of P_L being three fourth of the frequency yielding

$$P_L = \frac{3}{4} N f_\tau \tag{2}$$

where N is an integer value expressing the mode index. Equations (1) and (2) express the relationship between P_L and P_T for which the resonance is maximal.

An important physical constraint is that the tube length of the instrument must remain constant for each note. In addition, only seven different tube lengths can be used for an entire trumpet performance. Seven tube lengths were determined such that the optimal resonance was achieved for notes tuned to a frequency f_{ref} of 440 Hz. This constraint was not respected by the instance-based approach in previous work [5].

For each note, the player chooses a combination of tube length and mode in order to obtain the desired frequency. The choice of this combination is the *prior knowledge* that a player uses when he plays the instrument and is modelled by the series $P_{T,i}$ and N_i. When the fundamental frequency f is expressed in half tones using

$$I = 12 \log_2 \left(\frac{f}{f_{ref}} \right) + 16 \tag{3}$$

then index of the series i corresponds with the rounded value of I. The value 16 is added to make the index of the lowest trumpet note (low $F\sharp$) correspond to 1.

Evidently, the player does not always excite the tube at the exact frequency for which the resonance is maximal. For instance when vibrato is played, the lip frequency varies periodically. To express this deviation we introduce a real valued parameter $\Delta N \in [-0.5, 0.5]$ that expresses this deviation. In order to play a given note with index i when using a deviation ΔN, the lip frequency is computed by

$$P_L = \frac{3}{4} f_{\tau,i}(N_i + \Delta N) \tag{4}$$

where $f_{\tau,i}$ is computed from the control parameter $P_{T,i}$ using Eq. (1).

2. FORMALIZATION OF THE ESTIMATION TECHNIQUE

2.1. Distance Metrics

The perceived distance between two short time spectra is defined by two components. To express the perceptual similarity in timbre, the difference between the log spectral envelopes was used. This envelope was expressed in terms of *Mel Frequency Discrete Cepstrum Coefficients (MFDCC)* [2, 3, 7, 11]. In this work a stabilized version was used, computed with *posterior warping* and a *lower bound threshold* [4]. An elegant property of the discrete MFDCC's is that the log difference between the Mel scale spectral envelopes is equivalent with the Euclidean distance between the cepstrum coefficients. In other words, when two spectral envelopes are considered, defined by two cepstrum vectors $\bar{c}_1$ and $\bar{c}_2$ respectively, the spectral similarity D_1 is given by

$$D_1(\bar{c}_1, \bar{c}_2) = (\bar{c}_1 - \bar{c}_2)^T(\bar{c}_1 - \bar{c}_2) \tag{5}$$

The square difference of the log fundamental frequency yields the second distance metric.

$$D_2(f_1, f_2) = (\log(f_1) - \log(f_2))^2 \tag{6}$$

One can imagine to use a weighted combination of D_1 and D_2, but since the physical meaning of such a combined distance metric is questionable and it is not known how these weights should be determined, we choose to keep the criteria separated.

2.2. Data Set Design and Feature Extraction

The data set was designed by using a fixed set of seven tube lengths that were optimized for a given tuning frequency of 440 Hz (see [6]). This automatically imposes the physical constraints of the acoustic instrument. For every note, and for a range of values of ΔN from -0.06 to 0.06 in steps of 0.01, crescendos were synthesized by varying the pressure P_M slowly from the 0 Pa to 30000 Pa. This data set design guarantees that all the intensities for each note are available and that a variation in fundamental frequency can be realized for the synthesis of vibrato. These are all the elements that are needed to simulate an expressive trumpet performance.
After an additive analysis of the synthesized sounds [10], the discrete MFDCC's and fundamental frequencies were computed. The extraction of the discrete MFDCC's is represented formally by a 2 dimensional function of the control parameters for a given note i

$$\mathcal{C}^i(\Delta N, P_M) \tag{7}$$

and the estimation of the fundamental frequency as

$$\mathcal{F}^i(\Delta N, P_M) \tag{8}$$

2.3. Estimation

Given a vector of cepstrum coefficients $\bar{c}$ and a fundamental frequency f computed from a short time window of a recorded sound, the goal of the estimation consists of determining the values of ΔN and P_M which minimize

$$D_1\left(\bar{c}, \mathcal{C}^i(\Delta N, P_M)\right) \tag{9}$$

and

$$D_2\left(f, \mathcal{F}^i(\Delta N, P_M)\right) \tag{10}$$

Many parameter optimization techniques exist [1], but this case is quite particular since there are two criteria that need to be optimized.
The first step consists of classifying the fundamental frequency to the best note index i. This index is computed by taking the rounded value of I obtained from Eq. 3 and identifies the data that will be used for a given note. It follows directly from the inner working of the physical model that the pressure in the mouth P_M has the largest influence on the spectral envelope while ΔN has a predominant influence on fundamental frequency. Therefore, we propose an estimation procedure that consists of two steps. In each step one distance metric is optimized. First, the mouth pressure is optimized for each ΔN with respect to D_1. This results in a function $\mathcal{P}^i(\Delta N; \bar{c})$ yielding the mouth pressure P_M in function of ΔN for which D_1 is minimized given $\bar{c}$ and i. Note that this yields a *conditional optimum*, since it yields the best value of P_M for a given ΔN.

$$\mathcal{P}^i(\Delta N; \bar{c}) = \arg\min_{P_M} D_1\left(\bar{c}, \mathcal{C}^i(\Delta N, P_M)\right) \tag{11}$$

Inserting Eq. (11) in Eq. (10), the distance criterium D_2 only depends on ΔN for a given f and $\bar{c}$. When the value ΔN^* is determined which minimizes the second distance metric D_2 for a given f, the estimation procedure is completed.

$$\Delta N^* = \arg\min_{\Delta N} D_2\left(f, \mathcal{F}^i(\Delta N, \mathcal{P}^i(\Delta N; \bar{c}))\right) \tag{12}$$

Still, the control parameter values need to be computed from ΔN^* and i using

$$P_M^* = \mathcal{P}^i(\Delta N^*; \bar{c}) \tag{13}$$

$$P_L^* = \frac{3}{4} f_{\tau,i}(N_i + \Delta N^*) \tag{14}$$

$$P_T^* = P_{T,i} \tag{15}$$

It is important to note, that the optimization with respect to D_2 is optimal for a given f. By contrast, the value P_M^* only optimizes D_1 in a suboptimal way for $\bar{c}$. This is justified by the fact that small differences in spectral envelope, or timbre, are less disturbing than deviations of the fundamental frequency.

3. IMPLEMENTATION

3.1. Estimation Example

In the previous section, the estimation procedure was described in terms of continuous functions. Since no parametric or analytic forms of these functions are available they are practically realized by piecewise linear functions. Instead of repeating the entire derivation for these discrete sampled functions, a practical example is described for a given $\bar{c}$ and f. In this example, f has a value of 786,65 Hz for which Eq. (3) yields a value of 26,0588. This implies that $i = 26$, meaning that the 26th data set will be used (see Eq. (7) and (8)). This data set corresponds with a high G which has been produced by exciting the sixth mode of the tube with a length corresponding with the fingering where all valves are released. This is how the prior knowledge is taken into account and the physical constraints are imposed.
Fig.1 shows a plot of $D_1\left(\bar{c}, \mathcal{C}^{26}(\Delta N, P_M)\right)$ in function of P_M for different values of ΔN. One can observe that for each ΔN a global optimum is available but that the error function is quite noisy. In order to make an accurate and robust estimate of the minimum, D_1 is modelled locally by a quadratic approximation for each ΔN that is fit to the observed values by a least mean squares procedure. This results in

$$D_1\left(\bar{c}, \mathcal{C}^i(\Delta N, P_M)\right) \simeq a(\Delta N)P_M^2 + b(\Delta N)P_M + c(\Delta N) \tag{16}$$

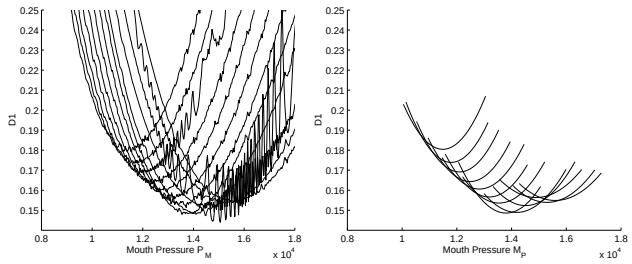

Figure 1: Spectral similarity in function of the mouth pressure for different ΔN values. Left, raw data. Right, local quadratic approximation.

In order to realize $\mathcal{P}^i(\Delta N; \bar{c})$ as defined in Eq.(11), the minimum of the fit is taken for each ΔN value yielding.

$$\mathcal{P}^i(\Delta N; \bar{c}) = \frac{-b(\Delta N)}{2a(\Delta N)} \tag{17}$$

The piecewise linear realization of this function is given on the left side of in Fig. 2. Now, the corresponding values of the fundamental frequencies can be retrieved from the data set which was expressed in the previous section by

$$\mathcal{F}^i(\Delta N, \mathcal{P}^i(\Delta N; \bar{c})) \tag{18}$$

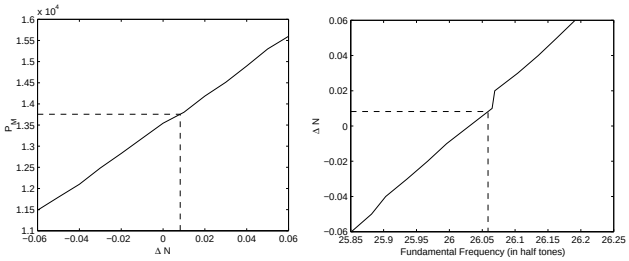

Figure 2: Left, piecewise linear function denoting $\mathcal{P}^i(\Delta N; \bar{c})$. Right, inverted piecewise linear function of $\mathcal{F}^i(\Delta N, \mathcal{P}^i(\Delta N; \bar{c}))$

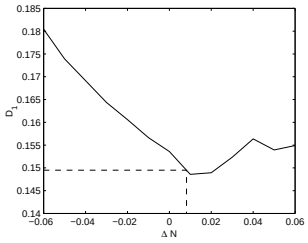

Figure 3: Suboptimal value for D_1 in function of ΔN

In Fig. 2, the inverse of this function was plot, since we wish to determine ΔN from the given f. This was expressed by Eq. (12) and is realized by evaluating the inverse piecewise linear function. The result is depicted by the dashed line in the figure. In this example, the value of f expressed in half tones was $26,0588$, and yielded a value of $\Delta N^* = 0.0082$. When $\mathcal{P}^i(\Delta N; \bar{c})$ is evaluated a value of P_M was obtained being 13756. Fig. 3 shows that ΔN^* yields a suboptimal value with respect to the spectral envelope similarity D_1.

3.2. Conclusion

It is shown that for each ΔN a global optimum can be found that can be determined in a robust manner using a local quadratic approximation [1]. This motivates the function $\mathcal{P}$ that expresses the optimal P_M in function of ΔN with respect to D_1. The function $\mathcal{F}^i(\Delta N, \mathcal{P}^i(\Delta N; \bar{c}))$ modelled by a piecewise linear function was observed to be increasing monotonously. Evidently, its inverse function is also increasing monotonous and therefore a unique solution of ΔN^* is obtained for a given f. Also, the function $\mathcal{P}$ returns a single value of P_M^*. This implies that the obtained solution is unique and that the optimization technique does not suffer from local minima.
In the example, it is shown that an exact solution with respect to D_2 was obtained for f using the inverse piecewise linear function. By contrast, the retrieved value of ΔN did not globally optimize D_1, and only a suboptimal solution

was obtained. We name this the *conditional optimum* with respect to D_1, since it yields the optimal value of P_M, *given the condition* that D_2 is optimized first for a given f. The motivation of this optimization is the fact that the accuracy of the fundamental frequency has a higher priority than the optimization of the spectral envelope.

4. RESULTS

In Fig. 4, the results are shown for a musical trumpet phrase. The phrase contains long notes with vibrato, slurred notes and attacked notes which were all simulated successfully. The top figure shows the original signal. The other figures show the estimated control parameters. From these figures, one observes how the mouth pressure follows the amplitude envelopes of the sounds while the lip frequency follows the melodic line of the excerpt including the vibrato at the end of the long sustained notes.

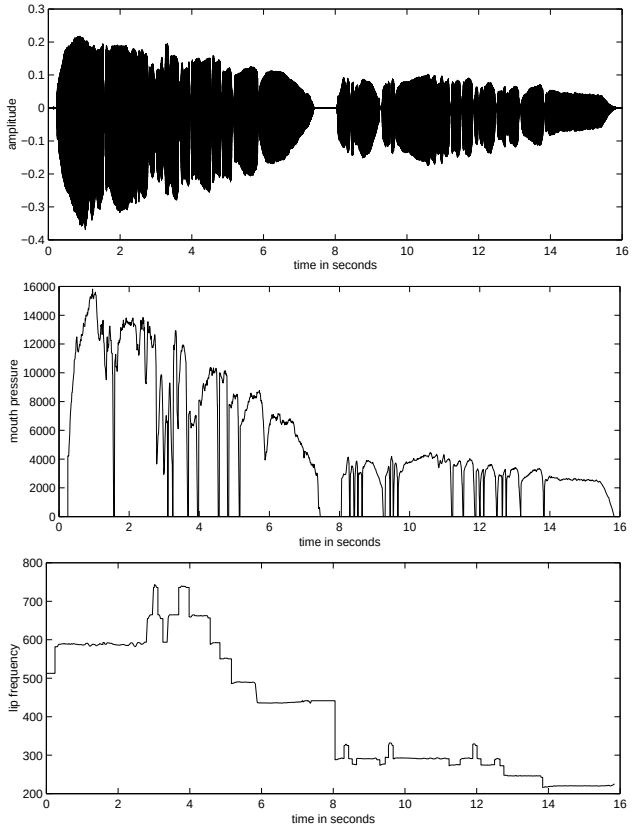

Figure 4: Top, original signal. Middle, estimated mouth pressure. Bottom, estimated lip frequency.

4.1. Posterior Tuning

During the derivation in the previous sections, a fixed set of seven tube lengths was assumed with respect to a given tuning frequency. This implied a unique solution for the mouth pressure and lip frequency. When tuning of the instrument is allowed, a solution will be obtained for every possible tuning frequency. Evidently, this tuning allows only slight variations in tube length, since large variations imply that the modes of the tube will fail to correspond with the desired note frequencies.

When the control parameters were estimated for a given sound, the value of the fundamental frequency was slightly adapted so that the median frequency of the note corresponded with the median frequency of data set. This was done in order to guarantee that the frequency range of the desired sound was available. However, this results in a simulation which is tuned slightly different than the original sound. This tuning can be compensated a posteriori using the following method. When the tube length and lip frequency are changed in manner so that the ratio

$$\frac{P_L}{f_\tau} = \frac{3}{4}(N_i + \Delta N) \tag{19}$$

remains constant, no variation in timbre is perceived. In addition, the expression

$$\Delta f_0 \equiv f_0 - (N + \Delta N) f_\tau \tag{20}$$

was observed to be nearly identical for every tube length. Therefore, $N + \Delta N$ and Δf_0 are kept constant while f_τ is adapted in order to be tuned to the desired frequency f_0'. The new tube length f_τ' is then obtained by taking the median value of

$$\frac{f_0' - \Delta f_0}{N + \Delta N} \tag{21}$$

for all notes that are played with this specific tube length. The new lip frequency values are finally obtained using $P_L' = \frac{3}{4}(N_i + \Delta N) f_\tau'$.

Fig. 5 shows that without posterior tuning, a systematic tuning deviation is obtained between the resynthesis and the original sound. When the tuning is applied, the matching is shown to be very accurate.

4.2. Transient Handling and Attack Improvement

The features that are extracted in section 2.2 implicitly assume that the signal is deterministic and stable during the windowed time frame. In the case of transients, this assumption does not hold implying that the feature extraction fails and the parameter estimation technique cannot be applied. However, a transient must always be considered in its context since it is the transition between two stable parts. Otherwise, we would speak of noise instead of a transient.

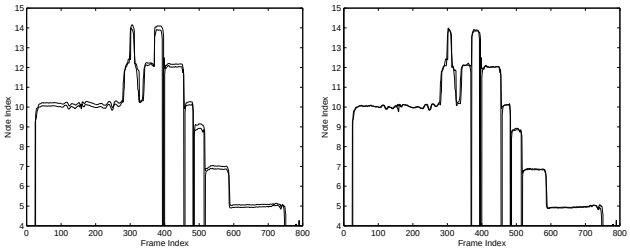

Figure 5: Comparison of the fundamental frequencies before and after posterior tuning.

In the case of the trumpet, the onset, offset and slur are the types of transients that can be distinguished. Therefore, a manual annotation of the sound was realized dividing the sound in silence, stable sound, onset, offset and slur. For the onset and offset, the same lip frequency and tube length were taken as for the preceding and consecutive stable part respectively. In the case of the slur, the response function of the tube was cross-faded between two different tube lengths and the lip frequency and mouth pressure were interpolated linearly.

In addition, a problem was observed at the attack being that the relationship between the control parameters and signal features was not instantaneous. In the case of a sustained sound, the lips open and close regularly. Fig. 6 shows that in the case of an attack, the lips are pushed open by the pressure in the mouth. Then, when the outgoing wave returns at the lips, an oscillation is initiated until finally the stable periodic state is reached. However, this procedure takes about 200 ms to complete implying that no sharp attack is obtained.

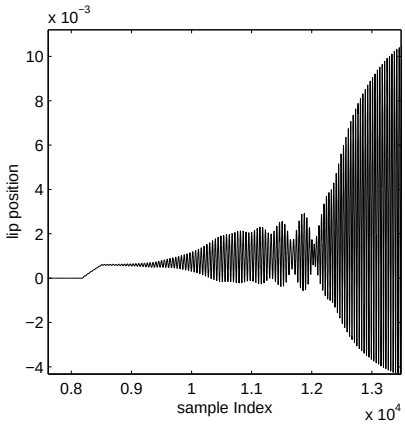

Figure 6: Lip positions at attack.

It can be questioned whether the lips are immobile at the beginning of a note since the trumpet player uses the tongue at the attack. The effect of the tongue does not only result in the fact that the pressure augments instantaneously, but also implies an initial speed of the lips when they open. Since the goal of the attack consists in obtaining the stable sustained state as soon as possible, an initial speed was given to the lips resulting in sharper and more realistic attacks.

5. CONCLUSIONS AND FURTHER RESEARCH DIRECTIONS

In this paper, a new automatic non parametric estimation technique is proposed for the control parameters of a physical model of a trumpet. An important aspect is that the control parameters respect the physical constraints of a real instrument and that the prior knowledge about how the instrument is played is incorporated. This means that a correct tube length and mode combination is selected in order to obtain a given note.

For each of these combinations, a data set was produced containing all possible intensities and variations in fundamental frequency in order to allow vibrato. The similarity between two short time segments was expressed by two complementary criteria being the difference in log fundamental frequency, and the difference between the log spectral envelopes. By using a conditional optimization technique and some posterior tuning of the tube length, an exact solution of the fundamental frequency was achieved while a conditional suboptimal solution was obtained for the spectral envelope. Due to a robust implementation using local quadratic approximations, the estimated control parameters were stable and did not need any post-processing.

Since the estimation can only be applied on stable portions of the sound, an alternative was searched for the transients. These transients were realized successfully by extrapolating control parameters from its context. Also the type of transient was taken into account. Furthermore, the model failed to produce sharp attacks and needed about 200 ms to yield a stable sound. This was improved greatly by adding an additional speed to the lips at the moment of the attack. This initial speed can be related to the effect of the tongue.

The simulation of an expressive trumpet phrase showed that the fundamental frequency could be simulated with a very high accuracy. The timbre on the other hand, is clearly still very different from the original recording. This is due to the fact that the data sets did not contain more similar timbres. Interestingly, the perceived loudness of the simulation was observed to be very similar, which confirms the validity and robustness of the distance metric based on the spectral envelope. However, this distance metric has still some limitations. Although the fundamental frequency and the energy distribution over the partials is characterized, the roughness of the sound and the noise component are not taken into account. One can conclude that the estimation technique allows to realize a simulation of the original sound with the

physical model that has a similar musical expression. The timbre however, can still be improved. Still, one must keep in mind that the computed signal by the physical model corresponds with the pressure wave at the bell of the instrument. This means that the effect of the room is not incorporated while this has an influence on the timbre. In addition, the estimation technique only allows to determine the gestures of the musician, while a large number of instrument parameters, like for instance the reflection function of the instrument, were assumed to be known. This implies that the resynthesis must be considered as a simulation played with a different instrument.

Finally, we remark that this work confronts the two major synthesis paradigms being the signal modelling paradigm and the physical modelling paradigm. For a wide range of signal models accurate parameter estimation techniques are available. Physical models are generally very difficult to invert. The trumpet model that is considered in this paper for instance is a non linear system with delayed feedback. The estimation technique that was proposed is on one hand robust and has a certain generality, but on the other hand it relies on well known parameter estimation techniques from the signal modelling domain. This is at the moment the strongest limitation of the technique. Since no adequate signal parameters can be computed at the transients, it is impossible apply the estimation technique.

6. ACKNOWLEDGEMENT

Wim D'haes is financially supported by the Flemish Institute for the Promotion of Innovation by Science and Technology (IWT).

7. REFERENCES

[1] Christopher M. Bishop. *Neural Networks for Pattern Recognition*, chapter 7. Parameter Optimization Algorithms, pages 253–294. Oxford University Press, 1995.

[2] Marine Campedel-Oudot, Olivier Cappé, and Eric Moulines. Estimation of the spectral envelope of voiced sounds using a penalized likelihood approach. *IEEE Transactions on Speech and Audio Processing*, 9(5):469–481, july 2001.

[3] Olivier Cappé, Jean Laroche, and Eric Moulines. Regularized estimation of cepstrum coefficients from discrete frequency points. *IEEE WASPAA*, October 1995.

[4] Wim D'haes, Dirk Van Dyck, and Xavier Rodet. Discrete cepstrum coefficients as perceptual features. *Proc. of the ICMC (to be published)*, september 2003.

[5] Wim D'haes and Xavier Rodet. Automatic estimation of control parameters: An instance-based learning approach. *Proceedings of the ICMC*, september 2001.

[6] Wim D'haes and Xavier Rodet. Physical constraints for the control of a physical model of a trumpet. *Int. Conference on Digital Audio Effects (DAFx-02)*, September 2002.

[7] Thierry Galas and Xavier Rodet. Generalized discrete cepstral method analysis for deconvolution of source-filter systems with discrete spectra. *IEEE WASPAA*, september 1991.

[8] Thomas Helie, Christophe Vergez, Jean Levine, and Xavier Rodet. Inversion of a physical model of a trumpet. *Proceedings of the ICMC*, pages 149–152, 1999.

[9] Axel Nackaerts, Bart De Moor, and Rudy Lawereins. Parameter estimation for dual-polarized plucked string models. *Proceedings of the ICMC*, September 2001.

[10] Xavier Rodet. Musical signal analysis/synthesis sinusoidal+residual and elementary waveform models. *IEEE Time-Frequency and Time-Scale Workshop (TSTF)*, August 1997.

[11] Diemo Schwarz and Xavier Rodet. Spectral envelope estimation and representation for sound analysis-synthesis. *Proceedings of the ICMC*, pages 351–354, 1999.

[12] Stefania Serafin, Julius O. Smith III, Harvey Thornburg, Frederic Mazzella, Arnaud Tellier, and Guillaume Thonier. Data driven identification and computer animation of a bowed string model. *Proceedings of the ICMC*, September 2001.

[13] Caroline Traube. Estimating the plucked point on guitar string. *Conference on Digital Audio Effects (DAFX-00)*, December 2000.

[14] Christophe Vergez. Trompette et trompettiste: un système dynamique non linéaire à analyser, modéliser et simuler dans un contexte musical. *Ph.D. thesis, Université Paris 6, IRCAM*, 1999.