DOI: 10.61782/fa.2023.0199

CONTRIBUTION OF MACHINE LEARNING AND PHYSICS-BASED SOUND SIMULATIONS FOR THE CHARACTERIZATION OF BRASS INSTRUMENTS

Jean-François Petiot[1*] **Misael Roatta**[1] **Vincent Fréour**[2,3] **Keita Arimoto**[3]

[1] LS2N UMR CNRS 6004, Ecole Centrale de Nantes, France

[2] LMA UMR CNRS 7031, Ecole Centrale Marseille, Aix Marseille Univ., France

[3] YAMAHA Corporation, Research and Development Division, Hamamatsu, Japan

ABSTRACT

Sound simulations by physical modelling are interesting to transcribe the physics underlying the functioning of a musical instrument. These simulations make it possible to listen to a virtual instrument with a mode of operation representative of the musician-instrument interaction. The work consists of studying the contribution of machine learning (ML) methods in the understanding of the relationships between the shape of a trumpet and the sound simulated. The physical model used is based on an acoustical modeling of the resonator, a mechanical model of the excitator, and an aeroelastic coupling between the excitator and the resonator. From different samples of the input impedance of the resonator, time domain simulations are generated to constitute a training set of sounds. Supervised learning is next trained to the data, with the impedance as input and sound descriptors as outputs, using classical ML methods (neural networks). The ML model is finally used to optimize the sound descriptors levels, according to the input impedance. To illustrate the approach, different "targets" for the sound features are considered (brightness, intonation), and a validation is conducted with the simulations. The approach is a first stage toward a "customization" of an instrument according to different perceptual dimensions.

[*]***Corresponding author***: jean-francois.petiot@ec-nantes.fr.

Keywords: *acoustics of brass instruments, time domain simulations, machine learning, sound features.*

1. INTRODUCTION

The study of the quality of musical instruments is particularly interesting to help their development. The quality can be evaluated by physical measurements on the instruments (objective quality) [1]. Concerning brasses, the main physical measurement is the input impedance of the bore [2]. In playing situation, the musician produces a note whose frequency (the playing frequency) is close to the resonance frequency of an impedance peak. But the timbre of a note is also conditioned by upper resonance frequencies of the resonator [3]. Although interesting information can be given by the impedance, it is a hard work to predict sound qualities of brasses only from the impedance.

A second interesting measurement that can be made on brass instruments concerns the analysis of sounds produced in playing situation. Various parameters of the signal can be extracted in order to characterise the sound. The main difficulties in this approach are to overcome the variability produced by the musician, in order to see the differences between the instruments. In this context, artificial player systems are interesting devices to generate sounds with brass instruments in a reproducible way [4]. Another mean to study the objective quality of instruments is to carry out sound simulations by physical modelling [5]. Assessing the brightness of trumpet sounds by a comparison of simulations, an artificial player system and a real musician, has been for example presented in [6]. Sound simulations by physical models constitute a very interesting approach because they allow, by working on a virtual prototype, the

exploration of the design space, by the creation of a large number of virtual instruments. The main interest of these simulations is that the sound result is driven by the causes that create the sound, as for a real instrument: if the physical model used is detailed enough to generate simulations in agreement with the real behavior (such as it is perceived by the musician), then the simulations can constitute a predictive tool for the development of the instrument (virtual acoustics) [7].
In recent years, machine learning (ML) has become an essential approach for the modeling of systems. Neural audio synthesis is nowadays a very active research field, used for example for the synthesis of musical sounds [8].
Our work is in this context. We focus on the ability of simulations by physical modelling to represent certain dimensions of the quality of a trumpet (mainly intonation and timbre). The objective of this preliminary paper is to show how machine learning (ML) methods can be implemented to model the acoustical behavior of a brass instrument (trumpet), the instrument being simulated with a physical model. This paper focuses on the intonation of trumpets and proposes an optimization of the intonation, based on the ML model.
Section 2 presents the physical model used for the simulations and the method to generate a learning dataset. In section 3, the different stages of the method are presented, with the different ML models considered, the method for the optimization of the models, and the validation process. Section 4 describes the first results obtained on the intonation of a Bb trumpet.

2. SOUND SIMULATIONS

2.1 Physical model of the trumpet

In this study, we utilize a classical elementary model of a brass instrument under playing conditions (see Fig. 1).

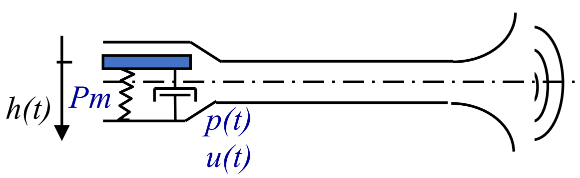

Figure 1. Representation of the outward striking model of the lips. P_m (pressure in the mouth), h (lip aperture), u (volume flow), p (pressure in the mouthpiece).

The vibrating lips are modeled as a one-degree-of-freedom (1-DOF) outward- striking valve, non-linearly coupled to the air column of the brass instrument [9]. This elementary model is a good compromise between simplicity and efficiency, and proved to model many properties of brass instruments [10]. The physics-based model of the trumpet is defined by a system of 3 equations (Eqn. (1)), which depends on three periodic variables: opening height $h(t)$ of the lips, volume flow $u(t)$ ($u(t) = 0$ if $h(t)<0$) and pressure $p(t)$ in the mouthpiece.

$$\begin{cases} P(j\omega) = Z_{in}(j\omega).U(j\omega) \\ \dfrac{d^2h(t)}{dt^2} + \dfrac{\omega_l}{Q_l}\dfrac{dh(t)}{dt} + (\omega_l)^2(h(t)-h_0) = \dfrac{P_m - p(t)}{\mu_l} \\ u(t) = b.h(t)\sqrt{\left|\dfrac{2(P_m - p(t))}{\rho}\right|}.sign(P_m - p(t)) \end{cases} \tag{1}$$

Several parameters are included in this model: air volumic mass ρ, input impedance Z_{in} of the trumpet, and the parameters concerning the musician embouchure (virtual musician). Typical values of the parameters of the virtual musician and their definition are given in table 1.

Table 1. Definition of the parameters of the virtual musician.

Definition	*Notation*	*Typical value*
Pressure in the mouth	*Pm* (Pa)	500 to 14000
Resonance frequency of the lips	$f_l = \omega_l/2\pi$ (Hz)	Variable, according to the regime simulated
Surface density of the lips	μ_l (kgm^{-2})	0.5 to 2
Width of the lips	b (mm)	10
Rest value of the opening height	h_0 (mm)	0.05
Quality factor of the resonance of the lips	Q_l	5

Numerical solutions of the physical model are obtained using an adaptation of the method presented in [11]. This adapted method computes the discrete time series p[n] and

u[n] using a discrete convolution of the resonator impulse response and an explicit numerical scheme of the discrete nonlinear coupled problem.
Different regimes of a Bb trumpet were simulated (Fig. 2). Results are presented in this paper for the regimes 2 (note Bb3), 3 (note F4), 4 (note Bb4), 5 (note D5), and 6 (note F5).

Figure 2. Notes corresponding to the five different regimes simulated on a Bb trumpet.

2.2 Definition of the data for the supervised learning

2.2.1 Variations of the input impedance Z_{in}

The input impedance in open fingering of a set of 9 real Bb trumpets was measured using a sensor developed and commercialized by the Center of Technology Transfer of Le Mans (CTTM). The input impedance is measured from 20 Hz to 2 kHz and the frequency axis corrected to 27°C, an estimate of the temperature inside the instrument.
The impedance is decomposed using modal analysis as a sum of complex modes (Eqn. 2) [12].

$$Z_{in}(\omega) = Z_c\left(\sum_n \frac{C_n}{j\omega - s_n} + \frac{C_n^*}{j\omega - s_n^*}\right) \qquad (2)$$

11 modes were considered to describe the impedance, leading to a set of 44 modal parameters, represented by vector *MP* (each mode is represented by 4 parameters: frequency (*fr_i)*, damping factor, and imaginary (*Im(C)_i*) and real part (*Re(C)_i*) of the residuals).
Given the ranges of each modal parameter for the 9 measured trumpets, a sampling of the modal parameters was made on the ranges with a uniform distribution. A set of *Tr* = 1000 impedances were generated, corresponding to virtual instruments. These instruments constitute the training set of the ML models.

2.2.2 Variations of the virtual musician VM

The parameters of the virtual musicians were adjusted manually by checking the convergence of the simulations toward auto-oscillations for a large majority of trumpets. The values of the parameters are given in table 2: only two parameters are relaxed: the quality factor and the resonance frequency of the lips, the other being fixed.

Table 2. Values of the parameters of the virtual musician *VM*.

Definition	*Notation*	*value*
Pressure in the mouth	*Pm* (Pa)	8000
Resonance frequency of the lips	$f_l = \omega_l/2\pi$ (Hz)	Variable, according to the regime simulated Range : +-30 cent
Surface density of the lips	μ_l (kgm^{-2})	1.3
width of the lips	*b* (mm)	10
rest value of the opening height	h_o (mm)	0.5
Quality factor of the resonance of the lips	Q_L	[4, 6]

Given the range of variations of the frequency f_l and the quality factor Q_l, a Latin Hypercube Sampling of 6 samples was defined, for each regime. For each regime, a set of M = 6 virtual musicians was then defined.

2.3 Characterization of the virtual instrument

2.3.1 Characterization of each simulated sound

To characterize each sound simulated, the fundamental frequencies was estimated using the YIN algorithm [13]. The power spectrum *S(F)* of the sounds was estimated using Fourier transform.

2.3.2 Characterization of all the regimes

For each regime simulated, the average playing frequency f_n for all the *M* virtual musicians was computed, so as the average power spectrum $S(F)_n$.
To characterize each regime of a virtual instrument, two criteria, calculated from the sounds generated, are considered.
The first one, the Equivalent Fundamental Pitch (*EFP* – Eqn. 3), represents the deviation in cent of the average playing frequency f_n from a reference frequency f_R, according to natural intervals. The reference frequency was chosen arbitrarily according to the common tuning note of the instrument (the regime 4, Bb4) (with $f_R = f_4/4$, the *EFP* of the regime 4 is then necessarily equal to “0”).

$$\overline{EFP_n} = 1200.\,log_2(\frac{\overline{f_n}}{n.\,f_R}) \qquad (3)$$

To characterize the global intonation of each virtual instrument, the global average *EFP* was computed. It corresponds simply to the average value of the absolute values of the EFP of all the regimes (2 to 6) (Eqn. 4).

$$\overline{EFP} = \frac{1}{5}.\sum_{k=2}^{6} |\overline{EFP_k}| \qquad (4)$$

The second one (not utilized in this paper) is the spectral centroid Sc_n of the sounds of regime n (Eqn. 5), calculated as the power-weighted average spectral frequency.

$$\overline{Sc_n} = \frac{\int_0^{\infty} f.\,|\overline{S(F)_n}|^2\,df}{\int_0^{\infty} |\overline{S(F)_n}|^2\,df} \qquad (5)$$

To characterize the "brightness" of each virtual instrument, the average spectral centroid is finally computed for all the regimes (Eqn. 6).

$$\overline{Sc} = \frac{1}{5}.\sum_{k=2}^{6} \overline{Sc_n} \qquad (6)$$

2.4 Summary of the simulations

The simulations were carried out for the *Tr* = 1000 impedances, the *M* = 6 virtual musicians and the 5 regimes. The CPU time on a personal computer (INTEL Core i7) to generate these 30 000 sounds was approximately 50 hours. The framework of the simulations is presented in Fig. 3.

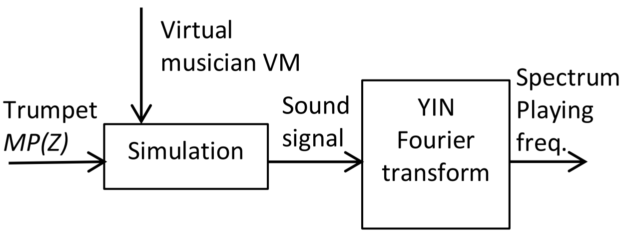

Figure 3. Block-diagram of the sound simulations

3. DATA MODELLING AND OPTIMIZATION

The general approach for the optimization of the impedance of a trumpet based on simulations and ML models consists of the following stages:

- **Generation of sounds (section 2)**. Sound simulations by physical model are used to create a database of sounds, with the input impedance of the instrument as an input (modal parameters), and the sound signal as an output. Different features can be considered to describe the sounds (EFP, spectral centroid, …),
- **Supervised learning (section 3.1)**. A model is fitted to the database, with the characterization of *Z* as an input (*MP*) and the different features as outputs. Different (classical) methods can be tested (regularized regression, Multilayer perceptron (MLP), neural networks, Support Vector Regression, …),
- **Optimization (section 3.2)**. Carry out an optimization of the input impedance, for a given target of the features, using the previous model. Different gradient-free methods can be implemented (Genetic Algorithm, Nelder Mead, …),
- **Validation**. The objective of this last stage is to verify that the optimized impedance, given in stage 3, produces sound features close to the target, using sound simulations.

3.1 Training of the ML models

The supervised learning consists in the prediction of an output, namely the average EFP of each regime EFP_n (of dimension 5), from the input of the system, namely the modal parameters of the impedance *MP* (of dimension 44). The framework of the model is presented in Fig. 4.

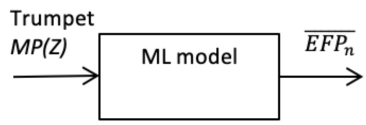

Figure 4. Block-diagram of the ML models

The dataset counts 1000 observations. 70% are used to train, 20% to test (*T*), and 10% to validate (*V*). A cross validation is performed during the train-test phase, with 10 iterations.

Different models are fitted to the data:

- ELASTIC NET regression,
- Multilayer Perceptron MLP with 1 or 2 hidden layers (MLP1 or MLP2).

The criterion used to optimize the hyper-parameters of each model (regime n) is the $RMSPE_n$ (root mean square percentage error), calculated on the test set T (Eqn. 7).

$$RMSPE_n = \sqrt{\sum_{j \in T} \left[\frac{(\overline{EFP_{j,n}} - \widehat{\overline{EFP_{j,\iota}}})}{\overline{EFP_{j,n}}} \right]^2} \quad (7)$$

The quality of the prediction of the ML models is represented by the $RMSPE$, summed over the 5 regimes, and calculated on the validation set V (Eqn. 8).

$$RMSPE = \sum_{i=2}^{6} \sqrt{\sum_{j \in V} \left[\frac{(\overline{EFP_{j,\iota}} - \widehat{\overline{EFP_{j,\iota}}})}{\overline{EFP_{j,\iota}}} \right]^2} \quad (8)$$

3.2 Optimization of Z

Using the ML models, the objective of this stage is to find the input impedance that optimizes the outputs, namely the equivalent fundamental pitch EFP_n.
Given that 4 regimes are present for each instrument (regime 4 is not considered as it has a null EFP – it corresponds to the reference), a multi-objective optimization is carried out. The algorithm used is a Genetic algorithm, NSGAII [14], with 500 individuals per generation, and a budget of 100 generations.
The design optimization problem of an instrument can be formulated as the search for the optimal modal parameters MP^* that minimizes the 4-dimensions objective function EFP_n (Eqn. 7) ($n = 2, 3, 5, 6$).

$$MP^* = argmin(\overline{EFP_n}) \quad (7)$$

The multiobjective optimization leads to the definition of a set of Pareto efficient solutions, called the Pareto front (or Pareto frontier). These Pareto efficient solutions are non-dominated, i.e. it does not exist another feasible solution better than the current one in some objective function without worsening other objective function. These efficient solutions represent different tradeoffs between the objectives.
A unique solution can be chosen on the Pareto front, according to an aggregate criterion (weighted sum for example) or to the proximity of the ideal solution (TOPSIS method) [15].

4. FIRST RESULTS

4.1 Convergence of the simulations

With the set of impedances considered and the set of virtual musicians, all the simulations converged toward an auto-oscillation. This is a sign that the virtual musicians were correctly tuned.
For very few regimes (5/30 000), the regime simulated was incorrect, just above the expected regime (for instance, a regime 3 was simulated instead of a regime 2). Even if these data leads to a very large EFP, they were kept in the training set.

4.2 Fitting of the models

The performances of the fitting of the models to the data are presented in table 3.

Table 3. performances of the different ML models.

Type ML model	*RMSPE*	$\overline{EFP^m_{min}}$ (model) - cent	$\overline{EFP^s_{min}}$ (simu) - cent
Elastic net	2.029	0.0976	4.047
MLP1	1.452	0.231	0.933
MLP2	1.381	0.0410	1.148

The prediction is the worst for the Elastic net method ($RMSPE$ = 2.029). The best ML models are obtained with neural networks (MLP1 or MLP2), that obtained the lowest $RMSPE$.
In order to interpret the ML models, and verify that the explanatory variables make sense from a physical point of view, it is interesting to examine the magnitude of the variables of the ML models. This is possible for the Elastic net model, but not for the MLP models. The magnitude of the variables in the Elastic-net models are plotted in Fig. 5 for each regime.

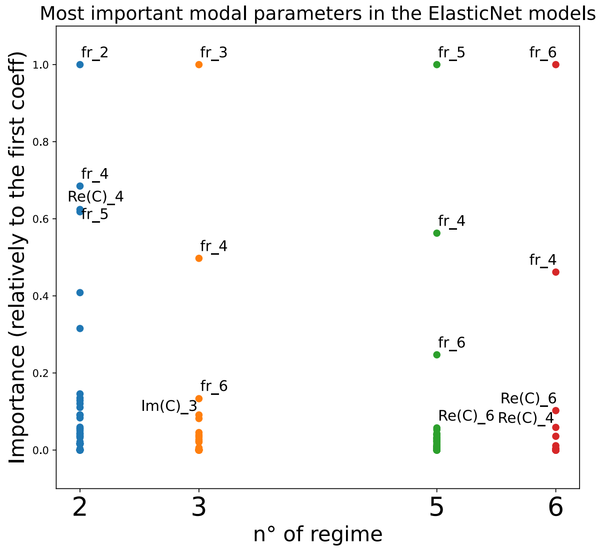

Figure 5. magnitude of the variables in the ML model of EFP for regimes 2, 3, 5 and 6.

Fig.5 shows that, for each regime, the most important variable is the frequency of the mode corresponding to this regime. For example, resonance frequency *fr_6* is the variable of the impedance that has a major effect on the EFP of regime 6. This is coherent from a physical point of view: the playing frequency of a sound is mainly governed by the corresponding resonance frequencies of the impedance. Given that the EFP is defined relatively to a tuning note (regime 4), it also makes sense that the resonance of the fourth mode *fr_4* is important to predict the EFP (second most important modal parameter).

4.3 Optimal instrument according to EFP

4.3.1 Efficiency of the optimization

After convergence of the optimization, the genetic algorithm provides the Pareto front. One solution was chosen on the Pareto front by minimizing the maximum of the EFP for all the regimes.

This solution gives the modal parameters MP^* of an "optimal instrument", characterized by the impedance Z^*. The minimum value of the average $\overline{EFP}$ (Eqn. 4), given by the ML model at this MP^*, is presented in table 3: it is labelled $\overline{EFP^m_{min}}$ (m for "model"). From MP^*, sound simulations can be performed to calculate the average EFP: it is labelled $\overline{EFP^s_{min}}$ (s for "simulation").

The deviation between $\overline{EFP^s_{min}}$ and $\overline{EFP^m_{min}}$ is representative of the quality of the modeling and the efficiency of the optimization.

Using Elastic net, the optimal solution, given by the model ($\overline{EFP^m_{min}} = 0.0976$) is finally not close to the simulated EFP ($\overline{EFP^s_{min}} = 4.047$). This is expected, due to the large prediction error of this method.

MLP models are more efficient. Table 3 shows that the best fitting (MLP2) doesn't lead necessarily to the best $\overline{EFP^s_{min}}$ (obtained by MLP1). MLP1 model provides in this case the most interesting solution.

4.3.2 Verification of the impedance of the optimal instrument

To illustrate the performance of this solution (optimal instrument with MLP1), it is interesting to calculate the "inharmonicity" of the resonances of the impedance corresponding to MP^*. This "inharmonicity" can be represented by the EFP, calculated on the resonance frequencies (with a similar definition than Eqn. 3).

The results show that the optimal instrument possesses resonance of the impedance that are close to harmonicity.

This first result is in agreement with a well-known intuition of instrument makers concerning the "harmonicity" of the resonances of the impedance. But further works are needed to investigate in detail the stability of this first result and the sensitivity to the choice of the virtual musicians.

5. CONCLUSIONS

The methodology presented in this paper is a first attempt to show how sound simulations and Machine learning models can be used to characterize the sound qualities of musical instruments. It was applied to the trumpet and to a dimension related to the intonation of the instrument,

represented by the EFP. The modeling of the EFP with a set of virtual instruments was successful, and the optimization led to a reliable result: the instrument that minimizes the EFP in playing conditions is an instrument that also minimizes the EFP of the resonances of the impedance ("harmonic" resonances).
The perspectives of this approach are numerous. It will be first interesting to assess the sensitivity of the results to the choice of the virtual musicians. It will be also very interesting to address a more complex percept than the intonation, the timbre of the instruments. The spectral centroid will be next the theme of the modelling, together with the integration of non-linear propagations in the resonator.
Another perspective of this study will be to take into account spectro-temporal descriptors, in order to study differences according to transient regimes of the instruments.

6. ACKNOWLEDGMENTS

This project is a direct continuation of the research works carried out with Joël GILBERT for more than 20 years [4, 6, 7, 10]. We wish to honor his memory and pay tribute for his major contributions to the acoustics of brass instruments.

7. REFERENCES

[1] Pratt R.L., Bowsher J.M. "The objective assessment of trombone quality". *Journal of Sound and Vibration* 65(4), 521-547, (1979).

[2] Dalmont, J.P. Acoustic impedance measurement, Part I: a review. *J. Sound and Vib*., 243, 427-439, (2001).

[3] Benade A. H. "Relation of Air-Column Resonances to Sound Spectra Produced by Wind Instruments," *J. Acoust. Soc. Am*. 40, 247-249, (1966).

[4] Petiot J.F., Tessier F., Gilbert J., Campbell M. "Comparative Analysis of Brass Wind Instruments With an Artificial Mouth: First Results". *Acta Acustica*. n°6, Vol 89, pp 974-979, (2003).

[5] Farina A., Tronchin L. "On the "virtual" reconstruction of sound quality of trumpets". *Acustica – Acta Acustica*. Vol. 86, 747-745, (2000).

[6] Poirson E., Petiot J.-F., Gilbert J. "Study of the brightness of trumpet tones". *Journal of the Acoustic Society of America*, Vol. 118(4), (2005).

[7] Petiot J-F., Gilbert J. "Comparison of trumpets' sounds played by a musician or simulated by physical modeling". *Acta Acustica united wih Acustica*. Vol. 99, (2013), 629-641. DOI 10.3813/AAA.918642.

[8] Hawley S.H., Chatziioannou V., Morrison A. "Synthesis of Musical Instrument Sounds: Physics-Based Modeling or Machine Learning?". Acoustics Today, Spring 2020, Vol. 16, Issue 1. https://doi.org/10.1121/AT.2020.16.1.20.

[9] Adachi S. and Sato M. "Time domain simulation of sound production in the brass instruments". *The Journal of the Acoustical Society of America*, 97(6):3850-3861, June 1995.

[10] Tournemenne R., Petiot J-F., Talgorn B., Kokkolaras M., Gilbert J. (2016). "Brass instruments design using physics-based sound simulation models and surrogate-assisted derivative-free optimization". *Journal of Mechanical Design*, April 2017, Vol. 139, 0141401-1-011401-9.

[11] Guillemain P., Kergomard J. and Voinier T. "Real-time synthesis of clarinet-like instruments using digital impedance models". *The Journal of the Acoustical Society of America*, 118(1): 483–494, 2005.

[12] V. Fréour, L. Guillot, H. Masuda, S. Usa, E. Tominaga, Y. Tohgi, C. Vergez, and B. Cochelin. "Numerical continuation of a physical model of brass instruments: Application to trumpet comparisons". *J. Acoust. Soc. Am*., vol. 148, no. 2, pp. 748–758, 2020.

[13] De Cheveigné A. and Kawahara H. "YIN, a fundamental frequency estimator for speech and music". *The Journal of the Acoustical Society of America,* 111(4):1917-1930, April 2002.

[14] Deb, K., Pratap, A., Agarwal, S. and Meyarivan, T. (2002). "A fast and elitist multiobjective genetic algorithm: NSGA-II". *IEEE Transactions on Evolutionary Computation*, 6(2), 182–197.

[15] Hwang, C.-L., and Yoon, K. (1981). Multiple Attribute Decision Making. Vol. 186, Berlin, Heidelberg: Springer Berlin Heidelberg. doi:10.1007/978-3-642-48318-9.