REAL-TIME VISUALISATION OF THE MUSICAL TIMBRE BASED ON THE SPECTRAL ESTIMATES OF THE SNAIL-ANALYSER

Thomas Hélie, Charles Picasso, Robert Piéchaud,
STMS laboratory (IRCAM-CNRS-SU)
1, place Igor Stravinsky
Paris, France
`firstname.name@ircam.fr`

Michaël Jousserand, Tom Colinot
Buffet Crampon
Mantes-La-Ville, France
`firstname.name@buffetcrampon.com`

ABSTRACT

This article presents a real-time software solution that allows musicians to visualise the timbre content of their musical tones. The timbre representation is based on the spectral estimates of the Snail-Analyser, for a high frequency precision, and on a harmonic-like representation. After a brief review on the derivation of these estimates, some second-stage estimates and the mapping used for the timbre representation are described. The visual representations in the application have been prototyped using the MAX software and developed with the Juce framework.

1. INTRODUCTION

The Snail-Analyser is a real-time software that allows the spectral analysis of a sound with a high frequency precision [1, 2]. It delivers a visual rendering on a musical scale made of a spiral (one turn is one octave, one angle is one chroma).

The ATRIM project[1] aims to design reliable tools with reactive visual renderings adapted to high precision pitch and timbre analysis of musical wind instruments. It is devoted to provide musicians with accurate and informative feedback in real time for tuning and timbre assessment, during performance and in several contexts such as: expert testing and improvement of manufactured instruments, tuning or comparison of instruments, musical practice by helping musicians adjust their motor control to reach their own or a teacher's target (intonation, timbre, vibrato, glissando, playing effects, etc).

A first tuning application has been designed in the ATRIM project to assess the frequency-deviation from the target pitch, in Hertz [3]. Its technology involves demodulated phase signals, low-pass filtered phase-constancy indicators, and other complementary signals calculated in the spectral estimates of the Snail-Analyser. The visual rendering mimics electro-mechanical strobe-tuners that exploit a stroboscopic technology (see e.g [4, 5] for original work and [3] and references therein for some brief descriptive historical elements and more recent non-mechanical versions). A second version of this application has been designed to assess a more comfortable deviation, in cents [6]. It involves a reactive local-in-time constancy indicator of the demodulated phase.

This paper addresses the assessment of timbre. This issue has a long history from the point of view of perception, cognition and signal processing (see e.g. [7, 8, 9]) and learning algorithms (see e.g. [10] and references therein for a recent work integrating results on timbre and instrumental playing techniques). A resulting general statement is that timbre is related to spectrotemporal indicators and their modulations, and that it intimately covaries with pitch and amplitude.

The paper is organised as follows. First, Section 2 presents a brief review on some spectral estimates used in the Snail-Analyser, the spiral visual rendering, a tuning spinner and in a rendering with harmonic structure that provides a starting basis to address musical timbre representation. Then, the analyser is complemented by an indicator overbuilt on that of local-in-time constancy of the demodulated phase, that characterises the harmonic synchronicity, in order to provide a set of spectrotemporal timbre indicators adapted to quasi-harmonic signals. Second, Section 3 presents the software prototype and the designed visual rendering based on these indicators. Section 4 ends with conclusive remarks and perspectives.

2. ANALYSIS METHOD

This section first presents a brief review on the principle used in the Snail-Analyser [1] with its spectral estimates and their use in the spiral rendering (in Section 2.1). These estimates are used in a harmonic view (in Section 2.2) of interest to represent timbre. They are complemented by an overbuilt estimate, the harmonic synchronicity (in Section 2.3), to be used in the final visualisation.

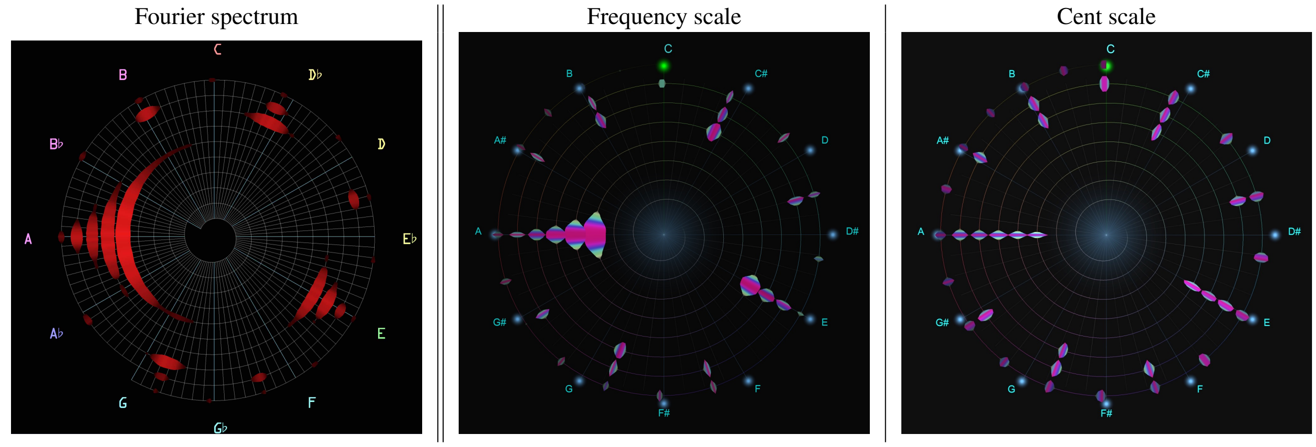
2.1. Review of the Snail-Analyser process

The analysis process of the input sampled signal is based on (see [2, Sec. 3.1] for more details) a spectral transform (here, a short-time Fourier transform) for successive frames, the complex values of which are interpolated according to a vector `Freq_v` of tuned frequencies that are exponentially spaced to correspond to equally spaced midi codes, and spectral estimates introduced in [1], some of them being recalled below to make this paper self-contained. Typical parameters are: $F_s = 44.1\,\mathrm{kHz}$ (sampling frequency), $T_{\mathrm{frame}} = 50\,\mathrm{ms}$ (duration of a classic-shape window) associated with the next power of 2 of the corresponding number of samples (number of FFT points), $T = 5\,\mathrm{ms}$ (incremental time step for overlap) and $R_m = 20$ frequency points per half tone (midi resolution) for `Freq_v` over the analysed midi range.

The downstream indicators of the Snail-Analyser are derived from the spectrum modulus and phase. They are presented by steps, denoted **(Si)** below, and defined for all frames starting at time $t \in \{nT \text{ s.t. } n \in \mathbb{N}\}$ and all frequencies f in `Freq_v`.

---

[1] ATRIM is the French acronym for "Analyseur Temps-Réel haute précision de justesse et de timbre pour Instruments Musicaux" (High precision real-time pitch and timbre analyser for musical instruments). This project is supported by the plan "France Relance" (see acknowledgements at the end of the paper).


Fourier spectrum

Frequency scale

Cent scale

Figure 1: *Spiral renderings of the spectrum of an ideal harmonic signal (harmonics synthetised with uniformly distributed amplitudes). From left to right: Fourier transform (50ms, Hanning), Snail-Analyser (frequency precision:* 6 *Hz), Snail-Analyser (cent precision:* 5 *cents).*

.

(S1) Demodulated phase (ϕ_d): this signal is defined by

$$\phi_d(t,f) = \phi_{Fourier}(t,f) - 2\pi f t, \qquad (1)$$

where $\phi_{Fourier}(t,f)$ denotes the spectrum phase.

(S2) Phase constancy indicator (C_{f_c}, set by f_c): this signal is defined as the squared modulus of the output of a low-pass filter (e.g. of Butterworth type) with cutoff frequency f_c, excited by the input complex signal $u_f : t \mapsto \exp\big(i\,\phi_d(t,f)\big)$, namely,

$$\xrightarrow{u_f(t)=\mathrm{e}^{i\,\phi_d(t,f)}} \boxed{\text{Low-Pass } (f_c)} \rightarrow \boxed{|\cdot|^2} \xrightarrow{C_{f_c}(t,f)} \qquad (2)$$

This indicator is used as a factor multiplying the spectrum loudness (ISO226:2003) to form the thickness represented on the spiral representation[2]. Compared to the Fourier spectrum (see Fig. 1 left), this improves the frequency accuracy by contracting the lobes to a precision of about $\pm f_c$ ($f_c = 6\,\text{Hz}$ in center figure and $f_c(f) = 2^{\frac{\kappa}{1200}} f$ with $\kappa = 5\,\text{cents}$ in the right figure). Due to the low-pass filter, this indicator also rejects time-varying components faster than f_c, which is of interest for tuning tasks in noisy environments.

To capture fast components (glissando, vibrato, etc) with the same precision ($\pm f_c$), a third (reactive) indicator is introduced.

(S3) Local-in-time constancy ($\delta\Phi_d$): this signal is defined by

$$\delta\Phi_d(t,f) = M[\phi_d(t,f) - \phi_d(t-T,f)], \qquad (3)$$

where M denotes the 2π-modulo centered on $(-\pi,\pi)$.

This deviation indicator is local-in-time, at the scale of the incremental time step T. It quantifies a deviation-speed indicator that can be converted in Hertz (frequency deviation δF), in cents (cent deviation $\delta\kappa$) and used to contract the spectral lobes (factor $\tilde{C}_{f_c}$) according to the shape inherited from the low-pass filter chosen to compute C_{f_c}, as described below.

(S4) Frequency deviation (δF): this signal defined by

$$\delta F(t,f) = \delta\Phi_d(t,f)/(2\pi T), \qquad (4)$$

estimates a frequency deviation from f to the central frequency $F \approx f + \delta F$ of the spectral lobe of a stationary component (see e.g. [11] for an overview of other estimates used for spectrum reallocation). Its robustness (to perturbations or an unsteady frequency component) is achieved through the averaging of the demodulated phase evolution over the frame duration T_{frame}. Its reliability is limited to the range $\frac{1}{2T}(-1,1)$ of unaliased frequency deviations (100 Hz for $T = 5\,\text{ms}$).

(S5) Cent deviation ($\delta\kappa$): this signal is defined by

$$\delta\kappa(t,f) = \mathcal{C}ents\left(\frac{f+\delta F(t,f)}{f}\right), \qquad (5a)$$

$$\text{where } \mathcal{C}ents(f_1/f_2) = 1200\log_2(f_1/f_2), \qquad (5b)$$

estimates the deviation of f_1 from a reference frequency f_2 in cents. A good approximation of $\delta\kappa$ is $\frac{1200}{\ln 2}\,\frac{\delta F}{f}$ for $|\delta F/f| \ll 1$. Note that according to the definition of δF, the unaliased range in cents depends on f and is given by $\pm 1200\log_2\left(1+\frac{1}{2Tf}\right)$.

(S6) Contraction factor (factor $\tilde{C}_{f_c}$): this signal defined by

$$\tilde{C}_{f_c}(t,f) = \Big| H\big(\delta F(t,f)/f_c\big)\Big|^2, \qquad (6)$$

restores the same shape as C_{f_c} for a stationary sound input signal, choosing $f \mapsto H(f/f_c)$ as the transfer function of the low-pass filter used to compute C_{f_c}.

The indicator ϕ_d mapped to the angle of a spinning pattern is the one used in the Snail-Analyser and in the strobe-tuner application [3], and that $\delta\kappa$ mapped to the angle velocity of the same pattern is the one used in the cent-sensitive version [6]. The contraction factor $\tilde{C}_{f_c}$ can be used in the Snail-Analyser: the combination of its high time reactivity and frequency precision makes it particularly suitable for enhancing real-time spectrogram-like rendering.

---

[2] The spiral skeleton (+1 round from the center is +1 octave, so that one angle is one chroma) is defined in polar coordinates by $\rho(f) = 1+\log_2(f/f_{min})$, $\theta(f) = \theta_{ref} + 2\pi\log_2(f/f_{ref})$ where f_{min} and f_{max} are the lowest and highest frequencies to be displayed and f_{ref} is the tuning reference displayed at angle θ_{ref}.

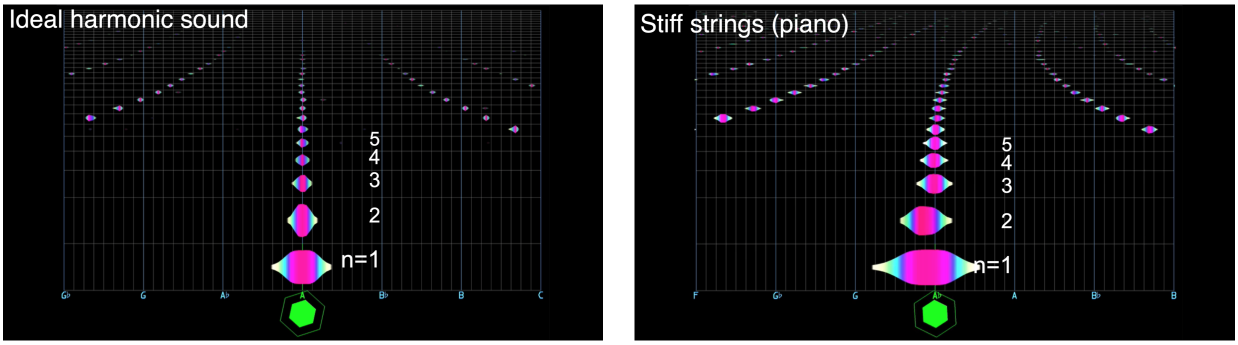


Figure 2: *Harmonic representation ($f_c = 6\,Hz$): (left) ideal harmonic signal of Fig. 1; (right) inharmonic string (acoustic upright piano).*

2.2. Harmonic view

A new rendering to be integrated into future Snail-Analyser distributions is a harmonic view that aligns harmonic components (see Fig. 2). Given a fundamental frequency F_1 (e.g. 440 Hz for note A4), each stage n represents a spectral area of harmonics $n \geq 1$ on a cent-scaled axis centered according to the abscissa

$$X^n_{F_1}(f) = \mathcal{C}ents\left(\frac{f}{nF_1}\right) = 1200\log_2\frac{f}{nF_1}. \qquad (7)$$

This representation is chosen as one of the bases of the timbre representation described in Section 3.

2.3. Harmonic synchronicity

The harmonic information is complemented by new descriptors of harmonic synchronicity in Hertz (ν_n) or in cents (σ_n), for each harmonic $n \geq 1$, as follows.

(S7) Harmonic frequency synchronicity (ν_n): it is defined by

$$\nu_n(t,f) = \delta F(t,nf) - n\,\delta F(t,f). \qquad (8)$$

This descriptor provides a local-in-time estimation of the harmonic synchronicity for nearly harmonic signals in the following sense.

Consider an input signal composed of nearly harmonic partials of frequencies $F_n = n(1+\epsilon_n)F_1$ (with $\epsilon_1 = 0$), where ϵ_n encodes the relative deviation of F_n compared to the expected harmonic frequency nF_1. Then, consider an analysis frequency f and assume that nf is within the frequency range of the spectrum lobe of the partial of frequency F_n. In this case, $\delta F(t,nf) \approx F_n - nf$ estimates the frequency deviation between the ground-truth frequency F_n and the analysis frequency nf, and $\nu_n(t,f) \approx n(1+\epsilon_n)F_1 - nF_1 = n\epsilon_n F_1$ estimates the deviation from the harmonic frequencies, independently from the analysis frequency f.

A conversion of ν_n into cents (relative to nf) is below.

(S8) Cent harmonic synchronicity (σ_n): it is defined by

$$\begin{aligned}\sigma_n(t,f) &= \mathcal{C}ents\left(\frac{nf + \nu_n(t,f)}{nf}\right)\\ &= 1200\log_2\left(1+\frac{\nu_n(t,f)}{nf}\right), \qquad (9)\end{aligned}$$

a good approximation of which is given for small frequency deviations by $\sigma_n(t,f) \approx \frac{1200}{\ln 2}\frac{\nu_n(t,f)}{nf}$.

Note that a version into cents, relative to an estimate of the target harmonic frequency $nF_1 \approx n\big(f+\delta F(t,f)\big)$, is $\tilde{\sigma}_n(t,f) = 1200\log_2\left(1+\frac{\nu_n(t,f)}{n\left(f+\delta F(t,f)\right)}\right) \approx \frac{1200}{\ln 2}\frac{\nu_n(t,f)}{n\left(f+\delta F(t,f)\right)}$.

3. VISUAL RENDERING AND SOFTWARE

3.1. Description of the visualisation

The software is a first prototype integrating a tuning spinner [6] and a harmonic representation inspired from Fig. 2 complemented by spinners cent-sensitive to the harmonic synchronicity (σ_n). The visual rendering is organised in two main parts (see Fig. 3).

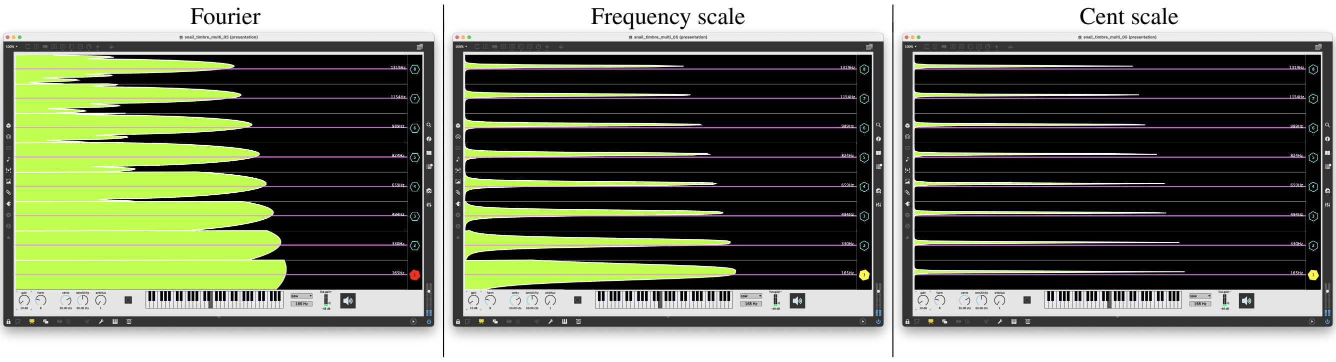
The left part displays a vertical representation of the harmonics, with the fundamental (of target pitch F_1) at the bottom of the representation and higher harmonics superimposed in tiers above. Note that this vertical harmonic representation is not the spectrum unrolled on the frequency Y-axis but a superposition of visual windows, centered around these harmonics, extracted from the Snail analysis kernel. Each visual window covers a range of ±1 semitones around the target harmonic frequency (nF_1), the target being represented by a pink line in the middle of the window.

The right part of each window displays a rotating spinning hexagon, that conveys pitch and harmonic synchronicity information. In the lowest one (pitch window), the rotation speed indicates the pitch deviation from the target F_1. Its color linearly varies from red (50 cents or more deviation) to green (0 cent). In the staged upper windows (harmonics $n \geq 2$), the rotation of the hexagons (of fixed blue color for now) indicates the harmonic synchronicity estimate, with an angular speed proportional to $\nu_n(t, nF_1)$.

The Figure 3 shows a sawtooth wave playing a E3 (normally 164.81Hz) slightly detuned to +20 cents (=166.73Hz). Each F0 spinner on the right has a yellow color due to the +20 cts deviation from the target frequency (E3=164.81Hz).

Consistently, the center of the lobes deviates above the reference pink line due to the slight detuning effect. The figure on the right shows the result using the cents scale processing of the contraction factor $\tilde{C}_{f_c}$ (f_c is set to 5 cents as in Fig. 1-right). This method significantly enhances accuracy and ensures equal sizing of the lobes on a cents scale. This is in contrast to the center figure, which represents the same signal but with f_c = 6,Hz (as in Fig. 1-center) and displays unequal lobe sizes.

To compare the precision stemming from the Snail analysis to the Fourier approach, Figure 3-left shows the loudness of the Fourier spectrum of the same signal.



Fourier

Frequency scale

Cent scale

Figure 3: *Harmonicity renderings of a sawtooth wave with a fundamental frequency at 166.73Hz (E3 +20 cents), 8 visible harmonics. From left to right: Fourier (50ms, Hanning), Snail-Analyser (frequency precision: 6 Hz), Snail-Analyser (cent precision: 5 cents).*

3.2. Rapid prototyping with MAX

The prototype Software and the visualisation have been developed using the Max environment [12], a visual programming paradigm for interactive multimedia applications. A Max external object (MXO), loaded as a dynamic library in the environment at startup, encapsulates the Snail kernel library and computes for the signal processing chain. The real-time visualisation is then built using the Javascript for User Interface (or JSUI [13]) in Max, which has internal bindings to the Max graphics engine. Each analysis signal from the Max object is passed on to the JSUI object to build the rendering. This first approach proves reliable enough for the design of proof of concept real-time renderings, running approximately at 25 frames per second and requiring fast development iterations [6]. Thus, other views for future developments are yet to be tested. Once the user interface prototypes reach a final state, a desktop version of the prototype should also be converted to the JUCE environment [14].

4. CONCLUSION AND PERSPECTIVES

The Snail analysis process appears relevant to design applications specifically aimed at real-time representations of the timbre of nearly harmonic sounds. The proposed visual rendering incorporates reactive descriptors pertaining to the amplitude of spectral components, tuning, and harmonic synchronicity. These descriptors ensure accurate frequency representation, making this prototype a reliable initial development. Further work will focus on exploring this tool and refining visualisations in collaboration with expert musicians, integrating new descriptors to enhance both information and legibility, and designing a desktop application.

5. ACKNOWLEDGMENTS

The authors thank the workplan "France Relance" for the financial support of the ATRIM project 21-PRRD-0001-01-11-009. This project is collaborative work between Buffet Crampon and the STMS laboratory.

6. REFERENCES

[1] T. Hélie, “Analyse fréquentielle par démduolation de phase d’un signal acoustique,” French Patent App. FR1455456A (2014), Int. Patent App. WO2015/189419 A1 (2015), Assignee: Centre National de la Recherche Scientifique, 2014.

[2] T. Hélie and C. Picasso, “The snail: a real-time software application to visualize sounds,” in *Proc. DAFx*, 2017.

[3] T. Hélie, C. Picasso, R. Piéchaud, M. Jousserand, and T. Colinot, “Variations on the Snail-Analyser and its spectral estimates for the accurate tuning of musical sounds: a strobe tuner like display,” in *31st ERK*, Portoroz, Slovenia, Sept. 2022.

[4] R. W. Young and L. Allen, “Stroboscope,” 1938, US2286030A.

[5] G. E. Arends and R. S. Dobrose, “Strobe tuner,” 1997, US5877443A.

[6] T. Hélie, C. Picasso, R. Piéchaud, M. Jousserand, and T. Colinot, “A Real-Time Cent Sensitive Strobe-like Tuning Software Based on Spectral Estimates of the Snail-Analyser,” in *Sound and Music Computing Conference*, Stockholm, Sweden, June 2023, vol. 19, pp. 1–7.

[7] J. M. Grey, “Multidimensional perceptual scaling of musical timbres,” *the Journal of the Acoustical Society of America*, vol. 61, no. 5, pp. 1270–1277, 1977.

[8] S. McAdams, S. Winsberg, S. Donnadieu, G. De Soete, and J. Krimphoff, “Perceptual scaling of synthesized musical timbres: Common dimensions, specificities, and latent subject classes,” *Psychological research*, vol. 58, pp. 177–192, 1995.

[9] K. Siedenburg, I. Fujinaga, and S. McAdams, “A comparison of approaches to timbre descriptors in music information retrieval and music psychology,” *Journal of New Music Research*, vol. 45, no. 1, pp. 27–41, 2016.

[10] V. Lostanlen, C. El-Hajj, M. Rossignol, G. Lafay, J. Andén, and M. Lagrange, “Time–frequency scattering accurately models auditory similarities between instrumental playing techniques,” *EURASIP Journal on Audio, Speech, and Music Processing*, vol. 2021, no. 1, pp. 1–21, 2021.

[11] F. Auger, P. Flandrin, Y.-T. Lin, S. McLaughlin, S. Meignen, T. Oberlin, and H.-T. Wu, “Time-frequency reassignment and synchrosqueezing: An overview,” *IEEE Signal Processing Magazine*, vol. 30, no. 6, pp. 32–41, 2013.

[12] Website Cycling74, “Max home page,” https://cycling74.com/products/max.

[13] Website Cycling74, “Max JSUI Home Page,” https://docs.cycling74.com/max7/refpages/jsui.

[14] Website, “JUCE Home Page,” https://www.juce.com.