

Fan-Chirp Transform with a Timbre-Independent Saliency Applied to Polyphonic Music Analysis

Isabela F. Apolinário and Luiz W. P. Biscainho

Abstract—This article presents the analysis of polyphonic music signals using the Fan Chirp Transform (FChT). Its main innovation consists in using a timbre-independent saliency function, instead of the classical approach where a partial accumulation is computed. A set of simulations allows one to conclude that the proposed approach yields better results for noise-corrupted signals besides lower computational complexity.

Keywords—Fan Chirp Transform, Music Signal Analysis, Saliency Function, Timbre.

I. INTRODUCTION

A recurrent task in the context of music information retrieval is to track predominant fundamental frequencies corresponding to the existing main melodies. This procedure is essential in many applications, such as automatic music transcription [1] and sound source separation [2]. A usual strategy to aid in this task is to compute a “saliency function” that indicates the relevance of each spectral component in the signal.

The classical approach to compute the saliency of a given frequency f_0 consists in accumulating the signal spectrum energy at integer multiple of f_0 [3]:

$$\rho(f_0) = \frac{1}{n_H} \sum_{i=1}^{n_H} \log |X(if_0)|, \quad (1)$$

where $X(f)$ is the spectrum of a discrete signal $x(n)$ and n_H a pre-determined number of harmonic partials. Since Equation (1) uses information of the spectrum magnitude, this saliency function is dependent of sound source timbres.

Another approach to compute a saliency function was presented in [4]. This method, here named as DLMP (after the authors’ initials), uses information of frequency location to determine the most prominent existing pitches. The peaks of the spectrum are estimated, and a saliency value is assigned to each detected peak. Such values reflect a “probability” of the considered peak being an existing fundamental frequency. For this purpose, a theoretical sequence of deviations is defined, based on the distance between the observed peak frequencies and the notes in a 12-note equal tempered scale [4].

Some modifications to the original DLMP method were proposed in [5], among which the most important are the following: a pre-processing step of spectrum noise floor removal was added in order to account for higher partials in the peak detection stage; the correlation coefficient is used as a measure of similarity between theoretical and observed

sequences, instead of their inner product; and a post-processing stage to remove ambiguity created by partials of existing fundamental frequencies was included.

As previously said, the saliency function informs the relevance of the existing pitches in a given discrete-time signal. However, the main interest here is to study the evolution in time of multiple fundamental frequencies of an audio signal. To that purpose, the signal under analysis is considered to be piecewise stationary and the spectral content of each time frame is evaluated by means of the Discrete Fourier Transform (DFT). This operation is known in the literature as the Short-Time Fourier Transform (STFT) [6].

The Fan Chirp Transform (FChT), first introduced in [7], models each fundamental frequency in a signal, along with its superior harmonic partials, as a linear function in time. In [8], the FChT was applied to music signals by means of the Short-Time Fan Chirp Transform (STFChT). This way, a fundamental frequency in a given music signal is considered to vary linearly in time for short time frames. This change allows sparser representations¹, and thereby possibly better precision when estimating the existing fundamental frequencies.

The main goal of this work is to study the performance attained by different strategies for computing the existing fundamental frequencies in polyphonic signals: A) combining the STFT with both presented saliency functions; and B) replacing the STFT by the STFChT as the chosen time-frequency transform. To that purpose, two measures of efficiency are adopted: the mean squared error between the annotated correct fundamental frequency and the estimated one, and a hit rate within a $\pm 3\%$ error tolerance. The same analysis is carried out when adding white noise to the signal in order to assess the methods’ robustness.

This paper is organized as follows. Section II briefly explains the timbre-independent saliency function, along with the proposed modifications. Section III introduces the FChT and explains its original implementation as defined in [8]. Section IV presents the performed experiments and respective results. Lastly, Section V draws some conclusions.

II. TIMBRE-INDEPENDENT SALIENCY FUNCTION

This section briefly explains the timbre-independent saliency function considered in this work [4], [5].

Figure 1 shows the block diagram of the DLMP saliency computation method [4]. The input signal $x(n)$ is assumed to be a discrete L -sample long audio signal, i.e., $n =$

Isabela F. Apolinário and Luiz W. P. Biscainho are with the Program of Electrical Engineering, COPPE/Federal University of Rio de Janeiro, Rio de Janeiro-RJ, Brazil, E-mails: isabela.apolinario@smt.ufrj.br, wagner@smt.ufrj.br. This work was partially supported by CNPq and CAPES.

¹Here, sparsity accounts for concentration around regions of the time-frequency plane corresponding to emitted tonal sounds.

$0, 1, \dots, L - 1$. The first step computes a time-frequency transform of $x(n)$, $X(k, j)$, where $k = 0, 1, \dots, K - 1$ and $j = 0, 1, \dots, J - 1$ are time and frequency indexes, respectively. The analysis window has N samples, and the hop size is τ . Two different time-frequency transforms, the STFT and the STFChT, will be considered in this work.

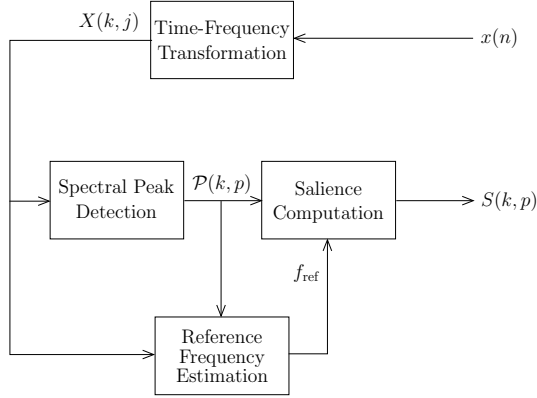


Fig. 1. Block diagram of the timbre-independent salience computation method.

The next step is the spectral peak detection. First, for each time frame k , the noise floor of the signal spectrum is estimated by the Stochastic Spectrum Estimator (SSE) method, detailed in [9], and then subtracted from the original magnitude spectrum in logarithmic scale. The peaks are then estimated from the resulting “flattened” spectrum through the same strategy used in [6], producing the sequence $\mathcal{P}(k, p) = \{(f_0(k), a_0(k)), \dots, (f_{P-1}(k), a_{P-1}(k))\}$, where $f_p(k)$ and $a_p(k)$ are respectively frequency and magnitude of peak p in frame k , and P is the number of peaks per frame (chosen as 80). At the end, signal $\mathcal{P}(k, p)$, $k = 0, \dots, K - 1$, encapsulates all relevant peak information along the K signal frames.

The obtained transform coefficients $X(k, j)$ and detected peaks $\mathcal{P}(k, p)$ are then utilized to estimate the reference frequency f_{ref} , essential to compute the note frequencies that compose the equal tempered scale. The procedure is described in [10].

To compute the salience function $S(k, p)$ for each frame k and peak p , the frequency locations of peak p and each note from the equal tempered scale with reference frequency f_{ref} are compared. These frequency differences obey a certain law and can, therefore, be used as a theoretical measure to determine which of the detected peaks correspond to sound sources and which correspond to harmonic partials. At this stage, four additional modifications were proposed and added to the method. They are addressed in [5].

At last, a post-processing stage was added to remove ambiguities created by multiples of the existing fundamental frequencies. The procedure is similar to that introduced in [3]. The main idea is to verify if, for each peak p with corresponding frequency f_p , the salience values in each submultiple frequencies f_p/q , for $q = 1, 2, \dots, p$, is significant. Further details can be found in [5]. This procedure with all proposed modifications added is called hereafter Modified DLMP (MDLMP) [5].

An inharmonic model is also included in order to improve fundamental frequency detection. The chosen model is the same used in [4], given by

$$f_h^{f_0}(\beta) = hf_0\sqrt{1 + \beta h^2}, \quad (2)$$

where f_0 is a given fundamental frequency, h is the harmonic number, and β is the inharmonicity coefficient². Since β is not given, it is estimated by performing an exhaustive search: 10 geometrically spaced values in the interval $[10^{-5}, 10^{-3}]$, plus the value $\beta = 0$ [11], are tried, and their corresponding saliences are computed. The chosen salience is the maximum among all obtained values.

III. THE FAN-CHIRP TRANSFORM

This section briefly exposes the concepts of the FChT and clarify each step in its computation.

A. Definition

The FChT is defined in [8] as

$$X(f, \alpha) \triangleq \int_{-\infty}^{\infty} x(t)\phi'_{\alpha}(t)e^{-j2\pi f\phi_{\alpha}(t)}dt, \quad (3)$$

where $\phi_{\alpha}(t)$ is a time linear warping function given by

$$\phi_{\alpha}(t) = \left(1 + \frac{1}{2}\alpha t\right)t. \quad (4)$$

By applying the variable change $\tau = \phi_{\alpha}(t)$ to Equation (3), one obtains

$$X(f, \alpha) = \int_{-1/\alpha}^{\infty} x(\phi_{\alpha}^{-1}(\tau))e^{-j2\pi f\tau}d\tau, \quad (5)$$

where α is the chirp rate parameter, and $\phi_{\alpha}^{-1}(t)$ is given by

$$\phi_{\alpha}^{-1}(t) = -\frac{1}{\alpha} + \frac{\sqrt{1 + 2\alpha t}}{\alpha}; \quad (6)$$

it is assumed that $x(t) = 0$ for $t \leq -1/\alpha$ to avoid aliasing [7].

From Equation (5), it is possible to notice that the FChT is actually the Fourier Transform of a time-warped version of the signal $x(t)$, $x(\phi_{\alpha}^{-1}(t))$. Therefore, the FChT can profit from the fast implementation of the Discrete Fourier Transform, the FFT algorithm [8].

B. Implementation

The implementation is done in short consecutive time frames, usually ranging from 20 to 100 ms, of the sampled audio signal $x(n)$. This procedure, named as Short-Time Fan Chirp Transform (STFChT), considers that the fundamental frequencies may be approximated by a linear chirp within each window [8].

The first step is the time warping caused by $\phi_{\alpha}(t)$, as seen in Equation (5). Since the considered signal is discrete in time, this step should be performed by means of a nonlinear

²This model was originally conceived for string instruments [11], in the case of which β is related to physical properties of the string. Mathematically, its use can be extended to model slightly inharmonic instruments in general.

sampling. Since only samples at time instances nT_s , where T_s is the sampling period, can be reached, an interpolation is carried out [8]. It should be pointed out that, in order to move on with this time warping operation, it is necessary to have a pre-determined α value. The next step is to calculate the FFT of the time warped signal, $x(\phi_\alpha^{-1}(t))$.

The estimation of the chirp rate α is a key step when calculating the FChT of a signal. It is responsible for the resolution of the resulting transform, since a poor estimation yields blurred representations [8]. To estimate α an exhaustive search is done, aiming at obtaining the representation with maximum sparsity. Originally in [8], the classical salience as defined in Equation (1) is used for that purpose. Many instances of the FChT for a number A of pre-determined values of α are firstly calculated, and then the salience function $\rho(f_0, \alpha_a)$ for each α_a , with $a = 1, 2, \dots, A$, and a grid of frequency values f_0^3 , is computed. As a result, a salience plane $\rho(f_0, \alpha)$ is obtained. The point (f_0^*, α^*) corresponding to the maximum value of $\rho(f_0, \alpha)$ is then chosen as the estimated fundamental frequency f_0^* and chirp rate α^* . Further details can be found in [8].

The same procedure can be applied to estimate the chirp rate α when combining the FChT to the timbre-independent salience function. Here, instead of an harmonic accumulation as sparsity measure, the salience method explained in Section II was used.

As in the MDLMP method, a post-processing stage is performed in order to attenuate ambiguities caused by multiples of f_0 . The procedure is the same as in [3]; here it is also applied to attenuate submultiples of f_0 . To remove submultiple spurious peaks in the obtained salience function [8], for each fundamental frequency candidate f_0 , the salience values at its multiples $\rho(qf_0)$, $q \in \mathbb{N}$ such that qf_0 is lower than the maximum considered fundamental frequency candidate, are weighted and subtracted from the salience at f_0 ($\rho(f_0)$) itself. Details and additional considerations are addressed in [8].

IV. EXPERIMENTS AND RESULTS

This section presents a set of informative experiments and their respective results.

A. Experiments

As previously said, the main target of this work is to extract the existing fundamental frequencies in a given audio signal $x(n)$. In particular, it aims to compare regarding efficiency: two time-frequency transforms, the STFT and the STFChT; and two different ways of computing the salience function, the classical and the MDLMP. To this end, a set of four experiments was designed in such a way that each of them contains a possible combination of transform and salience computation: STFT with original salience, STFT with MDLMP, STFChT with original salience, and STFChT with MDLMP.

In order to illustrate the effects of each experiment in the efficiency of f_0 extraction, two signals were considered. The

³The chosen frequency values are fundamental frequency candidates. Here, a grid of 192 geometrically-spaced values per octave were adopted in the range from 100 to 1600 Hz.

first one is a synthetic harmonic signal frequency-modulated by a sinusoid. Its fundamental frequency $f_0(t)$ is given by the following expression

$$f_0(t) = f_1(1 - 2^{1/12})\sin(2\pi f_2 t) + f_1, \quad (7)$$

where f_1 is the central frequency and f_2 , the modulation frequency. The values were chosen to mimic a typical vibrato, as found in singing voice performances, namely, 500 Hz and 6 Hz, respectively. The resulting signal consists of a total of 15 harmonic partials (considering the fundamental frequency), whose amplitudes are inversely proportional to the partial index. The second signal is an excerpt of a musical piece by Villa-Lobos, called *Bachianas Brasileiras* no. 6: a duet for flute and bassoon, often simultaneously played. Their fundamental frequencies were manually tracked and annotated departing from a specially computed STFT to serve as a reference and can be seen in Figure 2 (black and dashed black, respectively).

For the synthetic signal, only one fundamental frequency is present, and therefore a single fundamental frequency is extracted in each experiment. For the real signal, the presence of two simultaneous sources was considered to be known. Therefore, the two most prominent salience values were chosen as the ones corresponding to the existing fundamental frequencies in the signal.

To evaluate the robustness of the methods to background noise, Gaussian white noise was added to the clean signals. For the synthetic signal, six different SNR values were considered, namely, 0 dB, 10 dB, 20 dB, 30 dB, 40 dB and infinite (i.e., no added noise). For the real signal, the experiments were only performed twice: with clean and 30-dB SNR signals. Besides the SNR, the number of samples N of the analysis window was also varied: 1024, 2048, and 4096 for the synthetic, and 2048, 4096, and 8192 for the real signal.

Two different measures were adopted here. The first one is the mean squared error (MSE) between the reference and the estimated fundamental frequency along time. The second is the hit rate considering an error margin of $\pm 3\%$. The MSE has been included because the hit rate for the synthetic signal is usually 100%, and thus a more stringent measure is necessary. The same measure is not applied to the real signal, since it is not possible to quantify the human error inherent to the manual extraction of fundamental frequencies.

The remaining parameters used during simulations are listed below:

- Sampling frequency f_s of 44.1 kHz for both signals;
- Hop size τ of 256 samples;
- $A = 25$ values of the chirp rate α linearly spaced between -4.13 and 4.13 for the synthetic signal, and between -1.03 and 1.03 for the real signal; and
- Number of harmonics n_H of 10 for the synthetic signal and 15 for the real signal for both saliences;

B. Results

The first experiments are performed for the synthetic vibrato. Table I shows the hit rates obtained using the STFT combined to each salience computation method. For each cell, the upper value corresponds to the classic salience approach

and the lower value to the MDLMP method. It is possible to notice that the classical approach largely outperforms the MDLMP method for this choice of time-frequency transformation. For the specific case of $N = 1024$ samples and no added noise, however, it should be noted that the obtained MSE values are 1.42 and 0.29, respectively.

TABLE I

HIT RATES (IN %) FOR SYNTHETIC SIGNAL USING: STFT+ORIGINAL SALIENCE (UPPER VALUES); AND STFT+MDLMP (LOWER VALUES)

SNR (dB)	0	10	20	30	40	∞
N						
1024	79.6 36.1	100 49.0	100 57.7	100 72.2	100 87.8	100 100
2048	63.8 25.1	100 39.8	100 57.0	100 54.2	100 62.6	100 98.0
4096	47.6 19.8	97.9 37.5	98.8 48.6	100 54.3	100 39.5	100 77.8

Table II shows the results obtained with the STFChT combined to each salience computation method. Again, for each cell, the upper value corresponds to the classic salience approach and the lower value to the MDLMP method. Since the MDLMP method presented a hit rate of 100% for all proposed cases, and the same occurs for the classic salience for all SNR values above 10 dB (except for the case $N = 1024$ @ 10 dB SNR, with 99.6%), the MSE values are shown instead. Clearly, the MDLMP case brought a considerable overall improvement if one observes that the corresponding MSE values are lower than the ones obtained with the classical salience for N equal to 1024 and 2048. Since N is a flexible parameter that can be chosen according to the chosen signal and method, the most relevant result is that the minimum overall error was obtained by the combination of the STFChT with the MDLMP salience. It is worth mentioning that the hit rate results obtained for a 0 dB SNR with the classical salience were 82.3%, 78.9%, and 66.7% for analysis window sizes of 1024, 2048, and 4096, respectively.

TABLE II

MSE VALUES FOR SYNTHETIC SIGNAL USING: STFChT+ORIGINAL SALIENCE (UPPER VALUES); AND STFChT+MDLMP (LOWER VALUES)

SNR (dB)	10	20	30	40	∞
N					
1024	282.15 0.64	1.83 0.34	1.80 0.32	1.66 0.31	1.43 0.31
2048	1.01 0.59	0.88 0.56	0.89 0.55	0.91 0.55	0.88 0.55
4096	1.80 6.37	2.04 6.36	2.12 6.38	1.90 6.40	1.93 6.38

Now, the same experiments are performed for the real signal, the duet excerpt. Tables III and IV show the hit rates for the three chosen sizes of analysis windows (2048, 4096, and 8192), when combining each salience computation method to the time-frequency transforms STFT and STFChT, respectively. They contain the individual rates for each instrument, flute and bassoon, as well as the total hit rate. In

general, one can notice that the MDLMP method overcomes the classical salience computation method when it comes to estimating the two existing pitches. This is due to the coincidence of harmonics between the two fundamental frequencies. When performing the harmonic accumulation, the saliences obtained for common submultiples between both fundamental frequencies can end up having a greater value than the fundamental frequencies themselves. The post-processing stage mentioned in Section III does not help in this case, since the true fundamental frequencies are attenuated due to the high salience values of coincident submultiples. Furthermore, while the salience mean decreases, its variance increases in frequency⁴ [8], thus negatively affecting the flute's higher fundamental frequency estimation, which can be verified by the small hit rate success for this instrument.

TABLE III

HIT RATES (IN %) FOR REAL SIGNAL USING: STFT+ORIGINAL SALIENCE (UPPER VALUES); AND STFT+MDLMP (LOWER VALUES)

N	Flute		Bassoon		Total	
	Clean	30 dB	Clean	30 dB	Clean	30 dB
2048	37.6 86.1	36.1 86.3	73.9 63.8	71.5 63.8	55.5 75.1	53.6 75.1
4096	26.3 90.1	23.5 89.2	82.5 64.4	80.5 63.8	54.1 77.4	51.7 76.6
8192	21.8 87.8	19.4 88.2	84.7 66.1	82.7 65.7	52.9 77.1	50.8 77.1

TABLE IV

HIT RATES (IN %) FOR REAL SIGNAL USING: STFChT+ORIGINAL SALIENCE (UPPER VALUES); AND STFChT+MDLMP (LOWER VALUES)

N	Flute		Bassoon		Total	
	Clean	30 dB	Clean	30 dB	Clean	30 dB
2048	37.9 86.6	35.4 86.4	77.1 64.4	75.3 64.6	57.3 75.6	55.2 75.6
4096	28.5 89.2	22.7 88.0	86.7 64.6	84.4 63.5	57.3 77.0	53.3 75.9
8192	24.9 86.0	20.7 85.7	90.7 67.4	86.2 66.1	57.5 76.8	53.1 76.0

When observing the bassoon's hit rates for the MDLMP method, one notices that these values are relatively lower than the ones obtained by using the classical salience function. Figure 2 shows the estimated fundamental frequencies in red and blue for both salience functions. The upper figure shows the results for the classical salience approach, while the lower figure, for the MDLMP method. It is possible to see that for time instances between 1.4 and 2.5 seconds, the latter method was not capable of detecting the bassoon's fundamental frequency, thus decreasing the corresponding hit rates. The same issue was observed for the three sizes of analysis windows, and its explained by the fact that the second harmonic of

⁴In order to soften the negative impact caused by the mean decrease, a post-processing stage was added, where a noise floor is first estimated by means of the the SSE method (detailed in [9]) and then subtracted from the salience function.

that specific note is more prominent than the fundamental frequency itself, and the post-processing stage mentioned in Section II was not able to eliminate the ambiguity.

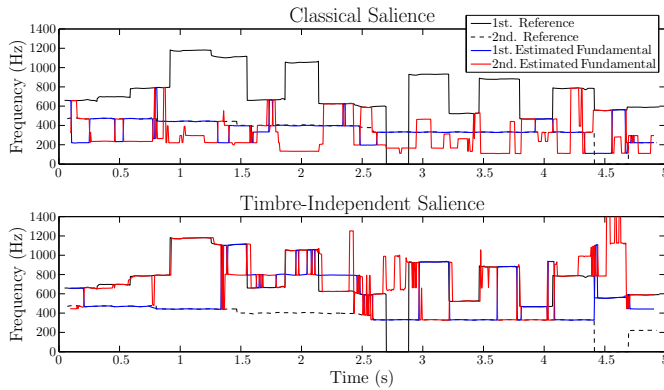


Fig. 2. Comparison between classical and timbre-independent saliences using an analysis window of 8192 samples. The first and second estimated fundamental frequencies (in blue and red, respectively) using the STFChT with the classical (upper figure) and the timbre-independent (lower figure) saliences are shown.

It is also worth mentioning that no significant differences were observed for the MDLMP method combined to the STFChT when compared to the same method combined to the STFT. Actually, slightly better results were obtained for the bassoon and slightly worse for the flute, resulting in (almost) no changes in the total hit rates. These results make sense in some level, since the bassoon was performing a vibrato, which can profit from the STFChT, and the flute played fairly stationary notes, meaning that the STFT (or a STFChT where the chirp rate α was set to zero) is the best choice. Another possible explanation is that the melodies of both existing fundamental frequencies were manually extracted and can therefore present some errors. Further investigation should be made in order to determine the best way of profiting from the sparsity provided by the STFChT when combined to the MDLMP method.

When adding Gaussian white noise to the clean signals, it is possible to notice that the hit rates attained by the classical salience decreased around 2% and 3.5% when using the STFT and STFChT, respectively. Considering that the mean hit rates for both cases were around 54% and 57%, respectively, this means a decrease of around 4% and 6%, respectively. In the case of the MDLMP method, these decreases were around 0.4% and 0.8% for the STFT and the STFChT, respectively, which are considerably smaller.

At last, it is important to point out that the computational complexity is considerably lower for the MDLMP method when compared to the classical salience. While the latter method assigns a salience value to each considered fundamental frequency candidate (769 values in the present setup) the MDLMP computes it for only a pre-determined number of detected peaks ($P = 80$ in the present setup).

V. CONCLUSIONS

In this paper, the performance of an f_0 extraction algorithm for polyphonic music combining one of two time-frequency transforms, the STFT and the STFChT, to one of

the salience functions, the original and the MDLMP, was evaluated. Its main innovation lies in using STFChT with the MDLMP method for fundamental frequency estimation. Some experiments were carried out in order to analyse the precision in terms of MSE and hit rate, and also to evaluate the performance degradation caused by additive Gaussian white noise.

As results, one can notice that better hit rates were obtained, in general, for the MDLMP when compared to the classical salience, specially for noise-corrupted signals. When using the STFChT+MDLMP method, better results were obtained for a synthetic monophonic signal when compared to the STFT; the classical salience outperformed the MDLMP only when combined to the STFT in the presence of noise.

When comparing the obtained hit rates for the real signal by using the MDLMP for both time-frequency transforms, no considerable improvements were introduced by the STFChT. This indicates that some further work is needed in order to find out better ways to combining the FChT to the timbre-independent salience. In this context, of course, more encompassing tests must be performed to better characterize the proposed STFChT+MDLMP strategy.

REFERENCES

- [1] A. Klapuri and M. Davy, Eds., *Signal Processing Methods for Music Transcription*, Springer, New York, USA, 2006.
- [2] Y. Li and D. Wang, "Separation of singing voice from music accompaniment for monaural recordings," *IEEE Transactions on Audio, Speech, and Language Processing*, v. 5, no. 4, pp. 1475–1487, May 2007.
- [3] M. Képesi and L. Weruaga, "Adaptive chirp-based time-frequency analysis of speech signals," *Speech Communication*, v. 48, no. 5, pp. 474–492, May 2006.
- [4] A. Degani, R. Leonardi, P. Migliorati, and G. Peeters, "A pitch salience function derived from harmonic frequency deviations for polyphonic music analysis," in *Proceedings of the 17th Conference on Digital Audio Effects (DAFx-14)*, Erlangen, Germany, September 2014.
- [5] I. F. Apolinário and L. W. P. Biscainho, "An improved method for determining pitch relevance in polyphonic audio," *Anais do 14o Congresso Nacional da AES Brasil*, São Paulo, Brazil, May 2016 (to appear).
- [6] U. Zölzer, Ed., *Digital Audio Effects*, Wiley, Chichester, United Kingdom, 2nd edition, 2011.
- [7] L. Weruaga and M. Képesi, "The fan-chirp transform for non-stationary harmonic signals," *Signal Processing*, v. 87, n. 6, pp. 1504–1522, June 2007.
- [8] P. Cancela, E. López, and M. Rocamora, "Fan-chirp transform for music representation," in *Proceedings of the 11th International Conference on Digital Audio Effects (DAFx-10)*, Graz, Austria, September 2010, pp. 1–8.
- [9] N. Laurenti, G. De Poli, and D. Montagner, "A nonlinear method for stochastic spectrum estimation in the modeling of musical sounds," *IEEE Transactions on Audio, Speech, and Language Processing*, v. 15, no. 2, pp. 531–541, February 2007.
- [10] K. Dressler and S. Streich, "Tuning frequency estimation using circular statistics," in *Proceedings of the 8th International Conference on Music Information Retrieval (ISMIR)*, pp. 2–5, Vienna, Austria, September 2007.
- [11] H. Fletcher, E. D. Blackham, and R. Stratton, "Quality of piano tones," *Journal of Acoustical Society of America*, v. 34, no. 6, pp. 749–761, June 1962.