INTERPRETATION AND CONTROL IN AM/FM-BASED AUDIO EFFECTS

Antonio José Homsi Goulart
Computer Music Research Group
University of São Paulo
São Paulo, Brazil
ag@ime.usp.br

Marcelo Queiroz
Computer Music Research Group
University of São Paulo
São Paulo, Brazil
mqz@ime.usp.br

Joseph Timoney
Sound and Digital Music Technology Group
Maynooth University
Maynooth, Ireland
joseph.timoney@mu.ie

Victor Lazzarini
Sound and Digital Music Technology Group
Maynooth University
Maynooth, Ireland
victor.lazzarini@mu.ie

ABSTRACT

This paper is a continuation of our first studies on AM/FM digital audio effects, where the AM/FM decomposition equations were reviewed and some exploratory examples of effects were introduced. In the current paper we present more insight on the signals obtained with the AM/FM decomposition, intending to illustrate manipulations in the AM/FM domain that can be applied as interesting audio effects. We provide high-quality AM/FM effects and their implementations, alongside a brief objective evaluation. Audio samples and codes for real-time operation are also supplied.

1. INTRODUCTION

In previous papers [1] [2] we presented our first studies on AM/FM Digital Audio Effects. The AM/FM decomposition enables an analysis-processing-resynthesis approach to work with audio signals. Effects can be implemented by manipulating signals obtained with a decomposition scheme that unravels the original audio time-based representation to an analogous representation based on a pair of new time-based signals with complementary information.

The AM/FM decomposition was firstly adopted in musical processing for synthesis purposes [3], where similarities with the FM synthesis [4] where drawn. More recently, a lot of research was devoted to AM/FM for speech analysis, in the area of works known as Modulation Filtering [5] [6] [7]. Back to the context of music signal processing, in the Modulation Vocoder series of works [8] [9] [10] the decomposition was adopted in order to explore applications like audio codification (compression), control of roughness of audio signals, and pitch transposition. AM/FM decomposition was also used as an extension of sinusoidal modelling for audio analysis/synthesis purposes [11].

In [1] the impact of smoothing in the AM/FM domain was assessed considering different configurations of smoothers and different psychoacoustics metrics. Then, in [2] we investigated effects based on manipulating the AM/FM domain signals with well-known time-domain manipulations of well established effects like octaver, chorus, wah-wah, etc. In the present paper, Section 2 will briefly review the AM/FM Hilbert-based decomposition with an intuitive explanation about the technique, so that manipulations in the AM/FM domain can lead to the design of interesting audio effects. In Section 3 AM/FM effects will be presented alongside audio examples that are available for downloading[1], where the reader will be able to assess the quality of the effects obtained from this study. In Section 4 a brief evaluation of the effects, based on comparisons using audio descriptors, will be presented. Finally, we conclude and point our current and future work. Audio files will be referenced in the paper with the symbol [▶filename].

2. AM/FM DECOMPOSITION

2.1. Envelope and instantaneous frequency

The idea behind the AM/FM decomposition is to understand the input signal as a single sinusoidal tone modulated both in amplitude (AM) and frequency (FM). Given an input signal $x(t)$, we want to find a pair of functions $(a(t), f(t))$ such that

$$x(t) = a(t)\cos\left(\int_0^t f(\tau)d\tau\right). \quad (1)$$

The amplitude modulator signal $a(t)$ is estimated with the decomposition as an envelope of the input signal, and the frequency modulator signal $f(t)$ is estimated as the instantaneous frequency (IF) of the input signal. In order to apply audio effects we might process $a(t)$ or $f(t)$ and consider the altered versions $a_{FX}(t)$ and $f_{FX}(t)$ in a resynthesis process

$$x_{FX}(t) = a_{FX}(t)\cos\left(\int_0^t f_{FX}(\tau)d\tau\right). \quad (2)$$

Notice that the argument for the cosine in Eq. 1 is the instantaneous phase, which is the integral of the instantaneous frequency. This is tied to the concept of a phasor (Figure 1), in which the phase (current angle) is given by increments from an initial position (initial angle) in the unit circle.

The increments in the phase are represented by the integral in Eq. 1. If $a(t)$ and $f(t)$ are constant, we will have equal steps around the circle (unit circle if $a(t) = 1, \forall t$), and thus a sinusoid will be obtained with the projection of the phasor onto the x axis (as in Fig. 1). However, if $a(t)$ or $f(t)$ vary, a different kind of signal will be obtained. This leads us to a useful interpretation for the IF, understanding this value as the frequency of a sinusoid that locally (at each time instant t) fits the original signal $x(t)$ [12].

[1] https://www.ime.usp.br/~ag/dl/dafx18.zip

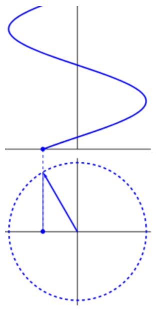

Figure 1: *A regular phasor.*

It is interesting to notice the local aspect of the instantaneous frequency, estimated from an infinitesimal neighborhood of each sample, as opposed to frequencies of sinusoidal components present in the signal spectrum, which have a global scope (the analysis window). It should be noted that the IF might even not be present in the spectrum of a signal, and is sometimes higher than the highest component present in a signal [13]. While additive synthesis [14] would provide a classic example of how to think globally about a signal, there are many situations where a local/instantaneous model of the signal is more appropriate, e.g. in the operation of adaptive devices such as limiters, which measure/change signal values constantly within a feedback loop (lacking knowledge of the whole process) [15]. In order to grasp a good intuition for developing AM/FM effects, this local information is the main concern of this study.

2.2. Ambiguity in the decomposition

Different techniques are available for obtaining an AM/FM decomposition, and as we are unraveling a single signal to a combination of two other signals, an inherent ambiguity permeates this decomposition. Given an input signal $x(t)$ we can find both

$$x(t) = a(t) \cos(\phi(t)) \tag{3}$$

and

$$x(t) = b(t) \cos(\theta(t)) \tag{4}$$

in such a way that $b(t) \neq a(t)$ and $\theta(t) \neq \phi(t)$ [13]. Actually we might think of two extreme (and undesirable) cases for the decomposition:

- $a(t) = x(t), \quad \forall t$; and
  $\phi(t) = 0 \rightarrow \cos(\phi(t)) = 1, \quad \forall t$;
  in this case all the information is coded in the AM portion. The resynthesis would be represented by a pure amplitude modulation, as the cosine value would be constant;
- $a(t) = 1, \quad \forall t$; and
  $\phi(t) = \cos^{-1}(x(t)), \quad \forall t$;
  in this case the information would go to the FM portion of the decomposition. The resynthesis would be a pure frequency modulation, as the envelope would be constant.

In the development of audio effects we are not really interested in these extreme cases, for in such cases we could work directly on the original time-domain signal. What we usually want is a decomposition that allocates non-trivial information both to the AM and FM portions, so we can develop processing routines that will bring interesting modifications to the dry signal after resynthesis.

2.3. Implementation

In our work we focused on the analytic signal based decomposition, although other techniques, for instance based on energy separation [16] [17] [18] are also available. The analytic signal is a complex signal without any negative frequency components [19]. Given a real signal, its analytic counterpart shows a similar spectrum considering the positive frequencies, but a null contribution from the negative frequencies. For instance, a regular sinusoidal signal $\cos(\omega_0 t)$ contains two components in its spectrum, localized at $+\omega_0$ and $-\omega_0$ [20]:

$$\cos(\omega_0 t) = \frac{1}{2}\left(e^{i\omega_0 t} + e^{-i\omega_0 t}\right). \tag{5}$$

Notice that if we consider only the positive component we get the regular phasor represented in Fig. 1 as the analytic signal for $\cos(\omega_0 t)$.

Now, by eliminating the negative components of any real signal $x(t)$ we get its analytic signal $z(t)$ as

$$z(t) = \frac{1}{2\pi}\int_0^{+\infty} X(\omega) e^{i\omega t} d\omega, \tag{6}$$

where $X(\omega)$ is the Fourier Transform of $x(t)$ [20]. Eq. 6 can be thought as a superposition of an infinite number of phasors, each of them spinning with its own frequency ω and radius $X(\omega)$. Figure 2 represents this view, considering three phasors; the projection onto the x axis gives the original signal $x(t)$.

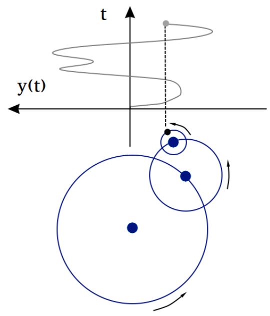

Figure 2: *Superposition of three phasors with different frequencies and radii.*

Considering this negative frequency components elimination perspective, one of the possibilities to obtain the analytic signal is via the Fourier Transform: we can transform the real signal $x(t)$, attribute zero to the negative portion of the spectrum, then apply the inverse transform to obtain $z(t)$ [21]. Another possibility is by applying the Hilbert Transform to $x(t)$, which gives $\hat{x}(t)$, a quadrature version of $x(t)$, where all the components are shifted by 90° [22]. We then can build the analytic signal as

$$z(t) = x(t) + i\hat{x}(t), \tag{7}$$

where $i = \sqrt{-1}$.

The AM/FM decomposition is a matter of finding the envelope and the instantaneous frequency of the analytic signal. Notice that $z(t)$ can be written as

$$z(t) = x(t) + i\hat{x}(t) = a(t)e^{i\phi(t)}, \tag{8}$$

where the envelope is given by

$$a(t) = \sqrt{x^2(t) + \hat{x}^2(t)} = |z(t)|, \tag{9}$$

the instantaneous phase is given by

$$\phi(t) = \arctan\left(\frac{\hat{x}(t)}{x(t)}\right), \tag{10}$$

and by differentiating this quantity we obtain the IF

$$f(t) = \dot{\phi}(t) = \frac{x(t)\dot{\hat{x}}(t) - \dot{x}(t)\hat{x}(t)}{x^2(t) + \hat{x}^2(t)}. \tag{11}$$

Eq. 8 helps us visualize the relation between the concepts of analytic signal, envelope and instantaneous phase and frequency. In the case of a sinusoidal $x(t)$ we will have $z(t)$ as a simple harmonic motion, with constant radius and frequency, so the increments in the angle are always the same. However, for a more general $x(t)$, $z(t)$ will exhibit unequal increments, i.e., in each sample the angle covered around the circle will not be the same; likewise, the radius will not be constant, so we will have a movement that alternates between shrinking and expanding spirals. Notice that the projection of this movement onto the x axis generates $x(t)$.

3. NEW AM/FM DAFX

In order to apply AM/FM effects we must modify $a(t)$ and/or $f(t)$ and then proceed to a resynthesis process. For the modifications we will be dealing with filters, compressors/expanders, and modulators. As these modifications depend on the choice of values for parameters like thresholds, cut-off frequencies, etc., it is important to have an idea about the ranges involved in the original signal. The values will then be chosen according to musical intentions.

In this paper our examples will be focused on a short guitar phrase consisting of a bend and a vibrato[2]. Its varying fundamental frequency represents an important test for the AM/FM decomposition. The waveform in shown in Fig. 3; the envelope and instantaneous frequency signals estimated with the method described in Section 2 are shown in Figures 4 and 5, respectively. Notice that extreme values for the IF are typically associated with the occurrence of very small values in the envelope (the denominator in Eq. 11 is essentially the envelope squared). Figure 6 shows a zoomed view of the IF, up to the value of 1200 Hz, showing that for the length of the signal the IF shows a trend going from (around) 700 Hz to (around) 500 Hz, contaminated with spikes at instants where the envelope values fade out.

The audio file [▶resynth-hilb] was created with an AM/FM analysis-resynthesis process, i.e., no manipulations on the envelope and instantaneous frequency were applied. Notice that the analysis is transparent, i.e., the reconstructed signal is identical to the original audio file ([▶bend-vibrato]).

3.1. Effects based on filtering

By filtering signals, we are choosing which components will remain unaltered and which will be amplified or attenuated to some extent [23]. For instance, by not allowing high frequencies in the signal we will prevent fast variations to occur, but slow fluctuations remain unaltered.

[2]In the audio files provided, other examples with different musical instruments and phrases are also available.

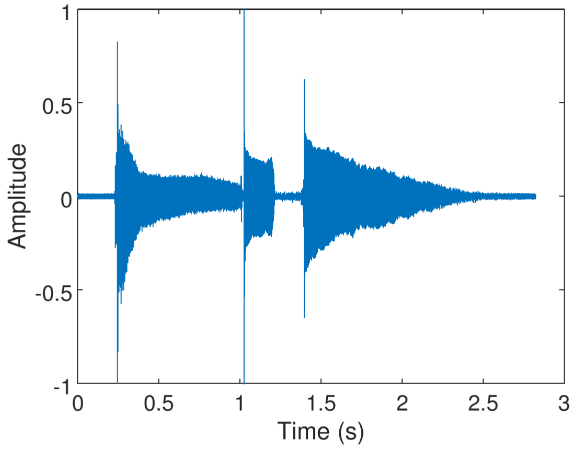

Figure 3: *Waveform of guitar phrase.*

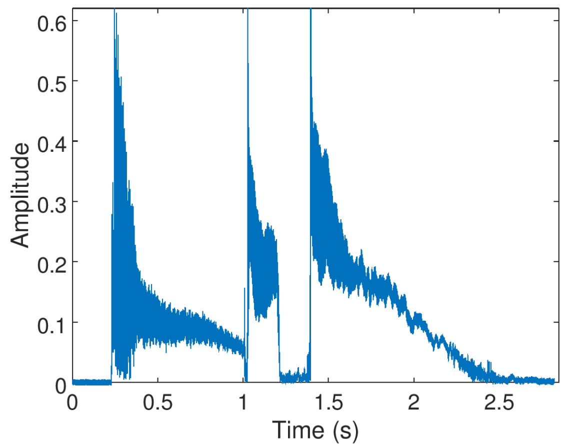

Figure 4: *Envelope of guitar phrase.*

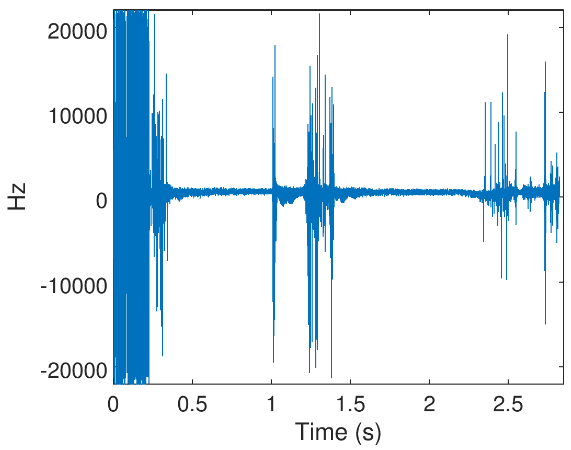

Figure 5: *Instantaneous frequency of guitar phrase.*

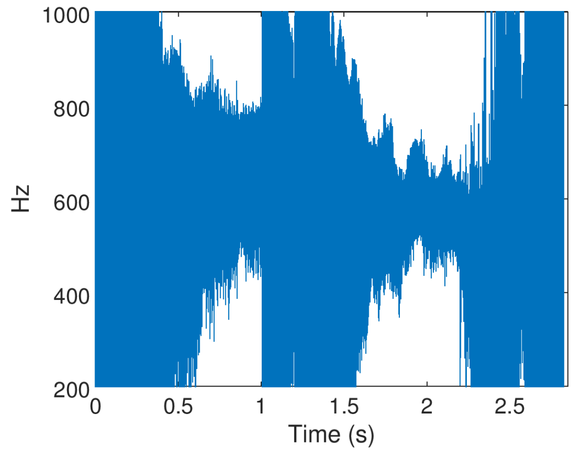

Figure 6: *Instantaneous frequency of guitar phrase (zoom).*

However, modifications in the IF of signals generally have a different meaning. If we, for instance, low-pass filter an IF signal, we will prevent sharp transitions occurring in the IF signal. This will result in a muffled sound, with less articulation, as the IF signal will only keep its slow variations. Thus, in the resynthesis, the phasor will have its increments constrained. We can check this effect by comparing the audio files [▶lowp-if-1000] (cut-off frequency set at 1 KHz) and [▶lowp-if-500] (cut-off at 500 Hz) with the dry signal [▶resynth-hilb]. Fig. 6 shows that the IF values lie around 600 Hz, so the cut-off at 1 KHz will not cause a huge impact, but at 500 Hz it will[3].

Notice that the range of variations in the IF signal is not the matter here, but the frequency with which they occur. Large variations will still exist, but will happen slowly over time. An interesting effect can be achieved by setting the cut-off at extreme values, e.g. 1 Hz. The file [▶lowp-if-1] will reveal a chirp, because the sweeping through the IF signal will happen with a limited speed.

By low-pass filtering the envelope component, a different kind of effect is achieved. High frequencies in dynamics are related to percussive sounds, which brings the sensation of the onset of a sound. So, by limiting the envelope only to low frequencies, sounds with a smooth onset, resembling a bowed violin, are obtained, as in the file [▶env-lowpf-10] (Butterworth low-pass filter with cut-off frequency at 10 Hz, applied to the envelope).

3.2. Effects based on dynamics processing

Instead of acting on the range of frequencies present in a signal, the manipulation of the dynamics of the envelope and IF signals imparts a selection of the actual values that we allow for these signals. For instance, if we use a limiter to prevent values for the IF higher than an specific threshold, we will prevent, in the resynthesis, angle increments higher than this threshold. In such a way, we can condition the excursion of the signals within a desired range.

Depending on the configurations, similar results can be obtained by using distortion, limiting, or compression [23]. These effects are all used to attenuate large values (higher than a threshold) in the input signal, differing only in the way they operate.

[3] Check also the audio examples considering different instruments (they come in separate folders and are named using the same convention).

Dynamics processing of the envelope or the audio signal itself will have similar results, but working on the IF signal brings interesting musical applications. For instance, as we know (Fig. 6) our IF values lie around 600 Hz, so applying a distortion with a threshold at 400 Hz ([▶if-ortion-400]) will result in a sound similar to the one obtained by changing the IF value to a constant equal to 400 Hz ([▶fix-if-400]). This will lead, in this example, to a perception of a drone note between a G4 and a G♯4.

3.3. Effects based on modulation

Another family of effects that we will describe is based on altering the value of the IF signal with a LFO (Low Frequency Oscillator) approach. We can both ring modulate the IF, i.e. directly multiply it with a modulator, or apply classic amplitude modulation instead, where the modulation will occur around the IF signal (an offset is added to the modulator signal) [23].

The former case represents the possibility for a very aggressive effect. As we saw in the previous sections, acting on the IF will probably result in pitch modifications. Therefore, a direct multiplication of the IF will result in aggressive transposition in the resynthesised signal.

A slow setting for the modulation frequency will result in a perception of glissandos ([▶gliss-if] [4]), while a higher modulation frequency will bring a very unstable kind of pitch variation, since the rapid excursion will sweep a range from the deep lows to the top highs ([▶aggress-if] [5]).

In the context of a classic amplitude modulation applied to the instantaneous frequency signal, different types of effects might be achieved depending on the modulator signal configuration. A deep modulation will tend to produce aggressive effects as well, but a more gentle variation might be interesting to create a detuning effect ([▶d-if-tune] [6]) or a vibrato ([▶v-if-brato] [7]).

4. EVALUATION

In addition to the intuition established with the theory and audition of the audio samples, an objective evaluation based on audio descriptors helps in the development and refinement of the effects.

Audio descriptors are quantities extracted directly from the audio signal, and might be related to models (e.g. psychoacoustic models) or to mathematical manipulations in order to derive some alternative perspective on the signal [24]. We will analyse two descriptors:

- Spectral centroid: indicates the position for the “center of mass” of a signal spectrum, the point that divides the spectrum in two balanced portions [25]. This quantity is strongly related to the brightness of a sound;
- RMS (root mean square): indicates the power, given by the averaged sum of the squared values of the signal [24], being therefore related to the perception of intensity.

The spectral centroid and RMS descriptors were extracted for both the dry signal (the audio with no effect applied) and the wet signals (the resynthesized signals) and compared. The Essentia [26] library was used via its Python API. All the audio samples were generated with Csound [27] [28].

[4] Glissando effect implemented via IF processing.
[5] Aggressive pitch modulation implemented via IF manipulation.
[6] Detuning effect implement via IF processing.
[7] Vibrato effect implemented via IF processing.

Despite the fact that low-pass filtering of the IF would not affect the range of values the IF signal, but only how fast variations can occur, Figure 7 shows that the operation lowered the spectral centroid, as a direct low-pass filtering of the original signal would do. The RMS (Figure 8) is not affected by such an operation; it is actually more influenced by manipulations on the envelope.

Figures 9 and 10 show the influence in the spectral centroid when the IF signal is fixed at a constant value. Knowing that the IF signal in our guitar signal example varies around 600 Hz, a 2 KHz fixed IF results in a higher centroid when compared to the dry signal, and an IF fixed at 200 Hz results in a lower centroid.

The gentle modulations effects do not seem to have much impact on the location of the spectral centroid, because the variations are small and are close to the original IF. Figure 11 shows the *v-ifbrato* case. A little less smooth version of the effect is represented in Figure 12. The slow ring modulation commences with null values for the modulator causing the centroid to start at null values, and the centroid progressively reaches the original centroid values as the modulator values become close to 1. Figure 13, however, shows a more extreme effect where the ring modulation is deep and fast, which causes the centroid values to oscillate between zero and the original centroid values for the dry signal.

5. CONCLUSIONS

The incoherent mono-component Hilbert Transform based decomposition has received lots of criticism, specially in the speech analysis literature. We emphasize that the decomposition is transparent, i.e., the signal obtained with an analysis-resynthesis procedure is identical to the original, but the intermediate step of processing in the AM/FM domain can be dangerous, regarding the potential of introducing noise or artifacts into the resynthesised version.

However, we generated many examples where the resynthesis produced a clean signal, without any noise that would invalidate its musical usefulness. In some cases where artifacts do appear after resynthesis, the so-called intelligibility requirement (important in speech analysis) could be loosened in many musical contexts, so the noisy sonorities obtained might also be interestingly explored.

There is a huge potential for the AM/FM approach to be considered as an alternative to other classic modulation effect techniques, like the vocoder [29]. The scheme considered here can be easily applied to a melodic signal, i.e., signals with melodic lines like those created with many wind instruments, guitar solos, bass lines, voice, among many other examples. However, a lot of care should be taken with the envelope and instantaneous frequency signals interpretation, which should not be regarded as simple amplitude (AM, envelope) and frequency/pitch (FM, instantaneous frequency) components, but instead should be acknowledged as signals controlling a single sinusoidal oscillator, which adapts itself to represent all sorts of complex musical signals.

In this paper we focused on low-pass and dynamic range compression examples, but high-pass, band-pass, and expansion of the dynamic range of the envelope or IF work in a similar way. Emphasis was given to processing of the IF signal in the AM/FM representation, but processing the envelope can also lead to interesting effects, especially exploring roughness issues.

The low-pass IF filtering examples were important not only to obtain audio effects per se, but also for showing that the perceptual brightness of a sound is somehow linked to the possibility of the IF signal to vary quickly in an AM/FM representation, in a sense that limiting this speed will produce a perception of a muffled sound. This would not be evident a priori, since the IF can still sweep through all the possible frequencies after being filtered.

Dynamics processing and modulations in the IF will result in pitch modifications on the original signal, a side-effect which should be taken into account when implementing these effects, specially considering the musical scenario where it will be used. Modulation can also modify the pitch of the signal.

The configuration for the parameters' values is challenging, since the very same technique can result in effects from the very subtle to the very aggressive. Evidently, both aesthetic concerns and practicalities of the instruments (dry audio signals sources) will come into play in the design of AM/FM DAFx.

6. FUTURE WORK

We are currently working on effects that explore the separation of the input signal in different bands. After the separation, the AM/FM decomposition can be applied to all bands, and multilayer effects based on different band-wise configurations of the same effect might lead to interesting results. We are also proceeding to a more thorough evaluation, considering more objective parameters that can elucidate our comprehension between the decomposition-manipulation of the estimated signals and the resulting effect. Subjective evaluations considering instrument players, DJs, and also music appreciators might lead to useful results about the quality and musicality of AM/FM-based effects.

7. ACKNOWLEDGMENTS

This work was partially supported by CAPES (proc. num. 8868-14-0) and was partially realized during Antonio Goulart's internship period at Maynooth University. We are grateful for the insightful comments by the reviewers.

8. REFERENCES

[1] Antonio José Homsi Goulart, Joseph Timoney, Victor Lazzarini, and Marcelo Queiroz, "Psychoacoustic impact assessment of smoothed am/fm resonance signals," in *Proceedings of the Sound and Musical Computing Conference*, Maynooth, Ireland, July 2015.

[2] Antonio José Homsi Goulart, Joseph Timoney, and Victor Lazzarini, "AM/FM DAFx," in *Proceedings of International Conference on Digital Audio Effects (DAFx)*, Trondheim, Norway, December 2015.

[3] James Justice, "Analytic signal processing in music computation," *IEEE Transactions on acoustics, speech and signal processing*, vol. ASSP-27, no. 6, pp. 670–684, 1979.

[4] John Chowning, "The synthesis of complex audio spectra by means of frequency modulation," *Journal of the Audio Engineering Society*, vol. 21, no. 7, pp. 526–534, 1973.

[5] S. Schimmel and L. Atlas, "Coherent envelope detection for modulation filtering of speech," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, 2005.*, March 2005, vol. 1, pp. 221–224.

[6] Qin Li and Les Atlas, "Over-modulated am-fm decomposition," in *Proceedings of the SPIE - Advanced Signal Processing Algorithms, Architectures, and Implementations*, Bellingham, WA, 2004, pp. 172–183.

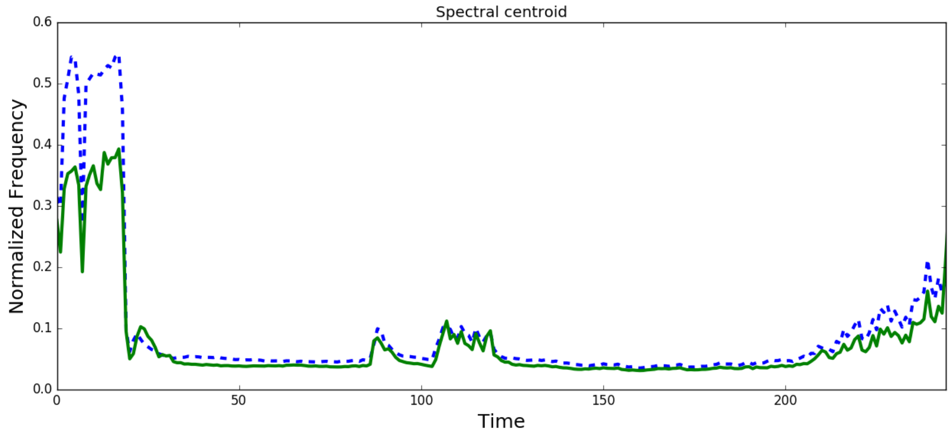

Figure 7: *Spectral centroid after low-pass (Butterworth) filtering the IF with a cut-off frequency at 1 KHz (solid green) versus spectral centroid of the dry signal (dashed blue).*

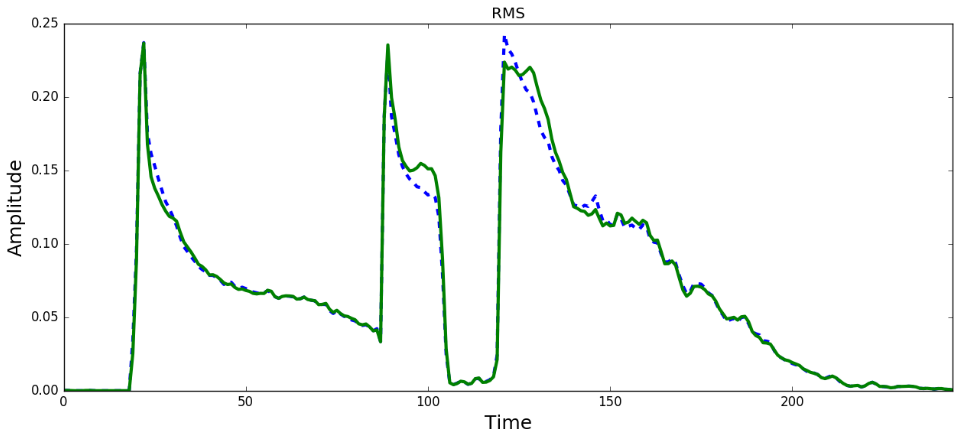

Figure 8: *RMS after low-pass (Butterworth) filtering the IF with a cut-off frequency at 1 KHz (green) versus RMS of the dry signal (blue).*

[7] P. Clark and L. Atlas, "Time-frequency coherent modulation filtering of nonstationary signals," *IEEE Transactions on Signal Processing*, vol. 57, no. 11, Nov 2009.

[8] Sascha Disch and Bernd Edler, "An amplitude and frequency-modulation vocoder for audio signal processing," in *Proceedings of the International Conference on Digital Audio Effects (DAFX-08)*, Espoo, Finland, September 2008.

[9] Sascha Disch and Bernd Edler, "An iterative segmentation algorithm for audio signal spectra depending on estimated local centers of gravity," in *Proceedings of the International Conference on Digital Audio Effects (DAFX-09)*, Como, Italy, September 2009.

[10] Sascha Disch and Bernd Edler, "An enhanced modulation vocoder for selective transposition of pitch," in *Proceedings of the International Conference on Digital Audio Effects (DAFX-10)*, Graz, Austria, September 2010.

[11] M. Caetano, G. P. Kafentzis, A. Mouchtaris, and Y. Stylianou, "Full-band quasi-harmonic analysis and synthesis of musical instrument sounds with adaptive sinusoids," *Applied Sciences*, vol. 6, no. 5, 2016.

[12] B. Boashash, "Estimating and interpreting the instantaneous frequency of a signal-part 1: Fundamentals," *Proceedings of the IEEE*, vol. 80, no. 4, Apr 1992.

[13] Patrick J. Loughlin, "Do bounded signals have bounded amplitudes?," *Multidimensional Systems and Signal Processing*, vol. 9, pp. 419–424, 1998.

[14] Charles Dodge and Thomas Jerse, *Computer Music: Synthesis, composition and performance*, Schirmer Books, New York, NY, USA, 2nd edition, 1997.

[15] D.E. Vakman and L.A. Vainshtein, "Amplitude, phase, frequency - fundamental concepts of oscillation theory," *Sov. Phys. Usp.*, vol. 20, no. 12, pp. 1002–1016, December 1977.

[16] P. Maragos, J. F. Kaiser, and T. F. Quatieri, "On amplitude and frequency demodulation using energy operators," *IEEE Transactions on Signal Processing*, vol. 41, no. 4, pp. 1532–1550, Apr 1993.

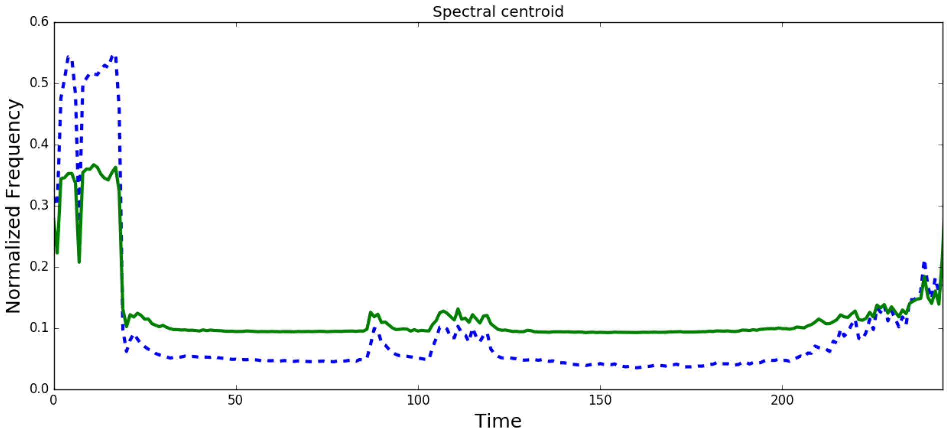

Figure 9: *Spectral centroid after fixing the IF at 2 KHz (green) versus spectral centroid of the dry signal (blue).*

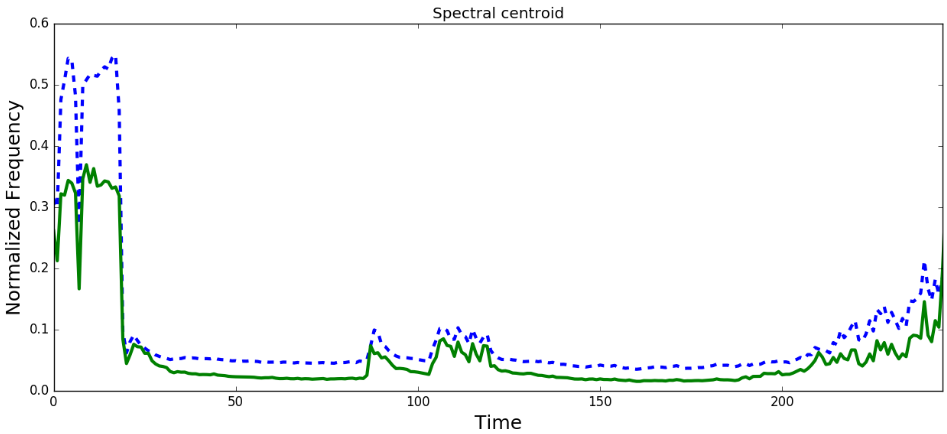

Figure 10: *Spectral centroid after fixing the IF at 200 Hz (green) versus spectral centroid of the dry signal (blue).*

[17] P. Maragos, J. F. Kaiser, and T. F. Quatieri, "Energy separation in signal modulations with application to speech analysis," *IEEE Transactions on Signal Processing*, vol. 41, no. 10, pp. 3024–3051, Oct 1993.

[18] Alexandros Potamianos and Petros Maragos, "A comparison of the energy operator and the hilbert transform approach to signal and speech demodulation," *Signal Processing*, vol. 37, no. 1, pp. 95 – 120, 1994.

[19] Julius Smith, *Mathematics of the Discrete Fourier Transform, with Audio Applications*, W3K Publishing, 2nd edition, 2007.

[20] A.V. Oppenheim and R.W. Schafer, *Digital signal processing*, Prentice Hall, New Jersey, USA, 1975.

[21] B. Picinbono, "On instantaneous amplitude and phase of signals," *IEEE Transactions on Signal Processing*, vol. 45, no. 3, pp. 552–560, Mar 1997.

[22] Stefan Hahn, *Hilbert Transforms in Signal Processing*, Artech House, Norwood, MA, 1996.

[23] Udo Zölzer, Ed., *DAFx: Digital Audio Effects*, Wiley & Sons, 2nd edition, 2011.

[24] Nick Collins, *Introduction to Computer Music*, Wiley and Sons, 2010.

[25] James Beauchamp, "Synthesis by spectral amplitude and 'brightness' matching of analyzed musical instrument tones," *J. Audio Eng. Soc*, vol. 30, no. 6, 1982.

[26] Dmitry Bogdanov, Nicolas Wack, E. Gómez, Sankalp Gulati, Perfecto Herrera, O. Mayor, Gerard Roma, Justin Salamon, J. R. Zapata, and Xavier Serra, "Essentia: an audio analysis library for music information retrieval," in *International Society for Music Information Retrieval Conference (ISMIR)*, Curitiba, Brazil, 04/11/2013 2013, pp. 493–498.

[27] Richard Boulanger, Ed., *The Csound Book: Perspectives in Software Synthesis, Sound Design, Signal Processing, and Programming*, MIT Press, 2000.

[28] Victor Lazzarini, Steven Yi, John ffitch, Joachim Heintz, Øyvind Brandtsegg, and Iain McCurdy, *Csound: A sound and music computing system*, Springer, 2016.

[29] J. L. Flanagan and R. M. Golden, "Phase vocoder," *Bell System Technical Journal*, vol. 45, no. 9, 1966.

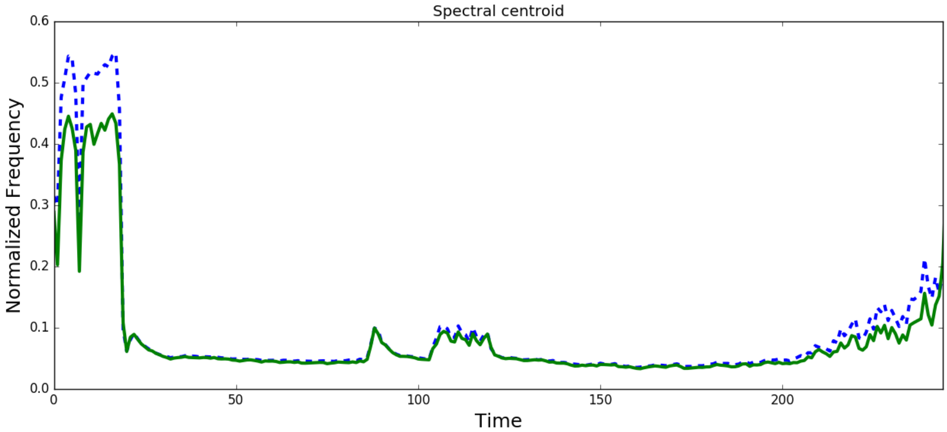

Figure 11: *Spectral centroid after amplitude modulating the IF (14 Hz of modulation depth and modulation frequency at 15 Hz) versus spectral centroid of the dry signal (blue).*

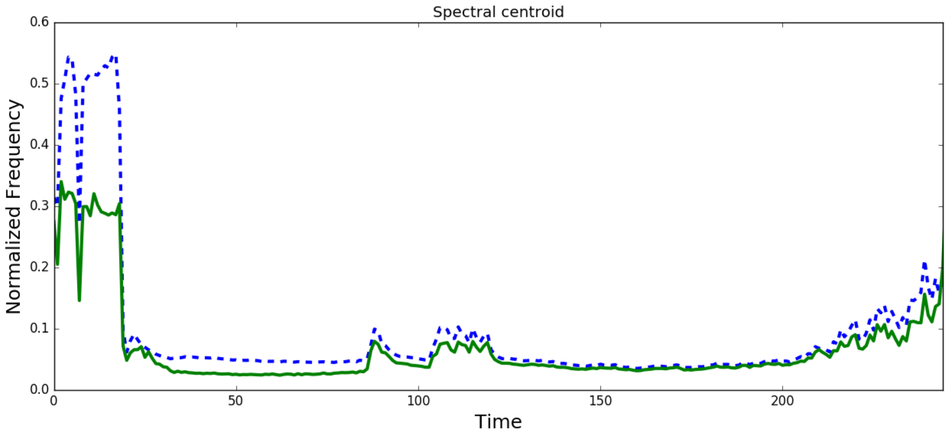

Figure 12: *Spectral centroid after ring modulating the IF (modulation frequency at 0.1 Hz) versus spectral centroid of the dry signal (blue).*

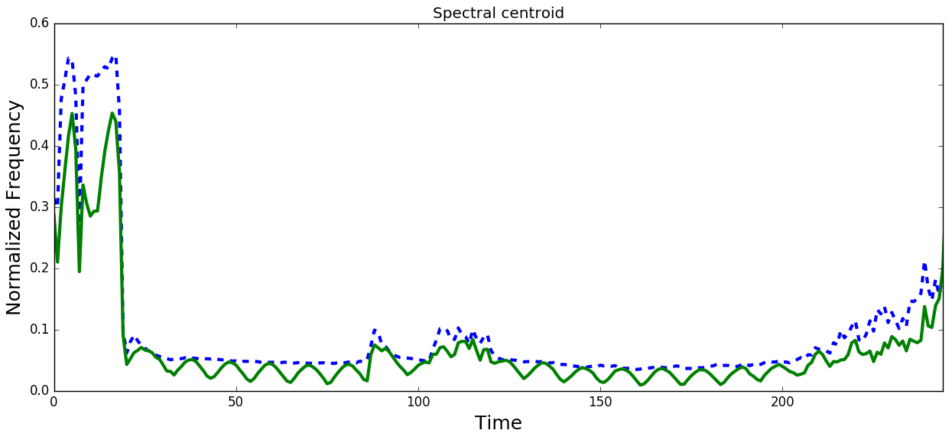

Figure 13: *Spectral centroid after ring modulating the IF (modulation frequency at 4 Hz) versus spectral centroid of the dry signal (blue).*