Perceptual Evaluation of Vibrato Models

Vincent Verfaille, SPCL, Faculty of Music, McGill University, Montréal, Qc, Canada
vincent@music.mcgill.ca http://www.music.mcgill.ca/musictech/

Catherine Guastavino, Department of Psychology, McGill University, Montréal, Qc, Canada
Catherine.Guastavino@mail.mcgill.ca http://www.psych.mcgill.ca/

Philippe Depalle, SPCL, Faculty of Music, McGill University, Montréal, Qc, Canada
depalle@music.mcgill.ca http://www.music.mcgill.ca/musictech/

Proceedings of the Conference on Interdisciplinary Musicology (CIM05)
Actes du Colloque interdisciplinaire de musicologie (CIM05)
Montréal (Québec) Canada, 10-12/03/2005

Abstract

We promote a clearer definition of vibrato (Seashore, 1932), based on a review of various vibrato features. We also propose a generalised vibrato effect generator that includes spectral envelope modulation, and a frequency-dependent hysteresis behaviour. We then investigate the influence of spectral envelope modulation on perceived quality with a double-blind randomized AB comparison task. Eight participants listened to 12 pairs of sounds with vibrato matched for loudness. Each pair included one sound with constant average spectral envelope (identical amplitude modulation over all frequencies) and one with modulated spectral envelope (frequency dependent amplitude modulation). Participants were asked to choose which version sounded the most natural. The statistical analysis revealed a significant preference for sounds with modulated spectral envelope ($p < 0.001$). Our results highlight the need to consider spectral envelope modulation for vibrato modelling.

Introduction

Vibrato was developed in the 17th century's Western music as an ornament to emphasize a particular note. It was originally used on the viola de gamba, the flute, and the singing voice, to enhance presence in musical ensembles and convey musical expression (Toff, 1996). It was imitated in the organ using a tremulant[1]. The regularity of this pulsation was then proposed as reference for the voice vibrato. In the 19th century, vibrato emerged in a more continuous form, thus becoming an attribute of musical timbre. This timbre effect, which is controlled/generated by performers, is now used on most musical instruments in Western music, including brass and wind instruments, intending to imitate the voice vibrato.

The present research aims to develop a generalised model that can account for the diversity of vibrato behaviour among different instruments (voice, string, brass and wind instruments). This model can be used to transform the vibrato of traditional instruments in the analysis/synthesis paradigm, and further to generate synthesis vibrato sounds on digital instruments.

We first present the state of the art about vibrato, from history, perception, acoustic and signal processing points of view. We then focus on a model of amplitude, frequency and spectral envelope modulation, simulating the complex behaviour of the frequencies and amplitudes of harmonics during vibrato. We finally present the perceptual evaluation of this model that was carried out to determine whether spectral envelope modulations were perceptible on saxophones sounds with vibrato, and to investigate the relevance of traditional models for adding vibrato to sounds. The implications of this study on musical practice and musicological interdisciplinarity are indicated, and we then conclude and indicate the futur directions of this research.

Vibrato State of the Art

Vibrato in Perception

Vibrato is generally defined as a vibrating quality related to pseudo-harmonic modulations of pitch, intensity or spectrum which alone or in combination serve to enrich the timbre of musical sounds. Indeed, a voice with vibrato is often denoted as bright or 'timbrée' (Garnier *et al.* , 2004). Vibrato can

[1] This air flow modulation system induces amplitude and frequency modulation, and then provides a good vibrato.

thus be considered as a timbre related perceptual attribute, since it may results from complex spectrum and spectral envelope modulations. This vibrating of pulsating aspect of vibrato can be attributed to at least one of these three components:

- fundamental frequency pulsations which are perceived as pitch pulsations, and then integrated as a vibrating quality (Frequency Modulation or pitch vibrato),
- intensity pulsations which are perceived as loudness pulsations, and then integrated as a vibrating quality (Amplitude Modulation or intensity vibrato),
- spectral enrichment cycles which correspond to spectral envelope pulsations, and are perceived as brightness modulation: the spectral centroid also varies periodicly and synchronously with AM and/or FM pulsations if any (Spectral Envelope Modulation).

Previous research investigated perceptual aspects of vibrato features that sound synthesis can benefit from, including pitch perception, vibrato rate (number of vibrato cycles per second), vibrato extent (difference between the mean and the extreme frequencies, sometimes denoted as vibrato deviation), and vibrato shape (shape of the waveform).

The pitch which is perceived for sounds with vibrato has been shown to depend on the duration of the notes. For sustained vibrato notes the perceived pitch can be estimated by the geometric mean between the two extreme frequencies (See (Shonle & Horan, 1980) for synthetic sounds and (Brown & Vaughn, 1996) for a replication with violin sounds). It has further been shown that the perception of pitch is accurate and independent from vibrato deviation (Järveläinen, 2002). However, for short notes with less than two vibrato cycles, the final part of the vibrato plays an important role, and the perceived pitch corresponds to a weighted time average where the note ending is weighted (see (d'Alessandro & Castellengo, 1994) for synthesized vocal vibrato).

The vibrato rate is generally around 6 Hz with with a variation of about $\pm 8\%$ (Prame, 1997), but it can range from 4 to 12 Hz (Desain *et al.* , 1999), and it increases towards note endings. This increase of vibrato rate towards note endings was estimated at around $+15\%$ by (Prame, 1997) for violin sounds and as an exponential increase for soprano singers (Bretos & Sundberg, 2003), who further showed that the vibrato rate differed significantly across notes.

The vibrato extent ranges between $0.6 - 2$ semitones for singers and between $0.2 - 0.35$ semitones for string players (see (Timmers & Desain, 2000) for a review). (Bretos & Sundberg, 2003) showed that the vibrato extent and the mean fundamental frequency were correlated with sound level. Results from similarity ratings indicate that the vibrato rate is perceptually more relevant than the vibrato extent (Järveläinen, 2002).

The use of vibrato by performers to convey musical expression was investigated in (Timmers & Desain, 2000). A strong effect of musical structure, particularly metrical stress, was observed on both vibrato rate and extent, yielding a consistent use of vibrato over repetitions.

The temporal evolution of vibrato has been investigated aspects during sustained notes and transition between notes. Results indicate that performers anticipate transition and that transitions occur in phase with vibrato, i.e. a note ascending towards the following note finishes with an ascending movement in the vibrato, and a note descending towards the following note finishes with a descending movement in the vibrato (d'Alessandro & Castellengo, 1994, Desain & Honing, 1996).

The perceptual prominence of amplitude modulation (AM) over frequency modulation (FM) for violin vibrato was investigated in (Mellody & Wakefield, 2000) using a same-different discrimination procedure and a multidimensional scaling task. The absence of frequency modulation had little effect on either task, while the absence of amplitude modulation affected both discrimination and sound quality scaling results.

The shape of the vibrato has received little attention. (Horii, 1989) quoted by (Timmers & Desain, 2000) proposed a classification of singer-vibrato-shapes into sinusoidal, triangular, trapezoidal, and unidentifiable. But the impact of vibrato shape of perceived sound quality remains to be studied.

Vibrato in Acoustics

Vibrato Sound Production We now explain vibrato production from an acoustical point of view, for various class of instruments.

For the singing voice, the vibrato is due to air flow modulations by the glottal source, coupled with resonances' modulations (Sundberg, 1987): variations in fundamental frequency (FM) are generated in the glottal source, and modify timbre (SEM) and amplitude (AM). The resonances' modulations are also responsible for SEM and AM, and are coupled to glottal source modulation, due to mechanical aspects of the voice production system.

For string instruments, the vibrato is obtained by moving the finger around a central position. The length of the string slightly varies, and the fundamental frequency varies accordingly (small FM). The finger motion adds a small amount of energy when moving, so the note can be sustained with no other excitation (*e.g.* the guitar), and this implies an AM. The body of the instrument does not move, so the spectral envelope is supposed constant (no SEM).

For wind and wood instruments (*cf.* Fig. 1), the vibrato is obtained by modulating the air flow: this varies amplitude (AM) and fundamental frequency (FM). Due to non-linearities inside the tube, a spectral enrichment appears when blowing louder, and disappears when blowing softer. This is the reason for the cycles of spectral enrichments (and then SEM). Depending on the instrument, modulations of the air flow can be obtained by different means. In the case of the saxophone for example, the instrumentist can apply the vibrato in two ways: by modulating the pressure on the reed on the mouth piece (soft vibrato, for soft notes) or the air pressure in the mouth.

Several observations can be made from Fig. 1, where six sound features are depicted[2]. The amplitude modulation (AM) is revealed by the modulation on the intensity, and is due to the production of the sound with vibrato. The frequency modulation (FM) is revealed by the modulation of the fundamental frequency F_0. The modulations of spectral centroid (SGC), high frequency content (HFC) and the inverse of the spectral slope (ISS) reveal the spectral envelope modulation (SEM). The odd/even balance modulation is also due to the spectral envelope modulation, but one can wonder if it is not also due to other effects in the tube, when the intensity is modulated. Indeed, the odd harmonics could be modulated in a slightly different way than the even harmonics, depending on the pressure node and non-linearities.

We note that some differences appear between instruments in that class. For example, the flute and the alto saxophone do not behave similarly during vibrato. Both have AM, FM and SEM, but the FM is in phase opposition for the saxophone, whereas it is not for the flute. For both sounds, the intensity, the SGC, the HFC and the ISS are phase synchronous. Also, modulations on SGC are more regular for the alto saxophone than for the flute. Concerning the odd/even balance modulation, it seems to always be in phase with FM for the alto saxophone, and sometimes in phase opposition for the flute. However, a further enquiry is necessary to generalise this to the whole frequency range.

For brass instruments, the vibrato is also obtained by modulating the air flow (*cf.* wind instruments). With the example given in Fig. 2, we notice how the FM is more regular than all the other modulations (AM, SGC, HFC, spectral slope). From our experience, the odd/even balance is less significative, and is sometimes in phase opposition, sometimes not (as for the flute).

To resume, a vibrato is made of at least one of these three kind of modulations:

- amplitude modulation (predominent in wind and brass instruments),
- frequency modulation (predominent in voice and string instruments),
- spectral envelope modulation and hysteresis (existing in wind, brass, voice).

Behaviour of Harmonics' Frequencies and Amplitudes With the information given below, we can depict how the amplitude and the frequency of each harmonic[3] is behaving, depending on which kind of modulation (AM, FM and SEM) is included in the vibrato.

[2]The features are defined in Appendix 1.

[3]We consider instrumental sounds, so the partials can be perfectly harmonic or nearly-harmonic for string instruments. We however confound the two cases by naming them 'harmonics'.

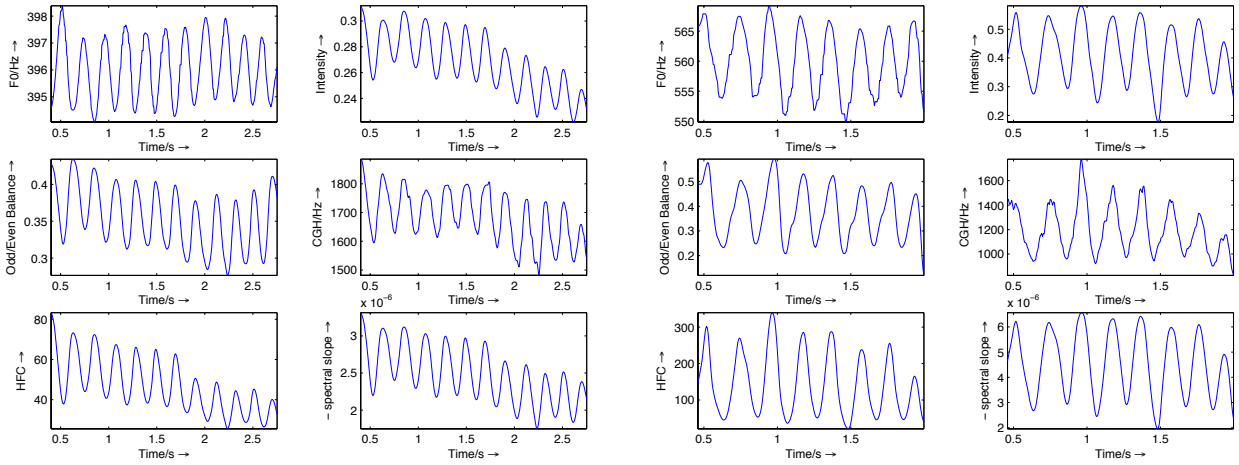

Figure 1. Left figure: G4 *ff* alto saxophone sound with vibrato. **Right figure:** Db5 *ff* flute sound with vibrato. **i)** fundamental frequency F_0, **ii)** intensity $\tilde{A}$, **iii)** odd/even harmonics balance, **iv)** spectral centroid (SGC, or spectral gravity center), **v)** high frequency content (HFC), **vi)** inverse of the spectral slope.

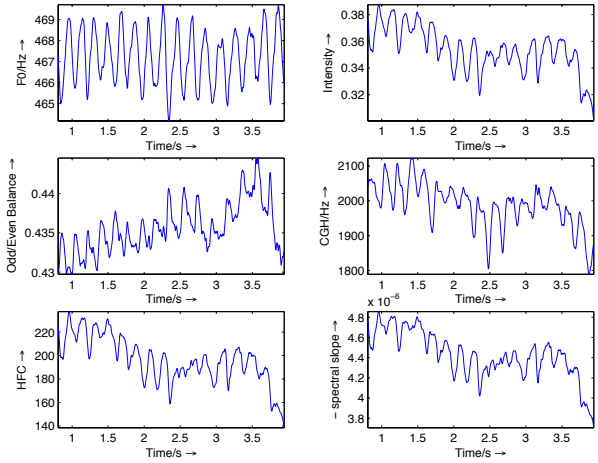

Figure 2. Bb4 *ff* trumpet sound with vibrato. **i)** fundamental frequency F_0, **ii)** intensity $\tilde{A}$, **iii)** odd/even harmonics balance, **iv)** spectral centroid (SGC, or spectral gravity center), **v)** high frequency content (HFC), **vi)** inverse of the spectral slope.

When only AM occurs, all partials' frequencies are unchanged by the vibrato, whereas all amplitudes have a pulsation (AM is equivalent to a global scaling of the spectral envelope). When only FM occurs, all partials' frequencies have a pulsation, and sweep the spectral envelope (FM is equivalent to a scaling of the source only, in a source-filter model). This also implies variations of partials' amplitudes (but not necessarily in a sinusoidal manner nor with the same periodicity). When AM and FM occur at the same time, all partials' frequencies have a pulsation, and the amplitudes are modulated twice: by sweeping the spectral envelope, and by modulating the amplitude.

When AM, FM and SEM occur at the same time, the harmonics sweep a cyclic time-varying spectral envelope, thus inducing more complex patterns. If we now take a further look at the frequency-magnitude diagram of some harmonics of a G5 played *ff* on an alto saxophone (Fig. 3 and 4), we note some well-know behaviour. Some harmonics vary on an ascending curve, since they are sweeping an ascending portion of the spectral envelope (*e.g.* harmonic number 11). Other harmonics vary on a descending curve, since they are sweeping a descending portion of the spectral envelope (*e.g.* harmonic number 1). Some harmonics follow a two part convexe curve (a sort of 'v'), because they sweep the spectral envelope around a node in the tube, that creates a zero in the frequency response of the spectral

envelope (*e.g.* harmonic number 4): they have a double period. Some other harmonics follow a two or three part concave curve (a sort of 'n'), because they sweep the spectral envelope around a small formant (*e.g.* harmonic number 10): they have a double or triple period. We already can notice that the path followed by harmonics is not a portion of curve that is swept forth and back: there is a hysteresis in that path, that we will explain and demonstrate this in Appendix 2.

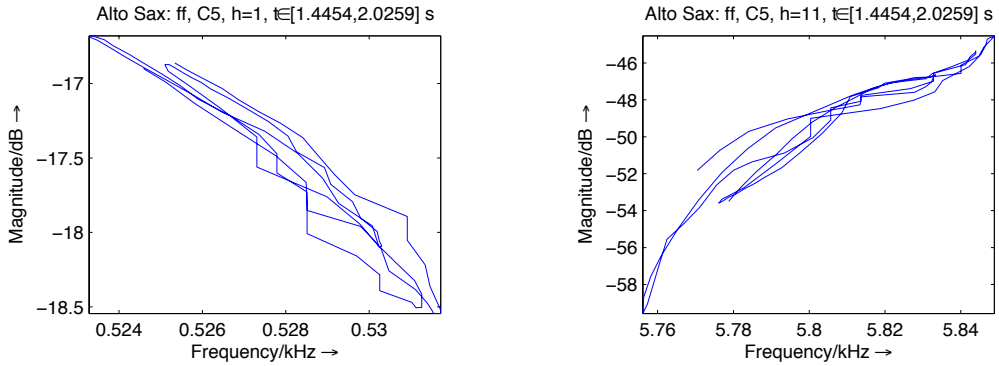

Figure 3. Behaviour of an alto saxophone C5 *ff* harmonics: **i)** on an descending part of the spectral envelope (harmonic number 1, left figure, with a single period) and **ii)** on an ascending part (harmonic number 11, right figure, with a single period).

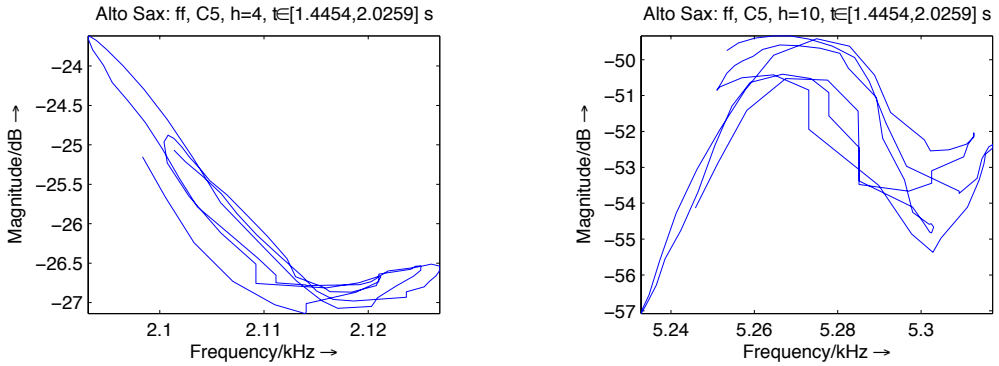

Figure 4. Behaviour of an alto saxophone C5 *ff* harmonics, **i)** sweeping around a valley of the spectral envelope (harmonic number 4, left figure, with a double period) and **ii)** sweeping around a formant (harmonic number 10, right figure, with a triple period).

Vibrato in Signal Processing

As previously said, various studies deal with vibrato analysis and perception. However, most signal processing models of vibrato rely on restricted definitions related (sometimes implicitly) to instrument-specific features, and often voice features. It has been shown that the vibrato of voice (Sundberg, 1987) as well as the vibrato of bowed string instruments (Mathews & Kohut, 1973) consists mainly of frequency modulation, whereas vibrato of wind instruments consists mainly in amplitude modulation. Several models have been recently developed to take into account these two modulations in a context of voice synthesis (Herrera & Bonada, 1998) and analysis/transformation/synthesis (Arfib & Delprat, 1998, Rossignol *et al.* , 1999).

Wind and brass instruments, however, exhibit more complex vibrato behaviour combining synchronized variations of not only frequency and amplitude but also spectral envelope. This does not mean that it is not the case for other instruments: spectral envelope modulation was introduced in a voice vibrato model (Maher & Beauchamp, 1990). This SEM was obtained by interpolating between two reference spectral envelopes (from two different loudness). A perceptual impact of spectral envelope modulation

on sound quality was observed, although not formally validated.

Formalisation of the Generalised Vibrato Model We consider the signal as a sum of modulated sinusoids, using the additive model (McAulay & Quatieri, 1986, Serra & Smith, 1990):

$$x(n) = \sum_{h=1}^{H} a_h(n) \cdot \cos\left(\Phi_h(n)\right) \quad (1)$$

The phase is given as the integral of the time-varying frequency $f_h(n)$:

$$\Phi_h(n) = \Phi_h(n-1) + 2\pi\frac{f_h(n)}{F_s} \quad (2)$$

with F_s the sampling rate or frequency and $\Phi_h(0)$ the initial phase. Vibrato is considered as a quasi-periodic feature, that can be expressed with a Fourier serie decomposition of the involved parameters. The amplitudes $\tilde{a}_h(n)$ and frequencies $\tilde{f}_h(n)$ are given as sum of sinusoids[4], according to the two-level sinusoidal model (Marchand & Raspaud, 2004):

$$\tilde{a}_h(n) = \sum_{l=1}^{M_h^a} \tilde{a}_l^a(n) \cdot \cos\left(\tilde{\Phi}_l^a(n)\right) \quad (3)$$

$$\tilde{f}_h(n) = \sum_{l=1}^{M_h^f} \tilde{a}_l^f(n) \cdot \cos\left(\tilde{\Phi}_l^f(n)\right) \quad (4)$$

Note that no assumption is made about the synchronisation between the modulations of amplitudes $\tilde{a}_h(n)$ and frequencies $\tilde{f}_h(n)$: the model of parameters given in Eq. (3) and (4) is able to represent any modulation, its accuracy depending on M_h^a the number of components to represent the amplitude $\tilde{a}_h(n)$, or M_h^f the number of components to represent the frequency $\tilde{f}_h(n)$. In practice, we use the same number for amplitudes and frequencies, and for all the harmonics:

$$M_h^a = M_h^f = M \quad (5)$$

We denote $\tilde{x}(n)$ for signal/parameters with vibrato, $x(n)$ for signal/parameters without vibrato, and $\overline{x}(n)$ for synthesis signal/parameters obtained by adding a vibrato of any type to a flat sound.

We also note the instantaneous amplitudes(signal intensity levels) as:

$$A(n) = \frac{1}{H}\sqrt{\sum_{h=1}^{H}\left(a_h(n)\right)^2} \quad (6)$$

$$\tilde{A}(n) = \frac{1}{H}\sqrt{\sum_{h=1}^{H}\left(\tilde{a}_h(n)\right)^2} \quad (7)$$

and $\mathcal{E}(f,n)$ (resp. $\tilde{\mathcal{E}}(f,n)$) the spectral envelope of the flat sound (resp. vibrated sound) estimated from the $\left(a_h(n), f_h(n)\right)$ values (resp. $\left(\tilde{a}_h(n), \tilde{f}_h(n)\right)$). The spectral envelope can be estimated either by linear interpolation (Serra & Smith, 1990) or by using the discrete cepstrum (Galas & Rodet, 1990). We now present the models for generating vibrato on flat sounds using the first component of the Fourier series on amplitudes and frequencies.

AM Pulsation In case the vibrato reduces to an AM pulsation, also called tremolo, the model of parameters reduces to:

$$\overline{f}_h(n) = f_h(n) \quad (8)$$

[4]These parameters are in practice estimated at the analysis step by block at $n = mR_A$ (with R_A the analysis step increment), and regenerated for each sample by linear interpolation for amplitudes $\tilde{a}_h(n)$ and cubic interpolation for the phases $\tilde{\Phi}_l^f(n)$ (frequencies $\tilde{f}_h(n)$ then obey to a quadratic interpolation) at the synthesis step (McAulay & Quatieri, 1986).

$$\overline{a}_h(n) = \gamma^a(n) \cdot a_h(n) \quad (9)$$

$$\gamma^a(n) = 1 + \tilde{a}_0^a(n) \cdot \cos\left(\tilde{\Phi}_0^a(n)\right) \quad (10)$$

where $\tilde{\Phi}_0^a(n)$ the phase of the AM is given as a function of $\tilde{f}_0^a(n)$ the frequency (or rate) of the AM and $\tilde{a}_0^a(n)$ the amplitude (or extent) of the AM, as:

$$\tilde{\Phi}_0^a(n) = \tilde{\Phi}_0^a(n-1) + 2\pi\frac{\tilde{f}_0^a(n)}{F_s} \quad (11)$$

Notice that the AM is globally applied to the signal, by giving the same ratio to all harmonics' amplitudes.

FM Pulsation In case the vibrato reduces to a FM pulsation with constant spectral envelope, as for the violin model (Mathews & Kohut, 1973) or the voice model (Sundberg, 1987, Arfib & Delprat, 1998), the model of parameters reduces to:

$$\overline{f}_h(n) = \gamma^f(n) \cdot f_h(n) \quad (12)$$

$$\gamma^f(n) = 1 + \tilde{a}_0^f(n) \cdot \cos\left(\tilde{\Phi}_0^f(n)\right) \quad (13)$$

$$\tilde{\Phi}_0^f(n) = \tilde{\Phi}_0^f(n-1) + 2\pi\frac{\tilde{f}_0^f(n)}{F_s} \quad (14)$$

$$\overline{a}_h(n) = \mathcal{E}\left(\overline{f}_h(n), n\right) \quad (15)$$

with $\tilde{f}_0^f(n)$ the frequency (or rate) of the FM and $\tilde{a}_0^f(n)$ the amplitude (or extent) of the FM. The amplitude modulation is a result of the spectral envelope scanning by the harmonics.

AM/FM Pulsation In case the vibrato reduces to an AM/FM pulsation (Herrera & Bonada, 1998, Rossignol *et al.* , 1999), the model of parameters reduces to:

$$\overline{f}_h(n) = \gamma^f(n) \cdot f_h(n) \quad (16)$$

$$\overline{a}_h(n) = \gamma^a(n) \cdot \mathcal{E}\left(\overline{f}_h(n), n\right) \quad (17)$$

with the assumption that the AM and FM pulsations are synchronous:

$$\tilde{\Phi}_0^f(n) = \tilde{\Phi}_0^a(n) = \tilde{\Phi}_0(n) \quad (18)$$

The amplitude modulation of each harmonic is a result of both the spectral envelope scanning by the harmonics and the global AM by $\gamma^a(n)$.

AM, FM and SEM Pulsation In order to apply a combined AM/FM/SEM, let us first express the time-varying modelling of the spectral envelope (SE). The time-varying SE can be obtained by scaling the original SE with a linear function of the frequency (thus changing its slope):

$$\overline{\mathcal{E}}\left(\overline{f}_h(n), n\right) = \gamma^e(n) \cdot \overline{f}_h(n) \cdot \mathcal{E}\left(\overline{f}_h(n), n\right) \quad (19)$$

where $\gamma^e(n)$ renders the spectral modulation:

$$\gamma^e(n) = c(n) + \tilde{a}_0^e(n) \cdot \cos\left(\tilde{\Phi}_0^e(n)\right) \quad (20)$$

where $c(n)$ and $\tilde{a}_0^e(n)$ must be estimated. To our knowledge, this SEM model is well suited for the singing voice, wind instruments such as flute, and brass instruments.

The time-varying SE can also be obtained by interpolating between two extrema spectral envelopes (Maher & Beauchamp, 1990), when the previous solution does not suit to the instrument:

$$\overline{\mathcal{E}}\left(\overline{f}_h(n), n\right) = \beta_e(n) \cdot \mathcal{E}_+\left(\overline{f}_h(n), n\right) + (1 - \beta_e(n)) \cdot \mathcal{E}_-\left(\overline{f}_h(n), n\right) \quad (21)$$

$$\beta_e(n) = \frac{1 + \cos\left(\tilde{\Phi}_0^e(n)\right)}{2} \quad (22)$$

Using this time-varying spectral envelope, the frequencies and amplitudes are determined accordingly using:

$$\overline{f}_h(n) = \gamma^f(n) \cdot f_h(n) \quad (23)$$

$$\overline{a}_h(n) = \gamma^a(n) \cdot \overline{\mathcal{E}}\left(\overline{f}_h(n), n\right) \quad (24)$$

with the assumption that AM, FM and SEM pulsations are synchronous, so they have the same phase:

$$\tilde{\Phi}^e_0(n) = \tilde{\Phi}^a_0(n) = \tilde{\Phi}^f_0(n) = \tilde{\Phi}_0(n) \quad (25)$$

Notice that it does not mean that the resulting amplitude modulation of harmonics occur at the same frequency. The scanning of a formant region might double the frequency of modulation.

Comparison of Vibrato Models The limit of the AM, FM and AM/FM models is that they consider vibrato modulation as made of only one modulated sinusoidal component. This exclude more realistic modulation curves: the two-level sinusoidal model provides a solution to this. Moreover, the AM/FM model consider phase synchronous modulations, whereas they can be in opposite phase (*e.g.* the saxophone, as explained in the acoustic part). None of these three models take into account the SEM, which is important, as we will show with the perceptual test.

Let us consider the example of the time-scaling of a voice sound with vibrato by using a model: if there is no SEM in the signal, then the approach proposed in (Arfib & Delprat, 1998), that consists in removing the FM vibrato by pich-shifting, time-scaling the flat sound, and then applying back the FM vibrato by pitch-shifting, is valid and similar to a real longer vibrated sound. However, if there is SEM in the signal, then the FM and SEM components are not processed in a coherent manner: SEM is time-scaled whereas the FM is not, thus resulting in a processed sound with artifacts that could be audible.

Author	Instrument(s)	FM	AM	random	SEM	Transitions
(Seashore, 1932, 1936)	voice, violin	yes	yes		yes	implicit
(Maher & Beauchamp, 1990)	voice	yes	yes	explicit	yes	no
(Arfib & Delprat, 1998)	voice	yes	no	no	no	no
(Herrera & Bonada, 1998)	any	yes	yes	no	no	implicit
(Rossignol *et al.* , 1999)	any	yes	no	no	no	implicit
(Järveläinen, 2002)	stringed	yes	no	no	no	implicit
(Marchand & Raspaud, 2004)	any	yes	yes	implicit	implicit	implicit
generalised vibrato model	any	yes	yes	implicit	explicit	implicit

Table 1. Vibrato models. AM stands for global amplitude modulation. FM stands for frequency modulation of the fundamental frequency. Random stands for harmonics' shimmer and jitter. SEM stands for spectral envelope modulation. This table also indicates if the model takes into account the transitions between vibrato notes.

As we can see in the comparison of the vibrato models (*c.f.* Table 1), only two models take into account the SEM: the panned-wavetable synthesis and the two-level sinusoidal model. The panned-wavetable synthesis method explicitely uses the SEM, and this SEM implicitely takes into account the AM. The vibrato control is composed of a time-varying sinusoidal component plus a random component. This model is good for synthesis and gives some clues for sound transformation. However, this model does not easily allow for AM modifications, since it is implicitely performed. The two-level sinusoidal model implicitely takes into acount the SEM, by modelling the AM and FM of each partial. This model allows for complex modulated amplitude and frequency of partials, and is good for sound transformation: for example, time-scaling can be performed in a good manner, without desynchronizing AM, FM and SEM.

In the model proposed in the next section, we overcome these limitations by combining the advantages of the two-level sinusoidal model and the explicit SEM of the panned-wavetable synthesis with cross-synthesis. We will explicit two ways of computing the SEM.

A Generalised Vibrato Model, with Explicit AM/FM/SEM

After proposing a definition of vibrato, we explain the signal processing model we developed and based on the panned-wavetable synthesis and the two-level sinusoidal model.

Vibrato Definition

Often in signal processing, only AM and FM vibrato are considered, and timbre modulation only concerns the complex behaviour of FM vibrato scanning the AM spectral envelope, but not the SEM. A clearer definition of vibrato is presented on the basis of a review of vibrato features (Seashore, 1932, Toff, 1996): **we define the vibrato as a vibrating quality of musical sounds, corresponding to simultaneous modulations of amplitude (AM), frequency (FM) and/or spectral envelope (SEM)**. Note that in order to take into account the specific behaviour of harmonics' amplitudes, we consider the modulations as simultaneous and not synchronous. However, we can already better define these modulations saying that the SEM, the FM and the global AM (not the AM of each frequency) are nearly sinusoidal and are synchronous or in phase opposition. We have established that the spectral envelope modulation implies a frequency-dependent hysteresis behaviour (see Appendix 2 for the demonstration).

Generalised Vibrato Model

Due to the limitations of the AM, FM and AM/FM vibrato models, it is clear that a generalised AM/FM/SEM model is needed: moreover, it is more adapted for transforming various instrument sounds with vibrato.

We developed a generalised vibrato model based on the definition given previously, for use in an analysis/transformation/synthesis context. We use the analysis by synthesis paradigm (Risset & Wessel, 1999): the quality of the model will be perceptually evaluated. Our model uses the two-level sinusoidal model[5] that represents the amplitudes and frequencies of harmonics as sums of sinusoids, thus implicitly integrating the spectral envelope modulation. It also uses the panned-wavetable synthesis technique[6] in order to explicitely represent the SEM. By doing so, we provide controls on the three components of vibrato (AM, FM and SEM) as follows:

1. we compute the two-level sinusoidal model data: $\left(\tilde{a}_h(n), \tilde{f}_h(n), \tilde{\phi}_h\right)$;
2. the explicit control over the FM is given by the modulated frequencies of harmonics: $\overline{f}_h(n) = T_f(\tilde{f}_h(n))$, with T_f a transformation of the frequencies (for example changing the frequency, the depth, the frequency composition of the vibrato controls);
3. the interpolation between two spectral envelopes (or the spectral envelope slope changes) allows for an explicit control over the SEM: $\overline{\mathcal{E}} = T_e\left(\tilde{\mathcal{E}}\right)$ with T_e a transformation of the spectral envelope;
4. the new amplitudes $\hat{a}_h(n)$ are computed by interpolation in the spectral envelope $\overline{\mathcal{E}}\left(\overline{f}_h(n), n\right)$;
5. the instantaneous amplitude $\hat{A}(n)$ is computed from the new amplitudes $\hat{a}_h(n)$.
6. when modelled as a sum of sinusoids, the instantaneous amplitude allows for an explicit control on the AM: $\overline{A}(n) = T_a(\hat{A}(n))$, with T_a a transformation of the instantaneous amplitude, applied by multiplying all the $\hat{a}_h(n)$ by the same ratio $r(n)$.
7. the final amplitudes are then given as: $\overline{A}(n) = r(n)\hat{A}(n)$.

Note that the two-level sinusoidal model also considers more than one component for the periodic vibrato, which is more realistic. It has however been shown in (Maher & Beauchamp, 1990) that the random component (jitter and shimmer) added to the vibrato control curve is not perceived by listeners. In the context of pure synthesis, this question may have importance for the realism of the synthetic sound. In our context of analysis/transformation/synthesis, the original instrument sound already have its harmonics' frequencies and amplitudes made of a general trend (due to the control) and random components (jitter and shimmer). It does not make any sense to add jitter/shimmer when adding a vibrato to a flat sound; it may however make sens to wonder what to do with that jitter/shimmer when time-scaling the vibrated sound. It does not seem that this question has been addressed yet.

[5] The two-level sinusoidal model was design in the analysis/transformation/synthesis context.

[6] The panned-wavetable synthesis technique was developed in order to produce a realistic vibrato for synthesis sounds.

In a context of vibrato sound synthesis, this model has to be combined with cross-synthesis, in order to take into account realistic vibrato control curves and spectral envelope modulations. Therefore, when synthesizing the test sounds, we combined this AM/FM/SEM model with cross-synthesis, since we needed to synthesize AM/FM sounds with and without SEM: a mean spectral envelope was needed, and extracted from a second sound.

Some More Insights in the Vibrato Model

We now develop some specific aspects of the generalised vibrato model, dealing with the hysteresis of harmonics' path in the frequency/amplitude domain, the difference of behaviour of SGC and HFC features depending on the exsitence of SEM, the way to implicitely take into account the AM in the SEM, the non linear coupling between the source and the filter, and finally the questions that may arise concerning the definition and perception of formants and valleys of the spectral envelope during vibrato.

Is There Hysteresis on the Harmonics' Path? This question deals with the symmetrical profile of the vibrato. We here question wether the vibrato has the same behaviour during the rise and during the fall. To answer this question, we consider the ideal case where all the vibrato parameters are constant with time. The demonstration is given in Appendix 2. The conditions on the spectral envelope for no hysteresis imply either oscillating spectral envelopes around each harmonics (SEM by changing the slope) or a flat spectral envelope (SEM by interpolation). In any other condition, the AM/FM/SEM pulsation implies hysteresis on the harmonics' path. In the general case where the pulsation rate and amplitudes are not constant with time, the AM/FM/SEM pulsation always have hysteresis.

Harmonics' amplitudes behaviour, with/without SEM To better understand the consequences of SEM harmonics' behaviour, let us plot the magnitudes and frequencies of harmonics, without SEM and with SEM (Fig. 5). As depicted, the 'no-SEM' harmonics have identical (and translated) amplitude patterns,

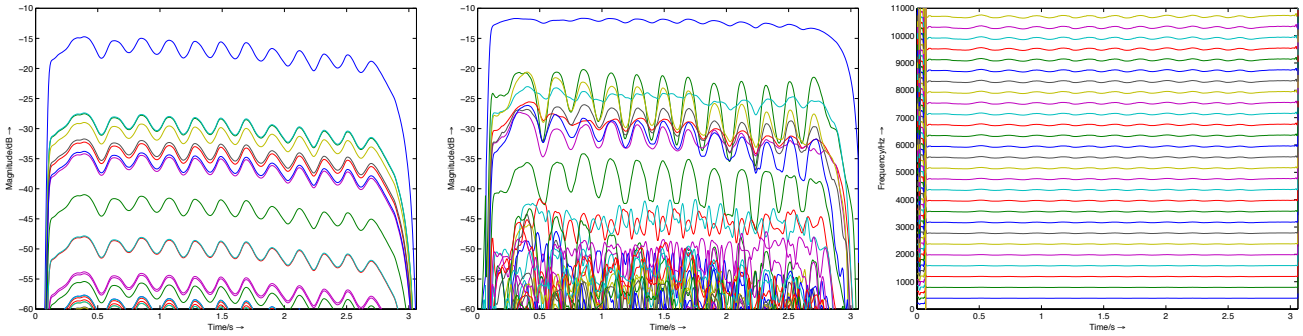

Figure 5. Comparison of harmonics' amplitudes when only preserving AM/FM (left figure) and when preserving AM/FM and SEM (middle figure). Frequencies are identical for both sounds (right figure).

whereas 'SEM' harmonics have partials with simple, double and triple period patterns, sometimes in opposite phase. This is due to the fact that the spectral envelope (without SEM) we used is constant, and not well enough discretized. Indeed, the spectral envelope estimation is smoother that the real spectral envelope: most of the zeros are not present, thus implying the absence of SEM specific behaviour, such as big range of magnitude variation around zeros when a harmonic sweeps around it.

This has some implications on the sound features. As depicted in Fig. 6, SGC and HFC have different behaviours depending if SEM is in the model or not. The extend of SGC is greater when there is a SEM, and the two SGC modulations are not in phase. These two effects are due to the way the SGC is computed as well as to the fact the spectral envelope is not modulated without SEM. The extend of HFC is greater when there is a SEM, and the two HFC modulations are in phase.

Setting the Parameters of the Two SEM Models We proposed two ways to modify the spectral envelope (SEM), we now explain how to set their parameters. When interpolating between two extrema spectral envelope (SE), these extrema correspond to the SE of notes at different loudness, and can be

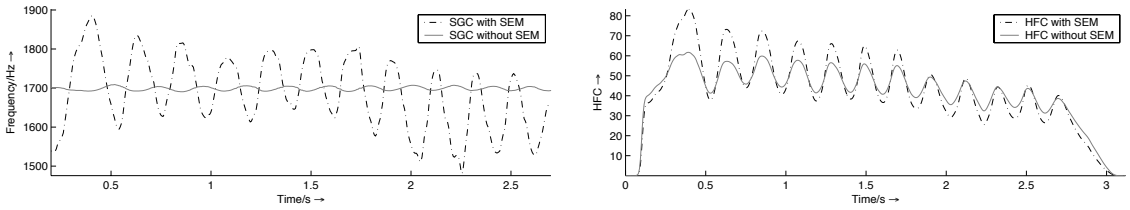

Figure 6. Comparison between spectral gravity centers (SGC, left figure) and high frequency contents (HFC, right figure) with and without SEM. The extend of SGC and HFC is greater when there is a SEM. The two SGC modulations are in opposite phase, and the two HFC modulations are in phase.

obtained precisely from estimated SE from flat sounds. For some instruments (*e.g.* brass and flute), these extrema can also be computed as the maximum SE with a slope change. Interpolating between two spectral envelopes is the most general case.

Concerning the slope changing, the $c(n)$ and $\tilde{a}_0^e(n)$ values have to be estimated. Considering that we know the extrema spectral envelopes $\mathcal{E}_-\left(\overline{f}_h(n),n\right)$ and $\mathcal{E}_+\left(\overline{f}_h(n),n\right)$, they are approximated by:

$$\mathcal{E}_+\left(\overline{f}_h(n),n\right) = (c(n)+\tilde{a}_0^e(n))\cdot\overline{f}_h(n)\cdot\mathcal{E}\left(\overline{f}_h(n),n\right) \quad (26)$$

$$\mathcal{E}_-\left(\overline{f}_h(n),n\right) = (c(n)-\tilde{a}_0^e(n))\cdot\overline{f}_h(n)\cdot\mathcal{E}\left(\overline{f}_h(n),n\right) \quad (27)$$

so $c(n)$ and $\tilde{a}_0^e(n)$ minimise the two following quantities:

$$\varepsilon_c(n) = c(n)\cdot\overline{f}_h(n)\cdot\mathcal{E}\left(\overline{f}_h(n),n\right) - \frac{\mathcal{E}_+\left(\overline{f}_h(n),n\right)+\mathcal{E}_-\left(\overline{f}_h(n),n\right)}{2} \quad (28)$$

$$\varepsilon_a(n) = \tilde{a}_0^e(n)\cdot\overline{f}_h(n)\cdot\mathcal{E}\left(\overline{f}_h(n),n\right) - \frac{\mathcal{E}_+\left(\overline{f}_h(n),n\right)-\mathcal{E}_-\left(\overline{f}_h(n),n\right)}{2} \quad (29)$$

In the optimal case where the extrema SE exactly corresponds to a SE with a given slope change, then:

$$\mathcal{E}_+\left(\overline{f}_h(n),n\right) = d_+(n)\cdot\overline{f}_h(n)\cdot\mathcal{E}\left(\overline{f}_h(n),n\right) \quad (30)$$

$$\mathcal{E}_-\left(\overline{f}_h(n),n\right) = d_-(n)\cdot\overline{f}_h(n)\cdot\mathcal{E}\left(\overline{f}_h(n),n\right) \quad (31)$$

and $c(n)$ and $\tilde{a}_0^e(n)$ are explicitely given as:

$$c(n) = \frac{d_+(n)+d_-(n)}{2} \quad (32)$$

$$\tilde{a}_0^e(n) = \frac{d_+(n)-d_-(n)}{2} \quad (33)$$

SEM Models with Implicit AM Both methods can implicitely combine AM and SEM in the SEM. We give the corresponding mathematical developments, in order to highlight the way the usual two-level sinusoidal model does this. When changing the slope of the SE:

$$\overline{\overline{\mathcal{E}}}\left(\overline{f}_h(n),n\right) = \gamma^a(n)\cdot\overline{\mathcal{E}}\left(\overline{f}_h(n),n\right) \quad (34)$$

$$= \gamma^a(n)\cdot\gamma^e(n)\cdot\overline{f}_h(n)\cdot\mathcal{E}\left(\overline{f}_h(n),n\right) \quad (35)$$

$$= \gamma^e(n)\cdot\overline{f}_h(n)\cdot\mathcal{E}_{\gamma_a}\left(\overline{f}_h(n),n\right) \quad (36)$$

with the notation:

$$\mathcal{E}_{\gamma_a}\left(\overline{f}_h(n),n\right) = \gamma^a(n)\cdot\mathcal{E}\left(\overline{f}_h(n),n\right) \quad (37)$$

When interpolating between two extrema SE:

$$\overline{\overline{\mathcal{E}}}\left(\overline{f}_h(n),n\right) = \gamma^a(n)\cdot\overline{\mathcal{E}}\left(\overline{f}_h(n),n\right) \quad (38)$$

$$= \gamma^a(n)\cdot\beta_e(n)\cdot\mathcal{E}_+\left(\overline{f}_h(n),n\right)+\gamma^a(n)\cdot(1-\beta_e(n))\cdot\mathcal{E}_-\left(\overline{f}_h(n),n\right) \quad (39)$$

$$= \beta_e(n)\cdot\mathcal{E}_{\gamma_a,+}\left(\overline{f}_h(n),n\right)+(1-\beta_e(n))\cdot\mathcal{E}_{\gamma_a,-}\left(\overline{f}_h(n),n\right) \quad (40)$$

with the new extrema spectral envelopes:

$$\mathcal{E}_{\gamma_a,+}\left(\overline{f}_h(n),n\right) = \gamma^a(n)\cdot\mathcal{E}_+\left(\overline{f}_h(n),n\right) \quad (41)$$

$$\mathcal{E}_{\gamma_a,-}\left(\overline{f}_h(n),n\right) = \gamma^a(n)\cdot\mathcal{E}_-\left(\overline{f}_h(n),n\right) \quad (42)$$

Non Linear Coupling Between the Source and the Filter The spectral enrichment (and so forth the SEM) when the loudness increases is due to a non linear effect, such as for trumpet and more generally for brass sounds (Risset, 1965). In the usual source-filter model, the filter being a linear system, there is no such non linear coupling between the source and the filter. The source/filter model is not ideal from that point of view, and physical modelling may give more accurate values of the parameters of the model. However, the generalised signal processing model of vibrato intends to take into account this non linear coupling between the source and the filter, using an additive/substractive representation of sound and the SEM to explicit the effect of non linear coupling.

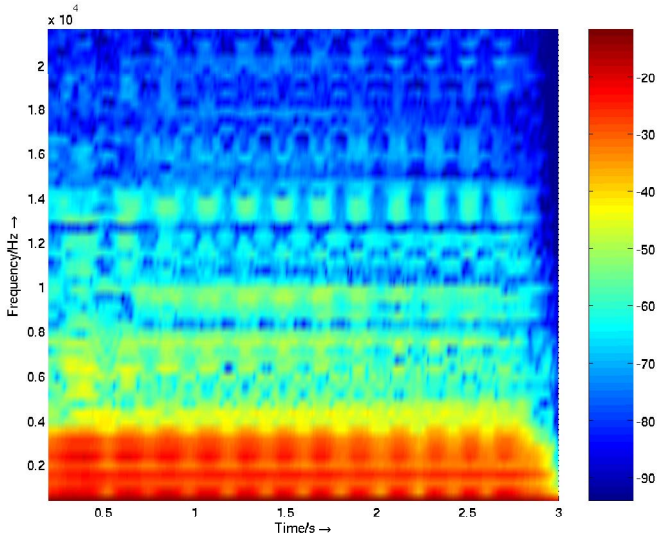

Figure 7. Sonagram of the spectral envelope filtered by linear interpolation on magnitudes.

Questions That Arise From the spectral envelope filtered sonagram (Fig. 7), we notice that formants and valleys' frequencies are preserved (but of course not magnitudes). It is well-known that a constant spectral envelope (SE) is better perceived thanks to jitter on harmonics, which then sweep the SE. In that case, the following open questions arise: How about the fact that SEM preserves formants and valleys' frequencies? Cannot this be of any help to perceive the formants? These questions are beyond the scope of this paper and will be adressed in future works.

Validation of this Vibrato Model

In the context of analysis/transformation/synthesis, flat sounds are added a vibrato by AM/FM. The model we propose combines AM/FM with SEM, as in (Maher & Beauchamp, 1990). We now must evaluate if the difference is perceptually relevant.

The influence of spectral envelope modulation on perceived quality was then investigated using a double-blind randomized AB comparison task. Eight participants listened to 12 pairs of sounds with vibrato. Each pair included one sound with constant average spectral envelope (identical amplitude modulation over all frequencies) and one with modulated spectral envelope (frequency dependent amplitude modulation). Both sounds in each pair were matched subjectively for loudness by 5 expert listeners in a preliminary experiment. In the main experiment, participants were asked to choose which version sounded the most natural and justify their choices in an open questionnaire. The statistical analysis (binomial test) revealed a significant preference for sounds with modulated spectral envelope ($p < 0.001$).

Methods

Synthesis of the Experimental Sounds Cross-synthesis techniques were used to create hybrid sounds from two saxophone sounds with and without vibrato. Using the knowledge and notations about AM, FM, SEM sounds described with the two-level sinusoidal model, we can explicit how we synthesized the experimental sounds.

We had the following constraints on material (sounds):

- sounds are created by cross-synthesis between a sound with vibrato and a sound without vibrato, so we need pairs of sound having the same nuance, pitch and duration. We selected sound pairs from the IOWA database (IOWA, 2005);
- in order to provide a good analysis and synthesis, the original sounds must be exempt of reverberation. This is the case for the sounds from the IOWA database, as they are recorded in an anechoic room;
- the frequency range must be representative of the instrument. We studied alto saxophone sounds, ranging in pitch from F3 to C5.

and on the synthesis:

- in order to use the two-level sinusoidal model, we first need to use an additive analysis/transformation/synthesis, the transformation being applied at the second level;
- the residual of analyzed sound was removed as we focus on the modification of the deterministic parts.

The hypothesis that we wanted to test is wether the spectral envelope modulation (SEM) can be heard or not. This implies to synthesize sounds for the experiment with and without this SEM. Another constraint is that any existing amplitude modulation and/or frequency modulation must be preserved.

The sound with SEM is directly synthesized from the analysis data, as:

$$\tilde{x}(n) = \sum_{h=1}^{N} \tilde{a}_h(n) \cdot \cos\left(\tilde{\Phi}_h(n)\right) \tag{43}$$

$$\tilde{\Phi}_h(n) = \tilde{\Phi}_h(n-1) + 2\pi \frac{\tilde{f}_h(n)}{F_s} \tag{44}$$

with $\tilde{a}_h(n)$, $\tilde{f}_h(n)$ and $\tilde{\phi}_h(n)$ provided by the analysis.

The sound without SEM are synthesized by cross synthesis between the sound with vibrato and the sound without vibrato, as follows:

1. we compute the two-level sinusoidal model data: $(a_h(n), f_h(n), \phi_h(n))$ and $\left(\tilde{a}_h(n), \tilde{f}_h(n), \tilde{\phi}_h(n)\right)$;

2. the synthesized harmonics' frequencies are the modulated frequencies of the sound with vibrato: $\overline{f}_h(n) = \tilde{f}_h(n)$;
3. the amplitudes are given by interpolation in the mean spectral envelope (constant SE) and then multiplication by the ratio of global instantaneous amplitudes, so that both sounds have the same amplitude modulation:

$$\overline{a}_h(n) = \tilde{\mathcal{E}}\left(\tilde{f}_h(n), n\right) \cdot \frac{\tilde{A}(n)}{A(n)} \tag{45}$$

The first level of sinusoidal analysis was performed using CLAM (Amatriain *et al.* , 2002a) and exported as SDIF data to Matlab. The second level of sinusoidal analysis was performed using Matlab, and the matlab SMS version (Amatriain *et al.* , 2002b). The synthesis data were then stored as SDIF, and the sounds were synthesized using CLAM.

Experimental Design Considerations One possibility for such a listening test would be to use one of a number of standard psychophysical tests to determine *discriminability*, that is, the ability of listeners to detect a difference between the two signals. These include AB-X, AAA XYZ, AX, etc. A potential drawback of such tests is that if differences are detectable between the two sets of sound samples, no data will exist as to which samples are preferred by listeners. After initial pilot testing, we determined that differences were readily apparent and easily detectable, even by unskilled listeners. Consequently, an A-B Preference Test was conducted. In an A-B Preference Test, listeners hear two samples (A and B) and are asked to indicate which they prefer. If they have no preference, they are instructed to pick an answer at random. Stimuli are presented in random, counterbalanced order, so that over the course of many trials and many subjects, systematic effects of presentation order ('order effects' or 'sequence effects') are nullified. As well, over the course of many trials, listeners who can discriminate one sound sound sample from another will choose each one an equal number of times (the samples were presented randomly and the listeners are choosing randomly) and hence inability to discriminate will be revealed as 'no preference' in the final results.

Apparatus Soundfiles were played through a MOTU 828mkII 24-bit 96KHz D/A convertor, attached to a MacIntosh Apple computer via Firewire. Listeners used AKG 240 Gold Professional Closed Ear 600Ω headphones. The test samples were presented with a graphical interface, programmed in Max/MSP (Cycling'74, 2003, Puckette, 1991).

Procedure

Loudness Matching Test Because the two synthesized sounds for each note contained unequal spectral distributions, automated methods for equating overall power in the soundfiles may not have yielded samples that would sound matched for loudness: the sample with the highest spectral centroid would always tend to sound louder. We thus employed a subjective evaluation experiment prior to the preference test to match for loudness. 5 listeners (4 males, 1 female, mean age 28; s.d.3.4) served without pay in the experiment. They were expert listeners, with a minimum of 7 years of musical training and familiar with loudness matching tasks. Participants were presented with pairs of sound samples in a double-blind, randomized listening test. Participants were instructed to set the level of one of them so that both sounds appeared equally loud. Each pair was presented twice in counterbalanced order. The graphical interface enabled participants to adjust the level on a slider in real time, and to switch back and forth between the two versions as many times as desired. Loudness judgments were consistent within and across subjects (s.d. < 1 dB for all samples). The volume settings were averaged over all participants, and the amplitude of sound samples were subsequently adjusted for each pair (average gain of 2.3 dB).

Preference Test A new set of 8 subjects (5 males, 3 females; mean age 29; s.d.10.5) participated without pay in the preference test. 5 participants were expert listeners with a minimum of 11 years of musical training, 3 were participants were non-musically trained. Participants were given instructions to choose which of two sounds they preferred, and to choose at random if they had no preference. Preferences were indicated by clicking the computer mouse inside a box underneath the icon for Sound *A* or Sound *B*. An additional on-screen button allowed participants to play the sound pairs as many times as they liked.

The sound pairs were always played in their entirety, and pressing the 'play again' button on the screen did not cause the currently playing sound to terminate. At the beginning of each trial, the two sounds were played sequentially. An icon on the computer monitor indicated which of the two sounds, *A* or *B* was playing and after participants indicated their preference by clicking the mouse in the appropriate box, the next trial started. Each sound pair was presented to each subject twice in counter-balanced order. Following the experiment, participants were asked to freely describe the difference between the 2 versions presented in each trial, and to justify their choices in an open questionnaire.

Results

A binomial test was conducted and revealed a significant preference ($p < 0.001$) for the sounds synthesized with SEM: vibrato sounds with SEM were selected on 76 trials out of 96 (see Fig. 8). The verbal comments collected were classified into categories emerging from the participants' spontaneous descriptions. These descriptions referred primarily to timbre (7 occ.), naturalness (7 occ.), vibrato depth (6 occ.), temporal structure (6 occ.) and attack/onset (4 occ.) and pleasantness (2 occ.). Sounds with SEM were described as having a 'full' timbre with a deep and slightly irregular vibrato and a round attack. Sounds without SEM, on the other hand, were described as 'harsh' and 'forced', too repetitive and predictable, thus being considered less natural and pleasant.

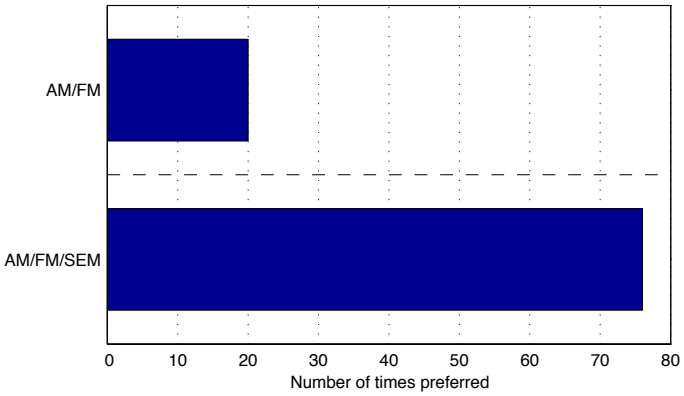

Figure 8. Results of the preference test reveal a significant preference for vibrato sounds with Spectral Envelope Modulation.

The analysis of the verbal data further highlights timbre differences: vibrato sounds with spectral envelope modulation were described as deeper and fuller, and thus more natural and pleasant; whereas sounds with constant average spectral envelope were described as forced, harsh, too repetitive and predictable.

Discussion

Implications on Musical Practice

This generalised model can enhance performers' awareness and understanding of various vibrato features for better analysis and sound production control. It further provides new and separated control parameters on the AM, FM and SEM components, that are perceptually relevant. This enables more intuitive interactions with the model to generate expressive novel sounds on digital instruments. Our model also opens new possibilities for audio processing, more specifically in electroacoustic composition, with more realistic sounds with vibrato or time-scaling modifications of sounds with vibrato.

Implications on Musicological Interdisciplinarity

In an attempt to bridge the gap between definitions of vibrato in various disciplinary fields (musicology, psychoacoustics, and signal processing), we provided a review of vibrato definitions and features. Based on this review, a generalized vibrato model including spectral envelope modulation was developed.

A perceptual evaluation of this vibrato model revealed the perceptual salience of spectral envelope modulation, which resulted in a significant improvement of vibrato sound modeling and synthesis. This

research also provides new insights for vibrato analysis and automatic recognition, since brightness modulation can be inferred from the spectral centroid and the high frequency content variations. This interdisciplinary approach could be beneficial for modeling other stylistic effects (trill, glissando, flatterzung).

Future Works
We propose here some future directions of research:

- analysis of the residual and its modulations during vibrato, as well as its effect on vibrato perception.
- analysis of other sounds: strings (violin), brass (trumpet).
- comparison of preference between AM, FM, AM/FM, AM/SEM, FM/SEM, AM/FM/SEM models.
- perceptual effect of the vibrato shape (perfectly sinusoidal versus natural), and of phase decay between harmonics.
- perceptual effect of time-scaling sounds with vibrato using the two-level sinusoidal model, when taking into account or not the scaling of the jitter/shimmer.

Acknowledgments
Daniel Levitin, Jean-Claude Risset and Gary Scavone for discussions about perception and acoustics.

Appendices

Appendix 1: Definition of Sound Features We define the six features of sounds with vibrato that are depicted on Fig. 1. The fundamental frequency $F_0(n)$ is given as the analysis frequency of the first harmonic $f_1(n)$. The intensity level is given by the instantaneous amplitude, as

$$A(n) = \frac{1}{H}\sqrt{\sum_{h=1}^{H}\left(a_h(n)\right)^2} \tag{46}$$

The odd/even balance is the square root of the ratio between the sum of the odd harmonics' power and the sum of all harmonics' energy:

$$b_{o/e}(n) = \sqrt{\frac{\frac{1}{H^2}\sum_{h=1}^{H/2}\left(a_{2h}(n)\right)^2}{\left(A(n)\right)^2}} \tag{47}$$

The spectral centroid (SGC or spectral gravity center) is correlated to the timbre attribute named brightness, and is computed as the gravity center of the harmonic spectrum, as:

$$cgs(n) = \sqrt{\frac{\sum_{h=1}^{H} a_h(n) f_h(n)}{\sum_{h=1}^{H} a_h(n)}} \tag{48}$$

The high-frequency content is usually used for attack detection ans is computed as:

$$hfc(n) = \frac{1}{H}\sum_{h=1}^{H}\left(a_h(n)\right)^2 f_h(n) \tag{49}$$

The spectral slope is the slope of the linear regression of the harmonic spectrum, *i.e.* the slope of the line that minimizes the distance between itself and the harmonic spectrum.

Appendix 2: Is There Hysteresis the Harmonics' Path? This question, deals with the symmetrical profile of its vibrato. We here wonder if the vibrato have the same behaviour during the rise and during the fall. To answer this question, we consider the ideal case where all the vibrato parameters are constant over time and where the vibrato period is an integer number of time indexes. The assumptions are:

- AM, FM and SEM vibrato parameters are constant: $\tilde{a}_0^a(n) = \tilde{a}_0^a$ and $\tilde{f}_0^a(n) = \tilde{f}_0^a = \tilde{f}_0$; $\tilde{a}_0^f(n) = \tilde{a}_0^f$ and $\tilde{f}_0^f(n) = \tilde{f}_0^f = \tilde{f}_0$; $\tilde{a}_0^e(n) = \tilde{a}_0^e$ and $\tilde{f}_0^e(n) = \tilde{f}_0^e = \tilde{f}_0$,
- partials have constant amplitude and frequency before applying vibrato: $a_h(n) = a_h$ and $f_h(n) = f_h$,

- the flat sound spectral envelope is constant with time: $\mathcal{E}\left(f(n), n\right) = \mathcal{E}\left(f(n)\right)$.

Let us note $\tilde{\phi}_0$ the red initial phase; the phase is then given for current time index n as:

$$\tilde{\Phi}_0(n) = 2\pi\tilde{f}_0\frac{n}{F_e} + \tilde{\phi}_0 \tag{50}$$

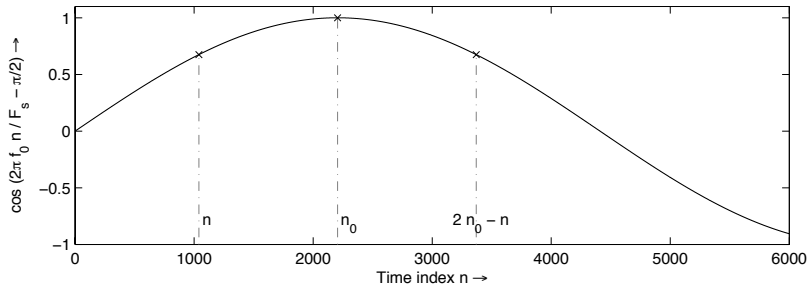

Figure 9. Symmetrical values of the sinusoidal control curve $\gamma(n) = \cos\left(10\pi n/F_e - \pi/2\right)$ at time indexes $n = 1040$ and $2n_0 - n$ (around the maximum at $n_0 = 2205$).

Let us note n_0 the location of the next maximum in the period, in the ideal case where the maximum happens exactly at the time n/F_e. Then, n and $2n_0 - n$ are symmetrical around n_0 (see Fig. 9), with n_0 defined as:

$$\cos\left(2\pi\tilde{f}_0\frac{n_0}{F_e} + \tilde{\phi}_0\right) = 1 \tag{51}$$

$$n_0 = \frac{F_e}{2\pi\tilde{f}_0}\left(2\pi M - \tilde{\phi}_0\right) \tag{52}$$

with M suitably chosen so that n and n_0 belong to the same period. This is equivalent to:

$$\cos\left(2\pi\tilde{f}_0\frac{2n_0 - n}{F_e} + \tilde{\phi}_0\right) = \cos\left(2\pi\tilde{f}_0\frac{n}{F_e} + \tilde{\phi}_0\right) \tag{53}$$

and then implies that:

$$\gamma^a(n) = \gamma^a(2n_0 - n) \tag{54}$$

$$\gamma^f(n) = \gamma^f(2n_0 - n) \tag{55}$$

$$\gamma^e(n) = \gamma^e(2n_0 - n) \tag{56}$$

$$\beta^e(n) = \beta^e(2n_0 - n) \tag{57}$$

$$\overline{f}_h(n) = \overline{f}_h(2n_0 - n) \tag{58}$$

When applying the SEM by *changing the slope*, the new SE is:

$$\overline{\mathcal{E}}\left(\overline{f}_h(n), n\right) = \gamma^e(n)\cdot\overline{f}_h(n)\cdot\mathcal{E}\left(\overline{f}_h(n)\right) \tag{59}$$

$$= \gamma^e(n)\cdot\overline{f}_h\cdot\mathcal{E}\left(\overline{f}_h(n)\right) \tag{60}$$

We compute it for $2n_0 - n$:

$$\overline{\mathcal{E}}\left(\overline{f}_h(2n_0 - n), 2n_0 - n\right) = \gamma^e(2n_0 - n)\cdot\overline{f}_h\cdot\mathcal{E}\left(\overline{f}_h(2n_0 - n)\right) \tag{61}$$

$$= \gamma^e(n)\cdot\overline{f}_h\cdot\mathcal{E}\left(\overline{f}_h(2n_0 - n)\right) \tag{62}$$

The condition for no hysteresis is:

$$\overline{\mathcal{E}}\left(\overline{f}_h(2n_0 - n), 2n_0 - n\right) = \overline{\mathcal{E}}\left(\overline{f}_h(n), n\right) \tag{63}$$

and is equivalent to:

$$\mathcal{E}\left(\overline{f}_h(2n_0 - n)\right) = \mathcal{E}\left(\overline{f}_h(n)\right) \quad (64)$$

This means that the only condition for no hysteresis is to have a flat spectral envelope $\mathcal{E}\left(\overline{f}_h(n)\right)$ on the frequency range $\overline{f}_h(n) \in \left[f_h \cdot \left(1 - \tilde{a}_0^f\right), f_h \cdot \left(1 + \tilde{a}_0^f\right)\right]$ of each partial. In any other condition, the AM/FM/SEM pulsation implies hysteresis on the harmonics' path. In the general case where the pulsation rates and extents are not constant with time, the AM/FM/SEM pulsation always have hysteresis.

When applying the SEM by *interpolating between two spectral envelopes*, the new SE is:

$$\overline{\mathcal{E}}\left(\overline{f}_h(n), n\right) = \beta_e(n) \cdot \mathcal{E}_+\left(\overline{f}_h(n)\right) + (1 - \beta_e(n)) \cdot \mathcal{E}_-\left(\overline{f}_h(n)\right) \quad (65)$$

We compute it for $2n_0 - n$:

$$\overline{\mathcal{E}}\left(\overline{f}_h(2n_0 - n), 2n_0 - n\right) = \beta_e(2n_0 - n) \cdot \mathcal{E}_+\left(\overline{f}_h(2n_0 - n)\right) + (1 - \beta_e(2n_0 - n)) \cdot \mathcal{E}_-\left(\overline{f}_h(2n_0 - n)\right) \quad (66)$$

$$= \beta_e(n) \cdot \mathcal{E}_+\left(\overline{f}_h(2n_0 - n)\right) + (1 - \beta_e(n)) \cdot \mathcal{E}_-\left(\overline{f}_h(2n_0 - n)\right) \quad (67)$$

The condition for no hysteresis given Eq. (63) is equivalent to:

$$\mathcal{E}_+\left(\overline{f}_h(2n_0 - n)\right) = \mathcal{E}_+\left(\overline{f}_h(n)\right) \quad (68)$$

$$\mathcal{E}_-\left(\overline{f}_h(2n_0 - n)\right) = \mathcal{E}_-\left(\overline{f}_h(n)\right) \quad (69)$$

This means that the only condition for no hysteresis is to have two flat spectral envelopes $\mathcal{E}_+\left(\overline{f}_h(n)\right)$ and $\mathcal{E}_-\left(\overline{f}_h(n)\right)$ on the frequency range $\overline{f}_h(n) \in \left[f_h \cdot \left(1 - \tilde{a}_0^f\right), f_h \cdot \left(1 + \tilde{a}_0^f\right)\right]$ of each partial. In any other condition, the AM/FM/SEM pulsation implies hysteresis on the harmonics' path. In the general case where the pulsation rates and extents are not constant with time, the AM/FM/SEM pulsation always have hysteresis.

References

Amatriain, X., de Boer, M., Robledo, E., & Garcia, D. 2002a. CLAM: an OO framework for developing audio and music applications. *In: Companion of the 17th annual ACM SIGPLAN Conf. on Object-oriented Prog., Systems, Languages, and Applications*. 14

Amatriain, X., Bonada, J., Loscos, A., & Serra, X. 2002b. *DAFX - Digital Audio Effects*. U. Zoelzer ed., J. Wiley & Sons. Chap. Spectral Processing, pages 373–438. 14

Arfib, D., & Delprat, N. 1998. Selective transformations of Sound using Time-frequency representations: An Application to the Vibrato Modification. *In: 104th Conv. of the Audio Eng. Soc., Amsterdam*. 5, 7, 8

Bretos, J., & Sundberg, J. 2003. Measurements of Vibrato Parameters in Long Sustained Crescendo Notes as Sung by Ten Sopranos. *Jour. of Voice*, **17**(3), 343–52. 2

Brown, J. C., & Vaughn, K. V. 1996. Pitch center of stringed instrument vibrato tones. *Jour. of the Ac. Soc. of America*, **100**(1), 1728–34. 2

Cycling'74. 2003. *Max/MSP, http://www.cycling74.com/*. 14

d'Alessandro, C., & Castellengo, M. 1994. The pitch if Sort-Duration Vibrato Tones. *Jour. of the Ac. Soc. of America*, **95**(3), 1617–30. 2

Desain, P., & Honing, H. 1996. Modeling continuous aspects of music performance: Vibrato and portamento. *In: Proc. Int. Conf. on Music Perception and Cognition*. 2

Desain, P., Honing, H., Aarts, R., & Timmers, R. 1999. *Rhythm Perception and Production*. P. Desain and W. L. Windsor (eds.), Lisse: Swets & Zeitlinger. Chap. Rhythmic aspects of vibrato, pages 203–16. 2

Galas, T., & Rodet, X. 1990. An improved cepstral method for deconvolution of source-filter systems with discrete spectra: Application to musical sounds. *Pages 82–8 of: Proc. of the Int. Computer Music Conf. (ICMC'90), Glasgow*. 6

Garnier, M., Henrich, N., Castellengo, M., Dubois, D., & Poitevineau, J. 2004. Perception et description acoustique de la qualité vocale dans le chant lyrique: une approche cognitive. *In: Journées d'Étude sur la Parole*. 1

Herrera, P., & Bonada, J. 1998. Vibrato extraction and parameterization in the Spectral Modeling Synthesis framework. *In: Proc. of the COST-G6 Workshop on Digital Audio Effects (DAFx-98), Barcelona, Spain*. 5, 7, 8

Horii, Y. 1989. Frequency modulation characteristics of sustained /a/ sung in vocal vibrato. *Jour. Speech and Hearing Research*, **32**, 829–36. 2

IOWA. 2005. *http://theremin.music.uiowa.edu/MIS.html*. 13

Järveläinen, H. 2002. Perception-based control of vibrato parameters in string instrument synthesis. *Pages 287–294 of: Proc. Int. Computer Music Conf.* 2, 8

Maher, R. C., & Beauchamp, J. 1990. An Investigation of Vocal Vibrato for Synthesis. *Applied Acoustics*, **30**, 219–45. 5, 7, 8, 9, 13

Marchand, S., & Raspaud, M. 2004. Enhanced Time-Stretching Using Order-2 Sinusoidal Modeling. *Proc. of the Int. Conf. on Digital Audio Effects (DAFx-04), Naples, Italy*, 76–82. 6, 8

Mathews, M., & Kohut, J. 1973. Electronic Simulation of Violin Resonances. *Jour. of the Ac. Soc. of America*, **53**(6), 1620–6. 5, 7

McAulay, R. J., & Quatieri, T. F. 1986. Speech Analysis/Synthesis Based on a Sinusoidal Representation. *IEEE Trans. on Acoustics, Speech, and Signal Processing*, **34**(4), 744–54. 6

Mellody, M., & Wakefield, G. 2000. The Time-Frequency characteristic of violon vibrato: modal distribution analysis and synthesis. *Jour. of the Ac. Soc. of America*, **107**, 598–611. 2

Prame, E. 1997. Vibrato Extent and Intonation in Professional Western Lyric Singing. *Jour. of the Ac. Soc. of America*, **102**(1), 616–21. 2

Puckette, M. 1991. Combining Event and Signal Processing in the MAX Graphical Programming Environment. *Computer Music Jour.*, **15**(3), 68–77. 14

Risset, J.-C. 1965. Computer Study of Trumpet Tones. *Jour. of the Ac. Soc. of America*, **33**, 912. 12

Risset, J.-C., & Wessel, D. L. 1999. *Exploration of timbre by analysis and synthesis*. D. Deutsch, Academic Press, New York. Pages 113–69. 9

Rossignol, S., Depalle, P., Soumagne, J., Rodet, X., & Collette, J.-L. 1999. Vibrato: Detection, Estimation, Extraction, Modification. *In: Proc. of the COST-G6 Workshop on Digital Audio Effects (DAFx-99), Trondheim, Norway*. 5, 7, 8

Seashore, C. E. 1932. The Vibrato. *University of Iowa studies, New series*, **225**. 1, 8, 9

Seashore, C. E. 1936. Psychology of the Vibrato in Voice and Speech. *Studies in the Psychology of Music*, **3**, 212–9. 8

Serra, X., & Smith, J. O. 1990. A Sound Decomposition System Based on a Deterministic plus Residual Model. *Jour. of the Ac. Soc. of America, sup. 1*, **89**(1), 425–34. 6

Shonle, J. I., & Horan, K. E. 1980. The Pitch of Vibrato Tones. *Jour. of the Ac. Soc. of America*, **67**, 246–52. 2

Sundberg, J. 1987. *The Science of the Singing Voice*. Dekalb, IL: Northern Illinois University Press. 3, 5, 7

Timmers, R., & Desain, P. 2000. Vibrato: Questions and Answers From Musicians and Science. *In: Proc. Int. Conf. on Music Perception and Cognition*. 2

Toff, N. 1996. *The Flute Book, A complete Guide for Students and Performers*. Oxford University Press, New York. Chap. Vibrato, pages 106–15. 1, 9