TRANSFORMING VIBRATO EXTENT IN MONOPHONIC SOUNDS

A. Roebel
Analysis/Synthesis team
STMS - IRCAM CNRS UPMC
Paris, France
axel(dot)roebel@ircam(dot)fr

S. Maller, *
Analysis/Synthesis team
STMS - IRCAM CNRS UPMC
Paris, France
simon.maller@audionamix(dot)com

J. Contreras
Analysis/Synthesis team
STMS - IRCAM CNRS UPMC
Paris, France

ABSTRACT

This paper describes research into signal transformation operators allowing to modify the vibrato extent in recorded sound signals. A number of operators are proposed that deal with the problem taking into account different levels of complexity. The experimental validation shows that the operators are effective in removing existing vibrato in real world recordings at least for the idealized case of long notes and with properly segmented vibrato sections. It shows as well that for instruments with significant noise level (flute) independent treatment of noise and harmonic signal components is required.

1. INTRODUCTION

Signal parameter transformation is one of the main interests in professional audio applications. Today there exist numerous methods allowing the independent transformation of pitch, spectral envelope, and duration achieving in many cases signal quality that is not distinguishable form a real recording, and accordingly new possibilities can be explored. One of the potentially very interesting objectives is expressive signal transformation. Under this term we will understand a dynamic signal transformation that has the objective to add or change the perceived affective state of the sound. Expressive speech signal synthesis has become an increasingly active research area [1, 2, 3, 4, 5] and expressive music signal synthesis has already become a reality for instrument sound signal synthesis [6][1]. Compared to the expressive sound synthesis the topic of expressivity transformation is significantly more complex because we do not have as much control and information about the sound signal to be transformed, and moreover, we may have to remove an existing expressive parameter variation to be able to introduce the desired expressivity. For expressive speech transformation there exist only rather few approaches [7, 8] that generally start with a neutral input signal.

In the present paper we will investigate into the problem of manipulating vibrato, which one of the important expressive means for the case of musical input signals. There exist quite a few studies that are related to the manipulation of vibrato [9, 10, 11, 12, 13, 14, 15]. A common starting point for all these studies is the estimation of the fundamental frequency, which is then followed by a separation into a slowly evolving fundamental frequency contour and the fast ornamentation contour representing the vibrato. The approaches then diverge and establish different control strategies for the adaption of the partial amplitudes and frequencies. The objective of vibrato manipulation can be diverse. The most ambitious case requires independent manipulation of the vibrato parameters *extent* (vibrato amplitude), *rate* (modulation frequency) and *form* (contour). While the order-2 sinusoidal model introduced in [13] provides theoretical means to control the vibrato form, to the best of our knowledge, this has not yet been considered in a practical implementation. Vibrato rate manipulation can be obtained by means of pitch shifting the modulated parameter contours using again either the order-2 sinusoidal model, or by means of the analytic signal of the modulated parameter [11], or by means of using time domain OLA techniques combined with resampling.

In the present study we limit ourselves to the problem of the modification of the vibrato extent. The basic idea of the method is conceptually simple. It consists of applying the separation into slowly varying contour components and modulated ornamentation not only to the fundamental

* Now with Audionamix, Paris

[1] This side by side presentation of expressive speech and music synthesis does not intend to imply that expressive effects will be generated by means of similar means for these two signal classes.

frequency but to all other sound parameters as well. The vibrato control manipulation will then become a simple mixing of the modulated parameter component. This strategy can in principle be applied to sinusoidal partial parameters (amplitude and frequency trajectories) as well as to the complete spectral envelope of the modulated sound. The spectral envelope based method allows to unify the treatment of the sinusoidal and noise modulations and therefore it will be used in the following discussion. When compared to methods using explicit sinusoidal models as for example [13, 14] the proposed method is simpler yet rather efficient. We note that nearly all existing studies focus entirely on the effect of the vibrato on the sinusoidal components. There exist only very few studies dealing with the effect of the vibrato on the noise components of the sound [16]. The present experimental investigation shows that the modulation of sinusoidal and noise components that is induced by vibrato is not synchronous. Therefore, the effect of the vibrato on the noise component cannot be neglected when a perceptually convincing removal of the vibrato is desired.

For the experimental investigation we have selected to use the problem of removing vibrato from a monophonic music instrument recording. The idea is that removing signal modulations is perceptively the most critical operation because any residual modulation will be easily perceived. Due to the fact that the presence or absence of modulation can be judged rather easily even when only a low level of modulation is present the problem allows for relatively simple perceptual evaluation. Please note as well, that the investigation is only concerned with the transformation of the extent of the vibrato ornamentation. Many other problems exist that should be addressed. Examples are the problems related to the detection of vibrato [17, 18], the change of the vibrato rate when time stretching a signal, the coherent transformation of note transitions, and the generation of concrete pitch and spectral envelope modulation contours if vibrato should be added to an unmodulated signal.

The organization of the article is as follows. In section **2** the model of vibrato generation is presented and discussed. Based on this model section **3** introduces dynamic signal transformation operators that allow to remove the effects related to vibrato with different levels of refinement. In section **4** these operators are experimentally evaluated using a real world flute signal and section **5** describes the conclusions that can be drawn from the investigation.

2. SIGNAL MODEL FOR VIBRATO ORNAMENTATION

Vibrato is an expressive ornamentation that is frequently used in music [19, 14]. The term vibrato is generally understood to refer to a quasi periodic modulation of the pitch or fundamental frequency. But, as will become clear in the following, pitch modulation without energy and timbre modulation does not exist in real world sound sources. Therefore, a realistic manipulation of vibrato requires a coherent transformation of timbre and energy modulations as well. We have selected vibrato for our experimentation with expressive signal transformation because: first it is widely used, second it includes effects like energy and timbre modulation and third, due to its periodic nature it is perceptually much easier to evaluate to access than for example note transitions. Transformation of expressive note transitions are currently under investigation and will not be addressed in the present study.

Vibrato is generally produced by means of a periodic variation of the fundamental frequency of the musical note. Accordingly, the vibrato signal has a quasi-periodic variation of the fundamental frequency around a central value. Perceptually, this modulation does not directly affect the perceived pitch of the note [19]. The perceived pitch of vibrato notes has been investigated repeatedly and has been found to be given by the a value close to the mean pitch over a vibrato period [20]. For short vibrato of less than 2 vibrato periods the perceived pitch is affected more by the end of the pitch evolution [21]

Different signal models including pitch variation with constant spectral envelope, pitch and amplitude modulation as well as pitch, amplitude and spectral envelope modulation have been compared in [14, 15]. In a natural physical musical instrument (including singing voice) the modification of the fundamental frequency will necessarily be accompanied by amplitude and timbre modulation. On one hand the partials are moving continuously through the different resonances of the sound source resonator which will generate an induced timbre modulation. The change of the pitch will moreover require a reconfiguration of the physical configuration of the excitation and the resonating parts leading to a more or less pronounced timbre modulation. Perceptual experiments have shown that the timbre modulation that comes with the vibrato modulation is perceptually more important than the fundamental frequency modulation itself [12] and that vibrato without independent timbre modulation does not sound natural and physically plausible [14]. An interesting question that arises is related to the modulation of the noise components of the signal. Noise components do not necessarily take the same acoustic path and therefore, we assume in the following that the modulation of the noise components will not be in sync with the modulation of the sinusoidal components.

Based on the results discussed so far we will establish our vibrato transformation algorithm taking into account pitch as well as amplitude and timbre modulation. For each partial we assume a amplitude trajectory that is given by

$$A_k(n) = \exp(S(w_k(n), n)), \tag{1}$$

where A_k represents the time varying amplitude of partial k, n is the discrete time, $S(w,n)$ is a quasi-periodically time-varying spectral envelope obtained from the log amplitude spectrum that modulates due to the physical reconfiguration of the instrument and w_k is the quasi-periodically time-varying frequency of partial k. We note that the existing amplitude modulation and also potential amplitude estimation errors that may be related to the modulation are absorbed into the time-varying spectral envelope S. Similarly for the noise component we assume a time varying distribution of the log amplitude spectrum given by $N(w,n)$ that may have a periodic structure in n.

3. VIBRATO TRANSFORMATION

The objective of the present section is to develop an algorithm that allows continuously controlling the vibrato extent. This control is performed for the part of the note that is located between attack and release and the region of attack and release are assumed to be known. In the following experiments the attack and release segments have been obtained by means of manual labeling.

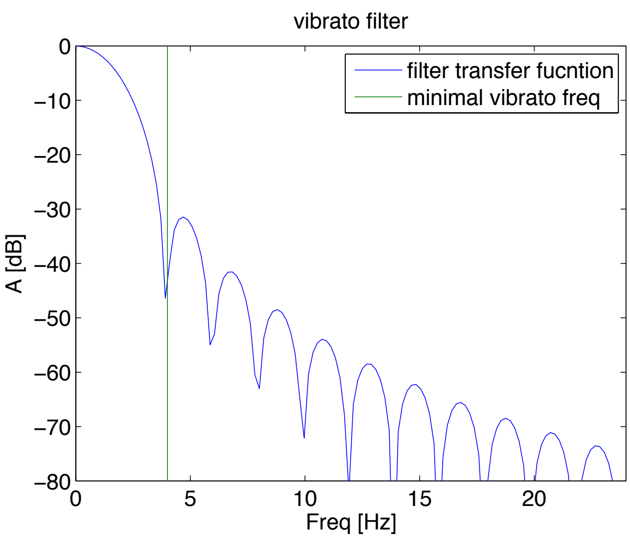

Figure 1: *Vibrato filter transfer function for a vibrato filter assuming minimum vibrato frequency of 4Hz and 2 filter length of 2 times the maximal vibrato period.*

To achieve this control the time-varying fundamental frequency and the time varying spectral envelope S or N will be represented by means of a superposition of a constant and a quasi periodic part. The constant part is derived by means of smoothing the time varying evolution of $w(n)$ and $S(w,n)$ over the local period of the fundamental and the quasi periodic part is then the difference between the original function and the smoothed version. Accordingly, the vibrato control can be achieved by means of synchronously scaling the dynamic part of the fundamental frequency and the spectral envelope as follows

$$\begin{aligned} w_k(n) &= \bar{w}_k(n) + \alpha\tilde{w}_k(n) && (2) \\ S(w,n) &= \bar{S}(w,n) + \alpha\tilde{S}(w,n). && (3) \\ N(w,n) &= \bar{N}(w,n) + \alpha\tilde{N}(w,n). && (4) \end{aligned}$$

Here $\bar{w}$, $\bar{S}$ and $\bar{N}$ are low pass filtered versions of w, $S(w,n)$, and $N(w,n)$ and $\tilde{w}$, $\tilde{S}$, and $\tilde{N}$ are the respective high-pass residuals that represent the modulated part of the related parameter. α is the control parameter that can be used to scale the modulation. Note, that the spectral envelopes are using log amplitude representation such that the variation is scaled geometrically. When compared to existing approaches for manipulation of the spectral envelop modulation the proposed scaling operator has a number of advantages. We discuss 2 examples. [9] proposed to find the target contour of the transformed envelope by means of interpolating between two waveform tables. The two tables were obtained at the time positions related to the extreme values of the fundamental frequency that were determined for each vibrato period. As interpolation control variable they used the instantaneous fundamental frequency. [15] uses a similar principle interpolating two spectral envelopes obtained from the extreme positions of the fundamental frequency. They proposed to control the linear interpolation by means of a sinusoidal interpolation contour. Both approaches suffer from 2 problems. First, they make a hypothesis about the form of the fundamental frequency contour, notably that there exist two robustly identifiable positions that can be used to anchor the interpolation. Second, and more importantly, they cannot represent the original signal. Even if no signal transformation is desired the model would still not create the unmodified spectral envelope. In contrast to this, the transformation proposed in eq. (3) will always be a neutral operation for $\alpha = 0$ and second it will preserve the local form and evolution of the envelope modulation even for $\alpha ! = 0$ as well as the phase relations between the modulations of the fundamental and all the partial amplitudes.

In the following we will assume that the sound sources are quasi harmonic such that the scaling of the frequency trajectories can be obtained by means of dynamic transposition. Accordingly, instead of operating the frequency trajectories of the individual partials only the fundamental frequency trajectory $w_0(n)$ needs to be treated[2].

3.1. Extracting modulated parameter contours

To achieve the separation between the two parameter contours that are required for each parameter to establish eq. (3) we propose to use a FIR filter that is given by a standard window function having at least 2 times the length of

[2] Note however, that the same principles could be applied to the individual amplitude and frequency parameter contours of a sinusoidal model.

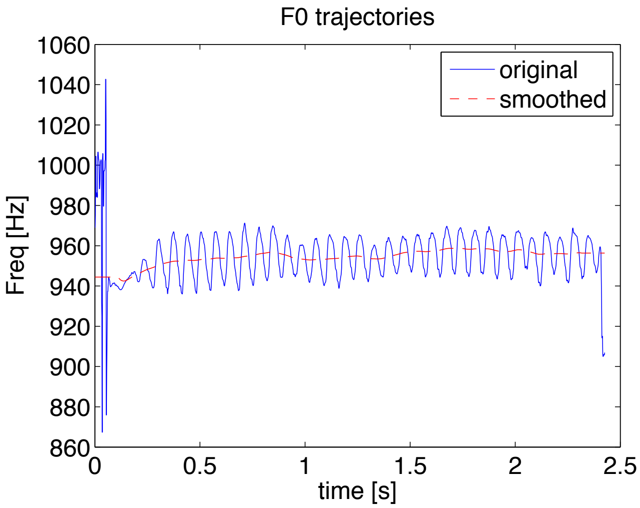

Figure 2: *Result of vibrato suppression on a real world vibrato note of a flute signal using vibrato filter shown in fig. 1. See text for more explanations.*

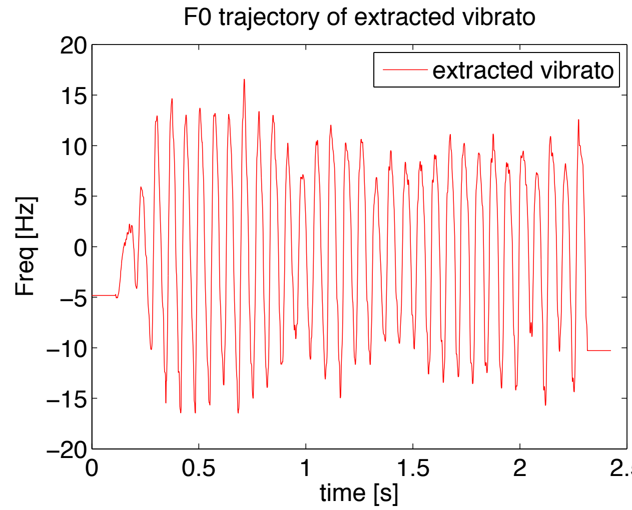

Figure 3: *The vibrato residual obtained by means of subtracting the smoothed f0 trajectory from the original trajectory shown in fig. 2. See text for more explanations.*

the period of the minimal vibrato frequency. The window is normalized to have a sum of 1. The advantages of this filter specification are the facts that

- the filter has linear phase
- the transfer function can be dynamically adapted to the vibrato frequency such that the zeros of the transfer function are located at the harmonics of the vibrato frequency.

In the following we use a Hanning window, but other window functions may be used. fig. 1 shows the transfer function obtained by means of using a Hanning window of 0.5s which is 2 times the period of a minimum vibrato frequency of 4Hz. The minimum amplitude rejection above the vibrato frequency limit is 30dB. Note that fundamental frequency variations of about 3Hz are already attenuated by about 20dB.

The application of the vibrato suppression filter to a flute vibrato of about 6Hz is shown in fig. 2. In this example the limiting parts of the smoothed f0 trajectory have been replaced by constant values for the range of values where the window is not completely inside the trajectory. The residual that results after subtracting the smoothed F0 trajectory from the original F0 trajectory is shown in fig. 3. The samplerate of the F0 trajectory in the present and all following cases is 100Hz.

3.1.1. Determining the target F0 contour

In the following section the relation between the original and smoothed fundamental frequency trajectories will be used to derive the signal transformations that are necessary to obtain a convincing control of the vibrato extent. To achieve the control of the vibrato extent the control parameter α is used and a segmentation of the fundamental frequency sequence into segments with and without vibrato is assumed [17]. The segments with vibrato are collected into a group of segments that are accessed individually by means of index k. This parameter α is then used to obtain the target fundamental frequency for segment k as follows

$$F_{k,target}(n, \alpha) = (F_k(n) - \bar{F_k}(n))\alpha + \bar{F_k}(n), \quad (5)$$

where $F_k(n)$ is the original fundamental frequency trajectory, $\bar{F_k}(n)$ its smoothed counter part and $F_{k,target}(n, \alpha)$ the fundamental frequency target trajectory. Accordingly, by means of setting $\alpha = 1$ the fundamental frequency is kept unchanged, with $\alpha = 0$ the vibrato is removed, $\alpha = -1$ the vibrato is inverted and with $\alpha = 2$ the vibrato extent can be amplified.

3.2. Signal transformation

In this section a number of vibrato transformation algorithms will be developed that provide an increasing level of refinement with respect to the effects that can be taken into account. The idea here is to experimentally evaluate the different strategies on the perceptually most demanding transformation that removes the vibrato completely ($\alpha = 0$). The different strategies will then be evaluated on a short example such that the required level of refinement for high quality vibrato control can be determined.

The transposition to be used to transform the pitch of the sound signal is given by

$$T_k(n) = \frac{F_{k,target}(n, \alpha)}{F_k(n)} \tag{6}$$

3.2.1. PS: preservation of spectral envelope

The first level of refinement consists of the application of the dynamic transposition specified in eq. (6) under preservation of the spectral envelope. This preservation is done by means of estimating the spectral envelope with the spectral envelope estimator proposed in [22, 23]. The related transformation operator is denoted as PS(α). This operator allows to take into account the timbre modulation that is induced by the modulation of the fundamental frequency assuming that the spectral envelope does not change with the modulation of the fundamental frequency.

3.2.2. PSCS: correction of spectral envelope modulation

In case that the modulation of the spectral envelope is perceptually significant after vibrato modification a refinement of the previous model is needed that allows to remove the modulation of the spectral envelope. This can be achieved by means of smoothing the spectral envelope. Similar to the procedure described for calculating the transformed fundamental frequency the PSCS(α) operator calculates the target spectral envelope by means of eq. (3). The extraction of the non modulated part $\bar{S}(w, n)$ of the spectral envelope is done similar as for the F0 trajectory besides that here we use a simple rectangular window having the length of approximately the spectral vibrato period. Again the modulated part of the spectral envelope $\tilde{S}(w, n)$ is obtained as residual after subtraction $\bar{S}(w, n)$ from the original envelope. Note, that the demodulation proposed here for spectral envelope smoothing will at the same time remove amplitude modulations and timbre modulations related to local deformation of the spectral envelope.

3.2.3. PSCTRT/PSCTNRT: correction of timbre modulation with and without residual transformation

The PSCS operator allows taking into account the modulation of the spectral envelope. In the experimental investigation we found, however, that noise and harmonic components do not follow the same modulation pattern. Accordingly, the smoothing of the spectral envelope will not remove the modulation in the non harmonic parts of the spectrum. To be able to handle 2 different modulation patterns in the noise and harmonic spectrum two variants for correction of timbre modulation have been developed. Both versions, the PSCTRT(α) as well as the PSCTNRT(α) operator establish a separation of the signal into harmonic components and residual noise. This separation of harmonic and non harmonic signal components is done with a standard harmonic sinusoidal model [24]. The harmonic component of the sound is treated with the PSCS(α) operator. The target envelope for the noise component is again calculated using eq. (3) but in the case the noise envelope is calculated using a small order AR model of the residual. In the present experiment we use an AR order of 20. The separation into modulated and non-modulated noise filter is done strictly equivalent to the method discussed above. The difference of the PSCTRT and the PSCTNRT operator is then that the former transposes the residual with the harmonic signal when the fundamental frequency modulation is removed and the latter extracts the residual before transposition such that the residual signal is not transposed. The difference is expected to be important when the residual signal contains sinusoidal components that may or may not be modulated with the fundamental frequency.

4. EXPERIMENTAL RESULTS

In this experimental evaluation the different versions of vibrato manipulation operators have been applied to different signals, e.g. singing, cello, and flute. In the following the results obtained for the flute signal will be presented because for this signal a version with and without vibrato was available. Moreover, for the flute signal a strong noise component is present such that all the different strategies for the treatment of timbre modulation can be evaluated. All signal transformations have been performed with an enhanced phase vocoder implementation that has been developed at IRCAM. Results obtained are not bound to this technology however, and many other approaches can be used to achieve the same results. The separation of sinusoidal and residual components has been performed using a harmonic sinusoidal model using a parameter estimation based on quadratically interpolated amplitude of spectral peaks [25] without any bias correction [26].

In the top row of fig. 4 the spectrogram of a short extract of two original flute signals played by the same player on the same flute with and without vibrato is shown. The pitch of both notes is approximately at 950Hz. The 4 spectrogram displayed below are extracts of the signals transformed using the 4 vibrato modification operators defined above. In the center left position the result of operator PS(0) is displayed[3]. The fundamental frequency modulation has been removed but all partials as well as the noise still present a clear amplitude modulation. Perceptually this modulation is rather strong leading to a very unsatisfying result. In the figure in center right position the results after additional transformation of the spectral envelope modulation by means of

[3]For sound examples please visit `http://anasynth.ircam.fr/home/english/media/dafx11-demo-vibrato-removal`

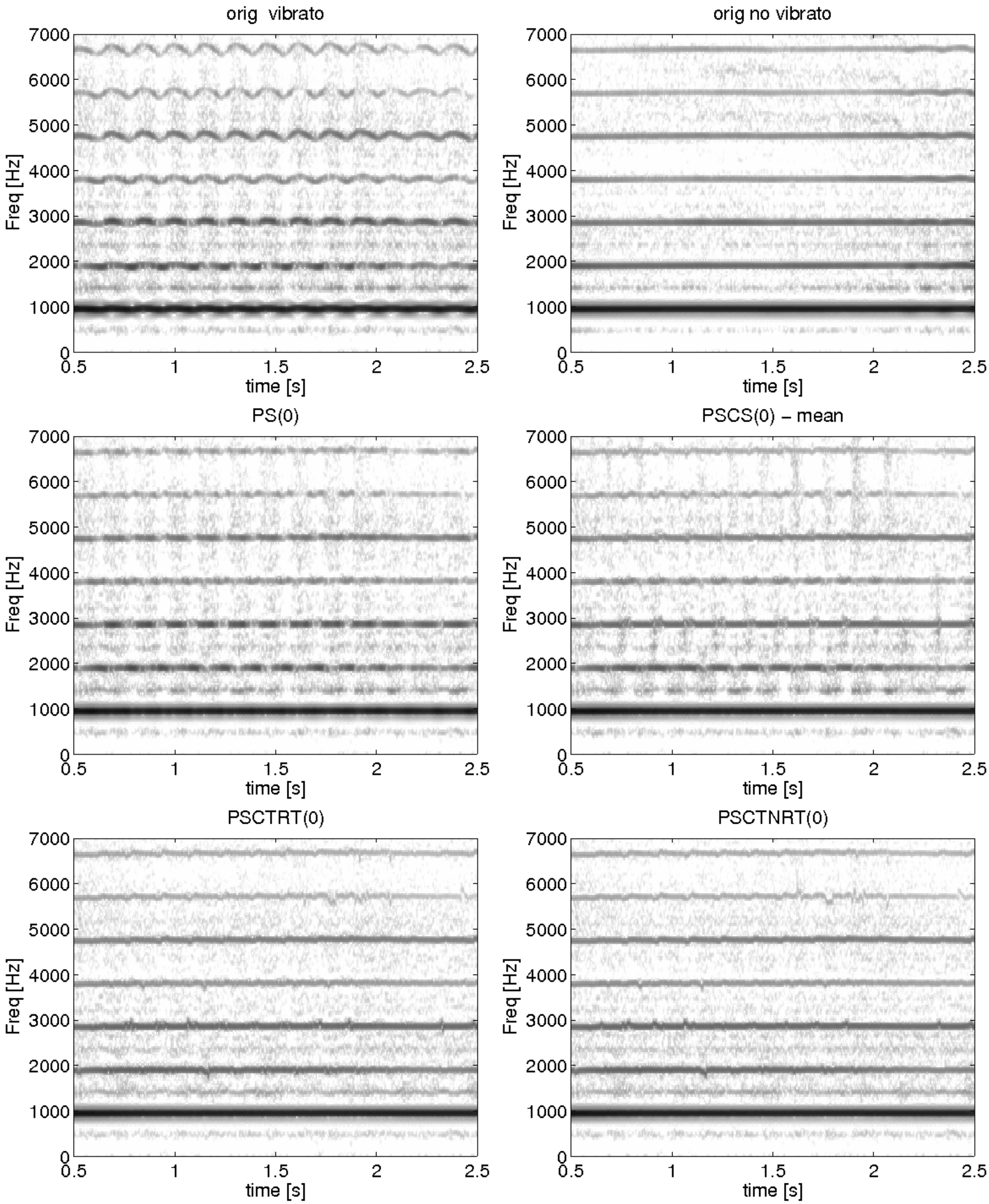

Figure 4: *Experimental results for vibrato removal using the different operators described in the text. The top left figure shows a spectrogram of 2 seconds of a flute signal with vibrato, on the top right the same flute playing the same note without vibrato. The bottom 4 figures show the spectrogram of 4 different vibrato extent transformation operators introduced in the text. See there for detailed explanations.*

the PSCS operator is shown. The amplitude modulation of the partials is significantly reduced. Note however, that the amplitude modulation of the noise is now significantly stronger. This is clearly visible for partials 3 and 7 that in the spectrogram of the original signal with vibrato both exhibit a strong amplitude modulation while the surrounding noise does not or hardly modulate in amplitude. After transformation the noise modulation around these 2 partials is significantly increased. Partial 5 in contrast does not suffer from this effect because partial and surrounding noise have approximately similar amplitude modulation.

The last row of fig. 4 shows the results obtained with the complex operators that separate harmonic and residual components and separately treat the spectral envelope modulation of both components. Both operators achieve a signal with a spectrogram that is very close to the flute signal with no vibrato in the top right figure. Both operators suffer from partial tracking problems in the areas where partials disappear in noise. This effect is perceptually weak, however, when carefully listening with headphones it can be perceived. It should be noted however, that it is easy to imagine a situation where this effect becomes a perceptual problem in which case a more complex transformation operator allowing for partial reconstruction would be required. The PSCSRT operator exhibits a slight modulation of the noise resonance that is located between partials 1 and 2 which is due to the fact that this operator applies the dynamic transposition that is required to remove the frequency modulation to both residual and harmonic component. In the informal listening tests with expert listeners this difference is hardly perceivable.

5. CONCLUSIONS

The present paper investigated into the problem of modification of the vibrato extent assuming that segments with vibrato are marked by a preceding analysis. The proposed signal operators are entirely based on spectral envelope smoothing operations, and therefore they are conceptually relatively simple. Especially they do not require a manipulation of individual partial parameters. Despite this simplicity the operators did allow us to achieve perceptually very convincing results when applied to vibrato manipulation. They have been evaluated on a task dealing with the removal of vibrato from a flute signal. The results demonstrate that for the flute signal the most complex operator is required to achieve optimal and natural results because the harmonic and noise envelope components do not follow the same amplitude modulation pattern. The requirement for the additional complexity that is related to the separate treatment of noise and sinusoidal modulations depends on the signal - it can be avoided if the noise component is weak. Further studies are needed that deal with the connection of the modified vibrato to the neighboring note transitions. In the near future we will investigate into generative instrument specific models that relate vibrato and timbre modulation and we will investigate into signal transformation operators that allow manipulating the vibrato rate.

6. ACKNOWLEDGMENTS

Many thanks to composer Alain Lithaud who provided the flute recordings for this investigation.

7. REFERENCES

[1] M. Schröder, “Emotional speech synthesis: A review,” in *Proc. EuroSpeech*, 2001, pp. 561–564.

[2] F. Burkhardt, “Emofilt: the simulation of emotional speech by prosody-transformation,” in *Interspeech*, 2005, pp. 509–512.

[3] J. Tao, Y. Kang, and A. Li, “Prosody conversion from neutral speech to emotional speech,” *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 14, no. 4, pp. 1145–1154, 2006.

[4] M. Bulut, S. Lee, and S. Narayanan, “Analysis of emotional speech prosody in terms of part of speech tags,” in *Interspeech*, 2007, pp. 626–629.

[5] C.-H. Wu, C.-C. Hsia, C.-H. Lee, and M.-C. Lin, “Hierarchical prosody conversion using regression-based clustering for emotional speech synthesis,” *IEEE Trans. Audio, Speech and Lang. Proc.*, vol. 18, no. 6, pp. 1394–1405, 2010.

[6] E. Lindemann, “Music synthesis with reconstructive phrase modeling,” *IEEE Signal Processing Magazine*, vol. 24, no. 2, pp. 80–91, 2007.

[7] G. Beller, *Analyse et Modèle Génératif de l’Expressivité, Application à la Parole et à l’Interprétation Musicale*, Ph.D. thesis, Université Paris VI UPMC, 2009.

[8] G. Beller, “Expresso: Transformation of expressivity in speech,” in *Proc. Speech Prosody*, 2010.

[9] R. Maher and J. Beauchamp, “An investigation of vocal vibrato for synthesis,” *Applied Acoustics*, vol. 30, pp. 219–245, 1990.

[10] E. Tellman, L. Haken, and B. Holloway, “Timbre morphing of sounds with unequal numbers of features,” *Journal of the Audio Engineering Society*, vol. 43, no. 9, pp. 678–689, 1995.

[11] D. Arfib and N. Delprat, "Selective transformations of sounds using time-frequency representations - application to the vibrato modification," AES Preprint 4652, 1998.

[12] M. Mellody and G.H. Wakefield, "The time-frequency characteristics of violin vibrato: Modal distribution analysis and synthesis," *Jour. of the Acoustic Society of America*, vol. 107, no. 1, pp. 598–611, 2000.

[13] S. Marchand and M. Raspaud, "Enhanced time-stretching using order-2 sinusoidal modeling," in *Proc. 7th Int. Conf. on Digital Audio Effects (DAFx)*, 2004, pp. 76–82.

[14] V. Verfaille, C. Guastavino, and P. Depalle, "Perceptual evaluation of vibrato models," in *Proceedings of the conference on interdisciplinaire musicology (CIM05)*, 2005.

[15] V. Verfaille, C. Guastavino, and P. Depalle, "Évaluation de modèles de vibrato," *Cahiers de la Société Québécoise de Recherche en Musique*, vol. 9, no. 1/2, pp. 49–62, 2007.

[16] V. Verfaille, C. Guastavino, and P. Depalle, "Control parameters of a generalized vibrato model with modulations of harmonics and residual," in *Acoustics'08*, 2008.

[17] S. Rossignol, X. Rodet, P. Depalle, J. Soumagne, and J.-L. Collette, "Vibrato: detection, estimation, extraction, modification," in *Proc. 2nd Int. Conf. on Digital Audio Effects (DAFx)*, 1999.

[18] P. Herrera and J. Bonada, "Vibrato extraction and parameterization in the spectral modeling synthesis framework," in *Proc. COST-G6 Workshop on Digital Audio Effects (DAFx)*, 1998.

[19] R. Timmers and P. Desain, "Vibrato: the questions and answers from musicians and science," in *Proceedings of the Sixth International Conference on Music Perception and Cognition*, 2000.

[20] J.I. Shonle and K.E. Horan, "The pitch of vibrato tones," *Journal Acoust. Soc. Am.*, vol. 67, no. 1, pp. 246–252, 1980.

[21] C. d'Alessandro and M. Castellengo, "The pitch of short-duration vibrato tones," *Journal Acoustic Society America*, vol. 95, no. 3, pp. 1617–1631, 1994.

[22] A. Röbel, F. Villavicencio, and X. Rodet, "On cepstral and all-pole based spectral envelope modeling with unknown model order," *Pattern Recognition Letters, Special issue on Advances in Pattern Recognition for Speech and Audio Processing*, pp. 1343–1350, 2007.

[23] A. Röbel and X. Rodet, "Efficient spectral envelope estimation and its application to pitch shifting and envelope preservation," in *Proc. of the 8th Int. Conf. on Digital Audio Effects (DAFx05)*, 2005, pp. 30–35.

[24] X. Amatriain, J. Bonada, A. Loscos, and X. Serra, "Spectral processing," in *Digital Audio Effects*, U. Zölzer, Ed., chapter 10, pp. 373–438. John Wiley & Sons, 2002.

[25] R. J. McAulay and T. F. Quatieri, "Speech analysis-synthesis based on a sinusoidal representation," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 34, no. 4, pp. 744–754, 1986.

[26] A. Röbel, "Frequency slope estimation and its application for non-stationary sinusoidal parameter estimation," in *Proc. of the 10th Int. Conf. on Digital Audio Effects (DAFx'07)*, 2007.