Provided for non-commercial research and education use.
Not for reproduction, distribution or commercial use.

This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

http://www.elsevier.com/copyright

Author's personal copy

Signal Processing 89 (2009) 24–30

Contents lists available at ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

Analyzing several musical instrument tones using the randomly modulated periodicity model

Shlomo Dubnov [a], Melvin J. Hinich [b,*]

[a] *Department of Music, University of California at San Diego, CA, USA*
[b] *Applied Research Laboratories, University of Texas at Austin, P.O. Box 8029, Austin, TX 78713-8029, USA*

ARTICLE INFO

Article history:
Received 13 December 2007
Received in revised form
9 June 2008
Accepted 3 July 2008
Available online 18 July 2008

Keywords:
Signal coherence
Signal-to-modulation-noise ratio
Random amplitude modulation
Vibrato
Musical instruments

ABSTRACT

The waveform of a simple sustained tone emitted from musical instruments such as a flute played by even the best musician is never exactly periodic. There is always some variation over time in the waveform of a single pitch that is characteristic of the instrument itself. In this paper, we employ the signal-coherence function introduced by Hinich [A statistical theory of signal coherence, J. Oceanic Eng. 25(2) (2000) 256–261] to study the subtle variation of tones from several instruments played by accomplished musicians. This measure characterizes the amount of variation in each Fourier component as a random amplitude-modulation component added to a coherent narrowband sinusoid. The signal-coherence function is computed from several digitized acoustic signals of several musical instruments. The signal-coherence functions show that there are important differences between the same notes produced from different instruments. The signal coherence of a vibrato, a deliberate modulation of a tone, is analyzed for the first time using the signal-coherence function. We show that for most practical playing conditions it has a small effect for lower frequencies. This allows characterization of modulation variations in sustained portions of sound with and without vibrato. The signal-coherence processing method applied to musical acoustics could lead to more realistic music synthesizers.

© 2008 Elsevier B.V. All rights reserved.

1. Introduction

This paper presents a model for studying the fluctuations of musical instrument tones that has never been used before to study the acoustics of musical instruments. The approach is based on the random modulated periodicity (RMP) model [1–4]. In this paper, we employ the signal-coherence function (SCF) to study the subtle variation of tones from several instruments played by accomplished musicians.

The waveform of a simple sustained tone emitted from a musical instruments such as a flute played by even the best musician is never exactly periodic [5–7]. There is always some variation over time in the waveform of a single pitch that is characteristic of the instrument itself. The RMP characterizes the amount of variation in each Fourier component as a random modulation component added to a coherent narrowband sinusoid. These deviations from periodicity of digital recordings of several musical instruments playing the same notes exhibit surprising variation in the shape of the waveform over time. Such notes recorded in laboratory conditions are important for understanding the random fluctuations that naturally appear in musical sounds. The RMP-based signal-processing method applied to musical acoustics could lead to more realistic music synthesizers.

Contrary to previous works on RMP, we explicitly include in this work the analysis of period errors, such as unknown vibrato, as a possible source of deviation from periodicity. In the paper we analyze the influence of both

* Corresponding author. Tel.: +1 512 471 5121; fax: +1 512 835 3259.
E-mail address: hinich@mail.la.utexas.edu (M.J. Hinich).

0165-1684/$ - see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.sigpro.2008.07.006

Author's personal copy

random and deterministic frequency modulations and show that for a perfectly periodic signal their (theoretical as well as simulated) effect on coherence is small for normal vibrato regimes. This allows carrying out signal-coherence analysis for different playing regimes without explicitly detecting or modeling their vibrato. The experimental analyses in the paper are performed using a set of sounds taken from McGill University Music Sound Database [8].

2. Randomly modulated periodicity

A varying periodic signal with a RMP is defined as follows:

Definition: A signal $\{x(t)\}$ is called a RMP with period T if it is of the form

$$x(t) = K^{-1} \sum_{k=-K/2}^{K/2} [\mu_k + u_k(t)] \exp(i2\pi f_k t) \quad \text{for} \quad f_k = \frac{k}{T} \tag{2.1}$$

where $\mu_{-k} = \mu*_k$, $\mu_{-k}(t) = \mu*_k(t)$, and $E[u_k(t)] = 0$ for each k and E is the expectation operation. The $K/2+1\{u_k(t)\}$ are jointly dependent random processes that represent the random modulation. This signal can be written as $x(t) = s(t)+u(t)$ where

$$s(t) = K^{-1} \sum_{k=-K/2}^{K/2} \mu_k \exp(2\pi f_k t),\ u(t) = K^{-1} \sum_{k=-K/2}^{K/2} u_k(t) \exp(i2\pi f_k t) \tag{2.2}$$

The periodic component $s(t)$ is the mean of $x(t)$. The term $u(t)$ is an unknown real valued zero mean non-stationary process. The RMP model is (1) sufficiently rich to describe the different amplitude and frequency modulations and (2) it allows simple characterization of coherence in terms of signal power relations between the mean and the modulation factors.

Suppose that we have M frames of the sampled process $x(t_n)$ of length $T = N\tau$ where $t_n = n\tau$ and τ is the sampling interval. The variation of the waveform from one frame to another is modeled in probabilistic terms by the joint distribution of the samples $\{x(\beta_m), \ldots, x(\beta_m+T-1)\}$, where the start of the mth frame is $\beta_m = (m-1)T$ for $m = 1, \ldots, M$. The discrete Fourier transform (DFT) of this frame, calculated at frequency $f_k = k/T$ for each $k = 1, \ldots, N/2$ is

$$X(k) = \mu_k + U(k), \quad \text{where} \quad U(k) = \sum_{n=0}^{N-1} u(t_n) \exp(-i2\pi f_k t_n) \tag{2.3}$$

To simplify the notation the index m is not used to subscript the variables $X(k)$ and $U(k)$. The variability of the complex Fourier amplitude $X(k)$ about its mean μ_k is $E[U^*(k)U(k)] = \sigma_u{}^2(k)$. If $\mu_k \neq 0$ and $\sigma_u(k) = 0$, then that complex amplitude corresponds to a true periodicity. The larger the value of $\sigma_u(k)$, the greater is the variability of that component from frame to frame. If $\mu_k = 0$ then that component does not contribute to periodicity at all.

In order to quantify the spectral variability we consider the dimensionless function $\gamma_k(k)$, called a SCF defined as follows for each $k = 1, \ldots, N/2$:

$$\gamma_x(k) = \sqrt{\frac{|\mu_k|^2}{|\mu_k|^2 + \sigma_u^2}} \tag{2.4}$$

If $\sigma_u(k) = 0$ and $\mu_k \neq 0$ then $\gamma_k(k) = 1$. This is the case when frequency component f_r has a constant amplitude and phase. If $\mu_k = 0$ then $\gamma_k(k) = 0$. This happens when mean component at that frequency is zero, which is true for any stationary random process with finite energy. A high-coherence value can be either due to large amplitude $|\mu_k|$ or small standard deviation of the frequency components regardless of their absolute values.

In the signal plus modulation-noise representation of $\{x(t)\}$ the signal-to-modulation-noise ratio (SMNR) is $\rho(k) = |\mu_k|^2 \sigma_u{}^{-2}$ for frequency f_k. Thus $\gamma_x{}^2 = \rho(k)/[\rho(k)+1]$ is a monotonically increasing function of SMNR. Inverting this relationship it follows that

$$\rho(k) = \frac{\gamma_x^2}{1 - \gamma_x^2} \tag{2.5}$$

For example, a signal-coherence value of 0.44 yields a SMNR of 0.24 which is −6.2 dB.

The RMP and the SCF are defined for an individual signal where it measures the variability of $X(k)$ about its mean μ_k. This also differs from sinusoidal models of audio signals [9] that try to represent individual signal waveforms rather than perform analysis of deviations from periodicity. The SCF holds for AM, FM, phase modulation, and situations when the carrier is broken when the modulation's amplitude exceeds the carrier's amplitude.

The method is sensitive to the frame length used to estimate the SCF. Coherence is lost if the frame length is different from the fundamental period. In order to find the period, a search over a range of SCF estimates for different pitch values is performed and the strongest value is taken as the SCF estimate. This method effectively removes the sensitivity of the method to errors in pitch estimation. This method will be described in more detail in the experimental section, including additional aspects of the algorithm such as using multiple periods in analysis frame and zero padding before performing DFT.

It should be noted that the RMP method does not distinguish between random and deterministic modulations, both considered as unknown sources of variation. In the paper we assess the influence of vibrato on SCF, showing it separately from other sources of modulation that might occur during sound production.

3. Estimating signal coherence

Recall that $\beta_m = (m-1)T$ for each $m = 1, \ldots, M$. The sample mean for each $t = 0, \ldots, T-1$

$$\bar{x}(t_n) = M^{-1} \sum_{m=1}^{M} x(\beta_m + t_n) \tag{3.1}$$

is an unbiased estimator of $s(t)$. Let $\bar{X}(k)$ denote the kth component of the DFT of $(\bar{x}(0), \ldots, \bar{x}(T-1))$. Let $Y_m(r)$ denote the kth DFT component of $(y(\beta_m), \ldots, y(\beta_m+T-1))$

Author's personal copy

where $y(\beta_m+t_n)=x(\beta_m+t_n)-\bar{x}(t_n)$. The estimate of the variance $\sigma_u{}^2(k)$ is $\hat{\sigma}_u^2=M^{-1}\sum_{m=1}^{M}|Y_m(k)|^2$. The estimate of $\gamma_x(k)$ is the statistic $\hat{\gamma}_x(k)$ defined by

$$\hat{\gamma}_x(k)=\sqrt{\frac{|\hat{X}(k)|^2}{|\hat{X}(k)|^2+\hat{\sigma}_u^2(k)}} \tag{3.2}$$

It is shown in [3] that $\hat{\gamma}_x(k)$is a consistent estimator of $\gamma_x(k)$ for frequency f_k with an error converging as $O(M^{-1/2})$. The expression $\bar{X}|\bar{X}(k)|\hat{\sigma}_u^{-2}(k)$ can be used as an estimator of the signal-to-noise ratio $\rho(k)$ for frequency f_k.

4. Period errors and average pitch estimation

The method of SCF estimation assumes that period of the fundamental is known so that the frame size and its appropriate DFT parameters could be set so as to match it. In practice, the exact frequency of a musical note is seldom perfectly known and often varies over duration of the sustained note. Playing with the so-called "vibrato" is considered as standard ways of producing a musical note on some instruments for some pieces.

In order to address these issues of pitch uncertainty and sensitivity of SCF to errors in period size we search for the best SCF value over a range of possible period values in order to refine the initial period estimate. Then we use multiple periods in one analysis window to increase the spectral resolution. Since the period is specified in integer number of samples, using multiple periods allows determining the period to precision of a fraction of the sampling time step. Moreover since short-term spectral analysis uses the DFT, changes in amplitude occur when frequency components move away from the DFT bin frequency f_k. This change can be large, increasing with the harmonic rank since the kth harmonic has a vibrato depth that is k times greater than vibrato of the fundamental.

Only integer number of samples can be used as a period value. Accordingly, re-sampling helps find the best SCF value by low-pass interpolation of the signal values in the time domain, thus allowing sub-sample precision in terms of the frequency resolution of the SCF function.

We investigate in the next section the effect of reduction in SCF due to pitch errors caused by deterministic or random frequency deviations. Our results indicate that pitch errors of the order of magnitude of few percent are negligible at least for the first several harmonics.

5. Influence of frequency modulation on signal coherence function

The coherence analysis of the previous sections is written out as a modulation component added to a coherent narrowband sinusoid. Mathematically one cannot distinguish between amplitude and phase modulation. In fact the variations of conjugate positive/negative components $u_{-k}(t)=u^*{}_k(t)$ include variations both in amplitude and phase. Since phase modulations, or more precisely their derivative, cause frequency changes, the model incorporates effects related to frequency deviations as well. In this section we explore the effect of random frequency deviations on SCF for a synthetic coherent signal. In natural playing conditions the signal contains quasi-periodic frequency deviation that could be generally termed as vibrato. In such case we can not determine a constant frame size that would match every period of the signal, but we may still do so using longer frames that correspond to multiple periods that deviate around average signal period. This views the effect of vibrato over a sufficiently long frame as a particular case of periodic modulation, as described below.

Consider a pure FM signal

$$x(t)=K^{-1}\sum_{k=-K/2}^{K/2}\mu_k\exp(\mathrm{i}2\pi f_k(1+Iv_k(t))t) \tag{5.1}$$

where $v_k(t)$ is a slowly varying unknown deterministic or random process[1] limited to the range $[-1,1]$ and I is a modulation index. Note that in our model the frequency deviation is proportional to frequency of every harmonic. We use the subscript k in the definition of $v_k(t)$ to allow for separate deviations in every harmonic, although in most practical situations it could be the same or highly correlated process.

Using the index r rather than k the DFT of component r is

$$X(r)=\frac{N}{K}\sum_{k=-K/2}^{K/2}\mu_k D_{k,I}(r) \tag{5.2}$$

where

$$D_{k,I}(r)=N^{-1}\sum_{n=0}^{N-1}\exp(\mathrm{i}2\pi t(k-r+kIv_k(t_n))/N) \tag{5.3}$$

is similar to Dirichlet kernel (with an additional multiplicative complex exponential) that reduces to Kronecker delta function for very small modulation indices.

Assume first that $v_k(t)$ is a bandlimited random deviation that can be approximated by piecewise constant function. In other words, we assume that it remains constant during the frame $\{x(\beta_m),\ldots,x(\beta_m+T-1)\}$. Defining $\theta_{k,I}(r)=k(1+Iv_k)-r$, we rewrite it as

$$\begin{aligned}D_{k,I}(r)&=N^{-1}\sum_{n=0}^{N-1}\exp(\mathrm{i}2\pi t_n\theta_{k,I}(r)/N)\\&=N^{-1}\exp\left(\mathrm{i}\pi\theta_{k,I}(r)\left(1-\frac{1}{N}\right)\right)\frac{\sin\pi\theta_{k,I}(r)}{\sin(\pi\theta_{k,I}(r)/N)}\end{aligned} \tag{5.4}$$

In order to evaluate the mean and variance values of the r-th DFT component we assume that the deviation is small relative to the spacing between the harmonics, i.e. we shall limit ourselves to deviations that are less than the width of the main lobe of $D_{k,I}(r)$ so that

$$E(X(r))\approx E(D_{r,I}(r))\mu_r \tag{5.5}$$

[1] It should be noted that $v_k(t)$ does not have dimensions of frequency. If we want to relate it to instantaneous phase deviations $\phi_k(\tau)$, then $v_k(t)$ is the "mean" phase accumulation $v_k(t)=t^{-1}\int_0^t\phi_k(\tau)\,\mathrm{d}\tau$.

Author's personal copy

$$E(X(r)X^*(r)) = \sum_k \sum_{k'} \mu_k \mu_{k'} E(D_{k,I}(r) D^*_{k',I}(r)) \approx \frac{1}{N^2} \mu_r^2 E\left(\frac{\sin^2(\pi\theta_{r,I}(r))}{\sin^2(\pi\theta_{r,I}(r)/N)}\right) \quad (5.6)$$

Using the estimate of $\hat{\gamma}_x(r)$ given in (3.2) we obtain an approximation to SCF for different values of modulation parameters. One may note that for $r = k$ we have $\theta_{r,I}(r) = rIv_r(t)$ and $D_{r,I}(r) = D_{1,rI}(1)$.

The above derivation assumed that the modulation $v_k(t)$ was random. A similar derivation also applies for unknown deterministic function $v_k(t)$. In such case the averaging of $D_{r,I}(r)$ in Eqs. (5.5) and (5.6) is replaced by integrating $D_{r,I}(r)$ as function of $\theta_{r,I}(r)$ over a range of r and I. Fig. 1 presents a graph of the theoretical evaluation of $\hat{\gamma}_x(r)$ as a function of $\eta = rI$ (that we call "vibrato depth"), obtained by numerical integration of $D_{r,I}(r)$ over different ranges of $\theta_{r,I}(r)$, i.e. as function of rI. This graph is compared to experimental signal-coherence values obtained from estimation of SCF calculated for 2000 instances of randomly modulated sinusoidal signal, with frequency modulation randomly chosen according to uniform distribution over the range $[-rI\ rI]$. Similar results were obtained by using deterministic modulation, such as sinusoidal frequency modulation with vibrato depth that matches, in root mean square sense, to the variance of the random modulation.

This graph can be used to evaluate the relative reduction in signal coherence as a function of increasing modulation index value I for a constant harmonic index r. It should be noted that Eq. (5.6) could be considered either as a function of harmonic index r for constant modulation index I or as function of I for a given r. Accordingly, this graph can be used to evaluate the reduction in signal coherence as function vibrato depth or as a function of harmonic index assuming a constant vibrato (i.e. modulation index). For modulation index of 1%, the 20th harmonic of a perfectly coherent signal can be looked up from the graph at vibrato depth 0.2 and it will lose coherence in the amount of less than 3% (coherence of the 20th harmonic is approximately 0.97). Harmonic number 50 that undergoes 50% modulation (vibrato depth 0.5) drops in coherence to 0.87, and so on.

6. Analysis of music instrument signals

In order to evaluate SCF on real signals, we applied coherence measure to instrumental sounds from the McGill University Master Samples that have a well-defined pitch during a sustained portion of their sounds. As mentioned above, these sounds were produced by natural playing, i.e. including vibrato and amplitude modulations, as well as in presence of recording room conditions (these are not anechoic room recordings). Assuming that recording conditions were similar for the different samples, we shall consider the relative changes in SCF among pairs of sounds. In first experiment we shall consider two Cello recordings playing a single note each, one with and the other without vibrato. In the second experiment we shall compare between SCF for sets of samples drawn from flute and French horn instruments. We shall compare the two sets in terms of the mean and the variance of SCF for the different harmonics.

A particularly interesting case for SCF analysis is that of the Cello. This instrument produces sounds through a periodic bow excitation that passes through a complicated resonance body. The analysis in Fig. 2 shows the waveforms and coherence analysis of an open-string Cello sound, which is indeed very coherent.

In view of such a coherent open-string sound, it is interesting to find that there is such a big difference in coherence between open and stopped sounds, as presented in Fig. 2. In term of the theoretical results of Section 3, the influence of frequency modulation on signal coherence should have been a reduction of SCF only by few percent. The Cello signal shown in Fig. 3 is played with vibrato depth of approximately 5.6%, which is less than a semitone deviation in pitch. Although in theory there should be very little change in SCF, in practice the signal exhibits a very significant loss of coherence.

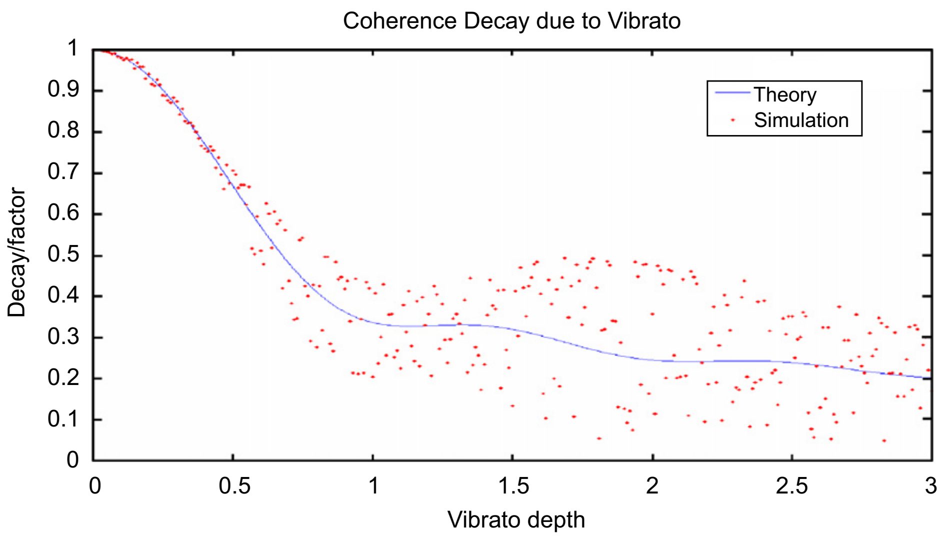

Fig. 1. Signal coherence as a function of vibrato depth.

Author's personal copy

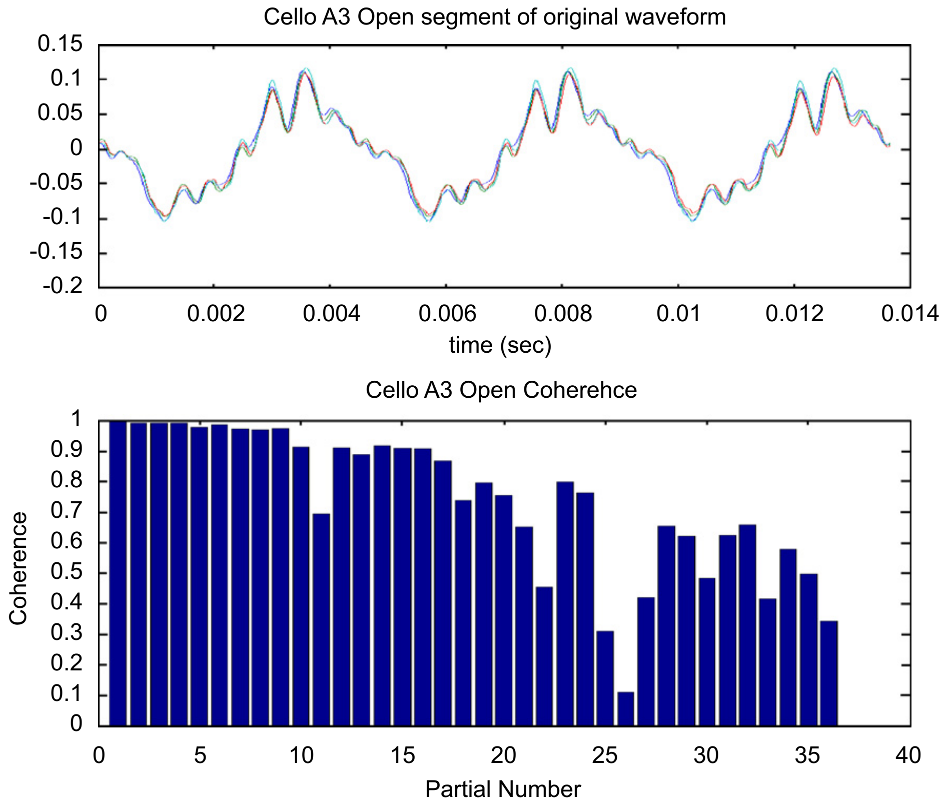

Fig. 2. Cello A3 open string (with no vibrato). Signal waveforms from several frames (top) and resulting coherence (bottom).

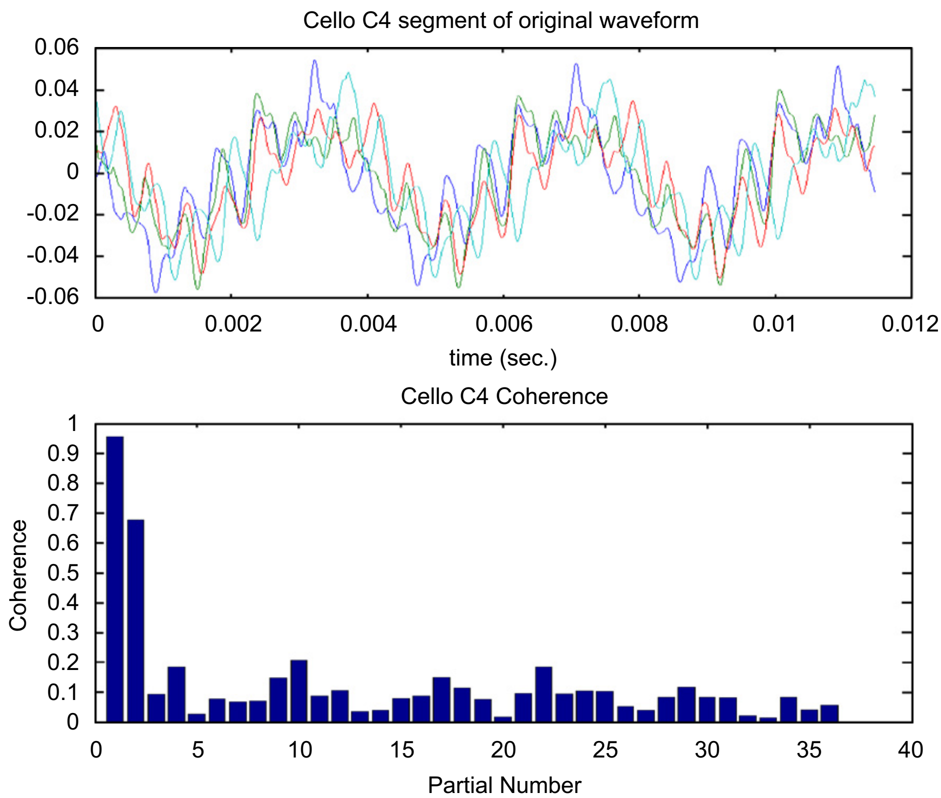

Fig. 3. Cello C4 (with natural vibrato). Signal waveforms from several analysis frames (top) and resulting coherence (bottom).

Author's personal copy

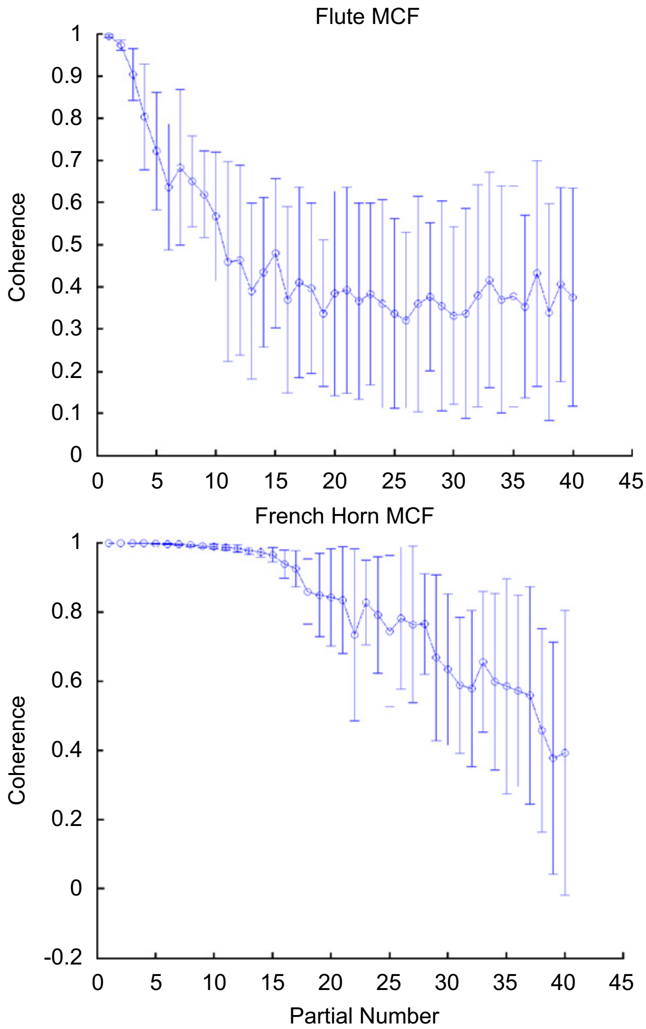

Fig. 4. SCF mean and standard deviation of a set of sounds of flute and French horn instruments.

This finding suggests that vibrato causes additional random modulations in the signal, reducing SCF beyond the level expected from deviations caused by frequency modulation alone. This finding is interesting in view of the rigid resonance body of the Cello, which does not allow strongly dynamic or interactive source-filter models.

We also performed SCF analysis for two sets of 20 sound clips each extracted from recordings of sustained notes played on the flute and the French horn. Each set was prepared by using recordings of 5 adjacent notes (with semitone differences in pitch between adjacent notes), with each note segmented into 4 sound clips of approx. 226 ms. In order to avoid errors in estimation of SCF due to imprecision of pitch estimators, we have performed a search over a range of possible periods around the theoretical frequency value of every note. To assure high-resolution search we used larger frames that contained 3 periods of the waveform. The signal was further over-sampled, allowing even greater sub-sample precision for period determination. For each sound clip an SCF with maximal coherence over the fist 20 harmonics was chosen as the final estimate, giving a total of 20 different SCF values for each instrument. Below (Fig. 4) we present the modulated coherence graphs for flute and French horn sounds, showing the mean and standard deviation errors around the mean for the two sound sets.

These results indicate that there is a significant difference between SCF of the two instruments over a range of pitches. Comparing SCF of the flute to theoretic SCF due to vibrato shows that an equivalent vibrato depth of 10% is required to cause such reduction in SCF. Considering the flute signal we find out that it was indeed played with vibrato, but the pitch deviation did not exceed 1.6 Hz, which is only 0.6% (0.006) of the mean pitch. This indicates that SCF of the flute is caused by additional mechanisms than vibrato. Pitch analysis of the French

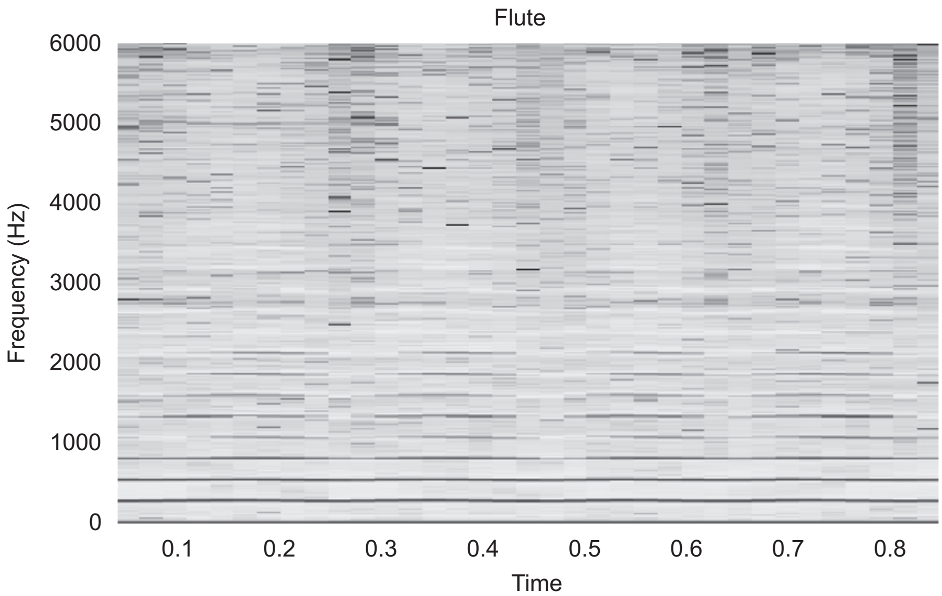

Fig. 5. Time–frequency plot of the flute sound.

Author's personal copy

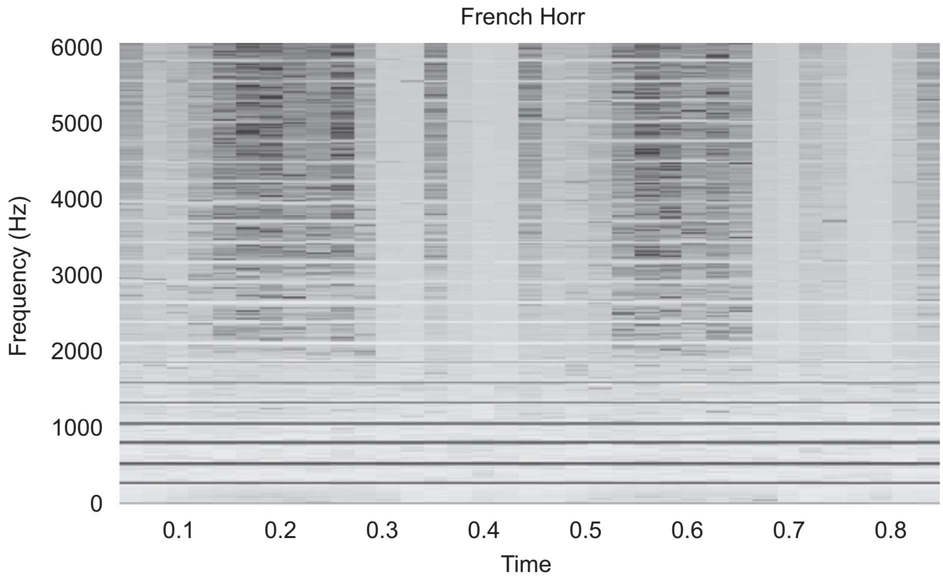

Fig. 6. Time–frequency plot of the French horn sound.

horn shows practically no vibrato, a situation that is confirmed by the high SCF values for the lower harmonics. It should be also noted that a graph with standard deviations for the Cello sounds could not be provided since the sound database contains a single recording of an open-string sound (A3).

Figs. 5 and 6 provide the time–frequency plots of the flute and French horn recordings, respectively. The plot shows the harmonics over time, with slight broadening visible in the lower harmonics of the flute due to modulation. It is evident that visual inspection of the time-frequency plot does not reveal the MCF properties of the signals. The plots were produced using a sliding Hamming window of 4096 samples with 75% overlap between successive windows. Sampling frequency is 44,100 Hz.

8. Conclusion

We applied a measure of the amount of variations that occur in Fourier components of acoustic signals due to unknown modulations that we call SCF. This measure is based on a model of RMP that considers the residual deviations of a signal after subtraction of a mean signal. The mean signal is obtained by averaging of the signal over multiple frames containing samples of individual periods. SCF is evaluated in terms of power relations between the mean and the difference signal.

Since this method is sensitive to errors in the period estimation we analyzed the effect of frequency deviations on SCF. We showed that reduction of SCF due to changes in pitch only (including the case of unknown vibrato) is relatively small for low harmonics and for normal playing conditions. We use these results to discuss the differences in SCF for real instrument sounds, suggesting that reduction in coherence due to pitch deviations alone is minor in comparison to other random modulations that appear during signal production.

The MATLAB programs used to generate the examples are available at http://music.ucsd.edu/~sdubnov/SigCoh. The program Spectrum and its executable for computing signal coherence is available on http://web.austin.utexas/hinich

References

[1] M.J. Hinich, A statistical theory of signal coherence, J. Oceanic Eng. 25 (2) (2000) 256–261.
[2] M.J. Hinich, P. Wild, Testing time series stationarity against an alternative whose mean is periodic, Macroecon. Dynam. 5 (2001) 380–412.
[3] M.J. Hinich, Detecting randomly modulated pulses in noise, Signal Processing 83 (6) (2003) 1349–1352.
[4] M.J. Hinich, P. Wild, Detecting finite bandwidth periodic signals, Signal Processing 85 (2005) 1557–1562.
[5] W. Beauchamp, Time-variant spectra of violin tones, J. Acoust. Soc. Am. 56 (3) (1974) 995–1004.
[6] M.E. McIntyre, R.T. Schumacher, J. Woodhouse, Aperiodicity in bowed string motion, Acustica 49 (1981) 13–32.
[7] S. Dubnov, X. Rodet, Investigation of phase coupling phenomena in sustained portion of musical instruments sound, J. Acoust. Soc. Am. 113 (1) (2003) 348–359.
[8] McGill University Master Samples, Faculty of Music, McGill University, 555 Sherbrooke St., West, Montreal, Quebec.
[9] X. Serra, J.O. Smith, Spectral modeling synthesis: a sound analysis/synthesis system based on deterministic plus stochastic decomposition, Comput. Music J. 14 (4) (1989) 12–24.