Calculation of a "narrowed" autocorrelation function

Judith C. Brown[a] and Miller S. Puckette[b]
Music and Cognition, Media Laboratory, Massachusetts Institute of Technology, 20 Ames Street, Cambridge, Massachusetts 02139

(Received 3 October 1987; accepted for publication 22 December 1988)

A method of calculating an autocorrelation function with extremely narrow peaks is described. This is done by including terms in the autocorrelation expression corresponding to delays at 2τ, 3τ, etc., in addition to the usual term with delay τ. Implications in the frequency domain are explored. Graphs of this autocorrelation function for a number of violin sound samples, including a two-octave scale, vibrato, and glissando, are presented. Graphs of the autocorrelation function for some synthetic sound samples are also included.

PACS numbers: 43.60.Gk, 43.75.De

INTRODUCTION

In assessing various fundamental frequency detection methods previously used in speech analysis to determine their applicability to the problem of extracting frequency from music, we have examined the method of autocorrelation. The use of a zero phase method, such as autocorrelation, is particularly promising for the study of musical signals since this means that contributions from all of the harmonics occur at the period of the fundamental, and any problem of a "missing" (or weak) fundamental is thus circumvented.

Although in one previous study[1] an extremely narrow peak in the autocorrelation function was obtained, this was accomplished after a complicated method of spectral flattening of the waveform achieved in hardware. Another study[2] employed nonlinear processing of the input signal to achieve spectral flattening. In general, however, autocorrelation studies have relied on the identification of a rather broad peak to determine the pitch period.[3]

We shall describe herein an elegant software method of obtaining a narrow peak in the autocorrelation function and shall demonstrate this narrowing quantitatively for a pure tone. Application to musical examples will be included, but the major focus of this article is on the calculation. In a later article, we shall give a detailed report on pitch tracking experiments that rely on the identification of this narrow peak, and these will be compared to several other methods used for musical pitch tracking.

Autocorrelation methods of analysis are most appropriate in the study of signals whose durations are at least several times the longest period studied. This requirement is met for most musical signals. The narrowing is obtained at the cost of an increase in the length of the sample analyzed and this results in less precision in time.

A narrow peak in the autocorrelation function is of interest for the study of musical signals as the identification of the peak can be thrown off by the presence of periodic noise with even a low amplitude. In Fig. 1 are graphs of the ordinary autocorrelation and two calculations with the autocorrelation function successively narrowed. The signal consists of two pure tones of frequencies 330 Hz (E4) and 392 Hz (G4), with the higher tone having one fourth the amplitude of the lower tone. This smaller component represents periodic noise in this simulation. For the ordinary correlation there appears to be only one tone present, and the peak position is shifted down (meaning a smaller period) relative to the "correct" position of the principal component. For the next graph the principal peak is in the correct position, but it is not clear that there is more than one periodic component. Finally, for the lowest graph, which has the narrowest peaks, the second component is clearly present, and both peaks are in their theoretical positions.

Thus it is clear that a narrowed autocorrelation function is useful both for determining peak position and for identification of components present. Periodic noise, such as that in this simulation, can arise, for example, with the violin, where one string might ring when another is bowed. Nonperiodic noise present in musical signals, such as the sound of breath in the flute or striking noise from the keys of the piano, is not a problem for autocorrelation.

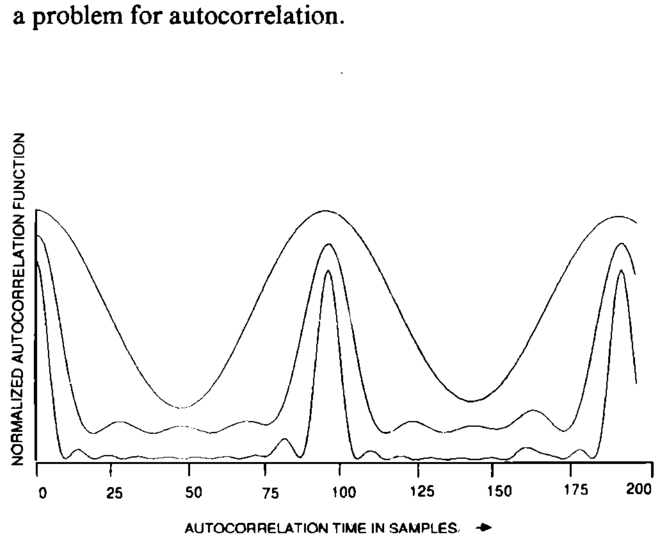

FIG. 1. Ordinary autocorrelation function (top) and two successively narrowed autocorrelation functions [calculation with Eq. (1b) using two terms, five terms, and ten terms] for two sounds of periods 97 and 81 samples with amplitudes in the ratio of 4 to 1.

[a] On sabbatical leave from Wellesley College, Wellesley, MA 02181.
[b] Present address IRCAM, 31 rue Saint-Merri, F-75004, Paris, France.

I. BACKGROUND

The autocorrelation function is defined as the product

$$g(\tau) = \int_{-\infty}^{\infty} f(t)\cdot f(t-\tau)dt.$$

It is convenient to calculate this as $\langle |f(t) + f(t-\tau)|^2\rangle$, where the cross term gives the autocorrelation.

If, instead, we consider the sum

$$S_N(\tau) = f(t) + f(t-\tau) + f(t-2\tau)\cdots f(t-(N-1)\tau), \quad (1a)$$

then its square

$$|S_N(\tau)|^2 = \langle |f(t) + f(t-\tau) + f(t-2\tau)\cdots f(t-(N-1)\tau)|^2\rangle \quad (1b)$$

will involve cross products of each pair of terms. The quantity defined in Eq. (1b) will be referred to as the narrowed autocorrelation. The case $N = 2$ gives the usual autocorrelation.

For a periodic function with period T, all of the terms in Eq. (1a) will be in phase for delays τ equal to an integral multiple of T. This is the same behavior as in the usual case with $N = 2$, but now for a small change in τ to $(T + \Delta)$ the nth term is $(n-1)\Delta$ out of phase with the first term. Thus these functions get out of phase very rapidly, leading to greatly narrowed peak. Here, $(n-1)\Delta < T$.

Quantitatively, if we consider a single harmonic component of unit amplitude $\exp(j\omega_0 t)$ in Eq. (1a),

$$S_N(\tau) = \sum_{n=0}^{N-1} \exp[j\omega_0(t - n\tau)],$$

$$= [1 - \exp(-j\omega_0 N\tau)]/[1 - \exp(-j\omega_0\tau)],$$

and its square

$$|S_N(\tau)|^2 = \sin^2(\omega_0 N\tau/2)/\sin^2(\omega_0\tau/2),$$

which has maxima of height N^2 at $\tau = 0$, nT where n is an integer and the period $T = 2\pi/\omega_0$. Most important, the peaks have a half-width of T/N measured from the peak to the first zero of the function. This means that the peaks are $2/N$ times narrower than for the usual autocorrelation. Examples are found in Fig. 2. Here, as predicted, the half-widths in samples of the functions are 4, 8, and 20 for the cases of $N = 10$, 5, and 2, respectively. Further details of the calculations for the figures are given in Sec. V.

A complex sound will demonstrate the same behavior as that described above for each of its Fourier components. The situation is, of course, complicated by the coincidence of peaks. For example, with the analysis of a sound consisting of a fundamental of period T and harmonics with periods $T/2$, $T/3$,..., etc., peaks of all harmonics coincide at multiples of the period T; peaks from all of the even harmonics will coincide at multiples of $T/2$ and so on.

In Fig. 3 we have included the analysis of a sound consisting of ten harmonics of equal amplitude calculated with five terms in Eq. (1b). Although the narrowed autocorrelation is quite complex, the position of the period of the fundamental is totally unambiguous. This figure is of interest, also, because it shows the loss of resolution due to discrete sampling, which is particularly apparent at $2T$. This problem becomes more troublesome the narrower the peak; it could, of course, be corrected with a higher sampling rate, but at a high cost in computation time.

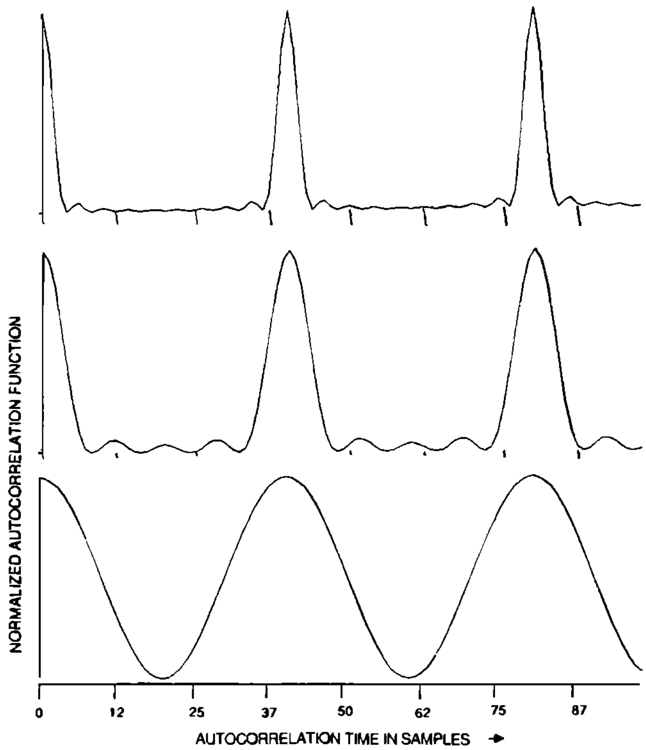

FIG. 2. Narrowed autocorrelation function for a sinusoid with period 40 samples calculated with Eq. (1b) using two terms, five terms, and ten terms.

The width of the peaks at T, $2T$, etc., is much narrower for a complex sound than for a pure tone, as it consists of the superposition of peaks due to the presence of each harmonic component. The half-width of each component is proportional to the inverse product of N times the number of the component. Thus the width of the peak in the general case depends on the exact amplitudes of the harmonics present and the greater the amplitude of higher harmonics, the narrower the peak.

It is interesting to note that the mathematical treatment of a single harmonic component described above is identical

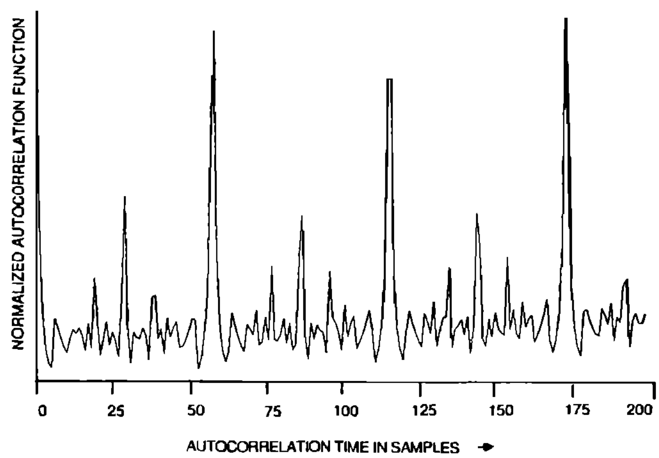

FIG. 3. Narrowed autocorrelation function for a signal with ten harmonics of equal amplitude. Five terms were included in Eq. (1b).

to that used to describe the diffraction of light by N slits where the beam from each slit is delayed (or advanced) by τ relative to that from adjacent slits.[4]

II. FOURIER TRANSFORM

Since it is well known that the Fourier transform of the autocorrelation function is the power spectrum,[5,6] it is of interest to calculate the Fourier transform of the narrowed autocorrelation. Our measured function is

$$\langle |S_N(\tau)|^2\rangle = \int_{-\infty}^{\infty} |f(t) + f(t-\tau) + f(t-2\tau)\cdots f(t-(N-1)\tau)|^2\, dt. \tag{2}$$

Carrying out the square, we find N terms that are time-shifted functions of the form $\int [f(t)]^2\, dt$. There are $2(N-1)$ terms that are time-shifted functions of the form $\int f(t)f(t-\tau)\,dt$, $2(N-2)$ terms that are time-shifted functions of the form $\int f(t)f(t-2\tau)\,dt$, and so on to two terms of the form $\int f(t)f(t-(N-1)\tau)\,dt$.

Collecting these terms and substituting the definition of the autocorrelation function $\int_{-\infty}^{\infty} f(t)f(t-\tau)\,dt = G(\tau)$ gives

$$\begin{aligned}\langle |S_N(\tau)|^2\rangle &= N\cdot G(0)\\ &\quad + 2(N-1)\cdot G(\tau) + 2(N-2)\cdot G(2\tau)\\ &\quad + \cdots + 2\cdot G((N-1)\tau).\end{aligned} \tag{3}$$

Taking the Fourier transform of Eq. (3), and noting that the Fourier transform of the autocorrelation function is the power spectrum,

$$F(\omega) = \int_{-\infty}^{\infty} G(\tau)\exp(-i\omega\tau)\, d\tau.$$

Since we are studying time-varying signals and are, in essence, interested in the time variation of the spectrum, this must be modified to a short-time Fourier transform.[7] We thus obtain

$$F(\omega) = \frac{1}{2M}\int_{-M}^{M} G(\tau)\exp(-i\omega\tau)\, d\tau.$$

For the terms involving $G(k\tau)$, this must be modified to give

$$\frac{1}{2M}\int_{-M}^{M} G(k\tau)\exp(-i\omega\tau)\, d\tau = \frac{1}{2P}\int_{-P}^{P} G(u)\exp\frac{-i\omega u}{k}\, du = F\left(\frac{\omega}{k}\right),$$

where $u = k\tau$ and $P = kM$.

Then, we find for the transform of our measured function $|S_N(\tau)|^2$,

$$\begin{aligned}F_S(\omega) &= N\cdot G(0)\delta(\omega) + 2(N-1)\cdot F(\omega)\\ &\quad + 2(N-2)\cdot F(\omega/2) + 2(N-3)\\ &\quad \cdot F(\omega/3)\cdots + 2\cdot F(\omega/(N-1)).\end{aligned} \tag{4}$$

Here, $F(\omega) = |\hat{f}(\omega)|^2$, where $\hat{f}(\omega)$ is the Fourier transform of $f(t)$.

The first term in Eq. (4) is the average of the square of the function $f(t)$. The second term is the power spectrum of $f(t)$ and is the only term that would be present with the conventional autocorrelation. Terms three–N arise due to the "narrowing" in the time domain and are expanded copies of the power spectrum at twice the frequency, 3 times the frequency,..., N times the frequency.

Examples of the Fourier transform of $S_N(\tau)^2$, as defined in Eq. (1), are found in Fig. 4 for $N = 2$ and $N = 5$ for an input sine wave. For this case, the autocorrelation function $G(\tau) = \cos(\omega_0\tau)$. The Fourier transform is

$$F(\omega) = [\delta(\omega + \omega_0) + \delta(\omega - \omega_0)]/2,$$

where $\delta(x)$ is the Dirac delta function. This can be substituted in Eq. (4), to give for positive frequencies

$$\begin{aligned}F_S(\omega) &= N + (N-1)\cdot\delta(\omega - \omega_0)\\ &\quad + (N-2)\cdot\delta(\omega/2 - \omega_0)\\ &\quad + \cdots\delta(\omega/(N-1) - \omega_0),\end{aligned}$$

in agreement with Fig. 4.

III. RESOLUTION

As discussed in Sec. I, the half-width of a peak in the narrowed autocorrelation for a sinusoidal component is T/N, where N is the number of terms in Eq. (1b). If we define the resolution as the correlation time divided by the half-width in analogy to the usual definition $f/\delta f$ for a peak at frequency f of width δf in the spectral domain, we obtain

$$\text{Resolution} = T/(T/N) = N.$$

Thus it would first appear that the resolution can be made arbitrarily large. For example, musical signals should be resolved to within a semitone. These two frequencies in the ratio of the 12th root of 2 ($= 1.059$) must thus have correlation times separated by $0.06T$. If we set $T/N = 0.06T$, we find that, for $N \simeq 17$, according to this theory, a semitone will be resolved.

The problem with this treatment is that it fails to take account of the discrete nature of the signal. The ultimate constraint in any discrete system is the requirement that two

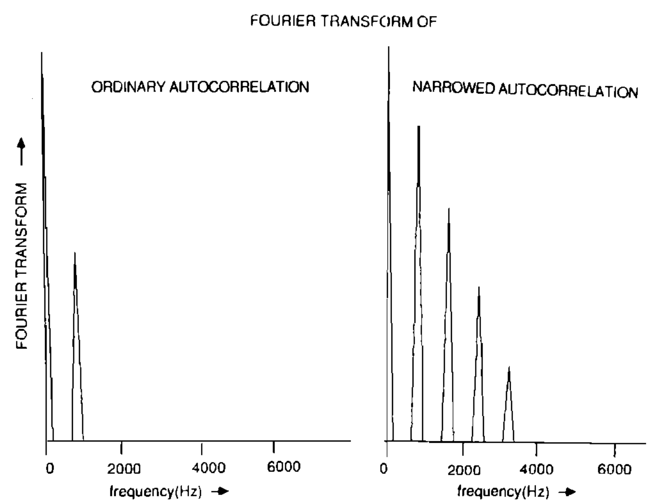

FIG. 4. Fourier transform of the narrowed autocorrelation function of a sine wave calculated with two terms (left) and with five terms (right) in Eq. (1b).

signal peaks be separated by at least two samples so that there can be a dip between the maxima. Again, examining the requirement that two tones a semitone apart be resolved, we set $0.06T \geqslant 2$ samples, which gives $T \geqslant 33$ samples. At a sample rate of 32 K, which is a relatively high sampling rate, this would mean that two sinusoids separated by a semitone at a frequency over 970 Hz (roughly B5) would not be resolved. In Fig. 5 is found the narrowed autocorrelation calculated with 17 terms in Eq. (1b) for two notes separated by a semitone. Each pair of notes has a duration of 0.1 s, indicated on the vertical axis, and the notes (identified with the convention that middle C is C4) have the periods given in Table I for sampling at 32 000 Hz.

Note that the predictions of the preceding discussion are roughly followed. Note, also, that the peak separation increases linearly with time, so that all peaks are clearly resolved for correlation times of $2T$, $3T$, etc. Finally, note that the time resolution for the changes of note pairs is poor due to the long integration times (about 16 ms) and the long time spanned by taking 17 terms. The latter means that we are looking at a time of 17τ for each point on the correlation function, and this is over 0.1 s for the highest correlation times.

Fortunately, it is usually not necessary to carry out the analysis with 17 terms since the correlation function of most musical signals is narrowed by the presence of higher harmonics by the mechanism discussed previously. In Fig. 6 we have calculated the narrowed autocorrelation, including only five terms in Eq. (1b) for the same pairs of notes as in the previous figure, but this time we have included ten harmonics of equal amplitude for each note. The resolution is, in fact, better here than in the previous figure calculated with 17 terms. The third and fourth pairs of notes at $t = 0.2$ s and $t = 0.3$ s show clearly the effects of discrete sampling. Following our previous discussion, the third pair should be resolved and the fourth pair should not be, whereas, in fact, the fourth pair is better resolved than the third.

TABLE I. Periods in samples at 32 000 Hz for the pairs of notes, as found in Fig. 5, with a duration of 0.1 s. The notes are identified with the convention that middle C is C4.

Time		Period		Period
0.0	C#5	57.7	D5	54.5
0.1	F5	45.8	F#5	43.2
0.2	G#5	38.5	A5	36.4
0.3	C#6	28.9	D6	27.2
0.4	F6	22.9	F#6	21.6
0.5	G#6	19.3	A6	18.2

IV. CALCULATIONS

All calculations were programmed in C and carried out on a Hewlett–Packard model 9000, series 300 "Bobcat" computer. Except where stated otherwise, five terms were included in Eq. (1b) since the extremely sharp peaks obtained with, e.g., ten terms, give rise to resolution problems due to discrete sampling, as well as requiring longer run times. The equation was integrated over 500 samples (about 15 ms) to give a 200-point autocorrelation. The latter was chosen so as to include G3 for the violin on the low-frequency end. The graphics software was written by Barry Vercoe at the MIT Experimental Music Studio for the Hewlett–Packard computer.

Sound examples were either generated using Barry Vercoe's C-sound software (Figs. 1–6, 13) or were recorded and digitized by Greg Tucker at the Experimental Music Studio with Chung Pei Ma playing the violin (Figs. 7–12). The sample rate was 32 kHz for all samples.

Most of the figures are plots of Eq. (1b) on the vertical axis against correlation time τ in samples on the horizontal axis. Although the vertical axis is not labeled with the value of the autocorrelation function, all examples were normalized to a value of 1 for zero delay. On the horizontal axis, one

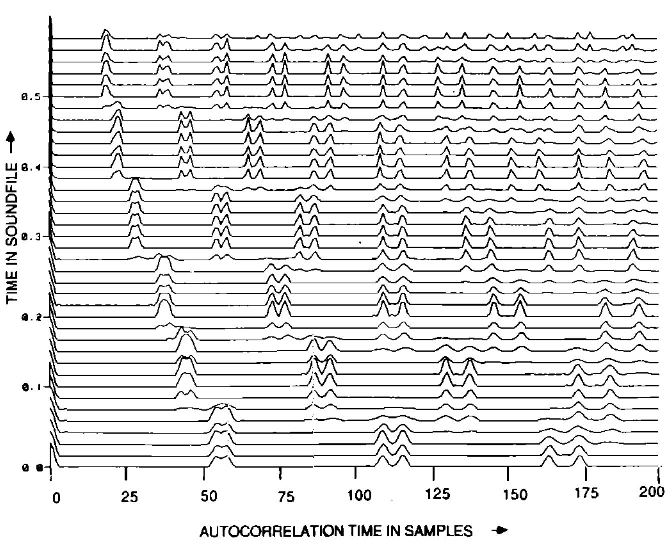

FIG. 5. Narrowed autocorrelation function of two ascending sinusoids a semitone apart calculated with 17 terms in Eq. (1b).

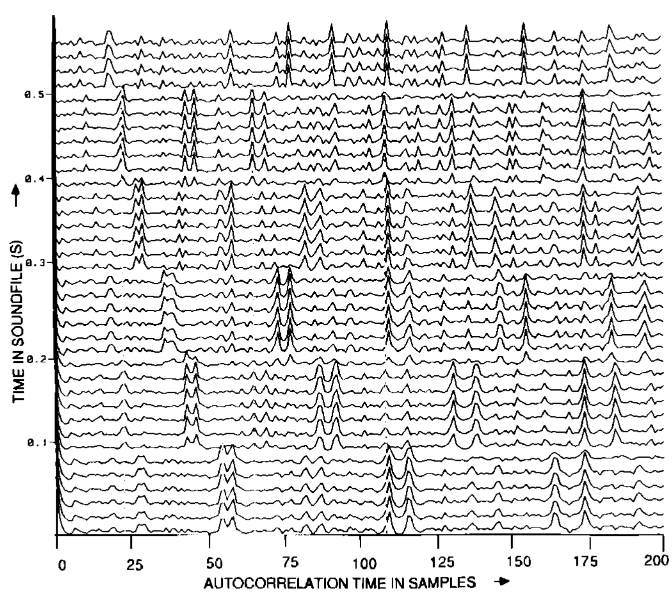

FIG. 6. Narrowed autocorrelation function of two ascending complex sounds a semitone apart calculated with five terms in Eq. (1b).

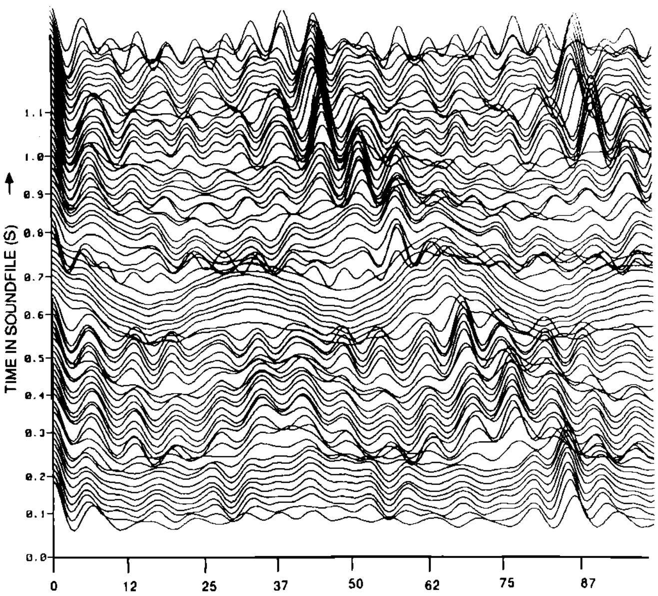

FIG. 7. Conventional autocorrelation function of a violin scale from G3 to G4.

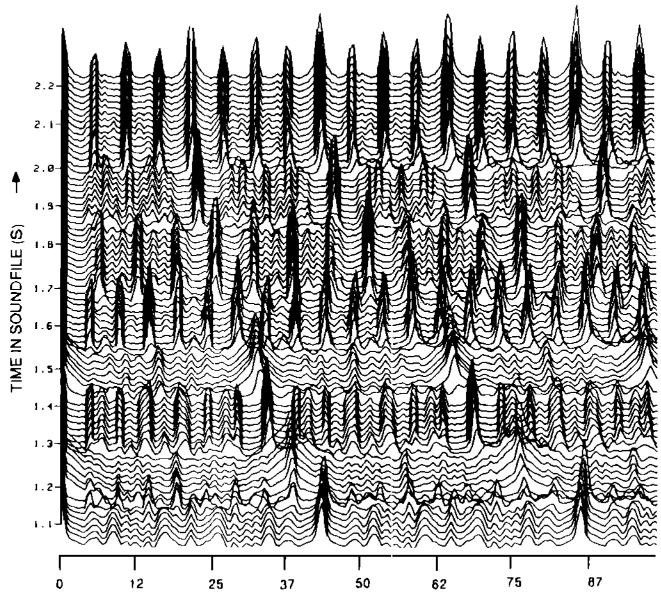

FIG. 9. Narrowed autocorrelation function of a violin scale from G4 to G5 calculated with five terms in Eq. (1b).

sample is equal to 31.25 μs at our sample rate. For the figures with multiple graphs (or frames), the labels on the vertical axis correspond to real time in the sound file. The frames were calculated sequentially so as not to overlap with the average, usually taken over 500 samples as mentioned.

V. MUSICAL EXAMPLES

A striking set of examples of the power of the technique of narrowed autocorrelation can be found in Figs. 7–13. Figure 7 is a conventional autocorrelation function [calculated with two terms in Eq. (1b)] of one octave of a violin scale from G3 to G4 included for comparison. Figures 8 and 9 show the narrowed autocorrelation function for the same scale. Figure 8 corresponds to the same octave as that of Fig. 7, and Fig. 9 corresponds to the next octave. The period of the fundamental is extremely clear in all cases. It is also easy to distinguish which notes are spectrally rich and which are relatively pure (for example, the fourth note has only two harmonics) by the complexity of the curves. Preliminary experiments in pitch tracking of this signal by picking the maxima in the numbers graphed in these two figures have been quite successful.

In Fig. 10 is the narrowed autocorrelation function of

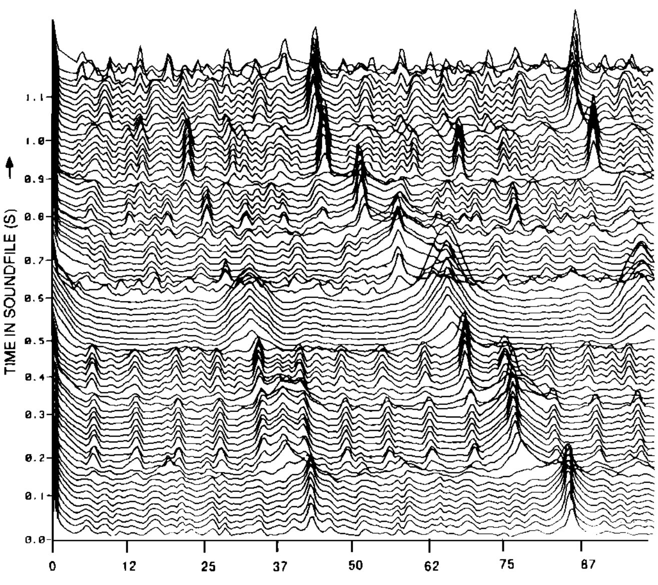

FIG. 8. Narrowed autocorrelation function of a violin scale from G3 to G4 calculated with five terms in Eq. (1b).

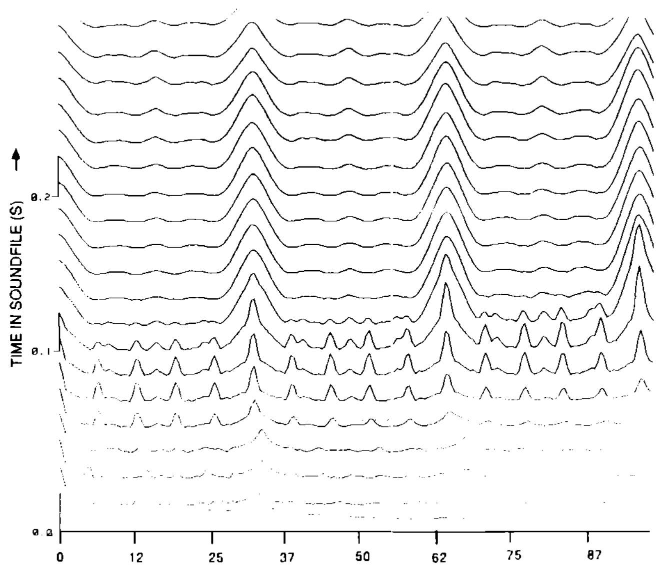

FIG. 10. Narrowed autocorrelation function of violin note bowed very quickly and its decay.

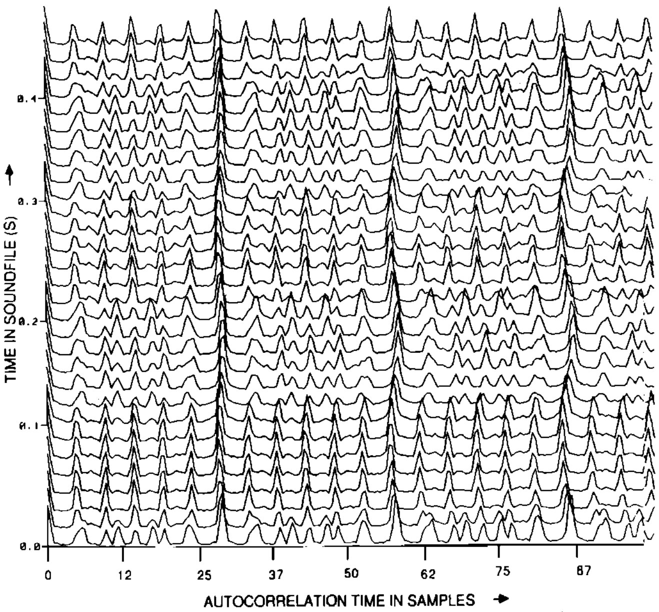

FIG. 11. Narrowed autocorrelation function of a violin note played with vibrato.

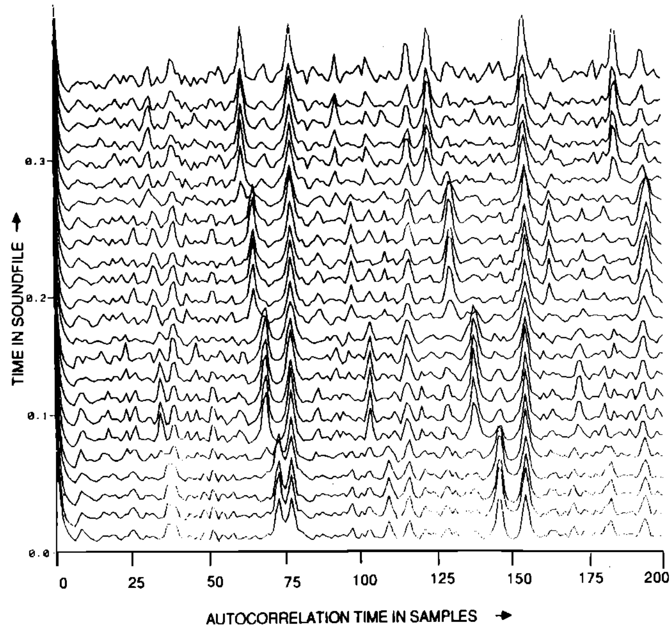

FIG. 13. Narrowed autocorrelation function of two complex sounds with frequency separation increasing from a semitone to two whole steps.

the note C5 bowed very quickly and then released to decay freely. Once the bow is released, the behavior is that of a plucked string. It is clear that the spectrum of the bowed string is more complex than that of the plucked string. Here, we have a particularly strong fifth harmonic (strong peak at $T/5$) for the bowed string and practically a pure tone for the decay. No decay in the amplitude of the signal is apparent because each frame is normalized to 1 for zero delay time τ.

Figure 11 shows the effect of vibrato on the narrowed autocorrelation for the violin note D5. It is interesting how much the spectrum changes as the frequency is lowered. The effect is that of having two distinct notes rather than the continuous coherent change up and down of each of the component harmonics that we might have anticipated for this small frequency change. This indicates that the body resonances of this violin are extremely narrow. Notice, also, that the vibrato is much more apparent at correlation times of two and three times the fundamental than at that of the fundamental itself.

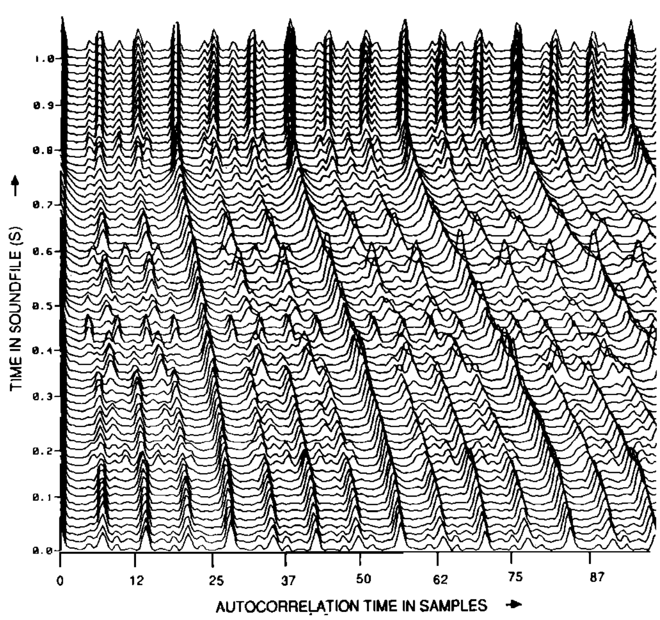

FIG. 12. Narrowed autocorrelation function of a violin glissando.

Figure 12 is the narrowed autocorrelation of a violin glissando. As in the case of vibrato, we note that there are discontinuous changes in the spectrum of the sound as the frequency of the fundamental rises continuously. Here, this might have been anticipated, however, as the frequency is increasing by a fifth (factor of 1.5).

Finally, we have included Fig. 13, which is a narrowed autocorrelation of two complex sounds to demonstrate the promise of this technique for polyphonic pitch tracking. The lower note remains on G#4, while the upper notes go from A4 to B4 with the separation increasing from a semitone to two whole steps. The maxima are clearly resolved and the pitches could easily be determined with a program to pick out maxima.

ACKNOWLEDGMENTS

JCB is grateful to Wellesley College for its generous Sabbatical leave support and to the Media lab at MIT for the use of its superb facilities. She would also like to thank the National Institutes of Health for a Biomedical Research Support Grant and the Marilyn Brachman Hoffman committee of Wellesley College for a summer research stipend for student assistants.

[1]M. M. Sondhi, "New Methods of Pitch Extraction," IEEE Trans. Audio Electroacoust. **AU-16**, 262–266 (1968).

[2]R. R. Rabiner, "On the Use of Autocorrelation Analysis for Pitch Detection," IEEE Trans. Acoust. Speech Signal Process. **ASSP-25**, 24–33 (1977).

[3]S. Seneff, "Pitch and Spectral Analysis of Speech Based on an Auditory Synchrony Model," Ph.D. thesis, MIT (1985).
[4]E. Hecht and A. Zajac, *Optics* (Addison-Wesley, Reading, MA, 1974).
[5]W. B. Davenport and W. L. Root, *An Introduction to the Theory of Random Signals and Noise* (McGraw-Hill, New York, 1958).
[6]R. Bracewell, *The Fourier Transform and Its Applications* (McGraw-Hill, New York, 1965).
[7]M. R. Schroeder and B. S. Atal, "Generalized Short-time Power Spectra and Autocorrelation Functions," J. Acoust. Soc. Am. **34**, 1679–1683 (1962).