Musical frequency tracking using the methods of conventional and "narrowed" autocorrelation

Judith C. Brown and Bin Zhang[a)]
Physics Department, Wellesley College, Wellesley, Massachusetts 01281 and Media Lab, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

(Received 20 August 1990; accepted for publication 23 December 1990)

In two recent papers, a description is given of a means of obtaining an arbitrarily narrow peak in the calculation of the autocorrelation function [J. C. Brown, and M. S. Puckette, "Calculation of a narrowed autocorrelation function," J. Acoust. Soc. Am. **85**, 1595–1601 (1989)] or of a narrow valley in the calculation of an inverse autocorrelation [J. C. Brown, and M. S. Puckette, "Musical information from a narrowed autocorrelation function," Proceedings of the 1987 International Conference on Computer Music, Urbana, Illinois, 84–88 (1987)]. These calculations are applied to the determination of the fundamental frequency of musical signals produced by keyboard, wind, and string instruments. These results are compared to frequency tracking results obtained on these sounds with conventional autocorrelation. In so doing it is determined first whether the method of autocorrelation is well-adapted to the problem of tracking the frequency of musical signals, and, second, under what conditions "narrowed" autocorrelation is advantageous.

PACS numbers: 43.60.Gk, 43.75.Yy

INTRODUCTION

Musical frequency tracking has been relatively little explored in comparison to the massive efforts that have been carried out by the speech community for use with various speech encoders for communications purposes. Musical applications have, for the most part, been in the area of intelligent systems, where an accurate frequency tracker is a necessity at the front end.

An initial effort at a musical transcription system was made by Piszczalski and Galler (1977). This is a system whose goal is to take an incoming audio signal, process it, and turn out a musical score. Their frequency tracker (Piszczalski and Galler, 1979) examined frequency ratios of spectral components to form a hypothesis for the fundamental frequency. Another musical frequency tracker proposed originally in the same year (Terhardt, 1979; Terhardt *et al.*, 1982) examined submultiples of spectral components to arrive at a fundamental frequency. Both of these methods are similar to the Schroeder (1968) histogram method.

In a series of excellent papers going back to the 1970s, the group at CCRMA in the Music Department at Stanford has worked on a musical transcription system. Various frequency trackers are discussed by Moorer (1975), Schloss (1985), Foster *et al.* (1982), Mont-Reynaud (1985), and Chafe *et al.* (1985, 1986). A recent report by Serra and Wood (1988) summarizes current work at CCRMA and gives extensive references to earlier work.

Barry Vercoe at MIT has worked on a machine intelligence problem called "the synthetic performer." Its goal is to be able to replace a member of a live ensemble of performers with a computer so that the remaining members cannot tell the difference. Early work by Vercoe (1984) and Vercoe and Puckette (1985) used a frequency tracker based on key sensors of a flute. More recently, as reported in a segment on the TV show "Discover," a spectral method similar to that of Schroeder (1968) and Amuedo (1985) has been used to track the audio signal of a violin.

I. BACKGROUND

It has been shown previously that the following equation (Brown and Puckette, 1989) can be used for the calculation of a narrowed autocorrelation function:

$$S_N(\tau)^2 = |f(t) + f(t+\tau) + f(t+2\tau) + \cdots + f[t+(N-1)\tau]|^2. \quad (1)$$

In the discussion, we showed that terms of the form $f(t)f(t+2\tau), f(t)f(t+3\tau)\ldots$, etc., in addition to the "ordinary" autocorrelation term $f(t)f(t+\tau)$ give rise to a narrowing of the autocorrelation function. Here, $f(t)$ is the time wave analyzed, t is the time, and τ is the autocorrelation time. For a periodic function, peaks occur at values of the autocorrelation time equal to multiples of the period T. When N is equal to 2 in the above equation, the conventional or standard autocorrelation function is given by the cross term with the other two terms contributing a constant shift. The equation can be normalized with the value of $S_N(\tau)^2$ for $\tau = 0$ so that the values of $S_N(\tau)^2$ range from 0–1. Most important, the width of the peaks measured from the maximum to the first zero is T/N for the time wave of a single harmonic component.

The equation for the inverted autocorrelation function was discussed by Brown and Puckette (1987). It was shown that the function

[a)] Present address: GW Instruments, 35 Medford St., Somerville, MA 02143.

 0001-4966/91/052346-09$00.80

$$p(t,\tau) = \{f(t) + f(t+\tau) + f(t+2\tau) + \ldots f[t+(N-1)\tau]\}/N$$

is the periodic function having period τ which best approximates $f(t)$ over the interval $N\tau$ in the mean-square sense. If we consider the difference

$$E(\tau) = \langle [f(t) - p(t,\tau)]^2 \rangle, \quad (2)$$

then $E(\tau)$ will have minima for τ equal to multiples of the period if $f(t)$ is a periodic function. The average over time indicated by the angle brackets is taken over a time equal to or greater than τ. Again we normalize so that the function varies from 0 to 1.

In the discussion that follows, we will refer to calculations with Eq. (2) as those of inverted autocorrelation.[1] Where necessary for distinction, calculations using Eq. (1) will be referred to as "erect" or standard autocorrelation. Within each of these categories, the calculations with $N = 2$ will be called conventional, and those with $N > 2$ will be called "narrowed."

An important property of musical sounds is that, for the most part, they have harmonic spectral components. The consequence for the autocorrelation function is that one of the peaks of each of the higher components occurs at the same position as that of the fundamental. For example, the second harmonic has a period equal to half of the fundamental so its peaks occur at $T/2$, $2(T/2)$, $3(T/2)$... and so on. The second peak of the second harmonic, then, will coincide with the first peak of the fundamental with similar reasoning for the other harmonics. Thus a large peak corresponding to the sum of all spectral components should occur at the period of the fundamental (and all integral multiples of the period of the fundamental). This is the property that makes the method of autocorrelation appear to be a good one for frequency tracking of musical signals. This property, of course, also holds true for the valleys for the inverse autocorrelation of Eq. (2).

II. CALCULATIONS

Examples of the conventional and narrowed autocorrelation functions for scales played by a piano, a violin, and a flute are shown in Figs. 1–6. Each curve is a 200-point autocorrelation function of sound from acoustic instruments sampled at 32 000 samples per second. An average is taken over 500 samples so each curve represents approximately 15 ms of sound. The time in the sound is given on the y axis with the autocorrelation time in samples on the x axis. Equation (1) was used for these calculations with $N = 2$ for the conventional autocorrelation and $N = 5$ for the narrowed autocorrelation. Examples calculated for the inverted autocorrelation function with Eq. (2) are not included.

The piano sound analyzed in Figs. 1 and 2 was from a 9-foot Bosendorfer which radiated surprisingly little energy in the higher harmonics. Since the peak widths are proportional to the period, a sound with higher harmonics present will have an innately narrow peak. As this is not the case here, this piano sound represents a good case to see the effect of

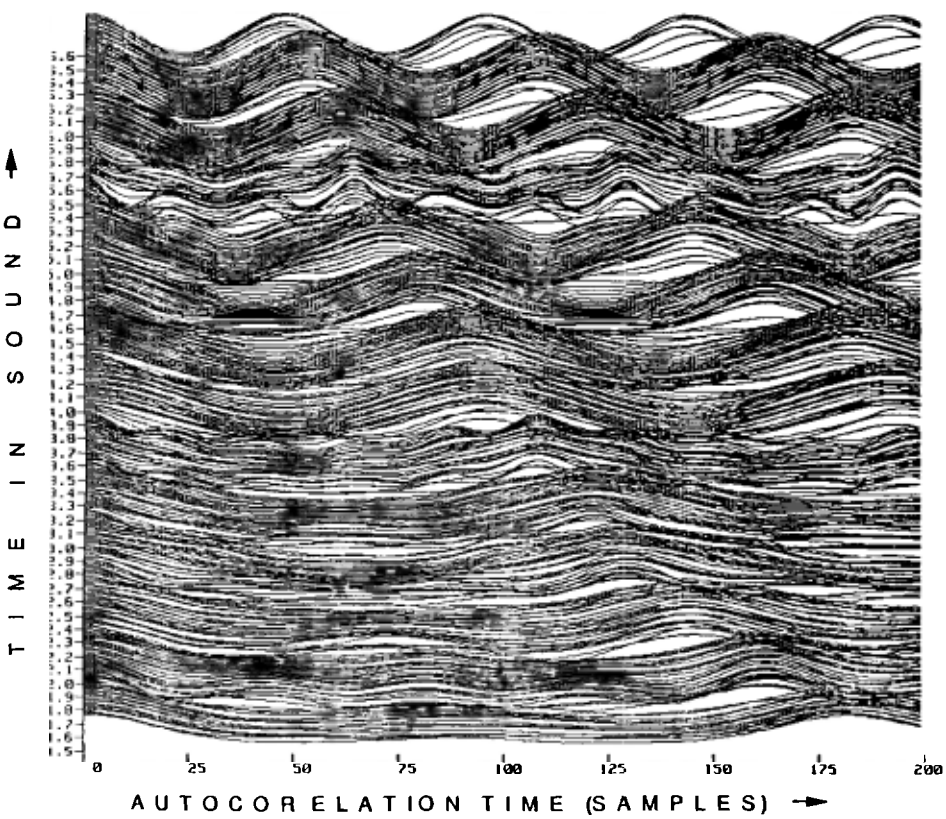

FIG. 1. Conventional autocorrelation for a C major piano scale over several octaves. This portion is from F3 to E5.

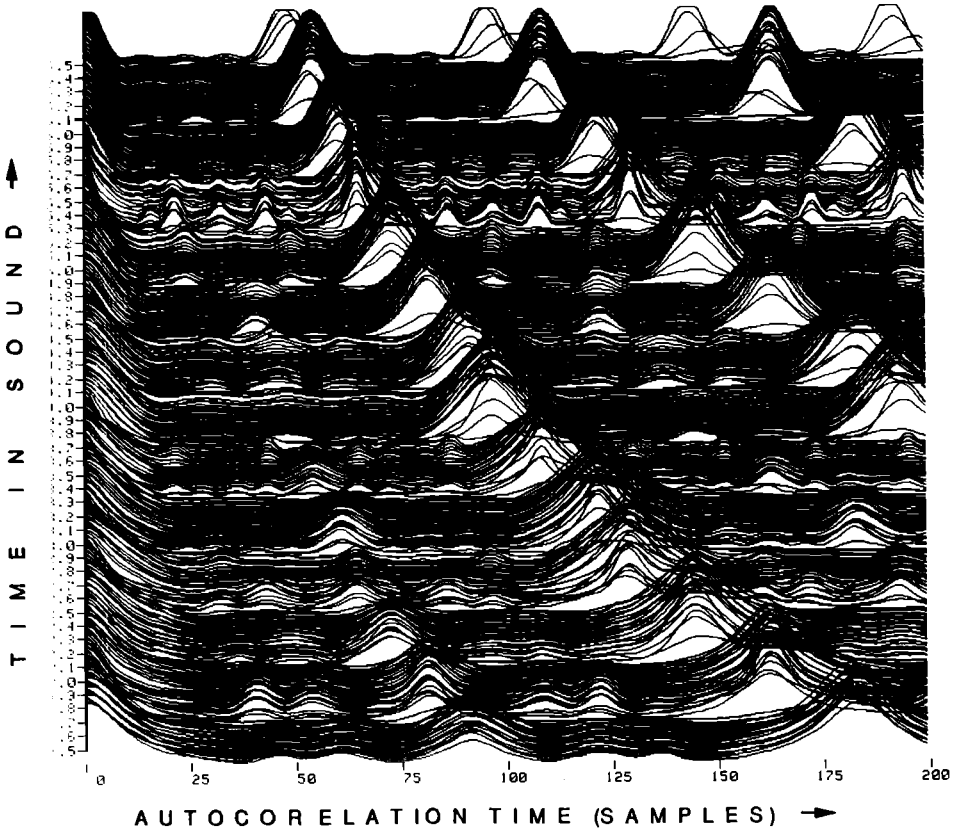

FIG. 2. Narrowed autocorrelation for the same piano sound as that of Fig. 1.

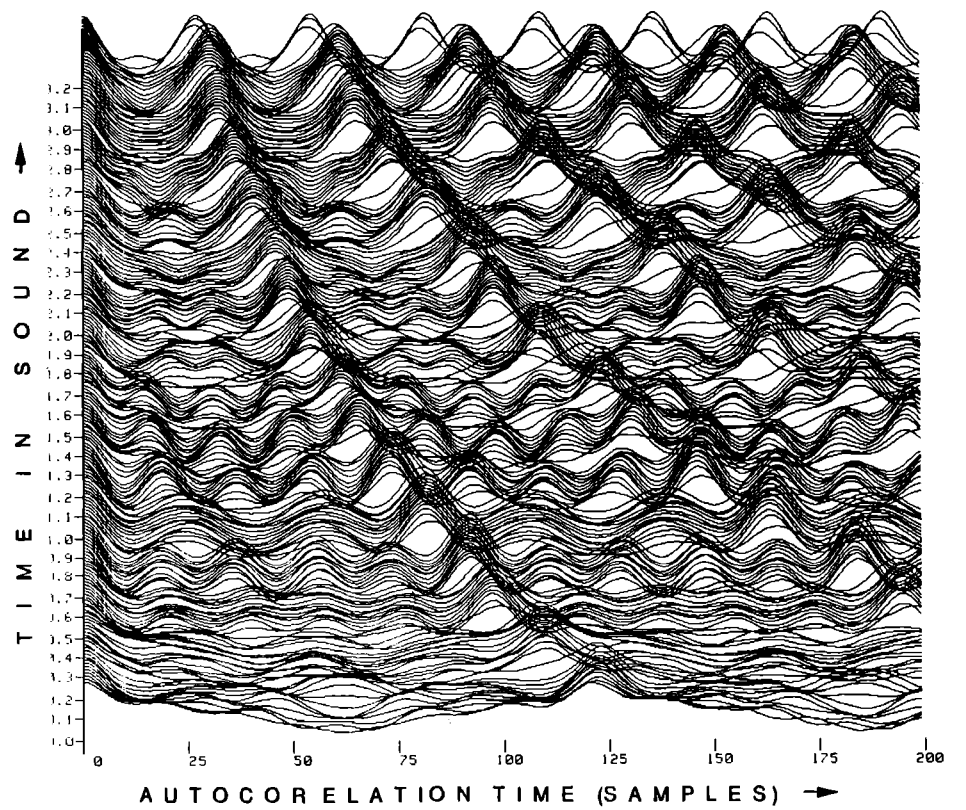

FIG. 3. Conventional autocorrelation of a flute scale from C3 to C5.

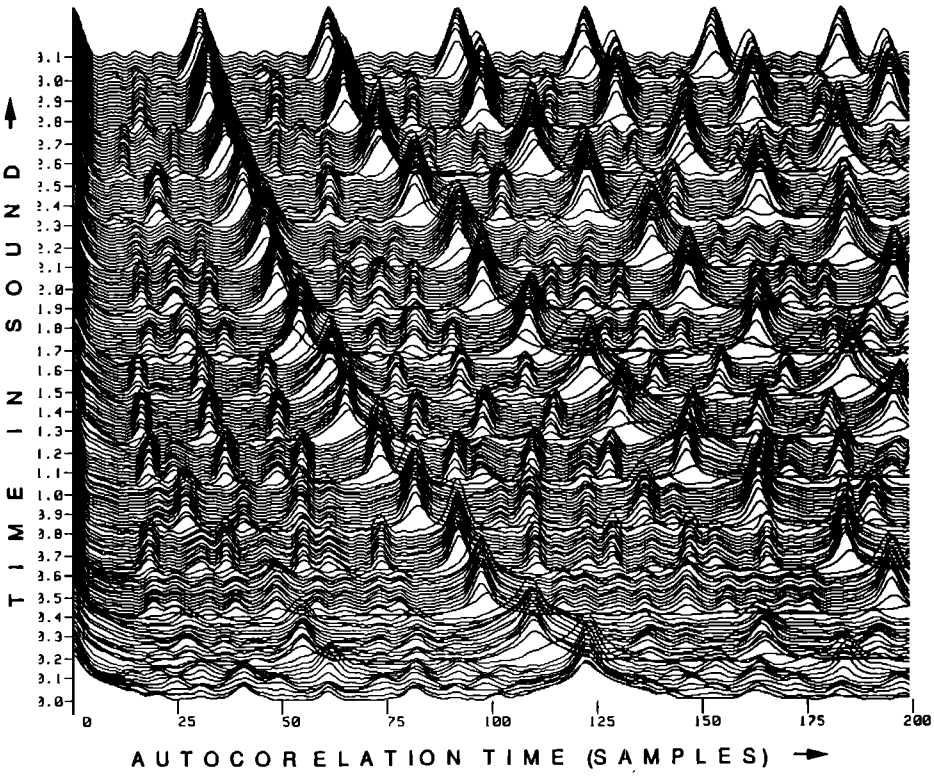

FIG. 4. Narrowed autocorrelation of a flute scale from C3 to C5.

narrowing on a broad peak. This is quite striking and is shown in Figs. 1 and 2.

The comparison of the flute scale in Figs. 3 and 4 and the violin scale in Figs. 5 and 6 show successively narrower natural peak widths (Figs. 3 and 5). The effect of narrowing in Fig. 6 compared to Fig. 5 is quite apparent and is visually preferable. One might, however, suspect that a computer could pick the natural peaks of Fig. 5 as easily as the nar-

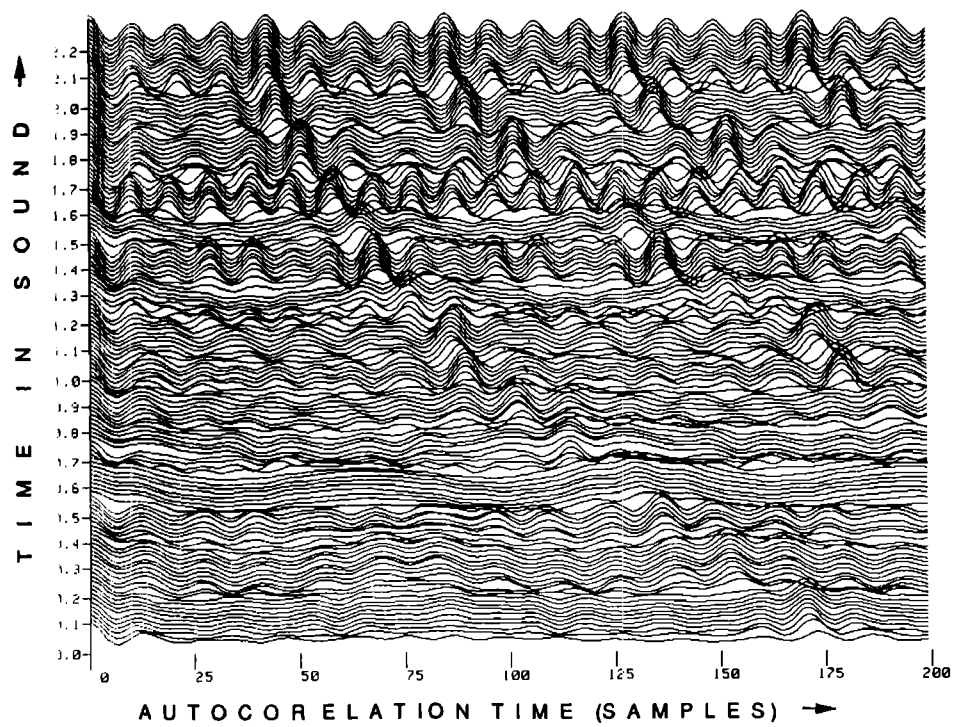

FIG. 5. Conventional autocorrelation of a violin scale from G3 to G5.

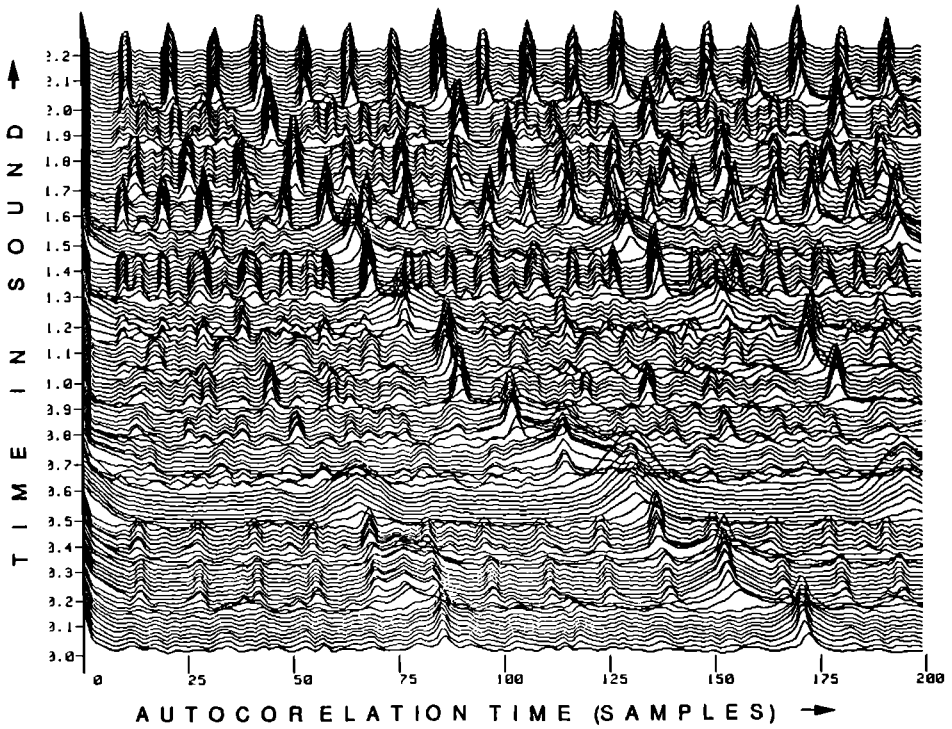

FIG. 6. Narrowed autocorrelation of a violin scale from G3 to G5.

rowed peaks of Fig. 6 since the peaks in Fig. 5 are already rather narrow.

The problem of determining the fundamental frequency for each of the curves of these figures is to find the position of the peak corresponding to one period of the fundamental. For the inverted autocorrelation, the problem is to identify the corresponding minimum. The problem is complicated by the fact that we have calculated a discrete autocorrelation, and there may not be a sample at the exact postion of the maximum (or minimum). An example is given in Fig. 7 where the valley corresponding to one period (occurring between samples 88 and 89) has a value greater than that of the valley corresponding to two periods (occurring at sample 179).

To circumvent this problem we have an adjustable parameter to denote how much higher (or lower) a peak (or valley) must be to replace the previous winner. This was 0.03 for inverted autocorrelation for both narrowed ($N>2$) and conventional ($N=2$) cases. For "erect" autocorrelation a percentage difference was used. An exhaustive study revealed that a lower value (approximately 15%) should be used for conventional than for narrowed (25%–35%) calculations. There was a small variation of the optimum value among the instruments. The larger value for the narrowed autocorrelation means that there is a greater difference in peak heights of the winning peak (sum of all harmonics) compared to other combinations of harmonics for the narrowed case. Since relative heights of individual harmonics are theoretically the same for both cases, this must result from less overlap of the narrower peaks, and is an advantage of the narrowed calculation.

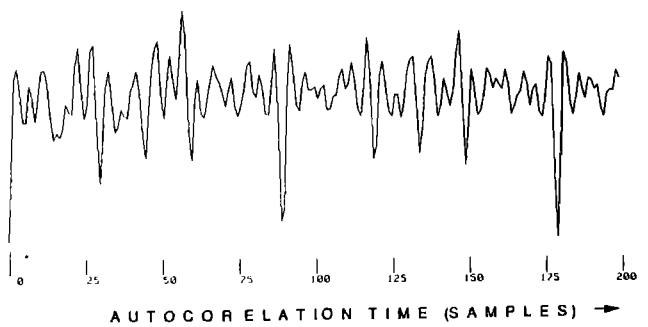

FIG. 7. Example of the narrowed inverted autocorrelation function for a violin sound where the peak at autocorrelation time equal one period is low due to discrete sampling.

We have also tried several methods of curve fitting for obtaining a better value of the maximum including a quadratic fit, cubic fit, linear extrapolation of the two points on either side of the maximum, and calculation of the maximum by considering sums of two adjacent points. This latter method was used for the narrowed autocorrelation for the violin as its results were as good as those of the other methods, and it is more efficient computationally. Other curves gave good results with no fitting procedure.

Another useful parameter is based on the normalization of these functions. Errors in frequency determination occur almost exclusively at note transitions where there is a mixture of both notes. There is often a change in amplitude of the waveform in these regions as well, and this affects the normalization. Both of these effects contribute to cause the peaks (or valleys for inverted autocorrelation) to differ from the ideal value of 1 (or 0). Our programs have the option of

returning a zero meaning no frequency determination if a peak was greater than 1.2 or less than 0.8 or if a valley was greater than 0.1. This property can be used by an intelligent system to signal note changes, if desired.

It is to be noted above that the quantity $E(\tau)$ from Eq. (2) is expected to be closer to the ideal value 0 than the quantity $S_N(\tau)^2$ from Eq. (1) is to 1 at τ equal to an integral number of periods. For example, if $f(t)$ differs from $f(t + T)$ by 10%, then $E(T)$ in Eq. (2) is $(0.1)^2 = 0.01$, whereas in Eq. (1), $S_N(\tau)^2$ is equal to 1.2. This is an advantage of inverted autocorrelation over the usual form.

Once determined, the period in samples for a given frame was passed to an array which converted the period to a midinote. The array could be "tuned" so that the tuning of the instrument studied corresponded to the appropriate indices of the array.

III. RESULTS

The frequency tracking program was run on the sounds graphed in Figs. 1–6 as representatives of the keyboard, wind, and string families of instruments. For each sound sample, we determined the maximum, subject to the con-

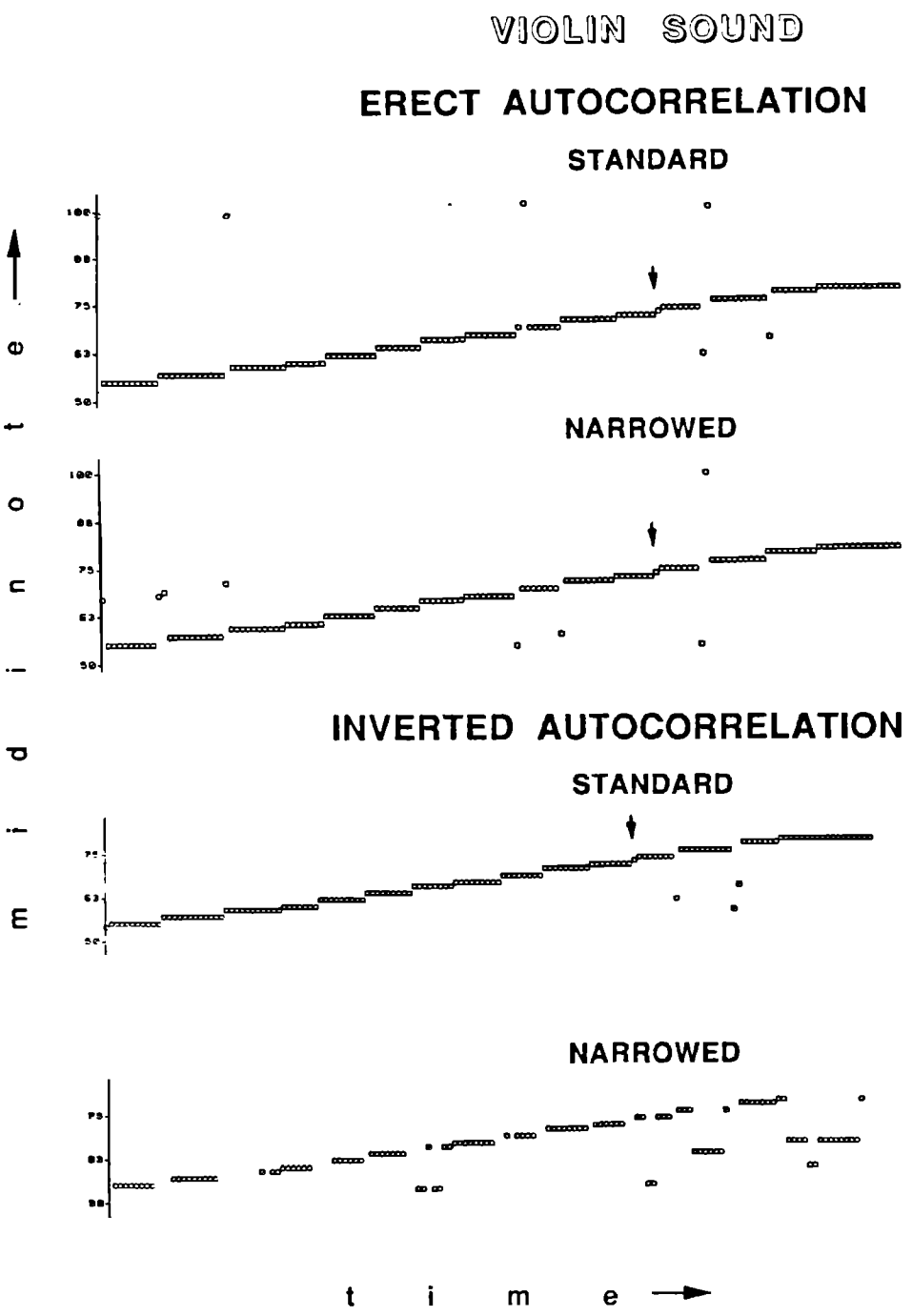

FIG. 8. Frequency tracking results using four autocorrelation methods on the violin.

straints previously described, for the autocorrelation function and for the narrowed autocorrelation function. We similarly determined the minimum for the inverted and narrowed inverted autocorrelation functions. These periods were then converted to midinotes to map the results to musical notes. Our results are shown as graphs of midinote against time in Figs. 8–10. Since each of the instruments was playing a scale, perfect results would consist of a sequential set of horizontal lines rising by one or two midinotes corresponding to a half or a whole step in the scale. Thus each error can be recognized as a point off the appropriate horizontal line. Errors are summarized in Table I. The poorest frequency tracking results of any of our calculations were obtained using narrowed inverted autocorrelation on the violin as shown at the bottom of Fig. 8. Conventional ($N = 2$) inverted did the best with less than a 3% error. Both narrowed and conventional "erect" autocorrelation did well for this case, with conventional slightly better. The violin spectrum is the richest of the instruments we studied with many high harmonics present. As mentioned previously, this means an autocorrelation line width that is already narrow, and it is not surprising that the calculations with narrowing are poor. This is due to the difficulties with discrete sampling discussed previously. Consistent with this

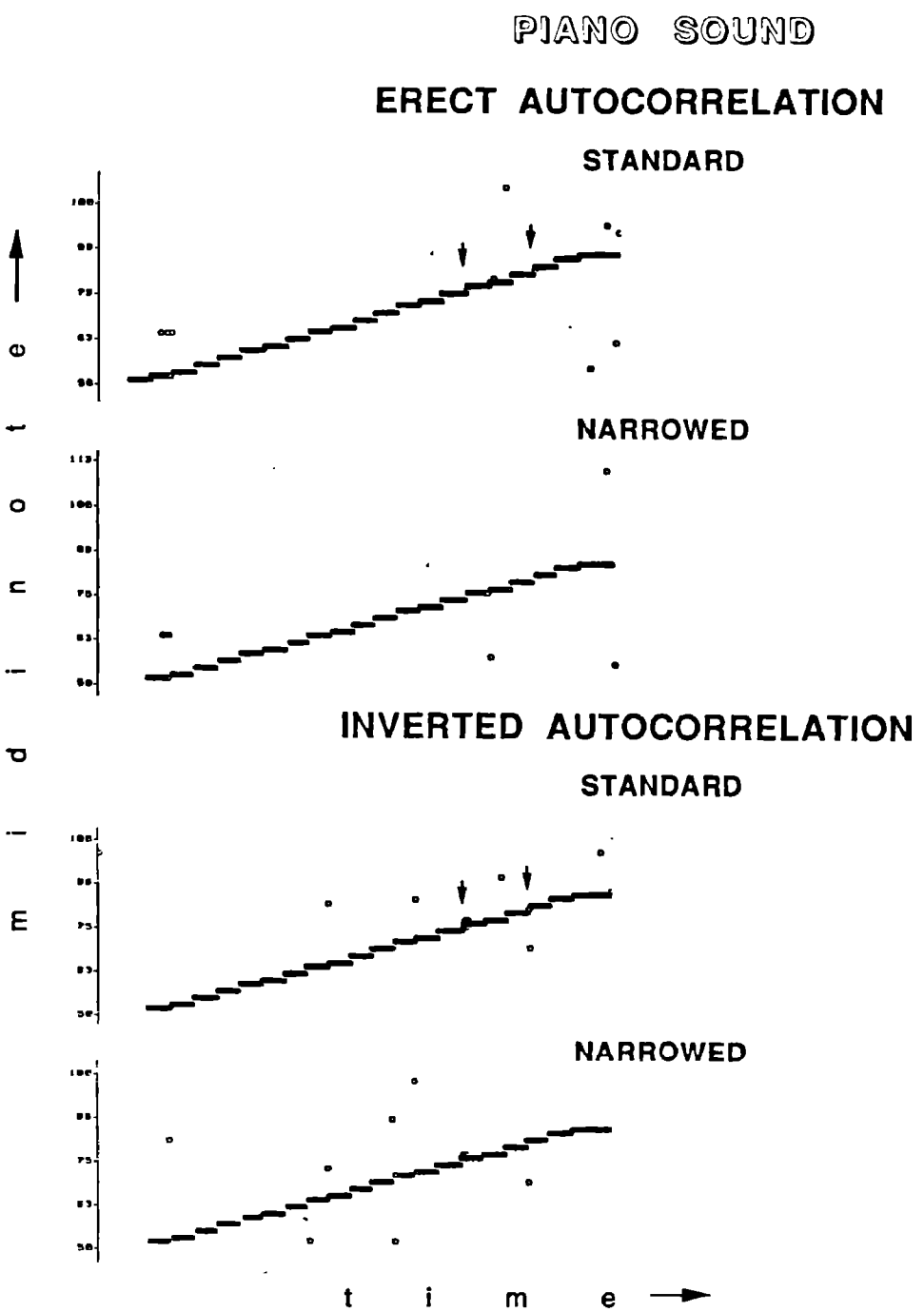

FIG. 9. Frequency tracking results using four autocorrelation methods on the piano.

mechanism, all errors have a midinote that is too low, usually by an octave. This means that the peak (or valley) corresponding to the period was missed and the second peak (or valley), corresponding to two periods, was reported.

We should note that, with the exception of the narrowed inverted results, both these and the following results could be perfect simply by requiring the frequency reported by two adjacent frames to agree. This was not done as it would eliminate the basis for comparisons in the results.

The results on the piano in Fig. 9 again show errors in transition regions. There is no significant difference in the results of narrowed over conventional for the inverted autocorrelation. Narrowed erect autocorrelation does best of all. Two errors on the graph using conventional erect autocorrelation are marked where an average of the two notes in a transition is obtained. Correct results are obtained by the narrowed calculation, and the mechanism for this improvement will be discussed further in the following section.

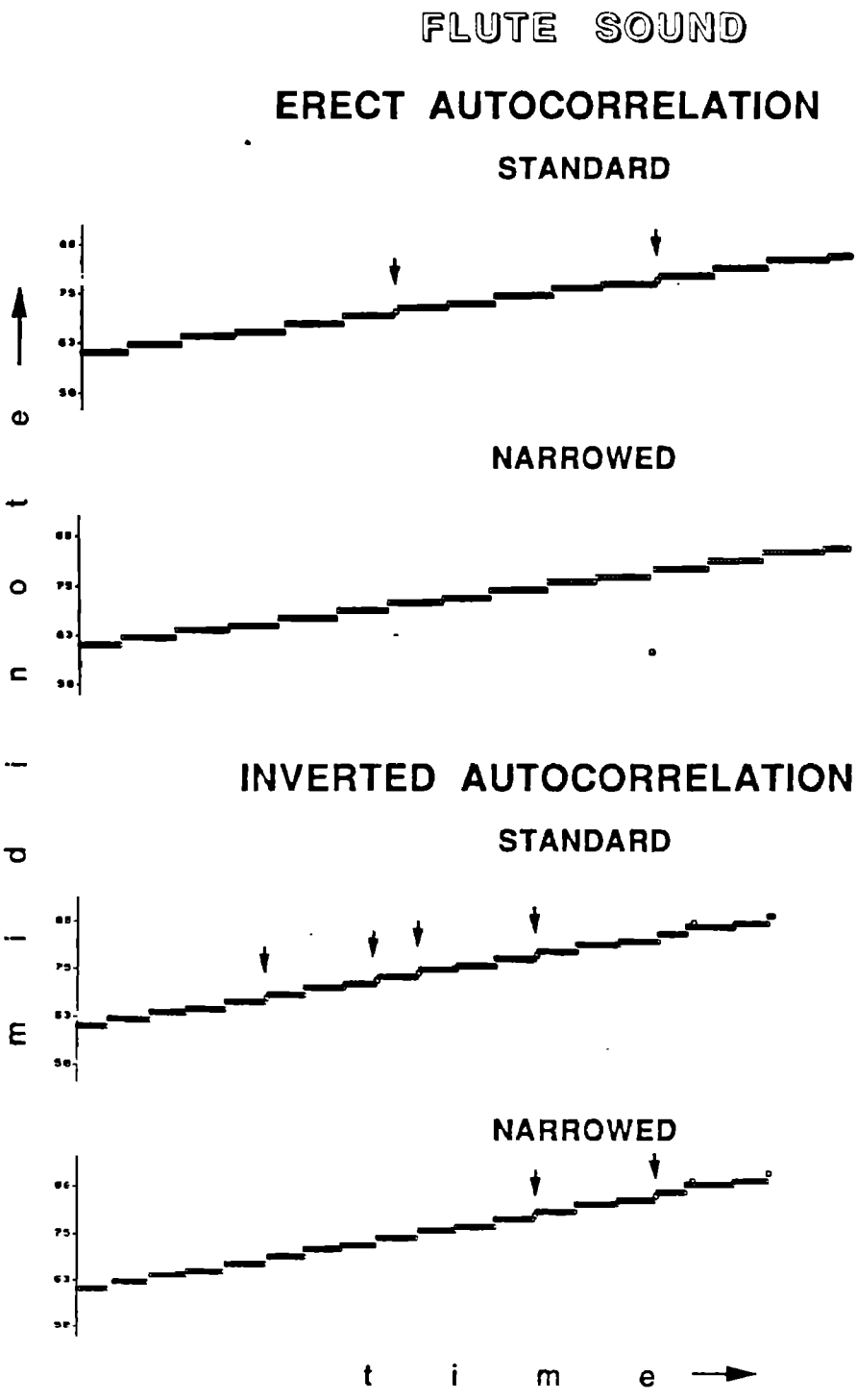

FIG. 10. Frequency tracking results using four autocorrelation methods on the flute.

TABLE I. Summary of errors by frequency trackers.

Instrument	Method	Terms	Frames	Errors
Violin	inverted	$N=2$	147	5
		$N=4$		26
	erect	$N=2$		6
		$N=5$		9
Piano	inverted	$N=2$	528	10
		$N=4$		9
	erect	$N=2$		11
		$N=5$		8
Flute	inverted	$N=2$	246	5
		$N=4$		3
	erect	$N=2$		2
		$N=5$		1

The results on the flute in Fig. 10 were the best of the instruments we studied. Again narrowed autocorrelation does a little better than conventional for both calculations with erect doing slightly better than inverted. Frames marked with arrows give an average of two notes in transition regions with narrowed doing better than conventional at reporting a single correct note.

IV. DISCUSSION AND CONCLUSIONS

As mentioned, conventional autocorrelation returns the average of two notes in some transition regions, whereas narrowed autocorrelation is better able to choose between the two notes. This mechanism was discussed in Brown and Puckette (1989). A simulation of a signal in the presence of low-amplitude noise was carried out. The signal consisted of a pure tone with another pure tone with a nearby frequency serving as noise. It was found that the presence of the noise shifted the position of the peak due to the signal for the conventional autocorrelation. With sufficient narrowing, however, the single shifted peak splits into two recognizable components with the signal peak at its proper position. It is this mechanism that is responsible for some of the improved results of narrowed over conventional autocorrelation in the transition regions for the flute and the piano.

The disadvantages of narrowed autocorrelation are twofold. First, the longer analysis time means less is known about the exact time for which the calculated midinote applies; i.e., we have the usual time/frequency trade-off. Second, the calculation is a little more expensive computationally as we have an extra $N-2$ additions, where N is the number of terms included in Eq. (1) or (2).

Finally, it should be emphasized that both conventional and narrowed autocorrelation have proven to be excellent frequency trackers for the musical sounds of this study. While single examples of musical instruments can be atypical, the spectra in this study varied widely and must thus indicate success for this method for a broad variety of musical sounds. With the exception of the narrowed inverted calculation for the violin sound, perfect results could have been obtained for the sounds in this study simply by requiring that results from two successive frames agree.

ACKNOWLEDGMENTS

JCB is very grateful to the Marilyn Brachman Hoffman Committee of Wellesley College for a fellowship for released time during which some of this work was accomplished. We would also like to thank Ken Malsky for some of the early work on frequency tracking with the inverted autocorrelation function.

[1] It is easy to verify that for the case $N=2$ with a sinusoid, one obtains a function identical to the usual autocorrelation but with the maxima and minima interchanged.

Amuedo, J. (1985). "Periodicity Estimation by Hypothesis-Directed Search," ICASSP-IEEE International Conference on Acoustics, Speech, and Signal Processing, 395–398.

Brown, J. C., and Puckette, M. S. (1987). "Musical Information from a Narrowed Autocorrelation Function," Proceedings of the 1987 International Conference on Computer Music, Urbana, Illinois, pp. 84–88.

Brown, J. C., and Puckette, M. S. (1989). "Calculation of a Narrowed Autocorrelation Function," J. Acoust. Soc. Am. 85, 1595–1601.

Chafe, C., Jaffe, D., Kashima, K., Mont-Reynaud, B., and Smith, J. (1985). "Techniques for Note Identification in Polyphonic Music," Proceedings of the International Conference of Computer Music, Vancouver, B.C., pp. 399–405.

Chafe, C., and Jaffe, D. (1986). "Source Separation and Note Identification in Polyphonic Music," Proc. ICASSP, Tokyo.

Foster, S., Schloss, W. A., and Rockmore, A. J. (1982). "Toward an Intelligent Editor of Digital Audio: Signal Processing Methods," Comput. Music J. 6, 42–51.

Mont-Reynaud, B. (1985). "Problem-solving Strategies in a Music Transcription System," Proc. IJCAI, 916–918.

Moorer, James A. (1975). "On the Segmentation and Analysis of Continuous Musical Sound by Digital Computer," Ph.D. dissertation, Stanford Department of Music Rep. No. STAN-M3.

Piszczalski, M., and Galler, B. F. (1977). "Automatic Music Transcription," Comput. Music J. 1, 24–31.

Piszczalski, M., and Galler, B. F. (1979). "Predicting Musical Pitch from Component Frequency Ratios," J. Acoust. Soc. Am. 66, 710–720.

Schloss, W. A. (1985). "On the Automatic Transcription of Percussive Music—from Acoustic Signal to High-Level Analysis," Ph.D. thesis, Department of Music Rep. No. STAN-M-27, Stanford University.

Schroeder, M. R. (1968). "Period Histogram and Product Spectrum: New Methods for Fundamental-Frequency Measurements," J. Acoust. Soc. Am. 43, 829–834.

Serra, X., and Wood, P. (1988). "Overview CCRMA (Recent Work)," Department of Music Tech. Rep. STAN-M-44, Stanford University.

Terhardt, E. (1979). "Calculating Virtual Pitch," Hear. Res. 1, 155–182.

Terhardt, E., Stoll, G., and Swann, M. (1982). "Algorithms of Extraction of Pitch and Pitch Saliance from Complex Tonal Signals," J. Acoust. Soc. Am. 71, 679–688.

Vercoe B., and Puckette, M. (1985). "Synthetic Rehearsal: Training the Synthetic Performer," Proceedings of the ICMC, pp. 275–289.

Vercoe, B. (1984). "The Synthetic Performer in the Context of Live Performance," Proceedings of the ICMC, pp. 199–200.