RESAMPLING IRREGULARLY SAMPLED DATA

Dave Hale

Abstract

Many data processing tools require that seismic data be sampled at regular intervals in space and time. Because uniform spatial sampling is often difficult, a resampling technique is needed to convert irregularly sampled data to uniformly sampled data. Two possible techniques are discussed in this paper.

The first resampling method discussed is exact, costly, and very sensitive to local error (i.e., a few bad samples). The second method is approximate, cheap, and insensitive to local error. Tests were conducted to compare the performance of the two methods in resampling data containing errors likely to be found in seismic data. The results of these tests indicate that the second method should be used to resample most irregularly sampled seismic data.

Introduction

Sampling seismic data uniformly in time is easy. Uniform spatial sampling, however, is more difficult and sometimes impossible to achieve. Geophone groups are misplaced when seismic lines cross rivers or highways. We usually try to organize land seismic surveys into a uniform grid of seismic lines, particularly in collecting 3-D data; but because of terrain or permitting problems, we are often unable to obtain this spatial uniformity. Marine seismic cables drifting sideways again result in irregular spatial sampling. We may even think of missing (dead) traces as a special case of non-uniform spatial sampling. Given then that irregular sampling exists, what can we do about it?

The simplest action is to ignore the problem by assuming that irregular samples are uniform. For low wavenumbers (spatial frequencies) this solution may be reasonable; certainly, a wavefield with a wavenumber of zero (a horizontal, constant-amplitude event) can be spatially sampled anywhere without error since all of the samples are identical. But for higher wavenumbers, corresponding to more steeply dipping events, the error in assuming uniform sampling becomes significant. Application of processing tools which assume uniform sampling (e.g., Fast Fourier Transforms, finite-difference migration, dip-filters, etc.) will further propagate and spread this error. Fortunately, we need not be content with this simple, do-nothing solution.

The remainder of this paper describes two possible methods for converting irregularly sampled data to uniformly sampled data. The first method was described in 1956 by Yen. His method will be summarized here not only because some geophysicists may be unaware of it, but also because the second method to be discussed is an approximation to Yen's solution which is more applicable to seismic data contaminated with truncations, aliasing, and noise. The second method is also faster computationally.

Yen's global replacement method

Let M be the set of indices or subscripts corresponding to m misplaced samples. The indices are integers which, when multiplied by the sampling interval Δt, yield true time, distance, or whatever is appropriate. Instead of $f(l\Delta t)$ we have $f(t_l)$ for $l \,\varepsilon\, M$ (l included in M). Assume $\Delta t = 1$ and that our data has no frequencies greater than Nyquist, *1/2* cycle/sample. Then the sampling theorem allows us to write:

$$f(t_l) \;=\; \sum_j f(j)\, sinc(t_l - j) \qquad ; l \,\varepsilon\, M$$

or

$$f(t_l) \;=\; \sum_{j \not\varepsilon M} f(j)\, sinc(t_l - j) \;+\; \sum_{j \varepsilon M} f(j)\, sinc(t_l - j) \qquad ; l \,\varepsilon\, M$$

which we can rearrange to obtain:

$$\sum_{j \varepsilon M} f(j)\, sinc(t_l - j) \;=\; f(t_l) \;-\; \sum_{j \not\varepsilon M} f(j)\, sinc(t_l - j) \;=\; g_l \qquad ; l \,\varepsilon\, M \tag{1}$$

or

$$S_{lj} f_j \;=\; g_l \qquad ; j, l \,\varepsilon\, M \tag{2}$$

where summation on the left-hand side is implied by the repeated subscript j. S_{lj} is a $m \times m$ matrix whose elements are $sinc(t_l - j)$ which may be inverted to obtain the uniform samples $f(l)$.

Yen (1956 and 1957) showed that S_{lj} is invertible and derived an *exact*, explicit formula for reconstructing the continuous signal $f(t)$ from its non-uniform and uniform samples:

$$f(t) = \sum_{j \varepsilon M} f(t_j)\, \psi_j(t) + \sum_{j \not\varepsilon M} f(j)\, \varphi_j(t) \tag{3a}$$

where

$$\psi_j(t) = \prod_{\substack{k \varepsilon M \\ k \neq j}} \left[\frac{t - t_k}{t_j - t_k} \right] \cdot \prod_{k \varepsilon M} \left[\frac{t_j - k}{t - k} \right] \cdot \frac{\sin \pi t}{\sin \pi t_j} \tag{3b}$$

and

$$\varphi_j(t) = \prod_{k \varepsilon M} \frac{(t - t_k)(j - k)}{(t - k)(j - t_k)} \cdot sinc\,(t - j) \tag{3c}$$

Because we want to replace the m misplaced samples with the uniform sample values $f(l)$, we must find the limits of $\psi_j(t)$ and $\varphi_j(t)$ as $t \to l$ for $l \varepsilon M$. Equations (3) then become:

$$f(l) = \sum_{j \varepsilon M} f(t_j)\, \psi_j(l) + \sum_{j \not\varepsilon M} f(j)\, \varphi_j(l) \qquad ; l \varepsilon M \tag{4a}$$

where

$$\psi_j(l) = \prod_{\substack{k \varepsilon M \\ k \neq j}} \left[\frac{l - t_k}{t_j - t_k} \right] \cdot \prod_{\substack{k \varepsilon M \\ k \neq l}} \left[\frac{t_j - k}{l - k} \right] \cdot \frac{1}{sinc\,(t_j - l)} \qquad ; l, j \varepsilon M \tag{4b}$$

and

$$\varphi_j(l) = \prod_{k \varepsilon M} \left[\frac{l - t_k}{j - t_k} \right] \cdot \prod_{\substack{k \varepsilon M \\ k \neq l}} \left[\frac{j - k}{l - k} \right] \cdot (-1)^{j+l+1} \qquad ; l \varepsilon M,\ j \not\varepsilon M \tag{4c}$$

Equations (4) provide the m solutions to equations (1) so Yen has, in fact, inverted S_{lj}. We can easily identify this inverse:

$$f_l = S_{lj}^{-1} g_j = R_{lj} g_j$$

$$f(l) = \sum_{j \varepsilon M} R_{lj} \left[f(t_j) - \sum_{k \not\varepsilon M} f(k)\, sinc\,(t_j - k) \right]$$

$$= \sum_{j \varepsilon M} f(t_j)\, R_{lj} - \sum_{k \not\varepsilon M} f(k) \sum_{j \varepsilon M} R_{lj}\, sinc\,(t_j - k)$$

$$= \sum_{j \varepsilon M} f(t_j)\, R_{lj} + \sum_{j \not\varepsilon M} f(j) \left[- \sum_{k \varepsilon M} R_{lk}\, sinc\,(t_k - j) \right]$$

Comparison with equation (4a) yields the identification:

$$\psi_j(l) = R_{lj} = S_{lj}^{-1} \quad ; l,j \in M$$

and

$$\varphi_j(l) = -\sum_{k \in M} R_{lk}\, sinc(t_k - j) \quad ; l \in M,\ j \notin M$$

Thus far nothing has been said about the limits of the second sum in equations (3a) and (4a). The correct procedure is to sum from $-\infty$ to $+\infty$, omitting the $j \in M$. But since our data is truncated, $f(j)$ is non-zero for only, say, n samples; and the sum may be taken over $j = 1,2,3, \cdots, n,\ j \notin M$. The assumption that $f(j)$ is zero outside this interval will naturally lead to resampling error near the edges of the data, and the magnitude and extent of this error are determined by the rate at which $\varphi_j(l) \to 0$ as $|j-l| \to \infty$. We might force $\varphi_j(l)$ to zero by tapering as a function of $|j-l|$. This approach is often used successfully in approximating the infinitely long sinc function with a finite-length, "local" function. However, whereas the envelope of $sinc(t)$ decays to zero monotonically as $|t| \to \infty$, the envelope of $\varphi_j(l)$ is not so well-behaved, tending to be large in the gaps created by misplaced samples and certainly not approaching zero monotonically as $|j-l| \to \infty$. Furthermore, whereas the symmetry of $sinc(t)$ about $t = 0$ leads us naturally to taper symmetrically, the asymmetry of $\varphi_j(l)$ about $|j-l| = 0$ leaves us wondering what sort of taper would be appropriate. Indeed, the taper might even need to be adjusted for every j since the $\varphi_j(l)$ differ dramatically (not simply shifted) for different j. All of these unfortunate properties of $\varphi_j(l)$ [and $\psi_j(l)$ as well] may not be obvious from equation (4c) [and (4b)], but they will be illustrated in examples discussed later.

Trouble in applying equations (4) is further expected when we have noisy or aliased data. Suppose that only one (the lth) of the n samples is misplaced. Then equations (4) yield:

$$f(l) = \frac{f(t_l)}{sinc(t_l - l)} + \sum_{\substack{j=1 \\ j \neq l}}^{n} f(j) \left[\frac{l - t_l}{j - t_l} \right] (-1)^{j+l+1}$$

Because all n samples are used in replacing the lth sample, we call Yen's procedure a "global" replacement method. The trouble in using a global method is that a single noisy or aliased (i.e., erroneous) sample, even one far away from the misplaced sample, will have some undesirable influence on the replaced sample value. The procedure may fail miserably if catastrophic error such as a tape "glitch" is present.

Intuition suggests that we should be able to replace a single, misplaced sample without using every other sample available. But we still do not know how to correctly "ignore" the remote samples. This problem provides motivation for a different replacement method.

A local replacement method

We begin by replacing the sinc functions in equation (1) with one of the finite-length approximations referred to earlier. Larner (1979) has shown that excellent approximations are possible with fairly short functions by sacrificing accuracy at the high frequencies. So we replace the sinc with:

$$s_{lj} \ = \ s(t_l - j) \ = \ h(t_l - j)\, sinc\,(t_l - j)$$

where $h(t)$ is a tapering function which decays to zero at $|t| = J$, to obtain:

$$\sum_{\substack{j \varepsilon M \\ |t_l - j| < J}} f(j)\, s_{lj} \ = \ f(t_l) \ - \sum_{\substack{j \not\varepsilon M \\ |t_l - j| < J}} f(j)\, s_{lj} \qquad ; l \ \varepsilon \ M \tag{5}$$

Again suppose that only one sample is misplaced. Then

$$f(l) \ = \ \frac{1}{s_{ll}} \left[f(t_l) \ - \sum_{\substack{j \neq l \\ |t_l - j| < J}} f(j)\, s_{lj} \right]$$

Now we can see why we are justified in calling this replacement procedure a "local" method. Only the samples closest to the lth sample are used in its replacement; and the method should, therefore, be less sensitive to remote catastrophic errors, such as truncations or bad data, than is the global method. Notice that we must choose J large enough to satisfy $|t_l - l| < J$ (to avoid $s_{ll} = 0$). We will assume for definiteness that $|t_l - l| < 1$ and investigate equations (5) more closely.

Equations (5) represent a *banded* system of m simultaneous linear equations for the m unknown uniform sample values $f(l)$; $l \ \varepsilon \ M$. The "bandwidth" (B) of the matrix on the left-hand side depends on the length J of our finite approximation to the sinc function and is given by: $B = 2J + 1$. Recall that Yen derived an explicit equation for the inverse of a matrix of sinc values. No comparable formula for the inverse of s_{lj} exists; the property used by Yen to obtain his inverse is destroyed by the tapering of sinc.

We should not, however, be discouraged by our inability to construct an explicit inverse. Suppose that all n samples have been misplaced (i.e., $m = n$). From equations (4) we see that replacement of all n samples will require roughly n^3 operations (multiplications and additions), assuming we have found no "correct" way to taper $\psi_j(l)$. The cost of solving equations (2) directly by Gaussian elimination is also about n^3 operations (Strang, 1980, p.5). *No computational effort is spared in having the inverse given by equation (4b)!*

The banded system of equations (5) now becomes attractive from a computational viewpoint. Solving this system for the n unknown $f(l)$ will require roughly J^2n operations (Strang, 1980, p.41), making the local method considerably cheaper than the global method.

The procedure for spatially resampling n misplaced traces of seismic data might be as follows:

1) Construct the banded, triangular decomposition of s_{lj} ($\sim J^2n$ operations).

2) For each time-slice of data:
   perform forward elimination and back-substitution to obtain the resampled time-slice ($\sim Jn$ operations).

One further simplification is worth mentioning. When misplaced samples occur sparsely, the $m \times m$ system of equations (5) may separate into systems of lower order. For example, if misplacements occur no more than once in every $2J$ samples, then all replacements require only the trivial inversion of 1×1 matrices. This occasional misplacement occurs when seismic lines cross rivers, highways, etc.

Testing the resampling methods

A series of tests were conducted to determine the performance of the global and local methods in resampling truncated, high-frequency (even aliased), and noisy data, the sort of data we collect in seismic experiments. The *known* signal used in most of these tests is given by:

$$x(t) = \begin{cases} \cos\left[\dfrac{2\pi f_{max}(t-1)^2}{100}\right] & ; 1 \leq t < 51 \\[2ex] \cos\left[\dfrac{2\pi f_{max}(101-t)^2}{100}\right] & ; 51 \leq t \leq 101 \end{cases} \tag{6}$$

This function is a chirp signal with an initial frequency of 0 cycles/sample at $t = 0$. Frequency increases linearly until $t = 51$ where frequency equals $f_{\max}$ cycles/sample, then decreases linearly back to 0 cycles/sample at $t = 101$.

A uniformly sampled version of this signal is shown in Figure 1. In this example, $f_{\max} = 0.4$ $(f_{Nyquist} = 0.5)$, and $n = 100$ uniform samples were taken at $t = l$; $l = 1,2,3, \cdots ,100$. The uniformly sampled trace is symmetric about $l = 51$ except for the missing 101st sample.

The next trace plotted in Figure 1 (entitled "sample shifts") is a sequence of 100 random numbers uniformly distributed between ±0.5. Irregular sampling times were then determined from the shifts by:

$$t_l = l + shift_l \qquad ; l = 1,2,3, \cdots ,100$$

The irregularly sampled trace, also plotted in Figure 1, was then given by equation (6) evaluated at these random sample times. This irregularly sampled trace is plotted as though it were sampled uniformly. If we were to incorrectly assume uniform sampling, we readily verify that our greatest error would be in the high-frequency, middle portion of the signal.

Three attempts were then made to construct the uniformly sampled trace from the irregular samples, and the results are shown in Figure 2. The first three traces are the resampled traces; the next three are the resampling errors, the *absolute*, sample-by-sample differences between the (known) uniformly sampled trace and the resampled traces.

The traces labeled "yen" correspond to results obtained using Yen's global method outlined by equations (4). "Local4" corresponds to the local method where J, the length of the sinc approximation, equals 4. "Local8" corresponds to $J = 8$. A quick examination of the errors and comparison of the traces in Figures 1 and 2 illustrate that all three methods produce reasonable uniform samples except near the ends of the traces. Detailed comparison of the errors is probably unwarranted since these errors correspond to one particular set of random sample shifts.

To provide a more reliable estimate of the errors expected in using these methods, the experiment described in Figures 1 and 2 was repeated 100 times, each time using a different set of random sample shifts. The absolute errors for each of the 100 trials were averaged to obtain the traces shown in Figure 3. Also shown in Figure 3 is the average, absolute error obtained by doing nothing, simply assuming that the irregular samples were uniform. Notice that this "do-nothing" error is approximately symmetric about sample 51. This symmetry, expected since the uniformly sampled trace is symmetric about sample 51, implies that an average of 100 trials reasonably approximates the expected error for a given trial.

Now examine the errors near the ends of the traces. Doing nothing happens to work well here only because the data is almost constant near the ends. Both global and local methods yield significant errors near the ends where the data is truncated. As expected, however, the *extent* of this error is less in local4 than in local8 than in yen. Compare, for example, the errors in constructing sample 5. (Additional tests not shown with $f_{max} = 0$ confirm that the error near the ends is indeed due to truncation and not to an inability of the methods to resample low-frequency signals.)

The slight asymmetry of the error in "yen" further emphasizes the extensive influence of truncations when using a global method. This asymmetry is due to the fact that the uniform signal is not truly symmetric about sample 51. The signal length is 100 samples; true symmetry would require the 101st sample, and its absence results in a slightly higher level of error in the last half of the trace than in the first half.

Now compare the errors near sample 51. The local4 method performs poorly at the higher frequencies. Why? The local4 method is based on a finite-length ($J = 4$) approximation to the sinc function. Because this particular approximation is known to break down at a frequency of about 0.3 cycles/sample (Larner, 1979), we expect the method to produce errors near sample 51 where signal frequencies are about 0.4 cycles/sample. The sinc approximation for the local8 method is known to be good from 0 to about 0.4 cycles/sample, so this method, not surprisingly, works well within this frequency range.

The results so far confirm what we might have expected, that the local method performs better than the global method in the presence of truncations but at the cost of performance at the higher frequencies.

Truncation is just one of the problems we are likely to encounter in resampling irregularly sampled data. Aliasing is another. To test the performance of the three methods in the presence of aliasing, the experiment described earlier was repeated with $f_{max} = 0.51$. Because $f_{Nyquist} = 0.5$ the uniformly sampled data is just slightly aliased and then only in the region near sample 51. This "local aliasing" occurs in seismic data, for example, when a steeply dipping diffraction tail overlays predominately horizontal reflection events.

Figure 4a illustrates the uniform, irregular, and resampled traces for one of the 100 trials conducted in this experiment. The average absolute errors for the different resampling methods are plotted in Figure 4b. The uniformly sampled trace in Figure

4a illustrates the aliased samples about sample 51. For these samples, *none of the resampling methods are valid;* equation (1) no longer holds for this data. So we should not find the errors near sample 51 in Figure 4b too surprising or interesting. More important is the error in the non-aliased samples. The average error in constructing samples 10 through 40 with the local methods is significantly less than the error obtained with the global methods in which the few slightly aliased samples around sample 51 influence the replacement of *all* samples. With the local method, the error is confined primarily to the aliased region.

In addition to truncations and aliasing, seismic data is usually contaminated with noise. This noise is often localized (e.g., one dead trace, a tape glitch, etc.) so we have reason to test the performance of a resampling method in the presence of localized noise. We, therefore, repeated the first experiment (with $f_{max} = 0.4$); but this time we set the value of the 26th irregular sample to zero. The absolute errors *for one trial* in constructing the uniform samples are shown in Figure 5.

As we should expect, all three resampling methods perform poorly near the 26th sample. Yen's global method, however, allows this one bad sample to significantly affect the replacement of a much wider range of neighboring samples than do either of the local methods. Local4, while still producing significant error at the high frequencies, produces the least amount of error near the bad 26th sample.

Other replacement methods

Many alternate methods for resampling irregularly sampled data exist, some of which are currently being used regularly by seismic data processors. But we know of no one using Yen's method. Sankur and Gerhardt (1973) tested Yen's method along with several other resampling techniques: low-pass filtering, Karhunen-Loeve interpolation, spline interpolation, and linear and zero-order-hold interpolation. Their conclusion was that Yen's technique is superior to any of the other methods tested. This result should not be too surprising since Yen's method is theoretically *exact* in the absence of truncations, aliasing, or noise. Sankur and Gerhardt, however, do not wholeheartedly recommend Yen's method because it is computationally expensive. In fact, to avoid the $\sim n^3$ operation count, they windowed the $\psi_j(l)$ functions in equations (4). In constructing the lth uniform sample, they let the indices k and j run from $l - W/2$ to $l + W/2$ in the products of equation (4b) and the sum of equation (4a), respectively, where W was their chosen window length. This windowing, while decreasing the operation count of Yen's method to $\sim W^2 n$, introduces resampling errors.

Recall that we have no reasonable way to choose W; we do not know how to taper $\psi_j(l)$, the inverse of S_{lj}, as a function of $j-l$.

To illustrate this problem, we generated two examples of $\psi_{51}(l)$ corresponding to two different sets of random sample shifts. These functions represent the weight applied to the 51st irregular sample in constructing the lth uniform sample and are labeled "yen" in Figures 6a and 6b. The largest value of $\psi_{51}(l)$ is $\psi_{51}(51)$ as we might expect, but both figures demonstrate that the 51st sample contributes significantly to the construction of remote samples as well. How do we taper (or window) these functions?

A fundamental purpose of this paper has been to demonstrate that, in practice, we should avoid this question by tapering the sinc matrix S_{lj} rather than its inverse $\psi_j(l)$. That tapering S_{lj} produces a localized $\psi_j(l)$ is perhaps not obvious. (After all, the inverse of a banded matrix is not necessarily a banded matrix.) So to illustrate the nature of $\psi_j(l)$ for the local methods, $\psi_{51}(l)$ is plotted for both local4 and local8 in Figures 6a and 6b. Comparison of $\psi_{51}(l)$ for the three methods reveals the relatively limited range of influence of sample 51 obtained with the local methods. The last two traces in Figures 6a and 6b are the sample-by-sample ratios of $\psi_{51}(l)$ for the local and global methods; they represent the unwieldy tapers one would need to apply to $\psi_{51}(l)$ for the global method to obtain $\psi_{51}(l)$ for the local methods. The message is clear: *do not taper or window $\psi_j(l)$; taper S_{lj} instead.*

Conclusions

The sizable errors obtained by "doing nothing" about misplaced samples suggest that we should do something. For "ideal" data, Yen's method is exact. For seismic data, contaminated with truncations, aliasing, or bad data, a local replacement method performs better and at much less computational cost than Yen's method. The error in using a local method is significant only for frequencies at which the approximation to the sinc function is in error. For the $J = 4$ and $J = 8$ approximations used in the tests discussed earlier, the error is negligible for frequencies below 60% and 80% of the Nyquist frequency, respectively. Longer sinc approximations increase the upper frequency limit of the approximation at the cost of making the method less local. For truncated data we should use very short approximations near the end samples and longer approximations near the middle samples.

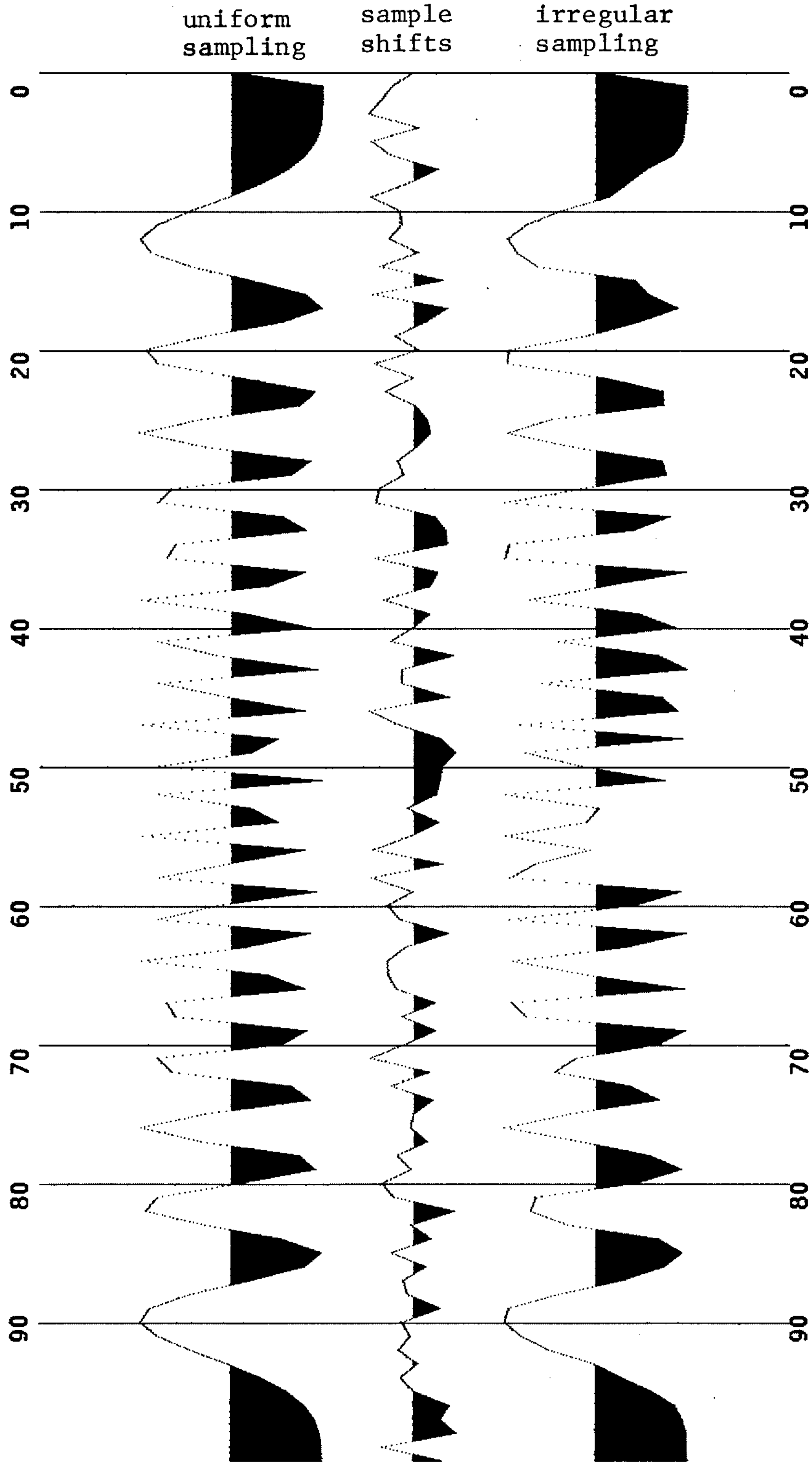

FIG. 1. Uniformly and irregularly sampled "chirp" signals used to test the resampling methods. The highest frequency in the chirp is 80% Nyquist. The middle trace displays the random shifts by which the irregular samples are misplaced from their uniform sample positions. The clip for this plot is 2 (meaning that an amplitude of 2 corresponds to 1 trace-separation and amplitudes greater than 2 are clipped).

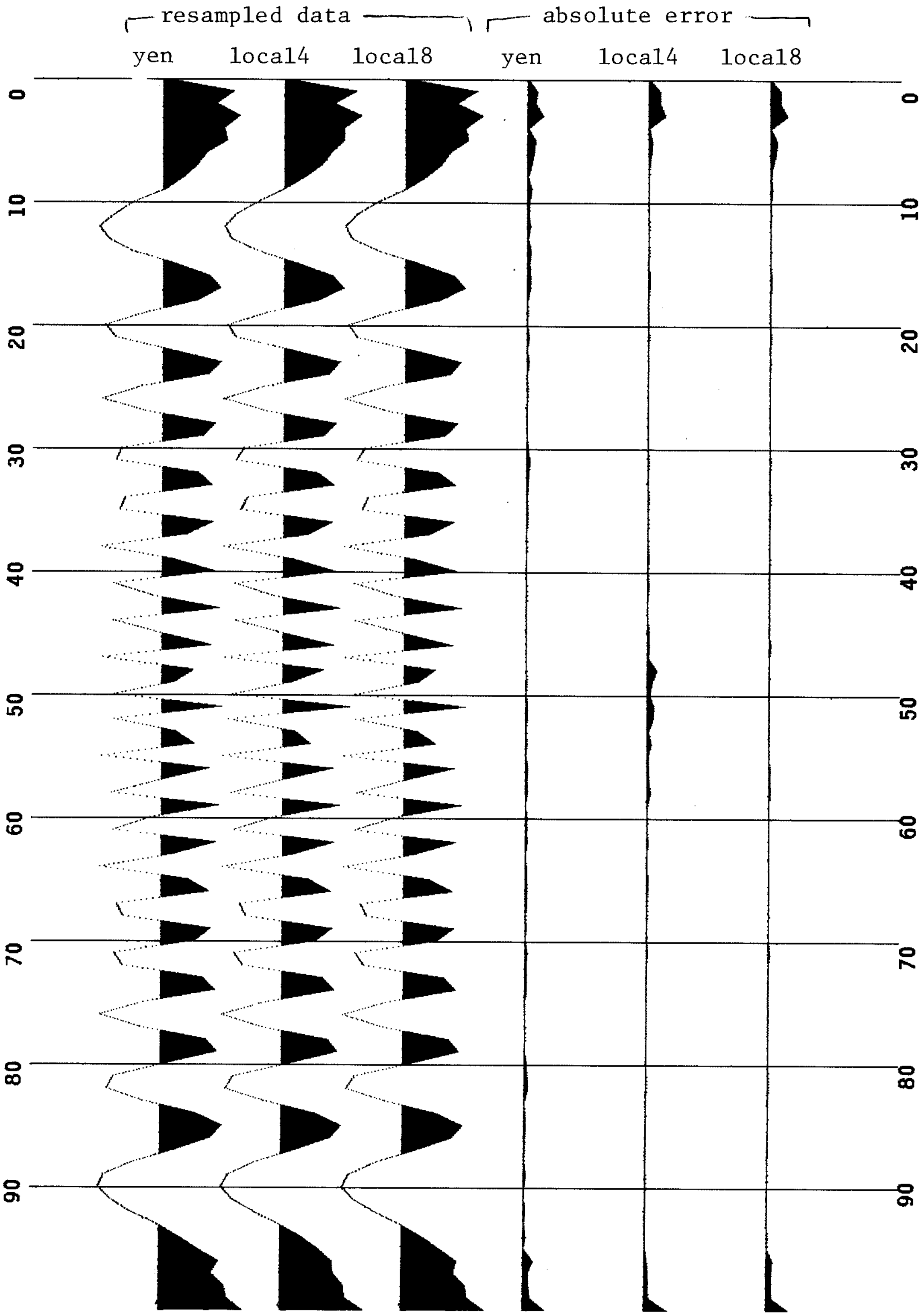

FIG. 2. Resampled traces and the absolute errors for the three methods tested. "Absolute error" is the absolute value of the difference between a resampled trace and the uniformly sampled trace of Figure 1. Clip = 2.

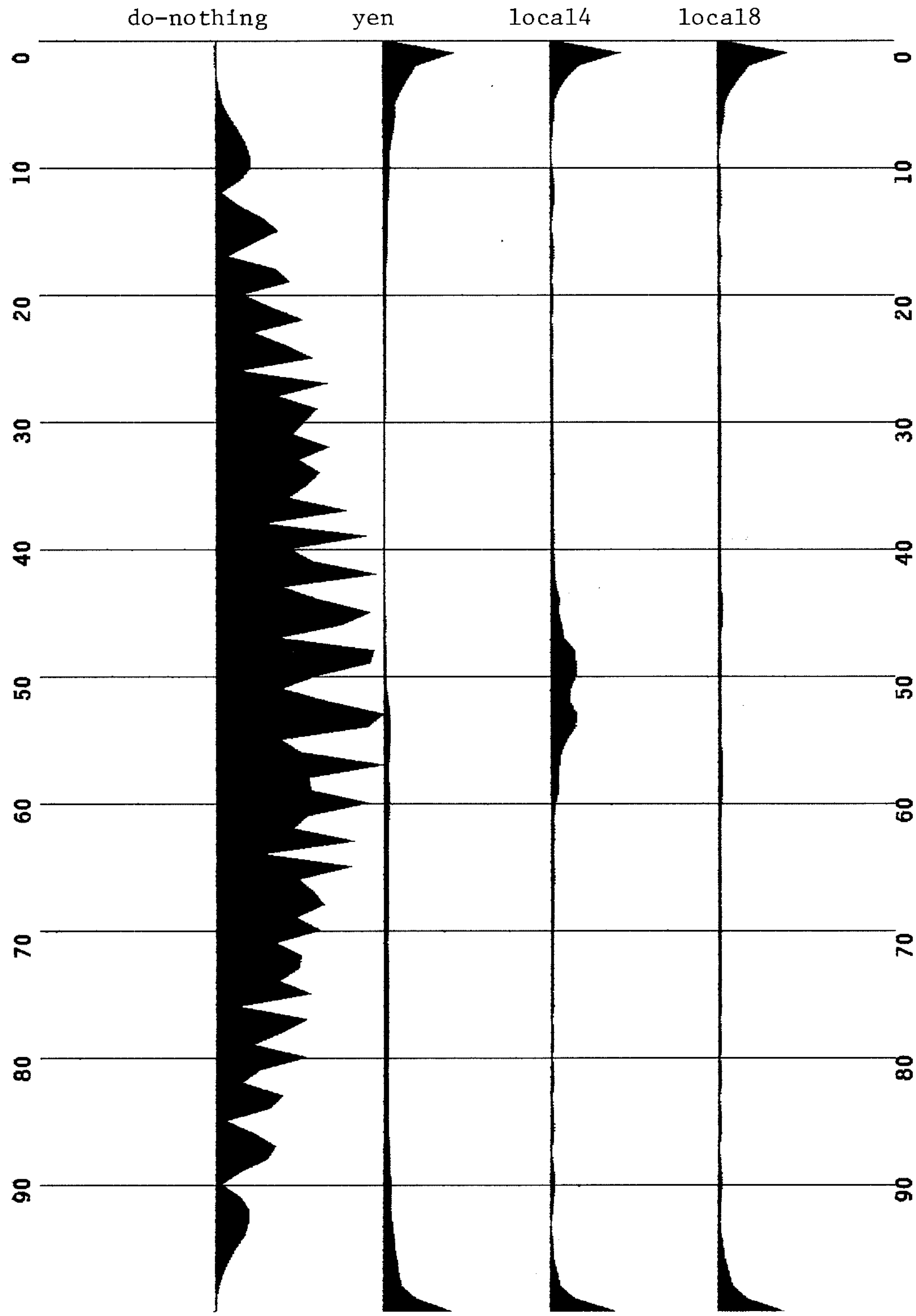

FIG. 3. Average absolute resampling errors for the three methods tested along with the average absolute error in not resampling. "Average" means the average over 100 trials, 100 different sets of random sample-shifts. Compare the extent of the error due to data truncations for the three methods and notice the error at high-frequencies in local4. Clip = 0.5.

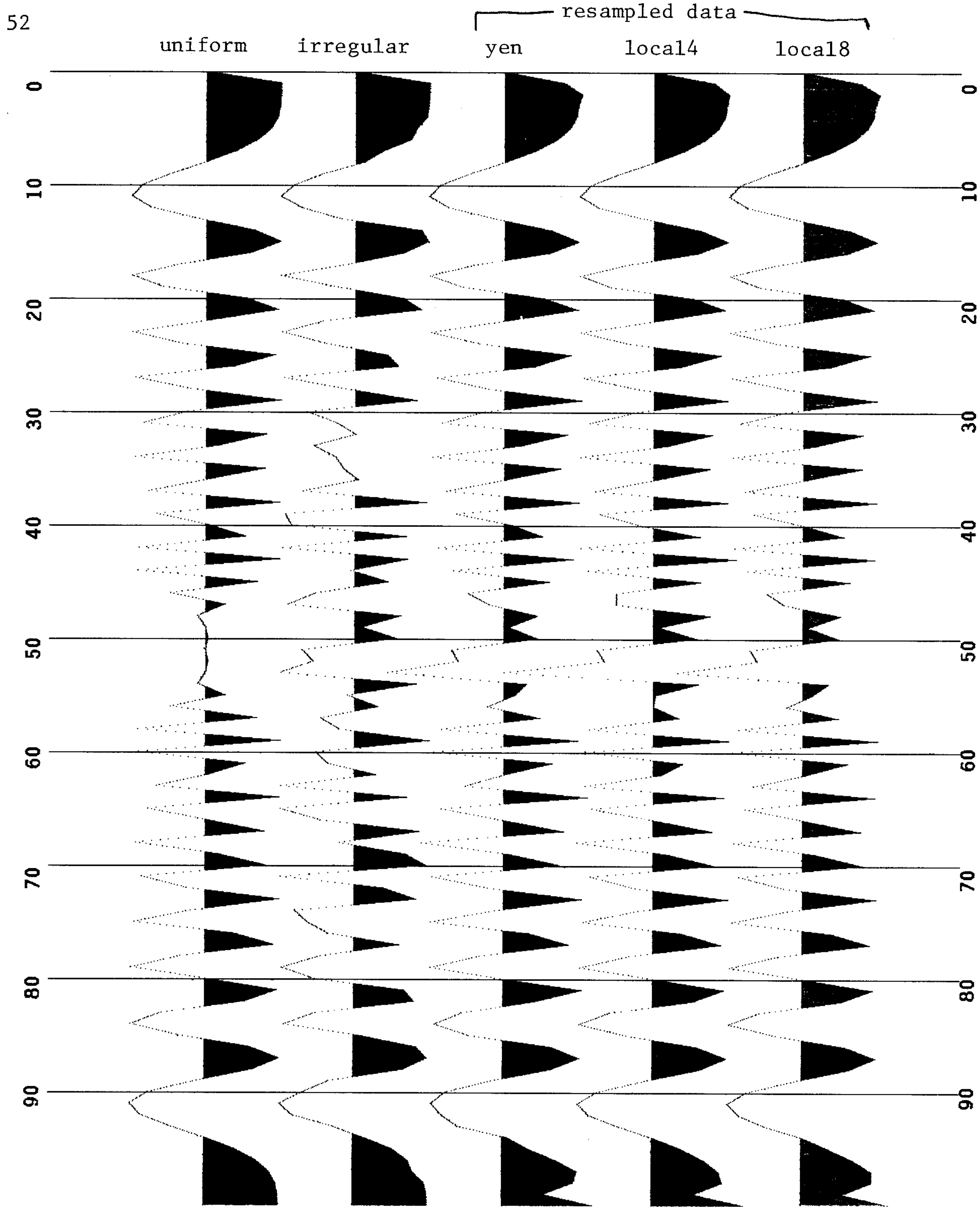

FIG. 4a. Uniform, irregular, and resampled traces from one trial where the uniformly sampled trace is slightly aliased at the middle samples. Clip = 2.

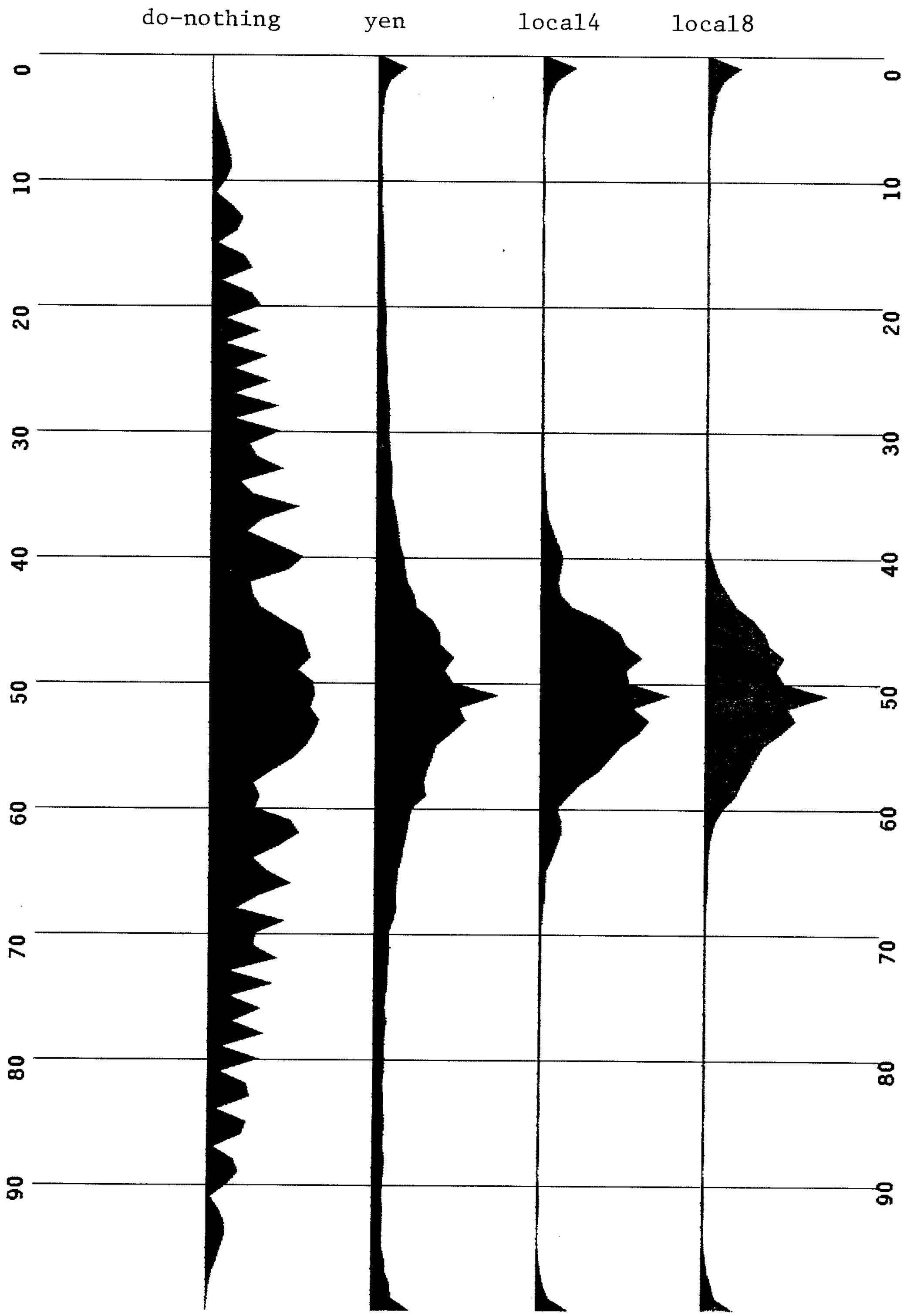

FIG. 4b. Average absolute resampling errors (100 trials) for the aliased data of Figure 4a. Compare the errors in the non-aliased regions. Clip = 1.

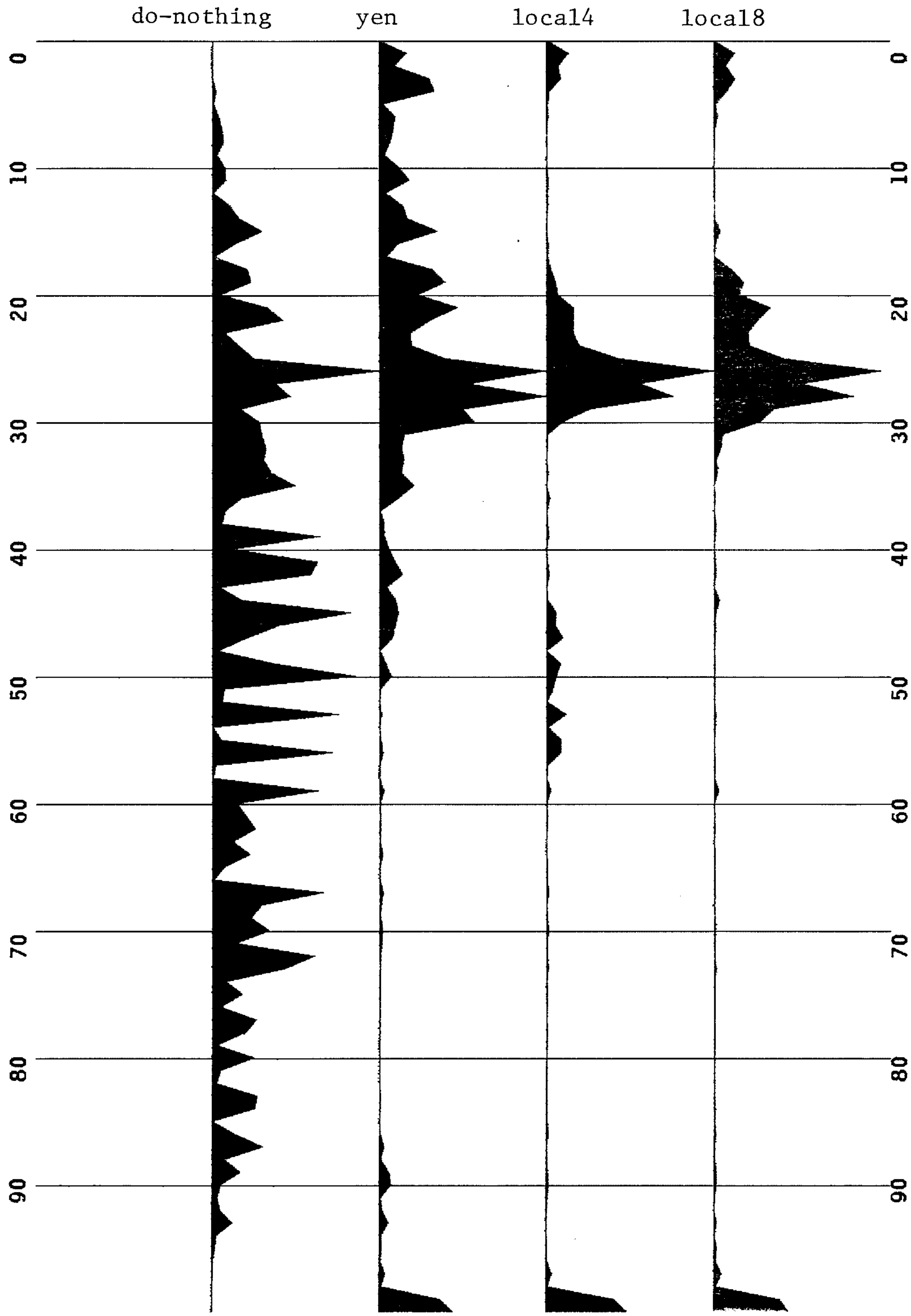

FIG. 5. Absolute resampling errors for one trial when the 26th irregular sample was set to zero (simulating a dead seismic trace). Notice the extent of the resampling errors for the different methods. Clip = 1.

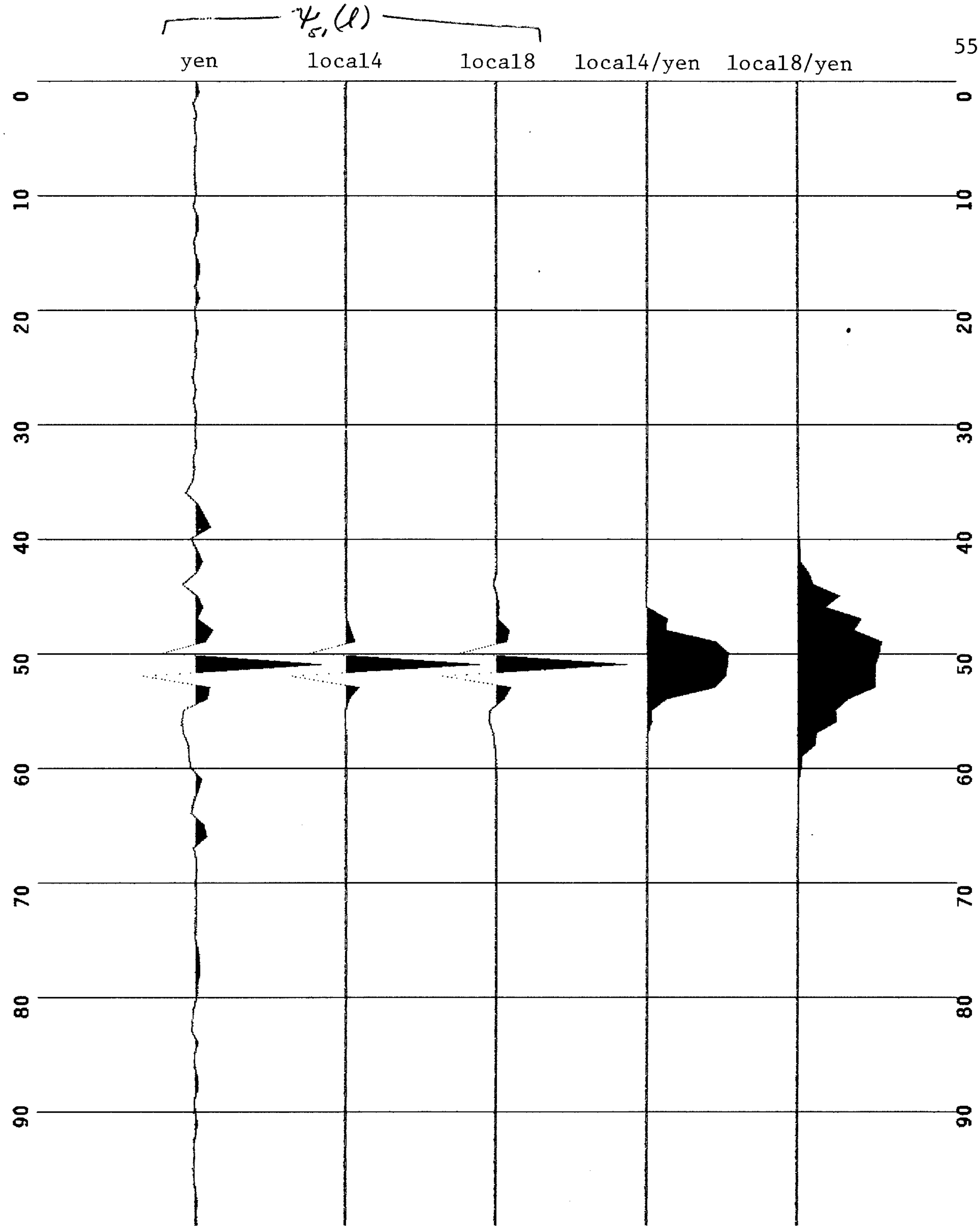

FIG. 6a. $\psi_{51}(l)$ corresponding to a particular set of random sample-shifts for the three resampling methods. The value plotted at sample 65, for example, is the weight applied to the 51st irregular sample in constructing the 65th uniform sample. Also plotted are the "taper" functions one would need to apply to yen to obtain local4 and local8. Clip = 2.

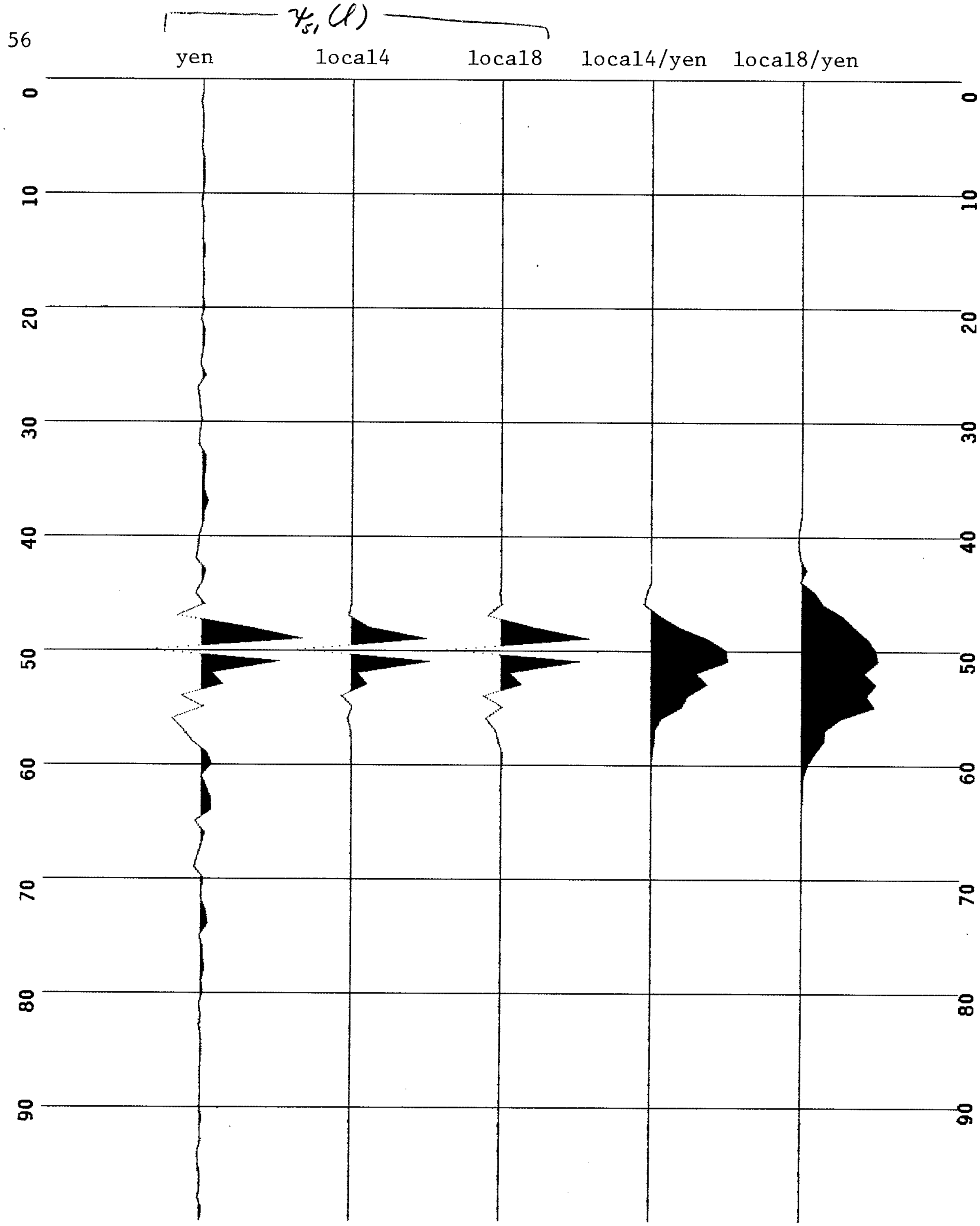

FIG. 6b. Same as Figure 6a except that the $\psi_{51}(l)$ and tapers correspond to a different set of random sample-shifts. Notice the considerable differences between the $\psi_{51}(l)$ in this figure and those in Figure 6a. Clip = 2.

REFERENCES

Larner, K.L., 1979, Report on interpolation operator design: Western Geophysical Co. Technical Document 79.08-13.01 WTR.

Sankur, B., and Gerhardt, L.A., 1973, Reconstruction of signals from non-uniform samples: IEEE Int. Conf. Commun., Conf. Rec., v. 9.1, p. 15.13-15.18.

Strang, G., 1980, Linear algebra and its applications: New York, Academic Press.

Yen, J.L., 1956, On the nonuniform sampling of bandwidth-limited signals: IRE Trans. Circuit Theory, v. CT-3, p. 251-257.

---------, 1957, On the synthesis of line-sources and infinite strip-sources: IRE Trans. on Antennas and Propagation, v. AP-5, p. 40-46.