

# A NONPARAMETRIC TEST FOR STATIONARITY BASED ON LOCAL FOURIER ANALYSIS

Prabahan Basu, Daniel Rudoy, and Patrick J. Wolfe

Statistics and Information Sciences Laboratory  
Harvard Engineering and Applied Sciences  
33 Oxford Street, Cambridge, MA 02138  
{pbasu, rudoy, patrick}@seas.harvard.edu

## ABSTRACT

In this paper we propose a nonparametric hypothesis test for stationarity based on local Fourier analysis. We employ a test statistic that measures the variation of time-localized estimates of the power spectral density of an observed random process. For the case of a white Gaussian noise process, we characterize the asymptotic distribution of this statistic under the null hypothesis of stationarity, and use it to directly set test thresholds corresponding to constant false alarm rates. For other cases, we introduce a simple procedure to simulate from the null distribution of interest. After validating the procedure on synthetic examples, we demonstrate one potential use for the test as a method of obtaining a signal-adaptive means of local Fourier analysis and corresponding signal enhancement scheme.

**Index Terms**— Hypothesis testing, stationarity, adaptive STFT, nonparametric spectral estimation, Wold decomposition

## 1. INTRODUCTION

Stationarity of random processes can be characterized in a variety of ways. When processes are completely described by parametric models (e.g., rational transfer functions), the constancy of model parameters over time is necessary and sufficient to guarantee their stationarity—an observation previously used to construct statistical tests for this property [1]. However, in many practical cases, a lack of prior knowledge about any underlying parametric model can render such parametric tests of limited use. On the other hand, time-invariance of coefficients arising in nonparametric (transform-based) representations is also sufficient to guarantee stationarity.

Therefore, we propose a nonparametric test for stationarity based on the well-known approach of checking that the statistics of transform coefficients over epochs of the signal do not deviate greatly from their sample mean. A similar approach based on the wavelet transform is described in [2]. Our interest here is in short-time spectral representations, as these are known to approximately diagonalize the covariance structure of signals which are roughly stationary over intervals [3]. As such, these representations are frequently used in the enhancement of locally stationary random processes. Accordingly, we define a test statistic that measures the variation of short-time spectral coefficients through time. Even though the sampling distributions of the test statistic are not analytically tractable, we can implement a constant false alarm (CFAR) hypothesis test using a Monte Carlo procedure, similar to one described in [2], to approximate the sampling distribution under the null hypothesis.

## 2. A TEST FOR STATIONARITY

Given  $N$  observations of a zero-mean discrete-time random process ( $x[n]$ ,  $n \in \mathbb{Z}$ ), we are interested in testing the hypothesis:

$$\begin{aligned} \mathcal{H}_0 : x[n] \text{ is wide-sense stationary (WSS)} \\ \mathcal{H}_1 : x[n] \text{ is nonstationary.} \end{aligned} \quad (1)$$

Wide-sense stationarity of  $x[n]$  implies that its autocorrelation function  $r[n, m] \triangleq E(x[n]x[m])$  depends only on the lag  $\tau = n - m$ , therefore we write  $r[n, m] = r[\tau]$  in this case. If we partition the  $N$  observations into  $M$  non-overlapping rectangular windows each of length  $L$ , then by the Wiener-Khinchine theorem, the power spectral density (PSD) corresponding to the  $m^{\text{th}}$  segment is:

$$S_{xx}^m[k] = \frac{1}{\sqrt{L}} \sum_{\tau=-L/2}^{L/2} r_m[\tau] e^{-i2\pi k\tau/L}. \quad (2)$$

Even when the process  $x[n]$  is not stationary, we can still loosely interpret the quantity  $S_{xx}^m[k]$  in (2) as the Fourier transform of the instantaneous or frozen-time autocorrelation of the process. Thus, a process is wide-sense stationary *only if* the relation  $S_{xx}^m[k] = S_{xx}^{m'}[k]$  holds for every frequency bin  $k$  and all  $m, m' \in [1, \dots, M]$ . Consequently, consider the function  $V(x)$  which measures the amount of spectral variation in windowed data segments over time:

$$V(x) \triangleq \frac{1}{ML} \sum_{k=0}^{L-1} \sum_{m=0}^{M-1} \left( S_{xx}^m[k] - \frac{1}{M} \sum_{p=0}^{M-1} S_{xx}^p[k] \right)^2. \quad (3)$$

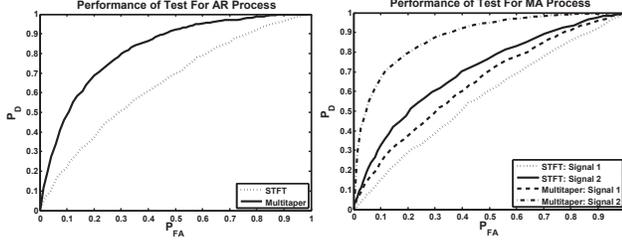
Clearly, if  $V(x) = 0$  then  $x[n]$  is a WSS process. We can turn (3) into a test statistic replacing  $S_{xx}^m[k]$  by its estimate from observed data. Specifically, we consider the periodogram and multitaper periodogram [4] estimators of the PSD which are given by:

$$\widehat{S_{xx}^m}[k] \triangleq |X_m^w[k]|^2 \quad \text{and} \quad \widetilde{S_{xx}^m}[k] \triangleq \frac{1}{R} \sum_{r=1}^R |X_m^{w_r}[k]|^2. \quad (4)$$

Here,  $X_m^w[k]$  is the discrete short-time Fourier Transform (STFT):

$$X_m^w[k] = \frac{1}{\sqrt{L}} \sum_{n=Lm+1}^{L(m+1)} w[n-Lm]x[n]e^{-i2\pi kn/L}, \quad (5)$$

where  $w$  is a rectangular window of length  $L$ , and  $X_m[k]$  is the  $k^{\text{th}}$  frequency component of the  $m^{\text{th}}$  STFT window. The multitaper spectrum  $X_m^{w_r}[k]$  is found by choosing the window  $w = w_r$  in (5)



**Fig. 1.** ROC curves summarizing test performance for time-varying AR (left) and MA (right) signals using STFT- and multitaper-based estimators. Signals were chosen to illustrate relative rather than absolute performance

to be the  $r^{\text{th}}$  of  $R$  discrete prolate spheroidal sequences [4]. Note that we have assumed that windows do not overlap.

Thus, the two test statistics we consider are given by:

$$\hat{V}(x) \triangleq \frac{1}{ML} \sum_{k=0}^{L-1} \sum_{m=0}^{M-1} \left( \widehat{S}_{xx}^m[k] - \frac{1}{M} \sum_{p=0}^{M-1} \widehat{S}_{xx}^p[k] \right)^2 \quad (6)$$

$$\tilde{V}(x) \triangleq \frac{1}{ML} \sum_{k=0}^{L-1} \sum_{m=0}^{M-1} \left( \widetilde{S}_{xx}^m[k] - \frac{1}{M} \sum_{p=0}^{M-1} \widetilde{S}_{xx}^p[k] \right)^2 \quad (7)$$

The main motivation behind considering two test statistics is that even though both  $\hat{V}(x)$  and  $\tilde{V}(x)$  are asymptotically unbiased [4], the latter may have lower variance since the variance of the multitaper PSD estimator is smaller than that of the periodogram estimator. Specifically, we have that as  $L$  grows,  $\text{Var}(\widehat{S}_{xx}^m[k])$  tends to  $(S_{xx}^m[k])^2$ , while  $\text{Var}(\widetilde{S}_{xx}^m[k])$  tends to  $(S_{xx}^m[k])^2/R$ . Thus, even for a fixed  $R$  and large  $L$ ,  $\text{Var}(\widetilde{S}_{xx}^m[k]) < \text{Var}(\widehat{S}_{xx}^m[k])$ <sup>1</sup>. This latter fact is crucial—lower variance estimators of the PSD imply higher sensitivity of (7) to changes in the signal statistics, and tests based on  $\hat{V}(x)$  can be expected to be more powerful than those based on  $\tilde{V}(x)$ .

To illustrate this point, we compare the *relative* detection performance of the STFT- and multitaper-based test statistics using synthetic time-varying MA(2) and time-varying AR(2) signals—corresponding to smooth and peaky spectra, respectively. The observations, each of length 5120, were obtained by changing the MA or AR coefficients at the midpoint of the signal. In particular, to generate data under  $\mathcal{H}_1$ , the initial MA coefficients of (1, 0.4) were changed to (1, 0.4 +  $\delta$ ) with  $\delta \in (0.1, 0.2)$  and the initial AR coefficients (−1.273, 0.81) were changed to (−1.196, 0.81) corresponding to a  $\pi/50$  Hz shift in the center frequency of the associated bandpass filter. To generate data under  $\mathcal{H}_0$  the initial parameters were simply left unchanged.

The test statistics of (6) and (7) were computed using 10 adjacent 512-sample rectangular windows;  $R = 6$  multitapers were used to obtain  $\tilde{V}(x)$ . Six hundred Monte Carlo simulations were done for these scenarios and the resultant ROC curves are shown in left and right panels of Figure 1 for the AR and MA examples, respectively. The performance gains associated with using the multitaper estimator are evident for this example.

### 3. CONSTRUCTING A CFAR TEST

Next we discuss how to implement the hypothesis test of (1) using a constant false alarm rate threshold, which requires knowledge of

<sup>1</sup>Choosing  $R$  large, however, may lead to greater bias and a weaker test.

$p(\hat{V}(x); \mathcal{H}_0)$  and  $p(\tilde{V}(x); \mathcal{H}_0)$ —the sampling distributions of (6) and (7), respectively, under  $\mathcal{H}_0$ . In this case,  $S_{xx}^m[k]$  is independent of  $m$  and so we may estimate  $S_{xx}$  by:

$$\widehat{S}_{xx}[k] = \frac{1}{M} \sum_m \widehat{S}_{xx}^m[k] \quad \text{or} \quad \widetilde{S}_{xx}[k] = \frac{1}{M} \sum_m \widetilde{S}_{xx}^m[k]. \quad (8)$$

There are other choices for how to estimate  $S_{xx}[k]$  under  $\mathcal{H}_0$  (e.g., median of the short-time spectra), but the estimator in (8) is natural and leads to good performance.

If the sampling distributions of (6) and (7) for the class of WSS signals with power spectra given by (8) are known—indicating how to set a CFAR threshold  $\gamma$ —the null hypothesis is rejected when the test statistic exceeds  $\gamma$ . Therefore, we characterize  $p(\hat{V}(x); \mathcal{H}_0)$  and  $p(\tilde{V}(x); \mathcal{H}_0)$  next.

#### 3.1. Asymptotic Analysis: White Noise Case

We begin by considering  $p(\hat{V}(x); \mathcal{H}_0)$  for the special case when  $x[n]$  is a white Gaussian noise process—not only do these calculations provide some intuition about the hypothesis test, but they shall also reappear in our analysis of the general case in Section 3.2. We begin by defining  $l \triangleq L/2 - 1$  and rewriting  $\hat{V}(x)$  of (6) as:

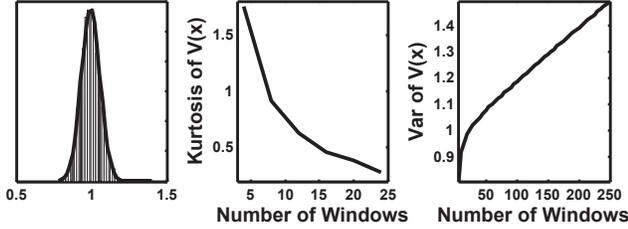
$$\hat{V}(x) = \frac{1}{M^2 l} \sum_{k=0}^l \left[ \sum_{m=0}^{M-1} (M-1) (\widehat{S}_{xx}^m[k])^2 - \sum_{r \neq s}^{M-1} \widehat{S}_{xx}^r[k] \widehat{S}_{xx}^s[k] \right].$$

It is well known that  $\widehat{S}_{xx}^m[k] \sim \frac{1}{2} \chi_2^2$  for  $0 < k \leq l$  and  $\widehat{S}_{xx}^m[0] \sim \chi_1^2$ . Letting  $\mu_{ki}$  denote the  $i^{\text{th}}$  moment of the  $k^{\text{th}}$  PSD bin, from (3.1) it follows that  $\mathbb{E}(\hat{V}(x)) = (M-1)/ML \sum_{k=0}^l (\mu_{k2} - \mu_{k1}^2)$  and:

$$\begin{aligned} \mathbb{E}(\hat{V}(x)^2) &= \frac{1}{(ML)^2} \sum_{k=0}^l \left[ \left( \frac{M-1}{M} \right)^2 [M\mu_{k4} + M(M-1)\mu_{k2}^2] \right. \\ &\quad \left. - 2 \frac{(M-1)^2}{M} [2\mu_{k3}\mu_{k1} + (M-2)\mu_{k2}\mu_{k1}^2] \right. \\ &\quad \left. + \frac{1}{M^2} [2P_{M-2}^M \mu_{k2}^2 + 4P_{M-3}^M \mu_{k2}\mu_{k1}^2 + P_{M-4}^M \mu_{k1}^4] \right] \\ &\quad + \frac{(M-1)^2}{(ML)^2} \sum_k \sum_j [\mu_{k2}\mu_{j2} - 2\mu_{k2}\mu_{j1}^2 + \mu_{k1}^2\mu_{j1}^2], \end{aligned} \quad (9)$$

where  $P_k^m \triangleq m!/(m-k)!$ . A histogram of  $p(\hat{V}(x); \mathcal{H}_0)$  obtained by simulation, overlaid with a Gaussian distribution fitted according to (9), is shown in the left panel of Figure 2 to illustrate this analysis. Even though the distribution of  $\hat{V}(x)$  is not Gaussian, the accuracy of the Normal approximation increases with the number of windows  $M$ . We have observed reasonable results when at least 15 – 20 windows are used in calculating  $\hat{V}(x)$ , as evidenced by the plot of empirical kurtosis as a function of the number of windows in the middle panel of Figure 2.

A key point is that we would ideally like to choose as large a window as possible, while still preserving the sensitivity of the periodogram to the presence of nonstationarity. Note that (9) implies that  $\text{Var}(\hat{V}(x))$  increases linearly in the number of analysis windows used, as confirmed by the plot in the right panel of Figure 2. Thus,  $L$  decreases as  $M$  increases and we conclude that using short analysis windows may increase the overall mean-square error of our spectral estimates and thereby decrease the power of the hypothesis test.



**Fig. 2.** Understanding  $p(\hat{V}(x); \mathcal{H}_0)$  when  $x[n]$  is WGN. Left: empirical (grey) and Gaussian (black) approximations of  $p(\hat{V}(x); \mathcal{H}_0)$  with  $M = 20$ . The kurtosis of  $V(x)$  under  $\mathcal{H}_0$  decreases (middle) and its variance grows linearly with  $M$  (right)

### 3.2. Wold Representation: General Case

In the general case, when the spectrum (8) is not white or is estimated using overlapping windows, deriving  $p(\hat{V}(x); \mathcal{H}_0)$  directly is either impossible or extremely tedious. Further, these calculations would be applicable only when  $M$  is large (e.g., middle panel of Figure (2)). Instead, we proceed via simulation, by leveraging an innovations representation of stationary processes. The Wold representation of a stationary process  $x[n]$  is given by:

$$x[n] = \sum_{l=0}^{\infty} h[l]\epsilon[n-l], \quad (10)$$

where  $\mathbb{E}(\epsilon[n]\epsilon[m]) = \delta[n-m]$  and  $h[n]$  is the impulse response of a stable and minimum-phase filter. This allows us to express the PSD of  $x[n]$  as follows:

$$S_{xx}[k] = |H[k]|^2 S_{\epsilon\epsilon}[k]. \quad (11)$$

Since  $S_{\epsilon\epsilon}[k] = 1$ , to estimate  $S_{xx}[k]$  and  $|H[k]|^2$  is equivalent. Thus, assuming that an observed signal  $x[n]$  has been partitioned according to the same tiling of the time-frequency plane that was used to define the test statistic in (6), we have that  $|\hat{H}[k]|^2 = \hat{S}_{xx}[k]$  as defined by (8).

The key idea behind our simulation approach is to use (11) in order to generate realizations of  $S_{xx}^m[k]$ , under the null, by multiplying  $|\hat{H}[k]|^2$  with different realizations of white noise spectra. In turn, this enables the computation of (6) for every realization so achieved. Specifically, let the  $i^{\text{th}}$  realization of a white noise PSD be denoted by  $\hat{S}_{\epsilon\epsilon}^{m,i}[k]$ , then the  $i^{\text{th}}$  realization of a WSS signal in the  $m^{\text{th}}$  window is given by:

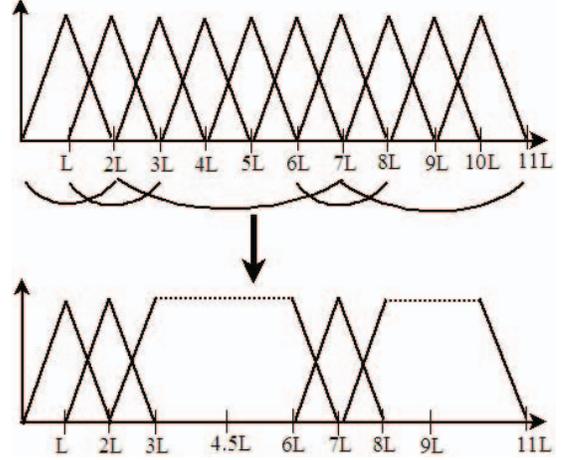
$$\widehat{S_{xx}^{m,i}}[k] = |\hat{H}[k]|^2 \widehat{S_{\epsilon\epsilon}^{m,i}}[k]. \quad (12)$$

The  $i^{\text{th}}$  instantiation of (6) under  $\mathcal{H}_0$  is, therefore, given by:

$$\begin{aligned} \hat{V}^i(x) &= \frac{1}{ML} \sum_{k=1}^L \sum_{m=1}^M \left( \widehat{S_{xx}^{m,i}}[k] - \frac{1}{M} \sum_{p=1}^M \widehat{S_{xx}^{p,i}}[k] \right)^2 \\ &= \frac{1}{ML} \sum_{k=1}^L |\hat{H}[k]|^4 \sum_{m=1}^M \left( \widehat{S_{\epsilon\epsilon}^{m,i}}[k] - \frac{1}{M} \sum_{p=1}^M \widehat{S_{\epsilon\epsilon}^{p,i}}[k] \right)^2. \end{aligned} \quad (13)$$

Similarly, we may obtain replicates of  $\tilde{V}^i(x)$  and characterize  $p(\tilde{V}(x); \mathcal{H}_0)$  by substituting multitaper estimates of the relevant PSDs into (12) and (13). In both cases, we may build up an empirical CDF of (6) or (7) and reject  $\mathcal{H}_0$  if the observed statistic lies in the tail beyond a specified false alarm threshold  $\gamma$ .

It is crucial to observe that since  $\hat{H}[k]$  is fixed, the only source of randomness in (13) is due to variance of the estimator for the PSD



**Fig. 3.** Example of how a fixed-resolution scheme using windows of length  $L$  (top) is modified to achieve an adaptive-resolution scheme by merging neighboring windows (bottom). See [5] for details.

of white noise. Therefore, in practice, once a windowing scheme has been chosen, only the distribution of  $\hat{S}_{\epsilon\epsilon}^{m,i}[k]$  has to be determined empirically (or via the Gaussian approximation of Section (3.1)!) for each  $k$ . Then  $p(\hat{V}(x); \mathcal{H}_0)$  is readily obtained by using  $\hat{H}[k]$  together with (13)—a powerful construct since all Monte-Carlo simulations may be done *offline*.

## 4. EXAMPLE APPLICATION

As an example of the applicability of the proposed testing framework to actual time series data, we evaluate its performance in the context of signal enhancement. Audio noise reduction is generally achieved through the attenuation of spectral coefficients using uniformly sized windows and local Fourier analysis. Instead, we propose an adaptive enhancement scheme in which the adaptive analysis scheme of [5] is modulated by the hypothesis test of Section 2.

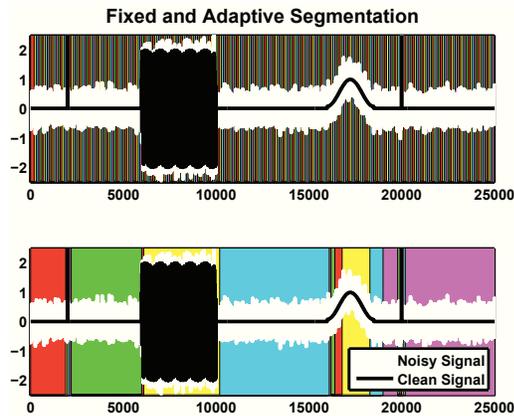
### 4.1. Enhancement System

Figure 3 illustrates the adaptive STFT scheme of [5], in which consecutive short-time analysis windows are to be merged based on a measure of time-frequency concentration. Here, we evaluate the performance of the proposed hypothesis test of (1) by using it in place of time-frequency concentration as the decision device in this adaptive scheme; the reader is referred to [5] for more details.

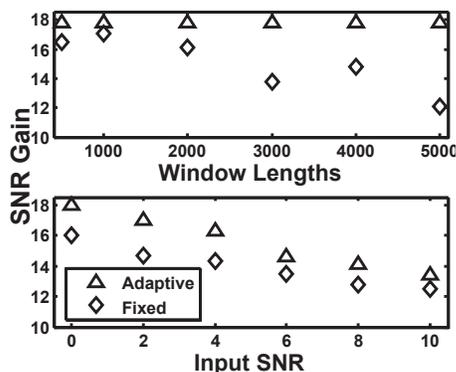
Assuming observations  $y[n]$  degraded by additive white Gaussian noise, the resulting spectral slice  $Y_m[k]$  corresponding to the  $m^{\text{th}}$  window is attenuated according to the standard Wiener suppression scheme:

$$\hat{X}_m[k] = \frac{S_{xx}^m[k]}{S_{xx}^m[k] + \sigma^2} Y_m[k]. \quad (14)$$

As we are interested in illustrating the performance of our nonparametric test for stationarity, rather than evaluating competing spectral estimators, we use the periodogram of the clean signal  $|X_m[k]|^2$  (instead of the usual  $|Y_m[k]|^2$ ) as the “oracle” estimate of  $S_{xx}^m[k]$ . Waveform reconstruction is achieved by overlap-add synthesis described in [5], as the adaptive analysis scheme still obeys the requisite partition of unity property.



**Fig. 4.** Fixed (top) and adaptive (bottom) segmentations of a test signal from [5], degraded with white Gaussian noise to yield an SNR of 5 dB. Boundaries of colored rectangles demarcate regions of stationarity



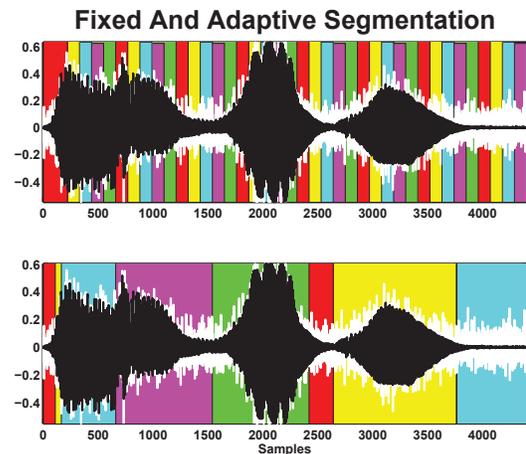
**Fig. 5.** Performance of the adaptive and fixed-rate systems for enhancement of the synthetic signal of Fig. 4. Above, SNR gain is shown as the lengths of the fixed-rate windows are varied; below, gain is shown as a function of input SNR for fixed window lengths

#### 4.2. Enhancement Example Results

We now apply the enhancement scheme to a synthetic test signal from [5], as well as to a short segment of clarinet music. Both signals were corrupted by additive white Gaussian noise to yield a signal-to-noise (SNR) ratio of 5 dB. In both examples, the fixed-resolution scheme employed 200-sample triangular windows with 50% overlap, as in Figure 3. To decide if two adjacent windows should be merged, each individual window is further subdivided into 4 parts and the test statistic of (6) is computed across all six windows tiling the two neighboring segments (the windows used to conduct the hypothesis are distinct from those used in the adaptive scheme). Using the techniques of Section 3.2 a 10% CFAR threshold was found and used in each decision of the adaptation scheme. Note that (13) allows us to approximate  $p(\hat{V}(x); \mathcal{H}_0)$  even if windows are overlapping.

The adaptation scheme is applied to one instance of the synthetic signal and the resultant segmentation is shown in Figure 4. The hypothesis test aids in identifying stationary regions which, in turn, leads to improved enhancement performance (leading to an additional 0.5–2 dB SNR gain) relative to the fixed-resolution scheme as shown in Figure 5.

Adaptive segmentation of the clarinet recording, shown in Fig-



**Fig. 6.** Fixed (top) and adaptive (bottom) segmentations of clarinet recording, degraded with white Gaussian noise to yield an SNR of 5 dB

ure 6, accurately captures dominant signal features even in the presence of severe noise. Here the adaptive enhancement scheme provides an additional 1.2 dB SNR gain over the fixed-resolution approach. Moreover, informal listening tests have indicated a significant reduction in musical noise—in agreement with results obtained for speech in [5].

#### 5. DISCUSSION

Here we have presented a nonparametric hypothesis test for stationarity. To this end, we proposed two test statistics based on the periodogram and multitaper estimators of the power spectral density and studied their sampling distribution under the null hypothesis using an efficient simulation scheme based on the Wold representation. We have evaluated our scheme in the context of enhancement for synthetic signals and explored its applicability for musical signals with promising results. Future work will consider testing procedures for more restricted classes of alternates.

**Acknowledgements:** Authors were supported in part by the Defense Advanced Research Projects Agency under Grant No. HR0011-07-1-0007. The first author was supported by the National Geospatial-Intelligence Agency Grant HM1582-08-1-0023. The second author was supported by the National Science Foundation Graduate Fellowship. We also thank the reviewers for their helpful comments.

#### 6. REFERENCES

- [1] S. M. Kay, “A new nonstationarity detector,” *IEEE Transactions on Signal Processing*, vol. 56, pp. 1440–1451, 2008.
- [2] A. Cardinali and G. P. Nason, “Stationarity and costationarity tests for stock index returns,” *Statistics Group, University of Bristol, UK, Technical Report 08:08*, 2007.
- [3] D. L. Donoho, S. Mallat, and R. von Sachs, “Estimating covariances of locally stationary processes: consistency of best basis methods,” *Proceedings of the IEEE International Symposium on Time-Frequency and Time-Scale Analysis*, pp. 337–340, 1996.
- [4] D. J. Thomson, “Spectrum estimation and harmonic analysis,” *Proceedings of the IEEE*, vol. 70, pp. 1055–1096, 1982.
- [5] D. Rudoy, P. Basu, T. F. Quatieri, B. Dunn, and P. J. Wolfe, “Adaptive short-time analysis-synthesis for speech enhancement,” *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. 4905–4908, 2008.