IEEE Int. Symp. Time-Freq. Time-Scale Analysis, Paris, France, pp. 449-452, 1996

TIME-CORRELATION ANALYSIS OF NONSTATIONARY SIGNALS WITH APPLICATION TO SPEECH PROCESSING

Ta-Hsin Li and *Jerry D. Gibson*

Texas A&M University, College Station, Texas 77843, USA

Abstract

This paper proposes a new method of displaying and analyzing evolutionary correlation structure of nonstationary signals. The method, called time-correlation analysis (TCA), is based on a filter-bank approach for stochastic signal characterization known as parametric filtering. Some properties of the TCA method are discussed that can be used to interpret the TCA plot. Examples of an application to speech analysis are given.

1. Introduction

Characterization of correlation structure is often the first step in second-order analysis of stochastic signals. There are two traditional ways of characterizing the correlation structure of a stationary signal: one is to use the autocorrelation function (ACF) in the time domain and the other is to use the spectral density function (SDF) in the frequency domain; both functions have distinct features that compensate for each other — the ACF describes linear relations among consecutive random samples in the time domain, whereas the SDF depicts the power distribution over different frequencies.

With a sliding analysis window for time-domain localization, the SDF has been successfully extended to nonstationary random signals, resulting in a three-dimensional time-frequency-intensity portrait known as the spectrogram. The ACF, however, is less successful when extended to nonstationary signals, due partly to the lack of effective visualization methods — sometimes a movie has to be used in order to effectively display time-varying sequences of ACF's [1].

This paper presents a new graphical method, called *time-correlation analysis* (TCA), for displaying the evolutionary correlation structure of nonstationary random signals [2]. The TCA method is based on a recently proposed approach of time series characterization, known as *parametric filtering* [3]. The gist of the parametric filtering approach can be summarized as follows: (i) take a parametric filter (a filter bank) $\mathcal{H}_\beta(z)$ indexed by a parameter β (discrete or continuous valued); (ii) filter the (zero-mean) signal X_t to obtain $X_t(\beta) = \mathcal{H}_\beta(z)\, X_t$; (iii) use an output statistic, such as the lag-one autocorrelation $\rho(\beta)$, of $X_t(\beta)$ to characterize the signal X_t. The filter bank is so designed that $\rho(\beta)$ uniquely determines the correlation structure of X_t. Classical examples of such filter banks are repeated summing and differencing [3], [4]. In this paper, we employ a more flexible parametric filter, called a complex exponential filter, and investigate some aspects of the resulting characterization function $\rho(\beta)$ when used to display time-varying correlation of nonstationary signals.

2. Time-Correlation Analysis

To obtain a TCA portrait of evolutionary correlation structure, the parametric filtering approach first passes a windowed signal (localized at time t) through the complex exponential filter

$$\mathcal{H}_\alpha(z) = \frac{1}{1-\alpha^* z^{-1}}\,,$$

where $\alpha = \eta \exp(-j\theta)$; then it calculates the lag-one autocorrelation, $\rho_t(\alpha)$, of the filtered signal; the new characterization function, known as the *demodulated lag-one autocorrelation*, is defined by [3]

$$\gamma_{\theta,t}(\eta) = \Re\{e^{-j\theta}\rho_t(\alpha)\} \quad (-1 < \eta < 1).$$

With η_k, $(k = 1, \cdots, m)$, uniformly spaced in $(-1, 1)$ and θ fixed in $[0, \pi]$, the TCA method plots $\gamma_{\theta,t}(\eta_k)$ as time series. An example of the TCA plot is shown in Fig. 1, where the signal is a waveform of the word "fish" spoken by a male speaker; the signal has added white Gaussian noise so that SNR = 25 dB [5].

For a stationary signal with ACF

$$R(\tau) = \frac{E\{X_{t+\tau}X_t\}}{E\{X_t^2\}},$$

some results have been obtained in [3] that justify the use of $\gamma_\theta(\eta) = \gamma_{\theta,t}(\eta)$ as characterization functions. In particular, it may be shown that the function $\gamma_\theta(\eta)$

T. H. Li is with the Department of Statistics.

J. D. Gibson is with the Department of Electrical Engineering and partially supported by NSF Grant NCR-9303805.

uniquely determines $R(\tau)$ for almost every θ. In other words, information about the correlation structure of any stationary signal is completely preserved in $\gamma_\theta(\eta)$ as a function of η. Similar properties still hold approximately for nonstationary signals under the assumption that the correlation structures evolve slowly relative to the length of the analysis window.

Unlike the spectrogram, the TCA plot describes nonstationary signals in terms of *correlation* (hence the name of the method) rather than power spectral density. However, with the frequency parameter θ controlling the center frequency of the filter $\mathcal{H}_\alpha(z)$, the TCA method also possesses frequency selectability, which is desirable from the viewpoint of (Fourier) spectral analysis. As such, the TCA method may be regarded as one that provides a new domain of signal characterization, sitting between the time and the frequency domains.

Also, unlike the spectrogram, which primarily emphasizes discrete spectral components (though desirable in some applications), the TCA plot combines both discrete and continuous spectral components of the signal. The combined information is integrated into the lag-one autocorrelation which can be easily interpreted as the *center of spectral mass* of the filtered signal.

3. Some Graphical Properties of TCA

Given a stationary signal with ACF $R(\tau)$, it may be shown [3] that, for any fixed θ, the mapping

$$R(\tau) \mapsto \gamma_\theta(\eta)$$

transforms the ACF into a monotone and smooth function of $\eta \in (-1, 1)$. This monotonicity makes the TCA plot easy to display – in fact, there is no need to introduce a third dimension, such as brightness or color, for displaying the TCA plots, because the trajectories of $\gamma_{\theta,t}(\eta)$ as time series almost never cross each other, as we can see from Fig. 1.

If X_t is white noise, then we have

$$\gamma_\theta(\eta) = \eta$$

for any θ. This property can be used as a benchmark in some applications for time series discrimination and change detection [3]. For the example in Fig. 1, the property is manifested as nearly uniformly spaced lines at the beginning of the TCA plot (about $t = 1$–350), corresponding to the background white noise before the word "fish" starts [5].

For the vowel sound /i/, the TCA lines cluster near $+1$, as can be seen in Fig. 1 for $t = 750$–1700. The *dominant frequency principle* is primarily responsible for this behavior. In fact, if the signal is a pure tone with $X_t = A\cos(\omega_0 t + \phi)$ for some $\omega_0 \in (0, \pi)$, then, it may be shown that

$$\begin{aligned} \gamma_\theta(+1^-) &= w_- \cos(\omega_0 - \theta) + (1 - w_-)\cos(\omega_0 + \theta), \\ \gamma_\theta(-1^+) &= w_+ \cos(\omega_0 - \theta) + (1 - w_+)\cos(\omega_0 + \theta), \end{aligned}$$

where $w_\pm \in (0, 1)$ are certain weights. Using the Taylor series expansion, one further obtains

$$\begin{aligned} |\gamma_\theta(+1^-) - \cos(\omega_0)| &\leq \theta, \\ |\gamma_\theta(-1^+) - \cos(\omega_0)| &\leq \theta. \end{aligned}$$

Because $\gamma_\theta(\eta)$ is monotone, we have

$$|\gamma_\theta(\eta) - \cos(\omega_0)| \leq \theta$$

for all $\eta \in (-1, 1)$. This implies that when θ is small the TCA lines will cluster near the value of $\cos(\omega_0)$. In Fig. 1, we took $\theta = 0.05\pi$. Thus the clustering of the TCA lines near $+1$ for $t = 750$–1700 indicates the presence of dominant low frequency components.

To further investigate the *range* of TCA lines, consider the case where X_t is a band-limited stationary signal with (normalized) SDF

$$f(\omega) = \sum_{\tau=-\infty}^{\infty} R(\tau)\, e^{-j\tau\omega}$$

and spectral support $\Omega \subseteq (-\pi, \pi]$. Assuming $f(\omega)$ is sufficiently smooth in Ω, it may be shown [2] that

$$\gamma_\theta(+1^-) = \begin{cases} 1 & \text{if } \theta \in \Omega, \\ 1 - a_\theta^{-1} & \text{if } \theta \notin \Omega, \end{cases}$$

where

$$a_\theta = \frac{1}{2\pi} \int_\Omega \frac{f(\omega)}{1 - \cos(\omega - \theta)}\, d\omega.$$

Similarly, we obtain [2]

$$\gamma_\theta(-1^+) = \begin{cases} -1 & \text{if } \theta' \in \Omega, \\ -1 + a_{\theta'}^{-1} & \text{if } \theta' \notin \Omega, \end{cases}$$

where $\theta' = \theta + \pi$ if $\theta \in (-\pi, 0]$ and $\theta' = \theta - \pi$ if $\theta \in (0, \pi]$. This implies that if both θ and θ' are inside the spectral support of X_t, then the TCA lines cover the entire interval $(-1, 1)$. The segment of background white noise in Fig. 1 ($t = 1$–350) exhibits this behavior. If θ is inside but θ' is outside the spectral support, then the TCA lines spread up to $+1$ but do not reach down to -1, indicating a lack of high frequency components in the signal when the TCA plot is constructed with $\theta \in (-\pi/2, \pi/2]$. An example of this case can be

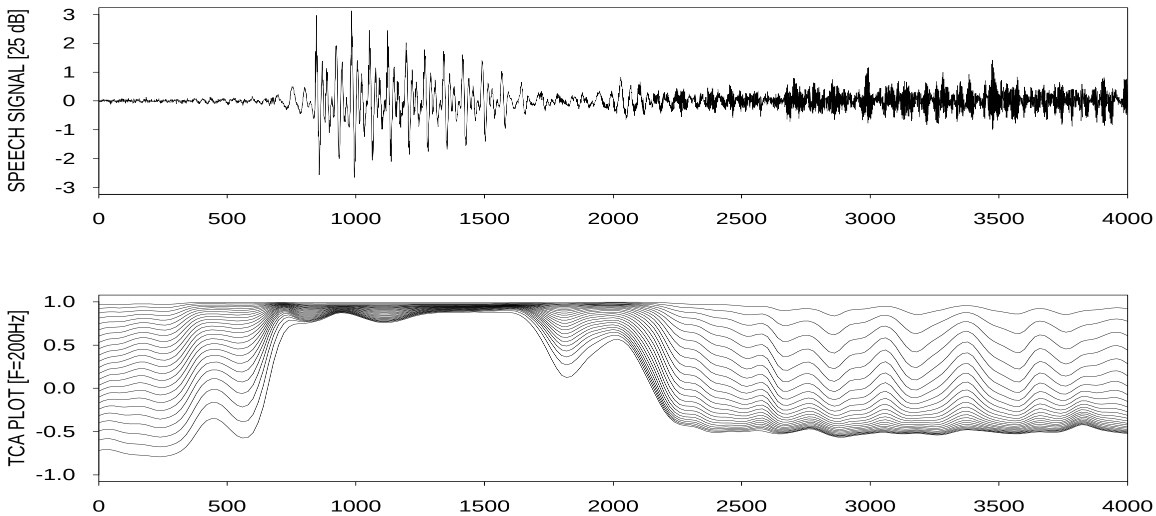

Fig. 1. A nonstationary speech waveform (top) and its TCA plot.

found in Fig. 1 for the segment $t = 1500$–2150, which corresponds to the transition period from /i/ to /sh/.

In addition to the range of TCA lines, the *density* of TCA lines also carries information about the signal. For example, in Fig. 1, the TCA lines are denser on the top for the segment /f/ ($t = 350$–750), but denser on the bottom for the segment /sh/ ($t = 2150$–4000). It is clear that, as η varies, dense TCA lines indicate slow changes in $\gamma_\theta(\eta)$ and coarse TCA lines reflect fast changes in $\gamma_\theta(\eta)$. The first derivative of $\gamma_\theta(\eta)$ measures the rate of change and thus is suitable for describing the density of TCA lines.

If the SDF $f(\omega)$ is smooth and positive at θ and θ', it may be shown [3] that

$$\lim_{\eta \to 1^-} \frac{d\gamma_\theta(\eta)}{d\eta} = \frac{1}{f(\theta)},$$
$$\lim_{\eta \to -1^+} \frac{d\gamma_\theta(\eta)}{d\eta} = \frac{1}{f(\theta')}.$$

This implies that when the signal has high (relative) energy concentration near θ so that $f(\theta)$ is large, the function $\gamma_\theta(\eta)$ changes slowly as η varies near $+1$, giving rise to dense lines on the top of TCA plot. In Fig. 1, this corresponds to the segment /f/ ($t = 350$–750), which is known to contain strong low frequency components [6]. On the other hand, when the signal has low energy concentration near θ, the function $\gamma_\theta(\eta)$ changes rapidly as η moves away from $+1$ and thus yields coarse TCA lines on the top. An example of this case is given by the segment /sh/ ($t = 2150$–4000), in which high frequency components are known to dominate the SDF [6].

More graphical properties of the TCA plot can be found in [2].

4. Change Detection Based on TCA

The TCA plot also provides a basis from which features can be selected for detecting significant correlation changes in nonstationary signals. For example, to measure the changes in correlation structure, one may use TCA-based *distortion measures* such as [5] [7]

$$\hat{\kappa}_t = \int \left\{ K\left(\frac{p_{\theta,t-s/2}(\eta)}{p_{\theta,t+s/2}(\eta)}\right) + K\left(\frac{p_{\theta,t+s/2}(\eta)}{p_{\theta,t-s/2}(\eta)}\right) \right\} d\mu,$$

where $K(r) = r - \log(r) - 1$ is the Kullback-Leibler kernel, $\mu = \mu(\theta, \eta)$ is a certain measure, and $p_{\theta,t}(\eta)$ is a function defined by

$$p_{\theta,t}(\eta) = \frac{1}{2}\left\{ \frac{d\gamma_{\theta,t}(\eta)}{d\eta} + (1 + \gamma_{\theta,t}(\eta_a))\,\delta(\eta - \eta_a) + (1 - \gamma_{\theta,t}(\eta_b))\,\delta(\eta - \eta_b) \right\}$$

for $\eta \in [\eta_a, \eta_b] \subset (-1, 1)$. The integral is defined on $(\theta, \eta) \in \theta \times [\eta_a, \eta_b]$ for some $\theta \subseteq (-\pi, \pi]$. Note that $p_{\theta,t}(\eta)$ combines the range and density information revealed in the TCA plot. For any given θ, the function $p_{\theta,t}(\eta)$ is equivalent to the function $\gamma_{\theta,t}(\eta)$ and hence preserves the local correlation structure of the signal; it is also a "probability" density function because $p_{\theta,t}(\eta) > 0$ and $\int_{\eta_a}^{\eta_b} p_{\theta,t}(\eta)\,\eta = 1$.

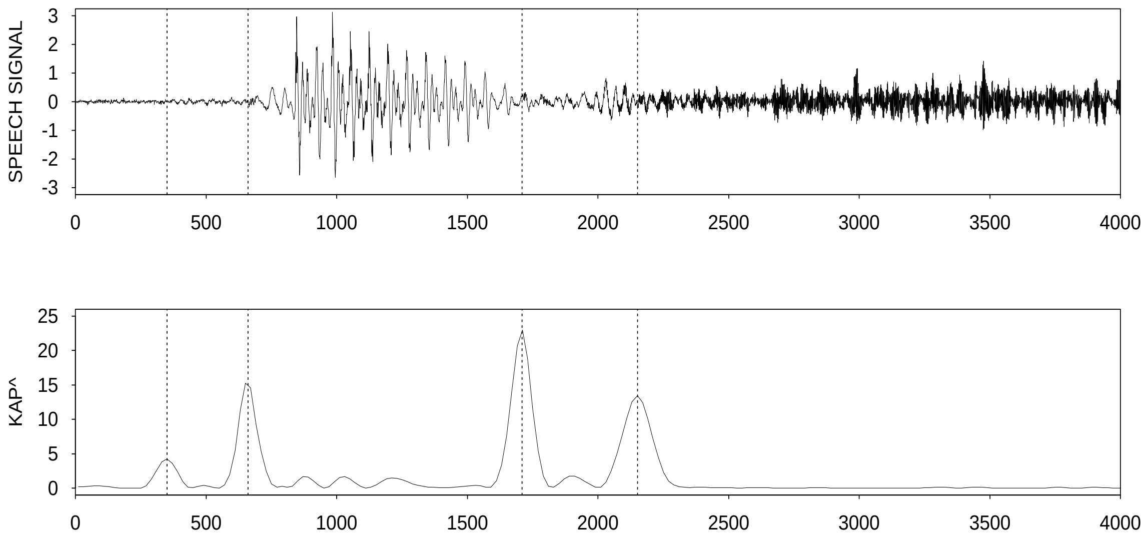

Fig. 2. The same speech waveform as in Fig. 1 and the trajectory of distortion measure $\hat{\kappa}_t$.

Using the TCA-based distortion measure $\hat{\kappa}_t$, an example of change detection is shown in Fig. 2 [5]. Note that sharp peaks in the trajectory of $\hat{\kappa}_t$ correspond to the significant correlation changes revealed by the TCA plot in Fig. 1. These changes can be easily detected using a peak-picking method. The vertical lines in Fig. 2 indicate the detected changes.

Because of the filtering and frequency selectability inherent in the TCA method, the TCA-based distortion measures can be made more robust than the traditional spectral density based distortion measures such as the Kullback-Leibler spectral divergence. The robustness is especially effective when the signal suffers from narrow-band contaminations such as spurious spectral peaks and notches. Details on this matter can be found in [7].

5. Concluding Remarks

In this paper, we proposed a new method for displaying and analyzing evolutionary correlation structure. The TCA method employs a filter-bank-based approach of signal characterization called parametric filtering. Graphical properties of TCA were investigated for the interpretation of TCA plots. The range and density features of TCA lines were considered in particular. These features were found to be related to the traditional spectral density in some interpretable ways which explain the behavior of TCA plots in some speech processing applications. The analysis also led to a distortion measure which is capable of detecting significant correlation changes depicted by TCA. The TCA method is proposed not to replace the traditional spectrogram or related time-frequency time-scale plots, but to serve as a complementary graphical tool that may suggest additional diagnostic features useful in the analysis of time-varying correlation structure. Future research should be devoted to the exploration of other filter banks which may be more flexible and versatile for time-correlation analysis.

References

[1] M. Slaney and R. F. Lyon, "On the importance of time – a temporal representation of sound," in *Visual Representation of Speech Signals*, M. Cooke, S. Beet, and M. Crawford, eds., pp. 95–116, New York, John Wiley, 1993.

[2] T. H. Li, "Time-correlation analysis of nonstationary time series," Tech. Rept. 236, Statistics Dept., Texas A&M Univ., College Station, 1995.

[3] T. H. Li, "Discrimination of time series by parametric filtering," *J. Amer. Statist. Assoc.*, vol. 91, pp. 284–293, March 1996.

[4] B. Kedem, *Time Series Analysis by Higher Order Crossings*, Piscataway, NJ: IEEE Press, 1994.

[5] T. H. Li and J. D. Gibson, "Speech analysis and segmentation by parametric filtering," *IEEE Trans. Audio, Speech Processing*, vol. 4, no. 3, pp. 203–213, May 1996.

[6] J. R. Deller, Jr., J. G. Proakis, and J. H. L. Hansen, *Discrete Time Processing of Speech Signals*, New York: Macmillan, 1993.

[7] T. H. Li, "Robust divergence measures for time series discrimination," Tech. Rept. 237, Statistics Dept., Texas A&M Univ., College Station, 1995.