New Distortion Measures for Speech Processing

TA-HSIN LI* and JERRY D. GIBSON†
Texas A&M University, Collage Station, TX 77843

Abstract – **New distortion measures are derived from a recently proposed characterization function of stationary time series and are shown to be more robust than some commonly-used distortion measures such as the Kullback-Leibler spectral divergence in speech processing.**

I. INTRODUCTION

Distortion measures are widely used in speech processing to quantify the deviations of speech signals in correlation structure, and among the most successful ones is the Itakura-Saito (IS) distance of spectral densities [2], also known as the Kullback-Leibler information divergence [3]. Although in many cases the IS distance is quite effective in discriminating signals and detecting special changes, its lack of robustness is also well known documented in the literature, especially when the signals are mixtures of narrow and wide band components such as voiced speech waveforms (e.g., [1]). On the basis of a method called *parametric filtering*, we propose some new distortion measures that are shown to be more robust than the IS distance.

II. NEW DISTORTION MEASURES

Given a zero-mean stationary signal $\{X_t\}$, the parametric filtering method characterizes the correlation structure of $\{X_t\}$ by the demodulated first-order autocorrelation of the form

$$\gamma_\theta(\eta) := \Re\{e^{-i\theta}\rho(\alpha)\} \quad (-1 < \eta < 1),$$

where $\rho(\alpha)$ is the first-order autocorrelation of the filtered signal $X_t(\alpha) := \overline{\alpha}\, X_{t-1}(\alpha) + X_t$ with $\alpha := \eta e^{-i\theta}$. Among other interesting properties of $\gamma_\theta(\eta)$, it can be shown [4], [5] that $\gamma_\theta(\eta)$ uniquely determines the correlation structure of $\{X_t\}$ for almost any θ and is infinitely differentiable in $\eta \in (-1, 1)$ even for mixed-spectrum signals of which the spectral density does not exist. Using these properties, we define

$$\begin{aligned} p_\theta(\eta) \;&:=\; \tfrac{1}{2}\left[\gamma'_\theta(\eta) + (\gamma_\theta(\eta_a) + 1)\,\delta(\eta - \eta_a)\right. \\ &\quad \left. + (1 - \gamma_\theta(\eta_b))\,\delta(\eta - \eta_b)\right], \end{aligned}$$

for any $-1 < \eta_a < \eta_b < 1$, where $\delta(\eta)$ is the Dirac delta. Clearly, the function $p_\theta(\eta)$ forms a (generalized) probability density in $[\eta_a, \eta_b]$ and, because of its equivalence to $\gamma_\theta(\eta)$, uniquely determines the correlation structure of $\{X_t\}$ for almost any θ. Therefore, we can define the following distortion measure using the Kullback-Leibler information divergence [3], namely

$$\begin{aligned} \kappa(p^0_\theta \| p^1_\theta) \;&:=\; \int_{\eta_a}^{\eta_b} p^0_\theta(\eta)\, K\big(p^1_\theta(\eta)/p^0_\theta(\eta)\big)\, d\eta, \\ \kappa(p^0_\theta; p^1_\theta) \;&:=\; \int_{\eta_a}^{\eta_b} K\big(p^0_\theta(\eta)/p^1_\theta(\eta)\big)\, d\eta, \end{aligned}$$

where $K(u) := u - \log u - 1$. Many other distortion measures can be defined, for instance, from the family of Renyi's information [6].

The IS spectral distance is known to be extremely sensitive to deviations of individual spectral peaks while less so to changes of overall spectral shapes (or envelopes). The new measures $\kappa(p^0_\theta \| p^1_\theta)$ and $\kappa(p^0_\theta; p^1_\theta)$ are potentially more robust than the IS distance because they are *finite* even when the spectral support changes. With this property, the new measures are able to avoid the disproportional sensitivity to frequency shifts and spectral peaks, and thus to discriminate correlation structures by treating the discrete and continuous components on an equal basis.

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

†J. D. Gibson is with the Department of Electrical Engineering. He is supported by NSF grant NCR-93-03805.

REFERENCES

[1] R. Andre-Obrecht, "A new statistical approach for the automatic segmentation of continuous speech signals," *IEEE Trans. ASSP*, vol. 36, pp. 29–40, 1988.

[2] F. Itakura and S. Saito, "A statistical method for estimation of speech spectral density and format frequencies," *Electron. Commun. Japan*, vol. 53-A, pp. 36–43, 1970.

[3] S. Kullback, *Information Theory and Statistics*, New York: Dover, 1968.

[4] T. H. Li, "Discrimination of time series by parametric filtering," Tech. Rep. 212, Dept. of Statistics, Texas A&M Univ., College Station, 1994.

[5] T. H. Li and J. D. Gibson, "Discriminant analysis of speech by parametric filtering," *Proc. 28th Conf. Inform. Sci. Syst.*, 1994.

[6] E. Parzen, "Time series, statistics, and information," in *New Directions in Time Series Analysis, Pt. I*, D. Brillinger *et al.* Eds., New York: Springer; pp. 265–286, 1992.