Benefits of Noncircular Statistics for Nonstationary Signals

Scott Wisdom[1], Les Atlas[1], James Pitton[1,2], and Greg Okopal[2]
[1]Dept. of Electrical Engineering, University of Washington
[2]Applied Physics Laboratory, University of Washington
{swisdom,atlas}@uw.edu, {pitton,okopal}@apl.uw.edu

***Abstract*—Conventional statistical signal processing of nonstationary signals uses circular complex Gaussian distributions to model the complex-valued short-time Fourier transform. In this paper, we show how noncircular complex Gaussian distributions can provide better statistical models of a variety of nonstationary acoustic signals. The estimators required for this model are computationally efficient, and also have a simple approximate finite-sample distribution. We also show that noncircular Gaussian models provide distinct benefits for statistical signal processing. In particular, we show how noncircular Gaussian models can improve detection of nonstationary acoustic events, and we explore how estimator parameter choices affect performance.**

I. INTRODUCTION

Many important natural physical signals in engineering are nonstationary and real-valued, especially acoustic signals. Such signals are often processed in the complex-valued frequency domain using Gaussian models, which only require second-order statistics. However, this conventional processing makes a fundamental assumption that the phase of these complex representations is uniform and thus uninformative. Such an assumption implies that the probability density functions (pdfs) of the observed random variables are *circular*, or invariant to rotation in the complex plane.

In this paper, we relax this conventional circular assumption, and demonstrate that using *noncircular* Gaussian distributions that are not phase-invariant can provide distinct benefits. In particular, we show using a rigorous hypothesis test that the short-time Fourier transform (STFT) of nonstationary signals exhibits significant noncircularity. Furthermore, we show that using noncircular Gaussian distributions can improve detection of nonstationary signals in the presence of stationary noise. Noncircular models are very simple, requiring only an additional second-order statistic, the complementary variance, in addition to the usual Hermitian variance.

This paper is organized as follows. First, we review prior work. Then we review the second-order statistics of noncircular complex-valued random variables and provide a hypothesis test for the connection between nonstationarity and noncircularity in the frequency domain. Finally, as our main contribution, we demonstrate the benefit of noncircular models on a realistic nonstationary acoustic event detection dataset, and explore the effect of estimator parameter choice on detection performance.

This work is funded by U.S. Office of Naval Research contract N00014-12-G-0078, delivery order 0013, and U.S. Army Research Office grant number W911NF-15-1-0450.

II. RELATION TO PRIOR WORK

Schreier and Scharf [1] were among the first to recognize that many real-valued nonstationary random signals exhibit second-order noncircularity (or "impropriety") in their analytic signals. Furthermore, they also examined the bifrequency Loève spectrum of such signals. Atlas [2] considered bifrequency spectral correlations and their application to separation and modification of speech from the perspective of modulation frequency. Douglas and Mandic [3] proposed the "panorama," which they define as the Fourier transform of a zero-mean signal's autoconvolution estimated using ensemble averaging. In concert with the conventional power spectrum–the Fourier transform of a zero-mean signal's autocorrelation–they proposed a detector for deterministic sinusoidal components in the presence of WSS noise. We go beyond previous work by using estimators of spectral noncircularity that do not require ensemble averaging.

Rivet et al. [4] observed spectral noncircularity in complex-valued speech spectra and explored the effect of noncircularity on the log magnitude spectra. Millioz and Martin [5] studied the circularity of the short-time Fourier transform (STFT) coefficients of nonstationary signals in white Gaussian noise, and used noncircularity to segment synthetic and natural signals in time-frequency. Clark [6] observed that speech signals exhibit spectral noncircularity. Spectral noncircularity is also related to the modulation frequency content of signals [7], [8]. Wisdom et al. [9] and Okopal et al. [10], [11] empirically showed that spectral noncircularity is useful for detection of speech signals and can be used for blind separation of multichannel mixtures of speech and noise. This paper goes beyond our recent work [12] on the connection between spectral noncircularity and nonstationary signals by applying approximate finite-sample distributions for noncircular statistics, performing acoustic event detection on a new, larger dataset, and exploring the effect of parameter choice on detection performance.

III. NONCIRCULAR GAUSSIAN RANDOM DATA

A complex-valued Gaussian random variable x requires not one, but two second-order moments to fully characterize its statistical behavior [13]. These two second-order moments are the Hermitian variance

$$R_x := E\{xx^*\} = E\{|x|^2\}, \tag{1}$$

and the complementary variance

$$\widetilde{R}_x := E\{xx\} = E\{x^2\}. \tag{2}$$

If $\widetilde{R}_x = 0$, then x is said to be *second-order circular*[1], or *proper*.

The degree of noncircularity of x is described by the circularity coefficient, defined as the ratio of the magnitude of complementary variance to Hermitian variance:

$$k = \frac{\left|\widetilde{R}_x\right|}{R_x}. \tag{3}$$

The circularity coefficient ranges from 0, which means x is circular, to 1, which means x is maximally noncircular, or *rectilinear*. A rectilinear complex random variable appears as a line in the complex plane.

The pdf of M i.i.d. samples $x_{0:M-1}$ of a noncircular Gaussian random variable x is given by

$$p(x_{0:M-1}; R_x, \widetilde{R}_x) = \frac{\exp\left(\frac{-\sum_{m=0}^{M-1}|x_m|^2 + \mathrm{Re}\left\{\frac{\widetilde{R}_x}{R_x}\sum_{m=0}^{M-1} x_m^2\right\}}{R_x(1-k_x^2)}\right)}{\left(\pi R_x\sqrt{1-k_x^2}\right)^M}. \tag{4}$$

Using this pdf, the maximum-likelihood estimators for Hermitian and complementary variance are, respectively,

$$\hat{R}_x = \frac{1}{M}\sum_{m=0}^{M-1}|x_m|^2, \quad \hat{\widetilde{R}}_x = \frac{1}{M}\sum_{m=0}^{M-1} x_m^2, \tag{5}$$

and the maximum-likelihood estimator for the circularity coefficient is

$$\hat{k}_x = \frac{\left|\hat{\widetilde{R}}_x\right|}{\hat{R}_x}. \tag{6}$$

IV. Spectral Noncircularity of Nonstationary Signals

In this paper, we consider N samples y_n, $n = 0, .., N-1$ of a real-valued time-domain signal with sampling frequency f_s. The short-time Fourier transform (STFT) is often used to analyze such signals, and is defined as

$$Y_{f,t} = \sum_{\ell=0}^{N-1} w_\ell \, y_{n+\ell+tN_{hop}} \, e^{-j\omega f \ell} \tag{7}$$

with $\omega := 2\pi f/N$ for $f \in \{0..F\}$ with $F = N/2+1$, where w_ℓ, $\ell = 0, .., N-1$ is a short-time window, N is the window length, and N_{hop} is the window hop.

To estimate noncircularity of the STFT, we use an estimator similar to Welch's method that averages adjacent STFT frames within a frequency bin. That is, to estimate the noncircularity of a STFT coefficient $Y_{f,t}$ of a real-valued time domain signal y_n, the M samples in the estimators (6) are $Y_{f,t+m}$ for $m = 0, .., M-1$. However, these STFT coefficients cannot be directly averaged together to estimate complementary variance, because the STFT hop N_{hop} causes deterministic phase progressions $e^{-j\omega N_{hop}}$ that destroy coherence of the complex-valued samples. Thus, to properly estimate complementary variance, the deterministic phase shifts must be compensated for, which means the samples used in the complementary variance estimator must be $Y_{f,t+m}e^{j\omega m N_{hop}}$. The estimators of time- and frequency-dependent variance from the STFT are

$$\hat{R}_y(f,t) = \frac{1}{M}\sum_{m=0}^{M-1}|Y_{f,t+m}|^2, \tag{8}$$

$$\hat{\widetilde{R}}_y(f,t) = \frac{1}{M}\sum_{m=0}^{M-1}(Y_{f,t+m}e^{j\omega m N_{hop}})^2. \tag{9}$$

V. Finite-Sample Approximation of Circularity Coefficient

To determine the suitability of a noncircular model for complex-valued random data, we can perform a hypothesis test:

$$\begin{aligned} \mathcal{H}_0 &: \; k_x = 0 \quad (x \text{ is circular}) \\ \mathcal{H}_1 &: \; k_x > 0 \quad (x \text{ is noncircular}) \end{aligned} \tag{10}$$

Assuming we have M i.i.d. samples $x_{0:M-1}$ of the random variable x, Delmas et al. [14, Remark 5] provide a finite-sample approximation for the sample circularity coefficient $\hat{k}_x$ under these two hypotheses, given by

$$\begin{aligned} \mathcal{H}_0 &: \; \hat{k}_x \sim \mathcal{R}\left(\frac{1}{\sqrt{M}}\right), \\ \mathcal{H}_1 &: \; \hat{k}_x \sim \mathcal{N}\left(k_x, \frac{(1-k_x^2)^2}{\sqrt{M}}\right). \end{aligned} \tag{11}$$

Using the distribution under the null hypothesis, we can choose a constant probability of false alarm (PFA) and compute a threshold T that we can test $\hat{k}$ against to determine statistically significant noncircularity. These pdfs are illustrated in figure 1 for $M = 10$ samples, a constant PFA of 0.05, and true circularity coefficient k equal to the threshold that achieves PFA $= 0.05$, which yields a threshold of $T = 0.77$. In practice, this finite-sample approximation has been empirically observed to be accurate for M as low as 10 [6, Appendix C].

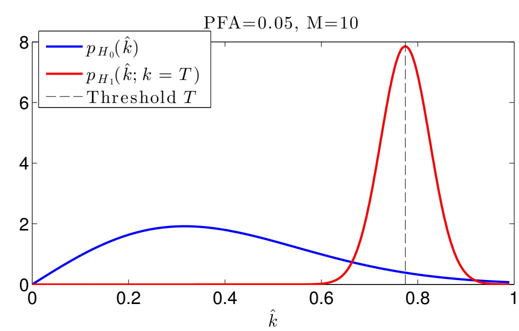

Fig. 1. Illustration of pdfs for finite-sample approximation of circularity coefficient estimator, for $M = 10$ samples, probability of false alarm (PFA) of 0.05. For illustration purposes, we choose a true circularity coefficient k equal to the theshold T, where $T = 0.77$ is chosen to achieve PFA $= 0.05$.

Next, we use the finite-sample approximation to demonstrate that a particular nonstationary speech signal exhibits substantial, statistically significant noncircularity. The left panel of figure 2 shows a spectrogram of a speech signal from a male speaker, which is the log magnitude of the complex-valued STFT. In the center panel, we plot the estimated circularity coefficient versus time and frequency bin computed using the

[1] Since we will assume Gaussian random variables, which are completely determined by their second-order moments, for brevity *circular* will refer to second-order circular throughout the remainder of this paper.

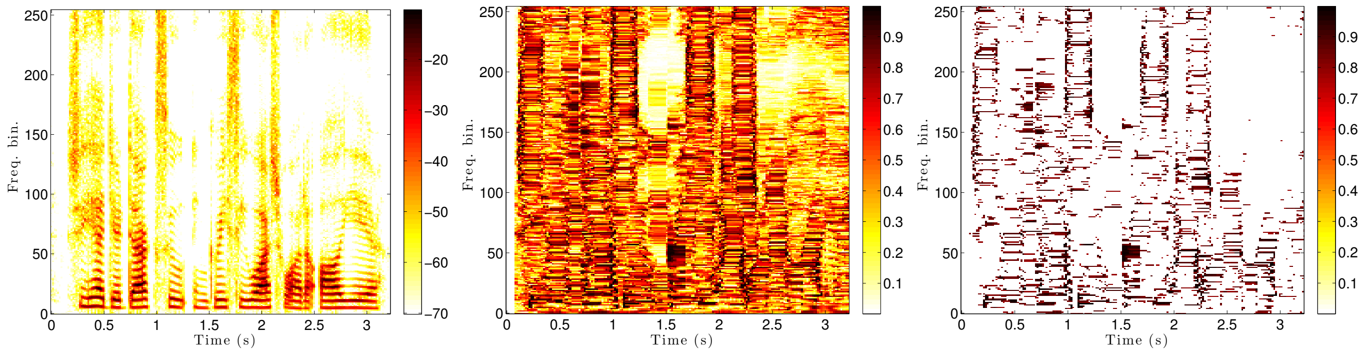

Fig. 2. Testing for noncircularity of a nonstationary speech signal. Left panel: spectrogram of speech signal. Center panel: estimated circularity coefficient versus time and frequency bin. Right panel: thresholded estimated circularity coefficient with PFA $= 0.05$.

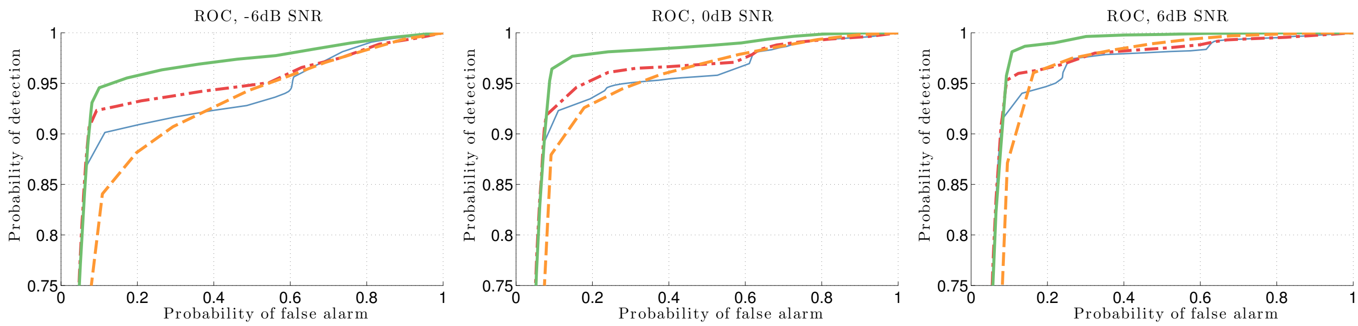

Fig. 3. Receiver operating characteristic (ROC) curves for DCASE2016 event detection with $N = 1024$ and $M = 64$ for various SNRs. Same legend as figure 4.

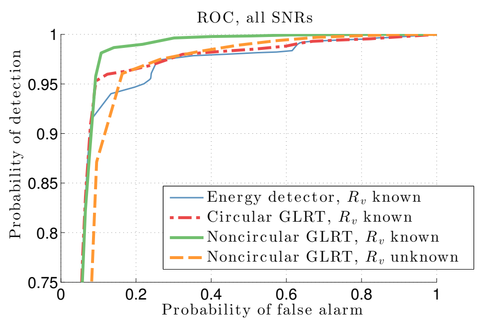

Fig. 4. ROC curve for DCASE2016 event detection with $N = 1024$ and $M = 64$ across all SNRs.

estimators (8) and (9). Using the finite-sample approximation (11), we determine the threshold for the hypothesis test (10) to be $T = 0.77$. In the right panel of figure 2, we apply this threshold to the estimated circularity coefficient shown in the center panel. Notice that the speech signal exhibits statistically significant noncircularity, especially at onsets and offsets.

VI. Detection of nonstationary acoustic events

To demonstrate the benefit of noncircularity for detection, we use Task 2 of the Detection and Classification of Acoustic Events 2016 (DCASE2016) dataset [15], available from http://www.cs.tut.fi/sgn/arg/dcase2016/. This dataset consists of eleven different types of acoustic events embedded in stationary background noise typical of an office environement. These events include clearing throat, coughing, door knock, door slam, drawer, human laughter, keyboard, keys put on table, page turning, phone ringing, and speech. Our goal is to detect when at least one acoustic event is active in 8 millisecond increments. Acoustic events are mixed with the background noise at -6dB, 0dB, and 6dB signal to noise ratio (SNR).

We will assume that the background noise v_n is stationary, which means its STFT coefficients $V_{f,t}$ are circular with frequency-dependent Hermitian variance $R_v(f)$. We will also assume that nonstationary acoustic events exhibit noncircularity in the STFT domain. Thus, the detection problem for a particular observed STFT frame $Y_{1:F,t}$ at time t is

$$\begin{aligned} \mathcal{H}_0 &: \; Y_{f,t} = V_{f,t}, && \forall f, \\ \mathcal{H}_1 &: \; Y_{f,t} = V_{f,t} + X_{f,t}, && \forall f, \end{aligned} \tag{12}$$

where the $X_{f,t}$ for $f \in \{1..F\}$ are the STFT coefficients of the acoustic event.

The generalized likelihood ratio test (GLRT) is defined as

$$G(y) \triangleq \left[\max_{\theta_1} p_{\mathcal{H}_1}(y|\theta_1)\right] / \left[\max_{\theta_0} p_{\mathcal{H}_0}(y|\theta_0)\right], \tag{13}$$

where θ_0 and θ_1 are the unknown parameters under hypotheses $\mathcal{H}_0$ and $\mathcal{H}_1$, respectively. If we assume that $X_{f,t}$ are circular Gaussians, the log GLRT for the detection problem (12) given observed STFT coefficients $\mathbf{Y}_t := Y_{1:F,t}$ is

$$\log G_C(\mathbf{Y}_t) = \sum_f \left(\frac{\hat{R}_y(f,t)}{R_v(f)} - \log \frac{\hat{R}_y(f,t)}{R_v(f)} \right). \tag{14}$$

When the $X_{f,t}$ are noncircular, the log GLRT is simply the circular log GLRT, $\log G_C(\mathbf{Y}_t)$ given by (14), plus an

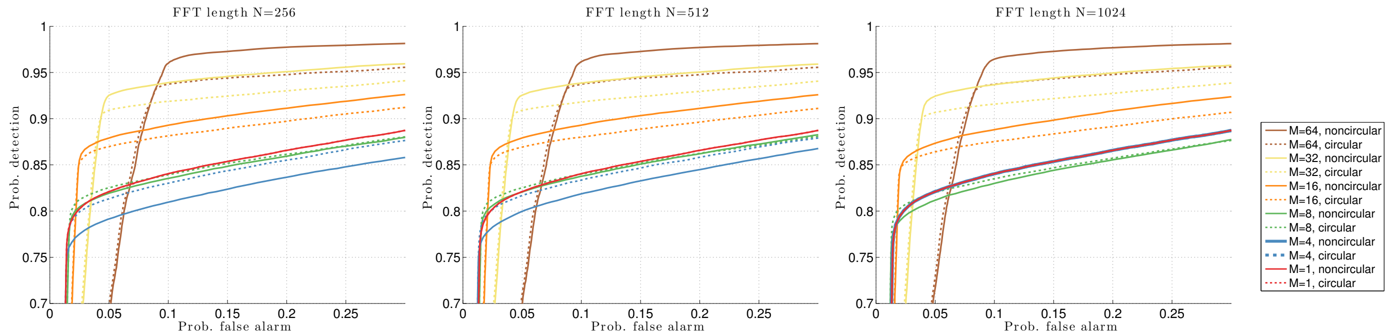

Fig. 5. ROC curvess for parameter sweep over FFT length N and averaging window length M. Notice that detection performance is relatively invariant to FFT length N for larger Ms, and that there is a tradeoff between probability of false alarm and probability of detection for larger Ms.

additional term:

$$\log G_{NC}(\mathbf{Y}_t) = \log G_C(\mathbf{Y}_t) - \frac{1}{2} \sum_f \log \left(1 - \left| \hat{k}_y(f,t) \right|^2 \right). \tag{15}$$

This additional term can also be used by itself as a blind detector, that is

$$\log G_{NC}^{\text{blind}}(\mathbf{Y}_t) = \log G_{NC}(\mathbf{Y}_t) - \log G_C(\mathbf{Y}_t). \tag{16}$$

This detector is blind in the sense that it does not require knowledge of the Hermitian noise variance $R_v(f)$. In practice, we estimate the noise variance $R_v(f)$ from all observation samples before the first acoustic event, which we know *a priori* to only contain noise. The blind detector is useful in situations where we do not know the when noise-only durations occur.

Using a FFT length of $N = 1024$ and averaging window length of $M = 64$, receiver operating curves (ROCs) for DCASE2016 event detection are shown in figure 3 for various SNRs, and the overall ROC curves across all SNRs are shown in figure 4. Note that the noncircular GLRT (solid thick green line) achieves the best performance versus all other detectors at all SNRs. The blind noncircular GLRT (dashed orange line) performs the worst, but has the advantage that it does not require knowledge of the Hermitian background noise variance $R_v(f)$. Note that if $R_v(f)$ is not known or cannot be estimated, detection using only circular Hermitian statistics is impossible.

VII. Effect of estimator parameters

In this section, we empirically explore the effect of the estimation parameters on DCASE2016 acoustic event detection. These parameters are the FFT length N, and the averaging window length M. Figure 5 shows ROC curves using the circular GLRT (14) and noncircular GLRT (15) for various settings of N and M. Notice that for larger M, the FFT length N does not have much effect on detection performance. Also, the noncircular detector does not start improving performance versus the circular detector until $M = 16$, which suggests that a certain minimum number of samples is required in order for a noncircular model to provide benefit.

To quantify the performance improvement and determine the best setting of estimator parameters, we use area under the curve (AUC), which is the integral of the ROC curve. For various FFT lengths N, figure 6 plots the AUCs for the ROCs in figure 5 versus averaging window length M. From the AUC plot, settings of $M = 32$ and $M = 64$ yield the best AUC, while FFT length N does not have much effect for larger M.

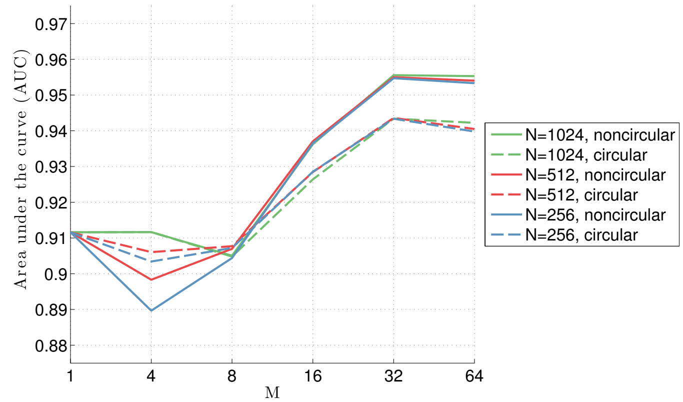

Fig. 6. Areas under the curve (AUCs) for parameter sweep ROCs in figure 5. Notice that $M = 32$ and $M = 64$ appear to be the best settings of averaging window length, while the AUC is not very dependent on FFT length N.

VIII. Conclusion

Real-valued nonstationary signals tend to exhibit second-order noncircularity in the complex-valued STFT domain. Given this property, we have shown that using noncircular Gaussian distributions to model the STFT provides distinct benefits for processing nonstationary signals. First, we described maximum likelihood estimators of the second-order statistics of noncircular Gaussian data. These estimators have a simple finite sample approximation, which allowed us to empirically test the statistical significance of noncircularity in the STFT. Then, we demonstrated the benefits of noncircular models for detection of nonstationary acoustic events and explored the effect of estimation parameter choice on these detection results.

Future work will explore multichannel extensions of the noncircular model. We also intend to explore the benefits of using noncircular non-Gaussian distributions of complex-valued random data, which require higher order statistical moments. Also, the application of nonstationary signals in the presence of nonstationary noise is another interesting extension.

References

[1] P. Schreier and L. Scharf, "Stochastic time-frequency analysis using the analytic signal: why the complementary distribution matters," *IEEE Transactions on Signal Processing*, vol. 51, no. 12, pp. 3071–3079, 2003.

[2] L. Atlas, "Modulation Spectral Transforms–Application to Speech Separation and Modification," in *Speech Dynamics by Ear, Eye, Mouth, and Machine, An Interdisciplinary Workshop, Technical Report of IEICE*, Kyoto, Japan, Jun. 2003. [Online]. Available: http://isdl.ee.washington.edu/papers/atlas-2003-ieice.pdf

[3] S. Douglas and D. Mandic, "Autoconvolution and panorama: Augmenting second-order signal analysis," in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2014, pp. 384–388.

[4] B. Rivet, L. Girin, and C. Jutten, "Log-rayleigh distribution: a simple and efficient statistical representation of log-spectral coefficients," *IEEE transactions on audio, speech, and language processing*, vol. 15, no. 3, pp. 796–802, 2007.

[5] F. Millioz and N. Martin, "Circularity of the STFT and Spectral Kurtosis for Time-Frequency Segmentation in Gaussian Environment," *IEEE Transactions on Signal Processing*, vol. 59, no. 2, pp. 515–524, 2011.

[6] P. Clark, "Coherent Demodulation of Nonstationary Random Processes," Ph.D. dissertation, University of Washington, 2012.

[7] P. Clark, I. Kirsteins, and L. Atlas, "Existence and estimation of impropriety in real rhythmic signals," in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2012, pp. 3713–3716.

[8] S. Wisdom, L. Atlas, and J. Pitton, "Extending coherence time for analysis of modulated random processes," in *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2014, pp. 340–344.

[9] S. Wisdom, G. Okopal, L. Atlas, and J. Pitton, "Voice Activity Detection Using Subband Noncircularity," in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Brisbane, Austrailia, Apr. 2015.

[10] G. Okopal, S. Wisdom, and L. Atlas, "Estimating the Noncircularity of Latent Components within Complex-Valued Subband Mixtures with Applications to Speech Processing," in *Proceedings of Asilomar Conf. on Signals, Systems, and Computers*, Pacific Grove, CA, Nov. 2014.

[11] ——, "Speech Analysis With the Strong Uncorrelating Transform," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 23, no. 11, pp. 1858–1868, Nov. 2015.

[12] S. Wisdom, L. Atlas, and J. Pitton, "On spectral noncircularity of natural signals," in *Sensor Array and Multichannel Signal Processing Workshop (SAM), 2016 IEEE*. IEEE, 2016, pp. 1–5.

[13] P. J. Schreier and L. L. Scharf, *Statistical Signal Processing of Complex-Valued Data: The Theory of Improper and Noncircular Signals*. Cambridge University Press, Feb. 2010.

[14] J.-P. Delmas, A. Oukaci, and P. Chevalier, "On the asymptotic distribution of GLR for impropriety of complex signals," *Signal Processing*, vol. 91, no. 10, pp. 2259–2267, Oct. 2011.

[15] D. Stowell, D. Giannoulis, E. Benetos, M. Lagrange, and M. Plumbley, "Detection and Classification of Acoustic Scenes and Events," *IEEE Transactions on Multimedia*, vol. 17, no. 10, pp. 1733–1746, Oct. 2015.