INTERSPEECH 2017
August 20–24, 2017, Stockholm, Sweden

Automatic time-frequency analysis of echolocation signals using the matched Gaussian multitaper spectrogram

Maria Sandsten[1], *Isabella Reinhold*[1], *Josefin Starkhammar*[2]

[1]Mathematical Statistics, Centre for Mathematical Sciences, Lund University, Sweden
[2]Department of Biomedical Engineering, Lund University, Sweden,

sandsten@maths.lth.se, isabella@maths.lth.se, josefin.starkhammar@bme.lth.se

Abstract

High-resolution time-frequency (TF) images of multi-component signals are of great interest for visualization, feature extraction and estimation. The matched Gaussian multitaper spectrogram has been proposed to optimally resolve multi-component transient functions of Gaussian shape. Hermite functions are used as multitapers and the weights of the different spectrogram functions are optimized. For a fixed number of multitapers, the optimization gives the approximate Wigner distribution of the Gaussian shaped function. Increasing the number of multitapers gives a better approximation, i.e. a better resolution, but the cross-terms also become more prominent for close TF components. In this submission, we evaluate a number of different concentration measures to automatically estimate the number of multitapers resulting in the optimal spectrogram for TF images of dolphin echolocation signals. The measures are evaluated for different multi-component signals and noise levels and a suggestion of an automatic procedure for optimal TF analysis is given. The results are compared to other well known TF estimation algorithms and examples of real data measurements of echolocation signals from a beluga whale (*Delphinapterus leucas*) are presented.

Index Terms: dolphin echolocation signals, multi-component signals, concentration measures, time-frequency analysis

1. Introduction

In marine biosonar research, it is important to determine the detailed origin and composition of echolocation signals from whales and dolphins [1, 2, 3, 4, 5]. The echolocation signals are short transient pulses, perhaps containing components originating from different locations within the sound generating mechanism or multiple reflections from internal tissue structures from the animal. Existing time-frequency (TF) representations are too noise sensitive or has too poor resolution for studying these multiple transient signal components individually. Therefore, there is a need for algorithms especially designed for TF visualization of short transient pulses.

From a TF concentration viewpoint, the Wigner distribution is the optimal choice of TF representation. The main drawback of the Wigner distribution is a phenomenon called cross-terms, which are large oscillating interference terms located midway between all components. Many methods have been proposed to reduce cross-terms at the cost of lower TF concentration [6]. Computationally efficient and noise robust algorithms can be found using multitaper spectrograms, especially if the number of averaged spectrograms are small [6, 7, 8]. The Hermite functions are optimal with regard to TF localization and orthogonality in the TF domain [9], and they are therefore often applied as multitapers for spectrogram estimation of non-stationary and transient signals [10, 11, 12, 13].

The matched Gaussian multitaper spectrogram (MGMT) derived in [14], gives the approximate Wigner distribution of a Gaussian shaped transient function. A fixed number of Hermite functions are used as multitapers and the weights of the different spectrogram functions are optimized to approximate the Wigner distribution. The number of spectrograms averaged determines the trade-off between resolution and cross-term suppression and with an automatic decision of the number an optimal TF representation and visualisation of a measured multi-component transient signal is achieved. Therefore we evaluate a number of well known concentration and sparsity measures for a variable number of tapers.

Many measures are based on information or sparsity optimization, where maybe the most applied in TF analysis is the normalized Rényi entropy [15, 16]. However, this measure is also well known to disregard cross-terms and the volume normalized Rényi entropy [17], which do consider cross-term suppression, is thus also evaluated. Two more measures are assessed, the time-frequency concentration measure suggested in [18], related to kurtosis, and the Gini index, which is found to satisfy a number of intuitive sparsity properties in [19]. The Gini index has recently been shown to be the optimal measure of a multitaper ambiguity function in [13].

The paper is organized as follows: section 2 introduces the matched Gaussian multitaper spectrogram and section 3 presents the evaluated concentration measures. The evaluation on simulated data and examples of real data are shown in section 4 and section 5. The paper is concluded in section 6.

2. The matched Gaussian multitaper spectrogram estimator

The multitaper spectrogram with windows $h_k(t)$ and weights α_k, $k = 1 \dots K$, is defined as

$$S_x(t,\omega) = \sum_{k=1}^{K} \alpha_k S_x^{h_k}(t,\omega), \tag{1}$$

where $S_x^{h_k}$ is the spectrogram,

$$S_x^{h_k}(t,\omega) = \left| \int_{-\infty}^{\infty} h_k(t-t_1) x(t_1) e^{-i\omega t_1} dt_1 \right|^2, \tag{2}$$

of the signal $x(t)$. The windows $h_k(t)$, $k = 1 \dots K$, are in this submission the Hermite functions, defined as

$$h_k(t) = \frac{1}{\sqrt{\pi^{\frac{1}{2}} 2^{(k-1)} (k-1)!}} H_{k-1}(t) e^{-\frac{t^2}{2}}, \quad k = 1, 2 \dots,$$

where

$$H_k(t) = (-1)^k e^{t^2} \frac{d^k}{dt^k} e^{-t^2}, \quad k = 0, 1 \dots.$$

http://dx.doi.org/10.21437/Interspeech.2017-119

A Gaussian windowed signal

$$x(t) = g(t - t_0)e^{-i\omega_0 t}, \tag{3}$$

where the unit-energy Gaussian function is

$$g(t) = \pi^{-\frac{1}{4}} e^{-\frac{1}{2}t^2}, \quad -\infty < t < \infty, \tag{4}$$

is often used to model a short non-stationary signal. As the quadratic class of distributions obey time-frequency shift-invariance $S_x(t - t_0, \omega - \omega_0) = S_g(t,\omega)$, further analysis can be restricted to $x(t) = g(t)$.

The Wigner distribution of the Gaussian function is

$$W_g(t,\omega) = 2e^{-(t^2+\omega^2)}, \tag{5}$$

and it can be shown that

$$W_g(t,\omega) = \sum_{k=1}^{\infty} \alpha_k S_g^{h_k}(t,\omega), \tag{6}$$

with $\alpha_m = 2$ and $\alpha_{m+1} = -2$ for $m = 2k - 1$ [14]. However, an infinite sum of multitaper spectrograms is of no use in practical calculations. If instead the number of terms in the sum is limited, i.e., K in Eq. (1) is assumed to be small, the computational effort is reasonable.

Fixing K and minimizing the total squared error with respect to α_k, $k = 1 \dots K$,

$$e_{min} = \min_{\alpha_k} \int_t \int_\omega \left(\sum_{k=1}^{K} \alpha_k S_g^{h_k}(t,\omega) - W_g(t,\omega) \right)^2, \tag{7}$$

will give the low-rank approximate Wigner distribution for the Gaussian function as a weighted multitaper spectrogram estimate, with the weights α_k and the Hermite functions $h_k(t)$ as window functions, $k = 1 \dots K$. The method is named Matched Gaussian MultiTaper (MGMT) spectrogram [14].

The resulting weights, for $K = 1 \dots 5$, are presented in Table 1. These weights are applicable for a Gaussian function and corresponding optimal Hermite functions disregarding the scaling of the Gaussian transient signal and the Hermite functions as long as they are matched according to the calculations above. When K is around $4 - 6$, the approximation to the Wigner distribution is satisfactory [14]. However, for close TF components, the cross-terms become more prominent for increasing value of K. A smaller value of K decreases the concentration but reduces the cross-terms. Therefore the trade-off between concentration and cross-term suppression need to be decided.

Table 1: *The weights for different values of* K.

α_k/K	1	2	3	4	5
α_1	1.333	1.778	1.926	1.975	1.992
α_2		-0.889	-1.481	-1.778	-1.909
α_3			0.593	1.185	1.580
α_4				-0.395	-0.922
α_5					0.263

3. Concentration measures

Many of the existing concentration measures in TF analysis are based on information or sparsity optimization where one of the most famous is the Rényi entropy (RE), [15], defined as

$$\text{RE} = \frac{1}{1-\alpha} \log_2 \sum_{n_0}^{n_1} \sum_{m_0}^{m_1} \left(\frac{S(n,m)}{\sum_{n_0}^{n_1} \sum_{m_0}^{m_1} S(n,m)} \right)^\alpha, \tag{8}$$

where the use of $\alpha = 3$ is the most applied formulation [16], and where all integrals are assumed to run over a chosen TF region of the time- and frequency discrete TF distribution, $S(n,m)$. The Rényi entropy adopt the lowest value for the Wigner distribution of a single Gaussian shaped transient signal, however it does not account for the cross-terms, if such are included in the chosen TF region. Instead the volume normalized Rényi entropy (VRE), is suggested in [17], defined as

$$\text{VRE} = \frac{1}{1-\alpha} \log_2 \sum_{n_0}^{n_1} \sum_{m_0}^{m_1} \left(\frac{S(n,m)}{\sum_{n_0}^{n_1} \sum_{m_0}^{m_1} |S(n,m)|} \right)^\alpha, \tag{9}$$

where the normalization is made with the absolute value of the TF representation. For both RE and VRE, a higher concentration of components results in smaller values of the measures.

A measure similar to the statistical kurtosis, which measures the sharpness of a distribution, has been suggested in [18] as a TF concentration (TFC) measure,

$$\text{TFC} = \frac{\sum_{n_0}^{n_1} \sum_{m_0}^{m_1} S^4(n,m)}{(\sum_{n_0}^{n_1} \sum_{m_0}^{m_1} S^2(n,m))^2}. \tag{10}$$

The fourth measure is the Gini index (GI), originally proposed in economics as measure of the inequality of wealth. In the application of TF analysis the values of the two-dimensional matrix $S(n,m)$ is organized into a sorted vector, $s(1) \leq s(2) \dots \leq s(N)$, $N = (n_1 - n_0 + 1) \cdot (m_1 - m_0 + 1)$ and the measure is defined

$$\text{GI} = 1 - 2 \sum_{l=1}^{N} \frac{s(l)}{||x||_1} \frac{N - l + 0.5}{N}, \tag{11}$$

where the l_1-norm is defined $||x||_1 = \sum_{l=1}^{N} |s(l)|$. For both TFC and GI, a higher concentration results in larger values of the measures.

4. Evaluation on simulated signals

The test signal is given as two complex valued Gaussian components that are moved closer together in time. The signal is exemplified in Figure 1, where the first component is located at $n = 50$ and $\omega = 2\pi 0.1$ and the second of half the amplitude at $n = 90$ and $\omega = 2\pi 0.04$, ($\omega = 2\pi f$), i.e., the time difference between the centers is 40 samples. The measures described in the previous section, Rényi entropy (RE), volume normalized Rényi entropy (VRE), TF concentration (TFC) and Gini index (GI) are applied to the MGMT spectrogram for different values of $K = 1 \dots 10$. For all measures the TF range of samples is $n_0 = 0$ to $n_1 = 500$ and frequency $m_0 = 0$ to $m_1 = 1023$ (corresponding to $\omega = 0$ to $2\pi 0.25$ with the number of FFT-samples set to 4096). The K_{opt} corresponding to the maximum values of the TFC and GI measures and the minimum values of the RE and VRE measures are extracted. The resulting MGMT for each measure is compared to the *sum of the Wigner distributions of the two individual components* (W_{sum}), which is the optimally concentrated, cross-term free, TF representation. The resulting mean square error (mse) is computed, giving the corresponding mse_{opt}.

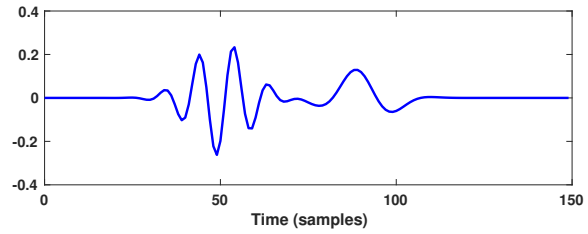

Figure 1: *A simulated signal of two Gaussian components of time difference* 40 *samples, where the first component is located at* $n = 50$ *and* $\omega = 2\pi 0.1$ *and the second at* $n = 90$ *and* $\omega = 2\pi 0.04$.

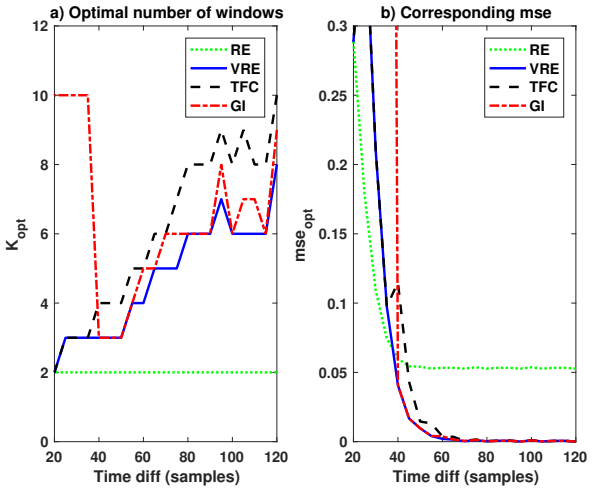

Figure 2: *Evaluation of measures for different time differences of two components; a) the resulting* K_{opt} *for different measures; b) the corresponding optimal mean square error,* mse_{opt}.

4.1. Different time differences

The results for different time differences, from 120 samples down to 20 samples, are depicted in Figure 2. For all measures except RE, K_{opt} is decreasing with the distance between the two components, Figure 2a). This follows from the fact that the MGMT only produces cross-terms if the Hermite functions used as windows overlap (in time- or frequency). A larger time distance, therefore welcomes a higher value of K and also a better approximation to W_{sum}, and accordingly a smaller mse, Figure 2b). RE gives $K_{opt} = 2$ independent of the distance, which corresponds to a rather high mse. It should also be noted that the GI measure fails for distances below 40 samples.

A thorough evaluation and comparison to other TF representations was done in [14] so in this submission, we restrict ourselves to evaluate for the best performance measure of the MGMT spectrogram. However, for comparison we mention the results of the Choi-Williams distribution, often applied in the context of biological signals, [20]. Optimizing the exponential parameter γ for time difference 100 samples gives the smallest $\text{mse}_{opt} = 0.015$ for $\gamma = 4$ and for time difference 40 samples, $\text{mse}_{opt} = 0.07$ for $\gamma = 2$. The optimal mse is given by the best performing kernel, when comparing to W_{sum}. We note that both optimal mse values are above the depicted mse values in Figure 2b).

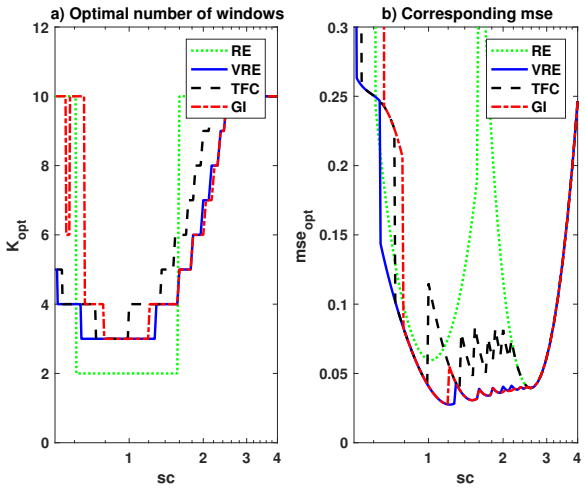

Figure 3: *Evaluation of measures for unmatched window and component lengths with fixed time difference* 40 *samples; a) the resulting* K_{opt} *for different measures; b) the corresponding optimal mean square error,* mse_{opt}.

4.2. Window length

For real data measurements, the length of the components are not known and therefore the matched Hermite function windows can not be chosen. Therefore we investigate the sensitivity of the choice of Hermite function window lengths in correspondence to the signal length. The relation is defined using a parameter sc which ranges from $sc = 0.5$, i.e., the window lengths are twice the length of the signal component, to $sc = 4$, where the window lengths are 4 times shorter than the actual component length. The time difference between components is fixed to 40 samples. The results are presented in Figure 3. VRE and GI perform similar for short window lengths. For longer window lengths, $sc < 1$, GI is somewhat more sensitive than VRE. The TFC measure gives a somewhat larger mse, and the performance of RE is very dependent of window lengths, performing poorly even when sc is close to one.

4.3. Noise

Usually measurements are corrupted by noise. In our recordings, the noise is of low frequency character and in this evaluation we simulate noise disturbances as low pass filtered white noise sequences using a FIR filter of order 50 and cut-off frequency 0.15. The SNR is defined as the averaged power of the signal in a reasonable time interval, divided by the averaged power of the noise. The two-component signal from previous simulations is used with fixed time difference of 40 samples. A number of 50 realizations are simulated where the phases of the components are uniformly distributed and low frequency noise realizations are added. Different SNR are simulated between 12 and 45 dB and the results are presented in Figure 4 as the average of the results of the 50 simulations. The best performance is given using VRE, which gives $K_{opt} = 3$ down to 20 dB. For SNR below 20 dB, $K_{opt} = 2$ which might be reasonable, as with decreasing SNR, the noise power is increasing and thereby also the cross-terms caused by spurious peaks of the noise.

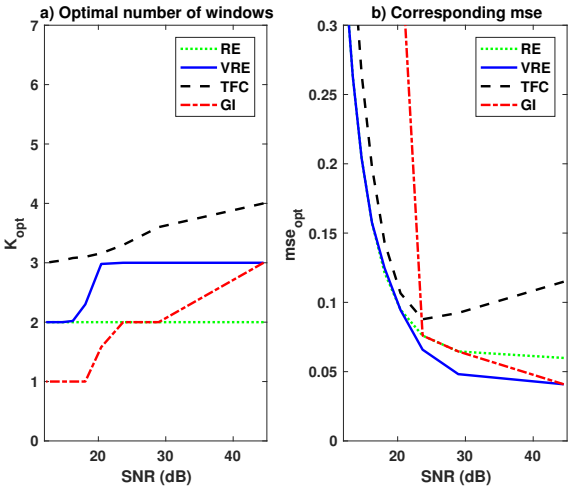

Figure 4: *Evaluation of measures for different SNR and time difference* 40 *samples; a) the resulting* K_{opt} *for different measures; b) the corresponding optimal mean square error,* mse_{opt}.

5. Evaluation on real data example

As example of the performance of the MGMT spectrogram and the automatic decision of K_{opt}, we show a measured echolocation signal from a beluga whale (*Delphinapterus leucas*), where the sample frequency is 1 MHz, depicted in Figure 5a). The approximative SNR is 20 dB where the mean signal power is measured between 280-360 μs and the noise power is estimated from 200 samples before the echolocation signal appears.

The different measures are evaluated for a TF range corresponding to the time interval 250-450 μs and 20-150 kHz, which is the range presented in Figure 5b-d). The VRE and GI measures perform similar in this example with $K_{opt} = 2$ and the corresponding multitaper spectrogram presented in Figure 5b). The strongest component is located at 300 μs and around 70 kHz, but a weaker component at 330 μs and 50 kHz is also visualized. This is not the case for the optimal multitaper spectrograms found using the TFC and RE measures. The MGMT with $K_{opt} = 6$ is given from the TFC measure and is shown in Figure 5c), where the second component is smoothed in the time direction from the use of higher order Hermite function windows. The spectrogram is also more influenced by noise. The optimal MGMT given from the RE measure is given for $K_{opt} = 10$ which is similar but worse than the spectrogram in Figure 5c). In Figure 5d) the Choi-Williams distribution is optimized using the VRE measure. The optimal mse is given with the parameter $\gamma = 0.3$.

6. Conclusion

The matched Gaussian multitaper (MGMT) spectrogram approximates the Wigner distribution of Gaussian shaped transient functions. The number of windows K controls the approximation and the cross-term suppression. In this submission we present an evaluation of different concentration or sparsity measures for automatic decision of K from real data measurements. Simulations show that volume normalized Rényi entropy (VRE) and Gini index (GI) have the most reliable performance to approximate the cross-term free Wigner distribution. For closely spaced components, VRE is superior to GI. The measures are

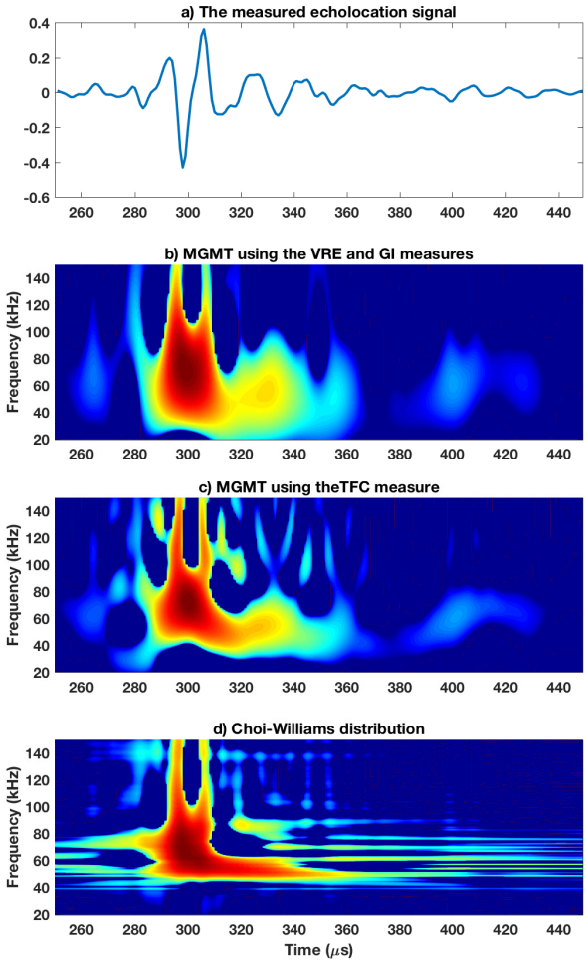

Figure 5: *Example of measure performances on a dolphin echolocation signal; a) the time signal; b) the MGMT spectrogram with* $K_{opt} = 2$ *from the VRE and GI measures; c) the MGMT spectrogram with* $K_{opt} = 6$ *from the TFC measure;d) the Choi-Williams distribution with* $\gamma_{opt} = 0.3$ *from the VRE measure.*

applied for TF estimation and representation of dolphin echolocation signals.

7. References

[1] C. Capus, Y. Pailhas, K. Brown, D.M. Lane, P. Moore, and D. Houser, "Bio-inspired wideband sonar signals based on observations of the bottlenose dolphin (Tursiops truncatus)," *J. Acoust. Soc. Am.*, vol. 121, no. 1, pp. 594–604, 2007.

[2] M.O. Lammers and M. Castellote, "The beluga whale produces two pulses to form its sonar signal," *Biol.Lett.*, vol. 5, pp. 297–301, 2009.

[3] W.W.L. Au, B. Branstetter, P.W. Moore, and J.J. Finneran, "Dolphin biosonar signals measured at extreme off-axis angles: Insights to sound propagation in the head," *J. Acoust. Soc. Am.*, vol. 132, no. 2, pp. 1199–1206, 2012.

[4] J. Starkhammar and M. Hansson-Sandsten, "Evaluation of seven time-frequency representation algorithms applied to broadband echolocation signals," *Advances in Acoustics and Vibration*, vol. 2015, pp. 1–13, 2015.

[5] J. Starkhammar, I. Reinhold, P. Moore, D. Houser, and M. Sandsten, "Intra-click time-frequency patterns across the echolocation beam of a beluga whale," *The Journal of the Acoustical Society of America*, vol. 140, no. 4, pp. 3239–3239, 2016.

[6] L. Cohen, *Time-Frequency Analysis*, Signal Processing Series. Prentice-Hall, Upper Saddle River, NJ, USA, 1995.

[7] M. G. Amin, "Spectral decomposition of time-frequency distribution kernels," *IEEE Trans. on Signal Processing*, vol. 42, no. 5, pp. 1156–1165, May 1994.

[8] D. J. Thomson, "Spectrum estimation and harmonic analysis," *Proc. of the IEEE*, vol. 70, no. 9, pp. 1055–1096, Sept 1982.

[9] I. Daubechies, "Time-frequency localization operators: A geometric phase space approach," *IEEE Trans. on Information Theory*, vol. 34, no. 4, pp. 605–612, 1988.

[10] S. Aviyente and W. J. Williams, "Multitaper marginal time-frequency distributions," *Signal Processing*, vol. 86, pp. 279–295, 2006.

[11] M. Hansson-Sandsten, "Optimal estimation of the time-varying spectrum of a class of locally stationary processes using Hermite functions," *EURASIP Journal on Advances in Signal Processing*, 2011, Article ID 980805.

[12] M. Hansson-Sandsten, "Evaluation of the optimal lengths and number of multiple windows for spectrogram estimation of ssvep," *Medical Engineering and Physics*, vol. 32, no. 4, pp. 372–383, 2010.

[13] B. Jokanovic and M. Amin, "Sparse and cross-term free time-frequency distribution based on Hermite functions," in *Proceedings of ICASSP*. IEEE, 2015, pp. 3696–3700.

[14] M. Hansson-Sandsten, "Matched Gaussian multitaper spectrogram," in *European Signal Processing Conference (EUSIPCO)*, Marrakech, Morocko, 2013, EURASIP.

[15] W. J. Williams, M. L. Brown, and A. O. Hero, "Uncertainty, information, and time-frequency distributions," in *Proc. SPIE Int. Soc. Opt. Eng.* SPIE, 1991, vol. 1566, pp. 144–156.

[16] R. G. Baraniuk, P. Flandrin, A. J. E. M. Janssen, and O. J. J. Michel, "Measuring time-frequency information content using Rényi entropies," *IEEE Trans. on Information Theory*, vol. 47, no. 4, pp. 1391–1409, May 2001.

[17] T.-H. Sang and W. J. Williams, "Rényi information and signal-dependent optimal kernel design," in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing*. IEEE, 1995, vol. 2, pp. 997–1000.

[18] D. L. Jones and T. W. Parks, "A high-resolution data-adaptive time-frequency representation," *IEEE Trans. on Acoustics, Speech, and Signal Processing*, vol. 38, pp. 2127–2135, 1990.

[19] N. Hurley and S. Rickard, "Comparing measures of sparsity.," *IEEE Transactions on Information Theory*, vol. 55, no. 10, pp. 4723 – 4741, 2009.

[20] H.-I. Choi and W. J. Williams, "Improved time-frequency representation of multi-component signals using exponential kernels," *IEEE Trans. on Acoustics, Speech and Signal Processing*, vol. 37, pp. 862–871, June 1989.