A Spectral Entropy-based Measure for Performance Evaluation of A First-order Differential Microphone Array

Ali Sarafnia, *Student Member, IEEE*, M. Omair Ahmad, *Fellow, IEEE* and M.N.S. Swamy, *Fellow, IEEE*
Department of Electrical and Computer Engineering
Concordia University, Montreal, PQ H3G 1M8, Canada
Email: sara_ali@encs.concordia.ca, {omair, swamy}@ece.concordia.ca

***Abstract*— For differential microphone arrays, most of the performance evaluation measures that are used in the context of noise reduction are based on the energy of the signal. In this paper, we propose a spectral entropy-based measure, which quantifies the ratio of the spectral information contained in the desired and actual outputs of the microphone array, and can evaluate the performance in terms of the average of lost/gain information. At the same time, the value of the spectral entropy-based measure shows whether a speech signal is noisy or if some information has been lost. The proposed measure provides some advantages over the energy-based measures, such as the array gain. Moreover, the performance of a first-order-differential microphone array designed based on the maximum value of the array gain is evaluated using the proposed spectral entropy-based measure.**

Keywords—Spectral Entropy; Differential Microphone Array; Array Gain; Beam-pattern

I. Introduction

Most of the current evaluation measures of microphone arrays emphasize the energy of the signal as a well-known criterion to find how good a microphone array is performing in the context of noise reduction. However, these measures are not capable of suggesting how perceptually good an utterance has been conveyed by a microphone array. In other words, the energy-based measures are not appropriate to determine the ability of a microphone array to convey captured information. Most of the current designs of differential microphone arrays are based on design parameter optimization and are driven by maximizing/minimizing the energy-based measures such as the array gain. This paper presents a spectral entropy-based measure to evaluate the performance of differential microphone arrays (DMAs) in the context of noise reduction and speech acquisition. Moreover, we design a first-order DMA based on maximizing the array gain and then evaluate the performance of such a first-order DMA using the proposed spectral entropy-based measure.

Entropy is usually used in the field of pattern classification and information technology. Originally, entropy was defined for information sources by Shannon [1]. It is a measure of disorganization or uncertainty in a random variable. The information can be interpreted as essentially the negative of the entropy, and the negative logarithm of its probability. The entropy of a random variable X with M states or symbol probabilities $[p_1, \ldots\ldots, p_M]$, where $P_X(x_i) = p_i$, is given by

$$H(X) = -\sum_{i=1}^{M} p_X(x_i)\, log_2 p_X(x_i) \qquad (1)$$

where H is the Shannon entropy. The application of the entropy concept for speech enhancement is based on the assumption that the speech spectrum is more organized during the speech segments compared to that during noise segments. In addition, the spectral peaks of the spectrum are supposed to be more robust to noise. Thus, a voiced region of speech would induce low entropy since there are clear formants in the region. The spectra of noise or unvoiced region would have a flatter distribution and thus a higher entropy [2, 3]. Simulations show that a wideband signal in time domain roughly contains less information rather than its noise-added version. Early efforts of employing entropy in noise reduction methods go back to Johnson's paper, where he implemented speech noise reduction by means of an information-theoretic spectrum estimation [4]. In addition, a generalization of minimum cross-entropy-spectrum analysis (MESA) was introduced by Misra, where he proposed the use of entropy features as speech features in speech (phoneme) recognition [3]. Moreover, Wang and Yi [5,6] proposed that an energy spectrum entropy method could be used to detect speech signals under various noise conditions. Also, Shen and Hang [7] presented an entropy-based algorithm for accurate and robust endpoint detection for speech recognition under noisy environments. They developed spectral entropy to identify the speech segments accurately, instead of using the conventional energy-based features. In this paper, we investigate the application of the entropy feature in the evaluation of first-

This work was supported in part by the Natural Sciences and Engineering Research Council (NSERC) of Canada and in part by the Regroupement Strategique en Microelectronique du Quebec (ReSMiQ).

978-1-5090-6389-5/17/$31.00 ©2017 IEEE

order differential microphone arrays. Utilization of the entropy concept in our proposed measure has the advantage of demonstrating the perceptual performance of a microphone array as opposed to the existing energy-based measures.

The paper is organized as follows: Section II discusses the theoretical background of the proposed entropy-based measure. Section III deals with the implementation of the spectral entropy-based measure for first-order DMA. In Section IV, experimental and performance assessment results for the proposed measure are provided. Conclusions and possible future work are presented in Section V.

II. Theoretical Background

A. Finding Probability Mass Function from Speech Spectrum

Spectral entropy, compared to time domain entropy, has the particular advantage that contributions to entropy from any particular frequency range can be explicitly separated. In fact, spectral entropy calculates information which is included in the various frequency components. As a result, if we define spectral entropy for each spectrum of the short-time Fourier transform (STFT) of the speech signal, then we are able to figure out the contribution of the perceptually important frequency components of the formants region rather than the time-domain entropy, which is just able to quantify the loss/gain of information. To compute the entropy of a short time spectrum, authors in [8] convert the spectrum into a probability mass function (PMF) by normalizing it over the sum of the energies of the frequency components of the short-time frame. By performing this, the area under the normalized spectrum in the full-band will sum up to unity. The following equation is used for the full-band normalization.

$$x_i = \frac{X_i}{\sum_{i=1}^{N} X_i} \; for \; i = 1 \; to \; N \tag{2}$$

where X_i is the energy of the i^{th} frequency component of the spectrum, $x = (x_1, \dots, x_N)$ is the PMF of the spectrum and N is the number of points in the spectrum (order of STFT/ number of DFT points). The entropy for each frame is computed from x by

$$H = -\sum_{i=1}^{M} x_i \, log_2 x_i \tag{3}$$

In applications such as ASR, it is important to separate the speech/silence by their spectral entropy; therefore, the energy of the frequency components or the magnitude squared of the spectrum is employed to magnify the difference between the smaller and larger values of the spectrum. This leads to lower spectral entropies for the speech parts. Thus, a higher dynamic range of spectral entropy difference between the speech parts and the silent parts is achieved, thus making the speech/silence detection easier. In other words, using magnitude squared of the spectrum reduces the uncertainty of speech spectrum while increasing the spectral entropy difference [8]. In this paper, we employ full-band spectral entropy in spectral information comparison. In other words, we do not intend to magnify the difference between the smaller and larger values of the spectrum in our proposed method. Accordingly, X_i can be defined as the magnitude of the i^{th} frequency component of the spectrum. Therefore, PMF is defined as the ratio of the magnitude to the sum of the total magnitudes of the spectrum, and is called the normalized magnitude. Similar to PMF of (2), PMF obtained from the magnitudes sum up to unity. The computational complexity will decrease as a result of using the magnitude instead of the magnitude squared value of the spectrum in the definition of PMF.

B. Definition of Spectral Entropy-based Measure (O_{DMA})

For every STFT of the speech frame, the proposed entropy-based measure is defined as

$$O_{DMA} = \frac{-\sum_{i=1}^{N} x_i^{output} log2(x_i^{output})}{-\sum_{i=1}^{N} x_i^{desired} log2(x_i^{desired})} \tag{4}$$

where x_i^{output} and $x_i^{desired}$ are the PMF components of their corresponding spectra. Three states are defined for O_{DMA} as follows:

$$Output_{DMA} = \begin{cases} Noisy & , \; O_{DMA} > 1 \\ Desired\ Speech & , \; O_{DMA} = 1 \\ Information\ lost & , \; O_{DMA} < 1 \end{cases} \tag{5}$$

III. Design of first-order DMA based on Maximizing the Array Gain

A. Review of First-order DMA Design

The filter parameters for M microphones are obtained by solving a linear system of M equations. The steering vector of length M is

$$d(\omega, cos\varphi) = [1 \; (e^{-j\omega\tau_0 cos\varphi})^1 \dots (e^{-j\omega\tau_0 cos\varphi})^{M-1}]^T \tag{6}$$

where φ is incident angle, τ_0 is the delay between two adjacent microphones at $\varphi = 0^\circ$ and, ω is the angular frequency. A first-order DMA requires a two-element microphone array. For the design of beam patterns, two constraints must be fulfilled, namely, a distortion-less response (a gain of unity at incident angle $\varphi = 0^\circ$) and a null within the interval $0^\circ < \theta < 180^\circ$. Accordingly, the design coefficients are $\alpha = cos\theta$, $\beta = 0$, where the angle θ represents the location of the null ($\beta = 0$) in the beam pattern. For a microphone spacing that is much smaller than the acoustic wavelength, the approximation ($e^x \approx 1 + x$) can be applied. Therefore, by approximating ($1 - e^{j\omega\tau_0(\alpha-1)}$), a first-order DMA can be designed as follows:

$$h(\omega) = \begin{bmatrix} h_1(\omega) \\ h_2(\omega) \end{bmatrix} = \frac{j}{(\alpha - 1)\tau_0\omega} \begin{bmatrix} 1 \\ -e^{j\omega\tau_0\alpha} \end{bmatrix} \tag{7}$$

where $h_1(\omega)$ and $h_2(\omega)$ correspond to the impulse responses of the two microphones.

B. Defining Array Gain for First-order DMA

In this section, we find the array gain of a first-order DMA. For a wideband desired signal $S(\omega)$, first-order DMA microphone signals are given by

$$\begin{aligned} X_1(\omega, \varphi) &= S(\omega) + V_1(\omega), \\ X_2(\omega, \varphi) &= e^{-j\omega\tau_0 cos\varphi} S(\omega) + V_2(\omega), \end{aligned} \tag{8}$$

where $V_1(\omega)$ and $V_2(\omega)$ are additive white noises. We define the input signal-to-noise ratio (SNR) as

$$iSNR(\omega) = \frac{\phi_{X(\omega)}}{\phi_{V_1(\omega)}} \tag{9}$$

where $\phi_{X(\omega)} = E[X(\omega)^2]$ and $\phi_{V_1(\omega)} = E[V_1(\omega)^2]$ are the variances of $X(\omega)$ and $V_1(\omega)$, respectively. The output SNR is defined as

$$oSNR[h(\omega)] = \frac{\phi_{X(\omega)}}{\phi_{V_1(\omega)}} . \frac{|h^H(\omega)d(\omega, cos0^{\circ})|^2}{h^H(\omega)\Gamma_v(\omega)h(\omega)} \tag{10}$$

where $\Phi_v(\omega) = E[v(\omega)v(\omega)^H]$ and $\Gamma_v(\omega) = \frac{\Phi_v(\omega)}{\phi_{V_1(\omega)}}$ are the correlation and pseudo-coherence matrices of $v(\omega)$, respectively. The definition of the array gain in SNR is easily derived from the two previous definitions:

$$\mathcal{G}[h(\omega)] = \frac{oSNR[h(\omega)]}{iSNR(\omega)} = \frac{|h^H(\omega)d(\omega, cos0^{\circ})|^2}{h^H(\omega)\Gamma_v(\omega)h(\omega)} \tag{11}$$

C. Re-design of First-Order DMA Based on Maximum Gain

In this section, we design a first-order DMA based on the maximum array gain and we find the corresponding angle of null, which is the design parameter of the first-order DMA. Taking the derivative of $\mathcal{G}[h(\omega)]$ with respect to θ and setting it to zero, we obtain $\theta = 90^{\circ}$ in the interval $0^{\circ} < \theta < 180^{\circ}$. Redesigning the first-order DMA by setting $\theta = 90^{\circ}$ and $\varphi = 0^{\circ}$ will lead to maximum array gain at this null angle. Also, it is desired to find the beam pattern of such a first order DMA at the incident angle $\varphi = 90^{\circ}$. Finally, we find the value of O_{DMA} at the null angle $\theta = 90^{\circ}$ to determine how O_{DMA} and the array gain interact. By defining the probability of each frequency component of the output (magnitude) P_Y

$$P_Y = \frac{P_S - M_2 V_2(\omega)}{M_1} \tag{12}$$

where P_S, the PMF component of the desired signal is

$$P_S = \frac{S_i(\omega)}{\sum_{i=1}^{N} S_i(\omega)} \tag{13}$$

and M_1, M_2 are given by

$$M_1 = 1 - j\frac{\alpha}{(\alpha - 1)}\frac{\sum_{i=1}^{N} V_2(\omega)}{\sum_{i=1}^{N} S_i(\omega)},$$
$$M_2 = j\frac{\alpha}{(\alpha - 1)\sum_{i=1}^{N} S_i(\omega)}. \tag{14}$$

equation on (4) can be written as

$$O_{DMA} = \frac{-\sum_{i=1}^{N} P_Y \log_2 P_Y}{-\sum_{i=1}^{N} P_S \log_2 P_S} = \frac{-\sum_{i=1}^{N}\left(\left(\frac{P_S - M_2 V_2(\omega)}{M_1}\right)\left(\frac{P_S - (M_1 + M_2 V_2(\omega))}{M_1 ln2}\right)\right)}{-\sum_{i=1}^{N}\left(\frac{P_S^2 - P_S}{ln2}\right)} \tag{15}$$

IV. Experimental Results

In this section, we provide experimental results for performance evaluation of first-order DMA. We have chosen 100 utterances spoken by both male and female speakers from the TIMIT database. The results in this section are obtained from frame-based analysis of one utterance, which is the wave sound “sa1.wav” with *frame length=20 ms* and fundamental frequency (F_0)= $235\ Hz$. Also, the microphone distance δ has been set to *2 cm* to prevent spatial aliasing. In Table I, O_{DMA} values of first-order DMA as well as the Mean Opinion Scores (MOS) are given for five different null-angles. The average of MOSs is ranked from 1 (Bad/Very Annoying) to 5 (Excellent), by five listeners. Also, the fundamental frequencies and the array gain are given for different θ. Table II includes the location of the first four formants of the desired signal and the output of the first-order DMA for five different null angles at SNR=0 dB. We observe that the formant values for $\theta = 90^{\circ}$ (Dipole Beampattern) are the most similar one to the formants of the desired speech. Fig. 1 illustrates the narrowband version of the array gain, beam pattern and O_{DMA} of first-order DMA while the input speech is “sa1.wav”. Also, the value of O_{DMA} over all the frames of “sa1.wav” is illustrated in Fig. 1. It can be concluded from Fig. 1 that the array gain reaches its maximum value at the null angle of 90°, beam pattern has its notches while the incident angle is 90°, and O_{DMA} value for all FFT points approach unity while null angle is set to 90°. Fig. 2 depicts the values of O_{DMA} at three different null angles over FFT points. It can be observed from Fig. 2 that O_{DMA} is equal to unity when the null angle is set to 90°. Finally, Fig. 3 shows the narrowband version of O_{DMA}, array gain and the beam pattern of first-order DMA at $F = 1.2\ kHz$.

TABLE I. VALUES OF O_{DMA} AS WELL AS ARRAY GAIN, PITCH AND MEAN OPINION SCORES (MOS) FOR FIVE DIFFERENT NULL ANGLES

ATTRIBUTE	Desired speech	Output speech $\theta = 0^{\circ}$, Null	Output speech $\theta = 90^{\circ}$, Dipole	Output speech $\theta = 120^{\circ}$, Hyper-Cardioid	Output speech $\theta = 135^{\circ}$, Super-Cardioid	Output speech $\theta = 180^{\circ}$, Cardioid
MOS	5	1	5	2	3	3
O_{DMA}	1	1.1415	1.0020	1.1199	1.1437	1.1391
Array Gain (dB)	N/A	-Inf	323.0561	7.889371	6.243918	5.420067
Pitch (Hz)	235.2941	N/A	242.4242	262.2951	242.4242	262.2951

TABLE II. LOCATION OF FIRST FOUR FORMANTS FOR DIFFERENT NULL ANGLES

Formant Number #	Desired speech (Hz)	Output (Hz), $\theta = 0^{\circ}$, Null	Output (Hz), $\theta = 90^{\circ}$, Dipole	Output (Hz), $\theta = 120^{\circ}$, Hyper-Cardioid	Output (Hz), $\theta = 135^{\circ}$, Super-Cardioid	Output (Hz), $\theta = 180^{\circ}$, Cardioid
1	706.6	912.6	704.6	707.9	707.9	707.0
2	830.1	1716.6	867.5	1571.6	1584.4	1585.9
3	1586.4	2492.4	1613.3	2493.6	2495.7	2497.6
4	2532.8	3302.9	2529.9	3340.6	3338.9	3353.4

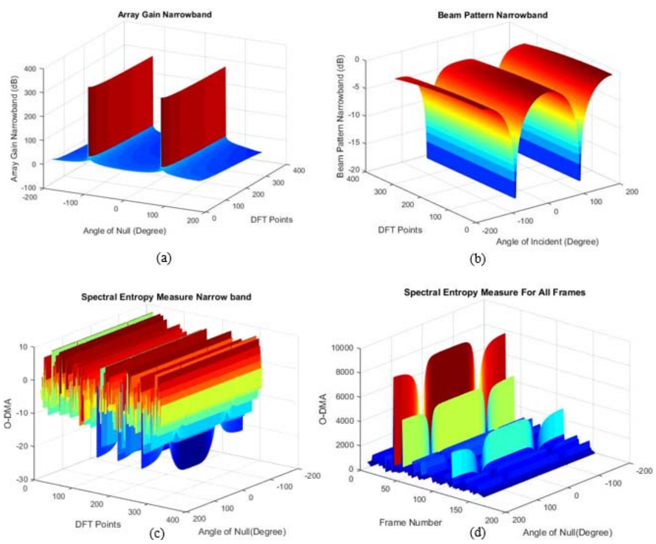

Fig. 1: For DFT Points=320: (a) Array Gain (Narrowband), (b) Beampattern (Narrowband), (c) O_{DMA} (Narrowband) (d) O_{DMA} for All Frames

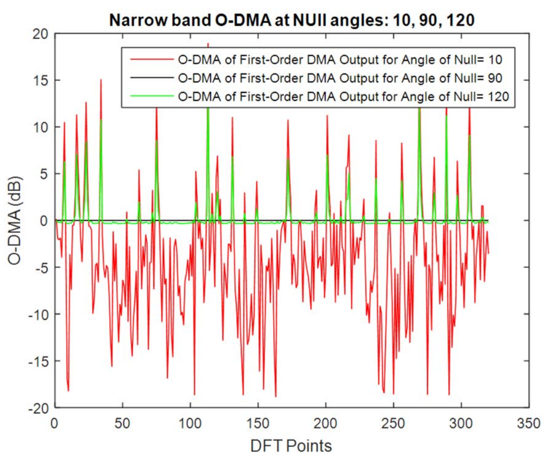

Fig. 2: Values of O_{DMA} for Three Different Null Angles Over DFT Points (Red line) 10 Degree, (Blue Line) 90 Degree, (Green Line) 120 Degree

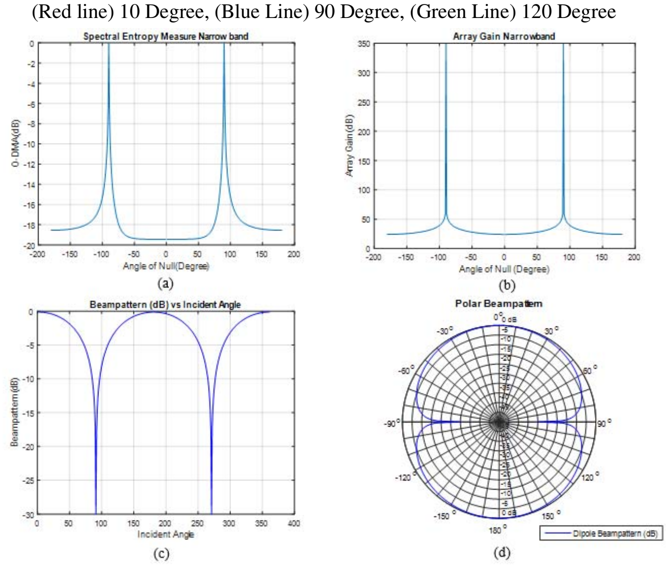

Fig 3: For $F = 1.2\ kHz$: (a) O_{DMA} (Narrowband), (b) Array Gain (Narrowband), (c) Beampattern (Narrowband), (d) Beampattern, Polar Form (Narrowband)

V. Conclusion

We have presented an entropy-based measure for the performance evaluation of first-order differential microphone arrays. We have also designed first-order DMA based on maximizing the array gain and it is found that the null angle of 90°, which is the design parameter of a first-order DMA, leads to the maximum value of the array gain. Moreover, for the null angle of 90°, O_{DMA} reaches the value of unity. Based on the definition of O_{DMA}, a value close to unity implies that the amount of information at the output of DMA is equal to the information contained in the desired speech signal. Another advantage of O_{DMA} compared to using the array gain as the metric is that the desired signal cancellation factor, which is an energy-based measure for evaluation of microphone arrays, can be replaced by the spectral entropy-based measure, since values less than unity for O_{DMA} indicate the desired signal cancellation. Consequently, there is no need to find the desired signal cancellation while evaluating the performance of DMA using O_{DMA}. The Array gain does not have such a characteristic and the desired signal cancellation has to be obtained separately.

We intend, in the future, to extend the concept of the spectral entropy-based measure for the evaluation of higher order differential microphone arrays.

References

[1] K.M. Pointing and S.M. Peeling, “The use of variable frame rate analysis in speech recognition,” Computer Speech and Language, vol. 5, no. 2, pp. 169–179, April 1991.

[2] P. Renevey and A. Drygajlo, “Entropy based voice activity detection in very noisy conditions,” in Proc.Eurospeech, USA, Sept. 2001, pp. 1887–1890.

[3] H. Misra, S. Ikbal, H. Bourlard, and H. Hermansky, “Spectral entropy based feature for robust asr,” in Proc. ICASSP, May 2004, pp. 193–196.

[4] R. Johnson, J. Shore, J. Buck and D. Burton, "Speech noise reduction by means of multi signal minimum-cross-entropy spectral analysis," Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP '83., 1983, pp. 1129-1132.

[5] Wang, K.C., Tsai,Y.H.:Voice activity detection algorithm with low signal-to-noise ratios based on spectrum entropy. In: Proceedings of the International Symposium on Universal Communication, pp. 423–428, 2008.

[6] Yi, Hu, Loizou, P.C.: Evaluation of objective quality measures for speech enhancement. IEEE Trans. Audio Speech Lang. Process. Vol. 16, no. 1, pp. 229–238, 2008.

[7] J.-L. Shen, J.-W. Hung, and L.-S. Lee, “Robust entropy-based endpoint detection for speech recognition in noisy environments.” in International Conference on Spoken Language Processing (ICSLP). ISCA, 1998.

[8] H. Misra, S. Ikbal, S. Sivadas and H. Bourlard, "Multi-resolution Spectral Entropy Feature for Robust ASR," Proceedings. (ICASSP '05). IEEE International Conference on Acoustics, Speech, and Signal Processing, 2005., 2005, pp. 253-256.