PBF-GSC BEAMFORMING FOR ASR AND SPEECH ENHANCEMENT IN REVERBERANT ENVIRONMENTS

Yi Ren Leng, Jonathan William Dennis, Ng Wen Zheng Terrence, and Huy Dat Tran

Human Language Technology Department, Institute for Infocomm Research,
A*STAR, Singapore

ABSTRACT

Reverberant noise is present in most practical applications of Automatic Speech Recognition (ASR) that do not rely on close-talk microphones. Unlike additive noise, there is limited progress in reducing the effect of reverberant noise in speech. For our REverberant Voice Enhancement and Recognition Benchmark (REVERB) challenge submission, we propose to use of a beamforming system combining the Generalized Sidelobe Canceller (GSC) with Phase-error based filtering (PBF). Our system is calibrated to make use of the baseline HMM recognizer with CMLLR and multi-condition training as it is the most practical setup for realistic applications.From the results we obtained, it appears that the main contribution to reverberant noise reduction comes from the combination of multiple audio streams while the adaptive filtering of GSC provides a slight improvement on top of this combination. For the speech enhancement task, we simply apply a previous noise reduction method, which is based on generalized gamma distribution modeling of speech spectral magnitude, on the output of developed PBF-GSC beamforming.

Index Terms— ASR, reverberant, beamforming, GSC, HMM, SGMM, DNN

1. INTRODUCTION

The REVERB [1] challenge is designed to compare the performance of Automatic Speech Recognition (ASR) systems in reverberant noise on a common database and evaluation metric. Reverberant noise is highly correlated with the target signal and is difficult to isolate compared to additive noise. In realistic applications of ASR systems where the speaker is not using a close-talk microphone, reverberant noise will always be present. Our submission for this challenge is focused mainly on addressing this problem: the reduction of reverberant noise in realistic environments.

We choose to concentrate on the multi-condition accoustic models instead of the clean models as it shows a significant improvement in Word Error Rate (WER) over the clean models in all of the noisy test conditions. It is shown in the challenge that the multi-condition training data can be easily generated from the clean data with arbitrary impulse responses and additive noise thus creating such data for real systems should not be an issue. Finally, our system is tuned to optimize the ASR performance on the RealData set as it offers the greatest room for improvement and is closest to realistic test environments.

Testing with the RealData development set, we found that beamforming [2] the 8-channel audio data into a single channel gives the best ASR performance. Our final system uses Phase-error based filtering (PBF) and Generalized Sidelobe Cancellation (GSC) in a combined PBF-GSC beamformer on the 8-channel audio to generate a single enhanced channel to be passed to the recognizer. We present our results for the proposed PBF-GSC front-end on the baseline MFCC-HMM recognizer and show that it is equally effective on the more advanced Subspace Gaussian Mixture Model (SGMM) and Deep Neural Network (DNN) recognizers.

For the speech enhancement task, we simply apply a previous noise reduction method, which is based on generalized gamma distribution modeling of speech spectral magnitude [3], on the output of developed PBF-GSC beamforming.

2. FRONT-END SELECTION

We tested conventional noise reduction methods such as RASTA filtering [4] with Perceptual Linear Predictive (PLP) features, ETSI Advanced Front End [5] using Mel-Frequency Cepstral Coefficients (MFCC) with Wiener filtering and noise robust Power-Normalized Cepstral Coefficients (PNCC) [6] on the RealData development set with both clean and multi-conditional models. All three of the above methods show significant improvements over the baseline MFCC system for clean training. However, when multi-condition training is used, the noise-robust systems actually fare worse than the baseline system as the great gains shown in clean training are significantly reduced in multi-condition training. Based on these findings, we decided to rely on the baseline MFCC for our feature front-end and seek other avenues for noise reduction.

With eight channels of audio data available, beamforming can be applied to reduce the effect of reverberation on speech. Firstly, Generalized Cross Correlation with Phase Transform

(GCC-PHAT) [7] is used to estimate the time difference of arrival (TDOA) for the 8 audio channels. The TDOA is used to align the individual channels in time before beamforming commences:

2.1. Generalized Sidelobe Canceller (GSC)

The GSC [8] algorithm is designed to steer the beamformer output to a fixed direction of interest by reducing the contribution of adjacent signals or sidelobes. We define this direction as the channel with the minimum TDOA x_r. It is implemented in two-steps: fixed beamforming where the individual channels are multiplied by fixed weights and summed to form a single reference signal, and adaptive filtering where the sidelobes are estimated and adaptively filtered from the reference signal.

The simplest implementation of the fixed beamformer (FBF) is simply to weigh each of the n channels x_i equally to obtain the reference signal y_f:

$$y_f = \sum_{i=1}^{n} w_i x_i = \sum_{i=1}^{n} \frac{x_i}{n} \tag{1}$$

A blocking matrix is constructed of $n-1$ linearly independent rows that sum to zero to estimate the sidelobes to be cancelled. We define the blocking matrix output b_i to be the difference of each other channel i with the reference channel x_r:

$$b_i = x_i - x_r, \qquad i \in [1, n], i \neq r \tag{2}$$

An adaptive Least Mean Squares (LMS) filter is used to iteratively filter the reference signal y_f at each time instant t by minimizing the $n-1$ blocking matrix outputs b_i. The adaptive LMS filter is an M order Finite Impulse Response (FIR) filter that minimizes the total power of the beamformer $p(t)$ using stochastic gradient descent.

$$p(t) = y_f(t) - \sum_{i=1, i\neq r}^{n} \sum_{j=1}^{M} w_{ij}(t) b_i(t+1-j) \tag{3}$$

$$||b_i(t)|| = \sum_{j=1}^{M} b_i(t+1-j) b_i(t+1-j) \tag{4}$$

$$w_{ij}(t+1) = \beta w_{ij}(t) + \mu \frac{p(t)}{||b_i(t)||} b_i(t+1-j) \tag{5}$$

In the above equations, the channel index i runs from $[1, n]$, excluding the reference channel r that has the minimum TDOA. β and μ are hyper-parameters governing the updating of the adaptive filter. Based on empirical testing with the RealData development set, the values $\beta = 0.99$ and $\mu = 0.02$ are found to give the best performance.

2.2. Phase-error based filtering (PBF)

The multiple-microphone PBF [9] uses the phase error between microphone channels to maximize the signal-to-noise ratio (SNR) for each time-frequency block and combines them into a single enhanced signal. A sliding Short Time Fourier Transform (STFT) is used to transform windows of the temporal signal into the time-frequency domain for phase estimation.

The phase error θ_{ij} for each window t is defined as the difference between the phase angles of channels X_i and X_j:

$$\theta_{ij}(t) = \angle X_i(t) - \angle X_j(t), \qquad i \neq j \tag{6}$$

For each of the n channels, the $n-1$ values of θ_{ij} are computed and multiplied into a single mask M_i:

$$M_i(t) = \left(\prod_{j=1, j\neq i}^{n} \frac{1}{1 + \gamma \theta_{ij}^2(t)} \right)^{\frac{1}{k}} \tag{7}$$

γ and k are hyperparameters that control the aggressiveness of the masking. Based on empirical testing with the RealData development set, the values $\gamma = 0.001$ and $k = 0.3n$ were chosen.

Each of the n masks M_i is applied to the corresponding STFT X_i and summed to give a filtered STFT X_f:

$$X_f = \sum_{i=1}^{n} M_i X_i \tag{8}$$

Finally, the Inverse Fast Fourier Transform (IFFT) is applied to X_f to transform it back into a temporal signal x_f which is the output of the PBF.

2.3. Proposed PBF-GSC system

The FBF in the GSC algorithm is used to obtain a cleaner reference signal y_f for the subsequent adaptive filtering. We propose to replace y_f with the PBF output x_f for a combined PBF-GSC system. As stated in [8], there are many possible ways to choose the weights for the FBF and the PBF can be seen as another solution to this problem. Applying the adaptive LMS filter on the PBF output should further reduce the noise present thus combining these two systems should produce better results than using them independently.

3. BACK-END SELECTION

3.1. Hidden Markov Model (HMM)

We use the provided baseline recognition system with Constrained Maximum Likelihood Linear Regression (CMLLR) adaption for our HMM system.

3.2. Subspace Gaussian Mixture Model (SGMM)

Here we briefly introduce the SGMM approach, which is an improvement over the conventional HMM-GMM framework [10]. The idea of SGMM is that while the HMM states share a common structure, the means and mixture weights of each state are allowed to vary within a subspace of the full parameter space. This is controlled by a global mapping from a vector space to the space of shared GMM parameters. The shared GMM is commonly referred to as a universal background model (UBM) that covers the total variability of the acoustic vector space.

Following the notation used in [10], the most basic form of the model, without speaker adaptation or sub-states can be denoted as follows. We use the index $1 \leq i \leq I$ for the Gaussians in the shared GMM, and the index $1 \leq j \leq J$ for the clustered phonetic states. For each state j, the probability model $p(x|j)$ is:

$$p(x|j) = \sum_{i=1}^{I} w_{ji} \mathcal{N}(x; \mu_{ji}, \Sigma_i) \quad (9)$$

$$\mu_{ji} = M_i v_j \quad (10)$$

$$w_{ji} = \frac{\exp(w_i^T v_j)}{\sum_{i=1}^{I} \exp(w_i^T v_j)} \quad (11)$$

where $M_i \in \mathbb{R}^{D \times S}$ are the GMM mean-projection matrices, $w_i \in \mathbb{R}^S$ are the weight-projection vectors, $v_j \in \mathbb{R}^S$ are the state-specific vectors that control the mapping, and $\Sigma_i \in \mathbb{R}^{D \times D}$ are the variances. D is the dimension of the input MFCC features, while the "subspace" of dimension S is a subspace of the total parameter space of the means of the GMM. In our experiments we set $S = 40$, $I = 700$ and $J = 9000$. An extension to the above model enables the use of sub-states, which are similar to using several mixtures to model each state in traditional HMM-GMM. The phonetic states are gradually split during the training iterations to produce 30000 sub-states in total for the model.

3.3. Deep Neural Network (DNN)

For comparison with the previous GMM-based systems a DNN-HMM hybrid system is also trained. The purpose of the DNN is provide posterior probability estimates for the HMM states through discriminative training. Using the notation from [11], for an observation o_{ut} corresponding to time t in utterance u, the output $y_{ut}(s)$ of the DNN for the HMM state s is obtained using the softmax activation function:

$$y_{ut}(s) \triangleq P(s|o_{ut}) = \frac{\exp(a_{ut}(s))}{\sum_{s'} \exp(a_{ut}(s'))} \quad (12)$$

where $a_{ut}(s)$ is the activation at the output layer corresponding to state s. In our case, the network is trained to optimise the cross-entropy objective function using the standard error back-propagation procedure, which is done through stochastic gradient descent (SGD). As found in previous work [11], it is common to use the negative log posterior as the objective, since this is also the expected cross-entropy between the distribution represented by the reference labels and the predicted distribution $y(s)$. The objective is therefore written as:

$$\mathcal{F}_{CE} = -\sum_{u=1}^{U} \sum_{t=1}^{T_u} \log y_{ut}(s_{ut}) \quad (13)$$

where s_{ut} is the reference state label at time t for utterance u. Finally, the necessary gradient is:

$$\frac{\partial F_{CE}}{\partial a_{ut}(s)} = -\frac{\partial \log y_{ut}(s_{ut})}{\partial a_{ut}(s)} = y_{ut}(s) - \delta_{s;s_{ut}} \quad (14)$$

where $\delta_{s;s_{ut}}$ is the Kronecker delta function. In our experiments we use a deep network with 4 hidden layers, each with 1200 hidden nodes. The input layer is composed from a context window of 9 MFCC frames, while the output of the network are the 2000 triphone tied-states that are force-aligned using the conventional HMM-GMM system.

4. EXPERIMENT SETUP

The data used in the REVERB challenge [1] is based on the noisy Multi-Channel Wall Street Journal Audio-Visual (MC-WSJ-AV) [12] corpus, which is in turn, developed from the Wall Street Journal text corpus [13]. The baseline HMM recognizer is developed using HTK [14] and the scripts provided for the challenge. The SGMM and DNN recognizers are developed using Kaldi [15].

Our multi-condition accoustic models are trained using the provided scripts, including CMLLR adapation, without any additional training data. The 8-channel test data is beamformed as described in the front-end selection into a single channel and used for testing.

5. RESULTS AND DISCUSSION

5.1. HMM recognizer

The results shown in Table 1 summarizes the results for the HMM recognizer using different front-ends. As the reverberation time of the simulated rooms increases from rooms one to three, the Word Error Rates (WER) is shown to increase. Based on this observation, it is likely that the reverberation time of the RealData room is higher than the 0.7s of the large-size simulated room (Room 3). The results shown for the fixed beamformer (FBF) refers to the simple mean of the 8-channel audio given by Equation (1).

Of the four alternate front-ends tested, the PBF system is the only one to outperform the baseline in all three simulated rooms. The aggressive adaptive filtering used in the GSC algorithm was calibrated to compensate for the increased noise

	Beamform	SimData							RealData		
		Near1	Far1	Near2	Far2	Near3	Far3	**Avg.**	Near	Far	**Avg.**
HTK	Baseline	16.76	18.35	20.78	32.75	24.79	39.96	25.55	49.92	47.54	48.73
	FBF only	16.66	17.35	16.88	23.35	18.94	27.95	20.18	37.11	38.15	37.63
	GSC only	17.50	18.65	17.14	21.93	18.84	27.71	20.29	37.66	37.44	37.55
	PBF only	16.59	17.64	16.93	23.01	19.01	27.69	20.14	36.92	37.85	37.38
	Proposed	17.43	18.67	17.14	22.03	18.89	27.52	20.27	35.68	36.66	36.17
Kaldi DNN		13.17	12.52	11.00	13.33	12.63	17.54	13.36	32.32	33.56	32.94
Kaldi SGMM		11.52	11.44	9.66	12.56	11.09	15.25	11.92	28.68	30.89	29.79

Table 1. Comparison of baseline system, systems using only the Fixed beamformer (FBF), Phase-error based filter (PBF) and Generalized Sidelobe Canceller (GSC), and the proposed (PBF-GSC) system on the evaluation set

present in the RealData environment. There is a mismatch with the short (0.25s) reverberation time of the first room resulting in the two systems using GSC to perform worse than the baseline. In the noiser rooms, the GSC systems are shown to fare better.

Individually, the FBF, GSC and PBF systems are shown to give an error rate around 37% in the RealData environment. The only difference between the FBF and GSC systems is the application of the adaptive LMS filter. The net effect of the adaptive filtering is to improve the far-condition WER at the expense of the near-condition WER, suggesting that the adaptive filter is overcompensating for the far-field condition. The average WER suggests that using the FBF alone instead of the computation-intensive GSC algorithm is sufficient since the adaptive filter only provides a marginal WER improvement. The PBF performance is a strict improvement over the FBF, again suggesting that beamforming in the form of Equation (1) is the main contributor to the improved performance of the beamforming methods studied here.

Applying the adaptive LMS filter from the GSC algorithm to the PBF output in our proposed PBF-GSC system is shown to lower the WER further to 36%. There is an absolute WER reduction of 1% in both near and far-field conditions in the combined system, suggesting that the adaptive LMS filter is more effective when a better reference signal y_f is provided. This suggests that it might be possible to further refine the system by calibrating the PBF output via its hyperparameters γ and k to better accommodate the adaptive LMS filter.

Strictly from the viewpoint of computation efficiency, the simple FBF of Equation (1) is probably the best solution as it requires no calculations beyond the simple mean of the individual channels. It is also easier to extend to different number of audio channels. In contrast, the PBF requires computations between each channel pair while the GSC requires an estimation of each additional sidelobe.

An alternate method for improving the performance is to increase the number of audio channels. Scaling the beamforming setup from 2 to 8 channels, we found that increasing the number of channels correlates to a reduction in WER. There is likely to be an upper bound to the possible improvements derived by this method and the need for additional hardware might not be feasible depending on the actual application of the system.

5.2. SGMM and DNN recognizer

The results in Table 1 also compare the performance for the proposed PBF-GSC beamforming for the baseline HTK against the SGMM and DNN recognizers. Comparing the SGMM and DNN results, it can be seen that the SGMM system consistently outperformed the DNN system. In particular, for the challenging real data, the performance of SGMM for the near and far conditions was 28.68% and 30.89% respectively. This compared well against the 32.32% and 33.56% that was achieved by the DNN system. The reason for this result may be caused by the relatively small amount of multiconditional data available for training, as specified by the competition. This is due to the large number of parameters to be trained in the DNN, which are difficult to optimise without a large amount of training data.

Comparing the results between the HTK and SGMM systems in Table 1, it can be seen that SGMM consistently outperforms the HTK system in each condition. The improvement is typically around a 6% absolute improvement, although the SGMM does particularly well in the simulated rooms with higher reverberation time (Room 3). This should be expected since the SGMM increases the modelling power of the GMM, by using a large UBM to cover the acoustic space and mapping this to a more specific subspace for each of the HMM states.

6. SPEECH ENHANCEMENT

For the speech enhancement task, we simply apply a previous noise reduction method [3] on the output of PBF-GSC beamforming, described in paragraph 2.3. The basic idea of the method proposed in [3] is to fit the parameterized distributions

	Measure	SimData							RealData		
		Near1	Far1	Near2	Far2	Near3	Far3	**Avg.**	Near	Far	**Avg.**
Single-channel	Cepst Distance	3.30	3.71	5.01	5.58	4.76	5.35	4.62	-	-	-
	Loglike Ratio	0.58	0.64	0.79	1.00	1.00	1.12	0.86	-	-	-
	Seg SNR	6.61	6.16	4.23	3.25	4.19	4.19	4.58	-	-	-
	SRMR	5.3	5.72	6.26	5.37	5.15	4.42	5.37	7.97	7.65	7.81
8-channel	Cepst Distance	2.15	2.48	2.75	3.88	2.88	3.94	3.01	-	-	-
	Loglike Ratio	0.26	0.31	0.39	0.57	0.48	0.67	0.45	-	-	-
	Seg SNR	11.08	10.1	6.84	5.76	7.35	4.53	7.61	-	-	-
	SRMR	4.09	4.15	3.5	3.67	3.90	3.21	3.75	3.67	3.79	3.73

Table 2. Results for the speech enhancement task for both the 8-channel and single-channel scenarios using the proposed generalized gamma distribution model of the speech spectral magnitude.

of speech and noise spectrum at each time frame and then employ the statistical estimation such as MMSE or MAP to estimate the clean speech spectral magnitude before resynthesize its waveform. Particularly, the generalized gamma distribution was proposed to model the speech spectral magnitude $|S|$

$$p\left(|S|\right) = \frac{b^a}{\Gamma\left(a\right)\sigma_S}\left(\frac{|S|}{\sigma_S}\right)^{La-1}\exp\left[-b\left(\frac{|S|}{\sigma_S}\right)^{L}\right], \qquad (15)$$

where σ_S^2 denotes the variance of speech spectrum and (a, b, L) are distribution parameters, while the Gaussian distribution is used to model the noise complex spectrum distributions. For the given task, we employ a special case of [3], with $L = 2, a = 1, b = 5$ and using MAP estimator, which yields a closed form solution for the spectral magnitude estimation as

$$G = \frac{1}{2\left(1+\frac{b}{\xi}\right)} + \sqrt{\frac{1}{4\left(1+\frac{b}{\xi}\right)^2} + \frac{4a-3}{4\gamma\left(1+\frac{b}{\xi}\right)}}, \qquad (16)$$

where ξ and γ denotes the prior and posterior SNR estimations at each time-frequency index [3].

Table 2 reports the evaluation results of proposed method for using 8-channel and single channel, respectively. At the multi-channel scenario, the noise reduction is applied on PBF-GSC output while at the single channel case, only the noisy channel number 1 was processed. Since the PESQ evaluation was not performed, it is hard to make any conclusion on the results. We just note that, the SRMR measurements could be tuned to getting very high value when changing the modeled distribution parameter a and b but it may not results in a better quality of the enhanced signals.

7. CONCLUSION

In this paper, we addressed the issue of reverberant noise using beamforming which combines multiple audio streams into a single stream that is less corrupted by noise compared to its individual components. The main objective of this study is to search for a viable solution to reverberant speech in realistic environments for actual ASR systems thus we have focused on using multi-condition training and the RealData set. Our final system combines Generalized Sidelobe Canceller (GSC) with Phase-error based filtering (PBF) into a PBF-GSC beamformer that operates on 8-channels of audio to generate a single enhanced signal. Although the PBF-GSC system is calibrated based on the results from the baseline MFCC-HMM recognizer built using HTK, it is found to perform well on the more advanced SGMM and DNN recognizers developed using Kaldi.

8. REFERENCES

[1] K. Kinoshita, M. Delcroix, T. Yoshioka, T. Nakatani, E. Habets, R. Haeb-Umbach, V. Leutnant, A. Sehr, W. Kellermann, R. Maas, S. Gannot, and B. Raj, “The reverb challenge: A common evaluation framework for dereverberation and recognition of reverberant speech,” in *Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, 2013.

[2] B.D. Van Veen and K.M. Buckley, “Beamforming: a versatile approach to spatial filtering,” *ASSP Magazine, IEEE*, vol. 5, no. 2, pp. 4–24, 1988.

[3] Tran Huy Dat, Kazuya Takeda, and Fumitada Itakura, “Gamma modeling of speech power and its on-line estimation for statistical speech enhancement,” *IEICE - Trans. Inf. Syst.*, vol. E89-D, no. 3, pp. 1040–1049, Mar. 2006.

[4] Hynek Hermansky and Nelson Morgan, “Rasta processing of speech,” *IEEE Trans. Speech Audio Process.*, vol. 2, no. 4, pp. 578–589, oct 1994.

[5] "Etsi es 202 212, speech processing, transmission and quality aspects (stq); distributed speech recognition; extended advanced front-end feature extraction algorithm; compression algorithms; back-end speech reconstruction algorithm," 2003.

[6] Chanwoo Kim and R.M. Stern, "Power-normalized cepstral coefficients (pncc) for robust speech recognition," in *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on*, 2012, pp. 4101–4104.

[7] C. Knapp and G. Clifford Carter, "The generalized correlation method for estimation of time delay," *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol. 24, no. 4, pp. 320–327, 1976.

[8] Lloyd J. Griffiths and Charles W. Jim, "An alternative approach to linearly constrained adaptive beamforming," *IEEE Trans. Antennas and Propagation*, vol. 30, no. 1, pp. 27–34, Jan. 1982.

[9] C. Y. Lai and P. Aarabi, "Multiple-microphone time-varying filters for robust speech recognition," in *Proc. ICASSP*, 2004, vol. 1, pp. 233–236.

[10] D. Povey, Luk Burget, Mohit Agarwal, Pinar Akyazi, Feng Kai, Arnab Ghoshal, Ondej Glembek, Nagendra Goel, Martin Karafit, Ariya Rastrow, Richard C. Rose, Petr Schwarz, and Samuel Thomas, "The subspace gaussian mixture modela structured model for speech recognition," *Computer Speech & Language*, vol. 25, no. 2, pp. 404 – 439, 2011.

[11] Karel Veselỳ, Arnab Ghoshal, Lukáš Burget, and Daniel Povey, "Sequence-discriminative training of deep neural networks," *Interspeech*, 2013.

[12] M. Lincoln, I. McCowan, J. Vepa, and H.K Maganti, "The multi-channel wall street journal audio visual corpus (mc-wsj-av): Specification and initial experiments," in *Proceedings of the 2005 IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU-05)*, 2005, pp. 357–362.

[13] T. Robinson, J. Fransen, D. Pye, J. Foote, and S. Renals, "Wsjcamo: a british english speech corpus for large vocabulary continuous speech recognition," in *Proceedings of the 1995 International Conference on Acoustics, Speech, and Signal Processing (ICASSP-95)*, 1995, vol. 1, pp. 81–84.

[14] S. J. Young, G. Evermann, M. J. F. Gales, T. Hain, D. Kershaw, G. Moore, J. Odell, D. Ollason, D. Povey, V. Valtchev, and P. C. Woodland, *The HTK Book, version 3.4*, Cambridge University Engineering Department, 2006.

[15] Daniel Povey, Arnab Ghoshal, Gilles Boulianne, Lukas Burget, Ondrej Glembek, Nagendra Goel, Mirko Hannemann, Petr Motlicek, Yanmin Qian, Petr Schwarz, Jan Silovsky, Georg Stemmer, and Karel Vesely, "The kaldi speech recognition toolkit," in *IEEE 2011 Workshop on Automatic Speech Recognition and Understanding*. Dec. 2011, IEEE Signal Processing Society.