

# MAXIMUM A POSTERIORI ESTIMATION OF SIGNAL RANK

Songsri Sirianunpiboon<sup>1</sup> Stephen D. Howard<sup>1</sup>, Douglas Cochran<sup>2</sup>

<sup>1</sup>Defence Science and Technology Organisation  
PO Box 1500, Edinburgh 5111, Australia

<sup>2</sup>School of Mathematical and Statistical Sciences  
Arizona State University, Tempe AZ 85287-5706 USA

## ABSTRACT

Estimating the number of linearly independent signals impinging on a collection of receivers is of increasing recent interest, particularly in connection with MIMO systems for communications and sensing. This paper derives maximum *a posteriori* estimators for signal rank from noisy data collected at multiple receivers. Situations in which noise variance is known and unknown are both treated. These estimators are shown to significantly outperform maximum-likelihood/BIC rank estimators.

**Index Terms**— Bayesian detection, MAP estimation, rank estimation, multi-channel sensing, MIMO communications.

## 1. INTRODUCTION

With the rise of transmit-diverse techniques in communications and sensing, rank is becoming an increasingly important descriptive parameter in detection and characterization of received signals. In [1], an algorithm is presented for estimating the number of signal sources from data received at multiple sensors with motivation drawn from cognitive radio applications. Cognitive radio is also the motivating application in [2, 3], where the focus is on detecting a signal of known rank in multiple receiver channels. The application context in [4, 5, 6] is less specific, accommodating multi-channel detection of signals having known rank in, for example, surveillance scenarios as well as communications.

It is important to note that most recent work on rank estimation uses one of two alternative signal models. The model used in [1, 2, 3, 5] takes the received signal to be gaussian with a rank- $K$  covariance matrix, so that the problem becomes one of testing to discriminate between different covariance structures. In this paper, and in [6, 7], the signal is assumed to occupy a fixed but unknown  $K$ -dimensional subspace. Although the second model can be applied to spectrum sensing in communications as an alternative to the first, it is necessary when addressing problems such as, for example, detection and parameter estimation of multi-element (MIMO) radar transmitters in electronic support and target detection using multi-static passive radar. Work presented in [7] uses this model when examining detection of radar emissions having specific rank, where it introduces a rank estimator for the purpose of characterizing MIMO radar transmitters by the rank of their emissions. This model is also used in [8], where Bayesian detectors based on comparison of signal rank were developed for use in multi-source, multi-static passive radar.

The primary objective of this paper is to derive the posterior distribution for signal rank based on minimally informative priors for the other signal parameters. Based on this posterior distribution a

maximum *a posteriori* (MAP) rank estimator is obtained. As mentioned above, the results presented here are anticipated to be relevant for both communications and sensing applications.

The remainder of the paper is organized as follows. Section 2 presents a precise mathematical description of the signal and measurement models. In this context, it sets forth the rank estimation problem addressed in the following sections. Section 3 gives maximum-likelihood (ML) solutions for this problem for the cases in which the noise variance is known and unknown, respectively. The need to regularize this estimator by, for example, the Bayesian Information Criterion (BIC) is also discussed. Section 4, which is the heart of this contribution, derives the posterior distribution of signal rank under minimally informative priors. This distribution supports MAP estimation of rank. Numerical results are summarized in Section 5, followed by concluding remarks in Section 6.

## 2. MODEL AND PROBLEM FORMULATION

The rank estimation problem introduced above is formulated precisely as follows. A signal of unknown rank  $K \geq 1$  is received at a set of  $M > K$  spatially distributed sensors. The sensor channels are suitably sampled (and possibly time aligned and Doppler compensated, depending on the specific scenario) to obtain  $M$  complex data vectors, each of length  $N$ . These are organized into a  $M \times N$  data matrix  $X$ , each element  $x_{mn}$  of which represents a sample of the noisy signal collected at the  $m^{\text{th}}$  sensor channel at time  $n$ .

The model of the data  $X$  from which  $K$  is to be estimated is

$$X = AS + \nu. \quad (1)$$

The  $K$ -dimensional signal subspace is defined by the matrix  $S \in \mathbb{C}^{K \times N}$ , whose rows are orthonormal vectors in  $\mathbb{C}^N$ .  $A$  is a complex  $M \times K$  matrix whose elements  $a_{mk}$  are the complex amplitudes of the components of the signal vector from sensor  $m$ . Beyond these properties,  $A$  and  $S$  are unknown. The noise matrix  $\nu$  is normally distributed with zero mean, and it is both spatially and temporally white; i.e., its  $MN \times MN$  covariance matrix is  $\sigma^2 I_{MN}$  where  $I_{MN}$  is the  $MN \times MN$  identity matrix.

Without constraints on either  $A$  or  $S$ , the model parameterization just described is redundant. To see this, let  $U$  be any  $K \times K$  unitary matrix. Then defining parameters  $A' = AU$  and  $S' = US$  yields an equivalent problem to the one defined by the original parameters  $A$  and  $S$ . A non-redundant parameterization is obtained by defining parameters  $A \in \mathbb{C}^{M \times K}$  and  $P = S^\dagger S$  where  $\dagger$  denotes hermitian transpose. The matrix  $P$  is an  $N \times N$ , rank- $K$  orthogonal projection matrix; i.e., it satisfies  $P = P^\dagger$ ,  $P^2 = P$  and  $\text{Tr}(P) = K$ . It is possible to associate a unique choice of  $S$  with

each  $P$ . The collection of all rank- $K$  orthogonal projection matrices constitute the Grassmannian  $G_{K,N}$ , which is a smooth complex manifold of complex dimension  $K(N - K)$  [9, 10].

This paper develops a MAP estimator of rank based on the derivation of a posterior distribution for the rank of the signal  $K$ . The multi-hypothesis problem considered is as follows:

$$\begin{aligned} H_0 &: X = \nu \\ H_K &: X = A_K S_K + \nu, \quad \text{for } K = 1, \dots, M-1. \end{aligned}$$

Under  $H_0$ , the joint probability density function (pdf) of  $X$  conditioned on  $\sigma^2$  is

$$p(X|H_0, \sigma^2) = (\pi\sigma^2)^{-MN} e^{-\frac{N}{\sigma^2} \text{Tr}(W)} \quad (2)$$

where  $W = X^\dagger X/N$ . Under  $H_K$  with  $K > 0$ , the joint pdf of  $X$  conditioned on  $A_K, S_K$  and  $\sigma^2$  is

$$\begin{aligned} p(X|H_K, A_K, S_K, \sigma^2) \\ = (\pi\sigma^2)^{-MN} e^{-\frac{N}{\sigma^2} \text{Tr}(W)} e^{-\frac{1}{\sigma^2} \text{Tr}((A_K A_K^\dagger - A_K S_K X^\dagger - X S_K^\dagger A_K^\dagger))}. \end{aligned} \quad (3)$$

In what follows, the notation  $p(X|K)$  to represent  $p(X|H_K)$  and  $p(X|K=0)$  for  $p(X|H_0)$  will be used.

### 3. RANK ESTIMATION BASED ON THE BAYESIAN INFORMATION CRITERION

First consider the situation in which the noise variance  $\sigma^2$  is known. The maximum likelihood estimate of  $K$  is given in [6] as

$$\begin{aligned} \hat{K} &= \arg \min_{K \in \{0, \dots, M-1\}} \min_{A_K, S_K} -\log p(X|H_K) \\ &= \arg \min_{K \in \{0, \dots, M-1\}} \frac{N \text{Tr}(W)}{\sigma^2} \left( 1 - \sum_{j=1}^K \lambda_j / \text{Tr}(W) \right), \end{aligned} \quad (4)$$

where  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_N$  are eigenvalues of  $W$ . Note that the non-zero eigenvalues of  $W$  are exactly the eigenvalues of the sample-covariance matrix  $\hat{R} = X X^\dagger / N$ . An unfortunate consequence of this observation is that, as the hypothesized rank  $K$  is increased, the hypotheses  $H_K$  become increasingly likely and the ML estimate of  $K$  defaults to  $K = M - 1$  irrespective of the data. It is shown below that the Bayesian detector is much better behaved in this regard.

Sensible estimates can be recovered by introducing a penalty function based on one of the so-called information criteria [11, 12]. With this approach, the estimate (4) is replaced by

$$\hat{K} = \arg \min_{K \in \{0, \dots, M-1\}} \min_{A_K, S_K} -\log p(X|H_K) + L(\nu(K), N),$$

where  $L(\nu(K), N)$  is a penalty function that depends on the number of parameters  $\nu(K)$  and the number of samples  $N$ . The penalty function for the Bayesian information criterion (BIC) is

$$L(\nu(K), N) = \frac{\nu(K)}{2} \log N,$$

where  $\nu(K)$  denotes the number of undetermined parameters in the likelihood function. In the model (1),  $A_K$  consists of  $2MK$  real parameters, the Grassmannian  $G_{K,N}$  presents  $2K(N - K)$  real parameters and there is one real parameter for the noise variance  $\sigma^2$  if

it is unknown. Therefore, the total number of undetermined parameters is  $2MK + 2K(N - K) + 1$  in the unknown noise case and  $2MK + 2K(N - K)$  if the noise variance is known. Thus,

$$\begin{aligned} \hat{K} &= \arg \min_{K \in \{0, \dots, M-1\}} \frac{N \text{Tr}(W)}{\sigma^2} \left( 1 - \sum_{j=1}^K \lambda_j / \text{Tr}(W) \right) \\ &\quad + (K(M + N) - K^2) \log N. \end{aligned} \quad (5)$$

If the noise variance is unknown, the BIC estimate is given by

$$\begin{aligned} \hat{K} &= \arg \min_{K \in \{0, \dots, M-1\}} MN \log \left( 1 - \sum_{j=1}^K \lambda_j / \text{Tr}(W) \right) \\ &\quad + (K(M + N) - K^2 + 1/2) \log N. \end{aligned} \quad (6)$$

### 4. POSTERIOR DISTRIBUTION FOR RANK

In [6], both Bayesian and generalized likelihood ratio tests for unknown signal of known rank  $K$ , using data collected at  $M$  spatially distributed receivers, were derived. In this section, a posterior distribution for the rank  $K$  given the data  $X$  is derived, thereby enabling a MAP estimator for rank to be defined. The likelihood under  $H_K$  given  $X$  is given by (3). This likelihood function is invariant under the transformations

$$X \rightarrow \mu U X V, \quad A \rightarrow \mu U A, \quad S \rightarrow S V, \quad \text{and} \quad \sigma \rightarrow \mu \sigma \quad (7)$$

where  $U$  and  $V$  are any unitary matrices of dimensions  $M \times M$  and  $N \times N$ , respectively, and  $\mu > 0$ . The priors for the nuisance parameters  $A$  and  $P$ , and for  $\sigma^2$  when it is unknown, are taken to be as non-informative as possible. As such, these priors need to be invariant under the transformations (7). The invariant non-informative prior measure on the space of unknown parameters is

$$dA d\sigma^{-2} d\mu(P),$$

where  $d\mu(P)$  is the normalized invariant measure on  $G_{K,N}$  and  $dA$  is Lebesgue measure in  $\mathbb{C}^{MK}$ . This prior is unfortunately not proper. The approach taken here to address this issue is to introduce proper priors that approach the invariant non-informative prior in an appropriate limiting sense. The prior is taken to have the form

$$\begin{aligned} p(K, A, \sigma^2) dA d\sigma^{-2} d\mu(P) \\ = p(K|\beta^2) p(A|K, \sigma^2, \beta^2) p(\sigma^{-2}|\tau) dA d\sigma^{-2} d\mu(P). \end{aligned}$$

The components of this prior are assigned as follows. The prior for  $A$  is chosen to be

$$p(A|K, \sigma^2, \beta^2) = (\pi\beta^2\sigma^2)^{-MK} e^{-\frac{1}{\beta^2\sigma^2} \text{Tr}(AA^\dagger)},$$

which is invariant and proper, and becomes less informative as  $\beta^2 \rightarrow \infty$ . The prior for  $\sigma^2$ , if it is unknown, is taken as the maximum entropy prior [13], i.e.,

$$p(\sigma^{-2}|\tau) = \tau^M e^{-\tau M \sigma^{-2}},$$

which becomes less informative as  $\tau \rightarrow 0$ . Finally,

$$p(K|\beta^2) = \frac{(1 + \beta^2)^{MK}}{\sum_{K=0}^{M-1} (1 + \beta^2)^{MK}}.$$

The form of this prior for  $K$  ensures that, as the prior for  $A$  becomes less informative, the posterior ratios for any two ranks  $K$  and  $K'$ ,

$p(K|X)/p(K'|X)$  approach a finite non-zero limit as  $\beta^2 \rightarrow \infty$ . Otherwise, as  $\beta^2 \rightarrow \infty$  the hypothesis  $H_K$  with the smallest value of  $K$  would dominate irrespective of the data.

For known  $\sigma^2$ , the marginalized likelihood for  $K = 1, \dots, M - 1$  is given by

$$p(X|K, \sigma^2, \beta^2) = \frac{p(X|K=0, \sigma^2)}{(1 + \beta^2)^{MK}} \int_{G_{K,N}} e^{\frac{N\alpha}{\sigma^2} \text{Tr}(WP)} d\mu(P) \quad (8)$$

where  $p(X|K=0, \sigma^2)$  is as given in (2) and  $\alpha = \beta^2/(1 + \beta^2)$ . The integral in (8) can be approximated using Laplace approximation. Using the method described in [6] for integration over  $G_{K,N}$  with respect to the normalized invariant measure (prior for  $P$ ), (8) becomes

$$p(X|K, \sigma^2, \beta^2) \approx \frac{\eta_K p(X|K=0, \sigma^2)}{(1 + \beta^2)^{MK}} \exp\left(\frac{N\alpha}{\sigma^2} \sum_{i=1}^K \lambda_i\right) \times \prod_{i=1}^K \prod_{j=1}^{N-K} (\lambda_i - \lambda_{K+j} + \frac{\sigma^2}{\alpha})^{-1}.$$

In this expression,

$$\eta_K = \frac{1}{\text{vol}(G_{K,N})} \left(\frac{\pi\sigma^2}{N\alpha}\right)^{K(N-K)}$$

and  $\text{vol}(G_{K,N})$  denotes the volume of the Grassmannian  $G_{K,N}$ ,

$$\text{vol}(G_{K,N}) = \frac{\prod_{n=N-K+1}^N A_{2n-1}}{\prod_{n=1}^K A_{2n-1}},$$

where  $A_n = 2\pi^{n/2}/\Gamma(n/2)$  is the area of the unit sphere in  $\mathbb{R}^n$  and  $\Gamma$  denotes the Gamma function. The posterior distribution for  $K$  is then

$$p(K|X, \sigma^2, \beta^2) = \frac{p(K|\beta^2)p(X|K, \sigma^2, \beta^2)}{\sum_{K=0}^{M-1} p(K|\beta^2)p(X|K, \sigma^2, \beta^2)}.$$

Taking the limit as  $\beta^2 \rightarrow \infty$  yields the posterior distributions for  $K = 1, \dots, M - 1$ , assuming known  $\sigma^2$ , as

$$p(K=0|X, \sigma^2) = C$$

and

$$p(K|X, \sigma^2) = \frac{C}{\text{vol}(G_{K,N})} \left(\frac{\pi\sigma^2}{N}\right)^{K(N-K)} e^{\frac{N}{\sigma^2} \sum_{i=1}^K \lambda_i} \times \prod_{i=1}^K \prod_{j=1}^{N-K} (\lambda_i - \lambda_{K+j} + \sigma^2)^{-1}.$$

The normalization constant  $C$  is defined so that

$$\sum_{K=0}^{M-1} p(K|X, \sigma^2) = 1.$$

When  $\sigma^2$  is unknown, the marginalized likelihoods are

$$p(X|K=0) = \frac{\tau^M \Gamma(p) \text{Tr}(\tilde{W})^{-\ell}}{N^\ell \pi^{MN}}$$

and

$$p(X|K, \beta^2, \tau) = \frac{p(X|K=0)}{(1 + \beta^2)^{MK}} \int_{G_{K,N}} \left(1 - \alpha \frac{\text{Tr}(WP)}{\text{Tr}(\tilde{W})}\right)^{-\ell} d\mu(P), \quad (9)$$

where  $\ell = MN + 1$  and  $\tilde{W} = W + \frac{M\tau}{N^2} I_N$ .

As in the case of known noise variance, the integral (9) can be approximated as in [6] to obtain

$$p(X|K, \beta^2, \tau) \approx \frac{\rho_K p(X|K=0) \gamma^{K(N-K)-p}}{(1 + \beta^2)^{MK}} \times \prod_{i=1}^K \prod_{j=1}^{N-K} (\tilde{\lambda}_i - \tilde{\lambda}_{K+j} + \frac{N\gamma}{p\alpha})^{-1}$$

where  $\gamma = (1 - \alpha \sum_{i=1}^K \tilde{\lambda}_i)$  with  $\tilde{\lambda}_i = \lambda_i/\text{Tr}(\tilde{W})$ , and

$$\rho_K = \frac{1}{\text{vol}(G_{K,N})} \left(\frac{\pi}{p\alpha}\right)^{K(N-K)}.$$

In the limit  $\beta^2 \rightarrow \infty$  and  $\tau \rightarrow 0$ , the posteriors with unknown  $\sigma^2$  are

$$p(K=0|X) = C$$

and

$$p(K|X) = C \frac{1}{\text{vol}(G_{K,N})} \left(\frac{\pi}{p}\right)^{K(N-K)} \gamma^{K(N-K)-p} \times \prod_{i=1}^K \prod_{j=1}^{N-K} (\tilde{\lambda}_i - \tilde{\lambda}_{K+j} + \frac{N\gamma}{p})^{-1}$$

for  $K = 1, \dots, M - 1$ , where in the limit,

$$\tilde{\lambda}_i = \lambda_i/\text{Tr}(W), \quad \gamma = 1 - \sum_{i=1}^K \tilde{\lambda}_i.$$

The MAP estimate of  $K$  is

$$\hat{K} = \arg \max_K p(K|X), \quad (10)$$

for which it is unnecessary to compute the constant  $C$ .

## 5. NUMERICAL RESULTS

In this section, the performance of the MAP rank estimator derived above is evaluated by simulation. Its performance is compared to that obtained using an information criterion approach (BIC,) as discussed in Section 3. Specifically, the performance of the MAP estimator (10) is compared with those of the corresponding BIC formulations (5) and (6). Simulations were carried out for  $M = 6$  antennas,  $N = 128$  samples, and with SNR defined as

$$\text{SNR(dB)} = 10 \log_{10} \frac{\sigma_a^2}{\sigma^2}$$

where the elements of  $A$  were taken to be zero-mean IID gaussian random variables with variance  $\sigma_a^2$ . The simulations were performed over 500,000 realizations with the rank  $K$  randomly generated between 0 and  $M - 1$ ; i.e.,  $K \in \{0, \dots, 5\}$  in each realization. Table 1 gives resulting estimates of the probability  $\Pr(\hat{K} = K|K)$  of the estimated rank given the true rank using the MAP estimator

**Table 1.**  $\Pr(\hat{K}|K)$  for the MAP estimator

$K \backslash \hat{K}$	0	1	2	3	4	5
0	1	0	0	0	0	0
1	0	1	0	0	0	0
2	0	0	1	0	0	0
3	0	0	0.0005	0.9995	0	0
4	0	0	0	0.0277	0.9723	0
5	0	0	0.0001	0.0232	0.4861	0.4906

**Table 2.**  $\Pr(\hat{K}|K)$  for the BIC estimator

$K \backslash \hat{K}$	0	1	2	3	4	5
0	1	0	0	0	0	0
1	0	1	0	0	0	0
2	0	0.0004	0.9996	0	0	0
3	0	0	0.0022	0.9978	0	0
4	0	0	0	0.0058	0.4461	0.5481
5	0	0	0	0	0.0021	0.9979

and assuming unknown noise variance. Table 2 gives corresponding figures for the BIC estimator. The SNR in both tests is 10 dB. In this “confusion matrix” format, the diagonal elements represent the probability of correct estimation of the rank; i.e.,  $\Pr(\hat{K} = K|K)$ .

Figure 1 shows the average probability of error for the MAP and BIC estimators. The figure clearly demonstrates that, in this scenario, the MAP estimator manifests substantial performance improvement over the BIC estimator in both the known and unknown noise variance cases. In fact, the MAP rank estimator with unknown noise variance is roughly comparable in performance to the BIC estimator that exploits known noise variance.

In this example, the BIC estimator for unknown noise variance provides slightly better performance than both the Bayesian estimator for unknown noise variance and the BIC estimator for known noise variance when the SNR is between 4 and 10dB. This is because the BIC estimator for unknown noise variance has a strong bias for maximal rank ( $K = 5$ ), which happens to be correct in this example. This bias is also responsible for plateau in performance at 0.1 probability of error as the SNR increases.

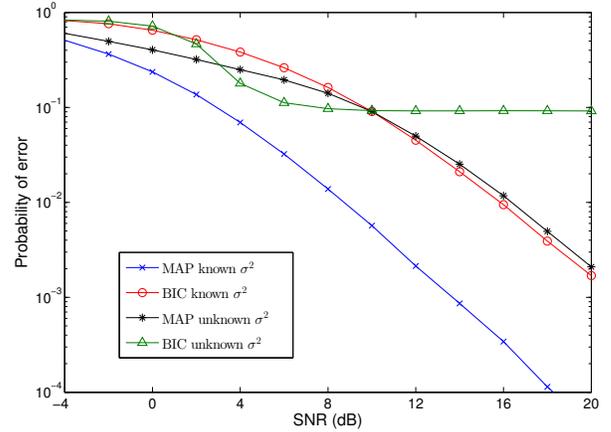
## 6. CONCLUSION

This paper has considered the problem of estimating signal rank from multiple channels of noisy sensor data. This problem is relevant in cognitive radio and is increasingly significant with the growth of MIMO systems in both sensing and communications.

MAP estimators for signal rank were derived for situations in which noise variance is known and unknown. These estimators were shown via simulation to significantly outperform ML/BIC rank estimators in a typical scenario.

## 7. REFERENCES

[1] M. Chiani and M. Z. Win, “Estimating the number of signals observed by multiple sensors,” in *Proceedings of the 2nd International Workshop on Cognitive Information Processing*, 2010, pp. 156–161.

**Fig. 1.** Probability of error versus SNR for the four estimators

[2] D. Ramírez, G. Vazquez-Vilar, R. López-Valcarce, J. Vía, and I. Santamaría, “Multiantenna detection under noise uncertainty and primary user’s spatial structure,” in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, April 2011, pp. 2948–2951.

[3] —, “Detection of rank- $p$  signals in cognitive radio networks with uncalibrated multiple antennas,” *IEEE Transactions on Signal Processing*, vol. 59, no. 8, pp. 3764–3775, 2011.

[4] P. O. Perry and P. J. Wolfe, “Minimax rank estimation for subspace tracking,” *IEEE Transactions on Selected Topics in Signal Processing*, vol. 4, no. 3, pp. 504–513, 2010.

[5] D. Ramírez, J. Iscar, J. Vía, I. Santamaría, and L. L. Scharf, “The locally most powerful invariant test for detecting a rank- $p$  gaussian signal in white noise,” in *Proceedings of the IEEE Sensor Array and Multichannel Signal Processing Workshop*, June 2012.

[6] S. Sirianunpiboon, S. D. Howard, and D. Cochran, “Multiple-channel detection of signals having known rank,” in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, May 2013, pp. 6536–6540.

[7] S. D. Howard, S. Sirianunpiboon, and D. Cochran, “Detection and characterization of MIMO radar signals,” in *Proceedings of the International Conference on Radar*, September 2013.

[8] S. D. Howard and S. Sirianunpiboon, “Passive radar detection using multiple transmitters,” in *Proceedings of the 47th Asilomar Conference on Signals, Systems, and Computers*, November 2013.

[9] J. Harris, *Algebraic Geometry, A First Course*, ser. Graduate Texts in Mathematics. Springer, 1992.

[10] L. Nicolaescu, *An Invitation to Morse Theory*. Universitext, Springer, 2007.

[11] H. Akaike, “A new look at the statistical model identification,” *IEEE Transactions on Automatic Control*, vol. 19, no. 6, pp. 716–723, 1974.

[12] M. Wax and T. Kailath, “Detection of signals by information theoretic criteria,” in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, April 1985, pp. 387–392.

- [13] S. Sirianunpiboon, S. D. Howard, and D. Cochran, "A Bayesian derivation of the generalized coherence detectors," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, March 2012, pp. 3253–3256.