

# SHIFT-INVARIANT SPARSE CODING FOR SINGLE CHANNEL BLIND SOURCE SEPARATION

*T. Blumensath, M. Davies*

Queen Mary, University of London  
Department of Electronic Engineering  
Mile End Road, London E1 4NS, UK

## ABSTRACT

In this paper we present results on single channel blind source separation based on a shift-invariant sparse coding model [1], [2] and [3]. This model learns a set of time-domain features from a single observation of the mixed signals. The found features can often be associated with a single source and can therefore be used to reconstruct the individual source signals. This is shown in this paper on two real world examples, the separation of fetal and maternal heartbeats from a single ECG recording and the separation of singing and accompanying guitar from a musical recording. In the first problem we learn two features, one representing the fetal heartbeat and one representing the maternal heartbeat. In the second example we learn a much larger set to model the more complex source signals and therefore introduce a clustering method to associate the different features with each of the sources.

## 1. INTRODUCTION

Single channel source separation is a difficult problem. The separation of different sources from a single channel observation requires the specification of source models. Often we are not able to exactly specify all model parameters a priori and are only able to specify a class of models. In this paper we address the problem of single channel source separation based on linear additive observation models. We assume that each source can be represented by a linear mixture of a set of time-domain features (possibly by a single feature) and that the observation is a linear mixture of the source signals. The source model can be written mathematically as:

$$\mathbf{x}_i = \sum_{k \in \mathcal{K}_i} \mathbf{a}_k s_{k,i},$$

where  $\mathbf{x}_i$  is the  $i^{\text{th}}$  source signal,  $\mathcal{K}_i$  is the set of indices of those features  $\mathbf{a}_k$  associated with the  $i^{\text{th}}$  source and  $s_{k,i}$  is the scalar coefficient that determines how much feature  $\mathbf{a}_k$  contributes to the source signal. The observation is modelled as a linear mixture of the sources:

$$\mathbf{x} = \sum_i \mathbf{x}_i + \epsilon = \sum_i \sum_{k \in \mathcal{K}_i} \mathbf{a}_k s_{k,i} + \epsilon,$$

where  $\epsilon$  is a vector of i.i.d. Gaussian noise.

For time-series, features can generally occur at arbitrary time location. If we define  $\mathbf{a}_k = \mathbf{a}_{k,0}$  to be the feature starting at the beginning of the observation block  $\mathbf{x}$ , then we can denote a shifted

feature by  $\mathbf{a}_{k,l}$ . For nonnegative values of  $l$  the feature starts at a later sample in the observation block  $\mathbf{x}$  while for negative  $l$  the feature starts before the current observation block  $\mathbf{x}$ . i.e. if we use  $a_{k,p}$  to denote the  $p^{\text{th}}$  sample of feature  $\mathbf{a}_k$ , the feature only contributes to the current observation with the samples for which  $p > -l$ . With this notation the shift-invariant linear source model becomes:

$$\mathbf{x}_i = \sum_{k \in \mathcal{K}_i} \sum_{l \in \mathcal{L}} \mathbf{a}_{k,l} s_{k,l,i},$$

where  $\mathcal{L}$  is the set of all possible feature shifts. The observation model is then:

$$\mathbf{x} = \sum_i \mathbf{x}_i + \epsilon = \sum_i \sum_{k \in \mathcal{K}_i} \sum_{l \in \mathcal{L}} \mathbf{a}_{k,l} s_{k,l,i} + \epsilon, \quad (1)$$

or, written more compactly in matrix notation,  $\mathbf{x} = \mathbf{A}\mathbf{s} + \epsilon$ .

For the general single channel blind source separation problem, both the set of features  $\{\mathbf{a}_k\}$  as well as the associated coefficients  $s_{k,l,i}$  are unknown. Furthermore, we do not know the sets  $\mathcal{K}_i$  required to assign the features to each source<sup>1</sup>. The first problem is to learn or adapt the features  $\mathbf{a}_k$  for a given set of observations. Once a set of features has been found, the second problem is to decompose the observation into a linear combination of those features, i.e. to estimate the coefficients  $s_{k,l,i}$ . Thirdly, if each source is represented by more than a single feature, the learned features need to be clustered into groups, i.e. we need to find the sets  $\mathcal{K}_i$ . In the next three sections we discuss possible solutions to these three problems before we present some experimental results in section 5.

## 2. LEARNING THE MODEL PARAMETERS

For a feature  $\mathbf{a}_k$  of length  $L$  and an observation  $\mathbf{x} \in \mathbb{R}^N$ , the set  $\mathcal{L}$  contains  $N + L - 1$  indices. In particular  $\mathcal{L} = \{-(L-1), -(L-2), \dots, -1, 0, 1, \dots, N-1\}$ . If we label the number of different features by  $K = |\cup_i \mathcal{K}_i|$ , then the number of terms in the summations in equation (1) is  $K(N + L - 1)$ . This means that the model is overcomplete. In order to find unique solutions for overcomplete systems, additional constraints are required. Two very strong constraints have been proposed to solve such overcomplete systems, the constraints of sparsity and non-negativity of the coefficients  $s_{k,l,i}$ .

<sup>1</sup>Generally, the number of sources is also unknown, but in this paper we do not deal with methods to estimate the number of sources.

## 2.1. Bayesian formulation

In order to solve the three problems stated above and to incorporate the sparseness and non-negativity constraints, we take a Bayesian approach. In the above model we have already specified the noise model  $p(\epsilon) \sim \mathcal{N}(0, \sigma_\epsilon \mathbf{I})$ . This defines the Gaussian likelihood  $p(\mathbf{x}|\mathbf{A}, \mathbf{s}) \sim \mathcal{N}(\mathbf{A}\mathbf{s}, \sigma_\epsilon \mathbf{I})$ . Sparsity of the coefficients  $\mathbf{s}$  can now be enforced by introducing the factorial prior  $p(\mathbf{s}) = \prod p(s_{k,l,i})$ , with  $p(s_{k,l,i})$  being a distribution with much of its probability mass concentrated at or around zero. If we require the coefficients  $s_{k,l,i}$  to be non-negative, we can simply restrict these distributions to non-negative values and change the normalising constant appropriately. Different prior formulations are possible (see for example [4], [5], [6] and [3]), with certain classes of priors requiring different learning algorithms.

## 2.2. The learning problem

In order to learn the features  $\mathbf{a}_k$  we can maximise the marginal posterior  $p(\mathbf{A}|\{\mathbf{x}\})$ , i.e. we can find the maximum of the posterior for the set of all available observations  $\{\mathbf{x}\}$ . If we assume a relatively flat prior  $p(\mathbf{A})$  we can instead maximise the marginalised likelihood

$$p(\{\mathbf{x}\}|\mathbf{A}) \propto \int p(\mathbf{x}|\mathbf{A}, \mathbf{s})p(\mathbf{s}) ds.$$

Unfortunately, for the priors  $p(\mathbf{s})$  of interest, this integral cannot be maximised analytically and approximations are required. One approach to maximise the marginalised likelihood can be based on a stochastic gradient descent optimisation strategy, which in each iteration requires the approximation of the following gradient (see [7] and [3] for the derivation):

$$\Delta a_{k,p} = \sigma_\epsilon^{-1} \int \sum_m \epsilon_m s_{k,p-m} p(\mathbf{s}|\hat{\mathbf{A}}, \mathbf{x}) ds,$$

where  $a_{k,p}$  is again the  $p^{\text{th}}$  element of feature  $\mathbf{a}_k$ .

Different methods have been proposed to approximate the above gradient; approximation of  $p(\mathbf{s}|\mathbf{x}, \mathbf{A})$  by a delta function [8] or a Gaussian [9], by importance sampling Monte Carlo [10] or by Gibbs sampling Monte Carlo [11].

## 3. FINDING THE SPARSE REPRESENTATION

In each iteration of the stochastic gradient procedure we need to either find the MAP estimate of  $p(\mathbf{s}|\mathbf{x}, \mathbf{A})$  or draw samples from  $p(\mathbf{s}|\mathbf{x}, \mathbf{A})$ . Furthermore, once the features have converged, it is generally required to calculate an estimate of the coefficients  $s_{k,l,i}$ . Again, different strategies are possible, gradient descent [9], annealing [11] or other sample based estimates such as sample mean and sample mode. As these approaches are discussed in detail in the cited references, they are not further explained here.

The use of the index  $i$  as a subscript of  $s_{k,l,i}$  in equation (1) allows for the same feature  $\mathbf{a}_{k,l}$  to contribute to different sources  $\mathbf{x}_i$  with different strengths. However, the methods mentioned above only determine the strength with which each feature contributes to the observed mixture and are not able to distinguish between individual sources. The clustering method proposed in the next section on the other hand uses global information to assign all occurrences of a feature to a single source independently from the local context in which this feature occurs. We therefore assume that each feature contributes only to a single source, i.e. we assume  $\cap_i \mathcal{K}_i = \emptyset$ .

## 4. CLUSTERING

To determine the sets  $\mathcal{K}_i$ , that is the sets of indices of the features for each source, we propose the use of clustering based on two features, one based on structure in the coefficients  $\mathbf{s}$  not exploited in the shift-invariant sparse coding model and one based on the harmonic structure in the features  $\mathbf{a}_k$ .

The first feature is an approximation of the probability of occurrence of a certain feature during a time interval:

$$p_t(\tilde{k}, i) = p(l \in [l_i, l_{i+1} - 1] : s_{kl} \neq 0, k = \tilde{k}),$$

which we approximate as:

$$p_t(\tilde{k}, i) \approx \frac{1}{\sum_i \sum_{l \in [l_i, l_{i+1} - 1]} |s_{\tilde{k}l}|} \sum_{l \in [l_i, l_{i+1} - 1]} |s_{\tilde{k}l}|.$$

Here we not only count the number of coefficients  $\mathbf{s}$  but also take their strength into account, which can be justified by thinking of stronger coefficients as a summation of smaller ‘quantum’ coefficients.

The second feature can be thought of as a probability distribution of the frequencies in a feature with a logarithmically spaced frequency partitioning:

$$p_f(\tilde{k}, i) \propto \sum_{l \in [2^i, 2^{i+1} - 1]} |\tilde{\mathbf{a}}_{\tilde{k}}(l)|^2, \quad (2)$$

where  $\tilde{\mathbf{a}}_{\tilde{k}}$  is the Fourier transform of feature  $\mathbf{a}_{\tilde{k}}$ .

Clustering can then be performed using K-means clustering with the symmetric Kullback-Leiber divergence. A more detailed description of this method can be found in [3].

## 5. EXPERIMENTAL EVALUATION

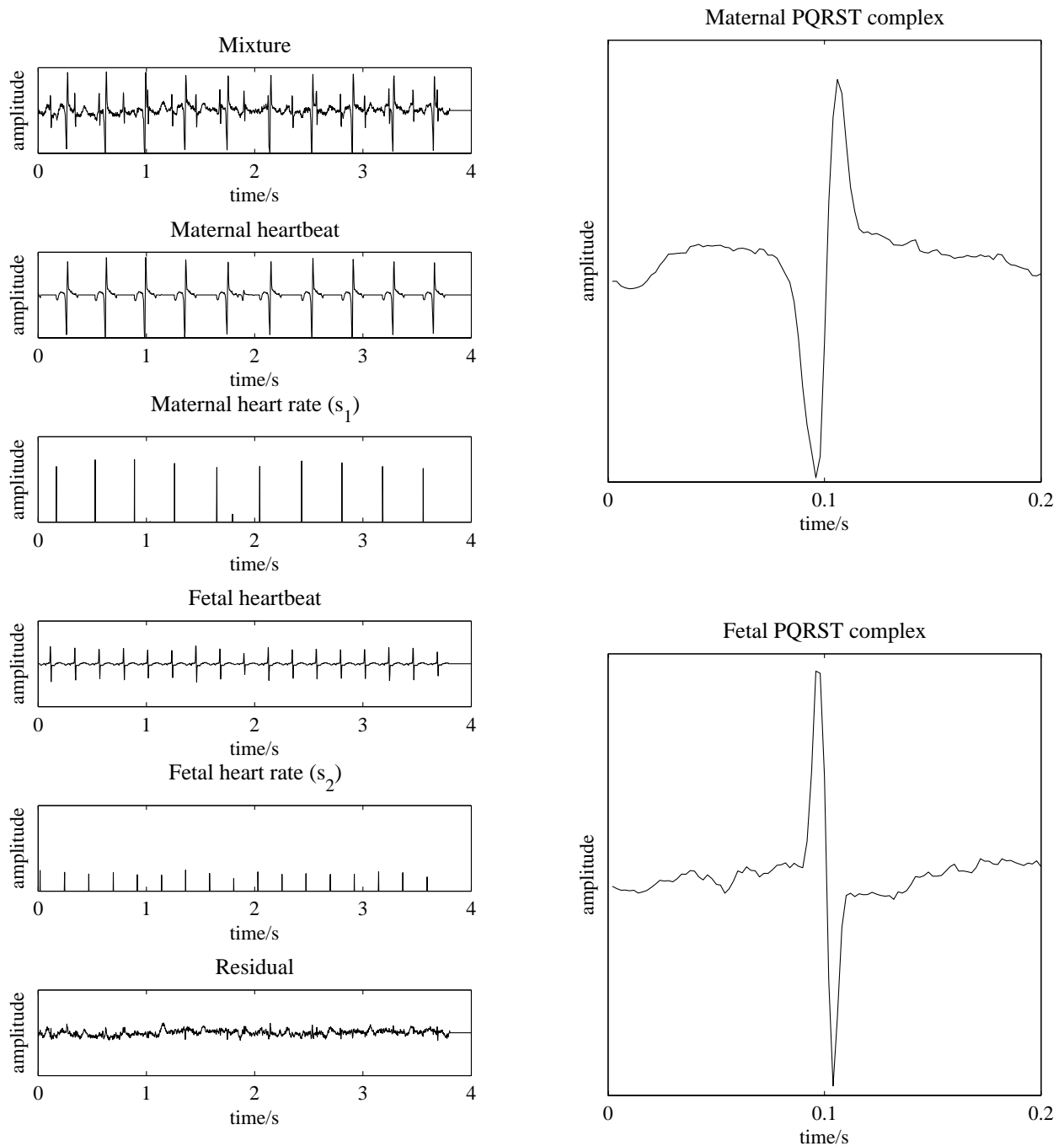
### 5.1. Fetal ECG

In medical diagnostics of cardiac conditions two important characteristics are the rate of the heartbeat and the different properties of the PQRST complex such as the T/QRS ratio [12]. Currently there are no non-invasive techniques available for prenatal diagnostics of fetal heartbeats that directly display these properties. Instead recordings can be taken from the maternal abdomen, but in these recordings the maternal heartbeat is much stronger than the fetal one and the signal to noise ratio is often poor. An example of one such recording is shown in the top left plot of figure 1. Recently ICA techniques have been used to separate the fetal and the maternal heartbeats by using recordings with several sensors [13] [14].

If we assume that each PQRST complex can be characterised by a single time-domain feature, then the shift-invariant sparse coding model can be used to learn these features for the PQRST complexes of the maternal and the fetal heartbeat respectively. The coefficients  $\mathbf{s}$  do then directly encode the heart rates. Furthermore a separation of the fetal and the maternal heartbeats is then possible.

To show this we used the data-set from [13] of multi-sensor recordings, but used the data from the first sensor only. From this data we learned two features of a length of 0.2 seconds with the algorithm as described in [6], which assumes that the probability of the strength of each heartbeat follows a modified Rayleigh distribution and therefore forces the coefficients  $\mathbf{s}$  to be non-negative.

The results of this experiment are shown in figure 1. The panels on the left show from top to bottom: the original signal used



**Fig. 1.** Separation of fetal and maternal heartbeat. On the left we show the original single channel recording, in which the fetal heartbeat is much weaker than the maternal one and in which the SNR is low (top), the separated maternal heartbeat signal (second panel), the heart rate (third panel), the separated fetal heartbeat (fourth panel), the fetal heart rate (fifth panel) and the residual noise (last panel). On the right we show the maternal PQRST complex (top) and the fetal PQRST complex (bottom).

for training, the reconstruction of the maternal heartbeat, the coefficients  $s$  associated with the maternal heartbeat, which encode the maternal heart rate, the fetal heartbeat, the coefficients  $s$  associated with the fetal heartbeat, which encode the fetal heart rate and

finally the residual noise term. On the right of figure 1 we show the PQRST complexes of the maternal (top) and the fetal (bottom) heart. From these results important diagnostic features such as the fetal heart rate or the fetal T/QRS ratio can be easily determined.

**Table 1.** Comparison between the features for clustering.

	$p_f$	$p_t$	$[p_t, p_f]$	Oracle
SIR vocal	11.5	12.6	11.8	15.2
SIR guitar	4.7	9	9.9	7.6
SAR vocal	-0.2	3	3.2	2.6
SAR guitar	3.7	3.3	3	4.0
SDR vocal	-0.8	2.3	2.4	2.3
SDR guitar	0.4	1.8	1.8	1.9

## 5.2. Music

For the ECG example of the previous experiment, each source was represented by a single feature so that source separation could be performed by using a single feature at a time. For more complex signals, more features are required to accurately describe a single source. In this case, a method is required to cluster the features. Such a clustering method was presented above. Here we present an experimental evaluation of this methods for single channel source separation. More details on the clustering method and the reported experiment can be found in [3].

As a test signal we recorded a guitar and vocal performance of the same musical piece. These signals were down-sampled to 8000 Hz and mixed linearly. The resulting mixture was used to learn a set of 500 features of 256 samples each. The exact training algorithm is described in [3]. After 500 000 iterations, 126 of the features had converged to harmonic features. The other features had not been updated significantly and were discarded.

After clustering, the sources can again be reconstructed by using only the features from each cluster. To evaluate the performance we use the signal to interference ratio SIR, the signal to artefact ratio SAR and the signal to distortion ratio SDR defined in [15]. The results for clustering based on the two features are shown in table 1, in which we give the results for clustering based on each of the features as well as on clustering based on the combination of both features. In addition, in the last column we show the results based on an oracle clustering method, in which features were assigned to the source for which this feature occurred most often in the decomposition of the individual sources [3].

## 6. CONCLUSION

In this paper we have proposed the use of a shift-invariant sparse coding model for single channel source separation. Three problems had to be addressed, learning of model parameters, inference of the model states for each observation and clustering of features into sources. We have presented a short review of previous work regarding the first two problems before proposing a novel solution to the third problem. The experimental section gave two practical examples. In the first example we have shown that the shift-invariant sparse coding model can successfully separate maternal and fetal heartbeats. For this problem we only used a single feature to model each source and did therefore not require clustering. In the second experiment we separated the vocal and guitar parts of a music recording. This example was based on source models that used several feature and we found that the proposed clustering method produced results that were close to the results obtained with an oracle clustering approach.

## 7. REFERENCES

- [1] B. A. Olshausen, "Sparse coding of time-varying natural images," in *Proc. of the Int. Conf. on Independent Component Analysis and Blind Source Separation*, 2000.
- [2] M. Plumbley, S. Abdallah, T. Blumensath, and M. Davies, "Sparse representations of polyphonic music," to appear in *EURASIP Signal Processing Journal*.
- [3] T. Blumensath and M. Davies, "Sparse and shift-invariant representations of music," to appear in *IEEE Transactions on Speech and Audio Processing*.
- [4] K. Kreutzer-Delgado, B. D. Rao, and K. Engan, "Convex/shure-convex (CSC) log-priors and sparse coding," in *Proc. of the 6th Joint Symposium on Neural Computation*, (Pasadena, California), May 1999.
- [5] P. Sallee and B. A. Olshausen, "Learning sparse multiscale image representations," in *Advances in Neural Information Processing Systems (NIPS)*, pp. 1327–1334, 2003.
- [6] T. Blumensath and M. Davies, "Enforcing sparsity, shift-invariance and positivity in a Bayesian model of polyphonic music," in *Proc. of the IEEE Workshop on Statistical Signal Processing*, July 2005.
- [7] M. S. Lewicki and B. A. Olshausen, "A probabilistic framework for the adaptation and comparison of image codes," *J. Opt. Soc. Am. A: Optics, Image Science, and Vision*, vol. 16, no. 7, pp. 1587–1601, 1999.
- [8] B. A. Olshausen and D. J. Field, "Emergence of simple-cell receptive field properties by learning a sparse code for natural images," *Nature*, vol. 13, pp. 607–609, Jun 1996.
- [9] M. S. Lewicki and T. J. Sejnowski, "Learning overcomplete representations," *Neural Computation*, no. 12, pp. 337–365, 2000.
- [10] T. Blumensath and M. Davies, "A fast importance sampling algorithm for unsupervised learning of over-complete dictionaries," in *Proc. of the Int. Conf. on Acoustics, Speech and Signal Processing*, pp. 213–216, March 2005.
- [11] B. A. Olshausen and K. Millman, "Learning sparse codes with a mixture-of-gaussians prior," in *Advances in Neural Information Processing Systems (NIPS)*, pp. 841–847, 2000.
- [12] C. Widmark, T. Jansson, K. Lindcrantz, and K. Rosen, "ECG waveform, short term heart rate variability and plasma catecholamine concentrations in response to hypoxia in intrauterine growth related guinea-pig fetuses," *Journal of Developmental Physiology*, vol. 15, no. 3, pp. 161–168, 1991.
- [13] L. De Lathauwer, B. De Moor, and J. Vandewalle, "Fetal electrocardiogram extraction by source subspace separation," in *Proc. IEEE SP / ATHOS Workshop on HOS*, (Girona, Spain), pp. 134–138, June 1995.
- [14] V. Vigneron, A. Paraschiv-Ionescu, A. Azancot, O. Sibony, and C. Jutten, "Fetal electrocardiogram extraction based on non-stationary ICA and wavelet denoising," in *Proceedings of the Seventh International Symposium on Signal Processing and Its Applications*, pp. 69–72, 2003.
- [15] R. Gribonval, L. Benaroya, E. Vincent, and C. Févotte, "Proposals for performance measurement in source separation," Tech. Rep. 1501, Institut de Recherche et Coordination Acoustique/Musique, 2003.