We are IntechOpen,
the world's leading publisher of
Open Access books
Built by scientists, for scientists

6,900
Open access books available

186,000
International authors and editors

200M
Downloads

154
Countries delivered to

Our authors are among the

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

BOOK CITATION INDEX
CLARIVATE ANALYTICS
INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

Sparse signal decomposition
for periodic signal mixtures

Makoto Nakashizuka

Graduate School of Engineering Science, Osaka University
Japan

1. Introduction

Periodicities are found in speech signals, musical rhythms, biomedical signals and machine vibrations. In many signal processing applications, signals are assumed to be periodic or quasi-periodic. Especially in acoustic signal processing, signal models based on periodicities have been studied for speech and audio processing.

The sinusoidal modelling has been proposed to transform an acoustic signal to a sum of sinusoids [1]. In this model, the frequencies of the sinusoids are often assumed to be harmonically related. The fundamental frequency of the set of sinusoids has to be specified for this model. In order to compose an accurate model of an acoustic signal, the noise-robust and accurate fundamental frequency estimation techniques are required. Many fundamental frequency estimation techniques are performed in the short-time Fourier transform (STFT) spectrum by peak-picking and clustering of harmonic components [2][3][4]. These approaches depend on the frequency spectrum of the signal.

The signal modeling in the time-domain has been also proposed to extract a waveform of an acoustic signal and its parameters of the amplitude and frequency variations [5]. This approach aims to represent an acoustic signal that has single fundamental frequency. For detection and estimation of more than one periodic signal hidden in a signal mixture, several signal decomposition that are capable of decomposing a signal into a set of periodic subsignals have been proposed.

In Ref. [7], an orthogonal decomposition method based on periodicity has been proposed. This technique achieves the decomposition of a signal into periodic subsignals that are orthogonal to each other. The periodicity transform [8] decomposes a signal by projecting it onto a set of periodic subspaces. In this method, seeking periodic subspaces and rejecting found periodic subsignals from the observed signal are performed iteratively. For reduction of the redundancy of the periodic representation, a penalty of sparsity has been introduced to the decomposition in Ref. [9].

In these periodic decomposition methods, the amplitude of each periodic signal in the mixture is assumed to be constant. Hence, it is difficult to obtain the significant decomposition results for the mixtures of quasi-periodic signals with time-varying amplitude. In this chapter, we introduce a model for periodic signals with time-varying amplitude into the periodic decomposition [10]. In order to reduce the number of resultant

periodic subsignals obtained by the decomposition and represent the mixture with only significant periodic subsignals, we impose a sparsity penalty on the decomposition. This penalty is defined as the sum of l_2 norms of the resultant periodic subsignals to find the shortest path to the approximation of the mixture. The waveforms and amplitude of the hidden periodic signals are iteratively estimated with the penalty of sparsity. The proposed periodic decomposition can be interpreted as a sparse coding [15] [16] with non-negativity of the amplitude and the periodic structure of signals.

In our approach, the decomposition results are associated with the fundamental frequencies of the source signals in the mixture. So, the pitches of the source signals can be detected from the mixtures by the proposed decomposition.

First, we explain the definition of the model for the periodic signals. Then, the cost function that is a sum of the approximation error and the sparsity penalty is defined for the periodic decomposition. A relaxation algorithm [9] [10] [18] for the sparse periodic decomposition is also explained. The source estimation capability of our decomposition method is demonstrated by several examples of the decomposition of synthetic periodic signal mixtures. Next, we apply the proposed decomposition to speech mixtures and demonstrate the speech separation. In this experiment, the ideal separation performance of the proposed decomposition is compared with the separation method obtained by an ideal binary masking [10] of a STFT. Finally, we provide the results of the single-channel speech separation with simple assignment technique to demonstrate the possibility of the proposed decomposition.

2. Periodic decomposition of signals

For signal analysis, the periodic decomposition methods that decompose a signal into a sum of periodic signals have been proposed. Most fundamental periodic signal is a sinusoid. In speech processing area, the sinusoidal modeling [1] that represents the signal into the linear combination of sinusoids with various frequencies is utilized. The sinusoidal representation of the signal $f(n)$ with constant amplitude and constant frequencies is obtained as the form of

$$f(n) = \sum_{j=1}^{J} A_j \cos(\omega_j n + \phi_j).$$
(1)

This model relies on the estimation of the parameters of the model. Many estimation techniques have been proposed for the parameters. If the frequencies $\{\omega_j\}_{1 \leq j \leq J}$ are harmonically related, all frequencies are assumed to be the multiples of the fundamental frequency. To detect the fundamental frequencies from mixtures of source signals that has periodical nature, multiple pitch detection algorithms have been proposed [2][3][4].

The signal modelling with (1) is a parametric modeling of the signal. On the contrast, the non-parametric modeling techniques that obtain a set of periodic signals that are specified in time-domain have been also proposed.

For time-domain approach of the periodic decomposition, the periodic signal is defined as a sum of time-translated waveforms. Let us suppose that a sequence $\{\mathbf{f}_p(n)\}_{0 \leq n < N}$ is a finite length periodic signal with a length N and an integer period $p \geq 2$. It satisfies the periodicity condition with an integer period p and is represented as

$$f_p(n) = a_p(n) \sum_{k=0}^{K} t_p(n - kp)$$
(1)

where $K = \lfloor (N\text{-}1)/p \rfloor$ that is the largest integer less than or equal to $(N\text{-}1)/p$. The sequence $\{t_p(n)\}_{0 \leq n < p}$ corresponds to a waveform of the signal within a period and is defined over the interval $[0, p\text{-}1]$. $t_p(n) = 0$ for $n \geq p$ and $n < 0$. This sequence is referred to as the p-periodic template. The sequence $\{a(n)\}_{0 \leq n < N}$ represents the envelope of the periodic signal. If the amplitude coefficient $a(n)$ is constant, the model is reduced to

$$f_p(n) = \sum_{k=0}^{K} t_p(n - kp). \tag{2}$$

Several periodic decomposition methods based on the periodic signal model (2) have been proposed [6] [7] [8] [9]. These methods decompose a signal f(n) into a set of the periodic signals as:

$$f(n) = \sum_{p \in P} \sum_{k=0}^{K} t_p(n - kp) \tag{3}$$

where P is a set of periods for the decomposition. This signal decomposition can be represented in the matrix form as:

$$\mathbf{f} = \sum_{p \in P} \mathbf{U}_p \mathbf{t}_p \tag{4}$$

where $\mathbf{t}_p$ is the vector which corresponds to the p-periodic template. The i-th column vector of $\mathbf{A}_p$ represent an impulse train with a period p. The elements of $\mathbf{U}_p$ are defined as

$$u_{n,i} = \begin{cases} 1 & \text{for } n = kp + i - 1 \text{ where } k = 0, 1, \cdots \\ 0 & \text{otherwise} \end{cases}. \tag{5}$$

The subspace that is spanned by the column vectors of $\mathbf{U}_p$ is referred to as the p-periodic subspace [8] [9].

If the estimations of the periods hidden in signal $\mathbf{f}$ are available, we can choose the periodic subspaces with the periods that are estimated before the decomposition. For MAS [6], the signal is decomposed into periodic subsignals as the least-squares solution along with an additional constrained matrix. In Ref. [8], the periodic bases are chosen to decompose a signal into orthogonal periodic subsignals. Therefore, these methods require that the number of the periodic signals and their periods have to be estimated before decomposition. Periodic decomposition methods that do not require predetermined periods have also been proposed. In Ref. [7], the concept of periodicity transform is proposed. Periodicity transform decomposes a signal by projecting it onto a set of periodic subspaces. Each subspace consists of all possible periodic signals with a specific period. In this method, seeking periodic subspaces and rejecting found periodic subsignals from an input signal are performed iteratively. Since a set of the periodic subspaces lacks orthogonality and is redundant for signal representation, the decomposition result depends on the order of the subspaces onto which the signals are projected. In Ref. [7], four different signal decomposition methods - small to large, best correlation, M-best, and best frequency - have been proposed. In Ref. [9], the penalty of sparsity is imposed on the decomposition results in order to reduce the redundancy of the decomposition.

In this chapter, we discuss the decomposition of mixtures of the periodic signals with time-varying amplitude that can be represented in the form of (1). To simplify the periodic signal model, we assume that the amplitude of the periodic signal varies slowly and can be approximated to be constant within a period. By this simplification, we define an approximate model for the periodic signals with time-varying amplitude as

$$f_p(n) = \sum_{k=0}^{K} a_{p,k} t_p(n-kp). \tag{6}$$

In order to represent a periodic component without a DC component, the average of $f_p(n)$ over the interval $[0, p\text{-}1]$ is zero. The amplitude coefficients $a_{p,k}$ are restricted to non-negative values.

These p-periodic signals can also be represented in a matrix form as well as the previous periodic signal model. The matrix representation of (6) is defined as

$$\mathbf{f}_p = \mathbf{A}_p \mathbf{t}_p \tag{7}$$

In this form, the amplitude coefficients and the template are represented in an N by p matrix $\mathbf{A}_p$ and a p-dimensional template vector $\mathbf{t}_p$, which is associated with the sequence $t_{\mathrm{p}}(n)$, respectively. $\mathbf{A}_p$ is a union of the matrices as

$$\mathbf{A}_p = \left(\mathbf{D}_{p,1}, \mathbf{D}_{p,2} \cdots \mathbf{D}_{p,K+1}\right)^{\mathrm{T}} \tag{8}$$

where superscript T denotes transposition.

{ $\mathbf{D}_{p,j}$ }$_{1 \le j \le K+1}$ are p by p diagonal matrices whose elements correspond to $a_{p,j-1}$. $\mathrm{D}_{p,K+1}$ is the p by $N\text{-}pK$ matrix whose non-zero coefficients that correspond to $a_{p,K}$ appear only in (i, i) elements. Since only one element is non-zero in any row of the $\mathbf{A}_p$, the column vectors of $\mathrm{A}p$ are orthogonal to each other. The $l2$ norm of each column vector is supposed to be normalized to unity. In (6), the average of the waveform over the interval $[0, p\text{-}1]$ must be zero. Hence, the condition

$$\mathbf{u}_p^{\mathrm{T}} \mathbf{t}_p = 0 \tag{9}$$

where $\mathbf{u}_p$ is a vector, of which elements correspond to the diagonal elements of $\mathbf{D}_{p,1}$.

Alternatively, the p-periodic signal in (2) can be represented as

$$\mathbf{f}_p = \mathbf{T}_p \mathbf{a}_p . \tag{10}$$

In this form, the amplitude coefficients and the template are represented in a N by $K+1$ matrix $\mathbf{T}_p$ and $K+1$-dimensional amplitude coefficients vector $\mathbf{a}_p$ whose elements are associated with the amplitude coefficients $a_{p,k}$, respectively. $\mathbf{T}_p$ consists of the column vectors that correspond to the shifted versions of the p-periodic template. As same as $\mathbf{A}_p$, only one element is non-zero in any row of $\mathbf{T}_p$. So, we defined $\mathrm{T}p$ as the matrix which consists of the normalized vectors that are orthogonal to each other.

In this study, we propose an approximate decomposition method that obtains a representation of a given signal $\mathbf{f}$ as a form:

$$\mathbf{f} = \mathbf{e} + \sum_{p \in \mathrm{P}} \mathbf{f}_p \tag{11}$$

where $\mathbf{e}$ is an approximation error between the model and the signal $\mathbf{f}$.

We suppose that the signal $\mathbf{f}$ is a mixture of some periodic signals that can be approximated by the form of (2), however, the periods of the source signals are unknown. So, we specify the set of periods P as a set of all possible periods of the source signals for the decomposition. If the number of the periods in P is large, the set of the periodic signals {$\mathbf{f}_p$ }$_{p \in \mathrm{P}}$ that approximate the signal $\mathbf{f}$ with small error is not unique. To achieve the significant decomposition with the periodic signals that are represented in the form of (2), we introduce the penalty of the sparsity into the decomposition.

3. Sparse decomposition of signals

In Ref. [15] [16] [17], sparse decomposition methods that are capableof decomposing a signal into a small number of basis vectors that belong to an overcomplete dictionary have been proposed. Basis pursuit (BP) [17] is a well known sparse decomposition method and decomposes a signal into the vectors of a predetermined overcomplete dictionary. The signal $\mathbf{f}$ is represented as $\mathbf{\Phi c}$, where $\mathbf{\Phi}$ and $\mathbf{c}$ are the matrix that contains the normalized basis vectors and the coefficient vector, respectively.

In sparse decomposition, the number of basis vectors in $\mathbf{\Phi}$ is larger than the dimensionality of the signal vector $\mathbf{f}$. For this decomposition, the penalty of the sparsity is defined as l_1-norm of $\mathbf{c}$. The signal decomposition by BP is represented as a constrained minimization problem as follows:

$$\min\|\mathbf{c}\|_1 \text{ subject to } \mathbf{f} = \mathbf{\Phi c} \tag{12}$$

where $\|\cdot\|_1$ denotes the l_1 norm of a vector.

Since the l_1-norm is defined as the sum of the absolutes of the elements in the coefficient vector $\mathbf{c}$, BP determines the shortest path to the signal from the origin through the basis vectors. The number of the basis vectors with nonzero coefficients obtained by choosing the shortest path is much smaller than the least square solution obtained by minimizing the l_2-norm [17].

Usually, (12) is solved by linear programming [17]. However, it is difficult to apply linear programming to the large number of samples that appear in signal processing applications. So, an approximation of the solution of BP is obtained from the penalty problem of (12) as follows:

$$\hat{\mathbf{c}} = \arg\min_{\mathbf{c}} \frac{1}{2}\|\mathbf{f} - \mathbf{\Phi c}\|_2^2 + \lambda\|\mathbf{c}\|_1 \tag{13}$$

where λ denotes a Lagrange multiplier. $\|\cdot\|_2$ denotes the l_2 norm of the vector. This unconstrained minimization problem is referred to as a basis pursuit denoising (BPDN) [17] [18]. When $\mathbf{\Phi}$ is specified as a union of orthonormal bases, an efficient relaxation algorithm can be applied [18].

From Bayesian point of view, the minimization (13) is the equivalent of MAP estimation of the coefficient vector $\mathbf{c}$ under the assumption that the probability distribution of each element of the coefficient vector is an identical Laplace distribution [15].

The dictionary $\mathbf{\Phi}$ is fixed for signal representation in the BP and BPDN. In a sparse coding strategy [15] [16], the dictionary $\mathbf{\Phi}$ is adapted to the set of the signals. The dictionary is updated with the most probable one under the estimated sparse coefficients and the set of the signals [15].

For our periodic decomposition, we also impose the sparsity penalty on the decomposition under the assumption that the mixture contains a small number of periodic signals that can be approximated in the form of (6). Our objective is to achieve signal decomposition to obtain a small number of periodic subsignals rather than basis vectors. In order to achieve this, we define the sparsity measure as the sum of l_2 norms of the periodic subsignals to find the shortest path to the approximation of the signal as well as BPDN.

4. Sparse periodic decomposition

4. 1 Cost function for periodic decomposition

For our periodic decomposition, we also impose the sparsity penalty on the decomposition under the assumption that the mixture consists of a small number of periodic signals that can be approximated in the form of (2). Our objective is to achieve signal decomposition with a small number of periodic subsignals rather than the basis vectors. In order to achieve this, the probability distribution of the l_2 norm of each periodic signal is assumed to be a Laplace distribution, and then the probability distribution of the set of the periodic signals is

$$P\left(\left\{\mathbf{f}_p\right\}_{p \in \mathrm{P}}\right) \propto \prod_{p \in \mathrm{P}} \exp\left(-\alpha_p \left\|\mathbf{f}_p\right\|\right). \tag{14}$$

The noise is assumed to be Gaussian, and then the conditional probability distribution of $\mathbf{f}$ is

$$P\left(\mathbf{f} \middle| \left\{\mathbf{f}_p\right\}_{p \in P}\right) \propto \exp\left(-\frac{1}{2\lambda}\left\|\mathbf{f} - \sum_{p \in P}\mathbf{f}_p\right\|_2^2\right). \tag{15}$$

Along with Bayes' rule, the conditional probability distribution of the set of the periodic signals is

$$P\left(\left\{\mathbf{f}_p\right\}_{p \in P} \middle| \mathbf{f}\right) \propto P\left(\mathbf{f} \middle| \left\{\mathbf{f}_p\right\}_{p \in P}\right) P\left(\left\{\mathbf{f}_p\right\}_{p \in P}\right). \tag{16}$$

Substituting the prior distributions of the periodic signals and the noise into (16), we can derive the likelihood function of the set of periodic signals. From the likelihood function, we define the cost function E for the periodic decomposition as:

$$E \equiv \frac{1}{2}\left\|\mathbf{f} - \sum_{p \in P}\mathbf{f}_p\right\|_2^2 + \lambda \sum_{p \in P}\alpha_p \left\|\mathbf{f}_p\right\|_2. \tag{17}$$

In our periodic decomposition, a signal $\mathbf{f}$ is decomposed into a set of periodic subsignals while reducing the cost E and maximizing the likelihood.

In the cost for BPDN (12), the sparsity penalty is defined as the l_1-norm of the coefficient vector that is identical the total length of the decomposed vector of the signal. In our periodic decomposition, the sparsity penalty is also defined as the sum of the decomposed vectors that are represented in the form of the periodic signal model shown in (6).

4. 2 Algorithm for sparse periodic decomposition

To find the set of the periodic subsignals $\{\mathbf{f}_p\}_{p \in P}$, we employ a relaxation algorithm. This relaxation algorithm always updates one chosen periodic subsignal while decreasing the cost function (17). The template vector $\mathbf{t}_p$ and amplitude vector $\mathbf{a}_p$ of the chosen period p are alternatively updated in an iteration. In the algorithm, we suppose that the set of the periods P consists of M periods which are indexed as $\{p_1 \dots p_M\}$.

The relaxation algorithm for the sparse periodic decomposition is as follows:

1) Set the initial amplitude coefficients for $\{\mathbf{A}_p\}$.
2) $i = 1$
3) Compute the residual

$$\mathbf{r} = \mathbf{f} - \sum_{j \neq i} \mathbf{f}_{p_j} \qquad (18)$$

4) Represent $\mathbf{f}_{p_i}$ as $\mathbf{A}_{p_i} \mathbf{t}_{p_i}$. If $\|\mathbf{f}_{p_i}\| = 0$, then the amplitude coefficients in $\mathbf{A}_{p_i}$ are specified to be constant. Update the template $\mathbf{t}_{p_i}$ with the solution of a subproblem:

$$\min_{\mathbf{t}_{p_i}} \frac{1}{2} \left\| \mathbf{f} - \mathbf{A}_{p_i} \mathbf{t}_{p_i} \right\|_2^2 + \lambda \alpha_{p_i} \left\| \mathbf{t}_{p_i} \right\|_2 \text{ subject to } \mathbf{u}_{p_i}^T \mathbf{t}_{p_i} = 0 \qquad (19)$$

5) Represent $\mathbf{f}_{p_i}$ as $\mathbf{T}_{p_i} \mathbf{a}_{p_i}$. Update the amplitude coefficient vector $\mathbf{a}_{p_i}$ with the solution of a subproblem:

$$\min_{\mathbf{a}_{p_i}} \frac{1}{2} \left\| \mathbf{f} - \mathbf{T}_{p_i} \mathbf{a}_{p_i} \right\|_2^2 + \lambda \alpha_{p_i} \left\| \mathbf{a}_{p_i} \right\|_2 \text{ subject to } \mathbf{a}_{p_i} \geq 0 \qquad (20)$$

where "$\mathbf{a} \geq 0$" denotes that the all elements of the vector $\mathbf{a}$ is positive.

6) If $i < M$, update $i \leftarrow i + 1$ and go to step 3). If $i = M$ and the stopping criterion is not satisfied, go to step 2).

For stable computation, the update stage of the amplitude coefficient in Step 5) is omitted when the l_2-norm of the template $\mathbf{t}_{p_i}$ becomes zero after Step 4).

The closed form solution of (19) is

$$\hat{\mathbf{t}}_{p_i} = \begin{cases} \dfrac{\|\mathbf{v}\|_2 - \lambda \alpha_{p_i}}{\|\mathbf{v}\|_2} \mathbf{v} & for \quad \|\mathbf{v}\|_2 > \lambda \alpha_{p_i} \\[2ex] 0 & for \quad \|\mathbf{v}\|_2 \leq \lambda \alpha_{p_i} \end{cases} \qquad (21)$$

where

$$\mathbf{v} = \mathbf{A}_{p_i}^T \mathbf{r}_{p_i} - \frac{\mathbf{u}_{p_i}^T \left(\mathbf{A}_{p_i}^T \mathbf{r}_{p_i} \right)}{\left\| \mathbf{u}_{p_i} \right\|_2^2} \mathbf{u}_{p_i} \quad . \qquad (22)$$

The solution of (10) is

$$\hat{\mathbf{a}}_{p_i} = \begin{cases} \dfrac{\|\mathbf{w}\|_2 - \lambda \alpha_{p_i}}{\|\mathbf{w}\|_2} & for \quad \|\mathbf{w}\|_2 > \lambda \alpha_{p_i} \\[2ex] 0 & for \quad \|\mathbf{w}\|_2 \leq \lambda \alpha_{p_i} \end{cases} \qquad (23)$$

where

$$\mathbf{w} = \left(\mathbf{T}_{p_i}^T \mathbf{r}_{p_i} \right)_+ \qquad (24)$$

$(\cdot)_+$ denotes replacing the negative elements of a vector with zero. The both solutions of the subproblems guarantee the decrement of the cost E. Thus, the cost E decreases until convergence. However, the set of the resultant periodic subsignals after the convergence of the iteration does not always obtain a minimum of the cost function E exactly. If any periodic subsignal becomes zero in iteration, the amplitude coefficients are specified to be

constant in step 4) of the next iteration. The proper search direction for $\mathbf{t}_{p_i}$ may not be obtained by these amplitude coefficients. However, the l_2 norms of the periodic signals that eliminated by the shrinkage in (21) and (23) is small enough to approximate the signal. Hence, we accept the periodic subsignals obtained by this algorithm as the result of the sparse decomposition instead of the proper minimiser of the cost E.

Tested set	Ave.	Std. Dev
28, 44, 52	14.6, 16.6, 12.6	2.4, 2.4, 2.2
30, 31, 32	16.9, 21.0, 20.7	3.1, 2.7, 2.7
50, 51, 52	10.8, 12.7, 10.8	1.7, 1.9, 1.7

Table 1. SNR improvements (dB) obtained by the sparse periodic decomposition for mixtures of three periodic signals.

5. Decomposition examples

In this section, we provide several examples of the sparse periodic decomposition. The examples demonstrate the decomposition of synthetic signals generated by adding three periodic signals. The length of the mixture and three source signals N is 256. Each source signal is generated with the model for the periodic signals shown in (1). Each waveform within a period is generated by Gaussian random variables. The average of the waveform of a period is normalized to zero. The amplitude envelope of one of the three source signals are specified as a constant. The envelopes of the other two source signals are specified as a decreasing Gaussian function

$$a(n) = \exp\left(-\left(\frac{2n}{N}\right)^2\right) \text{ for } n \geq 0$$

and an increasing Gaussian function

$$a(n) = \exp\left(-\left(\frac{2(n-N)}{N}\right)^2\right) \text{ for } n \geq 0,$$

respectively. The squared norms of the three source signals are normalized to unity. Since the three source periodic signals can be assumed to be independent to each other, the SNR of each source signal in the mixture is about -3.0 dB. The sets of three periods for mixtures are shown in the first column of Table. 1. The first set contains the periods have three divisors. The second and third consist of closely spaced periods. An example of the mixture is shown in Fig. 1(a). The three source periodic signals are shown in Fig. 1(b), (c) and (d), respectively.

For the sparse periodic decomposition, the sequence of the parameters $\{\alpha_p\}_{p \in P}$ and the sparsity parameter λ have to be specified. The shrinkage of the l_2-norm of the periodic component in the decomposition algorithm is performed with the threshold $\lambda\alpha_p$ in (21) and (23). The periodic signal $\mathbf{f}_p$ with the l_2-norm that is less than the threshold is eliminated by the shrinkage. Obviously, if the residual $\mathbf{r}$ in (18) can be assumed to be a noise that is small enough to approximate the input signal, its periodic approximation has to be eliminated during the decomposition. We assume that the noise as a Gaussian noise with a variance σ^2. The product $\lambda\alpha_p$ is specified as proportional value to the expected l_2 norm of the

approximated Gaussian noise with the periodic signal model. The expected l_2 norm of the periodic signal $\mathbf{f}_p$ that approximates a Gaussian noise, of which envelope is constant, is approximated as

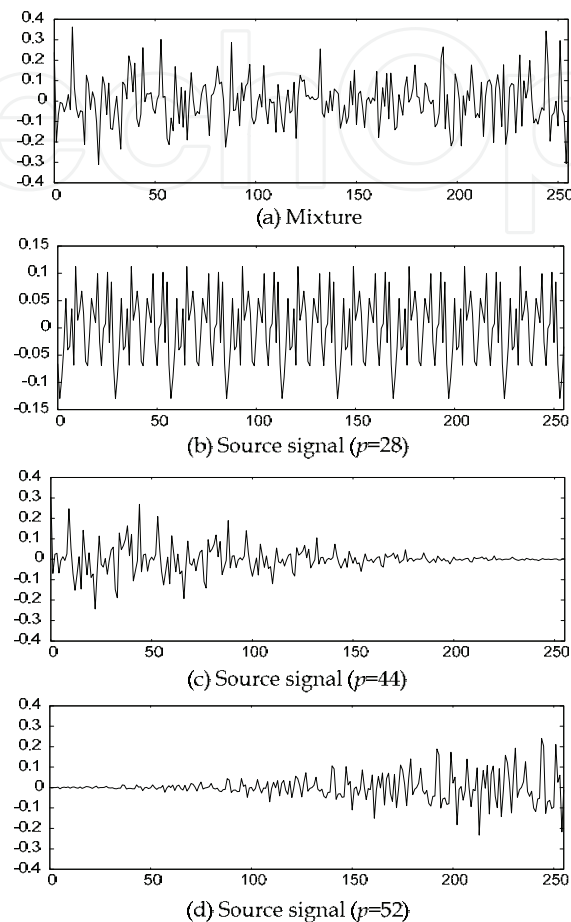$$E\left[\|\mathbf{f}_p\|_2\right] \approx \sigma\sqrt{p+1+\frac{N}{p}} \tag{25}$$

(a) Mixture

(b) Source signal (p=28)

(c) Source signal (p=44)

(d) Source signal (p=52)

Fig. 1. (a) Example of mixture of three periodic signals, the source periodic signals, (a) p = 28, (c) p = 44 and (d) p =52.

The product $\lambda\alpha_p$ is hence specified to a value that is proportional to this expectation. In actual decomposition, σ is assumed to be 1% of the l_2-norm of the input signal. $\lambda\alpha_p$ is specified as the expectation shown in (25).

In the experiments, we supposed that the period of the source signals are integer in the range [10, 59]. The periods for the decomposition are also defined as integers in this range. So, the number of the periodic signals that are obtained by the decomposition is 60. The iteration of the decomposition algorithm explained in Sect. 4. 2 is stopped when l_∞-norm of the difference of the periodic signals before and after updating is lower than a threshold value. The threshold is specified as $0.01 \times \lambda\alpha_p$ for all experiments.

In order to evaluate the decomposition, we compute the improvement in SNR. The improvement in SNR is computed as the difference of the SNRs of the mixture and decomposition results for each source period. We generate 1,000 mixtures to test the decomposition algorithm for each set of periods. Table 1 shows the averages and standard

deviations of the SNR improvements of the decomposed periodic signals for 1,000 tests. The average SNR improvements of the decomposition results exceed 10 dB. By these results, we see that the proposed decomposition can obtain significant decomposition results and separate three sources into its periods. In Fig. 2 and 3, an example of the mixture and its decomposition result are shown. The discrete Fourier transform (DFT) spectrum of the mixture (Fig. 1(a)) is shown in Fig. 2(a).
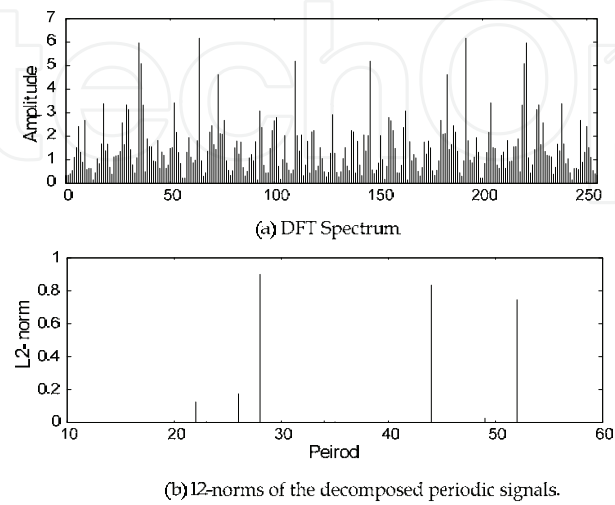

(a) DFT Spectrum

(b) l2-norms of the decomposed periodic signals.

Fig. 2. (a) DFT spectrum of the mixture in Fig. 1(a) and (b) distribution of the l_2 norm of the decomposed periodic signals.

(a) Decomposed signal (p=28)

(b) Decomposed signal (p=44)

(c) Decomposed signal (p=52)
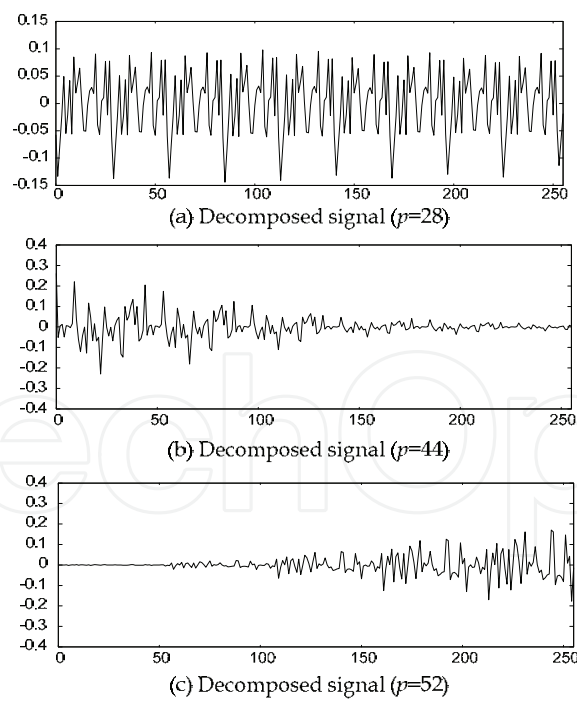
Fig. 3. Decomposed periodic signals, (a) p =28, (b) p = 44 and (c) p =52.

The distribution of l_2 norm of the resultant periodic signals of the mixture is shown in Fig. 2(b). As seen in Fig. 2(b), three periodic signals with large amplitude appear at the source periods. Small harmonics components are separated from the source periods due to the

weighting of the sparsity penalty, however, the almost energy of the mixture is decomposed into the three source periods. In Fig. 3, the periodic signals that appear in the decomposition result are also shown. In this set of the periods, the harmonics with periods 1, 2, and 4 which are the common divisors of the source periods cannot be separated accurately. However, the other harmonics are well collected to three fundamental periods.
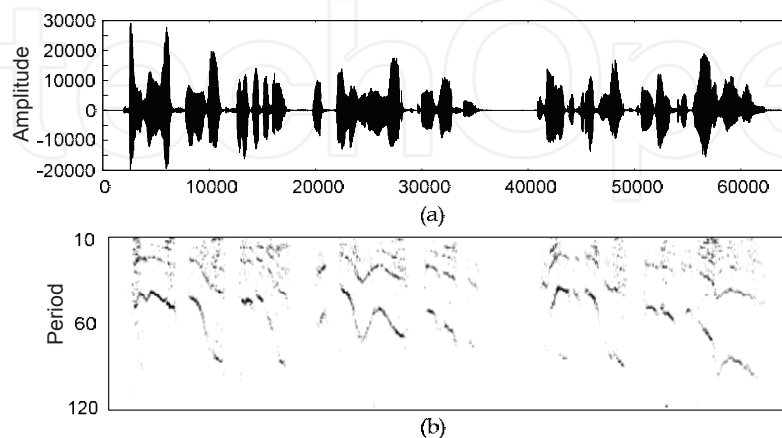

Fig. 4. (a) Speech signal (male, duration: 8.1 s, sampling freq. : 8 kHz) and (b) time-period energy distribution of (a).
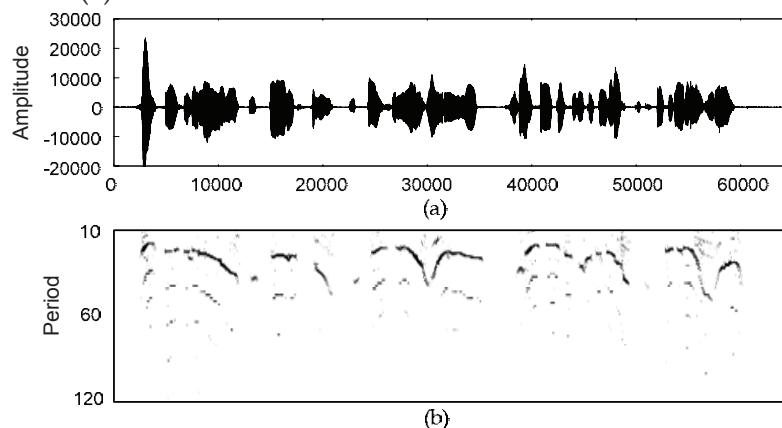
Fig. 5. (a) Speech signal (female, duration: 8.1 s, sampling freq. : 8 kHz) and (b) time-period energy distribution of (a).

6. Application to speech representation

In the synthetic signal examples, the signal mixtures consist of source periodic signals with integer periods. However, periods of many periodic signals that include speech and acoustic signals are not integer. In order to examine the sparse periodic decomposition for the signals with non-integer periods, we apply the proposed sparse decomposition to speech mixtures.

The speech signals for the experiments were selected 3 Japanese male and 3 female continuous speeches of about 8 s taken from ATR-SLDB (Spoken Language Database). The sampling rate of each speech signal is converted to 8 kHz. 15 speech mixtures that consist of two different speeches that are normalized to same power are generated.

For periodic decomposition, each mixture is divided into segments that contain 360 samples with 3/4 overlap. In each segment, the periods for decomposition are specified to be

integers in the range [10, 120] which corresponds to the range of the fundamental frequencies of most men and women. The stopping rule of the iteration of the relaxation method and the parameters are specified as the same rule that is mentioned in Sect. 5.

The examples of the male and female utterances and its time-period energy distributions are shown in Fig. 4 and Fig. 5, respectively. In Fig. 4(b) and 5(b), the brightness indicates the power of the resultant periodic signals for each segment and period. Darker pixels indicate higher powers of the resultant periodic subsignals.
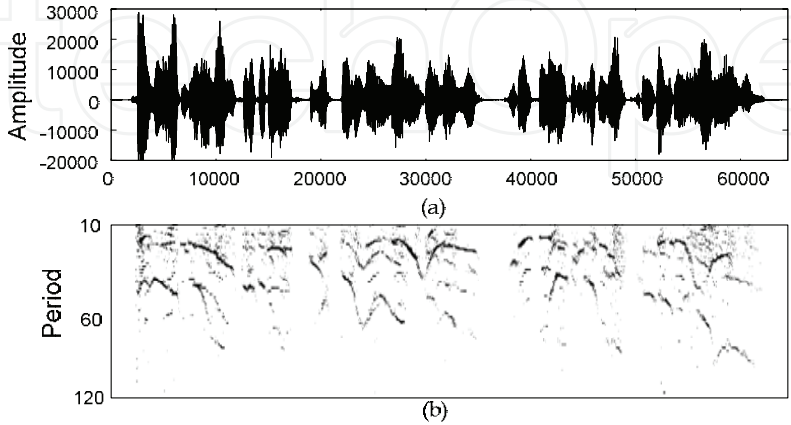
Fig. 6. (a) Mixture of female and male speeches and (b) time-period energy distribution of (a)

Speakers	Ave. SNR	Min. SNR	Max. SNR	Ave. num. of periods
(M, M)	20.1	10.4	28.9	16.1
(F, F)	20.2	10.3	27.6	11.2
(F, M)	20.2	10.2	28.9	14.0

Table. 2. Average, minimum and maximum SNRs (dB) of approximated speech segments and average numbers of periodic signals obtained by the sparse decomposition

Our method decomposes a signal into the periodic signals with only integer periods. Under this limitation, the speech components with non-integer periods and the frequency variations that occur in a segment are represented as the sum of some periodic signals. So, we see that the pitch contours are represented by some neighbouring periods in these time-period distribution. Moreover, small periodic components with periods that are multiples and divisors of the fundamental periods appear. These periodic components appear due to the non-integer periodic components of the speeches and the weighting of the sparsity measure in (17). However, the most of the signal energy is concentrated around the fundamental pitch periods of the speeches.

We also show the time-period energy distributions of the mixture of two speeches. Fig. 6(a) and (b) show the mixture of the source speech signals shown in Fig. 3(a) and Fig. 4(a) and its time-period energy distributions, respectively. We see that the time-period energy distribution of the mixture in Fig. 6 is almost equal to the sum of the two distributions of the source speeches shown in Fig. 4(b) and Fig. 5(b). The both of the pitch contours of the two source speeches are preserved in the distribution of the mixture. The proposed decomposition method can approximate the mixture while concentrating the energy of each speech to its pitch periods and provides sparse representation of the mixture. It is expected that the pitch periods of both the speech signals will be tracked in this time-period energy

distribution. Moreover, speech separation will be achieved by assigning the resultant periodic signals to the sources.

In order to evaluate the approximate decomposition, we compute the SNR and the number of the non-zero resultant periodic signals for each segment where the l_2 norm is greater than the noise level. The average, maximum and minimum SNRs over all voice active segments of mixtures are shown in Table 2. In this table, F an M denote female and male source speeches, respectively. The average numbers of periods for approximation of a segment are also shown. We see that the average approximation precision of the proposed decomposition is about 20 dB in the segmental SNR. The average number of the periods yield by the decomposition is about 14 for segments of speech mixtures consist of two speeches.

Speaker	Proposed (with sources)	DFT (with sources)	Proposed (with ref. sig.)
(M, M)	9.9±0.6	13.5±0.5	3.9±1.0
(F, F)	9.5±0.3	13.5±0.5	3.2±0.9
(F, M)	F: 10.1±1.5	F: 14.4±1.0	F: 6.5±2.5
	M: 9.8±1.0	M: 14.3±1.0	M: 6.7±2.7

Table 3. Average SNRs (dB) of separated speeches.

Next, we demonstrate the speech separation from a mixture with the sparse periodic decomposition. In this experiment, the speech separation is performed by assignment of the resultant periodic subsignals to the sources in each segment.

First, we use the clean source signals for assignment of the resultant periodic signals.

The separation is carried out by the following steps for each segment:

1. The segment of the mixture is decomposed into the set of the periodic signals $\{\mathbf{f}_p\}_{p \in \mathrm{P}}$.
2. The normalized correlations between the resultant periodic signals and the clean source segments $\{\mathbf{s}_i\}_{i = 1, 2}$ are computed.
3. Each resultant periodic signal $\mathbf{f}_\mathrm{P}$ are added to the separated output that is associated with the i-th source $\mathbf{s}_i$ that obtains larger correlation.

For recovering source signals, each resultant periodic signal is multiplied with a Hanning window in each segment. This assignment method does not obtain optimum separated results in terms of the SNR exactly. However, this experiment gives the rough ideal performance of the source separation by using the proposed sparse decomposition.

For comparison, the ideal separation results that are obtained by a STFT that is widely utilized for the sparse representation of speech signals are demonstrated. In the separation with the STFT, the ideal binary masks [20] are computed from the clean source speeches.

The mixture and the source signals are segmented by 512 points Hamming window with 3/4 overlap. In each segment, the DFT spectrum of the mixture and the source signals are computed. Each frequency bin of the DFT is assigned to the source whose amplitude of the frequency bin is larger than the other. The separation results obtained by the proposed decomposition and the DFT are shown in Table 3.

In this table, the SNRs of the separated speech signals are shown. We see that the SNRs of the separated speeches obtained by the proposed method are lower than the DFT by about

4dB. In the separation obtained by the proposed method, the approximation errors caused during the decomposition are involved in the separated output. Since the frequency resolution of the periodic decomposition is lower than the DFT at high frequency bands, the interferences between two speeches mainly occur at high-frequencies. However, the proposed representation is sparser than the DFT spectrum. In this experiment, the DFT yields 257 frequency bins for each segment. So, the DFT based separation is the problem of the assignment of the 257 frequency bins. In contrast, the average number of the periodic signals yield by the proposed method is about 14 for a segment. Comparing the proposed decomposition with the DFT, the separation problem can be reduced to relatively small size of a combinatorial optimization by the proposed decomposition.
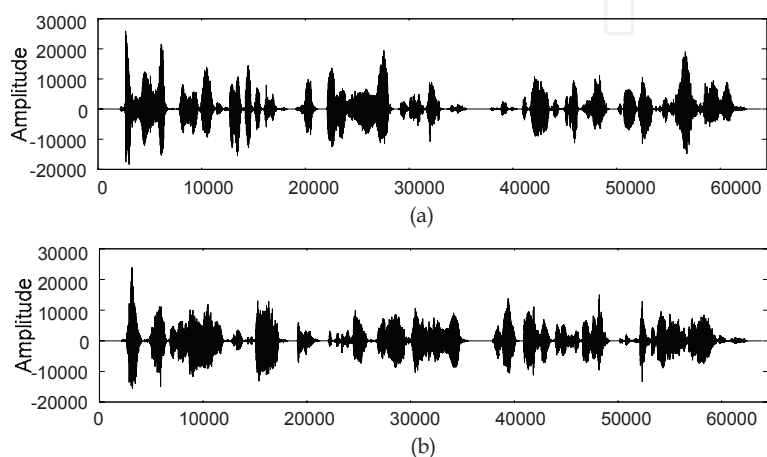
Fig. 7. Separated speech signals obtained by sparse periodic decomposition with reference speeches, (a) separated male speech (SNR: 7.2dB) and (b) female speech (SNR: 7.1dB) from the mixture shown in Fig. 5(a)

In above separation experiments, we assume that the source speeches are known. Next, we demonstrate the single-channel speech separation by referencing the clean speech segments. In this scenario of the separation, two speakers in a mixture are known and the clean speeches of the speakers are available, but the contents of the speeches in the mixture are unknown. In order to assign the periodic signals to the sources, a set of the clean speech segments of the i-th speaker is defined as $\{\mathbf{c}_{i,j}\}_{1 \leq j \leq Nr}$ where Nr is the number of the reference segments.

The resultant periodic signal $\mathbf{f}_p$ is assigned to the i-th speaker that gives the maximum of the normalized correlation as:

$$\max_{i,j} \frac{\mathbf{f}_p^{\mathrm{T}} \mathbf{c}_{i,j}}{\left\| \mathbf{f}_p \right\|_2 \left\| \mathbf{c}_{i,j} \right\|_2}$$

For this experiment, segments that are generated from a clean speech of 20 s are used for the references of each speaker. The segments where the voice is not active are rejected from the references. The references do not include the source utterances in the mixtures. The SNRs obtained by the separation with the references are also shown in Table. 3. Obviously, such a simple separation method causes many false assignments. For separation of the mixture consists of the speakers of same gender, the averages of the improvements of SNR are lower than 4dB. However, the averages of SNR close to the ideal results and are about 6.5dB for

the speakers of opposite gender. The separated signals from the mixture in Fig. 6(a) are shown in Fig. 7(a) and (b).

The single channel speech separation methods based on frequency masking of spectrum have been proposed [12] [13] [14]. In these methods, statistical models for the frequency spectra of the speakers are preliminary learnt. The separation is performed on the frequency spectrum of the mixture by using the statistical models. In our approach, the proposed sparse decomposition yields the small number of the periodic signals which approximate the source signal due to the sparsity penalty. So, the separation of two speeches that have less similarity can be performed by such a lazy assignment method.

7. Conclusions

In this chapter, we present a sparse decomposition method for periodic signal mixtures. The proposed decomposition is based on the model for the periodic signals with time-varying amplitude and the sparsity of the periods that appear in the decomposition result. In decomposition experiments of the synthetic signal and the speech mixtures, we demonstrated that the proposed decomposition has the ability of source separation.

The assignment method that is employed for the single-channel speech separation demonstrated in this paper is too simple to obtain good separation results. In our decomposition results, as seen in the figures in Sect. 4, the speech pitch contours are involved. We can use the temporal continuity of the speech pitches and spectra over the consecutive segments for improvement of the accuracy of the assignment. The accurate and robust assignment of the decomposed periodic signals is a topic for future research.

8. References

McAulay, R.; and Quatieri, T. Speech analysis/synthesis based on a sinusoidal representation, *IEEE Trans. on Acoustics, Speech and Signal Processing*, Vol. 34, No. 8, pp. 744-754, Aug. 1986. [1]

Wu, M.; Wang, D.; & Brown, G. J. A multipitch tracking algorithm for noisy speech, *IEEE Trans. on Speech and Audio Processing*, Vol. 11, No. 3, pp. 229-241, May 2003. [2]

Goto, M. A real-time music scene description system: Predominant F0 estimation for detecting melody and bass lined in real-world audio signals, *Speech Commun.*, Vol. 43, No. 4, pp. 311-329, 2004. [3]

Le Roux, J.; Kameoka, H.; Ono, H.; Cheveigne, A. & Sagayama, S. Single and multiple F0 contour estimation through parametric sepectrogram modeling of speech in noisy enviroments, *IEEE Trans. on Audio, Speech and Language Processing*, Vol. 15, No. 4, pp. 1135-1145, May 2007. [4]

Triki, T. & Slock T. Periodic signal extraction with global amplitude and phase modulation for musical signal decomposition, *Proc. on ICASSP*, Vol. 3, pp. 233-236, 2005. [5]

Santhanam, B.; & Maragos, P. Harmonic analysis and restoration of separation methods for periodic signal mixtures: Algebraic separation versus comb filtering, *Signal Processing*, Vol. 69, No. 1, pp. 81-91, 1998. [6]

Muresan, D. D. & Parks, T. W. Orthogonal, exactly periodic subspace decomposition, *IEEE Trans. on Signal Processing*, vol. 51, no. 9, pp. 2270-2279, Nov. 2003. [7]

Sethares, W. A.; & Staley, T. W. Periodicity transform, *IEEE Trans. on Signal Processing*, vol. 47, no. 11, pp. 2953-2964, Nov. 1999. [8]

Nakashizuka, M. A sparse decomposition method for periodic signal mixtures, *IEICE Trans. on Fundamentals*, Vol.E91-A, No.3, pp. 791-800, March 2008. [9]

Nakashizuka, M. A sparse periodic decomposition and its application to speech representation, Proc. on EUSIPCO 2008, Aug. 2008. [10]

Yilmaz, O. & Rickard, S. Blind separation of speech mixtures via time-frequency masking, *IEEE Trans. on Signal Processing*, Vol. 52, No. 7, pp. 1830-1847, July 2004. [11]

Roweis, T. S. Factorial models and refiltering for speech separation and denoising, *Proc. on Eurospeech*, Vol. 7, No. 6, pp. 1009-1012, Geneva, 2003. [12]

Reddy, A. M. & Raj, B. Soft mask methods for single-channel speaker separation, *IEEE Trans. on Audio, Speech and Language Processing*, Vol. 15, No. 6, pp. 1766-1776, Aug. 2007. [13]

Radfar, M. H. & Dansereau, D. M. Single-channel speech separation using soft mask filtering, *IEEE Trans. on Audio, Speech and Language Processing*, Vol. 15, No. 8, pp. 2299-2310, Nov. 2007. [14]

Lewicki, M. S. & Olshausen, B. A. A probabilistic framework for the adaptation and comparison of image codes, *J. Opt. Soc. Amer. A, Opt. Image Sci.*, Vol. 16, No. 7, pp. 1587-1601, 1999. [15]

Plumbley, M. D.; Abdallah, S. A.; Blumensath, T. & Davies, M. E. Sparese representation of polyphonic music, *Signal Processing*, Vol. 86, No. 3, pp. 417-431, March 2006. [16]

Chen, S. S.; Donoho D. L. & Saunders, M. A. Atomic decomposition by basis pursuit, *SIAM Journal on Scientific Computing*, Vol. 20, No. 1, pp. 33-61, 1998. [17]

Sardy, S.; Bruce, A. G. & Tseng, P. Block coordinate relaxation methods for nonparametric wavelet denoising, *Journal of Computational and Graphical Statistics*, vol. 9, no. 2, pp. 361-379, 2000. [18]
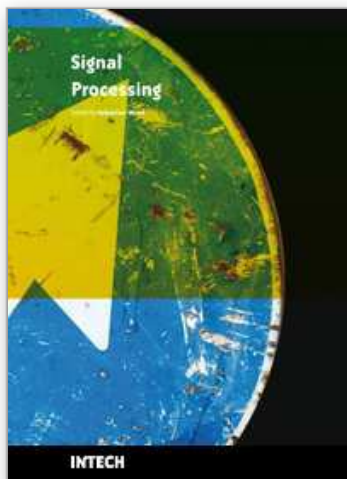
Signal Processing

Edited by Sebastian Miron

ISBN 978-953-7619-91-6

Hard cover, 528 pages

Publisher InTech

Published online 01, March, 2010

Published in print edition March, 2010

This book intends to provide highlights of the current research in signal processing area and to offer a snapshot of the recent advances in this field. This work is mainly destined to researchers in the signal processing related areas but it is also accessible to anyone with a scientific background desiring to have an up-to-date overview of this domain. The twenty-five chapters present methodological advances and recent applications of signal processing algorithms in various domains as telecommunications, array processing, biology, cryptography, image and speech processing. The methodologies illustrated in this book, such as sparse signal recovery, are hot topics in the signal processing community at this moment. The editor would like to thank all the authors for their excellent contributions in different areas of signal processing and hopes that this book will be of valuable help to the readers.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Makoto Nakashizuka (2010). Sparse Signal Decomposition for Periodic Signal Mixtures, Signal Processing, Sebastian Miron (Ed.), ISBN: 978-953-7619-91-6, InTech, Available from:
http://www.intechopen.com/books/signal-processing/sparse-signal-decomposition-for-periodic-signal-mixtures

INTECH
open science | open minds