統計的特徴量に基づく音声信号の分離・抽出に関する研究

Signal Separation and Extraction of Speech Signal Based on Its Statistical Characteristics

平成１９年３月

畔津 忠博

山口大学大学院理工学研究科

論文要旨

本論文は，雑音に埋もれた所望の音声信号から，それを分離・抽出するために音声の統計的特徴量を利用した幾つかの方法について述べる．

一般に，信号処理の分野において，センサーで受信された観測信号から雑音を除去して所望の信号を得ることは，重要な問題の１つである．この問題を解くために観測信号の統計的性質に応じてフィルタ係数を学習する適応フィルタが使われる．教師なし学習の適応フィルタを用いる代表的な例として，複数の信号源から生成される原信号が互いに混合されて複数のセンサーで観測されたとき，原信号や混合過程の情報を未知として，観測信号のみを用いて原信号を推定する問題がある．これは，ブラインド信号分離 (BSS) 問題と呼ばれる．例えば原信号が音声の場合は，複数話者の混合音声から特定の音声を取り出す問題，所謂カクテルパーティー問題として古くから考察されてきた．

近年，BSS 問題を解くための新しい考え方として，情報理論に基づいた情報論的アプローチがある．その中でも特に，独立成分分析 (ICA) と呼ばれる手法が盛んに研究されている．ICA は，混合過程が線形である場合に，未知の原信号が統計的に独立であるという仮定のみを用いて，観測信号から原信号を推定する統計的信号処理手法である．ICA では，学習により決定される線形変換を観測信号に施した分離信号により原信号を推定する．この線形変換を決定するための学習アルゴリズムを導くときに統計的独立性の規範が用いられる．代表的な ICA の学習アルゴリズムでは，この規範として分離信号の相互情報量最小化が用いられる．このとき，分離信号の周辺分布の確率密度関数を引数にもつ非線形関数が必要とされる．従来の ICA では，この非線形関数の記述にシグモイド型関数が用いられているが，原信号の確率密度関数の形状によっては，信号分離ができない場合が存在する．そこで，第 2 章では，関数近似能力に優れた動径基底関数 (RBF) ネットワークの特性を利用することにより，ICA に必要とされる非線形関数をできるだけ正確に記述する手法を提

案する．さらに，従来法と提案手法のそれぞれの利点を生かしたハイブリッドICA提案し，信号分離の収束スピードと精度の更なる向上を目指す．

一般的なICAの学習アルゴリズムは，瞬時混合のBSS問題に対して導出されている．しかしながら，音声のような伝搬遅延時間を無視できない信号に対しては，瞬時混合のICAをそのまま適用することはできない．伝搬遅延時間を含むBSS問題を解くためのICAの学習アルゴリズムは，時間領域で直接的に扱うものと周波数領域に変換して処理するものの2つに大きく分けられる．周波数領域でこの問題を扱った場合，周波数成分毎に瞬時混合のICAが利用できるため，時間領域で扱うよりも学習アルゴリズムを単純化できる．そこで，第3章では，周波数領域ICAの考えに立脚した信号分離手法を提案する．まず，周波数領域ICAを用いて正確に伝搬遅延時間を推定する方法を示し，次に，それを用いた信号分離手法について述べる．さらに，推定した混合過程の伝搬遅延時間と減衰係数を利用した音源定位の方法についても示す．

さて，今まで述べてきたICAはBSS問題を解くための手法であるが，原則として2つ以上の観測信号を必要とする．すなわち，ICAは1つの観測信号のみからでは，特定音声を抽出することはできない．そのため，ICAは非常に有用な手法ではあるが，本当の意味で完全にカクテルパーティー問題を解いているとは言えない．そこで，第4章では，特定話者の音声と他の複数話者の音声が混合した1つの観測信号から，特定話者の音声を抽出する方法を考える．このとき，ICAのような教師なし学習ではなく，抽出したい音声について事前に何らかの情報が得られている教師あり学習を用いる．提案手法では，教師信号として周波数領域における話者の個性が使われる．具体的には，特定話者の明瞭な音声から事前に作成される，スペクトル情報を蓄積した辞書を用いる方法を提案する．

最後に，時間領域の音声信号の特徴抽出にICAを用いる方法を検討してみる．ICAは，自然画像の特徴抽出に応用され成功している．そこでは，画像の特徴量として

小さなパッチ画像に相当する基底関数を使い，画像を基底関数の重み付き加算で表現している．そして，重み係数が互いに統計的に独立になるような基底関数をICAにより求めている．第5章では，ICAを音声信号の特徴抽出に利用し，どのような基底関数が得られるかを報告する．ICAによる特徴抽出は，スパースコーディングによる特徴抽出と密接な関係がある．スパースコーディングは，生物の視覚系の信号処理モデルに基づいて重みに疎(スパース)性の条件を課すことにより，少数の基底関数で信号を表現する手法である．そこで，ICAによって得られた基底関数が少数でも音声信号を再構成することが可能かどうか，情報圧縮の観点から検討を行う．

Abstract

This thesis describes several methods of signal separation and extraction of speech signal based on its statistical characteristics. The speech signal is often corrupted by other speech signals or background noise in a real environment. The extraction of the target speech signal from the mixed speech signal is important for smooth communication. This is called a cocktail party problem in speech signal processing. In general, the problem of separating or extracting the desired signals from the observed signals is important in the field of signal processing. Unsupervised or supervised adaptive filters are widely used to solve this problem. The adaptive filters determine the filter coefficients by using the statistical characteristics of the observed signals. The typical example of the unsupervised adaptive filters is the blind signal separation (BSS). The BSS is an effective approach to separate source signals from the observed ones, which are a linear mixture of the source signals. This separation is carried out without knowing the mixing coefficients and the properties of the source signals. The independent component analysis (ICA) is a statistical method to solve the BSS problem by applying a linear transformation to the observed signals so that the separated signals become statistically independent of each other.

The major ICA algorithm is derived by minimizing the mutual information of the separated signals. In this ICA algorithm, the nonlinear functions, concretely the derivatives of the logarithmic marginal probability density functions (PDFs) of the separated signals, are to be described explicitly in the learning process of the ICA. In general, the sigmoid functions are often employed for the description of those nonlinear functions. However, a simple function like the sigmoid function is not suitable to deal with a wide variety of probability distributions.

In chapter 2 of this thesis, we apply the RBF networks to describe those nonlinear functions as precisely as possible. The point of the proposed method lies in its description ability for the complicated probability distributions of signals. We also propose a hybrid ICA with the conventional ICA and the RBF-ICA. In the hybrid ICA, the conventional ICA is employed at the beginning stage of signal separation to get a rough separated signal with high speed, and then later it is switched to the RBF-ICA to get a better signal separation accuracy.

Many ICA algorithms are presented for the BSS problem of instantaneous mixtures. However, when, for example, the source signals are the speech signals, we have to consider the propagation time delays of the source signals in the BSS problem. Several methods for solving this problem have been reported in the time domain.

However, in case where this problem is treated in the time domain, the method becomes very complex just at that moment. On the other hand, the separation process is simplified by the frequency-domain ICA. The frequency-domain ICA is a straightforward extension of the commonly used ICA algorithm for instantaneous mixtures. In chapter 3, we propose a simple and effective method by extending the frequency-domain ICA. The proposed method estimates at first the relative propagation time delays and the propagation coefficient ratios. Then by making use of these estimates, the signal separation is carried out with a higher performance than the conventional frequency-domain ICA. Furthermore, the sound localization is realized by employing the estimated relative propagation time delays and the propagation coefficient ratios.

The ICA needs two or more observed signals because of using the statistical distance between signals, and thus, the ICA cannot extract the target signal from only one mixed signal. From this reason, the ICA is not a method to solve a real cocktail party problem. In chapter 4, we propose a speech extraction method from one mixed speech signal using the speaker individualities in the frequency domain. The speech signal has many features in the frequency domain, for example, the fundamental frequency, the formant frequency, and the spectral envelope and so on. These features represent the speaker individualities attributed to the human speech organs. In the proposed method, the speaker individualities are reflected in the dictionary, which is composed in advance from the clear speech signal of a target speaker. That is, this method is one of the supervised filters that are different from the ICA based methods mentioned above.

In chapter 5, we report the result of applying ICA to feature extractions of the speech signal in the time domain. The ICA is closely related to the sparse coding. The sparse coding is a method to represent an image signal by using a few basis functions for natural image. This is based on the perceptual system of the mammalian visual cortex. We first show the basis functions obtained by the ICA from the speech signal, and then investigate whether the target speech signal can be represented by using a couple of these basis functions obtained from the signal compression viewpoint.

Chapter 6 is devoted to conclusions.

目 次

図 目 次

第1章　序論

本論文は，雑音に埋もれた所望の音声信号から，それを分離・抽出するために音声の統計的特徴量を利用した幾つかの方法について述べる．

音声は，昔から現在に至るまで，人間が他者と相互に情報を交換するための最も重要な手段の１つとして用いられてきた．音声には，言語の内容を表す言語情報だけでなく，話者の個性や感情に由来する個人性情報や情緒的情報など様々な情報が埋め込まれている [1]．そのため，音声あるいは聴覚について色々な角度から研究することは，工学的応用に役立つだけでなく，人間の聴覚情報処理における生物学的知見も得られる可能性があり，非常に興味深い．

音声研究には，音声分析，音声合成，音声符号化 [2] や，音声認識 [3]，聴覚情景分析 [4][5] など様々な研究分野が存在するが，本論文では，音声信号分離・抽出を中心のテーマとする．音声信号分離・抽出とは，様々な雑音と混合して観測される信号から，所望の音声信号を取り出すことであり，円滑なコミュニケーションを行う上で必要とされる技術である．例えば，高騒音化におけるハンズフリー通話，通信会議室におけるマイクロホンアレー処理，機械による音声認識のための前処理，携帯電話などの移動体通信における雑音除去といった場面で適用できる．

一般的に，雑音に汚された所望の信号がセンサーで観測され，その信号 (観測信号) から雑音を除去することにより所望の信号を得ることは，信号処理の分野において重要な問題の１つである．この問題を解くためにウィナーフィルタやカルマンフィルタといった最適フィルタ理論がある．しかしながら，これらの理論では，所望の信号と雑音が明確に区別され，これらの統計的性質についての先験的な情報が

必要であったり，雑音の混合過程に微分方程式/差分方程式で表現される数学モデルを仮定したりするため，様々な雑音が存在する場合や，所望の信号と雑音に明確な区別がない場合，例えば音声同士が混合する場合などにはあまり適さない．

所望の信号と雑音に関する統計的性質やその混合過程が未知である場合，あるいは一部分しか分からない場合，観測信号の統計的性質に応じてフィルタ係数を学習する適応フィルタが使われる．これまで，Least Mean Square(LMS) アルゴリズムや Recursive Least Squares(RLS) アルゴリズム，線形予測 (Linear Prediction, LP) 法など様々な適応フィルタ理論が提案されている [6][7]．しかしながら，これら従来から提案されている適応フィルタは，最適フィルタの近似的側面が強く，さきほど指摘した問題点は残されたままである．

さて，適応フィルタは，教師あり学習と教師なし学習の 2 つに大きく分けることができる．教師あり学習は，所望の信号について何らかの情報が得られている場合に，それを教師信号として利用する方法であり，教師なし学習は，所望の信号について先験的情報が得られていない場合に，観測信号だけを利用して所望の信号を推定する方法である．

教師なし学習の適応フィルタを用いる代表的な例として，複数の信号源から生成された信号が，互いに混合されて複数のセンサーにより観測されたとき，信号や混合過程の情報を未知として観測信号のみを用いて信号源の信号 (原信号) を推定する問題がある．これは，ブラインド信号分離 (Blind Signal Separation, BSS) 問題と呼ばれる [7][11][12]．例えば原信号が音声の場合は，複数話者の混合音声から特定の音声を取り出す問題，所謂カクテルパーティー問題として古くから考察されてきた [8]-[10]．

人間は，高雑音下のカクテルパーティー会場などでも会話が可能であるので，何らかの方法でカクテルパーティー問題を解いていることになる．これには，信号源 (話者) とセンサー (聞き手の耳) との位置関係による音声信号の減衰や遅延の情報が

利用されている他，脳の高次情報処理も関係していると言われる．BSS 問題では前者のみを考察の対象とするが，後者も，人間の聴覚情報処理を解明する上で興味深い対象である．また，BSS 問題は，音声に限らず，画像処理における雑音の除去や特徴抽出，計測した脳波を脳内の機能毎の成分に分解することによる脳機能の分析，携帯電話などの移動体通信におけるエコーや干渉波の消去など，多くの分野において研究されている．

近年，BSS 問題を解くための新しい考え方として，シャノンに始まる情報理論 [13][14] に基づいた情報論的アプローチがある．その中でも特に，独立成分分析 (Independent Component Analysis, ICA) と呼ばれる手法が盛んに研究されている [15]-[31]．ICA は，混合過程が線形である場合に，未知の原信号が統計的に独立であるという仮定のみを用いて，観測信号から原信号を推定する統計的信号処理の 1 手法である．観測信号のみを使って所望の信号を推定することから，適応フィルタにおける教師なし学習の 1 つとしても位置づけられる [7]．原信号の統計的独立性のみ利用するため，様々な問題に適用でき汎用性が非常に高く，さきほど指摘した従来の適応フィルタではあまり適さない問題に対しても有効である．また，原信号に統計的従属性があるときは，多変量解析の正準相関分析と ICA を組み合わせる手法などが提案されている [26]．

ICA では，学習により決定される線形変換を観測信号に施した信号 (分離信号) により原信号を推定する．この線形変換を決定するための学習アルゴリズムを導くときに統計的独立性の規範が用いられる．この規範としては，分離信号の相互情報量の最小化 [19][21] や分離信号に非線形変換を施した信号のエントロピーの最大化 (Infomax)[20]，4 次キュムラントの絶対値またはネゲントロピーの最大化 (FastICA)[18][28]，4 次クロスキュムラントの最小化 (JADE)[16][17] などが用いられる．また，時間的な相関を用いる手法も提案されている [23][24]．

統計的独立性の規範として相互情報量の最小化を用いた場合，ICA の学習アルゴ

リズムにおいて，分離信号の周辺分布の確率密度関数を引数にもつ非線形関数が必要とされる [21]．ICA では，それぞれの原信号の特性は未知とするため，有限個の観測信号の値のみを使用して確率密度関数または非線形関数を推定しなければならない．確率密度関数を有限個の値から推定する方法として，モーメントやキュムラントを用いた多項式近似がよく知られているが [19][21][32]，複雑な計算を伴い，また，確率密度関数の種類によっては，近似できないものも存在する．

そこで，第 2 章では，関数近似能力に優れた動径基底関数 (Radial Basis Function, RBF) ネットワーク [33]-[36] の特性を利用することにより，ICA の学習過程で必要とされる非線形関数を正確に記述する手法を提案する．また，従来手法と提案手法のそれぞれの特徴を生かすため，この 2 つを併用する方法についても述べる．

さて，一般的な ICA の学習アルゴリズムは，瞬時混合の BSS 問題に対して導出されている．瞬時混合とは，混合過程に信号の伝搬遅延時間を考慮しないモデルのことである．伝搬遅延時間は信号源からセンサーまでの距離に依存するが，例えば，電波や光のような非常に高速に伝搬する信号の場合では，これを無視できるため瞬時混合として問題ない．しかしながら，音声のような伝搬遅延時間を無視できない信号に対しては，瞬時混合の ICA をそのまま適用することはできない．

伝搬遅延時間を含む BSS 問題を解くための ICA の学習アルゴリズムは，時間領域で直接的に扱うものと周波数領域に変換して処理するものの 2 つに大きく分けられる．前者は，瞬時混合の場合では 1 つの線形変換演算子行列を推定すれば良かったものが，この場合では FIR，IIR フィルタにおける多数のフィルタ係数を求める問題となり，アルゴリズムが複雑になる [37]-[39]．

一方，後者は，短時間フーリエ変換により時間領域の信号を周波数領域の信号に変換する．これにより，時間領域における伝搬遅延時間を含む BSS 問題を周波数帯域毎の瞬時混合問題に変換できるため，従来の ICA の学習アルゴリズムがそのまま適用できる [40]．そのため近年，周波数領域 ICA(Frequency-Domain ICA) を用いた

信号分離手法が盛んに研究されている [40]-[52]．さらに，人間の聴覚機構は高性能な周波数分析器として捉えることができるため [8]，今後，人間の聴覚情報処理の研究が脳の高次機能の解明とともに進めば，周波数領域ICAにこれらの知見を取り入れることも考えられる．

そこで，第3章では，周波数領域ICAの考えに立脚した信号分離手法を提案する．まず，周波数領域ICAを用いて正確に伝搬遅延時間を推定する方法を示し，次に，それを用いた信号分離手法について述べる．さらに，推定した混合過程の伝搬遅延時間と減衰係数を利用した音源定位の方法についても示す．

今まで述べてきたICAはBSS問題を解くための手法であるが，原則として2つ以上の観測信号を必要とする．それゆえ，ICAは1つの観測信号のみからでは，特定音声を抽出することはできない．人間の聴覚情報処理は，カクテルパーティー問題を何らかの方法で解いているが，このとき，センサーが2つの場合に相当する両耳による効果は必ずしも必要ではなく，それ意外の何らかの方法が使われていると推測される．この理由から，ICAは非常に有用な手法ではあるが，本当の意味で完全にカクテルパーティー問題を解いているとは言えない．

そこで，第4章では，特定話者の音声と複数話者の音声から作られる観測信号を，ICAのように複数の観測信号からでなく，ある1つの観測信号のみから特定話者の音声を抽出する方法を考える．このとき，ICAのような教師なし学習ではなく，抽出したい音声について事前に何らかの情報が得られている教師あり学習を用いる．音声信号は，周波数領域に多くの特徴をもっている．例えば，基本周波数，フォルマント周波数，スペクトル包絡などがある [53]．これらの特徴は，話者の発話器官に由来する話者の個性を表す [54]．それゆえ，提案手法では，教師信号として周波数領域における話者の個性が使われる．具体的には，特定話者の明瞭な音声から事前に作成される，スペクトル情報を蓄積した辞書を用いる方法を提案する．

最後に，時間領域の音声信号の特徴抽出にICAを用いる方法を検討する．ICAは，

自然画像の特徴抽出に応用され成功している [55][56]．信号を効率的に表現するためには，信号の性質をよく反映する特徴の抽出が非常に重要であるが，そこでは，画像の特徴量として小さなパッチ画像に相当する基底関数を使い，画像を基底関数の重み付き加算で表現している．そして，重み係数が互いに統計的に独立になるような基底関数を，ICA により求めている．

この考え方を音声信号に応用するため，ここでは，定常状態と見なせる音声信号の区間は，複数個の基底関数の重み付き加算で表わせると仮定する．このとき，基底関数は短時間の音声信号となる．そうすると，画像に対する特徴抽出の方法がそのまま音声信号にも適用できる．第 5 章では，音声信号において，どのような基底関数が得られるかを報告する．

また，ICA による特徴抽出は，スパースコーディング [57] による特徴抽出と密接な関係があり，考え方としては，ほとんど同じであることが示されている [58]．スパースコーディングは，生物の視覚系の信号処理のモデルに基づいて重みに疎 (スパース) 性の条件を課すことにより，少数の基底関数で信号を表現する手法である．

そこで，第 5 章では，ICA によって得られた基底関数が少数でも音声信号を再構成することが可能かどうか，再構成した音声信号の誤差とそれに使われた基底関数の数の関係を調べることにより，情報圧縮の観点から検討を行う．

本論文の構成は，図 1.1 のようにまとめられる．第 2 章では，瞬時混合 BSS 問題の定式化と代表的な ICA の学習アルゴリズムについて説明をした後，ICA の学習過程で必要とされる非線形関数の近似に RBF ネットワークを利用する手法を提案する．第 3 章では，伝搬遅延時間を含む BSS 問題の定式化と基本的な周波数領域 ICA の考え方を説明し，周波数領域 ICA に基づいた音声信号分離と音源定位の手法について，それぞれ提案する．第 4 章では，事前に得た特定話者のスペクトル情報を教師信号として，特定話者の音声と複数話者の音声から作られる 1 つの観測信号から，特定話者の音声を抽出する手法について提案する．第 5 章では，ICA を用いた音声

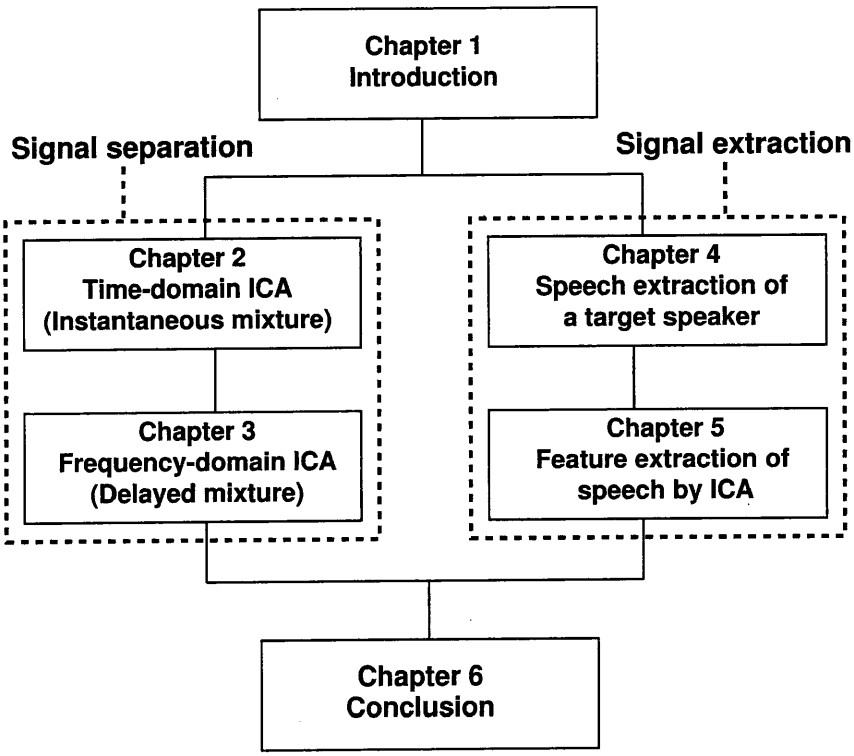


図 1.1: Organization of this thesis.

信号の特徴抽出とその特徴量を使った音声再構成を，情報圧縮の観点から検討する．

第 6 章で本研究の結論を述べる．

第2章　RBFネットワークを用いた独立成分分析

2.1　緒言

本章では，瞬時混合のブラインド信号分離 (BSS) 問題の定式化と代表的な独立成分分析 (ICA) の学習アルゴリズムについて説明をした後，提案手法であるICAの学習過程で必要とされる非線形関数の近似にRadial Basis Function(RBF) ネットワークを利用する手法について述べる．

BSS問題は，複数個の原信号の線形的な混合である観測信号から，混合過程や原信号の特性を未知として原信号を推定する問題である．ICAは，原信号の統計的独立性のみを手がかりとしてBSS問題を解くための統計的信号処理手法である．ICAでは，観測信号に線形変換を施すことにより，できる限り互いに統計的に独立になるような分離信号を得る．この得られた分離信号が推定された原信号となる．

代表的なICAの学習アルゴリズムは，相互情報量の最小化に基づいて，分離信号の結合確率密度関数と分離信号の周辺確率密度関数の積との統計的距離を最小にする線形変換を求めている [19][21]．これにより導かれたICAの学習アルゴリズムでは，分離信号の対数周辺確率密度関数の導関数を知ることがICAの学習過程で必要となる [21]．一般に，簡単な関数，例えばシグモイド型関数などが，この非線形関数の代わりに使われている [20][25][42]．しかしながら，シグモイド型関数などでは，原信号の確率分布の形状によっては近似関数として適当でない場合があり，このときは原信号の推定が正確にできない．BSSの分離性能を向上させるためには，この

非線形関数をできるだけ正確に記述することが求められる．

そこで，本章では，この記述にRBFネットワークを使用したICAの学習アルゴリズム(RBF-ICA)を提案する．さらに，従来のICAとRBF-ICAを併用した手法(ハイブリッドICA)を提案し，信号分離の収束スピードと精度の更なる向上を実現した．提案手法の有効性は，計算機シミュレーションにより確認された．

2.2 瞬時混合ブラインド信号分離

本節では，瞬時混合ブラインド信号分離(BSS)問題について定式化を行う．

今，N個の原信号が線形的に混合され，P個のセンサーで観測されたとする．このとき，観測信号$\boldsymbol{x}(t)$は以下のように与えられる．

$$\boldsymbol{x}(t) = \boldsymbol{A}\boldsymbol{s}(t) \quad (t = 0, 1, 2, \ldots), \tag{2.1}$$

ここで，$\boldsymbol{A}$は混合行列と呼ばれ，P行N列の行列である．tは$\boldsymbol{x}(t)$を観測する離散時間を表す．$\boldsymbol{s}(t)$は原信号を表し，大きさNのベクトルである．

$$\boldsymbol{s}(t) = (s_1(t), \ldots, s_N(t))^T, \tag{2.2}$$

ここで，$s_i(t)$は第i番目の信号源から生成される原信号を表す．また，Tは転置を表す．観測信号もベクトルで表現され，その大きさはPである．

$$\boldsymbol{x}(t) = (x_1(t), \ldots, x_P(t))^T, \tag{2.3}$$

ここで，$x_i(t)$は第i番目のセンサーで観測される観測信号を表す．瞬時混合BSS問題では，信号源やセンサーは固定されていると仮定されるため，混合行列$\boldsymbol{A}$は信号源とセンサーの距離に応じた原信号の減衰のみで決定される．そのため，混合行列$\boldsymbol{A}$は時間tによって変化しない実数行列となる．

今までの議論では，原信号の数Nとセンサーの数Pは異なるとした．しかし，センサーの数が原信号の数よりも多い場合，すなわち，$N < P$のときは，主成分分

析 (Principle Component Analysis, PCA) などの次元縮約の手法を使うと，P の次元を N に落とすことができる [28]．そのため，今後は $P = N$，すなわち，混合行列 A が正方行列であるとして議論する．$N < P$ の場合は，過完備表現 (Overcomplete Representations)[81][82] と呼ばれ，一般には解けない．

瞬時混合 BSS 問題を解くための一般的な手法では，観測信号 $\boldsymbol{x}(t)$ に N 行 N 列の実数行列 $\boldsymbol{W}$ を作用させて得られる信号 $\boldsymbol{y}(t)$ を原信号 $\boldsymbol{s}(t)$ の推定値とする．

$$\boldsymbol{y}(t) = \boldsymbol{W}\boldsymbol{x}(t), \tag{2.4}$$

ここで，$\boldsymbol{W}$ は分離行列，$\boldsymbol{y}(t)$ は分離信号と呼ばれ，

$$\boldsymbol{y}(t) = (y_1(t), \ldots, y_N(t))^T, \tag{2.5}$$

である．$N = 2$ の場合の瞬時混合 BSS 問題を図 2.1 に示す．もし分離行列が混合行列の逆行列になったとき，すなわち，$\boldsymbol{W} = \boldsymbol{A}^{-1}$ のとき，分離信号 $\boldsymbol{y}(t)$ と原信号 $\boldsymbol{s}(t)$ は一致し，瞬時混合 BSS 問題は解けたことになる．例えば，信号源とセンサーの位置が既知で原信号の減衰係数が計算できるならば，混合行列 $\boldsymbol{A}$ が分かるので分離行列 $\boldsymbol{W}$ は簡単に求めることができる．しかし，一般的には，原信号の確率分布や混合過程は未知であることが多い．このような先験的情報が得られない場合でも，次に述べる独立成分分析を用いると最適な分離行列を推定することができる．

2.3 独立成分分析

本節では，独立成分分析 (ICA) の代表的な学習アルゴリズムについて説明をするため，そのアルゴリズムの導出過程で使われる相互情報量最小化 [19][21] と自然勾配法 [65] について述べる．

ICA は，原信号 $\boldsymbol{s}(t)$ の確率分布などの特性や混合過程に関する先験的情報をもたずに，原信号の統計的独立性のみを手がかりとして，観測信号 $\boldsymbol{x}(t)$ から原信号 $\boldsymbol{s}(t)$

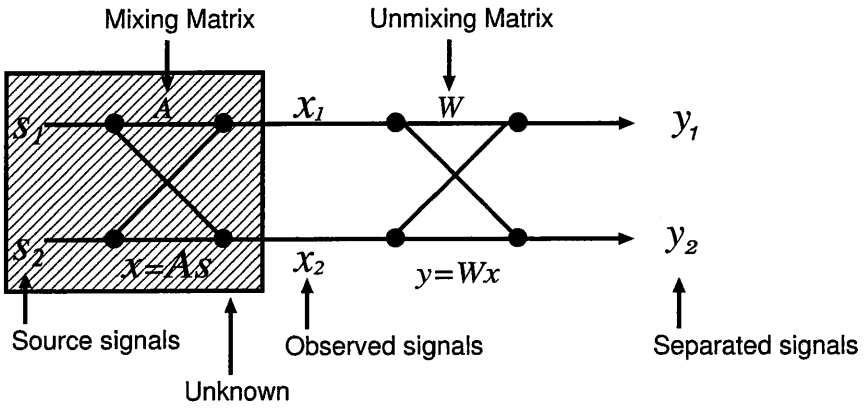


図 2.1: A schematic diagram of the BSS problem (N=2).

を推定する手法である．ICA の目的は，分離信号 $\boldsymbol{y}(t)$ が，できるだけ互いに統計的に独立になるような分離行列 $\boldsymbol{W}$ を決定することである．ただし，原信号が統計的に独立なガウス分布である場合，分離信号を統計的に独立にする分離行列は無数に存在するため，ガウス分布に対して ICA を適用することはできない．

2.3.1 相互情報量最小化

ICA の学習アルゴリズムは，統計的独立性の規範の選び方によって様々なものが提案されている．その中の代表的なものとして，分離信号の相互情報量を最小化する規範に基づいた学習アルゴリズムがある．まず，分離信号の結合エントロピーを $H(\boldsymbol{Y})$，その各成分のエントロピーを $H(Y_i)$ とすると，相互情報量は以下のように定義される [29][30].

$$D(\boldsymbol{W}) = \sum_{i=1}^{N} H(Y_i) - H(\boldsymbol{Y}), \tag{2.6}$$

ここで,

$$H(\boldsymbol{Y}) = -\int p_{\boldsymbol{Y}}(\boldsymbol{y}) \log p_{\boldsymbol{Y}}(\boldsymbol{y})\, d\boldsymbol{y}, \tag{2.7}$$

$$H(Y_i) = -\int p_{Y_i}(y_i) \log p_{Y_i}(y_i)\, dy_i. \tag{2.8}$$

式(2.6)は，分離信号の結合確率密度関数$p_{\boldsymbol{Y}}(\boldsymbol{y})$とその周辺確率密度関数の積$\prod_{i=1}^{N} p_{Y_i}(y_i)$との間のカルバック情報量(Kullback-Leibler Divergence, KLD)として以下のように表現することもできる.

$$D(\boldsymbol{W}) = \int p(\boldsymbol{y}) \log \frac{p(\boldsymbol{y})}{\prod_{i=1}^{N} p(y_i)}\, d\boldsymbol{y}. \tag{2.9}$$

今後は，簡略化のため確率密度関数の添字は省略する．ここでは，分離行列$\boldsymbol{W}$を適切に選ぶことによって相互情報量Dを最小化するのが目的であるため，Dの引数は$\boldsymbol{W}$になる.

KLDは2つの確率分布間の統計的距離を計る尺度の1つで，その値は0以上の値をとり，0のときは2つの確率分布は一致する．そのため，分離信号の相互情報量が0になるとき,

$$p(\boldsymbol{y}) = \prod_{i=1}^{N} p(y_i), \tag{2.10}$$

となり，分離信号は統計的に独立となる．よって，式(2.9)に最急降下法を適用して相互情報量を最小化する$\boldsymbol{W}$を求める．ただし，ICAでは統計的独立性のみを考慮するため，原信号の大きさと順番を決定することはできない．すなわち,

$$\boldsymbol{W} = \boldsymbol{P}\boldsymbol{D}\boldsymbol{A}^{-1}, \tag{2.11}$$

が解となる．ここで，$\boldsymbol{P}$は置換行列，$\boldsymbol{D}$は対角行列であり，この不定性が解の中に残る.

式 (2.9) は，$p(\boldsymbol{y}) = p(\boldsymbol{x})/|\det \boldsymbol{W}|$ に注意すると，以下のように変形できる．

$$D(\boldsymbol{W}) = \int p(\boldsymbol{x}) \log p(\boldsymbol{x})\, d\boldsymbol{x} - \log|\det \boldsymbol{W}| - \sum_{i=1}^{N} \int p(\boldsymbol{x}) \log p(y_i)\, d\boldsymbol{x}. \tag{2.12}$$

ここで，$\det \boldsymbol{W}$ は $\boldsymbol{W}$ の行列式，$|\cdot|$ は絶対値を表す．式 (2.12) の右辺の各項は，$\boldsymbol{W}$ に関して偏微分が計算できる．よって，分離行列の更新式は，基本的な最急降下法を適用することにより以下のように与えられる．

$$\begin{aligned} \Delta \boldsymbol{W} &= -\eta \frac{\partial D(\boldsymbol{W})}{\partial \boldsymbol{W}} \\ &= \eta \left(\boldsymbol{I} - \left\langle \boldsymbol{\phi}(\boldsymbol{y}) \boldsymbol{y}^T \right\rangle\right) \left(\boldsymbol{W}^T\right)^{-1}, \end{aligned} \tag{2.13}$$

ここで，$\Delta \boldsymbol{W}$ は $\boldsymbol{W}$ の更新量，η は学習係数，$< \cdot >$ は期待値であり，実際の計算では時間平均に置き換えられる．また，$\boldsymbol{\phi}(\boldsymbol{y})$ は，以下の成分をもつベクトルである．

$$\boldsymbol{\phi}(\boldsymbol{y}) = (\phi(y_1), \ldots, \phi(y_N)), \tag{2.14}$$

$$\phi(y_i) = -\frac{\partial \log p(y_i)}{\partial y_i}. \tag{2.15}$$

基本的な最急降下法はユークリッド空間の直交座標系で定義されており，偏微分方向と最急降下方向は一致する．しかしながら，行列の作る空間はリーマン空間であり，空間上の場所によって距離尺度が変化する．よって，リーマン空間上で基本的な最急降下法を適用した場合，通常その方向は最急降下の方向でない．リーマン空間における最急降下の方向は自然勾配と呼ばれ，この空間上で学習を行う場合，次に述べる自然勾配法がよく用いられる．

2.3.2 自然勾配法

スカラー関数 $D(\boldsymbol{W})$ における最急降下の方向は，$\boldsymbol{W}$ を現在の場所から $d\boldsymbol{W}$ だけ微小変化させとき，$d\boldsymbol{W}$ の大きさを一定として，$D(\boldsymbol{W})$ の変化量，

$$D(\boldsymbol{W} + \epsilon d\boldsymbol{W}) - D(\boldsymbol{W}), \tag{2.16}$$

が最も減少するような$d\boldsymbol{W}$の方向である．ϵは微小な定数である．$d\boldsymbol{W}$は行列なので，基本的な最急降下法を適用すると収束が遅くなってしまう．分離行列の学習を高速に行うためには，自然勾配法を適用する必要がある．以下に正則な行列空間における自然勾配法について簡単に説明する．

まず，単位行列$\boldsymbol{I}$における微少変化量$d\boldsymbol{X}$の大きさを，フロベニウスノルムにより以下のように定義する．

$$\begin{aligned} ||d\boldsymbol{X}||_{\boldsymbol{I}} &= \mathrm{tr}\left(d\boldsymbol{X}d\boldsymbol{X}^T\right) \\ &= \sum (dX_{ij})^2, \end{aligned} \tag{2.17}$$

ここで，trは行列のトレースを表す．分離行列は正則行列なので，リー群をなす[59]．リー群の性質から，$\boldsymbol{W}$における微少変化量$d\boldsymbol{W}$は，

$$(\boldsymbol{W}+d\boldsymbol{W})\boldsymbol{W}^{-1} = \boldsymbol{I} + d\boldsymbol{W}\boldsymbol{W}^{-1}, \tag{2.18}$$

により，$\boldsymbol{I}$における微少変化量$d\boldsymbol{W}\boldsymbol{W}^{-1}$に写される．この微少変化量の大きさは，リー群の変換に対して一定に保たれる[65]．よって，

$$\begin{aligned} ||d\boldsymbol{W}||_{\boldsymbol{W}} &= ||d\boldsymbol{W}\boldsymbol{W}^{-1}||_{\boldsymbol{I}} \\ &= \mathrm{tr}\left(d\boldsymbol{W}\boldsymbol{W}^{-1}(\boldsymbol{W}^{-1})^T d\boldsymbol{W}^T\right), \end{aligned} \tag{2.19}$$

となり，$\boldsymbol{W}$における微小変化量の大きさが計算できる．

$D(\boldsymbol{W})$における最急降下の方向を求めるには，この$||d\boldsymbol{W}||_{\boldsymbol{W}}$を一定とする拘束条件，すなわち，

$$\mathrm{tr}\left(d\boldsymbol{W}\boldsymbol{W}^{-1}(\boldsymbol{W}^{-1})^T d\boldsymbol{W}^T\right) = c, \tag{2.20}$$

の下，式(2.16)を最小にする必要がある．ここでcは1以下の定数である．式(2.16)においてϵの2次以上の項を無視すると，

$$D(\boldsymbol{W}+\epsilon d\boldsymbol{W}) - D(\boldsymbol{W}) = \epsilon \mathrm{tr}\left(\left(\frac{\partial D(\boldsymbol{W})}{\partial \boldsymbol{W}}\right)^T d\boldsymbol{W}\right), \tag{2.21}$$

となる．よって，式(2.20)と式(2.21)に対してラグランジュ未定乗数法を用いると，

$$\frac{\partial}{\partial(d\boldsymbol{W})}\left\{\epsilon\mathrm{tr}\left(\left(\frac{\partial D(\boldsymbol{W})}{\partial \boldsymbol{W}}\right)^{T} d\boldsymbol{W}\right)\right.$$
$$\left.-\lambda\left(c-\mathrm{tr}\left(d\boldsymbol{W}\boldsymbol{W}^{-1}(\boldsymbol{W}^{-1})^{T}d\boldsymbol{W}^{T}\right)\right)\right\}=0. \tag{2.22}$$

これを解くと，最適な方向の$d\boldsymbol{W}$が求まる．

$$d\boldsymbol{W}=-\frac{\epsilon}{2\lambda}\frac{\partial D(\boldsymbol{W})}{\partial \boldsymbol{W}}\boldsymbol{W}^{T}\boldsymbol{W}. \tag{2.23}$$

すなわち，正則な行列空間における自然勾配法は，基本的な最急降下法に右から$\boldsymbol{W}^{T}\boldsymbol{W}$を掛ければ良いことが分かる[7][29][31][65]．

よって，式(2.13)に右から$\boldsymbol{W}^{T}\boldsymbol{W}$を掛けると，

$$\begin{aligned}\Delta\boldsymbol{W} &= -\eta\frac{\partial D(\boldsymbol{W})}{\partial \boldsymbol{W}}\boldsymbol{W}^{T}\boldsymbol{W}\\ &= \eta\left(\boldsymbol{I}-\left\langle\boldsymbol{\phi}(\boldsymbol{y})\boldsymbol{y}^{T}\right\rangle\right)\boldsymbol{W},\end{aligned} \tag{2.24}$$

となり，これが相互情報量最小化によるICAの学習アルゴリズムになる[21]．式(2.24)に従い$\boldsymbol{W}$を更新すると分離行列が収束していき，やがて解が得られる．しかしながら，相互情報量最小化と自然勾配法により導かれた式(2.24)には，式(2.15)で表される$\phi(y_i)$の具体的な関数系が必要となる．

2.3.3 ICAの学習で必要な非線形関数の近似

一般的な方法は，$\phi(y_i)$を簡単な非線形関数，例えば，シグモイド型関数などで代用する方法である．また，$\phi(y_i)$に含まれる分離信号の周辺確率密度関数$p(y_i)$を，グラムシャリエ展開[21]やエッジワース展開[19]で近似したりする方法もある．しかしながら，いずれの方法も，原信号の確率分布の形状によっては信号分離を正確に行うことができない場合が存在する[60]．

また，確率密度関数をニューラルネットワークなどで近似して $\phi(y_i)$ を求める手法も提案されているが [61]，確率密度関数の微分を求めるときに離散的な差分で近似しているため，確率密度関数が大きく変化する場合などでは，近似精度に問題が生じる．確率密度関数をパラメトリックな方法で近似するアルゴリズムも提案されているが [62]，パラメータ毎に学習が必要となり，学習アルゴリズムが複雑になる．確率密度関数を近似することなしに，直接的に $\phi(y_i)$ を近似する方法も提案されているが [63]，スプライン関数が用いられているため滑らかな導関数を得ることが難しい．そこで，次節では，これらの問題点を解決するため，RBF ネットワークを用いて $\phi(y_i)$ を近似する手法 (RBF-ICA) を提案する．

2.4 提案する RBF ネットワークによる ICA

本節では，提案手法である ICA の学習過程で必要とされる非線形関数の近似に RBF ネットワークを利用する方法を説明する．

RBF ネットワークは，任意の非線形関数を動径基底関数 (Radial Basis Function, RBF) で展開する方法であり，複雑な関数を近似する能力に優れている [33]-[36]．動径基底関数とは，その関数値が関数の中心から描かれる等高線によって決まる関数であり，代表的な例としてガウス型関数がある．RBF ネットワークは中間層が 1 層のネットワークであるので，他のニューラルネットワークの手法に比べ，局所解をもたない利点がある．

RBF ネットワークを ICA に利用した例はあるが [64]，そこでは，非線形写像の逆写像の近似に RBF ネットワークを用いており，RBF ネットワークを使う目的が提案手法と全く異なる．提案手法における RBF ネットワークの利用法は，ガウス型関数を基底関数とする RBF ネットワークの出力の微分が，正確に，簡単に計算できることに注目し，分離信号の対数分布曲線を RBF ネットワークで近似し，さらに，それを微分することにより対数確率密度関数の微分，すなわち，式 (2.15) で表される

非線形関数を得ることである.

2.4.1 対数度数分布曲線の微分による非線形関数の近似

RBF ネットワークを利用した，式 (2.15) で表される $\phi(y_i)$ の近似は以下のように行われる．まず，分離信号 y_i のヒストグラムを計算するために，y_i の分布する領域を K 個の等間隔な区間に分割し，その区間内に属する y_i の数を調べる．次に，得られたヒストグラムから対数ヒストグラムを計算する．ただし，ヒストグラムの値が 0 のときは，その対数は 0 とする．そして，それを RBF ネットワークで近似することにより，微分可能な連続関数である対数度数分布曲線 $\log n(y_i)$ を得る．この様子を図 2.2 に示す.

対数度数分布曲線 $\log n(y_i)$ の微分は，以下に示す通り，対数確率密度関数の微分と等しくなるので,

$$\begin{aligned} \frac{\partial \log p(y_i)}{\partial y_i} &= \frac{\partial(\log n(y_i) - \log M)}{\partial y_i} \\ &= \frac{\partial \log n(y_i)}{\partial y_i}, \end{aligned} \tag{2.25}$$

$\phi(y_i)$ は，対数度数分布曲線 $\log n(y_i)$，すなわち，RBF ネットワークの出力を微分することによって得られる．ここで，M は度数の総数，すなわち，$p(y_i) = n(y_i)/M$ である．次節では，RBF ネットワークの具体的な構造ついて述べる.

2.4.2 RBF ネットワークによる非線形関数の表現と RBF-ICA

ここでは，RBF ネットワークによる非線形関数の表現方法と RBF-ICA の学習アルゴリズムについて述べる.

RBF ネットワークは中間層が 1 層のネットワークである．図 2.3 は，RBF ネットワークの構造を示す．ここでは，基底関数として次のようなガウス型関数を用いる.

$$h_l(y_i) = \exp\left(-\frac{(y_i - a_l)^2}{b_l}\right), \tag{2.26}$$

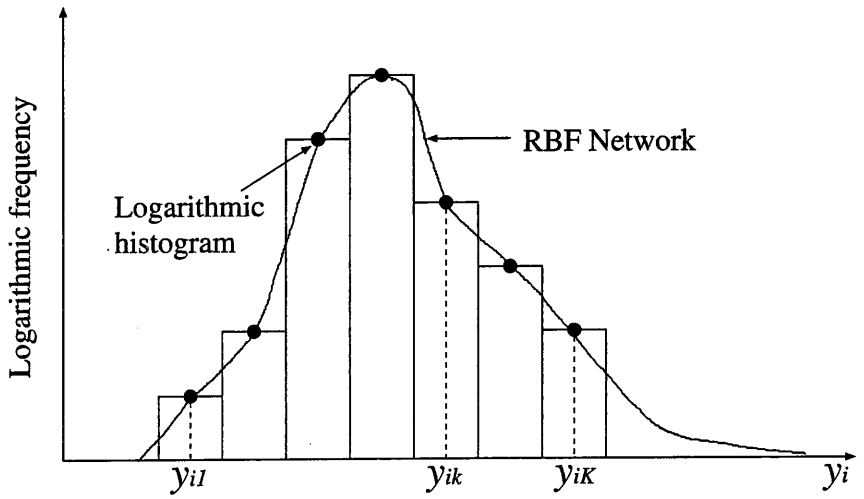


図 2.2: The logarithmic frequency curve approximated by the RBF network. y_{ik} is a middle point of the k-th interval ($k = 1, 2, \cdots, K$).

ここで，y_i はRBFネットワークへの入力信号であり，$h_l(y_i)$ は l 番目の基底関数，a_l と b_l はガウス型関数の形状を決めるパラメータである．a_l は基底関数の中心の位置を表し，b_l は基底関数の広がりを表す．RBFネットワークの出力は，この基底関数の重み付き加算で与えられる．

$$f(y_i) = \sum_{l=1}^{L} w_l h_l(y_i), \tag{2.27}$$

ここで，w_l は $h_l(y_i)$ におけるネットワークの結合の重みであり，L は基底関数の総数である．

任意の非線形関数は，重み w_l を適切に調節することによって近似される．その重み w_l は，次式のような誤差評価関数を最小化する教師あり学習によって調節される．

$$E(\boldsymbol{w}, \boldsymbol{a}, \boldsymbol{b}) = \frac{1}{2} \sum_{k=1}^{K} \left(\log n(y_{ik}) - f(y_{ik}) \right)^2, \tag{2.28}$$

ここで，$\log n(y_{ik})$ は入力信号 y_{ik} のための教師信号であり，K は教師信号の総数である．K はヒストグラムにおける区間の総数と等しくなる．また，$\boldsymbol{w}$ は重みベクト

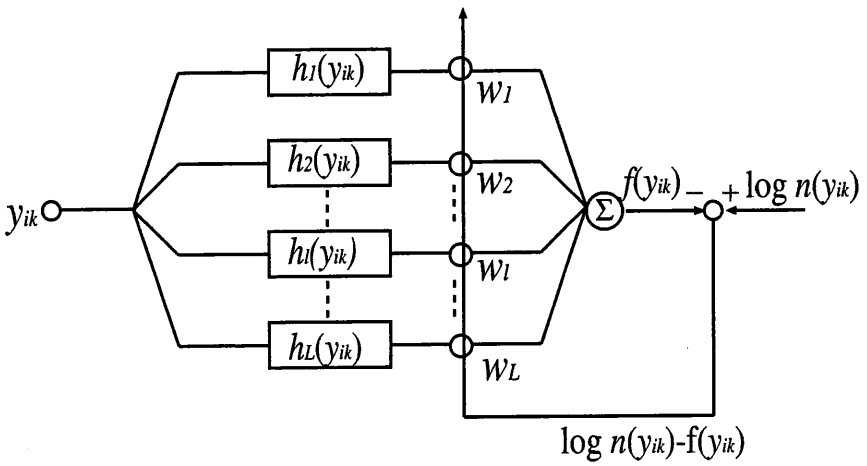


図 2.3: A Structure of the RBF network.

ル，$\boldsymbol{a}$ と $\boldsymbol{b}$ は基底関数のパラメータベクトルであり，以下のように表される．

$$\boldsymbol{w} = (w_1, \ldots, w_L)^T, \tag{2.29}$$

$$\boldsymbol{a} = (a_1, \ldots, a_L)^T, \tag{2.30}$$

$$\boldsymbol{b} = (b_1, \ldots, b_L)^T. \tag{2.31}$$

基本的な最急降下法により，w_l の更新式は以下のように与えられる．

$$\begin{aligned} \Delta w_l &= -\alpha \frac{\partial E(\boldsymbol{w}, \boldsymbol{a}, \boldsymbol{b})}{\partial w_l} \\ &= \alpha \sum_{k=1}^{K} (\log n(y_{ik}) - f(y_{ik}))\, h_l(y_{ik}), \end{aligned} \tag{2.32}$$

ここで，αは学習係数である．a_lとb_lもw_lと同様な方法で決定される．

$$\begin{aligned}\Delta a_l &= -\beta\frac{\partial E(\boldsymbol{w},\boldsymbol{a},\boldsymbol{b})}{\partial a_l} \\ &= 2\beta\sum_{k=1}^{K}\{(\log n(y_{ik}) - f(y_{ik}))\, w_l h_l(y_{ik})(y_{ik}-a_l)/b_l\}, \end{aligned} \tag{2.33}$$

$$\begin{aligned}\Delta b_l &= -\gamma\frac{\partial E(\boldsymbol{w},\boldsymbol{a},\boldsymbol{b})}{\partial b_l} \\ &= \gamma\sum_{k=1}^{K}\{(\log n(y_{ik}) - f(y_{ik}))\, w_l h_l(y_{ik})(y_{ik}-a_l)^2/b_l^2\}, \end{aligned} \tag{2.34}$$

ここで，βとγは学習係数である．

$\phi(y_i)$は，対数度数分布曲線$\log n(y_i)$をRBFネットワークによって近似した後，式(2.27)を微分することによって得られる．

$$\begin{aligned}\phi(y_i) &= -\frac{\partial \log p(y_i)}{\partial y_i} = -\frac{\partial \log n(y_i)}{\partial y_i} \simeq -\frac{\partial f(y_i)}{\partial y_i} \\ &= -2\sum_{l=1}^{L} w_l \frac{(y_i-a_l)}{b_l}\exp\left(-\frac{(y_i-a_l)^2}{b_l}\right). \end{aligned} \tag{2.35}$$

式(2.35)の値が，ICAの式(2.24)で表される学習過程で使われる．

図2.4は，$N=2$の場合のRBF-ICAによる信号分離過程を示している．図に示されている通り，RBF-ICAには，2つの異なる学習過程がある．1つは，ICAにおける分離行列$\boldsymbol{W}$の学習であり，もう1つは，非線形関数$\phi(y_i)$を近似するためのRBFネットワークにおける各パラメータの学習である．そのため，RBF-ICAは，学習時間が従来法に比べ大きくなる欠点がある．分離精度と学習時間は，いわゆるトレードオフの関係になる．そこで，次節では，信号分離の収束スピードと精度の更なる向上を目指すために，従来のICAとRBF-ICAの長所を生かした，ハイブリッドICAを提案する．

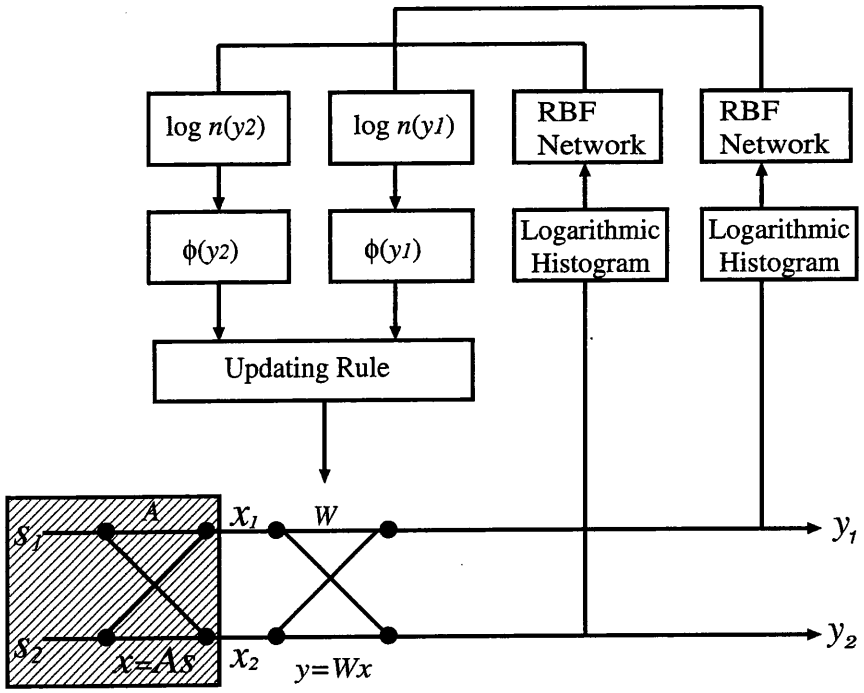


図 2.4: The learning process of the ICA by using the RBF networks (RBF-ICA).

2.4.3 ハイブリッドICA

ここで言うハイブリッドICAは，従来のICAとRBF-ICAの2つを組み合わせた手法である．前者は，シグモイド型関数により学習が高速に行える利点があり，後者は，正確に求めた非線形関数により精度の高い信号分離が行える利点がある．そのため，ハイブリッドICAは，2つの長所を生かして，最初に従来のICAを採用して高速に信号分離を行い，次にRBF-ICAに切り替えることによって，より精度の高い信号分離を実現する．

切り替えは，従来のICAの学習がほぼ収束した時点で行う．収束状況の判定は，次式のような分離行列 $\boldsymbol{W}$ の更新量の大きさを用いる．

$$||\Delta \boldsymbol{W}|| = \sum_{i=1}^{N}\sum_{j=1}^{N}\Delta W_{ij}^2, \tag{2.36}$$

ここで，$||\cdot||$はフロベニウスノルムを表す．フロベニウスノルムは，行列空間において2つの行列間の距離尺度の1つとして使われる．ICAの学習が収束するに従い$\Delta \boldsymbol{W}$は零行列に近くなるので，$||\Delta \boldsymbol{W}||$も最終的に0に近くなる．そのため，ハイブリッドICAでは，$||\Delta \boldsymbol{W}||$を，ICAの学習の収束状況を判定するための評価基準として採用する．従来のICAからRBF-ICAへの切り替えは，$||\Delta \boldsymbol{W}||$が，指定したしきい値$\epsilon$より小さくなったときに行われる．$\boldsymbol{W}$の値は，学習の切り替え時に従来のICAからRBF-ICAに引き継がれる．

2.5 計算機シミュレーション結果

本節では，提案手法の有効性を調べるために，RBF-ICAの計算機シミュレーション結果を，他手法との比較とともに述べる．また，実際の音声信号を用いたハイブリッドICAの結果を示す．

2.5.1 RBF-ICA

2.5.1.1 実験条件

計算機シミュレーションは，$N = 2$の場合で行った．図(2.5(a))は，正弦波$0.5\sin(30\omega_0 t)$ $(t = 0, \ldots, 48,000)$で表される，原信号$s_1(t)$の最初の1,600点を示す．ここで，ω_0は基本角周波数である．図2.5(b)は，範囲$[-0.3, 0.3]$の間で一様分布をする，もう1つの原信号$s_2(t)$を示す．観測信号$\boldsymbol{x}(t)$は，原信号$\boldsymbol{s}(t)$を次式のように計算機上で混合して作る．

$$x_1(t) = 0.7s_1(t) + 0.4s_2(t), \tag{2.37}$$

$$x_2(t) = 0.5s_1(t) + 0.6s_2(t). \tag{2.38}$$

すなわち，混合行列は，

$$\boldsymbol{A} = \begin{pmatrix} 0.7 & 0.4 \\ 0.5 & 0.6 \end{pmatrix}, \tag{2.39}$$

となる．表2.1に今回のRBF-ICAにおけるRBFネットワークの実験パラメータを示す．

表 2.1: Parameters for the RBF network

Range of histogram	$-1 \sim 1$
Number of intervals for histogram (K)	33
Number of the basis functions (L)	66
Initial value of the parameter b_l	0.005
Learning rate α for the wight w_l	0.1
Learning rate β for the parameter a_l	0
Learning rate γ for the parameter b_l	0

2.5.1.2 実験結果

図2.6は，観測信号$\boldsymbol{x}(t)$を示す．RBF-ICAによって得られた分離信号$\boldsymbol{y}(t)$は，図2.7に示される．図2.7より，分離信号は式(2.11)で示されるICAの不定性を除いて，原信号$\boldsymbol{s}(t)$とほぼ同一であることが分かる．これに対して，従来のICAで用いられるシグモイド型関数は，一様分布をもつ$\phi(y_i)$をうまく記述できないため，信号分離がうまくできていない．従来のICAにおける$\phi(y_i)$として，ここでは，

$$\phi(y_i) = \frac{1}{1+\exp(-y_i)}, \tag{2.40}$$

を用いた．その他，$-1+2(1/(1+\exp(-y_i)))$や$2\tanh(y_i)$[20]などの関数も試してみたが，結果はほとんど変わらなかった．

図2.8は，RBFネットワークによって近似した$\phi(y_1)$と$\phi(y_2)$を示す．ここで，図2.8に示されている点は，次式で表される$\Delta \log n(y_{1k})$と$\Delta \log n(y_{2k})$を示す．

$$\Delta \log n(y_{ik}) \;=\; \{\log n(y_{ik}+\Delta_k/2) - \log n(y_{ik-1}-\Delta_k/2)\}/\Delta_k, \tag{2.41}$$

ここで，$\Delta_k(=y_{ik}-y_{ik-1})$ は，図 2.2 で示されたヒストグラムの区間を示す．図 2.8 から，RBF ネットワークの出力の微分値は，$\Delta \log n(y_{ik})$ とよく一致していることが分かる．

比較のため，非線形関数の近似を，関数近似でよく用いられるスプライン関数を使って行ってみた．ここでは，自然スプライン関数を採用した [66][67]．図 2.9 は，スプライン関数で近似した $\phi(y_1)$ と $\phi(y_2)$ を示す．自然スプライン関数は，3 回微分可能であるが，一方，RBF は無限回微分可能である．この違いが，図 2.8 と図 2.9，及び，図 2.14 と図 2.15 で示されているような，非線形関数の近似結果に表れていることが分かる．すなわち，これらの図から RBF ネットワークによる近似結果は，スプライン関数のそれよりも，より滑らかである．このことが，以下に述べる信号分離の精度に貢献していると思われる．

ICA による信号分離精度を評価するために，次式で定義される Performance Index(PI) がよく使われる [21]．

$$\mathrm{PI} = \sum_{i=1}^{N}\left(\sum_{j=1}^{N}\frac{|P_{ij}|}{\max_q |P_{iq}|}-1\right)+\sum_{j=1}^{N}\left(\sum_{i=1}^{N}\frac{|P_{ij}|}{\max_q |P_{qj}|}-1\right), \qquad (2.42)$$

ここで，P_{ij} は行列 $\boldsymbol{P}(=\boldsymbol{WA})$ の成分である．PI は常に正であり，小さい値ほど良い分離性能を示す．

図 2.10(a) と図 2.10(b) は，従来の ICA，スプライン関数を用いた ICA，提案する RBF-ICA を用いて得られた PI を示す．このとき，式 (2.24) における学習係数は，$\eta=0.1$ と $\eta=0.01$ とした．表 2.2 に，ICA の学習回数 2,000 回での，それぞれの手法の PI を示す．

図 2.10 と表 2.2 から，RBF-ICA が，他の 2 つの手法よりも分離性能の点で優れていることが分かる．スプライン関数を用いた ICA は確かに分離はできているが，その精度は図 2.10(a) と図 2.10(b) で示されているように，学習係数 η に敏感であることがわかる．ただし，計算量は RBF-ICA よりもスプライン関数を用いた ICA の方

が少ないので．精度をあまり重視せず計算時間を節約したい場合には，スプライン関数を用いても問題ないであろう．

表 2.2: PIs after 2,000 steps of the ICA learning.

	RBF-ICA	Spline function	Conventional ICA
$\eta = 0.1$	0.00377	0.162	2.99
$\eta = 0.01$	0.00379	0.00410	2.99

2.5.2 ハイブリッドICA

2.5.2.1 実験条件

計算機シミュレーションは，ハイブリッドICAと，比較のため，従来のICAをハイブリッドせずにそのまま継続したものと，ハイブリッド時にRBFネットワークの代わりにスプライン関数を用いたものの，計3つで行った．

計算機シミュレーションでは，2人の男性話者の音声信号で，原信号 $s_1(t)$ として「こころ」(夏目漱石) の「私はその人を常に先生と呼んでいた」，原信号 $s_2(t)$ として「城ノ崎にて」(志賀直哉) の「山手線の電車に跳ね飛ばされて怪我をした」を朗読したものを用いた [31]．ともに，サンプリング周波数16kHz，量子化ビット数16ビット，3秒間のモノラル音声であった．

観測信号 $\boldsymbol{x}(t)$ は，前節と同じ式 (2.39) で与えられる混合行列により原信号 $\boldsymbol{s}(t)$ を計算機上で混合して作る．図2.11と図2.12は，それぞれ，原音声信号 $\boldsymbol{s}(t)$ と観測信号 $\boldsymbol{x}(t)$ を示す．今回のシミュレーションでは，しきい値 ϵ を 10^{-5} として，学習の切り替えを行った．従来のICAの $\phi(y_i)$ としては，前と同様に式 (2.40) を用いた．式 (2.24) における学習係数は，従来のICA，RBF-ICAともに $\eta = 0.1$ とし，スプライン関数の場合は $\eta = 0.01$ とした．表2.3に今回のハイブリッドICAのRBF-ICAにおけるRBFネットワークの実験パラメータを示す．

表 2.3: Parameters for the RBF network

Range of histogram	$-1 \sim 1$
Number of intervals for histogram (K)	33
Number of the basis functions (L)	66
Initial value of the parameter b_l	0.01
Learning rate α for the wight w_l	0.1
Learning rate β for the parameter a_l	0
Learning rate γ for the parameter b_l	0.000001

2.5.2.2 実験結果

図 2.13 は，ハイブリッド ICA によって得られた分離信号を示しており，この図から，信号分離が十分に行われていることが分かる．図 2.14 は，このシミュレーションにおいて RBF ネットワークによって近似された $\phi(y_1)$ と $\phi(y_2)$ を示す．一方，図 2.15 は，RBF ネットワークの代わりにスプライン関数によって近似された $\phi(y_1)$ と $\phi(y_2)$ を示す．RBF ネットワークによって近似された $\phi(y_i)$ の形状は，前節と同じくスプライン関数のものよりも，より滑らかであることが分かる．

図 2.16 は，従来の ICA のみ，RBF-ICA によるハイブリッド ICA，スプライン関数によるハイブリッド ICA，によって得られた PI を示す．この図から，RBF-ICA によるハイブリッド ICA が，他の手法よりも良い分離性能を与えていることが分かる．表 2.4 は，ICA の学習回数 1,000 回での，それぞれ手法の PI を示す．RBF-ICA によるハイブリッド ICA の信号分離精度が，他手法と比較して良いことが分かる．

表 2.4: PIs after 1,000 steps of the ICA learning.

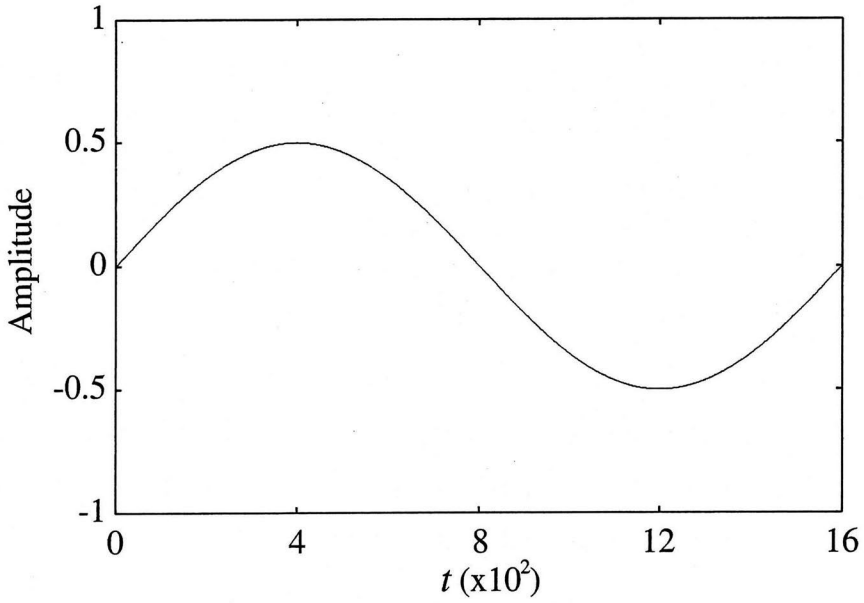
Hybrid (RBF network)	Hybrid (spline function)	Conventional ICA
0.00169	0.0273	0.0273

2.6 結言

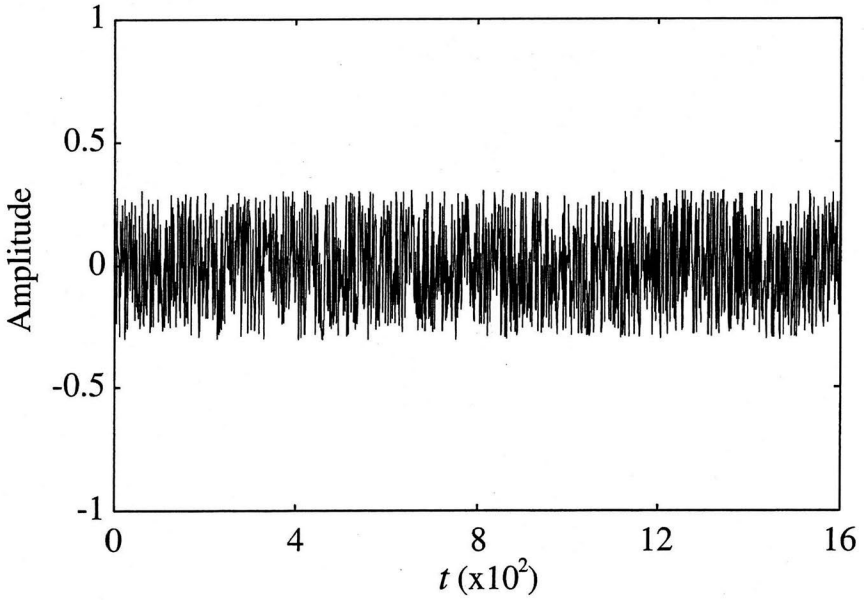
本章では，瞬時混合BSS問題を解くために，相互情報量最小化によって導かれたICAの学習アルゴリズを説明した後，そこで必要とされる非線形関数の記述をできるだけ正確に行うためにRBFネットワークを用いたRBF-ICAを提案し，信号分離精度の向上を目指した．また，従来のICAとRBF-ICAのそれぞれの長所を生かしたハイブリッドICAを提案した．

計算機シミュレーション結果は，提案したRBF-ICAが，シグモイド型関数を用いる従来のICAやスプライン関数を用いる手法よりも，良い分離性能を与えることを示した．また，RBF-ICAは，従来のICAではうまく分離できないような確率分布をもつ信号に対しても，正確に分離できることを示した．さらに，ハイブリッドICAを用いると，信号分離の収束スピードと精度の更なる向上が実現できることを示した．

今後の課題は，RBF-ICAの計算量をできるだけ削減し，次章で述べる周波数領域ICAにハイブリッドICAを適用することである．

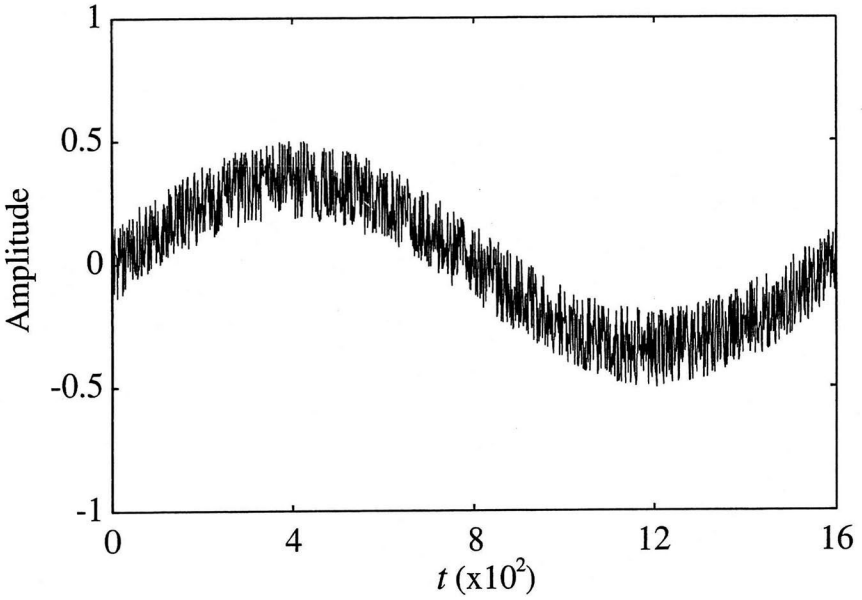


(a)

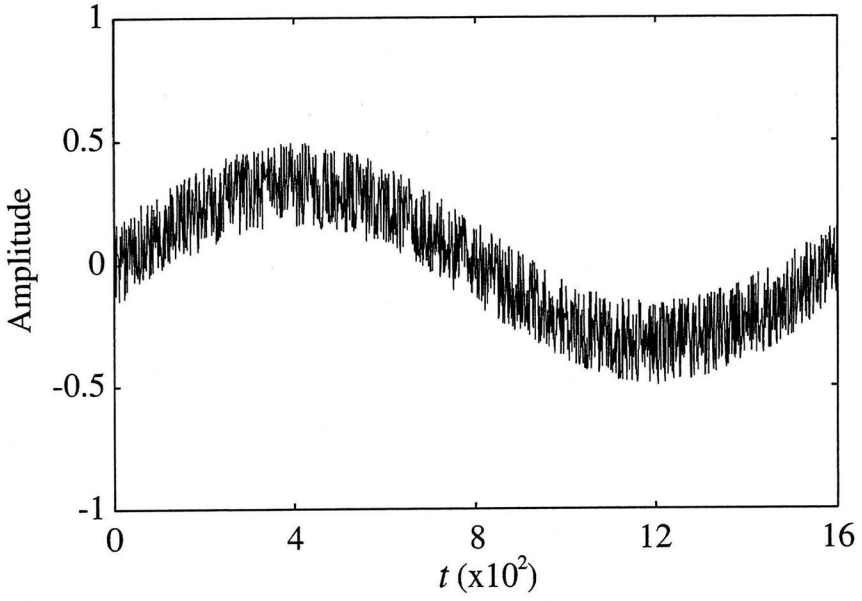


(b)

図 2.5: The source signals. (a) $s_1(t)$. (b) $s_2(t)$.

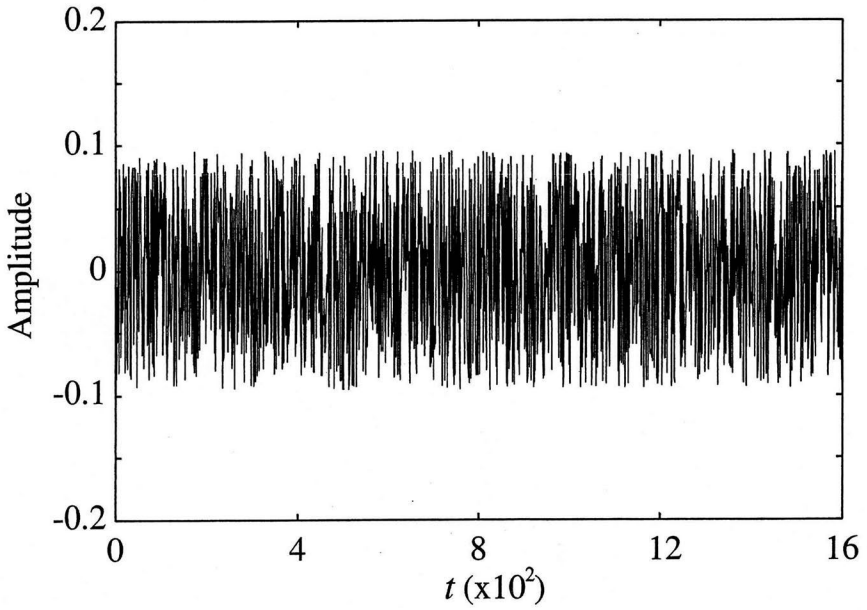


(a)

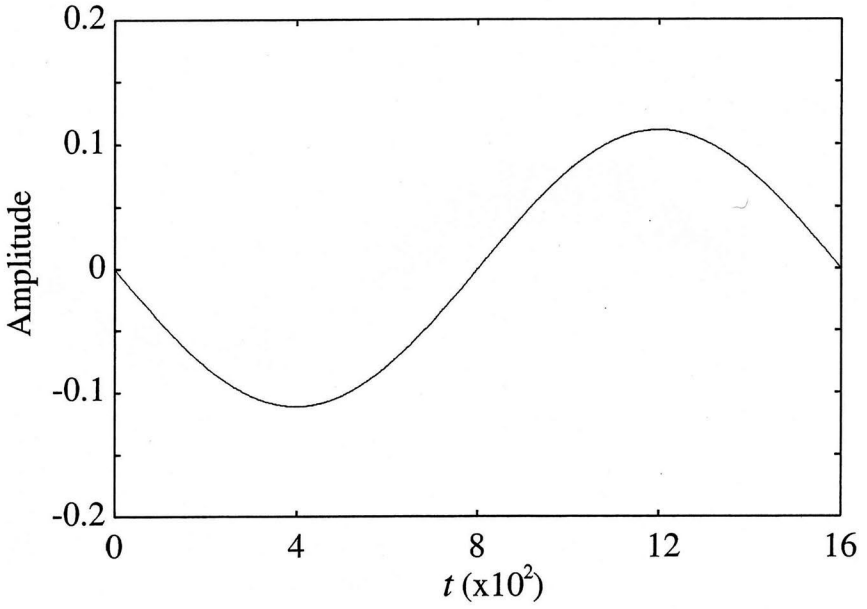


(b)

図 2.6: The mixed signals. (a) $x_1(t)$. (b) $x_2(t)$.

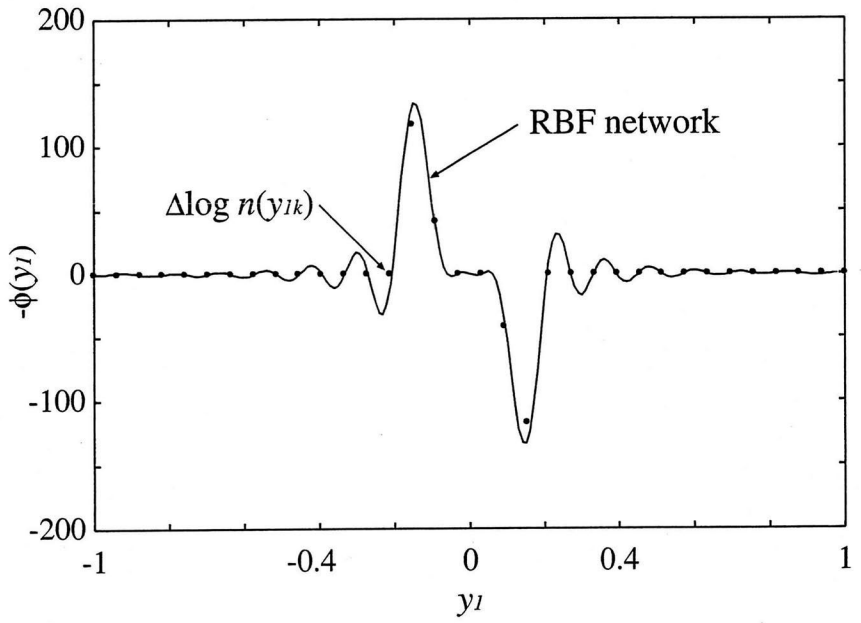


(a)

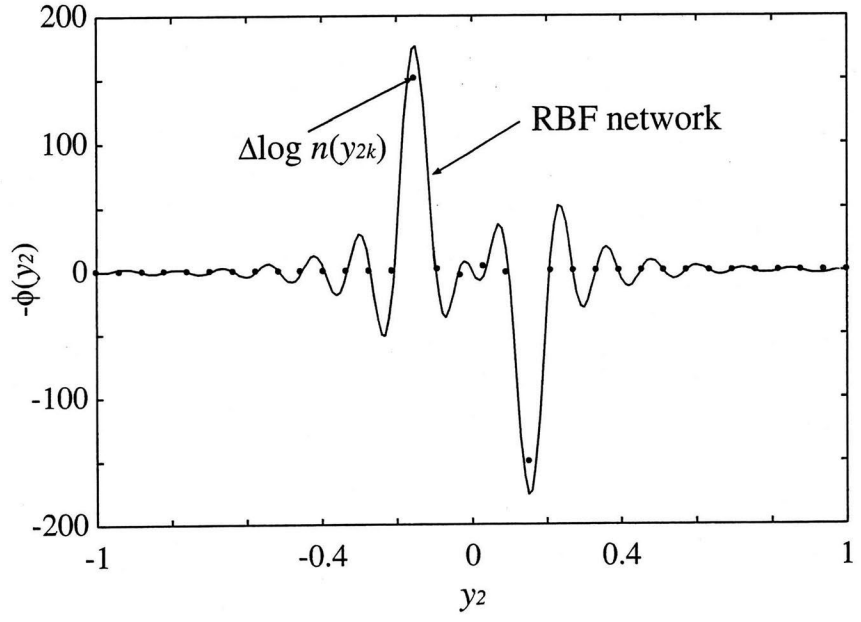


(b)

図 2.7: The separated signals obtained by using the RBF-ICA. (a) $y_1(t)$. (b) $y_2(t)$.



(a)



(b)

図 2.8: The nonlinear functions approximated by the RBF networks. (a) $\phi(y_1)$. (b) $\phi(y_2)$.

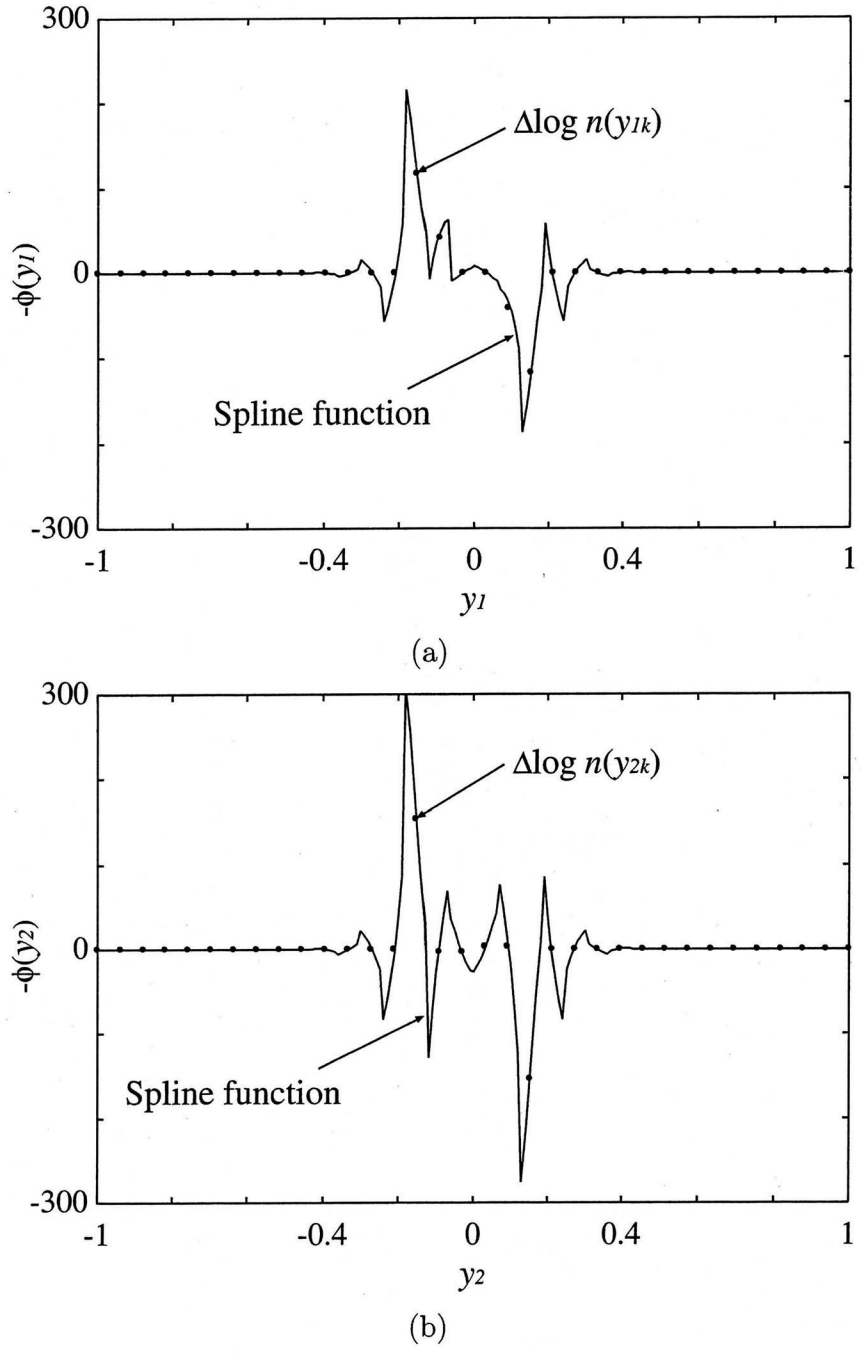
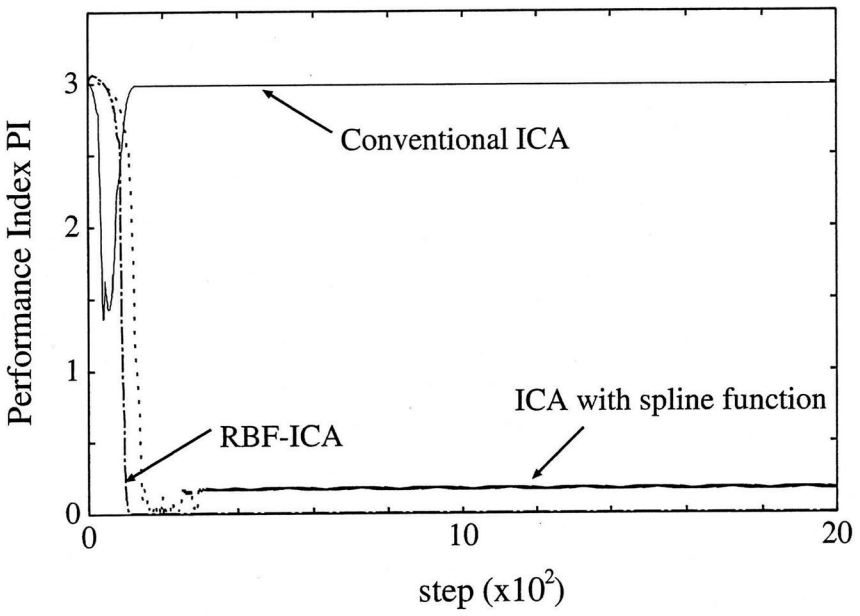
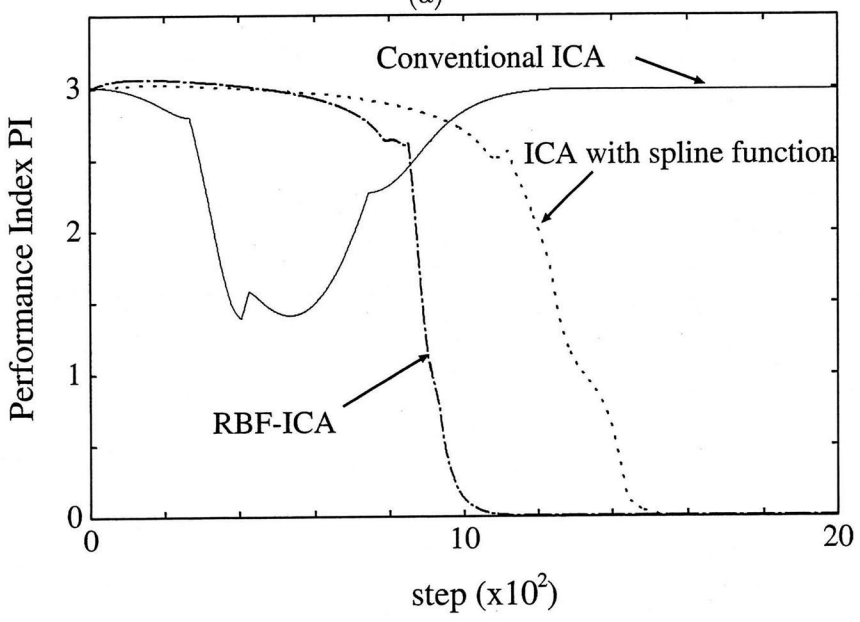


図 2.9: The nonlinear functions approximated by the spline functions. (a) $\phi(y_1)$. (b) $\phi(y_2)$.

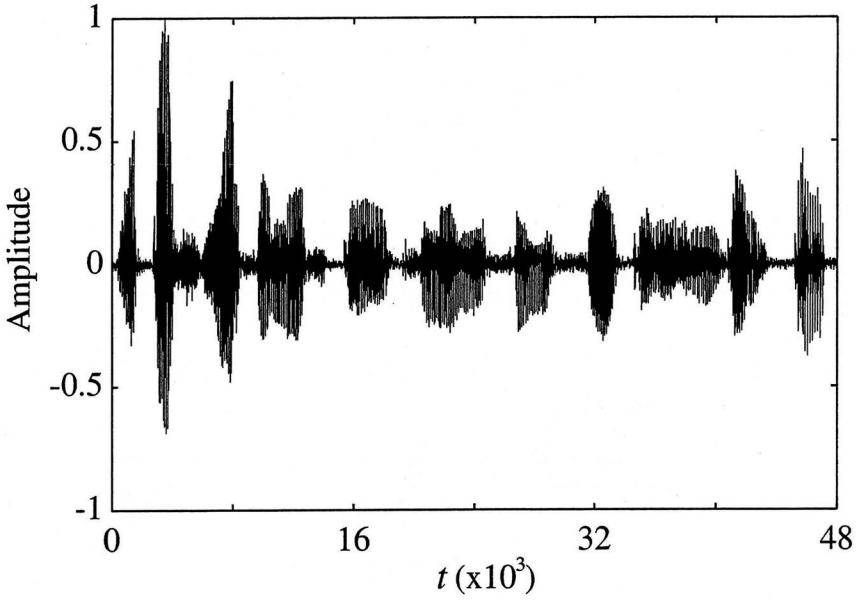


(a)

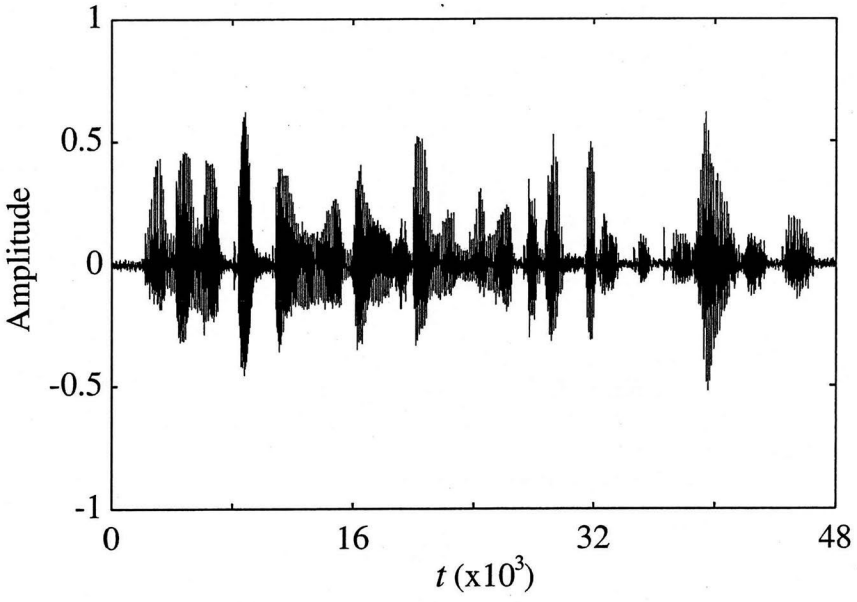


(b)

図 2.10: The results of PIs obtained by using the conventional ICA, ICA with spline function, and the proposed RBF-ICA. (a) $\eta = 0.1$. (b) $\eta = 0.01$.

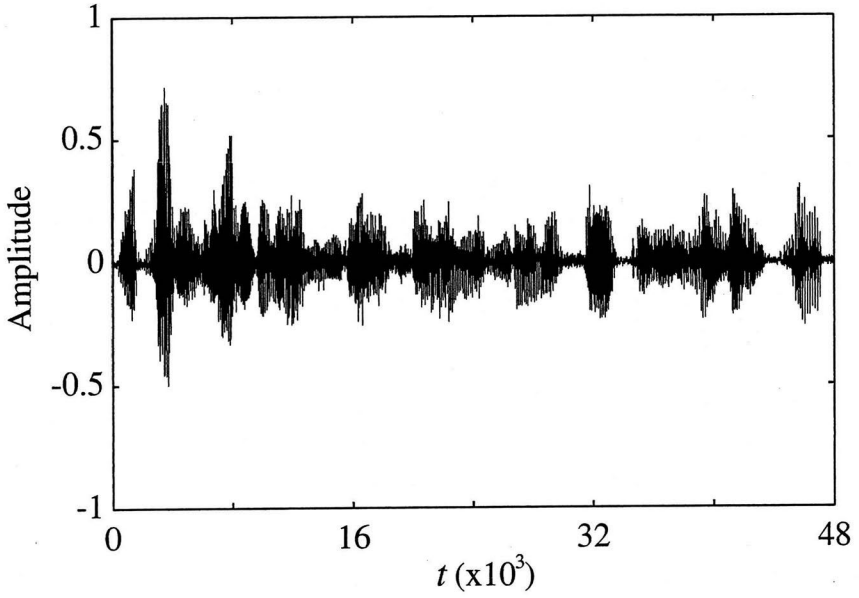


(a)

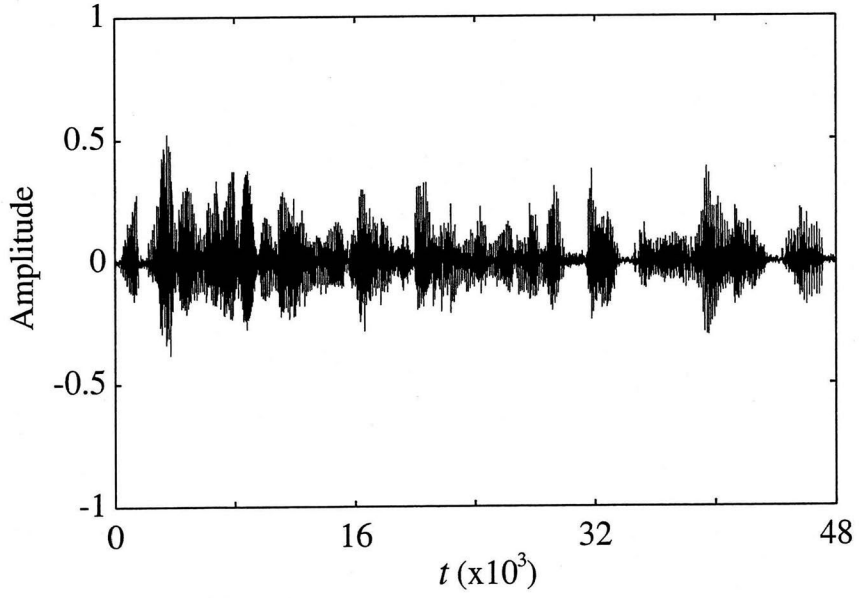


(b)

図 2.11: The source speech signals. (a) $s_1(t)$. (b) $s_2(t)$.

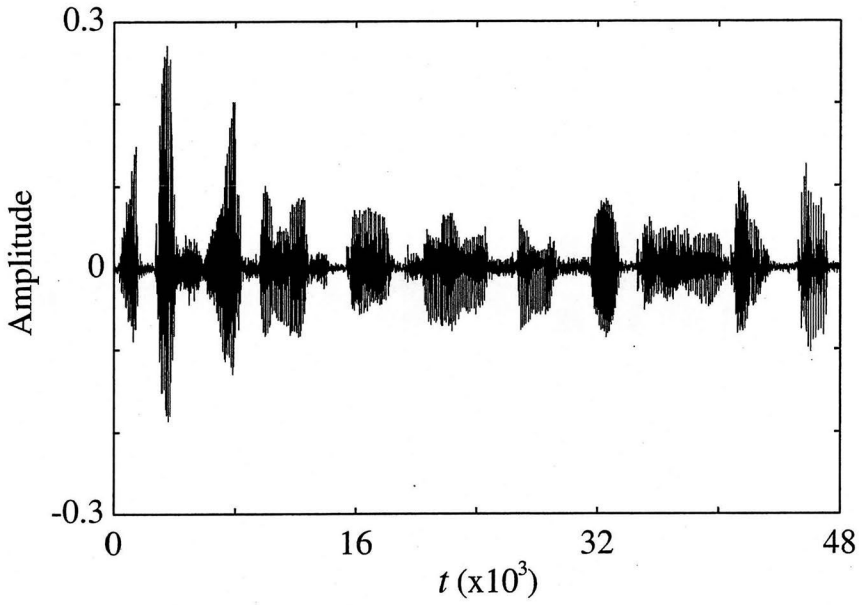


(a)

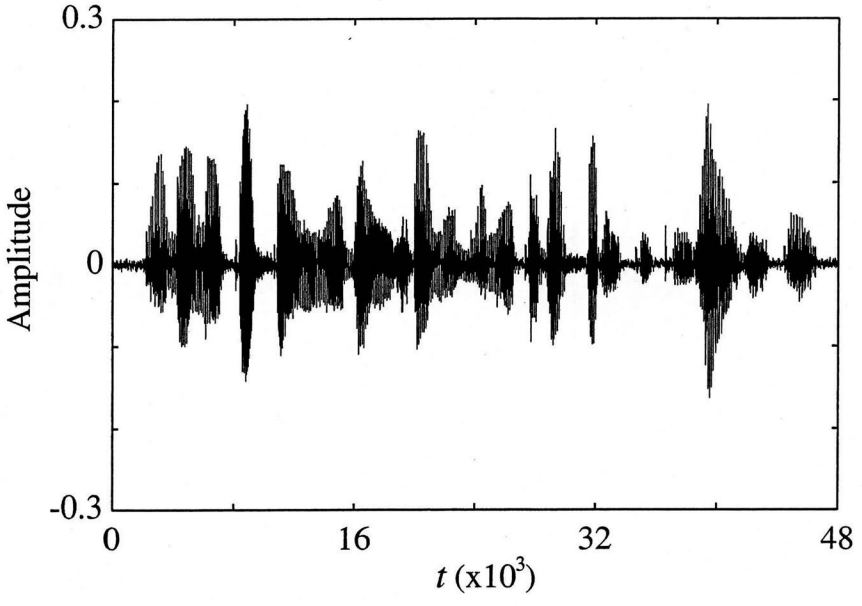


(b)

図 2.12: The mixed speech signals. (a) $x_1(t)$. (b) $x_2(t)$.

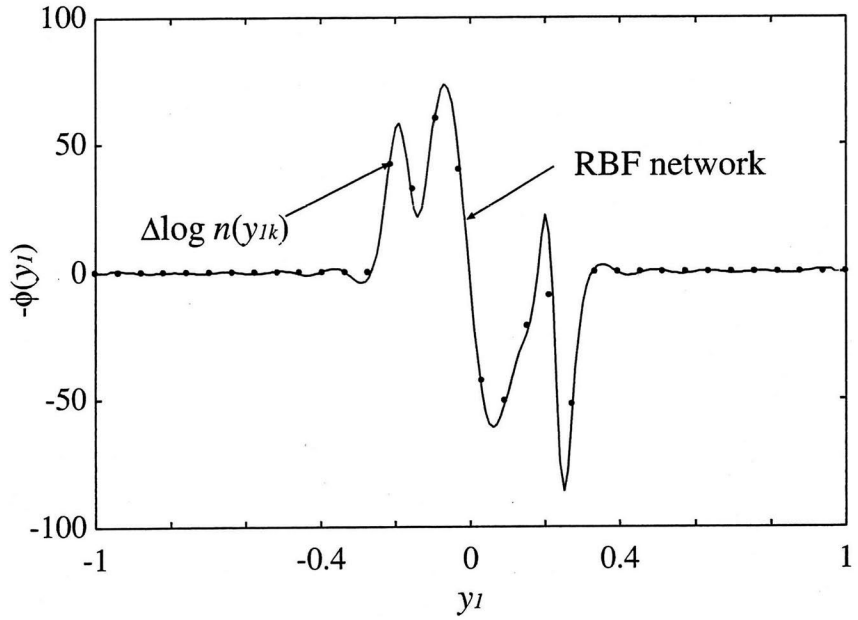


(a)

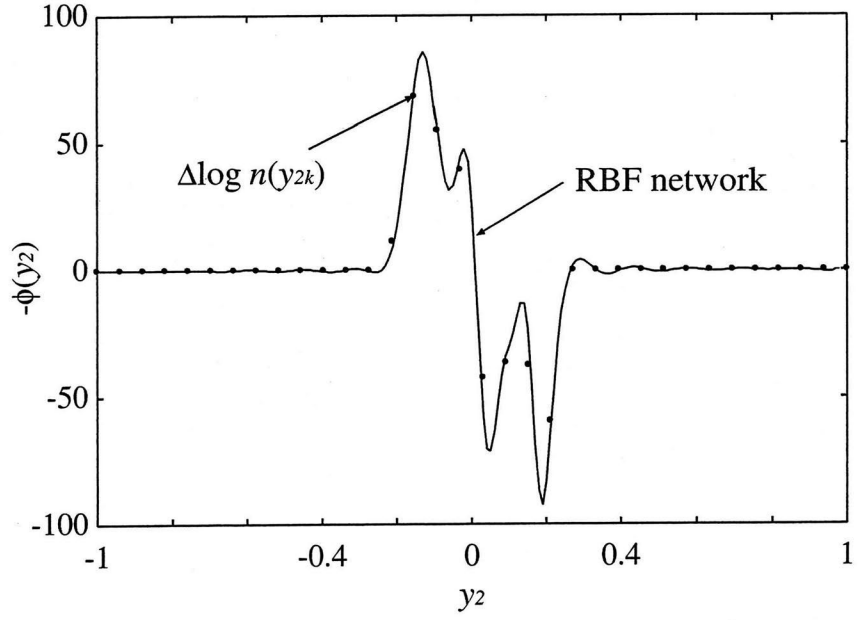


(b)

図 2.13: The separated speech signals obtained by using the hybrid method followed by the RBF-ICA. (a) $y_1(t)$. (b) $y_2(t)$.



(a)



(b)

図 2.14: The nonlinear functions approximated by the RBF networks. (a) $\phi(y_1)$. (b) $\phi(y_2)$.

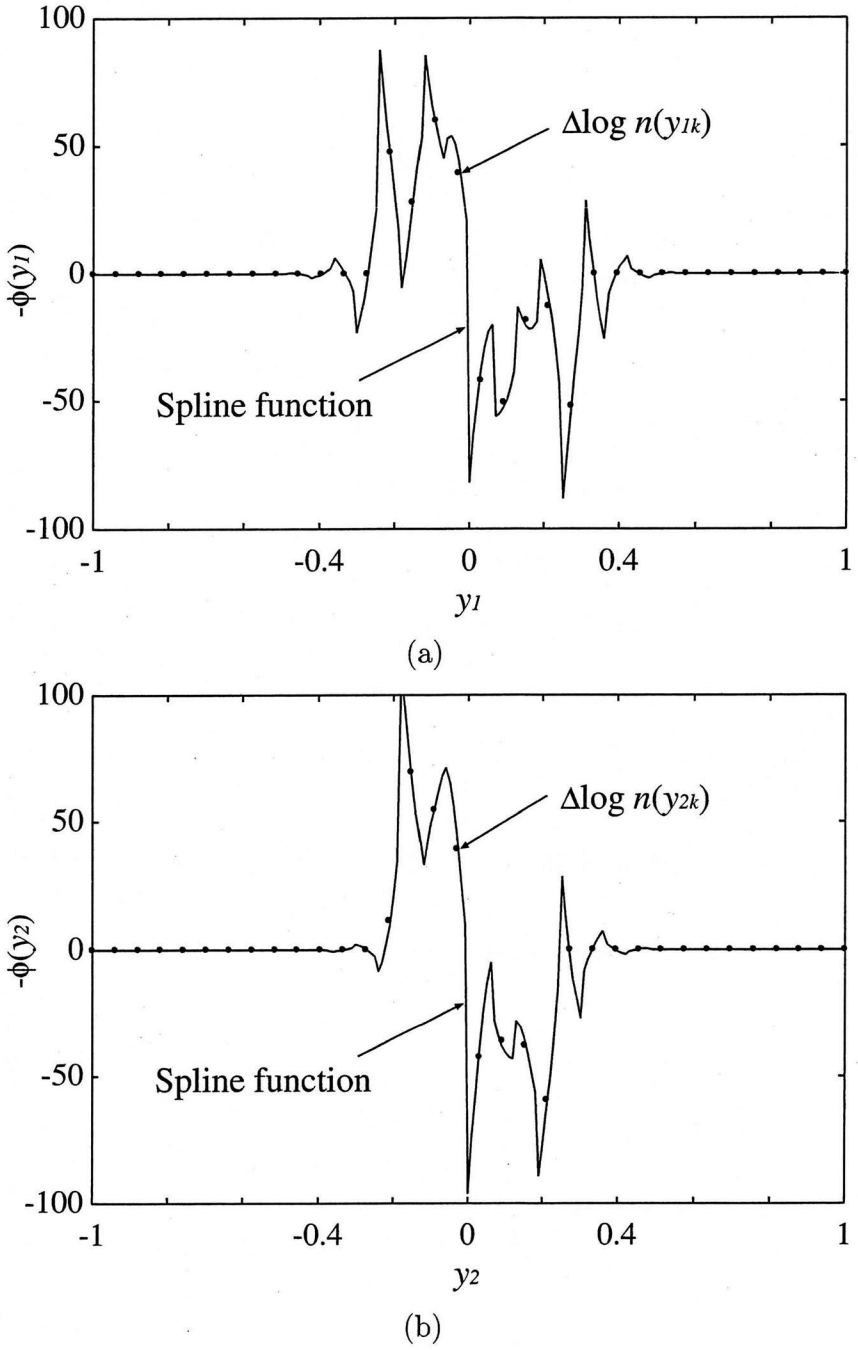


図 2.15: The nonlinear functions approximated by the spline functions. (a) $\phi(y_1)$. (b) $\phi(y_2)$.

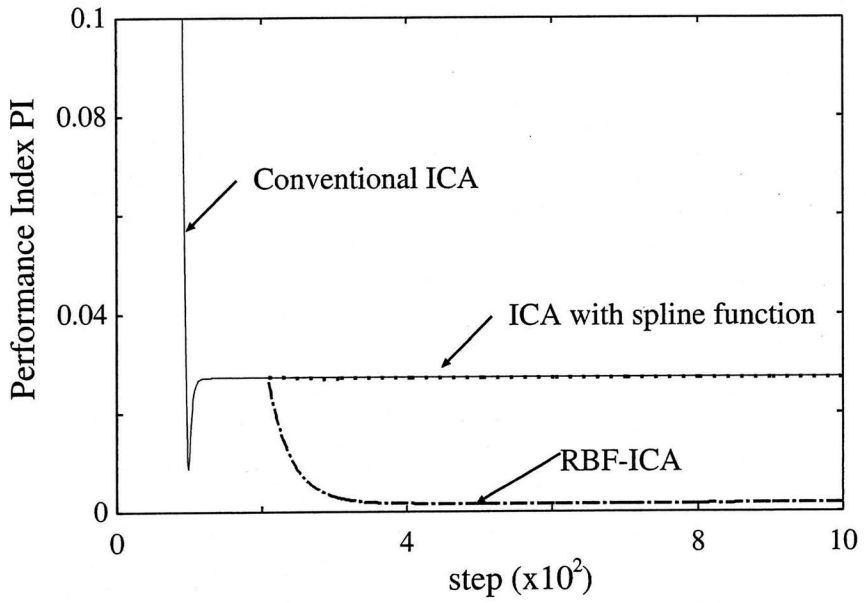


図 2.16: The results of PIs by using only the conventional ICA, and by using the hybrid methods followed by the RBF-ICA and by the ICA with spline function.

第3章　周波数領域ICAに基づいた音声信号分離と音源定位

3.1　緒言

本章では，伝搬遅延時間を含むブラインド信号分離 (BSS) 問題の定式化と基本的な周波数領域での独立成分分析 (ICA)，すなわち，周波数領域ICA(Frequency-Domain ICA) の考え方を説明した後，提案手法である周波数領域ICAに基づいた音声信号分離と音源定位の手法について述べる．

前章で述べたように，ICAの学習アルゴリズムは，一般的に瞬時混合 (Instantaneous Mixture) のBSS問題に対して導出されている．これを瞬時混合ICAと呼ぶことにする．しかしながら，音声のような伝搬遅延時間を無視できない信号に対しては，瞬時混合ICAをそのまま適用することはできない．この問題を解くための方法の1つに周波数領域ICAがある [40]．周波数領域ICAは，時間領域における観測信号を短時間フーリエ変換を用いて周波数領域に変換することにより，この問題を周波数帯域毎の瞬時混合問題として扱う手法である．そのため，従来の瞬時混合ICAがそのまま適用できることとなり，時間領域で直接的に定式化するよりも [20][37]-[39]，この問題を単純化することができる．

そこで，本章では，信号の混合過程において時間遅れを考慮する必要があるBSS問題に対して，周波数領域ICAに基づいた信号分離手法を提案する．まず，周波数領域ICAを用いて正確に伝搬遅延時間を推定する手法を提案する．次に，その得られた伝搬遅延時間に基づいた信号分離手法を提案する．さらに，推定した混合過程

の伝搬遅延時間と減衰係数を利用した音源定位の手法についても提案する．提案手法の有効性は，計算機シミュレーションにより確認された．

3.2 伝搬遅延時間を含むブラインド信号分離

本節では，信号の混合過程で伝搬遅延時間を含むブラインド信号分離 (BSS) 問題について定式化を行う．

今，N 個の原信号が各々伝搬遅延時間を含んで線形的に混合され，N 個のセンサーで観測される場合を考える．このとき，k 番目の観測信号 $x_k(t)$ は以下の式で表される．

$$x_k(t) = \sum_{l=1}^{N} a_{kl} s_l(t - d_{kl}) \quad (k = 1, \ldots, N), \tag{3.1}$$

ここで，s_l は l 番目の原信号，t は $x_k(t)$ を観測する時間である．d_{kl} と a_{kl} は伝搬遅延時間と減衰係数をそれぞれ表す．これらの混合係数の値は，l 番目の信号源から k 番目のセンサーまでの距離で決まる．図 3.1 は，$N = 2$ の場合における伝搬遅延時間を含む混合過程を示している．

次に，この問題を周波数領域で考えるため，式 (3.1) をフーリエ変換すると次式のようになる [42].

$$x_k(f) = \sum_{l=1}^{N} A_{kl}(f) s_l(f), \tag{3.2}$$

$$A_{kl}(f) = a_{kl} \exp(-j2\pi f d_{kl}), \tag{3.3}$$

ここで，f は周波数，j は虚数単位を表す．式 (3.2) を行列形式で表現すると，次式のように簡潔にまとめることができる．

$$\boldsymbol{x}(f) = \boldsymbol{A}(f)\boldsymbol{s}(f), \tag{3.4}$$

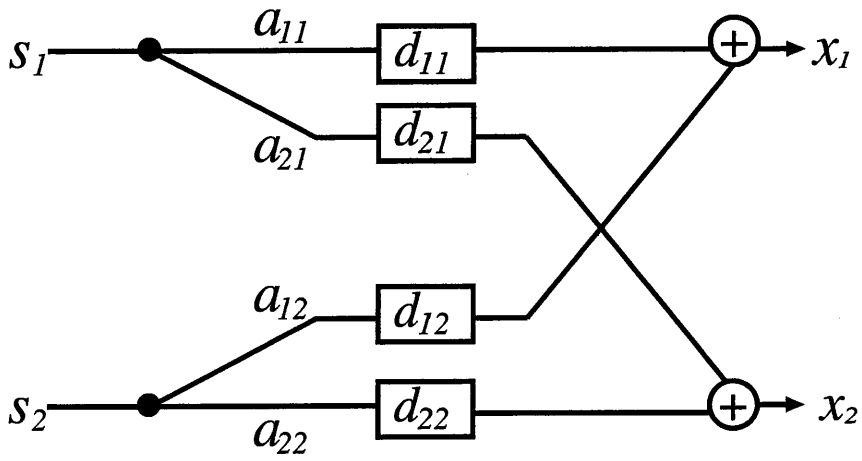


図 3.1: The BSS problem of two signals and two sensors with the propagation time delays included in the mixing process.

ここで，$\boldsymbol{A}(f)\Big(=\big(A_{kl}(f)\big)\Big)$ は，周波数領域における N 行 N 列の混合行列である．$\boldsymbol{x}(f)$ と $\boldsymbol{s}(f)$ は，

$$\boldsymbol{x}(t) = \big(x_1(t),\ldots,x_N(t)\big)^T, \tag{3.5}$$

$$\boldsymbol{s}(t) = \big(s_1(t),\ldots,s_N(t)\big)^T, \tag{3.6}$$

をそれぞれフーリエ変換したベクトルである．ここで，T は転置を表す．

今までの議論では，時間 t を $-\infty$ から ∞ の範囲で考えてフーリエ変換を行った．しかしながら，実際の処理では $-\infty$ から ∞ までの時間は扱えないので，音声信号処理で非常によく用いられるフレーム処理を使う [54][68]．フレーム処理では，次式で表されるハミング窓のような窓関数を用いる．

$$w(t) = \begin{cases} 0.54 - 0.46\cos(2\pi t/L_F) & 0 \le t \le L_F \\ 0 & \text{otherwise} \end{cases}, \tag{3.7}$$

ここで，L_F はフレームの長さを表す．この窓関数を使って，観測信号を次式のよう

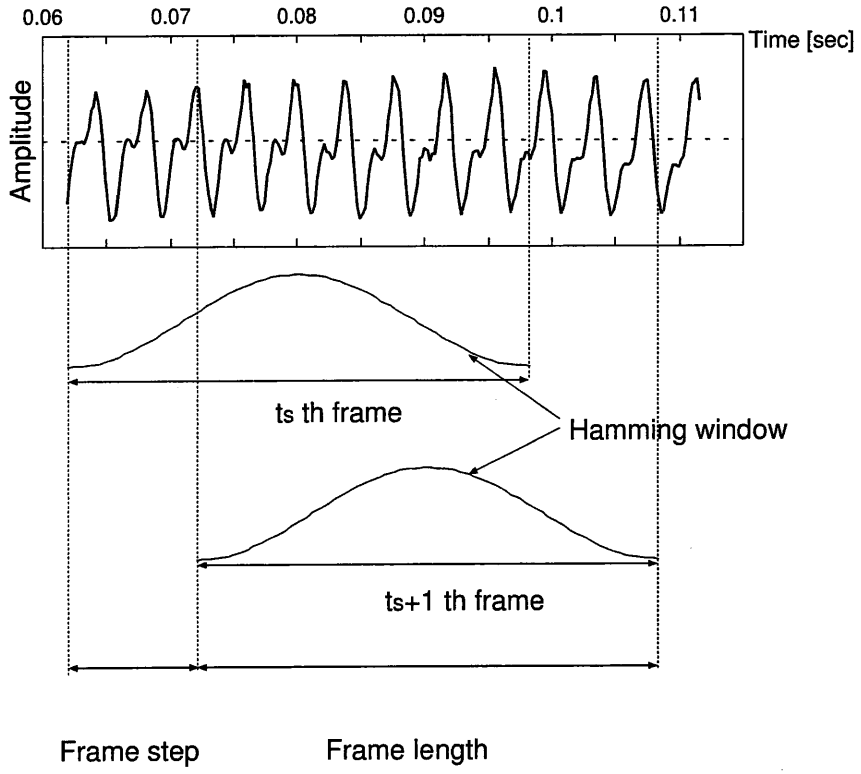


図 3.2: Speech signals divided into frames by Hamming windows.

に有限長のフレーム毎に切り出す.

$$x_i(t, t_s) = w(t - t_s)x_i(t), \tag{3.8}$$

ここで，t_s はフレーム番号を表す．フレーム処理の様子を図 3.2 に示す．この $x_i(t, t_s)$ を短時間フーリエ変換することにより，式 (3.4) を近似する [40].

$$\boldsymbol{x}(f, t_s) = \boldsymbol{A}(f)\boldsymbol{s}(f, t_s), \tag{3.9}$$

ここで，$\boldsymbol{x}(f, t_s)$ と $\boldsymbol{s}(f, t_s)$ は，それぞれ，フレーム番号 t_s で切り出した $\boldsymbol{x}(t, t_s)$ と $\boldsymbol{s}(t, t_s)$ の短時間フーリエ変換 (Short-Time Fourier Transform, STFT) である．

周波数領域 ICA は，$\boldsymbol{x}(f, t_s)$ を時間 t_s で表現される時系列信号と考えて，周波数 f 毎に従来の瞬時混合 ICA を式 (3.9) に適用する手法である．瞬時混合 ICA と周波

数領域ICAの相違点は，瞬時混合ICAでは混合行列が唯一つだけ存在するのに対し，周波数領域ICAでは周波数毎に混合行列が存在することである．

次節では，式(3.9)で表される観測信号の周波数成分$\boldsymbol{x}(f,t_s)$に対して，ICAを適用する方法について説明する．

3.3 周波数領域ICA

本節では，周波数領域ICAの基本的な考え方について説明する．

周波数領域ICAでは，観測信号のフレーム番号t_sにおける周波数成分$\boldsymbol{x}(f,t_s)$にN行N列の複素行列$\boldsymbol{W}(f)$を作用させて得られる分離信号$\boldsymbol{y}(f,t_s)$を，原信号のフレーム番号t_sにおける周波数成分$\boldsymbol{s}(f,t_s)$の推定値とする．

$$\boldsymbol{y}(f,t_s) = \boldsymbol{W}(f)\boldsymbol{x}(f,t_s). \tag{3.10}$$

fを固定すれば，式(3.10)は瞬時混合BSS問題の定式化と同じである．そのため，これを解くためには前章で述べた瞬時混合ICAが利用できるが，信号が複素数であることが異なる．そこで，前章で述べた式(2.24)で与えられる分離行列の更新式をそのまま複素形式に書き直してみる．

$$\Delta\boldsymbol{W}(f) = \eta\left(\boldsymbol{I} - <\boldsymbol{\Phi}(\boldsymbol{y}(f,t_s))\boldsymbol{y}(f,t_s)^H>\right)\boldsymbol{W}, \tag{3.11}$$

ここで，$\boldsymbol{\Phi}$は，式(2.24)の$\boldsymbol{\phi}$が複素ベクトルであることを強調するため特に大文字で表した．また，$<\cdot>$はt_sに関する平均操作となり，Hは複素転置を示す．

もし，分離行列の学習が収束したとすると，

$$\boldsymbol{I} - \left\langle\boldsymbol{\Phi}(\boldsymbol{y})\boldsymbol{y}^T\right\rangle = 0, \tag{3.12}$$

となる．これを行列の成分で表すと次式のようになる．

$$\left\langle\boldsymbol{\Phi}(\boldsymbol{y}_i)\boldsymbol{y}_j^*\right\rangle = 0 \quad (i \neq j), \tag{3.13}$$

$$\left\langle\boldsymbol{\Phi}(\boldsymbol{y}_i)\boldsymbol{y}_i^*\right\rangle = 1, \tag{3.14}$$

ここで，$*$は複素共役を表す．式(3.13)は$\boldsymbol{y}_i$と$\boldsymbol{y}_j$が統計的に独立となるための条件として必要である．一方，式(3.14)が成り立つためには$\boldsymbol{y}_i$の実部と虚部同士が統計的に独立となることが要求されるが，他の信号と統計的に独立となるための条件としては必要ない．そのため，式(3.13)の条件だけを満たすように，式(3.11)を書き直すと次式のようになる．

$$\begin{aligned}\Delta \boldsymbol{W}(f) &= \eta\Big[\mathrm{diag}\left(<\boldsymbol{\Phi}(\boldsymbol{y}(f,t_s))\boldsymbol{y}(f,t_s)^H>\right) \\ &\quad - <\boldsymbol{\Phi}(\boldsymbol{y}(f,t_s))\boldsymbol{y}(f,t_s)^H>\Big]\boldsymbol{W}(f), \end{aligned} \tag{3.15}$$

ここで，$\mathrm{diag}(\boldsymbol{X})$は，行列$\boldsymbol{X}$の対角成分をもつ対角行列である．$\boldsymbol{\Phi}(\cdot)$は，次式で与えられる複素ベクトルである[40][41].

$$\boldsymbol{\Phi}(\boldsymbol{y}(f,t_s)) = \boldsymbol{\phi}(\mathrm{Re}(\boldsymbol{y}(f,t_s))) + j\ \boldsymbol{\phi}(\mathrm{Im}(\boldsymbol{y}(f,t_s))), \tag{3.16}$$

$$\boldsymbol{\phi}(\boldsymbol{y}) = \big(\phi(y_1),\ldots,\phi(y_N)\big), \tag{3.17}$$

ここで，$\mathrm{Re}(\cdot)$と$\mathrm{Im}(\cdot)$は，それぞれ$\boldsymbol{y}(f,t_s)$の実部と虚部を示しており，$\phi(\cdot)$はシグモイド型関数のような任意の非線形関数である．$\boldsymbol{\Phi}(\cdot)$については，式(3.16)の形式ではなく複素平面の極座標で表現する方法もある[49]．最後に，分離信号$\boldsymbol{y}(t)$は，逆短時間フーリエ変換(ISTFT)を$\boldsymbol{y}(f,t_s)$に適用することによって得ることができる．

$$\boldsymbol{y}(t) = \sum_{t_s}\mathrm{ISTFT}(\boldsymbol{y}(f,t_s)). \tag{3.18}$$

3.4 提案する信号分離手法と音源定位

本節では，提案手法である伝搬遅延時間と減衰係数の推定方法と，これらの推定値を用いた信号分離手法と音源定位について説明する．

3.4.1 伝搬遅延時間と減衰係数の推定

最初に，式(3.10)の分離行列$\boldsymbol{W}(f)$から混合過程における伝搬遅延時間と減衰係数が推定できることを示す．正確には，これらの推定値は，原信号とそれぞれのセンサー間の相対的な位置関係で決まる相対伝搬遅延時間と減衰係数比であることが示される．

サンプリング周波数F_sで観測された観測信号$\boldsymbol{x}(t)$にM点STFTを適用すると，式(3.3)の混合係数は次式のように離散形式で表される．STFTでは，高速フーリエ変換(Fast Fourier Transform, FFT)を用いる．

$$A_{kl}(f_n) = a_{kl}\exp(-j2\pi f_n d_{kl}), \tag{3.19}$$

ここで，

$$f_n \;=\; (F_s/M)n \quad (n=0,\ldots,M-1). \tag{3.20}$$

混合行列$\boldsymbol{A}(f_n)$のための分離行列$\boldsymbol{W}(f_n)$は式(3.15)によって学習されるが，$\boldsymbol{A}(f_n)$が複素行列のため一意に決まらない．そのため，学習後の分離行列$\boldsymbol{W}(f_n)$の収束値は，その初期値に強く依存してしまう[7]．そこで，まず，式(3.19)を次式のように分解してみる．

$$\boldsymbol{A}(f_n) \;=\; \boldsymbol{A}'(f_n)\mathrm{diag}(\boldsymbol{A}(f_n)), \tag{3.21}$$

$$A'_{kl}(f_n) \;=\; \begin{cases} 1 & k=l \\ a'_{kl}\exp(-j2\pi f_n d'_{kl}) & k\neq l \end{cases}, \tag{3.22}$$

ここで，

$$a'_{kl} \;=\; a_{kl}/a_{ll}, \tag{3.23}$$

$$d'_{kl} \;=\; d_{kl}-d_{ll}, \tag{3.24}$$

ここで，a'_{kl} は減衰係数比，d'_{kl} は相対伝搬遅延時間である．

また，$\boldsymbol{A}(f_n)$ の逆行列 $\boldsymbol{A}'^{-1}(f_n)\Big(= \big(A'^{-1}_{kl}(f_n)\big)\Big)$ の各々の要素を級数展開し，a'_{kl} の1次の項までで近似すると次式が得られる．

$$A'^{-1}_{kl}(f_n) \simeq \begin{cases} 1 & k = l \\ -a'_{kl}\exp(-j2\pi f_n d'_{kl}) & k \neq l \end{cases}, \tag{3.25}$$

ここで，$\simeq$ は近似的に等しいことを意味する．式 (3.25) から，$\boldsymbol{A}'^{-1}(f_n)$ は，離散周波数 f_n の変化によって単位行列の周りを振動することが分かる．ただし，この近似では a'_{kl} の2次以上の項を無視するために，l 番目の原信号が l 番目の観測信号に最も寄与すると仮定した．すなわち，$d_{kl} - d_{ll} > 0 \ \ (k \neq l)$ と $a_{kl}/a_{ll} < 1 \ \ (k \neq l)$ であるとした．計算機シミュレーションでは，それぞれの周波数で単位行列を初期値として学習した $\boldsymbol{W}(f_n)$ が，学習後に $\boldsymbol{A}'^{-1}(f_n)$ に収束する傾向があることが確認された．また，$\boldsymbol{W}(f_n)$ の初期値が単位行列で $d_{kl} - d_{ll} > 0 \ \ (k \neq l)$ と $a_{kl}/a_{ll} < 1 \ \ (k \neq l)$ が仮定されるとき，周波数領域 ICA における式 (2.11) で与えられる ICA の不定性による問題 [40][52] はあまり問題とならなかった．

学習後の $\boldsymbol{W}(f_n)$ は $\boldsymbol{A}'^{-1}(f_n)$ の近似となっているが，$A'_{kl}(f_n) \ \ (k \neq l)$ の ISTFT が次式のように与えられることに注目する．

$$\begin{aligned} &\mathrm{ISTFT}(A'_{kl}(f_n)) \\ &= \frac{1}{M}\sum_{m=0}^{M-1} a'_{kl}\exp(-j\frac{2\pi m}{M}(F_s d'_{kl} - \frac{f_n M}{F_s})) \\ &= \frac{1}{M}\sum_{m=0}^{M-1} a'_{kl}\cos(\frac{2\pi m}{M}(F_s d'_{kl} - \frac{f_n M}{F_s})) \\ &\quad -j\frac{1}{M}\sum_{m=0}^{M-1} a'_{kl}\sin(\frac{2\pi m}{M}(F_s d'_{kl} - \frac{f_n M}{F_s})). \end{aligned} \tag{3.26}$$

式 (3.26) は $\mathrm{ISTFT}(A'_{kl}(f_n))$ の実部が $f_n = F_s^2 d'_{kl}/M$ のとき，すなわち，$n = F_s d'_{kl}$ のとき最大値をとることを示す．そのため，もし，$n = p_{kl}$ のとき $\mathrm{ISTFT}(A'_{kl}(f_n))$

の実部が最大値をとるならば，相対伝搬遅延時間 $\hat{d}'_{kl}$ は次式のように推定される．

$$\hat{d}'_{kl} = p_{kl}/F_s, \tag{3.27}$$

ここで，$\hat{d}'_{kl}$ の時間解像度は $1/F_s$ である．$\hat{d}'_{kl}$ の最小値と最大値は，それぞれ，0 と $(M-1)/F_s$ である．このことは，相対伝搬遅延時間が周波数領域 ICA から得られた分離行列の逆行列 $\boldsymbol{W}'^{-1}(f_n)\Big(= \big(W'^{-1}_{kl}(f_n)\big)\Big)$ によって推定できることを示している．

減衰係数比は，次のように簡単に推定される．すなわち，$\boldsymbol{W}^{-1}(f_n)$ が $\boldsymbol{A}'(f_n)$ の近似であることを考慮すると，式 (3.22) より，推定される減衰係数比 $\hat{a}'_{kl}$ は，次式のように与えられる．

$$\hat{a}'_{kl} = < |W^{-1}_{kl}(f_n)|/|W^{-1}_{ll}(f_n)| >, \tag{3.28}$$

ここで，$<\cdot>$ は f_n に関する平均操作であり，$|\cdot|$ は複素数の絶対値である．$|W^{-1}_{ll}(f_n)|$ によって割る理由は，f_n に関する $\boldsymbol{W}(f_n)$ のゆらぎに関して，正規化を行うためである．

3.4.2 推定した混合係数による信号分離手法

先程得られた相対伝搬遅延時間と減衰係数比を使った信号分離手法について述べる．

推定混合行列 $\hat{\boldsymbol{A}}'(f_n)$ は，次式のように $\hat{d}'_{kl}$ と $\hat{a}'_{kl}$ によって作られる．

$$\hat{A}'_{kl}(f_n) = \hat{a}'_{kl} \exp(-j2\pi f_n \hat{d}'_{kl}). \tag{3.29}$$

再構成された分離行列 $\hat{\boldsymbol{W}}(f_n)$ は，$\hat{\boldsymbol{A}}'(f_n)$ の逆行列を計算することによって得られ

る．そのとき，分離された周波数成分は次式のように与えられる．

$$\begin{aligned}\hat{\boldsymbol{y}}(f_n) &= \hat{\boldsymbol{W}}(f_n)\boldsymbol{x}(f_n) \\ &= \hat{\boldsymbol{W}}(f_n)\boldsymbol{A}(f_n)\boldsymbol{s}(f_n) \\ &= \hat{\boldsymbol{W}}(f_n)\boldsymbol{A}'(f_n)\mathrm{diag}(\boldsymbol{A}(f_n))\boldsymbol{s}(f_n) \\ &\simeq \mathrm{diag}(\boldsymbol{A}(f_n))\boldsymbol{s}(f_n). \end{aligned} \tag{3.30}$$

さらに，$\hat{\boldsymbol{y}}(f_n)$ に ISTFT を適用することにより，分離信号 $\hat{\boldsymbol{y}}(t)$ が原信号の推定値として次式のように得られる．

$$\hat{y}_l(t) \simeq a_{ll}s_l(t - d_{ll}). \tag{3.31}$$

上記の $\hat{\boldsymbol{y}}(t)$ により信号分離が行われるが，ここでは，さらに良い分離性能を得るために，その $\hat{\boldsymbol{y}}(t)$ と推定した混合係数を観測信号に作用させる信号分離手法を提案する．まず，$a_{kl}s_l(t - d_{kl})$ が，式 (3.31) の $\hat{y}_l(t)$ と式 (3.27) の $\hat{d}'_{kl}$ と式 (3.28) の $\hat{a}'_{kl}$ を使うことにより，次式のように近似されることに注意する．

$$\begin{aligned} a_{kl}s_l(t - d_{kl}) &= a'_{kl}a_{ll}s_l(t - d_{ll} - d'_{kl}) \\ &\simeq a'_{kl}\hat{y}_l(t - d'_{kl}) \\ &\simeq \hat{a}'_{kl}\hat{y}_l(t - \hat{d}'_{kl}). \end{aligned} \tag{3.32}$$

すると，式 (3.1) と式 (3.32) を組み合わせることにより，最終的な分離信号 $\boldsymbol{z}(t)$ が得られる．

$$\begin{aligned} z_k(t) &= x_k(t) - \sum_{l=1, l\neq k}^{N} \hat{a}'_{kl}\hat{y}_l(t - \hat{d}'_{kl}) \\ &\simeq a_{kk}s_k(t - d_{kk}). \end{aligned} \tag{3.33}$$

式 (3.33) から，この修正された手法は，k 番目のセンサーで観測される k 番目の原信号のスケールを正確に特定できることが分かる．

上記の手法は次のようにまとめることができる．

Step 1. 周波数領域ICAから分離行列 $\boldsymbol{W}(f_n)$ を得る.

Step 2. $\boldsymbol{W}(f_n)$ から相対伝搬遅延時間 d'_{kl} と減衰係数比 a'_{kl} を推定する.

Step 3. $\hat{d}'_{kl}$ と $\hat{a}'_{kl}$ から混合行列 $\boldsymbol{A}'(f_n)$ を推定する.

Step 4. $\hat{\boldsymbol{A}}'(f_n)$ から分離行列 $\boldsymbol{W}(f_n)$ を再構成する.

Step 5. $\hat{\boldsymbol{W}}(f_n)$ と $\hat{d}'_{kl}$ と $\hat{a}'_{kl}$ から最終的な分離信号 $\boldsymbol{z}(t)$ を得る.

提案手法は，上のStep 1からStep 5までの過程で構成される．提案手法の主な計算コストは，Step 1での周波数領域ICA[40]の学習とStep 2での M 回の逆行列の計算と $N(N-1)$ 回のFFTの計算とStep 4の M 回の逆行列の計算である．そのため，提案手法の計算量は N と M が増えるにつれ増加する．ただし，Step 4においては，式(3.22)の逆行列の計算を直接行わなくても，式(3.22)の特徴を使うことにより計算量を削減することが可能となる．

本章と同じ話題である，伝搬遅延時間を含むBSS問題に対し観測信号の周波数領域の情報を利用して，混合係数の推定や音源分離を行う幾つかのアルゴリズムが提案されている[73]-[76]．その中で代表的なものとして，Degenerate Unmixing Estimation Technique(DUET)アルゴリズム[73][74]がある．DUETアルゴリズムは，原信号がW-disjoint orthogonal条件[74]を満たすとき，2つの観測信号のみを使用して任意の数の原信号を分離することができる．しかしながら，W-disjoint orthogonality条件は，ある時刻のある周波数においては，1つの原信号しか観測されないとする仮定のため，長い連続的な信号同士では満足することは難しい．また，DUETアルゴリズムは混合係数を推定するために2次元ヒストグラムを使うが，2次元ヒストグラムの形状はヒストグラムの区間の選び方に強く依存するので，そこから鋭いピークを得ることは簡単ではない．一方，提案手法は，式(3.26)の実部のピークを1次元上で探すことと式(3.28)の値を計算することによって，混合係数を簡単に推定することができる．

また，位相差の偏角から遅延時間を推定する方法も提案されているが [75]，偏角を求めるときに直線の勾配が利用されるため誤差が生じやすく，提案手法のように時間分解能を明確にすることができない．

3.4.3 推定した混合係数による音源定位

推定された相対伝搬遅延時間と減衰係数比を利用した音源定位の手法について提案する．音源定位の代表的な手法として，Multiple Signal Classification(MUSIC) 法がある [69][70]．しかしながら，この手法は事前に任意の方向の位置ベクトルを必要とする [71]．この位置ベクトルを得るためには，任意の場所における仮想的な原信号とセンサー間の伝達関数を多数求めなければならない．一方，提案手法は ICA に基づいた手法であるため，観測信号の情報のみを利用した音源定位が可能である．ここでは，簡単のため $N = 2$ の場合，すなわち，2 信号 2 センサーの場合を考えるが，N は 2 以上の任意の数に容易に拡張できる．

図 3.3 で示されている通り，センサー 1 とセンサー 2 を，それぞれ，座標 $(S, 0)$ と座標 $(-S, 0)$ に配置させた直交座標系を考える．

L_{k1} $(k = 1, 2)$ を音源 1 からセンサー k までの距離とすれば，以下のように表すことができる．

$$L_{11} = \sqrt{(X_1 - S)^2 + Y_1^2}, \tag{3.34}$$

$$L_{21} = \sqrt{(X_1 + S)^2 + Y_1^2}. \tag{3.35}$$

d'_{21} は，音源 1 から 2 つのセンサーまでの伝搬遅延時間の差であるので，L_{21} と L_{11} の差は，次式のように d'_{21} によって表現される．

$$L_{21} - L_{11} = vd'_{21}, \tag{3.36}$$

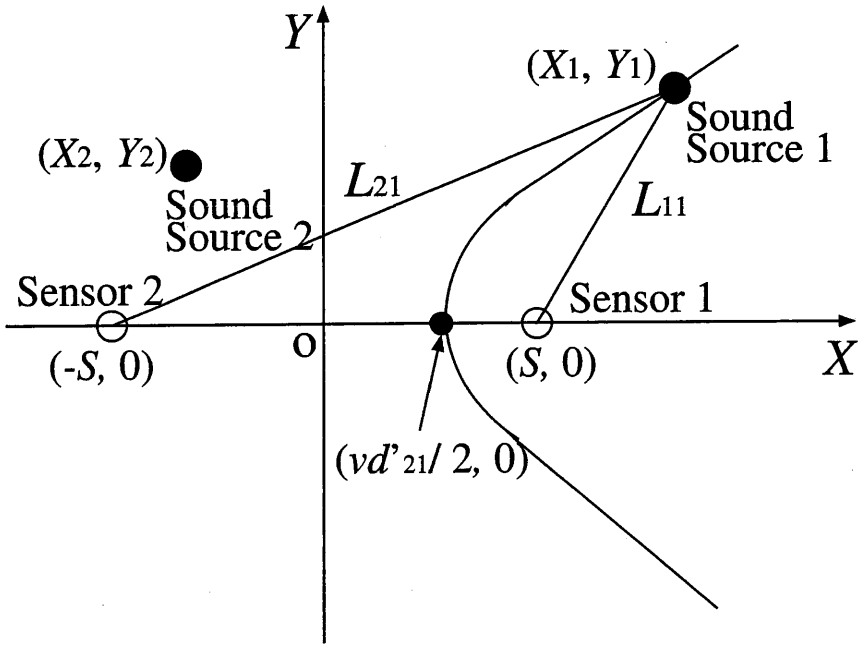


図 3.3: Rectangular coordinates for the sound localization.

ここで，v は原信号の速度である．式 (3.36) は，次式のように双曲線の方程式の標準形に書き直すことができる．

$$\frac{X_1^2}{(vd'_{21}/2)^2} - \frac{Y_1^2}{R^2} = 1, \tag{3.37}$$

$$R^2 = S^2 - (vd'_{21}/2)^2, \tag{3.38}$$

ここで，

$$vd'_{21}/2 < S. \tag{3.39}$$

音源 1 は，図 3.3 で示される双曲線上に位置する．ここで，空気を伝搬する音声の減衰係数は，次のように距離に逆比例することが知られている [72].

$$a_{kl} \propto \frac{1}{L_{kl}}. \tag{3.40}$$

よって，式(3.40)により次式が得られる．

$$\frac{L_{11}}{L_{21}} = a'_{21}. \tag{3.41}$$

式(3.37)と式(3.38)と式(3.41)の連立方程式を解くことにより，X_1 と Y_1 は次式で与えられる．

$$X_1 = \frac{1}{4}\frac{(1+a'_{21})}{(1-a'_{21})}\frac{(vd'_{21})^2}{S}, \tag{3.42}$$

$$Y_1 = \frac{2R}{vd'_{21}}\sqrt{X_1^2 - (\frac{vd'_{21}}{2})^2}. \tag{3.43}$$

X_2 と Y_2 も，X_1 と Y_1 と同様な方法で与えられる．

$$X_2 = -\frac{1}{4}\frac{(1+a'_{12})}{(1-a'_{12})}\frac{(vd'_{12})^2}{S}, \tag{3.44}$$

$$Y_2 = \frac{2R}{vd'_{12}}\sqrt{X_2^2 - (\frac{vd'_{12}}{2})^2}. \tag{3.45}$$

よって，音源1と音源2の定位が可能であることが分かる．

3.5 計算機シミュレーション結果

本節では，伝搬遅延時間を含むBSS問題に対する提案手法の結果を音声信号を用いて述べる．まず，伝搬遅延時間の推定結果と信号分離結果について示し，その後，音源定位の結果を示す．

3.5.1 混合係数の推定と信号分離

3.5.1.1 実験条件

計算機シミュレーションには，前章と同じく，2人の異なる男性話者のサンプリング周波数16kHz，量子化ビット数16ビット，3秒間のモノラル音声信号 $s_1(t)$ と $s_2(t)$

を用いた．観測信号 $\boldsymbol{x}(t)$ は，原信号 $\boldsymbol{s}(t)$ を次式のように計算機上で混合して作る．

$$x_1(t) = 0.6s_1(t) + 0.4s_2(t - 0.005), \tag{3.46}$$

$$x_2(t) = 0.5s_1(t - 0.003) + 0.6s_2(t - 0.001). \tag{3.47}$$

表 3.1 に今回の周波数領域 ICA における実験パラメータを示す．

表 3.1: Parameters for the frequency-domain ICA

Frame step	160 points(10 ms)
Frame length	512 points(32 ms)
Window function	Hamming window
FFT length	2,048 points
Number of iterations for ICA learning	1,000 steps
Learning rate η	0.1
Nonlinear function ϕ	$1/(1+\exp(-y_i))$

3.5.1.2 実験結果

図 3.4 と図 3.5 は，それぞれ，原信号 $\boldsymbol{s}(t)$ と観測信号 $\boldsymbol{x}(t)$ を示す．図 3.6 は，$W_{12}^{-1}(f_n)$ と $W_{21}^{-1}(f_n)$ の ISTFT を示す．この図から，相対伝搬遅延時間に対応するピークが高解像度で明白に確認でき，提案手法による推定が良い精度で行われていることが分かる．図 3.7 は，$W_{11}^{-1}(f_n)$ と $W_{22}^{-1}(f_n)$ の ISTFT を示す．これらの図では，0 の場所にピークが存在し，式 (3.22) の関係を満たしていることが分かる．

相対伝搬遅延時間の推定値は $\hat{d}'_{12} = 4.00 \times 10^{-3}$ と $\hat{d}'_{21} = 3.00 \times 10^{-3}$ であり，一方，真値は $d'_{12} = 4.00 \times 10^{-3}$ と $d'_{21} = 3.00 \times 10^{-3}$ であり，推定結果は正確であった．また，減衰係数比において，その推定値は $\hat{a}'_{12} = 0.658$ と $\hat{a}'_{21} = 0.805$ であり，一方，真値は $a'_{12} = 0.667$ と $a'_{21} = 0.833$ であった．減衰係数比の推定精度は相対伝搬遅延時間のそれよりも良くなかったが，ある程度推定できていることが分かる．

他手法の比較として，観測信号 $x_1(t)$ と $x_2(t)$ の相互相関関数 [77] を利用して，相対伝搬遅延時間の推定を行ってみた．その結果を図 3.8 に示す．この図から，真値である 4.00×10^{-3} と 3.00×10^{-3} の場所付近にピークが存在するが，提案手法ほど高解像度でなく，また，別の場所にもピークが存在しており，それらを真値と区別することは難しい．

図 3.9(a) と 3.10(a) は原信号の一部分を示す．図 3.9(b) と図 3.10(b) は，式 (3.18) で与えられる従来の周波数領域 ICA による分離信号 $\boldsymbol{y}(t)$ を示す [40]．一方，図 3.9(c) と図 3.10(c) は，式 (3.31) で与えられる提案手法による分離信号 $\hat{\boldsymbol{y}}(t)$ を示す．また，図 3.9(d) と図 3.10(d) は，式 (3.33) で与えられる提案手法による分離信号 $\boldsymbol{z}(t)$ を示す．

これらの図から，提案手法は，観測信号からほぼ正確に原信号を分離していることが分かる．聞き取りテストでも，これらの分離信号は原信号と比較して満足できるものであり，$\boldsymbol{z}(t)$ が最も良かった．

さらに，分離性能を定量的に比較するために，各々の方法は，次式で定義される信号対雑音比 (Signal to Noise Ratio, SNR) によって比較された [43]．

$$\mathrm{SNR}_k = 10 \log_{10} \frac{\sum_t (a_{kk} s_k(t - d_{kk}))^2}{\sum_t (\chi_k(t) - a_{kk} s_k(t - d_{kk}))^2}, \tag{3.48}$$

ここで，$\chi_k(t)$ には従来の周波数領域 ICA である式 (3.18) で与えられる $y_k(t)$，もしくは，提案手法である式 (3.31) で与えられる $\hat{y}_k(t)$，または，式 (3.33) で与えられる $z_k(t)$ が代入される．表 3.2 は，従来の周波数領域 ICA と提案手法の SNR_k を示す．提案手法の SNR は，従来の周波数領域 ICA と比較して改善されていることが分かる．

3.5.2 音源定位

3.5.2.1 実験条件

式 (3.46) と式 (3.47) で与えられる観測信号から原信号の音源定位の結果を示す．ここで，$S = 3$，$v = 1{,}000$ とした．

表 3.2: SNR_k (dB) of the conventional and the proposed methods.

Separated signal	SNR_1	SNR_2
y (Conventional method using Eq. (3.18))	6.33	5.27
$\hat{y}$ (Proposed method using Eq. (3.31))	10.2	7.96
z (Proposed method using Eq. (3.33))	12.7	11.0

3.5.2.2 実験結果

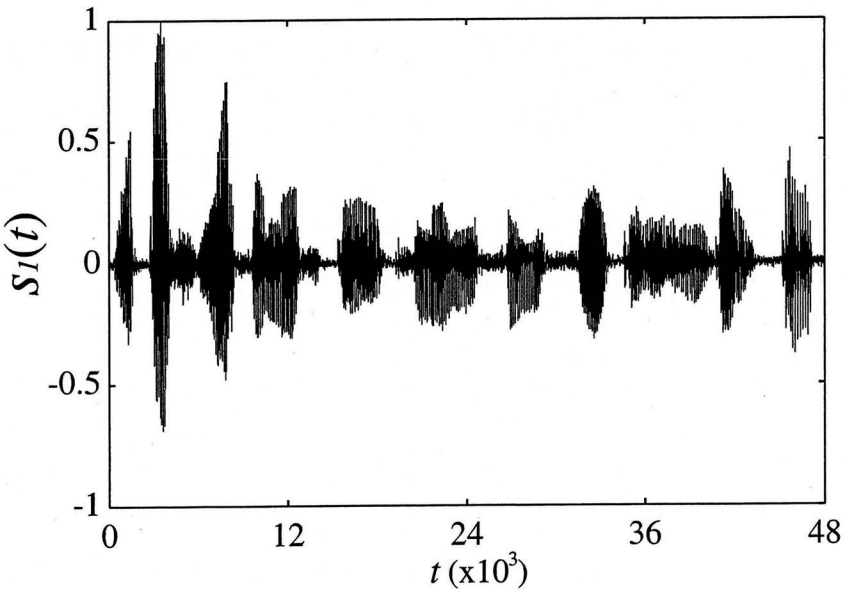
音源位置の推定値は $(\hat{X}_1, \hat{Y}_1) = (6.94, 11.7)$ と $(\hat{X}_2, \hat{Y}_2) = (-6.46, 6.87)$ であり，一方，真値は $(X_1, Y_1) = (8.25, 14.1)$ と $(X_2, Y_2) = (-6.67, 7.11)$ である．これらの結果を図 3.11 に示す．図 3.11 から，音源 2 の定位は，ほぼ成功しているが，音源 1 の定位は，あまり良くない．これは，減衰係数比の推定精度が相対伝搬遅延時間に比べて低いことが原因である．そのため，さらに音源定位の性能を向上させるためには，減衰係数比の推定精度を向上させることが必要である．

3.6 結言

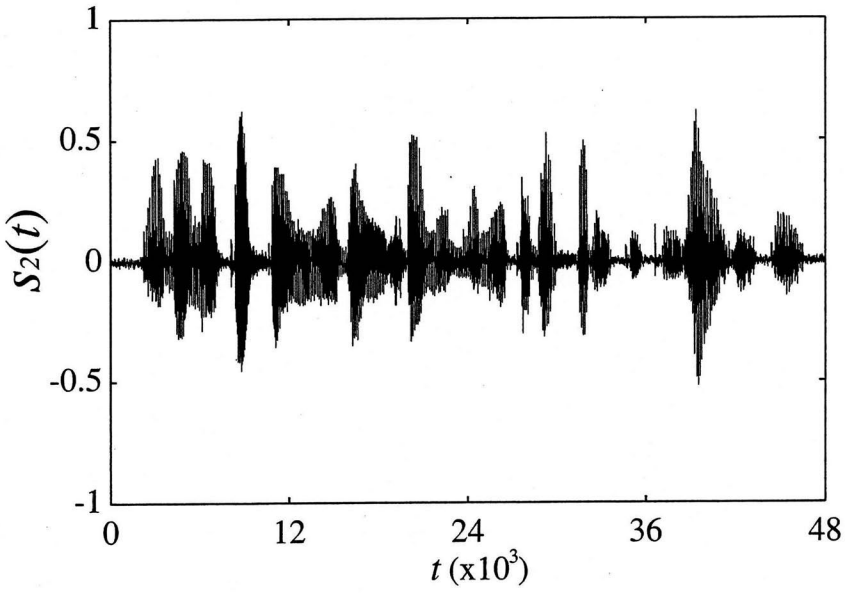
本章では，伝搬遅延時間を含む BSS 問題を解くために，周波数領域 ICA に基づいた効果的な信号分離手法を提案した．そこでは，推定された相対伝搬遅延時間と減衰係数比が使われた．また，これらの推定値を利用した音源定位の手法について提案した．

計算機シミュレーション結果は，推定された相対伝搬遅延時間が，サンプリング周波数で決まる時間解像度内で正確に推定できることを示した．また，提案手法が，信号分離精度において従来の周波数 ICA よりも良い結果を与えることを示した．音源定位に関しては，減衰係数比の推定値の精度により，ほぼ推定できたものと誤差が生じたものがあり，音源によって精度にバラツキがあった．

今後の課題は，減衰係数比の推定精度を向上させることや，センサーの数が原信号の数よりも少ない場合や室内などの残響音下での環境にも適用できるように提案手法を拡張することなどである．

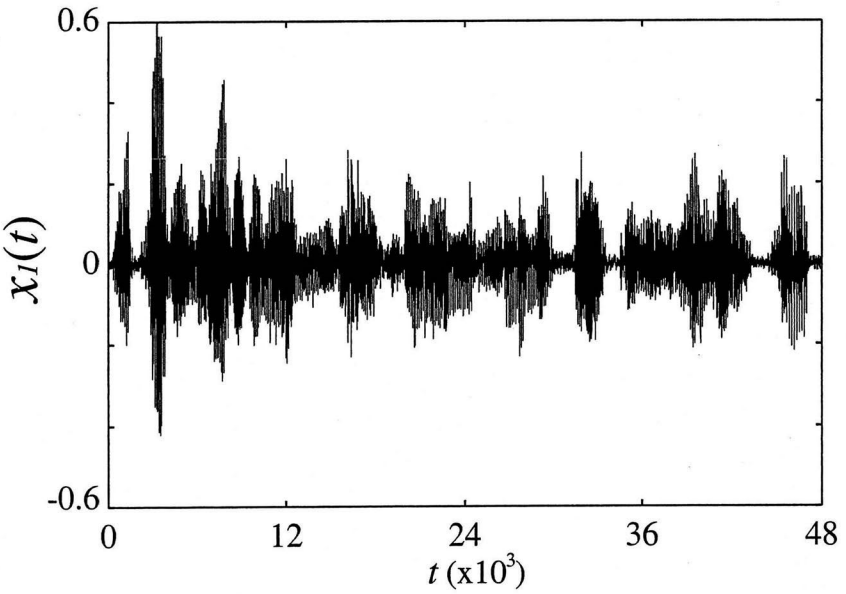


(a)

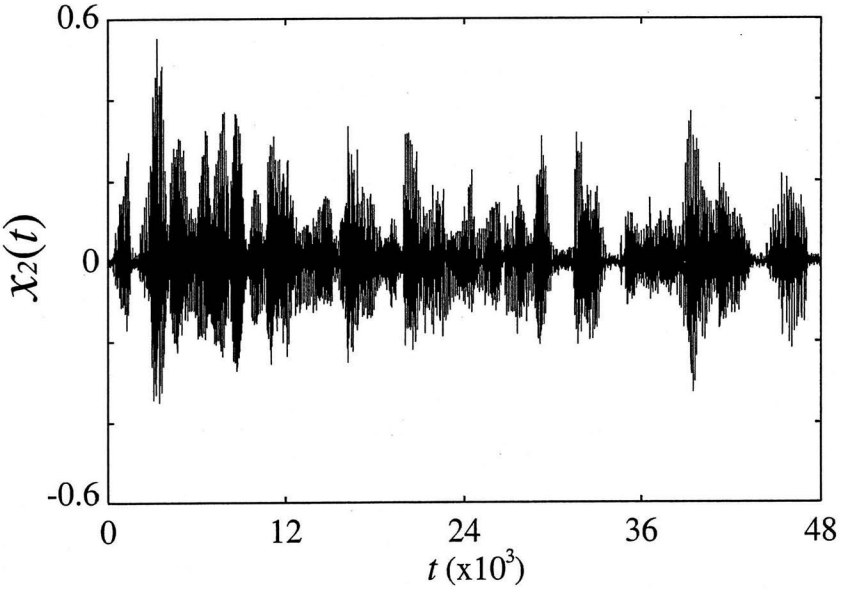


(b)

図 3.4: The source speech signals. (a) $s_1(t)$. (b) $s_2(t)$.

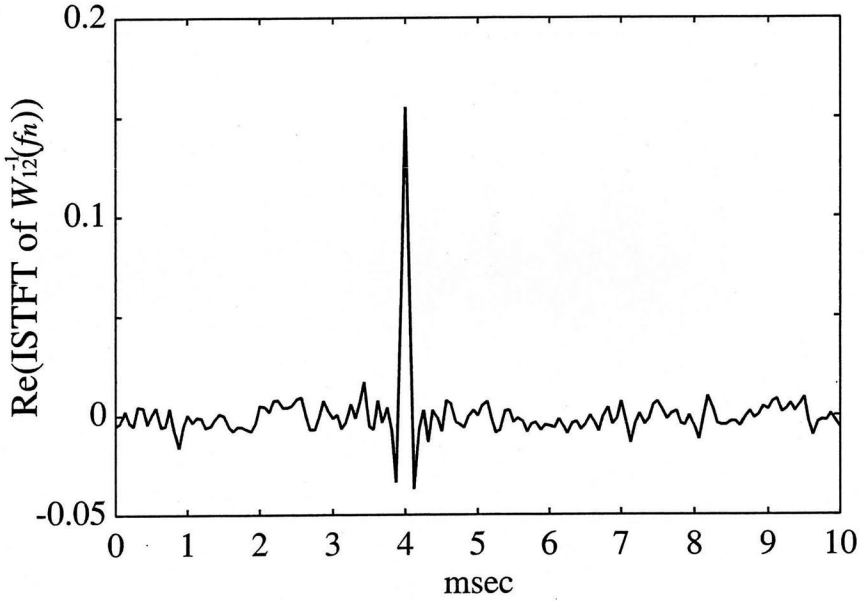


(a)

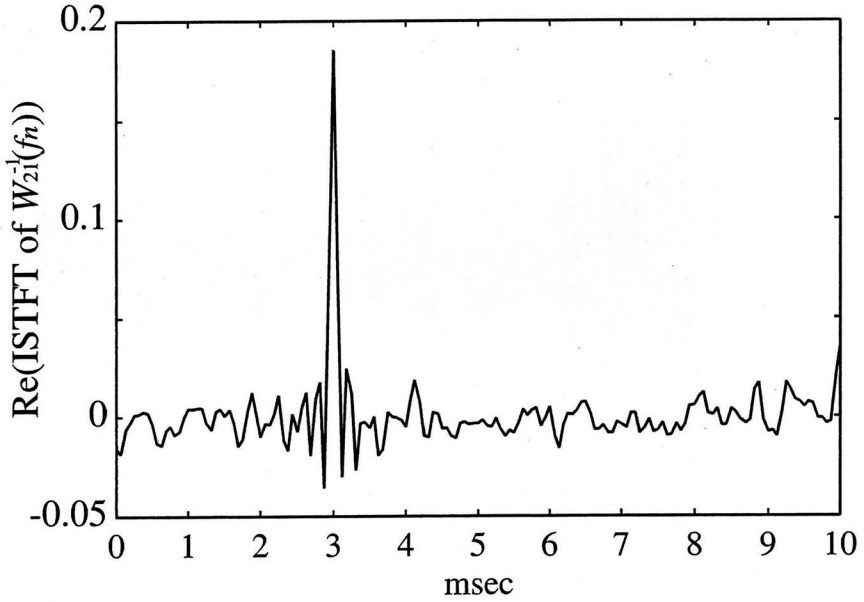


(b)

図 3.5: The mixed speech signals. (a) $x_1(t)$. (b) $x_2(t)$.

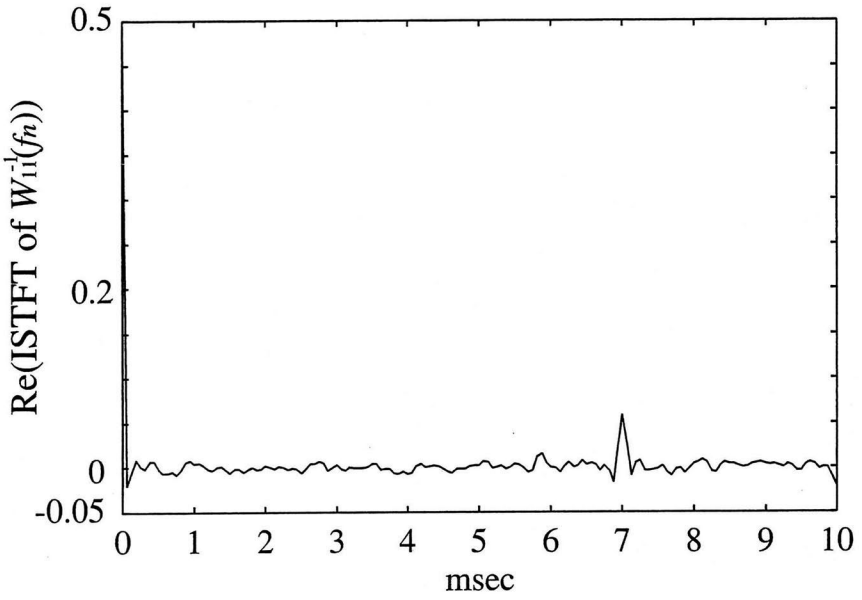


(a)

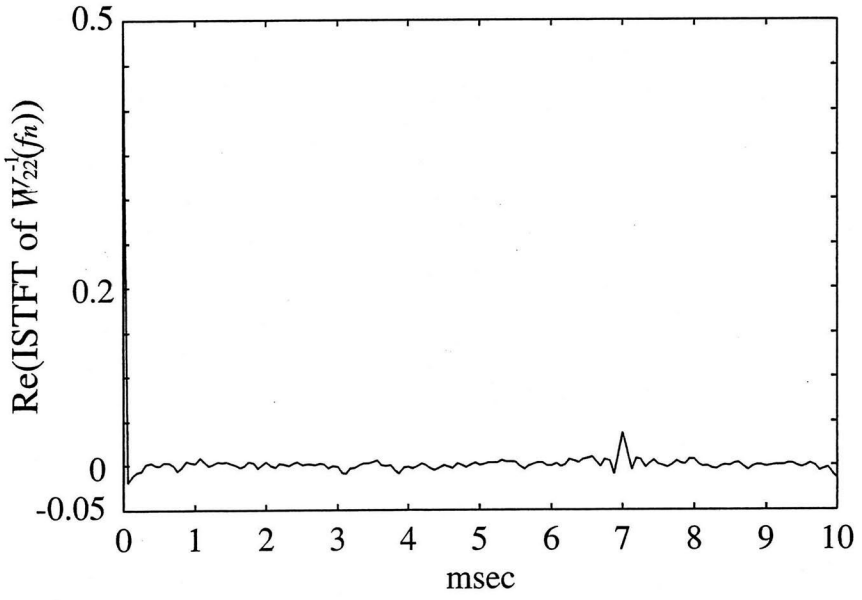


(b)

図 3.6: The real parts of the ISTFT of (a) $W_{12}^{-1}(f_n)$ and (b) $W_{21}^{-1}(f_n)$.



(a)



(b)

図 3.7: The real parts of the ISTFT of (a) $W_{11}^{-1}(f_n)$ and (b) $W_{22}^{-1}(f_n)$.

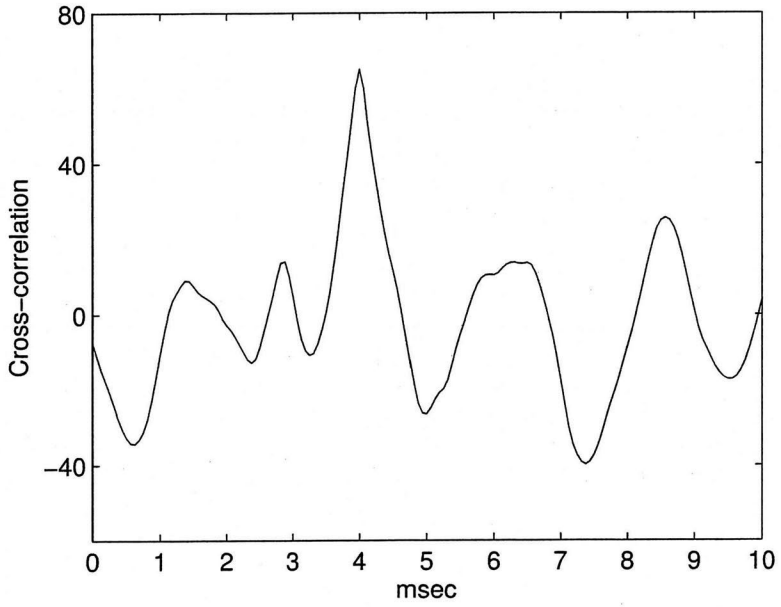


図 3.8: The cross-correlation function of $x_1(t)$ and $x_2(t)$.

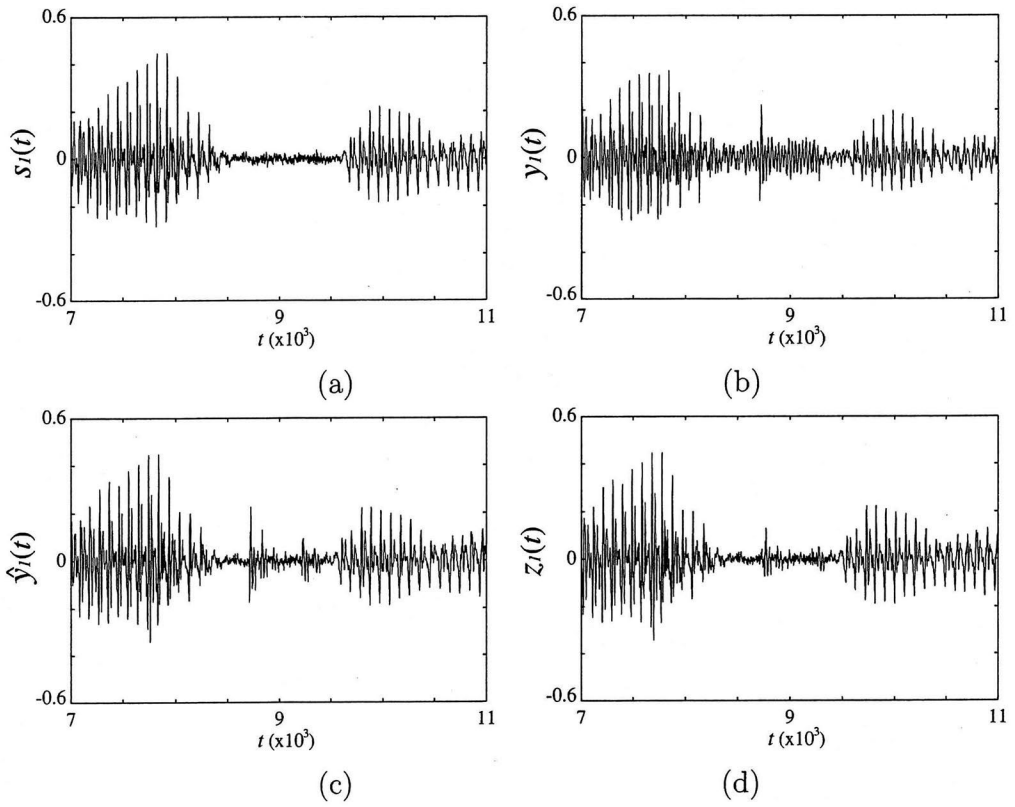


図 3.9: Signal separation results corresponding to $s_1(t)$. (a) The source speech signal $s_1(t)$. (b) The conventional frequency-domain ICA, $y_1(t)$ of Eq. (3.18). (c) The proposed method, $\hat{y}_1(t)$ of Eq. (3.31). (d) The proposed method, $z_1(t)$ of Eq. (3.33).

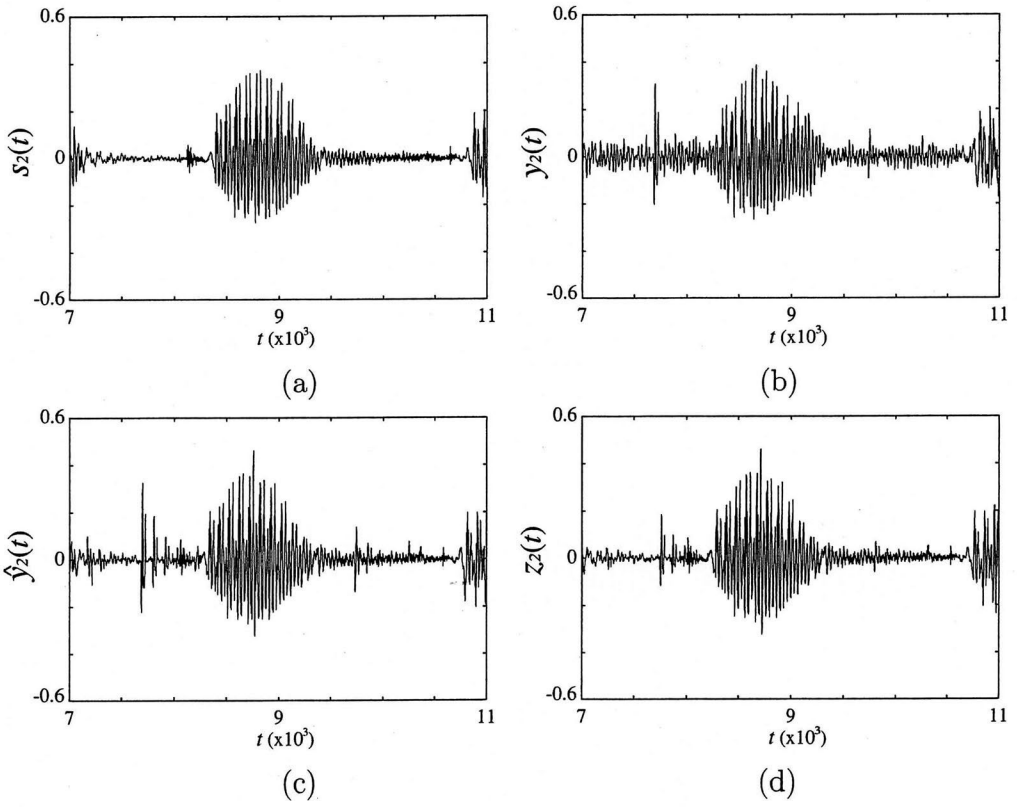


図 3.10: Signal separation results corresponding to $s_2(t)$. (a) The source speech signal $s_2(t)$. (b) The conventional frequency-domain ICA, $y_2(t)$ of Eq. (3.18). (c) The proposed method, $\hat{y}_2(t)$ of Eq. (3.31). (d) The proposed method, $z_2(t)$ of Eq. (3.33).

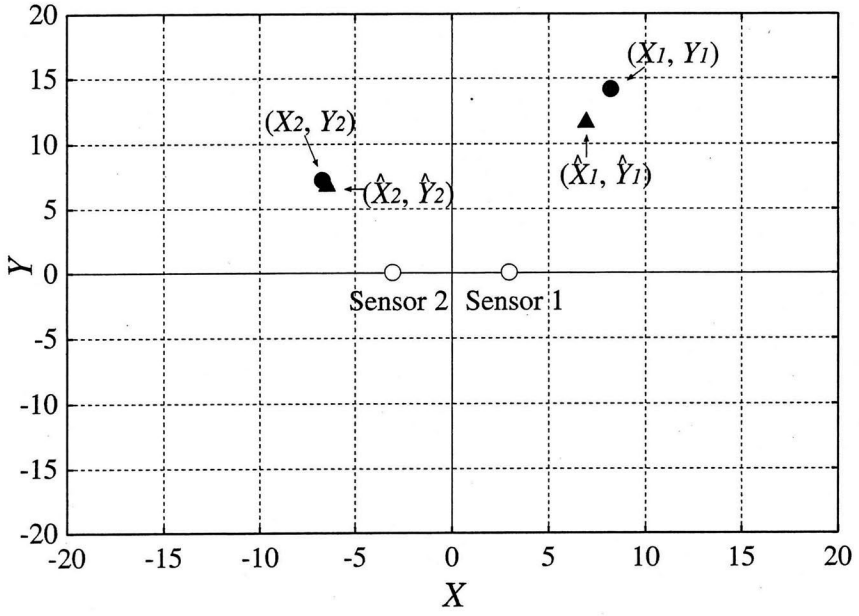


図 3.11: Results of the sound localization by using the proposed method.

第4章　1つの混合音声からの特定話者の音声抽出

4.1　緒言

本章では，特定話者の音声が背景雑音や他の話者の音声などにより汚されて観測されるとき，事前に得た特定話者のスペクトル情報を教師信号として，その観測信号から特定話者の音声を抽出するための手法を提案する．ここで，観測信号は1つだけとする．

提案手法では，音声信号の特徴ベクトルから構成される辞書を使用する．特徴ベクトルの集合は，すべての音韻を含む音声のスペクトル情報である．辞書は，特定話者の明瞭な音声から事前に作成され，観測信号から特定話者の音声を抽出するために使われる．提案手法の有効性は，計算機シミュレーションにより調べられた．

4.2　カクテルパーティー効果と独立成分分析

実環境下における音声信号は，少し離れた場所から空気を伝播してセンサーに入力されるため，観測信号に多数の雑音が混入されるのは避けられない．それゆえ，観測信号から所望の音声信号を抽出することは，騒音下における明瞭な音声抽出，電子媒体などで録音された音からの音声復元，機械による音声認識や個人認証などの前処理など様々な工学的な応用が考えられる．

第1章で述べたように，音声信号処理の分野で複数話者の混合音声から特定話者の音声を取り出すことは，カクテルパーティー問題と呼ばれる [8][9]．今まで述べて

きた独立成分分析 (ICA) は，この問題を解くための 1 つの手法である．しかしながら，ICA は信号間の統計的距離を使っているため，原則として 2 つ以上の観測信号を必要とする．それゆえ，ICA は 1 つの観測信号のみからでは，特定音声を抽出することはできない．人間の聴覚情報処理は，カクテルパーティー問題を何らかの方法で解いている．このとき，センサーが 2 つの場合に相当する両耳による効果は必ずしも必要ではなく，それ以外の何らかの方法が使われていると推測される．この理由から，ICA は非常に有用な手法ではあるが，本当の意味で完全にカクテルパーティー問題を解いているとは言えない．

ここでは，特定話者の音声と複数話者の音声から作られる混合音声を観測した場合，ICA のように複数の観測信号からでなく，ある 1 つの観測信号のみから特定話者の音声を抽出する方法について検討する．このとき，提案手法では，周波数領域における話者の個性が使われる．音声信号は，周波数領域に多くの特徴をもっている．例えば，基本周波数，フォルマント周波数，スペクトル包絡などがある [53]．これらの特徴は，話者の発話器官に由来する話者の個性を表す [54]．提案手法において，話者の個性は，特定話者の明瞭な音声から事前に作成される辞書に反映される．

4.3 提案する音声抽出法

本節で述べる提案手法は，以下に述べる 2 つのステップからなる，すなわち，辞書作成とその辞書を使った音声抽出である．

4.3.1 辞書作成

今回は，高速フーリエ変換 (FFT) によって得られたスペクトルが特徴ベクトル，すなわち，辞書の要素として採用される．辞書は，特定話者のすべての音韻を含む明瞭な音声によって作られる．この辞書は，特定話者にとっては普遍的なものであり，特定話者の音声が様々な雑音に汚された場合に適用できる．辞書の作成は，前

章で述べたような図3.2で表されるフレーム処理を使用する．各々のフレームでのFFTによって得られたスペクトルは，特徴ベクトルとして辞書に蓄積される．

4.3.2 辞書を利用した音声抽出

混合音声は連続的にフレームに分割され，各々のフレームでのスペクトル情報が，辞書作成の過程と同様にして計算される．このスペクトルは入力ベクトルとなり，辞書へ連続的に入力される．このとき，最もよく入力ベクトルと類似した特徴ベクトルを探すために，入力ベクトルと辞書内の特徴ベクトルの距離尺度が使われる．

今，入力ベクトルとi番目の特徴ベクトルを以下のように表す．

$$\boldsymbol{x} = (x_1, \ldots, x_N)^T, \tag{4.1}$$

$$\boldsymbol{v}_i = (v_{i1}, \ldots, v_{iN})^T, \tag{4.2}$$

ここで，Nはベクトルの次元であり，Tは転置を表す．音声信号のスペクトル特性は，話者それぞれの発話器官の性質によって決まるため，2つのスペクトル間の距離尺度は，これらの性質を考慮しなければならない．それゆえ，これらの特性を表現する距離尺度として，次式で与えられるベクトル間の各要素の重み付き2乗誤差の和$d_i(\boldsymbol{x})$を採用する．

$$d_i(\boldsymbol{x}) = \sum_{n=1}^{N} w_n (x_n - v_{in})^2, \tag{4.3}$$

ここで，w_nは特定話者の個性を表す重みとなる．

重みをどのように与えるかが問題となるが，今回は以下のように考えてみる．話者の個性は，基本周波数やフォルマント周波数といった特定の周波数領域において顕著に表れる．それゆえ，この特徴は，特定話者のスペクトル情報を蓄積した辞書内の特徴ベクトルの統計量で表現できると考え，今回は次式のように重みw_nとして，すべての特徴ベクトルの平均を採用してみる．そのとき，話者の個性は式(4.3)

に反映される.

$$w_n = \frac{1}{M}\sum_{i=1}^{M} v_{in}, \tag{4.4}$$

ここで，M は特徴ベクトルの総数である.

辞書の出力 $\boldsymbol{y}(\boldsymbol{x})$ は，入力ベクトルと最近接する K 個の特徴ベクトルの重み付き加算として次式のように与えられる.

$$\boldsymbol{y}(\boldsymbol{x}) = \frac{\sum_{k=1}^{K} \mu_k(\boldsymbol{x})\boldsymbol{v}_k}{\sum_{k=1}^{K} \mu_k(\boldsymbol{x})} \tag{4.5}$$

ここで,

$$\mu_k(\boldsymbol{x}) = 1 - \frac{d_k(\boldsymbol{x})}{\sum_{j=1}^{K} d_j(\boldsymbol{x})}. \tag{4.6}$$

時間領域の音声は，この出力を逆 FFT(IFFT) することにより得られる.

4.4 計算機シミュレーション結果

本節では，提案手法の有効性を調べるために，計算機シミュレーション結果を述べる.

4.4.1 実験条件

辞書作成は，女性話者の ATR 音素バランス文 50 文のうち 49 文，残り 1 文は観測信号として用いた．ATR 音素バランス文は，新聞，雑誌，小説，手紙，教科書等からの 50 個の文章で，音素バランスがとれるように構成したものである．サンプリング周波数は 16kHz，量子化ビット数は 16 ビットでモノラル音声信号であった．辞書作成で用いた 49 文の時間の合計は 283.904 秒であった．図 4.1 は，今回作成した辞書に式 (4.4) を適用して得られた重みを表す.

1つの観測信号 $x_1(t)$ は，図 4.2(a) に示される辞書作成に用いなかった ATR 音素バランス文の1文「野球のあとのビールぐらいうまいものはない」の朗読音声 $s_1(t)$ と他の1人の男性話者の「私はその人を常に先生と呼んでいた」(「こころ」(夏目漱石)) の朗読音声 $s_2(t)$ と他の1人の女性話者の「あらゆる現実をすべて自分の方へねじ曲げたのだ」の朗読音声 $s_3(t)$ を次式のように計算機上で混合して作る．

$$x_1(t) = s_1(t) + 0.4s_2(t) + 0.5s_3(t). \tag{4.7}$$

$s_1(t)$ が抽出したい特定話者の音声である．$s_1(t)$，$s_2(t)$，$s_3(t)$ ともにサンプリング周波数 16kHz，量子化ビット数 16 ビット，3秒間のモノラル音声信号であった．図 4.2(b) は，観測信号 $x_1(t)$ を示す．表 4.1 は，各々の音声信号の基本周波数の平均と標準偏差を示す．

表 4.1: Fundamental frequency (Hz) of each speech signal.

	Mean	SD
$s_1(t)$ (Target speech signal of a feamale)	229.0	28.9
$s_2(t)$ (Speech signal of a male)	133.5	25.6
$s_3(t)$ (Speech signal of another female)	308.1	53.5

4.4.2 実験結果

図 4.2(c) は，提案手法により抽出された音声信号 $y(t)$ を示す．図 4.3(a) と図 4.3(b) と図 4.3(c) は，それぞれ，図 4.2(a) と図 4.2(b) と図 4.2(c) の音声信号のサウンドスペクトログラムを示す．

これらの図より，抽出音声は，特定話者の音声信号の特徴をよく捉えていることが分かる．また，抽出音声を実際に聞いてみると，他の話者の音声がほとんど取り除かれており，効果的に音声抽出ができていることを確認できた．

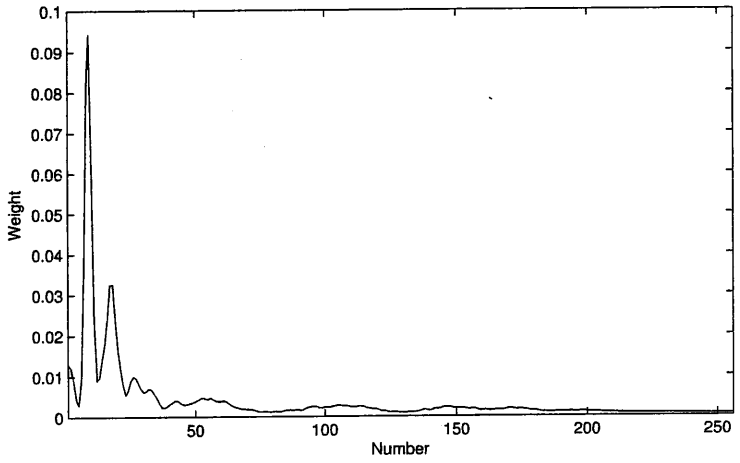


図 4.1: The weight obtained by averaging the feature vectors in the dictionary.

さらに，抽出性能を定量的に比較するために，次式で定義される信号対雑音比 (SNR) を計算した.

$$\mathrm{SNR} = 10\log_{10}\frac{\sum_t s_1(t)^2}{\sum_t(\chi(t) - s_1(t))^2}, \tag{4.8}$$

ここで，$s_1(t)$ は特定話者の音声信号，$\chi(t)$ には式 (4.7) で与えられる観測信号 $x_1(t)$，または，提案手法により抽出された音声信号 $y(t)$ が代入される．表 4.2 は，SNR の結果を示す.

表 4.2: SNR (dB) of the speech signal defined by Eq. (4.8).

$\chi(t)$	SNR
$x_1(t)$ (Mixed speech signal)	6.390
$y(t)$ (Extracted speech signal)	6.932

また，パワースペクトルに対しても次式で定義されるSNRを計算した．

$$\mathrm{SNR} = 10\log_{10}\frac{\sum_{f_n,t_s}|s_1(f_n,t_s)|^2}{\sum_{f_n,t_s}(|\chi(f_n,t_s)|-|s_1(f_n,t_s)|)^2}, \tag{4.9}$$

ここで，f_nは離散周波数，t_sはフレーム番号，$|\cdot|$は複素数の絶対値である．$s_1(f_n,t_s)$は原信号のフレームt_s番目の離散スペクトルであり，$\chi(f_n,t_s)$には$x_1(t)$の離散スペクトル$x_1(f_n,t_s)$，または，$y(t)$の離散スペクトル$y(f_n,t_s)$が代入される．表4.3は，SNRの結果を示す．表4.2と表4.3から，抽出された音声信号のSNRは時間領域，パワースペクトルの場合ともに改善されていることが分かる．

表 4.3: SNR (dB) of the power spectrum defined by Eq. (4.9).

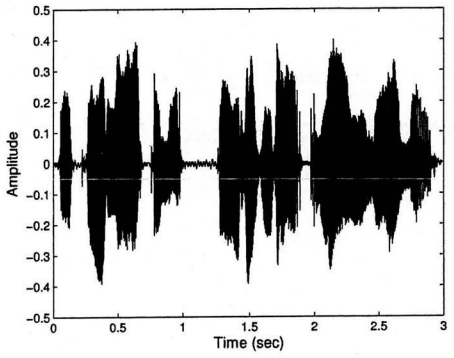
$\chi(f_n,t_s)$	SNR
$x_1(f_n,t_s)$ (Mixed speech signal)	9.57
$y(f_n,t_s)$ (Extracted speech signal)	9.94

4.5 結言

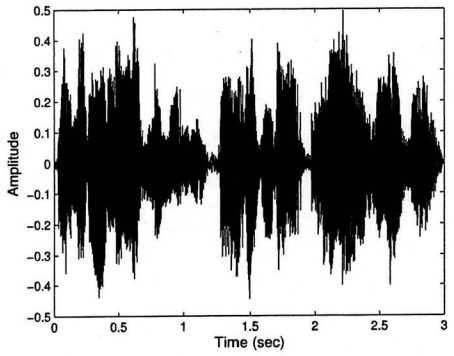
本章では，事前に得た特定話者のスペクトル情報を特徴ベクトルとして辞書内に蓄積し，その辞書を使って観測信号から特定話者の音声を抽出する方法を提案した．具体的には，辞書内の特徴ベクトルと観測信号の特徴ベクトルの距離尺度として，話者の個性を考慮できるようなものを考え，それを用いて選ばれた数個の特徴ベクトルから抽出音声を復号する方法を提案した．

計算機シミュレーション結果は，提案手法が非常に簡単な方法にもかかわらず，3話者の混合音声から作られる1つだけの観測信号から特定話者の音声を抽出できることを示した．

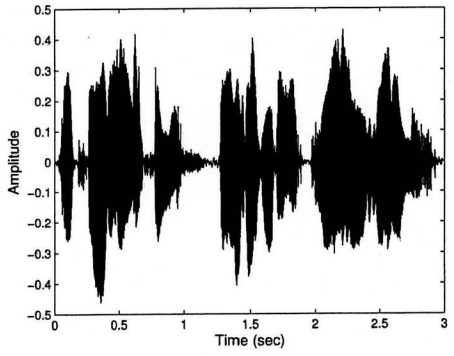
今後の課題は，ベクトル量子化[78][79]を用いることにより，抽出音声の品質をあまり損ねることのない特徴ベクトル辞書の圧縮方法とそれを用いた復号方法を考案すること，また，話者の個性をさらによく表現する距離尺度を探すことなどである．



(a)

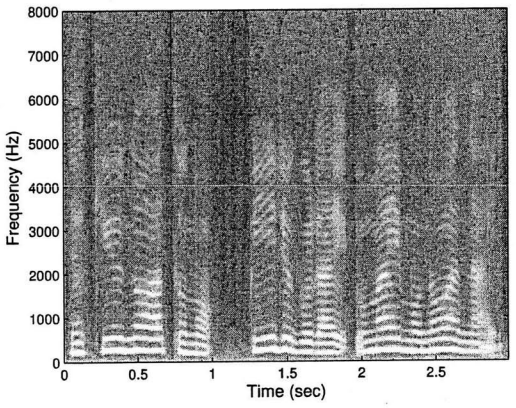


(b)

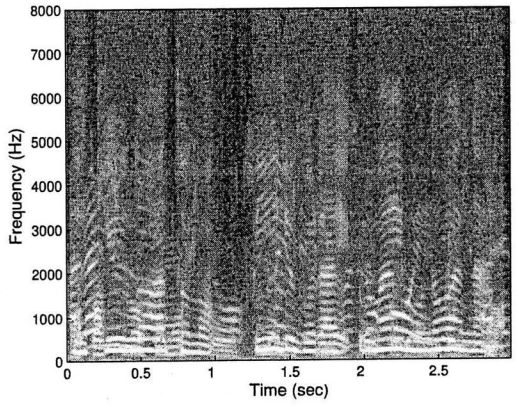


(c)

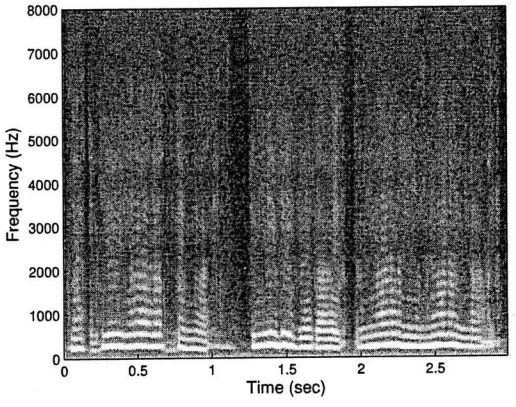
図 4.2: Results of the speech extraction. (a) The target speech signal of a female $s_1(t)$. (b) The mixed speech signal of three people (two females and one male) $x_1(t)$. (c) The extracted speech signal by the proposed method $y(t)$.



(a)



(b)



(c)

図 4.3: Sound spectrograms corresponding to the signals shown in Figs.4.2(a), (b), and (c), respectively.

第5章　独立成分分析を用いた音声信号の特徴抽出と再構成

5.1　緒言

本章では，画像信号の特徴抽出にも応用されている独立成分分析 (ICA) を，時間領域の音声信号の特徴抽出に用いた結果を報告する．

信号を効率的に表現するためには，信号の性質をよく反映する特徴の抽出が非常に重要である．信号の特徴量として色々なものが考えられるが，ここでは最も簡単なモデルを考え，信号がある基底関数の重み付き加算で表わせると仮定する．ここでの基底関数は，第2章で述べたRBFネットワークの基底関数と直接の関係はない．このとき，信号の特徴量は基底関数とその重みになる．ここで，情報圧縮の観点から，できるだけ少ない基底関数で信号を表現することを考える．従来は，数学的簡明さから，このモデルでの特徴量を計算するためにフーリエ展開が用いられてきた．しかしながら，フーリエ展開は基底関数を三角関数に固定しているため，信号の統計的性質が反映されず，多くの基底関数を用いなければ信号を表わせない欠点がある．

また，信号の統計量から基底関数を決定する手法として代表的なものに主成分分析 (PCA) がある [83]．PCAは，信号の2次統計量を用いて基底関数を決定するため，ある程度信号の統計的性質を考慮した基底関数を得ることが可能である．しかしながら，基底関数に直交条件を付加しているため，得られる基底関数は三角関数に類似したものになる．また，基底関数の数は，基底関数の次元より多くすること

はできない.

一方, 最近注目されている手法として, スパースコーディングがある [31][57][58][84]. スパースコーディングは, 自然画像における特徴量抽出のために考え出された手法である. このスパースコーディングは, 重みに疎 (スパース) 性の条件を付加する. また, 基底関数に直交条件を課さず, 基底関数の数を基底関数の次元より多くすることができる. これにより, 少数の基底関数の加算で信号を表現することが可能になる. 得られた基底関数の形状は視覚系の受容野によく似ていると報告されており, 信号の統計的性質をよく反映していることが分かる.

このスパースコーディングはICAと密接な関係があり, 考え方としては, ほとんど同じであることが示されている [58]. 実際, スパースコーディングの結果を受けて, ICAを自然画像の特徴抽出に適用した結果, 同じような基底関数が得られたと報告されている [55][56].

そこで, 本章ではICAを音声信号の特徴抽出に適用し, どのような基底関数が得られるかを調べ, また, 得られた基底関数が少数でも音声信号を再構成することが可能かどうか, 再構成した音声信号の誤差とそれに使われた基底関数の数の関係を調べることにより, 情報圧縮の観点から定量的検討を行う.

5.2 独立成分分析による特徴抽出

本節では, 独立成分分析による音声信号の特徴抽出の方法について述べる. 基本的には, 画像信号の特徴抽出の方法に準じている [55].

5.2.1 音声信号の表現

音声信号は非定常な信号であるが, 10ms程度の時間区間でみると定常と近似することもできる [54]. そのため, その程度の短い時間区間の音声信号 $\boldsymbol{x}$ は, 基底関数

$\boldsymbol{a}_i$ の重み付き加算，

$$\boldsymbol{x} = \sum_{i}^{N} s_i \boldsymbol{a}_i, \tag{5.1}$$

で表現されると仮定する [85]．N は基底関数の総数，係数 s_i は基底関数 $\boldsymbol{a}_i$ の選択によって決定される重みである．また，$\boldsymbol{x}$ と $\boldsymbol{a}_i$ は，それぞれ，

$$\boldsymbol{x} = (x_1, \ldots, x_M)^T, \tag{5.2}$$

$$\boldsymbol{a}_i = (a_{i1}, \ldots, a_{iM})^T, \tag{5.3}$$

で与えられる．M は離散化された音声信号の次元である．N と M は当然異なっても良く，$M < N$ の場合は過完備基底と呼ばれ [81][82]，特徴抽出では一般的である．しかし，ここでは式 (2.24) を用いて学習するため M=N とする．

5.2.2 基底関数と重みの学習

式 (5.1) を行列形式で書き直すと，

$$\boldsymbol{x} = \boldsymbol{A}\boldsymbol{s}, \tag{5.4}$$

となる．ここで，

$$\boldsymbol{s} = (s_1, \ldots, s_N)^T, \tag{5.5}$$

$$\boldsymbol{A} = (\boldsymbol{a}_1, \ldots, \boldsymbol{a}_N)^T, \tag{5.6}$$

である．式 (5.4) を式 (2.1) と対応させると，基底関数の集合が混合行列，基底関数の係数が原信号，離散化された音声信号の各要素が観測信号となる．ICA を用いる目的は，係数 s_i が互いに統計的に独立で，線形の重ね合せによって短区間の音声信号を構成できる基底関数 $\boldsymbol{a}_i$ を発見することである．

ICA の学習は特定話者の比較的長い音声を，第 3 章，4 章のように短区間のフレームに切り出してを行う．ただし，切り出し位置は乱数により決定する．

5.3 計算機シミュレーション結果

本節では，音声信号の特徴抽出にICAを用いた結果，得られた基底関数について報告する．次に，それを使った再構成音声の誤差について情報圧縮の観点から示す．

5.3.1 実験条件

基底関数の作成には，前章の辞書作成と同じく女性話者のATR音素バランス文50文のうち49文，サンプリング周波数16kHz，量子化ビット数16ビット，283.904秒間のモノラル音声信号を，ダウンサンプリングして8kHzにしたものを用いた．この基底関数を使用して再構成する音声信号として，同話者の基底関数の作成に用いていないATR音素バランス文の1文，3秒間のモノラル音声信号を，ダウンサンプリングして8kHzにしたものを用いた．

ICAの学習は，式(2.24)を確率降下学習にしたものを使う[29]．

$$\Delta \boldsymbol{W} = \eta \left(\boldsymbol{I} - \boldsymbol{\phi}(\boldsymbol{y}) \boldsymbol{y}^T \right) \boldsymbol{W}. \tag{5.7}$$

式(2.24)は全データが入力されてから$\boldsymbol{W}$を更新するが，式(5.7)はデータが入力されるたびに$\boldsymbol{W}$を更新する．また，ϕとして，ここでは，

$$\phi(y_i) = -1 + \frac{2}{1 + \exp(-y_i)}, \tag{5.8}$$

を用いた[20][55]．フレーム長は0.008秒，切り出したフレームの総数は10,000個，学習回数は1,000回である．

5.3.2 実験結果

図5.1は，ICAによって得られた基底関数を示す．この基底関数の横軸は時間で64点(0.008秒)あり，縦軸は振幅である．周期性をもつもの，雑音のような不規則な

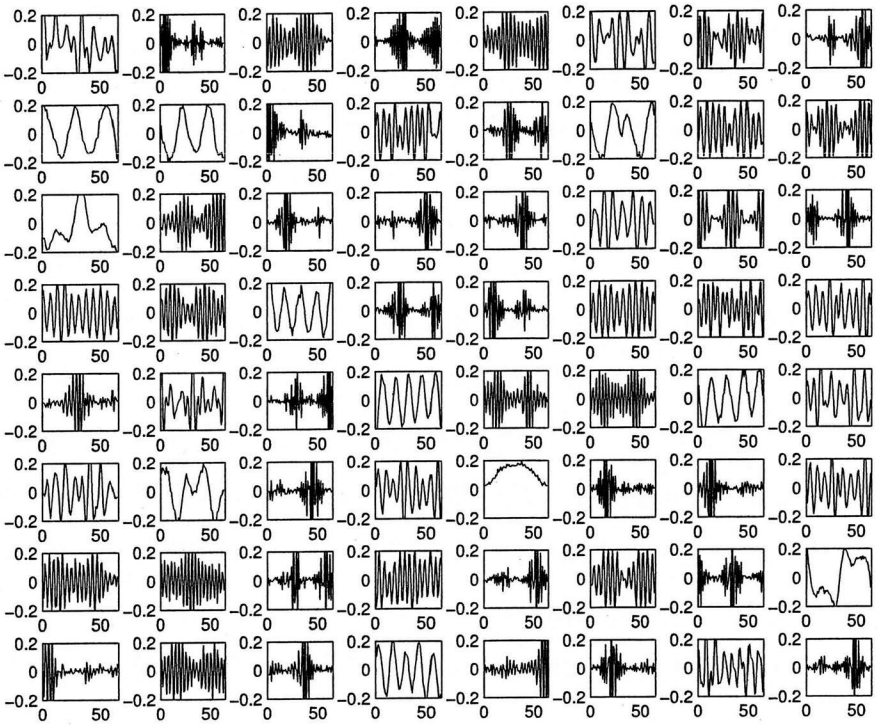


図 5.1: The basis functions obtained by the ICA from the speech signal.

もの，局所的な構造をもつものなど，様々な特徴的な波形があることが確認できる．

図 5.2(a) は原信号を示す．図 5.2(b) と図 5.2(c) と図 5.2(d) は，それぞれ，基底関数を 3 個，9 個，すべて (64 個) を用いて再構成した音声信号を示す．3 個では原信号の特徴は捉えているものの誤差が大きいが，9 個ではある程度再構成できていることが分かる．基底関数をすべて使うと原信号を完全に復元できる．

再構成音声に用いた基底関数の数と再構成音声と原音声間の誤差との関係を定量

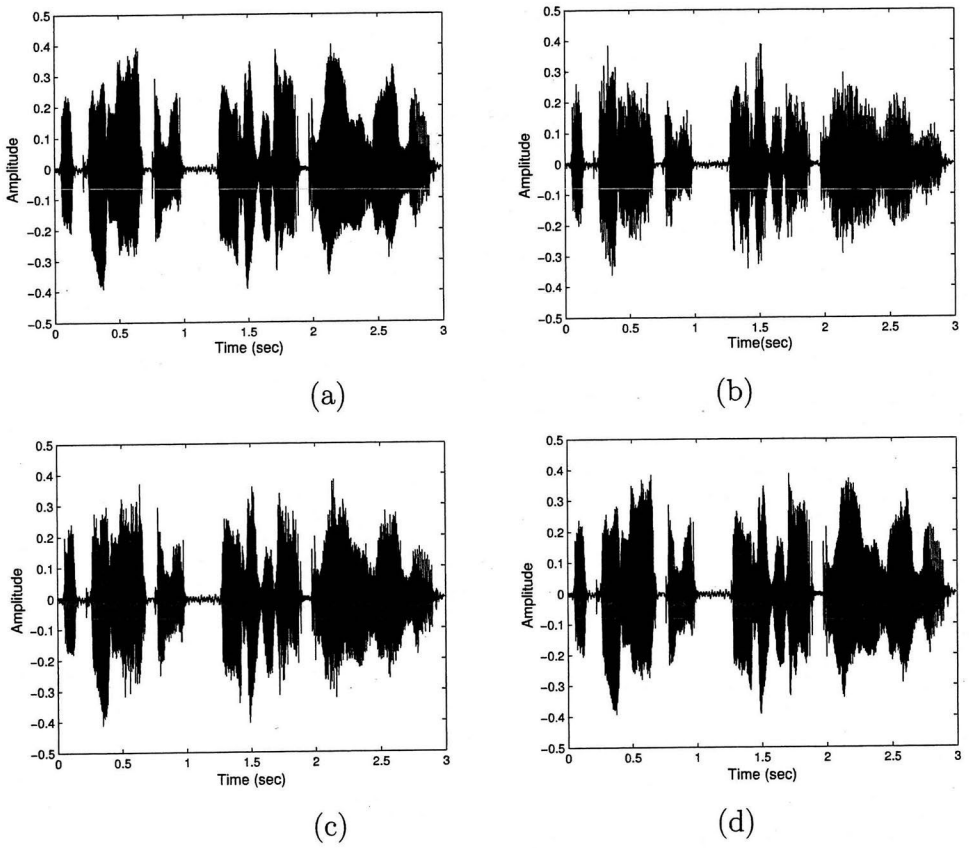


図 5.2: Results of the speech reconstruction. (a) The source speech signal. (b) The reconstructed speech signal by three basis functions, (c) nine basis functions, (d) all basis functions.

的に調べるために，次式で定義される正規化した 2 乗誤差を使用した．

$$\mathrm{NSE} = \frac{\sum_t (s(t) - y(t))^2}{\sum_t s(t)^2}. \tag{5.9}$$

結果を図 5.3 に示す．最初の 10 個程度まで急激に減少し，後はゆるやかに減少していることが分かる．

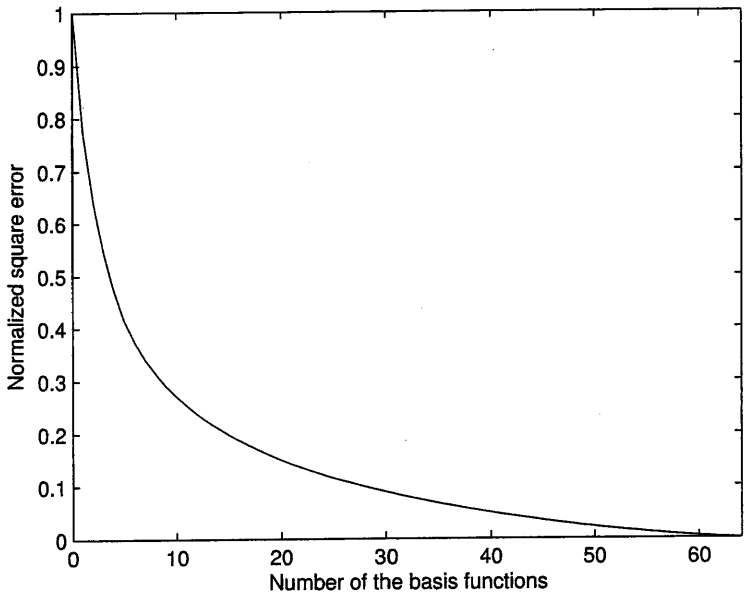


図 5.3: The normalized square error defined by Eq. (5.9) versus number of the basis functions.

5.4 結言

本章では，ICA による特徴抽出の方法を説明した後，それを音声信号に適用した結果を報告した．

計算機シミュレーション結果により，ICA により得られた基底関数が，周期性をもつもの，雑音のような不規則なもの，局所的な構造をもつものなど，様々な特徴的な波形であることを確認できた．また，比較的少ない基底関数でも，ある程度音声信号を再構成することができ，情報圧縮への応用が可能であることも示された．

今後の課題は，情報圧縮をさらに進めるため，画像に対して成功している ICA の特徴抽出の方法を，音声の統計的な性質を十分に反映できる形に改良することが挙げられる．改良する点としては，音声と自然画像の統計的な性質において，類似した点や異なる点を比較検討し，音声信号の基底関数による表現を単なる加算から変

更することなどが考えられる．また，得られた基底関数の意味づけを音素レベルで行い，個人認証などへの応用が可能かどうか検討することである．

第6章　結論

本論文では，カクテルパーティー問題として知られている，雑音に埋もれた所望の音声信号からその音声信号を分離・抽出する問題を解くために，音声の統計的特徴量を利用した幾つかの方法について述べた．

最初の第2章，3章では，2つのブラインド信号分離 (BSS) 問題に対して，独立成分分析 (ICA) に基づく信号分離アルゴリズムを提案した．1つめは，混合過程に時間遅れがないと仮定できる場合，すなわち，瞬時混合 BSS 問題に対するアルゴリズムで，もう1つは，混合過程で伝搬遅延時間を考慮しなければならない BSS 問題に対するアルゴリズムである．

第2章では，瞬時混合 BSS 問題を解くために利用される瞬時混合 ICA において，その学習過程で必要とされる非線形関数をできるだけ正確に記述するために，RBF ネットワークを使用した ICA の学習アルゴリズムを提案した．計算機シミュレーション結果により，提案手法は従来手法に比べ計算量は増加するが信号分離の精度が向上し，また，従来手法では分離ができないような確率分布をもつ原信号に対しても分離を行えることを確認した．さらに従来手法と提案手法の複合型モデルも提案し，信号分離の収束スピードと精度の更なる向上を目指した．この有効性も音声信号に対する計算機シミュレーションにより確認された．

第3章では，伝搬遅延時間を含む BSS 問題に対して，周波数領域 ICA に基づく手法を提案した．提案手法では，周波数領域 ICA で得られた分離行列に注目し，分離行列から相対伝播遅延時間と減衰係数比が求まることを見出した．計算機シミュレーション結果により，特に相対伝播遅延時間は，サンプリング周波数で決まる時

間解像度内の正確さで求まることを確認した．また，これらの量を信号分離に用いると，従来手法に比べ分離性能が向上することを確認した．さらに，相対伝播遅延時間と減衰係数比から信号源の位置を求めること，すなわち，音源定位が可能であることも示した．しかし，減衰係数比の推定値の精度が相対伝播遅延時間ほど正確ではなく，そのことが，推定された音源位置の精度に影響を及ぼした．

次の第4章，5章では，音声信号の特徴量抽出とそれを用いた音声抽出・再構成について，2つの方法を検討した．1つめは，音声信号の周波数領域における特徴を使うもので，もう1つは，ICAやスパースコーディングに基づいた時間領域の音声信号の特徴抽出である．

第4章では，事前に得た特定話者のスペクトル情報を特徴ベクトルとして辞書内に蓄積し，その辞書を使って観測信号から特定話者の音声を抽出する方法を提案した．提案手法では，辞書内の特徴ベクトルと観測信号の特徴ベクトルの距離尺度として，話者の個性も考慮できるようなものを考え，その距離尺度を用いて選ばれた辞書内の数個の特徴ベクトルから，抽出音声を復号する方法を示した．計算機シミュレーション結果により，提案手法が非常に簡単な方法にもかかわらず，3話者の混合音声から作られる1つだけの観測信号から特定話者の音声を抽出できることを確認した．

第5章では，ICAによる特徴抽出を時間領域の音声信号に適用した結果を報告した．計算機シミュレーション結果により，得られた基底関数が，周期性をもつもの，雑音のような不規則なもの，局所的な構造をもつものなど，様々な特徴的な波形であることを確認した．それぞれの基底関数の意味づけは行っていないが，これらは，音声の統計的性質を反映しているのではないかと考えられる．また，情報圧縮の観点から，再構成した音声信号の誤差とそれに使われた基底関数の数の関係を調べた．その結果，比較的少ない基底関数でも，ある程度音声信号が再構成できることを確認した．

人間の聴覚情報処理では，脳が自然界の中で獲得してきた様々な事前知識が使われていると考えられる．どのような事前知識が脳内に蓄積されているかは，音声信号の統計的特徴量と深い関連があると推測できる．これらの事前知識とICAなどの統計的信号処理手法を組み合わせることで，人間の聴覚情報処理を模擬した，さらに精度の高い音声信号のための信号分離・抽出アルゴリズム(カクテルパーティー効果)を考案することが，今後の最終的な課題となる．

謝辞

本研究の遂行にあたり，本研究テーマに導いて下さり，懇切丁寧な御指導を賜りますとともに，終始暖かい励ましを頂きました山口大学大学院理工学研究科 内野 英治 教授に謹んで感謝の意を表します．また，熱心な御指導と様々な御配慮を賜りました同大学院理工学研究科 末竹 規哲 助教授に心から感謝の意を表します．

さらに，本論文の副査をお引き受け下さり，多くの有益なご教示を賜りました，同大学院理工学研究科 増山 博行 教授，吉川 学 教授，松野 浩嗣 教授に厚くお礼申し上げます．

最後に，研究を進める上で多くの御意見，御協力を頂きました矢野 和昭 氏，浮田 宏真 氏，村田 和俊 氏，川崎 隆介 氏，筒井 俊太 氏をはじめ，内野，末竹研究室の皆さんに深く感謝致します．

参考文献

[1] 中田和男, 音声, コロナ社 (1995).

[2] L. R. Rabiner, R. W. Schafer, 鈴木久喜 (訳), 音声のディジタル信号処理 (上) (下), コロナ社, (1983).

[3] 今井聖, 音声認識, 共立出版, (1995).

[4] A. S. Bregman, Auditory scene analysis, MIT Press, Cambridge MA (1990).

[5] D. Wang and G. J. Brown (Ed.), Computational auditory scene analysis, IEEE Press, Piscataway (2006).

[6] S. ヘイキン, 武部幹 (訳), 適応フィルタ入門, 現代工学社 (1987).

[7] S. Haykin (Ed.), Unsupervised adaptive filtering, Volume I: Blind source separation, John Wiley, New York (2000).

[8] 三浦種敏 (監修), 聴覚と音声, 電子情報通信学会 (1980).

[9] B. Arons, "A review of the cocktail party effect," *Journal of the American Voice I/O Society*, Vol.12, pp.35–50 (1992).

[10] S. Haykin and Z. Chen, "The cocktail party problem," *Neural Computation*, Vol.17, pp.1875–1902 (2005).

[11] A. Cichocki and S. Amari, Adaptive blind signal and image processing, John Wiley, New York (2002).

[12] 松岡清利, "ブラインド信号分離," 日本ファジィ学会誌, Vol.10, No.3, pp.394–400 (1998).

[13] C. E. Shannon, "A mathematical theory of communication," *Bell System Technical Journal*, Vol.27, pp.379–423,623–656 (1948).

[14] 佐藤洋, 情報理論, 裳華房 (1983).

[15] C. Jutten and J. Herault, "Blind separation of sources, Part I: An adaptive algorithm based on neuromimetic architecture," *Signal Processing*, Vol.24, pp.1–10 (1991).

[16] J. -F. Cardoso and A. Souloumiac, "Blind beamforming for non Gaussian signals," *IEE Proceedings-F*, Vol.140, No.6, pp.362–370 (1993).

[17] J. -F. Cardoso and A. Souloumiac, "Jacobi angles for simultaneous diagonalization," *SIAM Journal on Matrix Analysis and Applications*, Vol.17, No.1, pp.161–164 (1996).

[18] A. Hyvärinen, "Fast and robust fixed-point algorithms for independent component analysis," *IEEE Transactions on Neural Networks*, Vol.10, No.3, pp.626–634 (1999).

[19] P. Comon, "Independent component analysis, a new concept?," *Signal Processing*, Vol.36, pp.287–314 (1994).

[20] A. J. Bell and T. J. Sejnowski, "An information-maximization approach to blind separation and blind deconvolution," *Neural Computation*, Vol.7, pp.1129–1159 (1995).

[21] S. Amari, A. Cichocki and H. H. Yang, " A new learning algorithm for blind signal separation," *Advances in Neural Information Processing Systems*, Vol.8, pp.757–763, MIT Press, Cambridge MA (1996).

[22] H. H. Yang and S. Amari, " Adaptive on-line learning algorithm for blind separation: Maximum entropy and minimum mutual information," *Neural Computation*, Vol.9, pp.1457–1482 (1997).

[23] L. Molgedey and H. G. Schuster, " Separation of a mixture of independent signals using time delayed correlations," *Physical Review Letters*, Vol.72, No.23, pp.3634–3637 (1994).

[24] K. Matsuoka, M. Ohya and M. Kawamoto, " A neural net for blind separation of nonstationary signals, " *Neural Networks*, Vol.8, No.3, pp.411–419, (1995).

[25] 池田思朗, " 独立成分解析の信号処理への応用, " 計測と制御, Vol.38, No.7, pp.461–467 (1999).

[26] 赤穂昭太郎, 梅山伸二, " マルチモーダル独立成分分析ー複数情報源からの共通特徴抽出法ー, " 電子情報通信学会誌, Vol.J83-A, No.6, pp.669–676 (2000).

[27] T. -W. Lee, Independent component analysis: Theory and applications, Kluwer Academic Publishers, Boston (1998).

[28] A. Hyvärinen, J. Karhunen and E. Oja, Independent component analysis, John Wiley, New York (2001).

[29] 甘利俊一, “独立成分分析とその周辺,” 多変量解析の展開, pp.1–63, 岩波書店 (2002).

[30] 甘利俊一, 村田昇, 独立成分分析 多変量データ解析の新しい方法, サイエンス社 (2002).

[31] 村田昇, 入門 独立成分分析, 東京電機大学出版局 (2004).

[32] 竹内啓, 確率分布の近似, 教育出版 (1975).

[33] E. Uchino and T. Yamakawa, “Soft computing based signal prediction, restoration, and filtering,” in: Intelligent Hybrid Systems: Fuzzy Logic, Neural Networks, and Genetic Algorithms, Kluwer Academic Publishers, pp.331–351 (1997).

[34] E. Uchino, S. Nakamura and T. Yamakawa, “Nonlinear modeling and filtering by RBF network with application to noisy speech signal,” *Journal of Information Sciences*, Vol.101, pp.177–185 (1997).

[35] 坂和正敏, 田中雅博, ニューロコンピューティング入門, 森北出版 (1997).

[36] 堀口剛, 佐野雅己, 情報数理物理, 講談社 (2000).

[37] K. Torkkola, “Blind separation of delayed sources based on information maximization,” *Proc. ICASSP*, pp.3510–3513, Atlanta, GA, May (1996).

[38] K. Torkkola, “Blind separation of convolved sources based on information maximization,” *Proc. IEEE Workshop on Neural Networks for Signal Processing*, pp.423–432, Kyoto, Sep. (1996).

[39] T. -W. Lee, A. J. Bell and R. Lambert, "Blind separation of delayed and convolved sources," *Advances in Neural Information Processing Systems*, Vol.9, pp.758–764, MIT Press, Cambridge MA (1996).

[40] P. Smaragdis, "Blind separation of convolved mixtures in the frequency domain," *Neurocomputing*, Vol.22, pp.21–34 (1998).

[41] N. Murata and S. Ikeda, "An on-line algorithm for blind source separation on speech signals," *Proc. of 1998 International Symposium on Nonlinear Theory and Its Application*, pp.923–926 (1998).

[42] S. Kurita, H. Saruwatari, S. Kajita, K. Takeda and F. Itakura, "Blind signal separation using directivity pattern," *Technical Report of Japanese Society for Artificial Intelligence*, pp.21–26, Nov. (1999).

[43] N. Murata, S. Ikeda and A. Ziehe, "An approach to blind source separation based on temporal structure of speech signals," *Neurocomputing*, Vol.41, pp.1–24 (2001).

[44] 澤田宏, 向井良, 荒木章子, 牧野昭二, 猿渡洋, "周波数領域Blind Source Separationにおける帯域分割数の最適化," 電子情報通信学会技術報告, Vol.EA2000-95, pp.53–59 (2001).

[45] 澤田宏, 向井良, 荒木章子, 牧野昭二, "周波数領域ブラインド信号分離のための極座標表示に基づく活性化関数," 日本音響学会2001年秋季研究発表会講演論文集, pp.615–616, Oct. (2001).

[46] D. E. Schobben, Real-time adaptive concepts in acoustics, Kluwer Academic Publishers, Dordrecht (2001).

[47] S. Araki, S. Makino, R. Mukai, Y. Hinamoto, T. Nishikawa and H. Saruwatari, "Equivalence between frequency domain blind source separation and frequency domain adaptive beamforming," *Proc. ICASSP*, Vol.II, pp.1785–1788, May (2002).

[48] T. Nishikawa, H. Saruwatari and K. Shikano, "Blind source separation of acoustic signals based on multistage ICA combining frequency-domain ICA and time-domain ICA," *IEICE Trans. Fundamentals*, Vol.E86-A, No.4, pp.846–858 (2003).

[49] H. Sawada, R. Mukai, S. Araki and S. Makino, "Polar coordinate based nonlinear function for frequency-domain blind source separation," *IEICE Trans. Fundamentals*, Vol.E86-A, No.3, pp.590–596 (2003).

[50] S. Makino, H. Sawada, R. Mukai and S. Araki, "Blind source separation of convolutive mixtures of speech in frequency domain," *IEICE Trans. Fundamentals*, Vol.E88-A, No.7, pp.1640–1655 (2005).

[51] H. Saruwatari, H. Yamajo, T. Takatani, T. Nishikawa and K. Shikano, "Blind separation and deconvolution for convolutive mixture of speech combining SIMO-model-based ICA and multichannel inverse filtering," *IEICE Trans. Fundamentals*, Vol.E88-A, No.9, pp.2387–2400 (2005).

[52] 清水浩毅, 伊藤雅紀, 竹内義則, 松本哲也, 工藤博章, 大西昇, "指向性マイクロホンの近接配置における周波数領域ブラインド音源分離の性能評価," 電子情報通信学会論文誌 A, Vol.J89-A, No.6, pp.485–493 (2006).

[53] T. Kitamura and M. Akagi, "Speaker individualities in speech spectral envelopes," *Journal of the Acoustical Society of Japan (E)*, Vol.16, No.5, pp.283–289 (1995).

[54] D. O'shaughnessy, Speech communications, IEEE Press, New York (2000).

[55] A. J. Bell and T. J. Sejnowski, "The 'independent components' of natural scenes are edge filters," *Vision Research*, Vol.37, pp.3327–3338 (1997).

[56] A. Hyvärinen and P. O. Hoyer, "Emergence of phase and shift invariant features by decomposition of natural images into independent feature subspaces," *Neural Computation*, Vol.12, No.7, pp.1705–1720 (2000).

[57] B. A. Olshausen and D. J. Field, "Emergence of simple-cell receptive field properties by learning a sparse code for natural images," *Nature*, Vol.381, pp.607–609 (1996).

[58] B. A. Olshausen and D. J. Field, "Sparse coding with an overcomplete basis set: A strategy employed by V1?," *Vision Research*, Vol.37, pp.3311–3325 (1997).

[59] 熊原啓作, 行列・群・等質空間, 日本評論社, (2001).

[60] N. Suetake, Y. Nakamura and T. Yamakawa, "Maximum entropy ICA constrained by individual entropy maximization employing self-organizing maps," *Proc. Int. Joint Conf. on Neural Networks (IJCNN'99)*, In CD-ROM (1999).

[61] S. Fiori, "Hybrid independent component analysis by adaptive LUT activation function neurons," *Neural Networks*, Vol.15, pp.85–94, (2002).

[62] L. Xu, C.C.Cheung and S. Amari, "Learned parametric mixture based ICA algorithm," *Neurocomputing*, Vol.22, pp.69–80, (1998).

[63] A. Chen and P. Bickel, "Efficient independent component analysis (II)," *Technical Report 645, Department of Statistics, UC Berkeley*, Jan. (2003).

[64] Y. Tan, J. Wang and J. M. Zurada, "Nonlinear blind source separation using a radial basis function network," *IEEE Trans. on Neural Networks*, Vol.12, No.1, pp.124–134 (2001).

[65] S. Amari, "Natural gradient works efficiently in learning," *Neural Computation*, Vol.10, pp.251–276 (1998).

[66] W. H. Press, S. A. Teulolsky, B. P. Flannery and W. T. Vetterling, Numerical Recipes in C 日本語版, 技術評論社 (1993).

[67] 市田浩三, 吉本富士市, スプライン関数とその応用, 教育出版 (1979).

[68] 古井 貞熙, 音声情報処理, 森北出版 (1998).

[69] R. O. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. on Antennas and Propagation*, Vol.AP-34, pp.276–280 (1986).

[70] 菊間信良, アレーアンテナによる適応信号処理, 科学技術出版 (1998).

[71] F. Asano, K. Yamamoto, I. Hara, J. Ogata, T. Yoshimura, Y. Motomura, N. Ichimura and H. Asoh, "Detection and separation of speech event using audio and video information fusion and its application to robust speech interface," *EURASIP Journal on Applied Signal Processing*, Vol.11, pp.1727–1738 (2004).

[72] C. E. Speaks, Introduction to sound, Singular Publishing, San Diego (1999).

[73] A. Jourjine, S. Rickard and Ö. Yilmaz, "Blind separation of disjoint orthogonal signals: Demixing N sources from 2 mixtures," *Proc. of ICASSP*, Vol.5, pp.2985–2988, Istanbul, Turkey (2000).

[74] Ö. Yilmaz and S. Rickard, "Blind separation of speech mixtures via time-frequency masking," *IEEE Trans. on Signal Processing*, Vol.52, No.7, pp.1830–1847 (2004).

[75] 中迫昇, 小倉久直, 米森裕典, "複素 ICA の推定パラメタに基づく音源位置同定," 第１２回計測自動制御学会中国支部学術講演論文集, pp.236–237 (2003).

[76] J. Yamashita and Y. Hirai, "Blind source separation using orientation histograms in joint mixture distributions," *Proc. of Neural Networks and Computational Intelligence*, pp.152–157 (2004).

[77] 南茂夫, 科学計測のための波形データ処理, CQ 出版社 (1986).

[78] A. Gersho and R. M. Gray, Vector quantization and signal compression, Kluwer Academic Publishers, Boston (1992).

[79] M. Abe, S. Nakamura, K. Shikano and H. Kuwabara, "Voice conversion through vector quantization," *Journal of the Acoustical Society of Japan (E)*, Vol.11, No.2, pp.71–76 (1990).

[80] X. -H. Han, Z. Nakao, Y. -W. Chen, R. Kodama, "An ICA-domain shrinkage based Poisson-noise reduction algorithm and its application to penumbral imaging," *IEICE Trans. Information and Systems*, Vol.E88-D, No.4, pp.750–757 (2005).

[81] T. -W. Lee, M. S. Lewicki, M. Girolami and T. J. Sejnowski, "Blind source separation of more sources than mixtures using overcomplete representations," *IEEE Signal Processing Letter*, Vol.6, No.4, pp.87–90 (1999).

[82] M. S. Lewicki and T. J. Sejnowski, "Learning Overcomplete Representations," *Neural Computation*, Vol.12, No.2, pp.337–365 (2000).

[83] T. D. Sanger, "An optimality principle for unsupervised learning," *Advances in Neural Information Processing Systems* , Vol.1, D. Touretzky (Ed.), pp.11–19 (1989).

[84] 阪口豊, 樺島祥介, "脳内情報表現への情報理論的アプローチ," 脳の情報表現, pp.69–86, 朝倉書店 (2002).

[85] 小谷学, 白田康伸, 前川聡, 小澤誠一, 赤澤堅造, "スパース・コーディングによる音声の表現," 電気学会論文誌C, Vol.120-C, No.12, pp.1996–2002 (2000).