修士論文

動的モード分解を用いた音源分離の基礎的検討

宮崎 陽祐

主指導教員 鵜木 祐史

北陸先端科学技術大学院大学
先端科学技術研究科
（情報科学）

令和６年６月

Abstract

A single source usually produces a single sound. Based on this assumption, it is possible to separate sources by separating individual sounds from a signal waveform in which multiple sounds are mixed. This processing technology is called source separation. It is important to study methods to improve source separation performance and to enable separation under conditions that have been difficult to achieve in the past, because these methods can improve the performance of other acoustic signal processing applications, such as speech recognition. In this study, we focus on the separation of two sound sources. Therefore, we study techniques for separating individual acoustic signals from a mixture of two different acoustic signals.

The source separation strategy depends on the number of sources and the number of channels in the recorded sound. If the number of sources exceeds the number of channels, it is relatively easy to perform source separation. However, separation becomes more difficult if the number of sources is greater than the number of channels. Attempting source separation on a monaural sound is always difficult because the number of channels is always less than the number of sources. However, the monaural source separation method can be implemented with a single microphone, which provides a much greater degree of freedom in implementing the methods in real-world environments. Moreover, it is possible to extend the separation model that works on monaural sounds to multichannel sounds, or to integrate it with existing multichannel separation methods. Because of these potential applications, the study of source separation methods for monaural sounds is a very important topic because of its significant contribution to the overall sound source separation technology.

The separation of non-stationary signals is a long-standing problem in source separation of monaural signals. Research is still ongoing to solve this problem. Currently, the latest methods are based on Deep Neural Network (DNN), but these methods are limited by the large amount of training data required. On the other hand, the source separation method based on Principal Component Analysis (PCA) does not require any training. Therefore, PCA can be applied even in situations where sufficient training data is not available, making it an effective method in this regard. However, by its nature, PCA cannot capture local temporal changes in the amplitude and frequency of signals. Therefore, it is not suitable for analyzing non-stationary signals with such local changes. This is expected to hurt the source separation of non-stationary signals. Since the separation of non-stationary signals is an issue in the source separation of monaural signals, it is desirable to process sounds with a method suitable for the analysis of non-stationary signals.

This study focuses on a method called Dynamic Mode Decomposition (DMD), which, like PCA, decomposes signals into orthogonal components in a data-driven manner. In addition, DMD has a feature not found in PCA: the time evolution of the decomposed components can be observed as a parameter. This feature makes it possible to capture local changes in the amplitude and frequency of the sound. DMD has been used to analyze non-stationary signals in various research fields and has been applied in the field of acoustics. Therefore, it is expected that DMD can be effectively used in source separation tasks. The purpose of this study is to investigate the feasibility of source separation using DMD in the task of extracting only the target sound from a monaural sound. To this end, this study investigated how sounds are analyzed by DMD. Based on the results, this study designed a method of source separation using DMD and evaluated its separation performance.

The relationship between DMD and sounds was investigated, and the following three results were found: (1) The signal to be analyzed is represented by DMD as a linear sum of modes with unique frequencies and attenuation rates; (2) The distribution of time evolution features obtained by DMD analysis is significantly different between acoustic signals and noise; and (3) When a noise mixed signal is mode-decomposed by DMD, the distribution of time evolution features is divided into two groups, and it is possible to separate the sound by extracting only one of the groups. Based on the above findings, this study designed a source separation method using DMD by the following steps: (1) Mode decomposition and extraction of features related to the time evolution of the sounds/noise mixture by DMD; (2) Grouping the features estimated to correspond to the acoustic signal by focusing on the distribution of the features; (3) Synthesize the signal using only the modes associated with the features of grouping.

To investigate the feasibility of the DMD source separation method, DMD and PCA source separation methods were compared using computer simulations to compare the performance. As a result, the DMD source separation method showed higher performance than the PCA source separation method, indicating that DMD has good potential to be applied to source separation. However, it was also found that the low signal-to-noise ratio (SNR) of the noise mixture sounds distorts the features.

If this distortion of the features can be compensated, the performance of the DMD source separation method can be further improved. Therefore, it is necessary to investigate the relationship between noise level and the features distortion in the future. In addition, although this simulation investigated the separation of non-stationary artificial sound from stationary noise, it is necessary to further investigate whether this separation can also be performed for non-stationary

noise. For this purpose, it is necessary to investigate the relationship between non-stationary noise and DMD in the future.

目 次

図 目 次

表 目 次

第1章　序論

1.1　はじめに

我々の身の回りの環境では，様々な音源から音が発生している．我々はその音を聴取し，様々な情報を得て活用することで日々生活している．例として，会話によって生活に必要な情報を共有したり，個々の感情を他者へ伝達すること，周囲で発生する自動車の音や警笛を聴取し，自身が危険な状況に置かれないよう回避することが挙げられる．

こうした音を聴取するという営みの中で，我々ヒトは所望の音のみに注目して聞き取る能力を持っている [1][2]．この能力があることで，音を聴取することの目的を，精度高く達成することができる．例えば，会話するときには会話相手とのコミュニケーションに集中することができる．また，警笛など重要な報知音を正確に聞き取り，自身に危険が迫っているかの判断を的確に行うことができる．この優れた能力を計算機上で模擬するとすれば，それは図 1.1 のように，収録音から目的とする音響信号のみを如何にして分離するか，という仕組みを構築することになる．このような仕組みは音源分離と呼ばれる [3][4]．音源分離が実現することで，計算機上で稼働する様々な音響信号処理をより高精度にすることができる．例えば，我々が発話した音声の内容を計算機上で認識する技術を音声認識と呼ぶが，これを実現する際，音源分離技術によって収録音から音声以外の不要な音を取り除くことで，認識の精度を高めることができる [3]．さらに，音源分離技術の性能を高めたり，従来では難しかった条件下で分離を可能にする手法が開発されれば，音声認識といったその他の音響信号処理アプリケーションの性能をより高めることも可能だと考えられる．

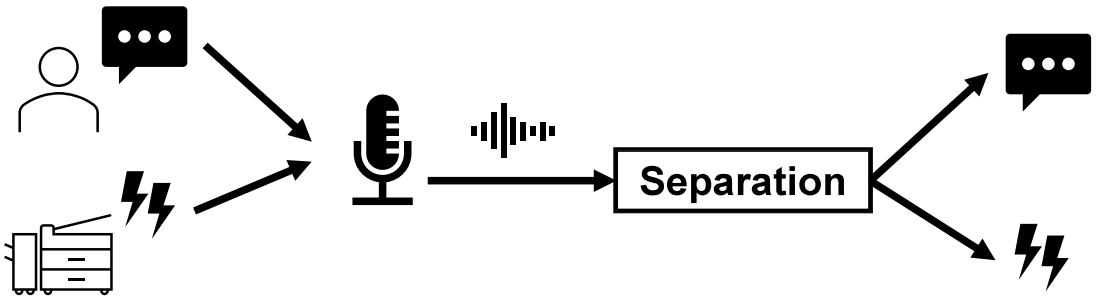

図 1.1: 音源分離のイメージ図.

1.2 研究背景

1つの音源からは，通常1つの音が発生している．このことを仮定したとき，複数の音が混ざり合った信号波形から一つ一つの音を分離すれば，音源どうしの分離も行えると考える．これが音源分離である．本研究では，2つの音源どうしを分離することを対象とする．したがって，2種類の音響信号が混合した信号から，個別の音響信号を分離するための技術について研究を行う．

音源分離を行うための技術は，想定する音源数と収録した音響信号のチャネル数によって取るべき方策が大きく異なる．音源数がチャネル数より少なければ，音源分離を行うこと比較的容易である．しかし音源数がチャネル数以上の場合，分離の難易度は高くなる [3]．さらに，音源の空間的な位置情報を利用するか，未知として分離を試みるか，という観点もある．一般に前者の方が，位置情報の条件に合致する場合は精度良い音源分離を行うことができるが，逆に条件から外れる場合の分離性能は期待できない．後者の技術は，特にブラインド音源分離と呼ばれる [5][6]．これは技術的難易度は高いが，幅広い条件下での適用が期待できる．

一方で，単一チャネル信号，則ちモノラル信号を対象にした音源分離法も様々研究・提案されている．モノラル信号上で音源分離を試みることは，常に対象音源数よりチャネル数が少ない条件での問題設定であり，分離の難易度は高くなる．また，複数チャネル信号の分離技術のように，音源の空間的情報を用いることはできないため，観測信号のみから如何にして特定音源由来の信号を推定するかということが重要となる [7]．これらの理由から，モノラル信号での音源分離は技術的難易度は高いと言えるが，マイクロホン一つあれば実現できるため，補聴器といった小型デバイスへの搭載など実環境への導入の自由度は格段に高くなるというメリットがある．その上，モノラル信号上で動作する分離モデルを複数チャネル信号用に拡張したり，既存の複数チャネル信号の音源分離技術と融合させることも可能である．これらの応用可能性から音源分離技術全体への貢献度は高く，非常に重要な研究テーマと言える．

モノラル信号の音源分離を行う上で，音声と非定常信号どうしの分離が難しく [7]，長年の課題である．これを解決するため，非定常な背景雑音を追従して推定できる方法 [8]，非定常信号の分析に適している手法を応用する提案 [9][10][11]，深層学習（DNN）を用いて，大量の学習データから音声といった非定常信号の分離モデルを学習させる方法 [12][13][14] など，この課題を解決するするために現在も研究が進められている．現在，最新の技術は DNN ベースの手法であるが，大量の学習用データが必要であるという制約がある．これに対し，主成分分析 (PCA) を用いた音源分離法 [4][15][16] は，そうした学習が不要のため，学習データが十分に用意できないような状況でも適用できるという点で非常に強力である．しかし PCA はその性質上，音響信号がもつ振幅や周波数の局所的な時間変化を捉えることができない．このため，そうした局所的な時間変化が起こり得る非定常信号の分析には適さない．よって，非定常信号の音源分離において悪影響をもたらすこ

とが予想される．非定常信号どうしの分離が課題とされているという背景がある中では，非定常信号の分析に適した手法で音響信号を処理することが望ましいと言える．

本研究では非定常信号の分析に適し，さらに音源分離に有効活用できる可能性をもつ手法として，動的モード分解（DMD）という手法に着目する [17][18][19][20]．DMD は流体解析の分野で提案された手法である [18]．元々流体解析では，複雑な流体データの中から PCA を用いて主要な構造を抽出するという解析方法が用いられてきた．しかし，PCA は局所的な時間変化を捉えることができないという問題があり，DMD はこの点をクリアする手法として提案・発展した背景がある [17][19]．DMD は PCA と同じく，行列データを決定論的に成分分解する解析手法である．しかし，DMD と PCA で異なるのは，分解成分が振動数や振動数といった時間発展に関するパラメータを持っていることである [20]．この特徴は，信号の性質が時々刻々と変化する非定常信号の分析に適応的であり，非定常信号の分析にも有利に働くと考えられる．DMD は，流体解析の分野に限らず，神経科学 [21]，金融 [22]，映像処理 [23]，疫学 [24], など多種多様な分野において利用されている．音響分野においては，機械音のパワースペクトルに対し DMD を適用し，異常音検知を試みた研究や [25]，音の発生による空気の粗密を光学的に観測する「空力音」に対し DMD を適用することでノイズ除去を行った研究がある [26]．その他，音響信号ではないが，単変量の時系列データに対する DMD の適用方法を紹介し，低周波のトレンド成分抽出の性能を検証した研究がある [27]．DMD は，神経科学分野における脳波や [21]，金融工学分野における株価の推移など [22]，非定常信号を分析するために様々な研究分野で用いられてきた実績があり，また音響分野における応用例もあることから，音源分離タスクにおいても活用が期待できる．

1.3 研究目的

本研究の目的は，DMD を用いた音源分離の実現可能性を検討することである．モノラル信号における音源分離において，PCA は学習データを用いずに収録データのみを用いて動作するという魅力がある．しかし，非定常信号どうしの分離が課題とされているにも拘らず，PCA は非定常信号の分析には適していない．この点を問題と考え，これを改善し得る方法として，本研究では DMD に着目する．DMD を用いた音源分離の実現可能性を検討するにあたり，DMD によってどのような音響信号の分析が可能であるかの調査を行う．その上で，そこで得られた知見を基に音源分離法を考案し，この分離性能を評価することでその実現可能性を検討する．

1.4 論文構成

本論文は6章構成である．図1.2に本論文の構成を示す．第1章では，本研究の研究背景および研究目的について述べた．第2章では，今日までの音源分離技術の概観をレビューし，今回の検討で比較対象とするPCAの音源分離法についても説明する．第3章では，本研究で着目するDMDについて，並びにどのような信号分析がなされるかを具体例と共に説明する．第4章では，DMDによる音響信号の分析の特徴をもとに音源分離法を考案し，その分離性能をコンピュータシミュレーションによって検討した結果を示す．5章で，得られたシミュレーション結果からDMDによる音源分離の可能性について考察を展開する．この結果と考察を踏まえ，第6章にて全体のまとめを述べ，残された課題を整理する．

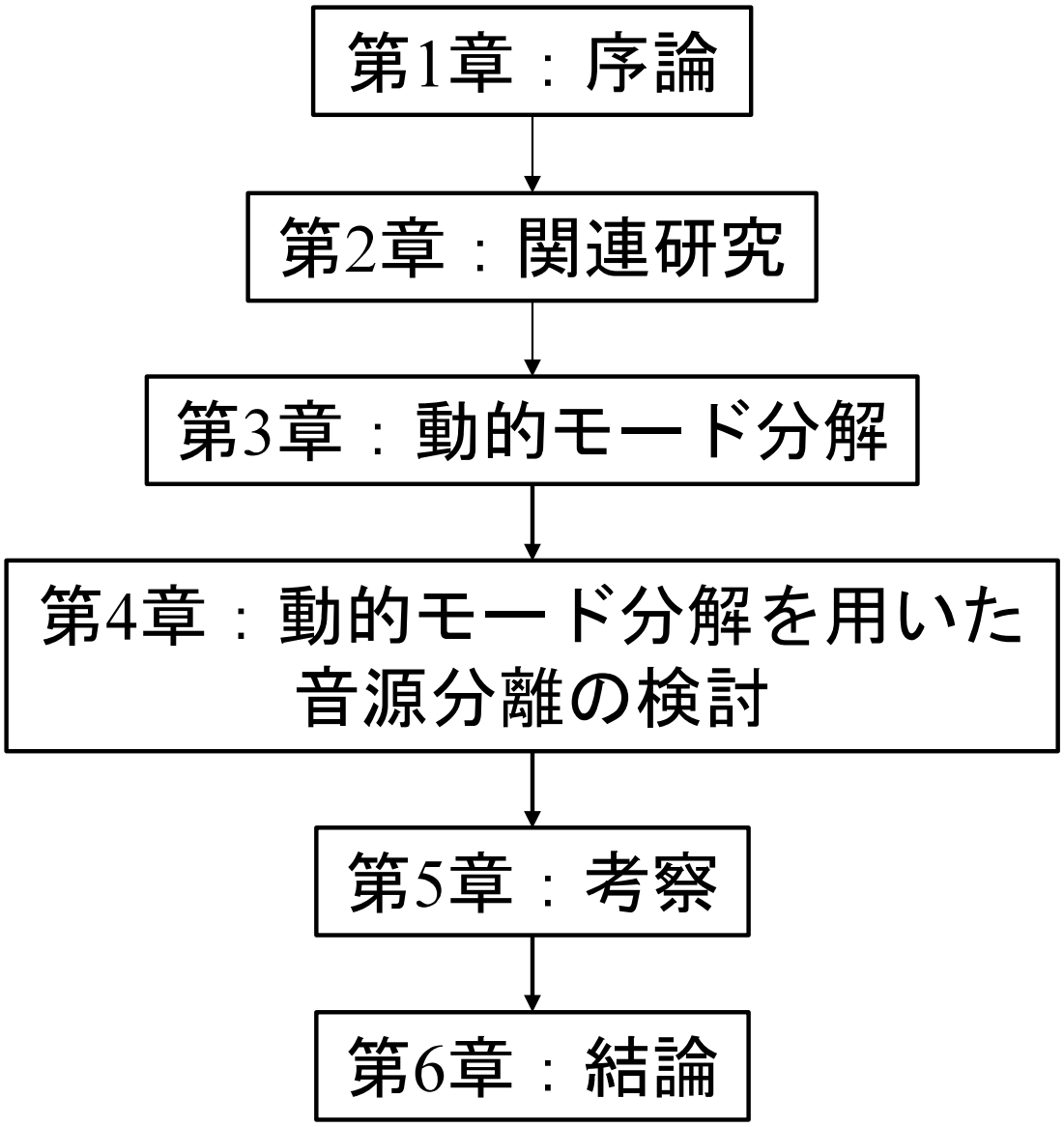

図 1.2: 本論文の構成.

1.5 記号の定義

本論文の内容に先立ち，使用する記号を予め表 1.1 のように定義する.

表 1.1: 本論文で使用する記号の定義.

記号	定義
t	時刻
$x(t)$	原信号
$\hat{x}(t)$	$x(t)$ を復元した信号
$\boldsymbol{R_x}$	$x(t)$ の自己相関行列
$\boldsymbol{\Lambda_p}$	$\boldsymbol{R_x}$ の固有値行列
$n(t)$	雑音
$\boldsymbol{R_n}$	$n(t)$ の自己相関行列
σ_n^2	$\boldsymbol{R_n}$ の分散
$y(t)$	雑音混合信号
$\boldsymbol{R_y}$	$y(t)$ の自己相関行列
$\boldsymbol{V}$	$y(t)$ の固有ベクトルを束ねた行列
$\boldsymbol{\psi_k}$	DMD へ入力する時系列データ
$\boldsymbol{A}$	$\boldsymbol{\psi_k}$ 間に仮定する変換行列
$\boldsymbol{\phi_j}$	モードの次元方向の重みベクトル
α_j	モードの振幅
$\phi_{1,j}$	モードの初期位相
$\boldsymbol{\Phi}$	$\boldsymbol{\phi_j}$ を束ねた行列
μ_j	モードの時間発展
$\boldsymbol{\mu}$	μ_j を対角成分に持つ対角行列
σ_j	モードの減衰率
ω_j	モードの角周波数
$\boldsymbol{\Psi}$	DMD への入力行列
$\boldsymbol{\Psi_0}$	$\boldsymbol{\Psi}$ の部分行列
$\boldsymbol{\Psi_1}$	$\boldsymbol{\Psi}$ の部分行列
$\boldsymbol{U}$	$\boldsymbol{\Psi_0}$ の左特異ベクトルの行列
$\boldsymbol{W}$	$\boldsymbol{\Psi_0}$ の右特異ベクトルの行列
$\boldsymbol{\Sigma}$	$\boldsymbol{\Psi_0}$ の特異値行列
$\boldsymbol{F}$	$\boldsymbol{A}$ を基底変換した行列
$S(\omega)$	$x(t)$ のパワースペクトル
$\hat{S}(\omega)$	$\hat{x}(t)$ のパワースペクトル
τ	シミュレーションで用いる信号の減衰率を操作する時定数

第2章　関連研究

2.1　音源分離技術の概観

音源分離は長く研究され続けており，従来より様々な手法が提案されてきた．単一チャネル信号（モノラル信号）を扱うか複数チャネル信号を扱うかで技術体系が大きくことなることから，この 2 つに分けて概観を述べる．

2.1.1　複数チャネル信号の音源分離

複数チャネル信号を対象とした音源分離法について，元は無線通信の分野で用いられてきたビームフォーミング技術を利用した音源分離法がまず発展した [4]．これは収録用マイクロホンを中心とした空間にバンドパスフィルタを形成し，特定方向から到来する音響信号を推定することで特定音源を分離するものである．最も早くから存在するのは固定ビームフォーマと呼ばれる技術で，マイクロホンに対し特定の角度から到来する信号を強める遅延和ビームフォーマや，逆に特定の角度から到来する信号を受け付けない死角を形成することで目的音の分離を行うヌルビームフォーマが有名である [4]．これら固定ビームフォーマの技術は，運用する音空間が異なったり，マイクロホンの性能が異なったりすると分離性能に影響するため，この点が実用上の制約であった．この後，ウィーナーフィルタをビームフォーミングに応用した適応ビームフォーマが発展した [4]．これは，マイクロホンからの観測信号を用いることで特定方向以外から到来する音響信号の影響を最小化するような空間フィルタを形成するものである．空間情報に加え観測信号の情報も使うことで，運用する音空間やマイクロホンの条件に応じて適応的なフィルタを形成することができ，固定ビームフォーマの難点をクリアし得るものである．

これらビームフォーミング技術を用いた音源分離法は現在でも利用され続けているものであるが，いずれも，マイクロホンに対し目的音源がどこから到来してくるかの情報が必要となる．例えばマイクロホンがヒトやロボットのような移動するものに搭載されている場合，マイクロホンに対する音源位置が変化し得るため，この場合に適用するには課題があった．これに対し，マイクロホン位置や音源位置を用いずに音源分離を行う技術の研究が進み，これらは特にブラインド音源分離と呼ばれる [6]．その嚆矢となったのは，1990 年代に提案された独立成分分析（ICA）である [28][5][6]．これは，異なる音源から発生した音響信号は互いに独立であるという仮定を置くことで，収録した混合信号をチャネル別に独立性を最

大化することで信号を分離する．この技術を基にして，周波数帯域別にICAを行う周波数領域ICA（FDICA）が導入されたことを皮切りに[28]，音響信号処理分野でのブラインド音源分離の研究は活発化した．ブラインド音源分離を実現する上でFDICAは画期的な技術であるが，いくつか解決が難しい課題も抱えている．観測信号を信号要素へ分離できたとしても，その信号要素と音源との対応関係は自明でなく，ラベル付けのための技術が必要であり，これはパーミュテーション問題と呼ばれている[5]．これを解決するため，周波数帯域の情報をベクトルとして扱うよう拡張した独立ベクトル分析（IVA）が提案されるなど[29]，改善提案も様々為されてきた．近年では深層学習と融合させた手法も提案されるなど[30]，現在においても研究が進められている．

2.1.2　モノラル信号の音源分離

モノラル信号の音源分離はマイクロホン数が1つに対し分離対象は2つ以上となるため，常に劣決定の問題設定であり，複数チャネル信号の音源分離より技術的難易度は高いと言える．また，モノラル信号の音源分離では，複数チャネル信号での音源分離におけるビームフォーミングのように音源の空間的情報を用いることはできないため，観測信号のみから如何にして特定音源を分離するための情報を掴み取るか，ということが重要である．早くから存在する手法としては，スペクトルサブトラクション法が有名である[31]．これは観測信号から雑音のスペクトルを何らかの方法で推定し（目的音が発生していない時間帯の信号情報を用いるなど），これを観測信号のスペクトルから減算するというものである．これは雑音を低減するフィルタの形成方法の一手法とも言え，同じく雑音抑圧フィルタを形成する手法であるウィーナーフィルタ[32]や，カルマンフィルタ[33]も，音源分離のための技術として応用されてきた．こうした雑音抑圧フィルタ推定による音源分離技術において，抑圧しきれずに残ってしまった雑音成分はミュージカルノイズと呼ばれ，これが不快感を催す音であるため問題視されている[3]．また，雑音が非定常であると，雑音の性質が時々刻々と変化するため，これに追従できるような雑音推定法が必要とされ，提案されてきた[8]．

この後，1990年代に入り，汎用コンピュータの性能が向上したことで，より大規模な代数計算を用いた信号分析が可能となった．これによって，信号を複数の信号成分に分解することで分析・分離を行う手法が提案されるようになった．代表的にはPCAの利用があり[15]，これについては次節で説明する．他には経験的モード分解（EMD）という手法を利用する分離法も提案された[9][10][11]．経験的モード分解は非定常信号の分析に適しており[9]，この性質を非定常信号である音声の抽出や性質改善に役立てようとしたものである．汎用コンピュータの性能が向上したことにより，もう一つ技術潮流が生まれた．それは非負値行列因子分解（NMF）の利用である[34][35][36]．NMFは画像処理分野で生まれた技術であり，これを観測信号の振幅スペクトログラムに適用する形でモノラル音源分離に

導入された [37]．NMF は，1 つの非負値行列をより低次元な 2 つの非負値行列へ分解するという手法である．これを振幅スペクトログラムに適用すると，短時間スペクトルのパターンとその時間的な発生タイミングの系列に分解できる．これにより，特定音源のスペクトルパターンのモデル化が可能となる．楽器音等は低ランクの行列モデルとして良く近似できることから，楽曲から楽器パートを分離したり，歌声を分離する技術として盛んに用いられてきた [36]．NMF は非負値行列を扱うため，本来は複素行列であるスペクトログラムに含まれる位相情報を原理的に扱えない．これを解決するため，複素スペクトログラムを扱えるようモデルを拡張した複素 NMF が提案されるなど [38]，スペクトログラムの位相情報を如何にして扱うかが NMF の課題とされている．なお，複数チャネル信号の技術であるが，上述の IVA と組み合わせた独立低ランク行列分析（ILRMA）という手法が提案されるなど [39]，NMF は近年でも他手法との融合が模索され続けている．

この後，汎用コンピュータの計算速度は益々向上し，さらにインターネットの利用拡大によって大規模なデータへのアクセスも容易になったことから，大規模な計算機資源と大量の学習データを要する深層学習（DNN）の利用ハードルが低下した．この潮流は 2010 年代にモノラル音源分離にも訪れた．アプローチとしては，スペクトログラムを入力として，特定の音源のみを抽出するための時間周波数マスクを学習する方法の研究が進み [14]，そのブレイクスルーとして有名なのが Deep Clustering という手法である [12]．これ以降，モノラル音源分離の研究では DNN ベースの手法提案が主流となっていった．その後，スペクトログラム入力型の学習モデルを踏み台としつつ，波形信号そのものから分離のための特徴量を学習する Time-domain Audio Separation Network（TasNet）が提案され [13]，以降は波形入力型の DNN ベース音源分離手法も流行した．現在，音源分離技術の性能を競うコンペティションとして Deep Noise Suppression Challenge（DNS Challenge）が最新で [40]，その名に冠する通り DNN が用いられることを前提として学習用データとテストデータが提供されている．このことからも，モノラル音源分離の分野では DNN ベースの手法が先端的であると言える．

2.2 音源分離技術が抱える課題

以上に述べた音源分離手法について，表 2.1 に年代別にまとめた．これら音源分離技術全体を通じて，異なる話者の音声の分離といった非定常信号どうしの分離の難しさが課題としてある [7]．モノラル信号の音源分離の分野においては，スペクトルサブトラクション法といった雑音抑圧フィルタの手法に始まり，音声と定常雑音との分離においては実用レベルの性能がある．しかし，雑音が非定常の場合や，音声どうしの分離には適用が難しかった．NMF の導入により，楽器音と歌声の分離は可能になったものの，音源モデルから外れた場合の分離は依然として課題である．近年は DNN を用いて大量の学習データを用いることで，この課題に立ち向かおうとしてる最中である．この課題から今すぐ解き放たれるためには，複数

マイクロホンを用いるビームフォーミングを用いることも一つの手である．ビームフォーミングにおいては音源の到来位置によって音源を分けるため，音響信号が定常か非定常かはそもそも大きな問題にならないからである．しかし，ビームフォーミングは導入のための制約が大きい技術である．これを解決するために発展したのが ICA を始めとするブラインド音源分離技術であった．とはいえ，ICA においては信号どうしの独立性を利用することから，非定常信号どうしではその独立性の仮定がうまく成り立たない場合も多い．このように，非定常信号の分離が課題であり続けるために，今日まで，長年音源分離の研究が続けられている．

表 2.1: 音源分離技術の概観.

年代	モノラル信号を対象	複数チャネル信号を対象
1940 年〜1990 年	スペクトルサブトラクション法 ウィーナーフィルタ カルマンフィルタ	遅延和ビームフォーマ ヌルビームフォーマ 適応ビームフォーマ
1990 年〜2000 年	PCA（部分空間法） EMD	PCA（MUSIC 法） ICA FDICA
2000 年〜2010 年	NMF 複素 NMF	IVA
2010 年以降	Deep Clustering TasNet	ILRMA

2.3　主成分分析を用いた音源分離法

本研究の目的は，DMD を用いた音源分離の実現可能性を検討することであり，検討を進めるフィールドはモノラル信号上の目的音・妨害音の分離とする．その上で，本研究では PCA を検討のための比較対象とした．理由は 2 つある．一つは，PCA がモノラル信号上の音源分離法として既に応用されていることである．もう一つは，DMD と PCA とで類似する点が多く，DMD を用いた音源分離法は PCA と類似した仕組みで設計することが可能であると考えられ，比較対象として適していると考えられることである．本節では，PCA の音源分離法について述べる．

PCA を用いることで，入力を互いに直交する主成分ごとに分解することが可能である．これをモノラル信号に適用した場合，直交成分は固有の周波数を持つ正弦波成分へと分解されることになる．これを利用し，入力に含まれる際立った周波数成分を推定することで音源分離のための周波数フィルタを形成することができ，この手法は MUSIC（MUltiple SIgnal Classification）法と呼ばれる [4][41]．また，音響信号と雑音との分離問題を考えたとき，音響信号と雑音とは基本的に無相関であると考えられる．この仮定の下，信号に PCA を適用することで信号部分空間と雑音部分空間との切り分けが可能である．この仕組みを用いた音源分離法は部分空間法とも呼ばれる [4][15][16]．MUSIC 法も部分空間法も，複数チャネル信号の音源分離にも活用できる手法であるが [4]，モノラル信号を対象とした部分空間法の端緒は Ephraim と Trees の研究 [15] である．

信号音を $x(t)$，対する雑音を $n(t)$ とし，混合信号 $y(t)$ が以下の式 2.1 で与えられるとする．

$$y(t) = x(t) + n(t) \tag{2.1}$$

このとき，$x(t)$ の自己相関行列を $\boldsymbol{R_x}$，$n(t)$ の自己相関行列を $\boldsymbol{R_n}$，$y(t)$ の自己相関行列を $\boldsymbol{R_y}$ としたとき，以下の式 2.2 のように自己相関行列に変換しても線形性が成り立つ．

$$\boldsymbol{R_y} = \boldsymbol{R_x} + \boldsymbol{R_n} \tag{2.2}$$

ここで，$\boldsymbol{R_x}$，$\boldsymbol{R_n}$ が以下の式 2.3 のように固有値分解されるとする．

$$\begin{aligned} \boldsymbol{R_x} &= \boldsymbol{V} \begin{bmatrix} \boldsymbol{\Lambda_p} & 0 \\ 0 & 0 \end{bmatrix} \boldsymbol{V^T} \\ \boldsymbol{R_n} &= \boldsymbol{V}(\sigma_n^2 \boldsymbol{I}) \boldsymbol{V^T} \end{aligned} \tag{2.3}$$

ただし，$\boldsymbol{V}$ は固有ベクトルを束ねた直交行列，$\boldsymbol{\Lambda_p}$ は $\boldsymbol{R_x}$ の固有値を対角成分にもつ対角行列，σ_n^2 は $\boldsymbol{R_n}$ の分散である．このとき，時間的相関の少ない雑音のエネルギーが観測空間全体に広がる一方，目的とする信号成分のエネルギーはより低ランクの部分空間に集中するという性質に基づくことで，$\boldsymbol{R_y}$ は以下の式 2.4 のよ

うに表現できる [16].

$$\boldsymbol{R_y} = [\boldsymbol{V_p}\boldsymbol{V_{q-p}}]\left(\begin{bmatrix}\boldsymbol{\Lambda_p} & 0\\ 0 & 0\end{bmatrix} + \sigma_n^2\boldsymbol{I}\right)\left[[\boldsymbol{V_p}\boldsymbol{V_{q-p}}]^T\right] \tag{2.4}$$

ただし，$\boldsymbol{V_p}$ は $\boldsymbol{V}$ の 1 から p 列目までの固有ベクトルを束ねた行列，$\boldsymbol{V_{q-p}}$ は $\boldsymbol{V}$ の p から q 列目までの固有ベクトルを束ねた行列を表す．すなわち，$\boldsymbol{R_y}$ の固有値の内固有値と対応する固有ベクトルを 1 から p 個選択することで，信号 $x(t)$ に対応する $\boldsymbol{R_x}$ の分離が可能となる．

PCA は，与えられたデータから特徴的な構造を自動的に分析・抽出するデータ駆動型の手法である．DNN を用いる手法は学習データが必要となる一方，PCA であれば学習データが利用できない状況でも音源分離に活用することができる．このデータ駆動型の手法であることが，音源分離における PCA の強みと考える．しかし，PCA を音響信号のような時系列信号に適用する場合，音響信号のもつ特徴を統計的に扱うことで，局所的な変動の特徴を捉えることができなくなる．この点は，音声といった非定常信号を分析する上で問題である．このため，PCA による音源分離は入力を信号フレームに分割して各々に適用することが前提となる [16].

第3章　動的モード分解

3.1　概要

本研究では，非定常信号の分析に適し，さらに音源分離に有効活用できる可能性をもつ手法として，DMD に着目する．本章で，DMD の概要の説明と，音源分離へ向けた信号分析の議論を展開する．

DMD とは，多次元の時系列データを対象に，次元方向の情報と時間方向の情報を持ち合わせた成分に分解する手法である [20]．ある時刻 k の状態がベクトル $\boldsymbol{\psi_k}$ で表現されるような時系列データを対象としたとき，DMD によって $\boldsymbol{\psi_k}$ は以下の様な分解が行われる．

$$\boldsymbol{\psi_k} = \sum_{j=1}^{r} \alpha_j \boldsymbol{\phi_j} e^{(\sigma_j + i\omega_j)k} \tag{3.1}$$

式 3.1 における分解成分 $\alpha_j \boldsymbol{\phi_j} e^{(\sigma_j + i\omega_j)k}$ を本研究でモードと呼称する．すなわち，DMD によって時系列データ $\boldsymbol{\psi_k}$ を r 個のモードの線形和として情報表現することが可能である．α_j はモードの係数，$\boldsymbol{\phi_j}$ はモードの次元方向の重みベクトルである．また $e^{(\sigma_j + i\omega_j)k}$ はモードの時間発展を表現しており，σ_j はモードの減衰率，ω_j はモードの角周波数を表す．この σ_j と ω_j を，本論文では時間発展特徴と呼称する．

3.2　モード分解の原理

DMD では，分析対象のデータ $\boldsymbol{\psi_k}$ に以下の様な関係性が成り立つと仮定する．

$$\boldsymbol{\psi_{k+1}} = \boldsymbol{A}\boldsymbol{\psi_k} \tag{3.2}$$

すべての k で $\boldsymbol{A}$ は一定とする．この $\boldsymbol{A}$ を固有値分解し，固有値行列を $\boldsymbol{\mu}$，固有ベクトルを束ねた行列を $\boldsymbol{\Phi}$ としたとき，$\boldsymbol{\psi_k}$ の解は以下の式で得られる．

$$\boldsymbol{\psi_k} = \boldsymbol{\Phi}\boldsymbol{\mu}^k\boldsymbol{\Phi}^\dagger\boldsymbol{\psi_0} \tag{3.3}$$

ただし，$\cdot^\dagger$ は疑似逆行列を表す．この $\boldsymbol{\mu}$ と $\boldsymbol{\Phi}$ を推定することが，DMD の計算上の目標となる．ここで，$\boldsymbol{\Phi}$ の各列ベクトル，則ち固有ベクトルが式 3.1 の $\boldsymbol{\phi_j}$ に対応する．また，$\boldsymbol{\Phi}^\dagger\boldsymbol{\psi_0}$ の演算によって初期値ベクトルが得られるが，この各要素が係数 α_j と対応する．そして，$\boldsymbol{\psi_0}$ から k ステップ先の時間的変化は $\boldsymbol{\mu}^k$ で表現さ

れることとなり，ここに着目することで $\boldsymbol{\psi_k}$ が持つ時間発展の特徴をパラメータとして捉えることができる．各固有値 $\mu_j = diag(\boldsymbol{\mu})$ として，時間発展特徴との対応は $\mu_j = e^{(\sigma_j + i\omega_j)}$ となっている．したがって，σ_j と ω_j は以下のように算出される．

$$\sigma_j = Re\left[\frac{\mu_j}{\Delta t}\right], \quad \omega_i = Im\left[\frac{\mu_j}{\Delta t}\right] \tag{3.4}$$

ただし，$Re\left[\cdot\right]$ は括弧内の実部，$Im\left[\cdot\right]$ は括弧内の虚部を表す．また，Δt は入力のサンプル間時間幅，則ちサンプリング周期である．

3.3 アルゴリズム

本節で，DMD による信号の分析合成のアルゴリズムを説明する．サンプル数 $N+1$ の時系列信号 $x(t) = \{x(t_0), x(t_1), \cdots, x(t_N)\}$ を DMD で分析するとき，次のような行列 $\boldsymbol{\Psi}(\in \mathbb{R}^{h\times(N-h+1)})$ を作成する．

$$\boldsymbol{\Psi} = \begin{bmatrix} x(t_0) & x(t_1) & \cdots & x(t_{T-h+1}) \\ x(t_1) & x(t_2) & \cdots & x(t_{T-h+2}) \\ \vdots & \vdots & \ddots & \vdots \\ x(t_{h-1}) & x(t_h) & \cdots & x(t_N) \end{bmatrix} \tag{3.5}$$

ただし，h は $2 \leq h \leq \frac{N}{2}$ の自然数で，分析者が予め決定するパラメータである．ここで，$\boldsymbol{\Psi}$ を以下の式 3.6 の様な 2 つの行列 $\boldsymbol{\Psi_0}, \boldsymbol{\Psi_1}(\in \mathbb{R}^{h\times(N-h)})$ に分ける．ただし，$\boldsymbol{\Psi}$ の l 列目ベクトルを $\boldsymbol{\psi_l} = \begin{bmatrix} x(t_l) & x(t_{l+1}) & \ldots & x(t_{l+h-1}) \end{bmatrix}^T$ と表す．

$$\boldsymbol{\Psi_0} = \begin{bmatrix} | & | & \cdots & | \\ \boldsymbol{\psi_0} & \boldsymbol{\psi_1} & \cdots & \boldsymbol{\psi_{T-h}} \\ | & | & \cdots & | \end{bmatrix}, \boldsymbol{\Psi_1} = \begin{bmatrix} | & | & \cdots & | \\ \boldsymbol{\psi_1} & \boldsymbol{\psi_2} & \cdots & \boldsymbol{\psi_{T-h+1}} \\ | & | & \cdots & | \end{bmatrix} \tag{3.6}$$

上記の $\boldsymbol{\Psi_0}, \boldsymbol{\Psi_1}$ の間に，以下の関係が成り立つとする．

$$\boldsymbol{\Psi_1} = \boldsymbol{A}\boldsymbol{\Psi_0} \tag{3.7}$$

この $\boldsymbol{A}$ は式 3.2 のものと同じである．$\boldsymbol{A}$ を推定することで，DMD の計算目標である固有値と固有ベクトルを求められる．$\boldsymbol{A}$ の推定値としては，例えば，$\boldsymbol{\Psi_0}$ の疑似逆行列 $\boldsymbol{\Psi_0^\dagger}$ を計算し，$\boldsymbol{A} \approx \boldsymbol{\Psi_1}\boldsymbol{\Psi_0^\dagger}$ とすることで得られる．しかし，$\boldsymbol{\Psi_0}, \boldsymbol{\Psi_1}$ は一般にはフルランク行列でなく，したがって $\boldsymbol{A}$ もフルランクとならず，$\boldsymbol{A}$ の固有値分解を直接計算することは難しい．そのため，DMD においては $\boldsymbol{A}$ のランク数を予め考慮した上で固有値と固有ベクトルを求めることを考える．その一般的な方法は，特異値分解 (SVD) を用いたアプローチである．

まず，$\boldsymbol{\Psi_0}$ の SVD を計算する．

$$\boldsymbol{\Psi_0} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{W}^* \tag{3.8}$$

ただし，$\boldsymbol{U} \in \mathbb{C}^{h\times r}, \boldsymbol{W} \in \mathbb{C}^{r\times(N-h)}$ は正規直交行列，$\boldsymbol{\Sigma} \in \mathbb{C}^{r\times r}$ は特異値行列である．なお，$\cdot^*$ は随伴行列を表す．ここで，$\boldsymbol{A}$ を $\boldsymbol{U}$ で基底変換した行列を $\boldsymbol{F} \in \mathbb{C}^{r\times r}$ とおく．

$$\boldsymbol{F} := \boldsymbol{U}^*\boldsymbol{A}\boldsymbol{U} \tag{3.9}$$

式 3.9 を変形することで次式が得られる．

$$\boldsymbol{A} = \boldsymbol{U}\boldsymbol{F}\boldsymbol{U}^* \tag{3.10}$$

式 3.7 に式 3.8 及び式 3.10 を代入する．

$$\begin{aligned}\boldsymbol{\Psi_1} &= \boldsymbol{A}\boldsymbol{\Psi_0} \\ &= \boldsymbol{U}\boldsymbol{F}\boldsymbol{U}^*\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{W}^* \\ &= \boldsymbol{U}\boldsymbol{F}\boldsymbol{\Sigma}\boldsymbol{W}^*\end{aligned} \tag{3.11}$$

したがって，$\boldsymbol{F}$ は次式で推定される．

$$\hat{\boldsymbol{F}} = \boldsymbol{U}^*\boldsymbol{\Psi_1}\boldsymbol{W}\boldsymbol{\Sigma^{-1}} \tag{3.12}$$

この $\hat{\boldsymbol{F}}$ はランク数 r のフルランク行列であり，固有値分解が可能である．

$$\hat{\boldsymbol{F}}\boldsymbol{\Phi_F} = \boldsymbol{\Phi_F}\boldsymbol{\mu_F} \tag{3.13}$$

このとき，$\hat{\boldsymbol{F}}$ の固有値 $\mu_{Fj} \in \mathbb{C}$ は $\boldsymbol{\mu_F}$ の対角成分 $diag(\boldsymbol{\mu_F}) = \{\mu_{F1}, \mu_{F2}, \cdots, \mu_{Fr}\}$，固有ベクトル $\boldsymbol{\phi_{Fj}} \in \mathbb{C}^{\boldsymbol{h}}$ は $\boldsymbol{\Phi_F}$ の列ベクトル $\boldsymbol{\Phi_F} = \begin{bmatrix}\boldsymbol{\phi_{F1}} & \boldsymbol{\phi_{F2}} & \cdots & \boldsymbol{\phi_{Fr}}\end{bmatrix} (1 \leq j \leq r)$ となる．

さて，DMD によって最終的に求める $\boldsymbol{A}$ の固有値を $\mu_j \in \mathbb{C}$，固有ベクトルを $\boldsymbol{\phi_j} \in \mathbb{C}^h$ とする．まず μ_j について，元々 $\boldsymbol{F}$ と $\boldsymbol{A}$ は基底ベクトルを変換した関係にあり，従って互いの固有値は等しく，$\mu_j = \mu_{Fj}$ である．μ_j を対角成分にもつ対角行列を $\boldsymbol{\mu}$ とするなら，以下の式 3.14 となる．

$$\boldsymbol{\mu} = \boldsymbol{\mu_{Fj}} \tag{3.14}$$

ここで，式 3.13 は，式 3.12 と式 3.14 を代入して以下のように変形できる．

$$\begin{aligned}&\hat{\boldsymbol{F}}\boldsymbol{\Phi_F} = \boldsymbol{\Phi_F}\boldsymbol{\mu_F} \\ \Rightarrow\ &\boldsymbol{U}^*\boldsymbol{A}\boldsymbol{U}\boldsymbol{\Phi_F} = \boldsymbol{\Phi_F}\boldsymbol{\mu} \\ \Rightarrow\ &\boldsymbol{A}\boldsymbol{U}\boldsymbol{\Phi_F} = \boldsymbol{U}\boldsymbol{\Phi_F}\boldsymbol{\mu}\end{aligned} \tag{3.15}$$

したがって，$\boldsymbol{A}$ の対角化に要する行列は $\boldsymbol{U}\boldsymbol{\Phi_F}$ となる．これが式 3.3 における $\boldsymbol{\Phi}$ にあたり，求めるべき固有ベクトル $\boldsymbol{\phi_j}$ は $\boldsymbol{\Phi}$ の各列ベクトルである．

$$\begin{bmatrix} | & | & & | \\ \boldsymbol{\phi_1} & \boldsymbol{\phi_2} & \cdots & \boldsymbol{\phi_r} \\ | & | & & | \end{bmatrix} = \boldsymbol{\Phi} = \boldsymbol{U}\boldsymbol{\mu_F} \tag{3.16}$$

以上の $\mu_j, \boldsymbol{\phi_j}$ が，DMD の出力である．

上記の $\boldsymbol{\phi_j}$ の 1 行目要素を $\phi_{1,j}$ とおいたとき，$\phi_{1,j}$，μ_j 及び係数 α_j を用いることで，入力 $x(t)$ を再構成することが可能である．このとき，α_j はモードの振幅，$\phi_{1,j}$ はモードの初期位相を表す係数である．

$$x(t_n) = \sum_{j=1}^{r} \alpha_j \phi_{1,j} \mu_j^n \tag{3.17}$$

なお，α_j は次式で算出される．

$$\begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_r \end{bmatrix} = \boldsymbol{\Phi}^{\dagger} \boldsymbol{\psi_0} \tag{3.18}$$

3.4 信号の分析例

3.4.1 減衰する純音の分析例

本節では，DMD へ音響信号を入力したとき，どのような時間発展特徴が得られるのかを，いくつかの信号例から説明する．

まず，以下の式 3.19 で表されるような，減衰する 100 Hz の純音 $x_1(t)$ を考える．

$$x_1(t) = e^{-10t} \cos(2\pi \cdot 100t) \tag{3.19}$$

ただし，信号のサンプリング周波数は 16000 Hz とした．この $x_1(t)$ から得られた時間発展特徴の減衰率 σ_j を横軸，周波数 $\frac{\omega_j}{2\pi}$ を縦軸にとった分布図を図 3.1 に示す．このとき，式 3.5 のように入力を行列に変形する際の行数 $h = 2$ とした．プロット点横の添え字はモードの番号 j である．図 3.1 の様に，1 つの周波数成分に対し，DMD によって 2 つの時間発展特徴が得られる．モードの周波数は $\frac{\omega_j}{2\pi}$ によって，負の周波数も含めた形で取得できる．また，周波数成分が減衰する場合，その減衰率は σ_j によって取得できる．モノラル信号のような 1 次元の入力に対し，DMD への入力に用いる行列の行数は少なくとも 2 以上が必要と言える．なぜなら，1 つの周波数成分に対しモードは 2 つ抽出されるが，モード抽出数の上限は行数 h となるためである．

DMD によるモード分解の後，問題なく信号再合成が行えるかどうかを，原信号と再合成信号との信号対誤差比（SER : Signal to Error Ratio）[42] と，対数スペクトル距離（LSD : Log Spectral Distance）[43] で評価した．原信号を $x(t)$，再合成信号を $\hat{x}(t)$ としたとき，SER を以下の式 3.20 で算出する．単位は [dB] で，値が大きいほど信号間の誤差は小さいと評価できる．

$$SER\{x(t), \hat{x}(t)\} = 10 \log_{10} \frac{\sum_{n=0}^{N} x^2(t_n)}{\sum_{n=0}^{N} \{x(t_n) - \hat{x}(t_n)\}^2} \tag{3.20}$$

LSD は，$x(t)$ のパワースペクトルを $S(\omega)$，$\hat{x}(t)$ のパワースペクトルを $\hat{S}(\omega)$ とおいたとき，以下の式 3.21 で算出する．単位は [dB] で，値が 0 に近いほど信号間の誤差は小さいと評価できる．

$$LSD\{S(\omega), \hat{S}(\omega)\} = \sqrt{\frac{1}{N} \sum_{n=0}^{N} \Big[10 \log_{10}\{S(\omega_n)\} - 10 \log_{10}\{\hat{S}(\omega_n)\}\Big]} \tag{3.21}$$

この結果，SER は 220.54 dB，LSD は 1.0742×10^{-7} dB であり，原信号と再合成信号との誤差は非常に小さいと言える．すなわち，DMD によって原信号の情報をほぼ損なうことなく分析・合成が可能だと言える．図 3.2 に，原信号 $x_1(t)$，再合成信号 $\hat{x}_1(t)$ の波形を表示する．図 3.2 から，波形としても問題なく再合成が行えていることが分かる．

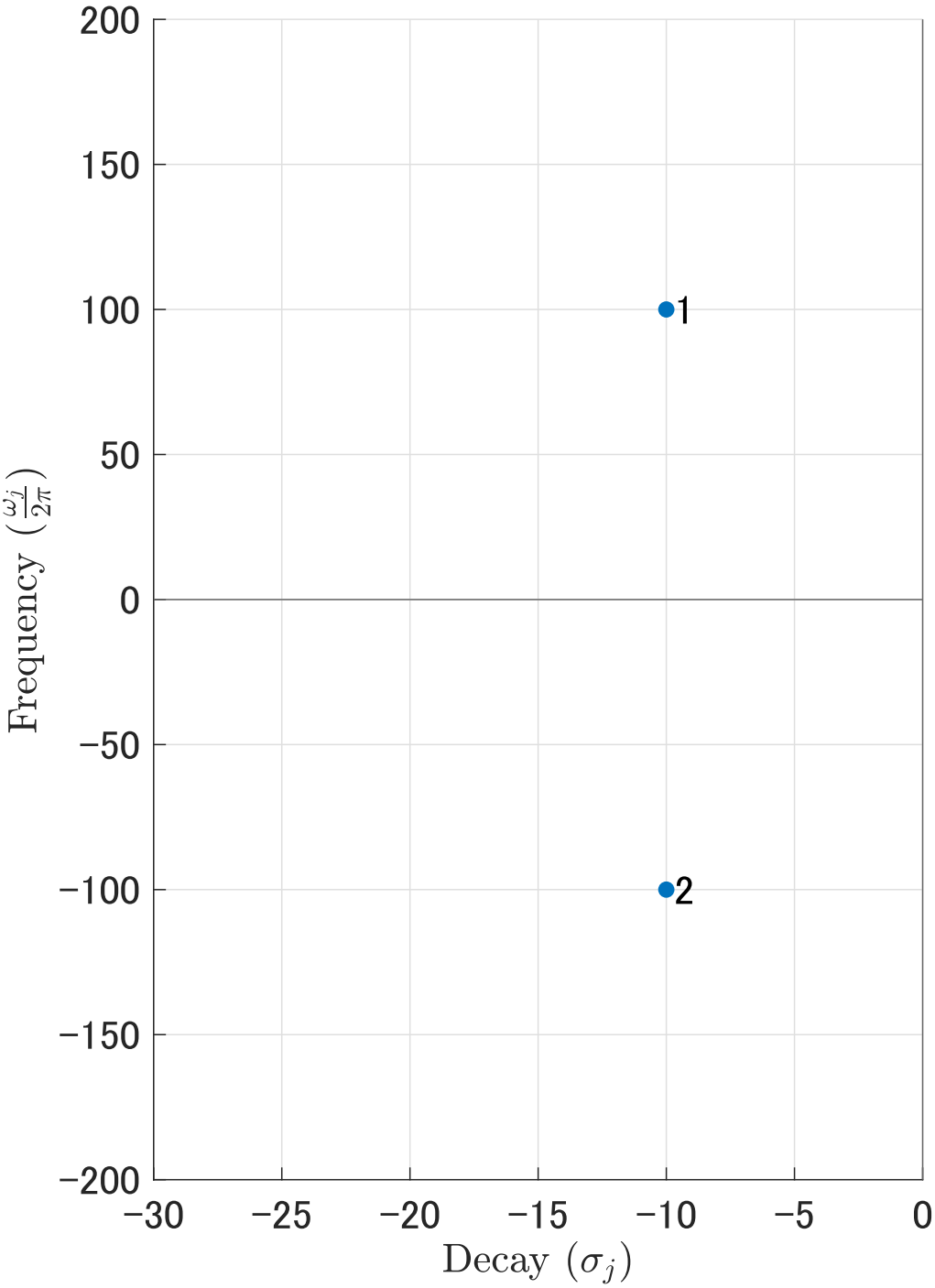

図 3.1: $x_1(t)$ の時間発展特徴の分布図.

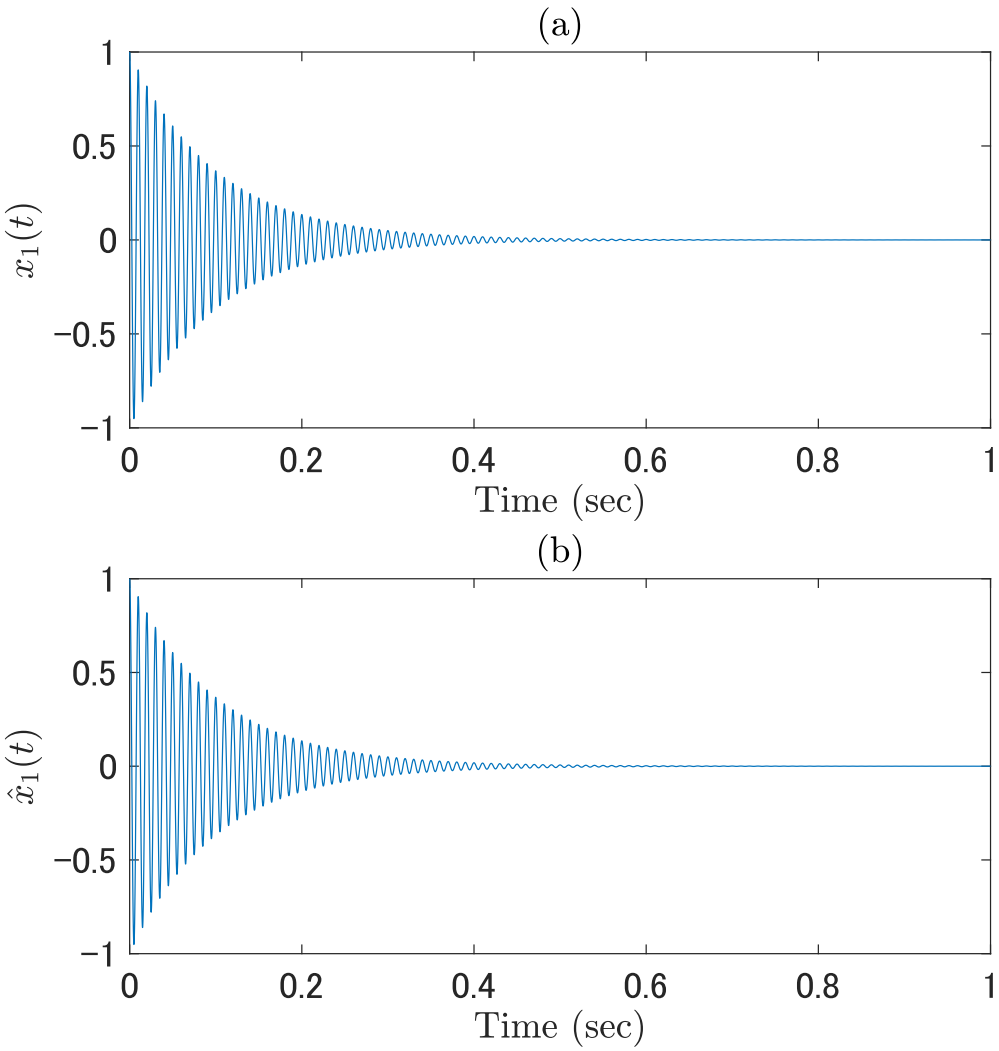

図 3.2: 原信号と再合成信号の波形の比較：(a) 原信号 $x_1(t)$ と (b) 再合成信号 $\hat{x_1}(t)$.

3.4.2 減衰する調波複合音の分析例

次に，以下の式で表されるような，減衰項をもち，基本周波数が 100 Hz で，調波成分が 20 個の調波複合音 $x_2(t)$ を考える．

$$x_2(t) = \sum_{p=1}^{20} e^{-10t} \cos(2\pi 100pt) \tag{3.22}$$

この $x_2(t)$ に対し，得られた時間発展特徴を前節と同様の形で可視化したものを図 3.3 に示す．信号のサンプリング周波数，軸ラベルは前節と同様である．行数は $h = 160$ とした．このとき，時間発展特徴が 40 個抽出された．1 つの周波数成分に対し 2 つモードが抽出されるため，周波数成分が 20 個であれば，図 3.3 に示す通り 40 個のモードが抽出される．$\frac{\omega_j}{2\pi}$ の値は $\{\pm 100, \pm 200, \pm 300, \cdots, \pm 2000\}$ であり，$x_2(t)$ の各調波成分の周波数を取得することができる．また，周波数成分が減衰するとき，その減衰率をその減衰率を σ_j によって取得できる．

前節と同じく，原信号と再合成信号との SER と LSD を計算した結果，SER は 238.72 dB，LSD は 1.4449×10^{-7} dB であった．このように，音響信号が調波複合音の場合でも，情報を損なうことなく DMD で分析・合成が可能と言える．図 3.4 に，原信号 $x_2(t)$，再合成信号 $\hat{x}_2(t)$ の波形を表示する．図 3.4 から，波形としても問題なく再合成が行えていることが分かる．

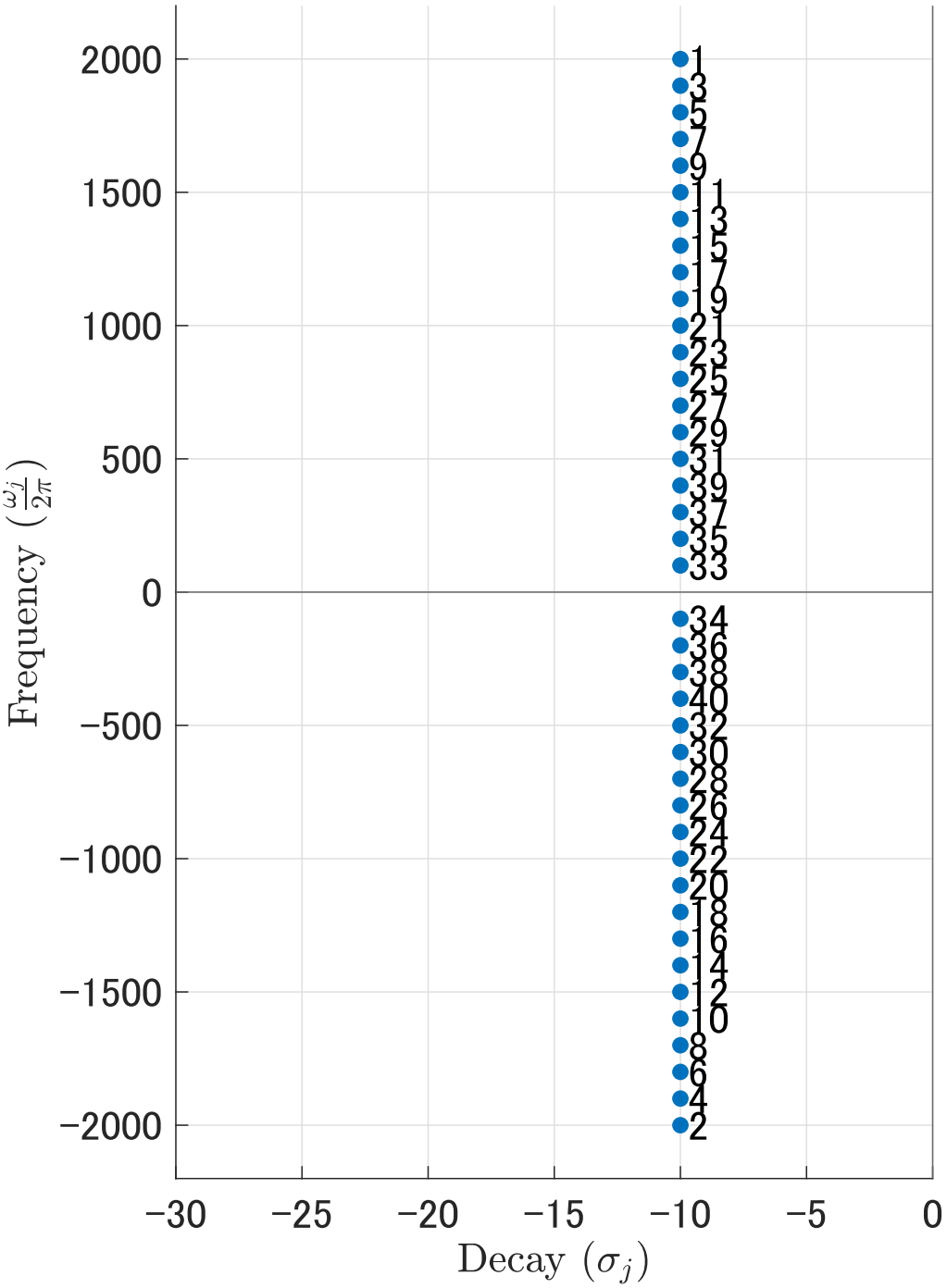

図 3.3: $x_2(t)$ の時間発展特徴の分布図（DMD への入力行列の行数 160 の場合）.

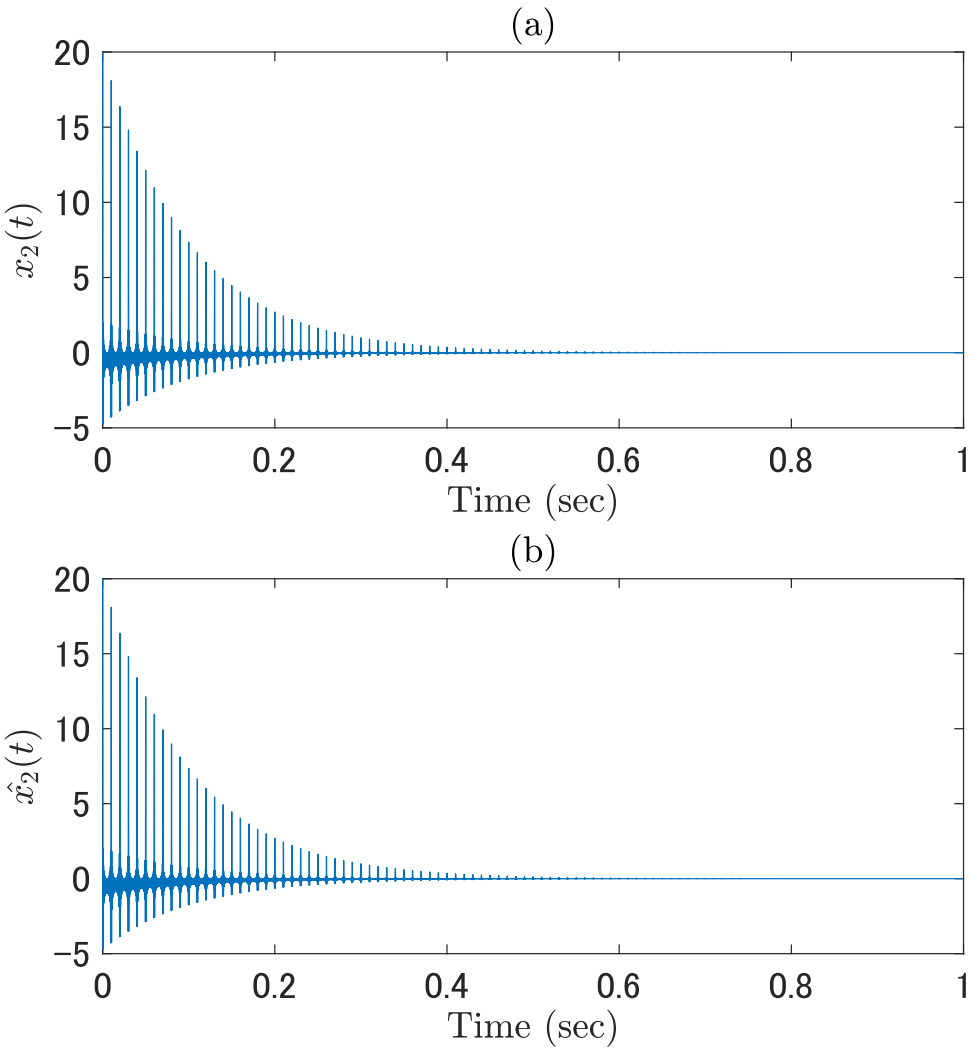

図 3.4: 原信号と再合成信号の波形の比較：(a) 原信号 $x_2(t)$ と (b) 再合成信号 $\hat{x}_2(t)$.

ここで，同じ $x_2(t)$ に対し，$h = 40$ とした場合に抽出された時間発展特徴の分布図を図 3.5 に示す．$x_2(t)$ は 20 個の周波数成分を持ち，モードは 40 個抽出される．このため入力行列の行数は 40 あれば十分だと予想されるが，実際には図 3.5 のように，$x_2(t)$ のもつ減衰率や周波数を正しく捉えることができない．このことから，DMD 分析の際に設定する行数 h は，対象の信号が持つモード数より大きな値を設定する必要があると考えられる．経験的には，対象の信号が持つ最長周期分を含めるような h であれば，DMD による時間発展特徴の分析が問題なく行える．例えば $x_2(t)$ の場合，サンプリング周波数は 16000 Hz，基本周波数は 100 Hz のため，この周波数成分の周期 0.01 sec が最長であり，この 1 周期分を収録するには 160 サンプルとなる．このため，$h \geq 160$ と設定すれば，図 3.3 に示したような時間発展特徴の分析結果が得られる．

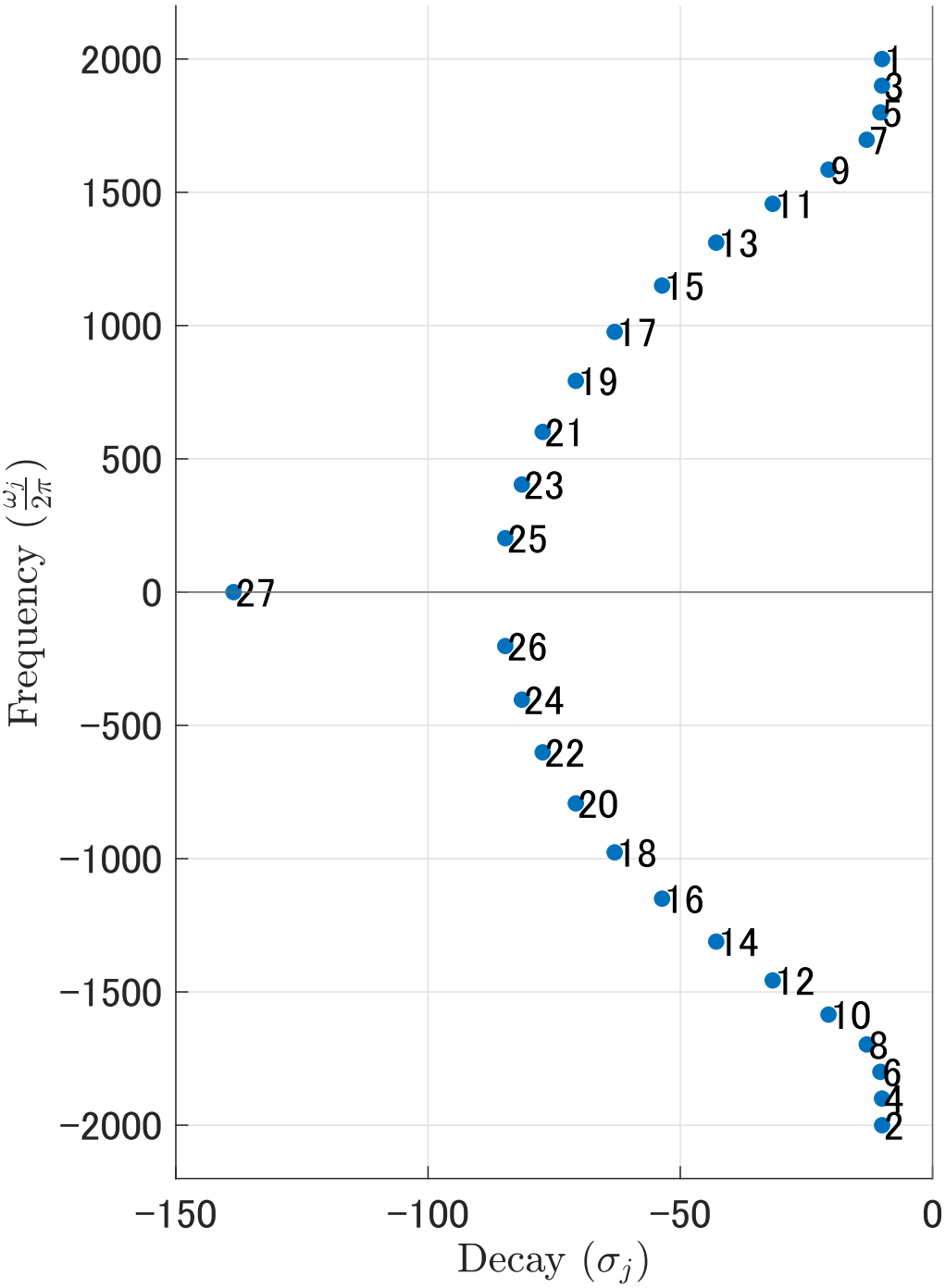

図 3.5: $x_2(t)$ の時間発展特徴の分布図（DMD への入力行列の行数 160 の場合）.

3.4.3 白色雑音の分析例

次に，白色雑音についてDMDを適用した場合にどのような時間発展特徴が得られるかを例示する．サンプリング周波数は前節と同じく16000 Hzとし，$h = 160$で分析した場合を図3.6に示す．プロット点横の添え字は省略しているが，$h = 160$のときモードは160個抽出された．$\frac{\omega_j}{2\pi}$は-8000から8000までの値域に満遍なく分布している．白色雑音は全帯域の周波数成分をもつと解釈すれば，時間発展特徴の$\frac{\omega_j}{2\pi}$はナイキスト周波数8000 Hzを上限として抽出されたと考えられる．また，σ_jが-500近くの値を持ち，非常に急速に減衰する成分として抽出されている．対象の白色雑音は時間減衰しないため，この分析結果は妥当とは言えない．ここで，図3.7は$h = 16000$としたときの分析結果を示しており，時間発展特徴及びモードは16000個抽出された．$h = 160$の場合と比較すると，$\sigma_j = 0$周辺にプロットが集中している．このように，抽出可能なモード数を増加させると，σ_jの値は0へ漸近していく．このことから，白色雑音の分析のためには多数のモードが必要となると考えられる．

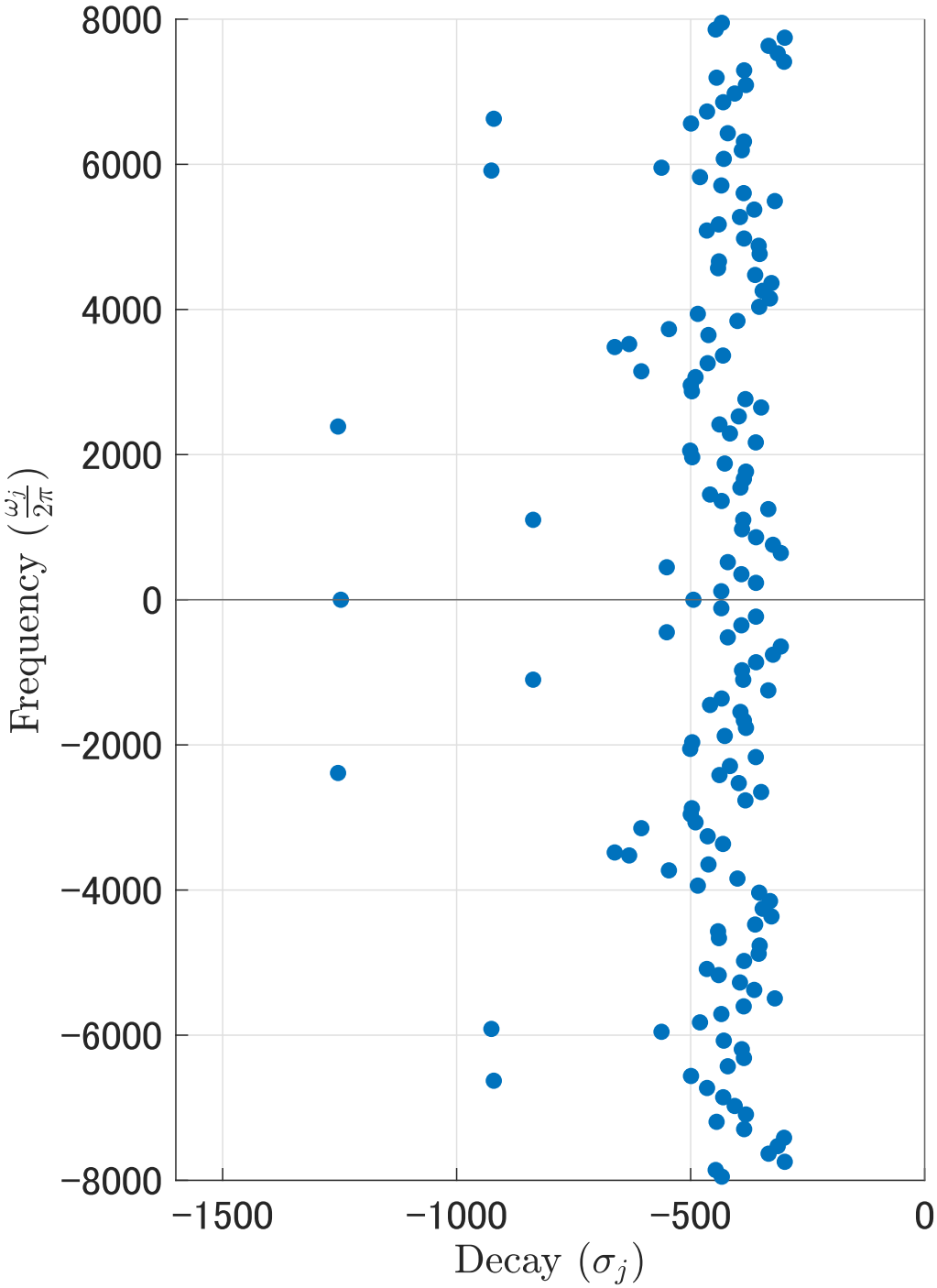

図 3.6: 白色雑音の時間発展特徴の分布図（DMD への入力行列の行数 160 の場合）.

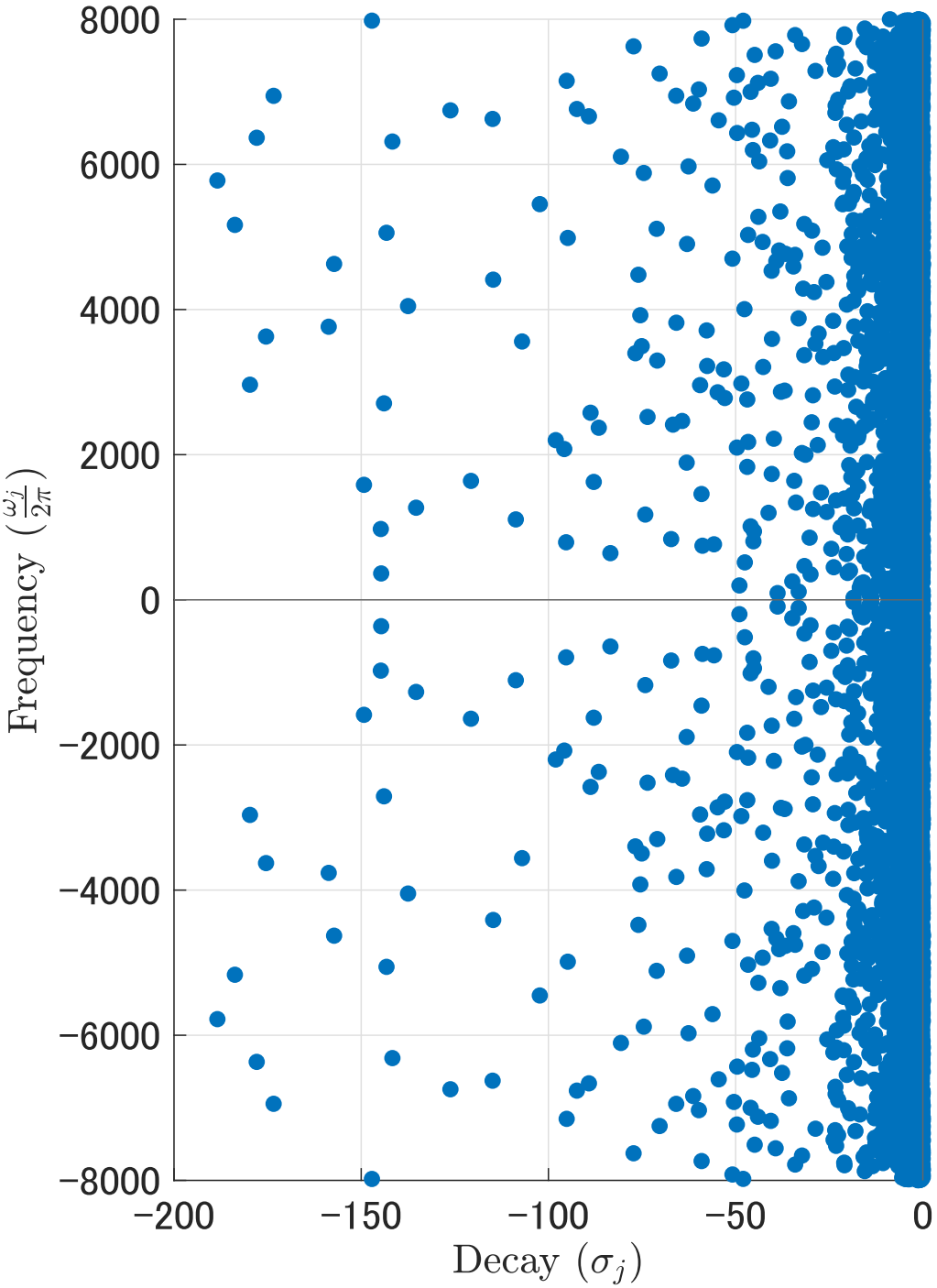

図 3.7: 白色雑音の時間発展特徴の分布図（DMD への入力行列の行数 16000 の場合）.

3.4.4 雑音混合信号の分析例

最後に，雑音混合信号の時間発展特徴の分析結果を例示する．前節の $x_2(t)$ に白色雑音を付加した雑音混合信号 $y(t)$ に対し，$h = 200$ の条件で DMD 分析を施し，抽出された時間発展特徴の可視化図を図 3.8 に示す．サンプリング周波数は 16000 Hz，$x_2(t)$ と白色雑音の SN 比は 15 dB である．時間発展特徴の分布が，赤色の枠で囲んだ $\sigma_j = -10$ 近辺に分布するグループと，それ以外のグループに分離していることが見受けられる．このうち，$\sigma_j = -10$ 近辺に分布する時間発展特徴が，元の $x_2(t)$ がもつ減衰率，周波数を捉えているものと考えられる．ここで，赤色の枠で囲んだ時間発展特徴に対応するモードのみから再合成した信号 $\hat{x_2}(t)$，原信号 $x_2(t)$，雑音混合信号 $y(t)$ の信号波形を図 3.9 に示す．特定のモードのみによって再合成した $\hat{x_2}(t)$ は，$y(t)$ における雑音の影響を除去できている．このように，時間発展特徴の分布図を利用すれば，雑音混合信号から音響信号の分離が行えると考えられる．

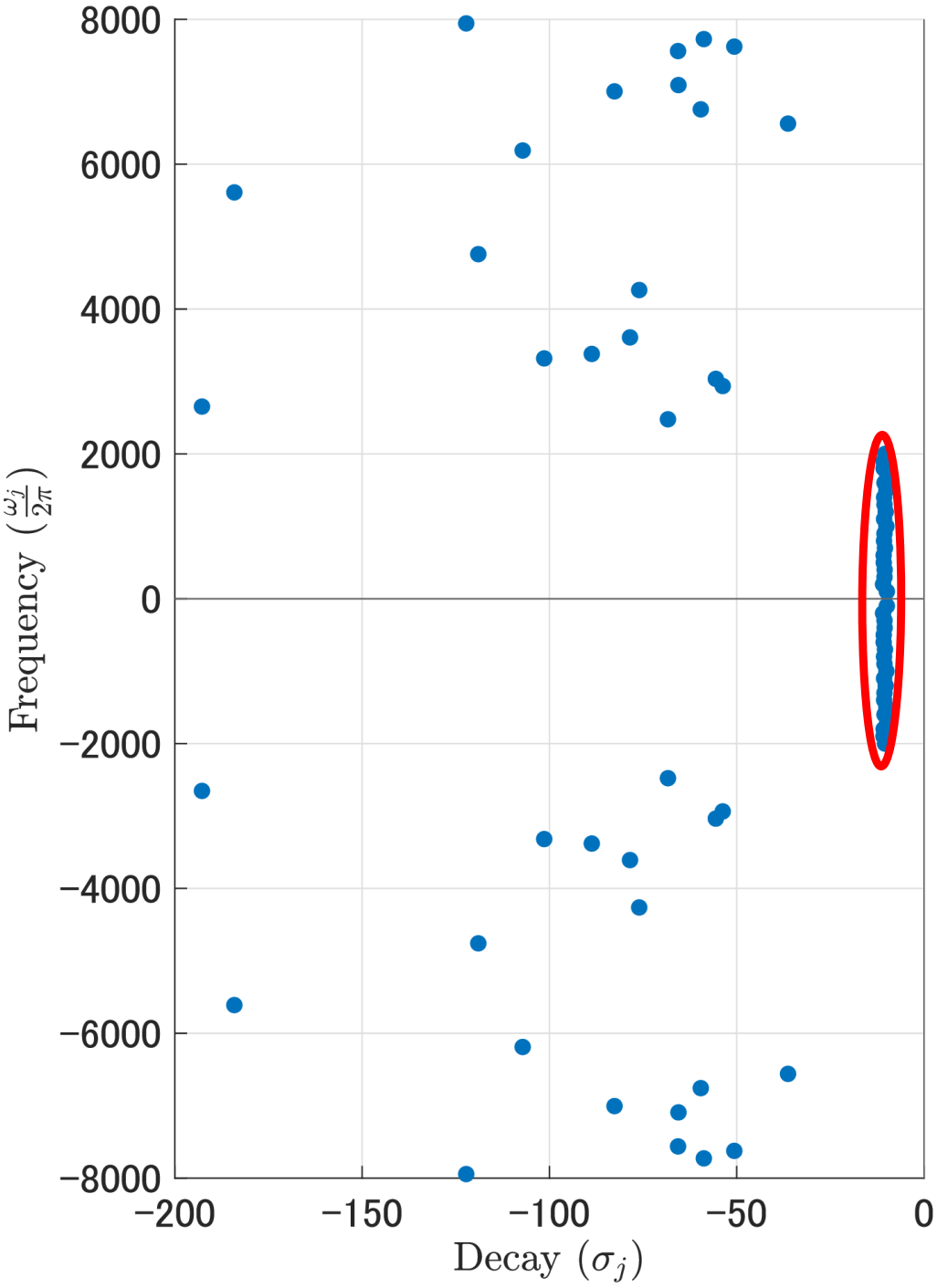

図 3.8: $y(t)$ の時間発展特徴の分布図.

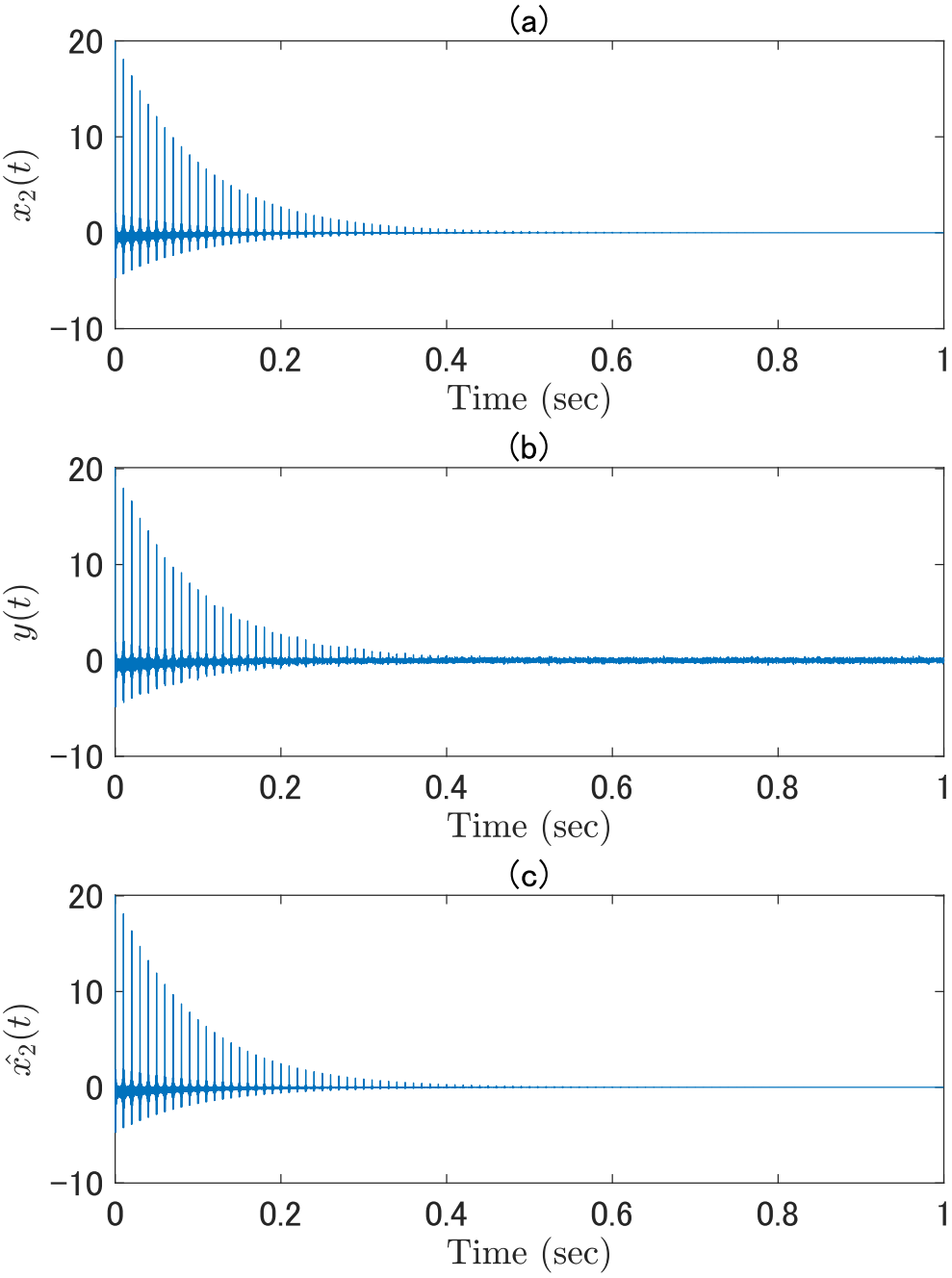

図 3.9: 原信号，雑音混合信号及び再合成信号の波形の比較：(a) 原信号 $x_2(t)$, (b) 雑音混合信号 $y(t)$, (c) 再合成信号 $\hat{x}_2(t)$.

第4章　動的モード分解を用いた音源分離の検討

4.1　検討する分離法

前章で，調波複合音のような音響信号と，白色雑音のような雑音とでは，時間発展特徴の減衰率・周波数の分布に違いがあることを述べた．この分布の違いに基づいて，特定の時間発展特徴に紐付くモードのみ選択することで，音源分離が可能であると考えられる．DMD による音源分離法の実現可能性について，コンピュータシミュレーションを用いて検討した．

図 4.1 に DMD 音源分離法の分離処理のブロックダイアグラムを示す．対象とする信号は，減衰する調波複合音 $x(t)$ と雑音 $N(t)$ の混合信号 $y(t)$ とする．この混合信号に対し DMD を適用することで，一部のモードが $x(t)$ のもつ減衰率を捉え，それは時間発展特徴 σ_j の一部 σ_p に反映される．この σ_p に対応するモード成分 $\alpha_p, \phi_{1,p}, \mu_p$ のみを用いて信号再合成を行うことで，元の音響信号の復元を試みる．σ_p の選択方法としては，σ_j を降順にソートした $\sigma_{j'}$ について，元の調波複合音の調波成分数が q 個と既知であるという条件の下，$1 \leq j' \leq 2q$ の範囲に該当するものを選択した．この基準は，雑音混合信号の時間発展特徴の内，音響信号に該当しない σ_j は負方向に大きいという観察から採用した．

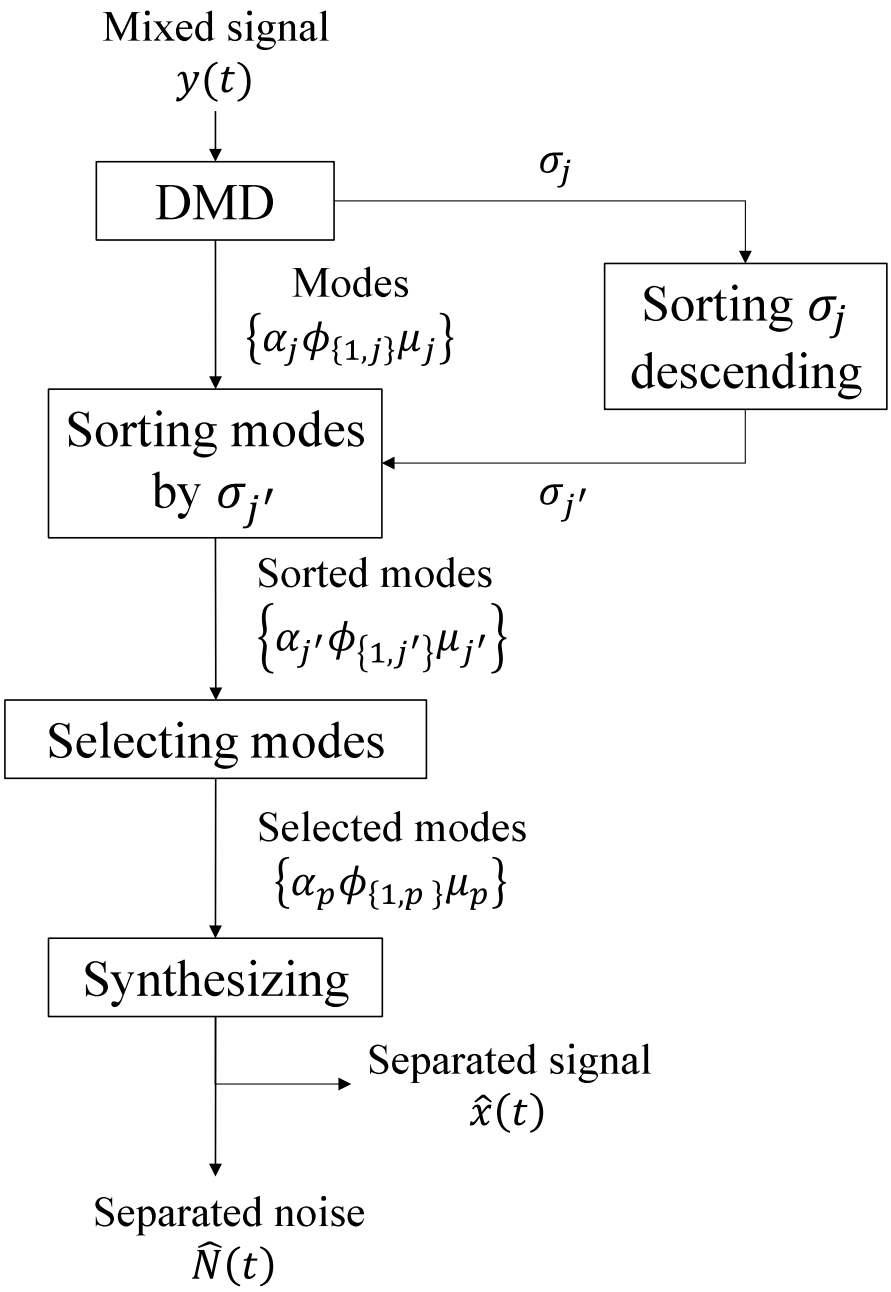

図 4.1: DMD の時間発展特徴を利用した音源分離法のブロックダイアグラム.

4.2 評価シミュレーション

4.2.1 シミュレーション条件

音源分離シミュレーションの設定としては，音源が 2 つで，それぞれ目的音と妨害音の 2 種の音を発生させているものとする．両音源から発せられた音が混合し，モノラル信号として観測されるとする．これを，目的音と妨害音の信号の加算で表現する．そして，同一音源から出た音は，同一の減衰率を持つと仮定し，DMD による音源分離を試みる．目的音 $x(t)$ には，以下の式 4.1 で表される信号を用いた．

$$x_(t) = \sum_{p=1}^{20} \frac{1}{20} e^{-\frac{6.9}{\tau}t} \cos(2\pi p f_0 t) \tag{4.1}$$

f_0 はすべての信号で 100 Hz とした．減衰項として指数関数を採用しているが，その指数 $-\frac{6.9}{\tau}$ は τ 秒で信号パワーが約 60 dB 減衰するように設定した．τ は $\{1, 0.5\}$ としたため，減衰率としては $\{-6.9, -13.8\}$ の 2 条件である．音響信号に付加する妨害音には，白色雑音とピンク雑音を採用した．以下，シミュレーション条件を表に示す．

表 4.1: シミュレーション条件.

対象信号	目的音に妨害音を加算したモノラル信号
目的音	調波複合音 $x(t)$，減衰率 $\{-6.9, -13.8\}$
妨害音	白色雑音，ピンク雑音
SNR 条件	$\{20, 15, 10, 5, 0, -5, -10\}$
行数の条件	$\{200, 400, 600, 800, 1000\}$ dB
検討する手法	DMD による音源分離法
比較する手法	PCA による音源分離法
評価指標	SER，LSD

4.2.2 評価方法

今回の実験シミュレーションは，人工音と雑音との混合信号からクリーンな人工音を分離するというタスクであるため，分離性能の評価尺度として SER[42] と LSD[43] の 2 つを採用した．その計算式は，3 章の式 3.20，式 3.21 と同一である．SER は，原信号に対する復元信号の誤差の割合を SN 比で表現したもので，値が大きいほど分離性能は高いと評価できる．また，LSD は対数表示したパワースペクトルどうしの距離を示し，値が 0 に近いほど分離性能は高いと評価できる．1 つの音響信号 $x(t)$ に対し 100 通りの雑音を混合した $y(t)$ について，SER と LSD の平均値と標準偏差を取り，これを評価した．

これに対し，PCA を用いた音源分離でも同様の条件でシミュレーションを実施し，DMD の分離法と比較する．PCA の分離法について，DMD と同じく調波複合音の調波成分数は 20 個と既知であるとして，取得した固有値の内で値が大きい上位 40 個の固有値に対応する主成分を選択し，この主成分のみを用いて信号を再合成することで音響信号を分離した．なお，PCA を計算する際に入力信号の自己相関行列を推定するが，その次数は DMD の列数条件 h と同じ値とした．

4.2.3 結果

音源分離シミュレーション結果を図 4.2〜図 4.21 に示す．図 4.2〜図 4.11 が白色雑音との混合信号のシミュレーション結果，図 4.12〜図 4.21 がピンク雑音との混合信号のシミュレーション結果である．まず，入力行列の行数 $h = \{600, 800, 1000\}$ の結果に注目すると，DMD の音源分離法は，PCA の音源分離法より SER，LSD ともに上回る分離性能を記録したことが分かる．白色雑音との混合信号の場合，SER，LSD とも標準偏差は 0.5 dB 未満とばらつきは小さい．ピンク雑音との混合信号の場合，SER の標準偏差は白色雑音の場合と比較してやや大きい．DMD の音源分離法は PCA の音源分離法を上回る分離性能ではあるものの，雑音の種類によって分離性能の安定性に差が生じた．また，SNR が -5 dB 以下の条件では，SER は両者とも大きくは変わらないものの，LSD では DMD の音源分離法が良い分離性能を記録している．$h = \{200, 400\}$ の SER の結果に注目すると，SNR が 10 dB 以上の高 SNR 条件では DMD の音源分離法が優位な性能であるものの，低 SNR では PCA の音源分離法の性能が高いことが分かる．このことから，DMD による音源分離法の分離性能には h が強く影響したと考えられる．今回のシミュレーションの場合，$h = 600$ 以上が良い h の設定であったと言える．

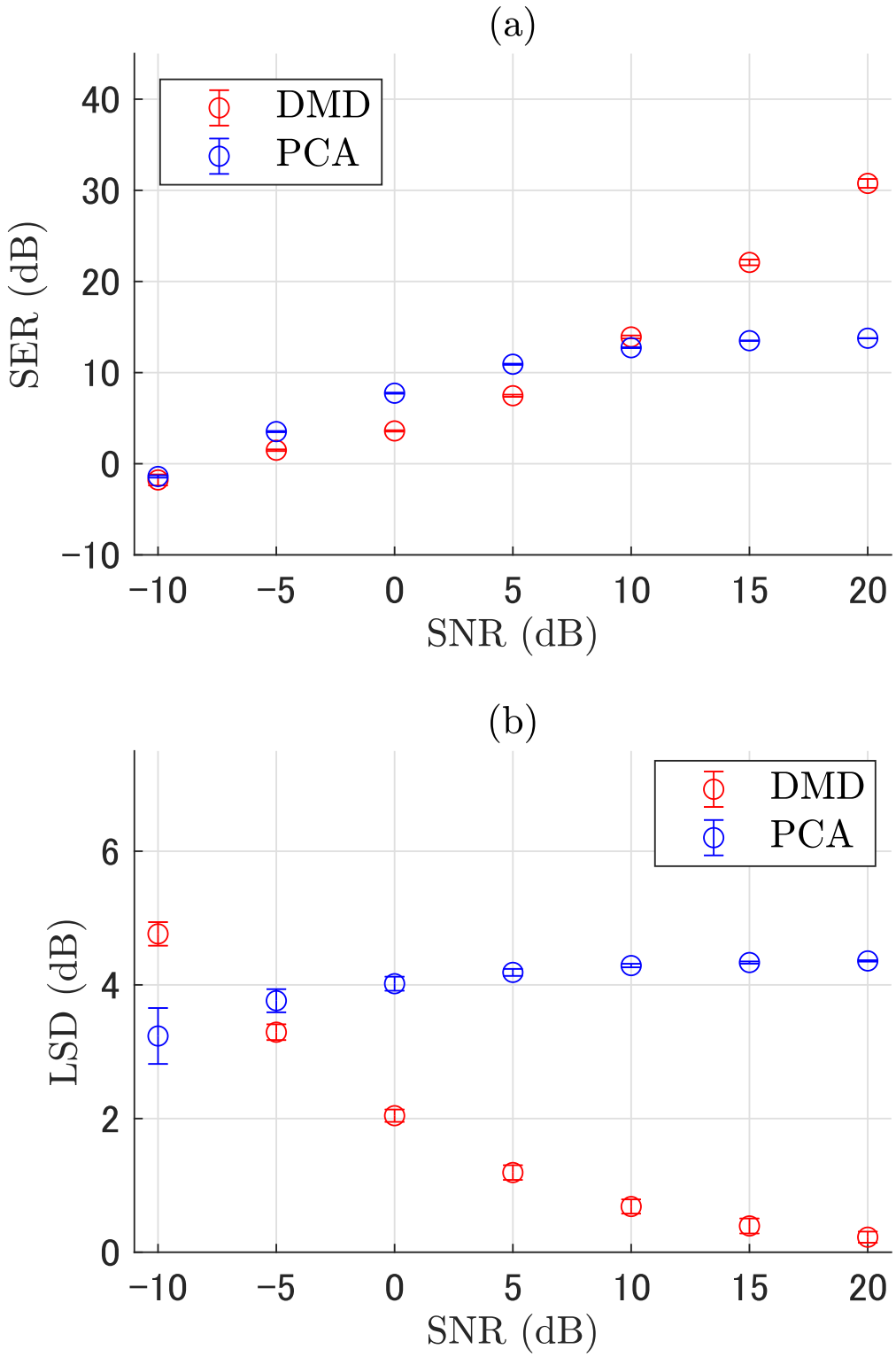

図 4.2: 減衰率 −6.9 の調波複合音と白色雑音との混合信号における，DMD と PCA の音源分離シミュレーションの結果（入力行列の行数 200 の場合）：(a) 原信号と混合信号との SER, (b) 原信号と混合信号との LSD.

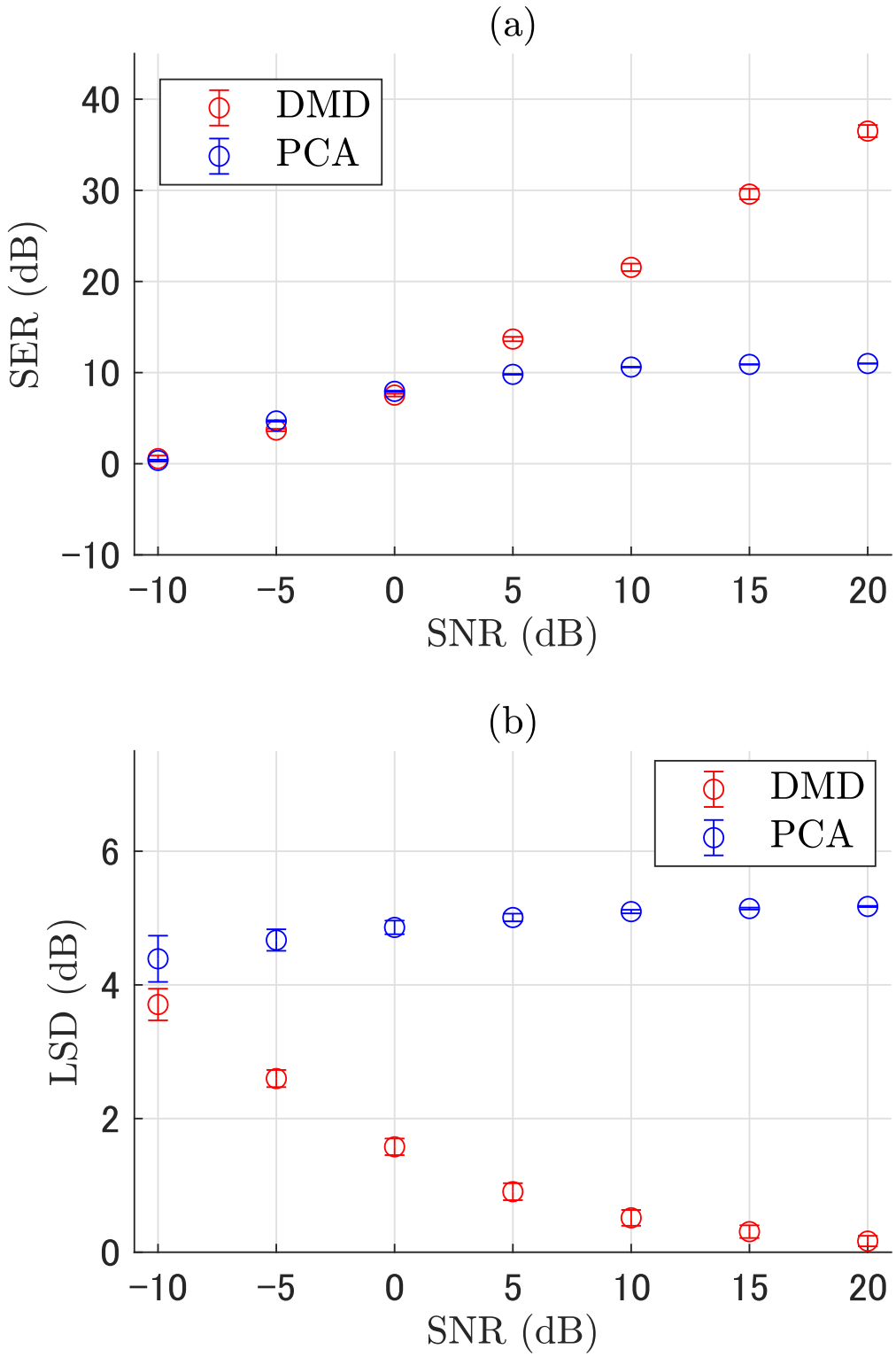

図 4.3: 減衰率 −6.9 の調波複合音と白色雑音との混合信号における，DMD と PCA の音源分離シミュレーションの結果（入力行列の行数 400 の場合）：(a) 原信号と混合信号との SER, (b) 原信号と混合信号との LSD.

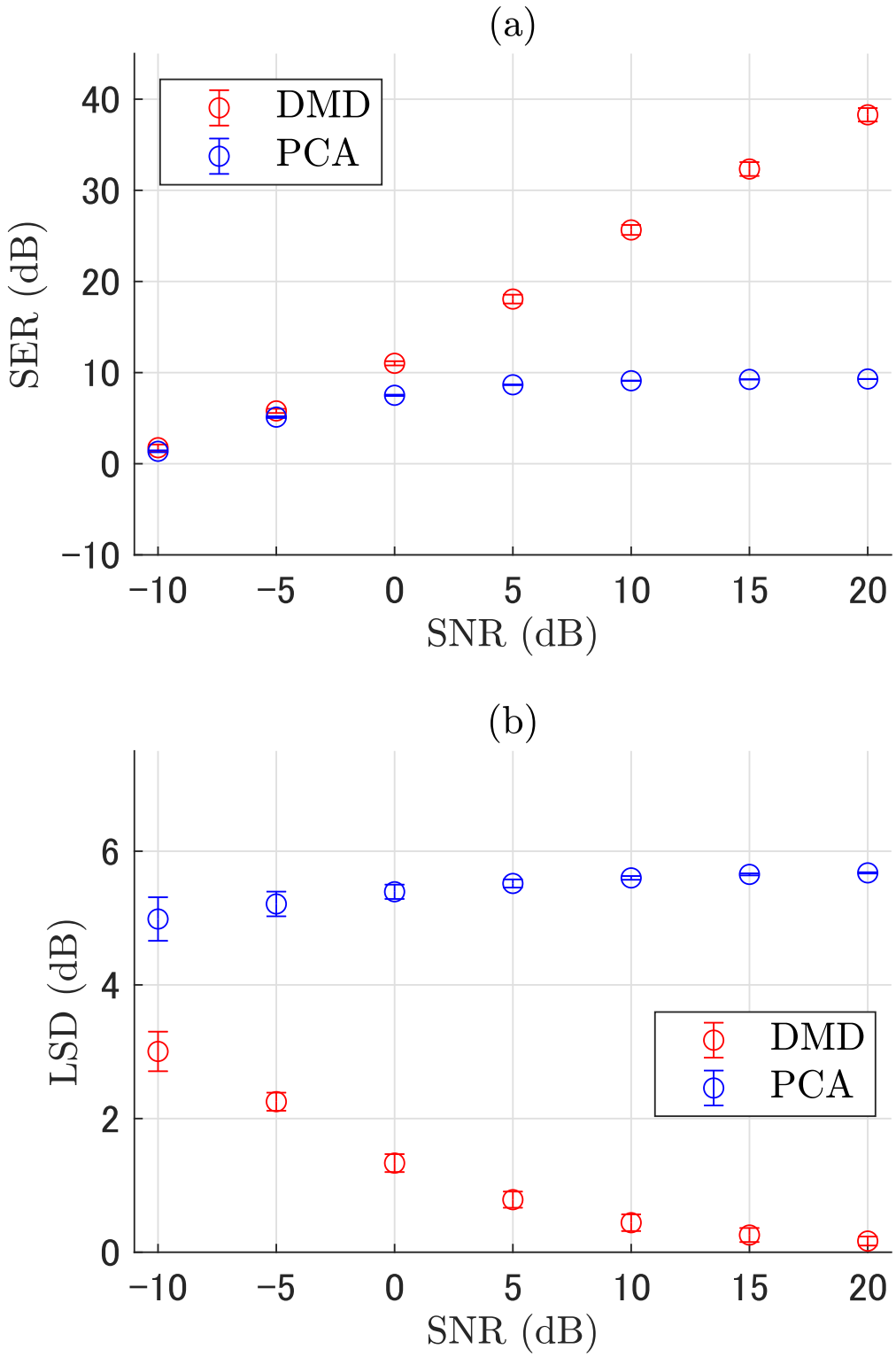

図 4.4: 減衰率 −6.9 の調波複合音と白色雑音との混合信号における，DMD と PCA の音源分離シミュレーションの結果（入力行列の行数 600 の場合）：(a) 原信号と混合信号との SER, (b) 原信号と混合信号との LSD.

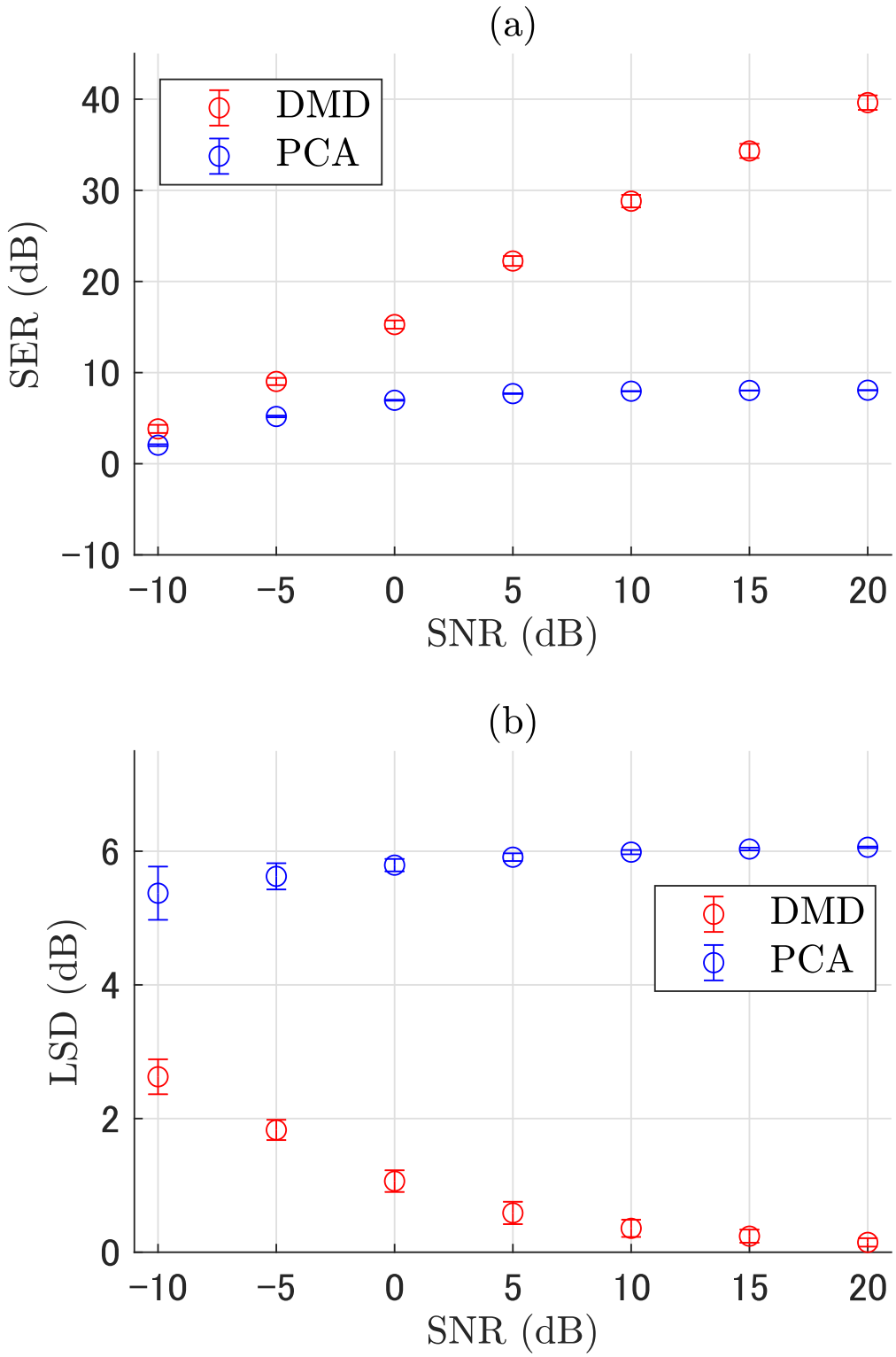

図 4.5: 減衰率 −6.9 の調波複合音と白色雑音との混合信号における，DMD と PCA の音源分離シミュレーションの結果（入力行列の行数 800 の場合）：(a) 原信号と混合信号との SER, (b) 原信号と混合信号との LSD.

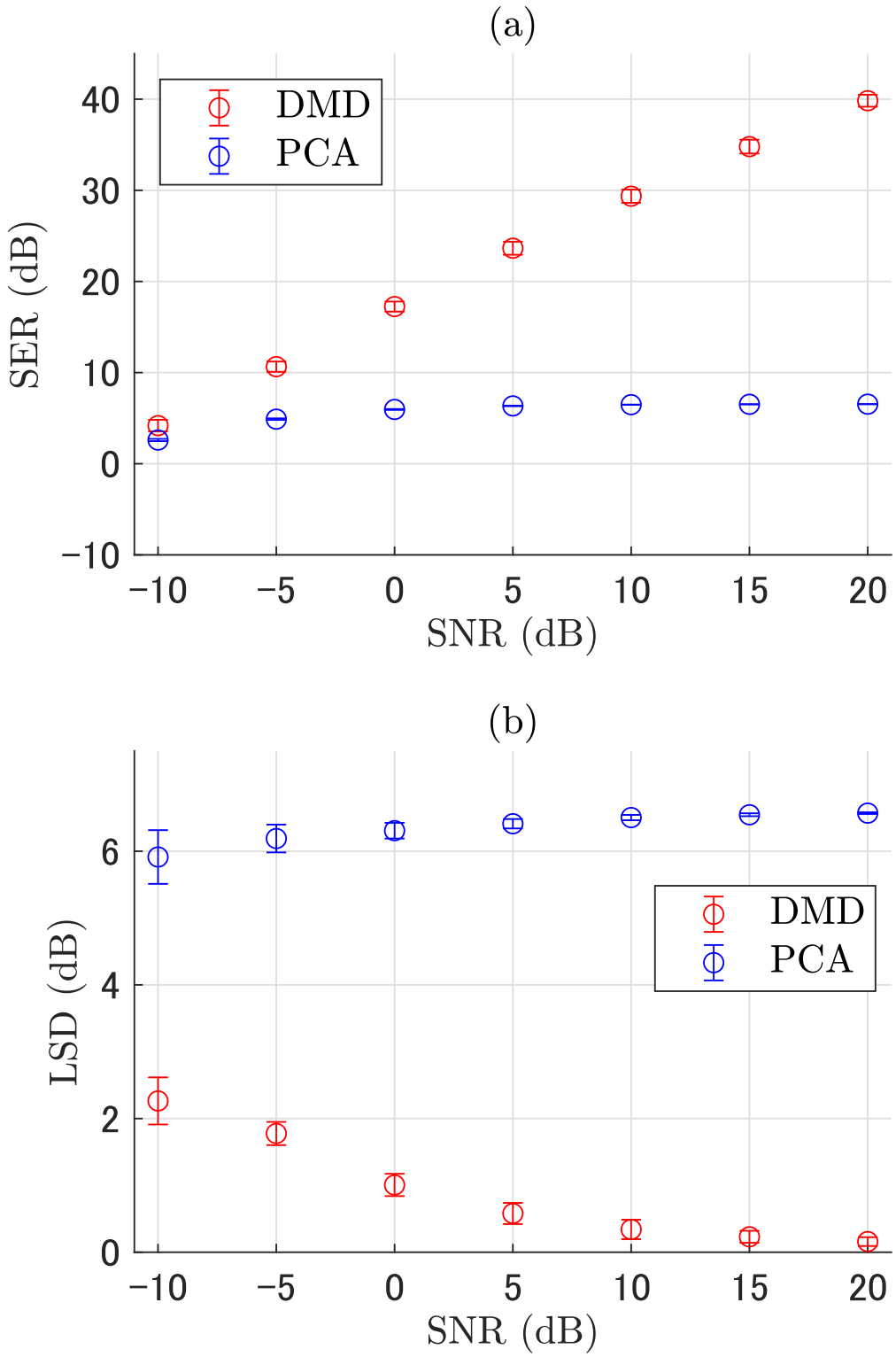

図 4.6: 減衰率 −6.9 の調波複合音と白色雑音との混合信号における，DMD と PCA の音源分離シミュレーションの結果（入力行列の行数 1000 の場合）：(a) 原信号と混合信号との SER, (b) 原信号と混合信号との LSD.

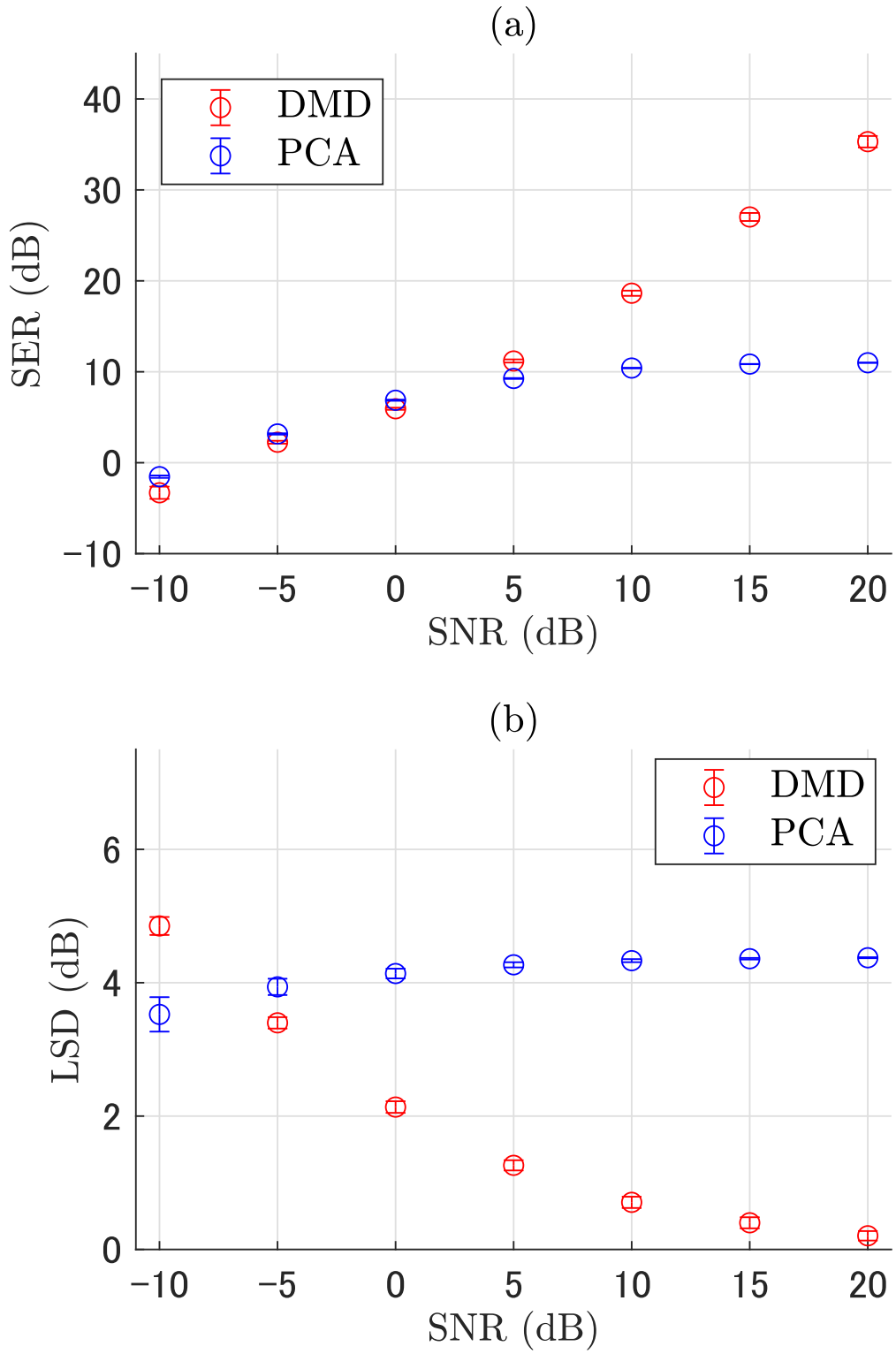

図 4.7: 減衰率 −13.8 の調波複合音と白色雑音との混合信号における，DMD と PCA の音源分離シミュレーションの結果（入力行列の行数 200 の場合）: (a) 原信号と混合信号との SER, (b) 原信号と混合信号との LSD.

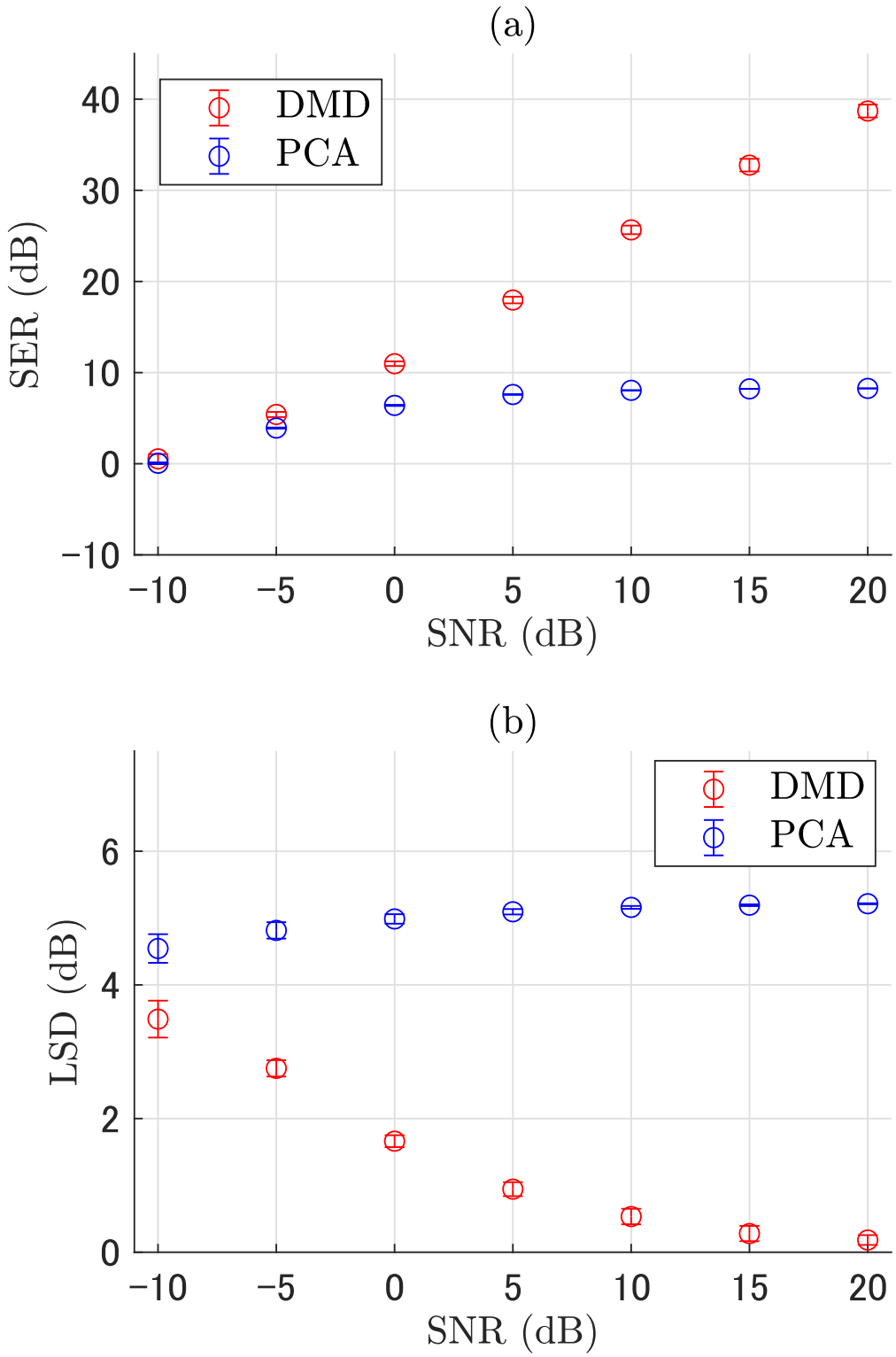

図 4.8: 減衰率 −13.8 の調波複合音と白色雑音との混合信号における，DMD と PCA の音源分離シミュレーションの結果（入力行列の行数 400 の場合）: (a) 原信号と混合信号との SER, (b) 原信号と混合信号との LSD.

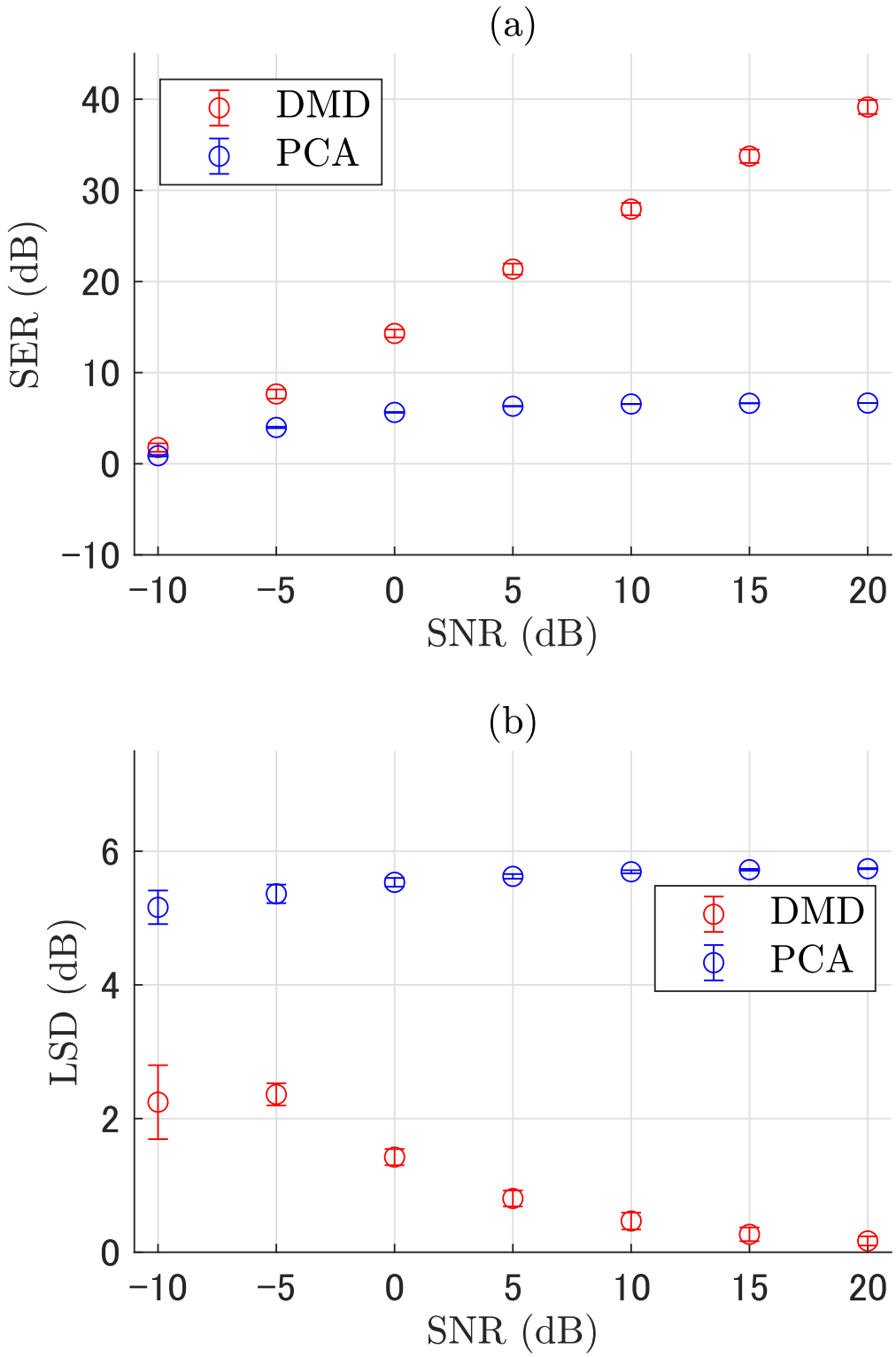

図 4.9: 減衰率 −13.8 の調波複合音と白色雑音との混合信号における，DMD と PCA の音源分離シミュレーションの結果（入力行列の行数 600 の場合）: (a) 原信号と混合信号との SER, (b) 原信号と混合信号との LSD.

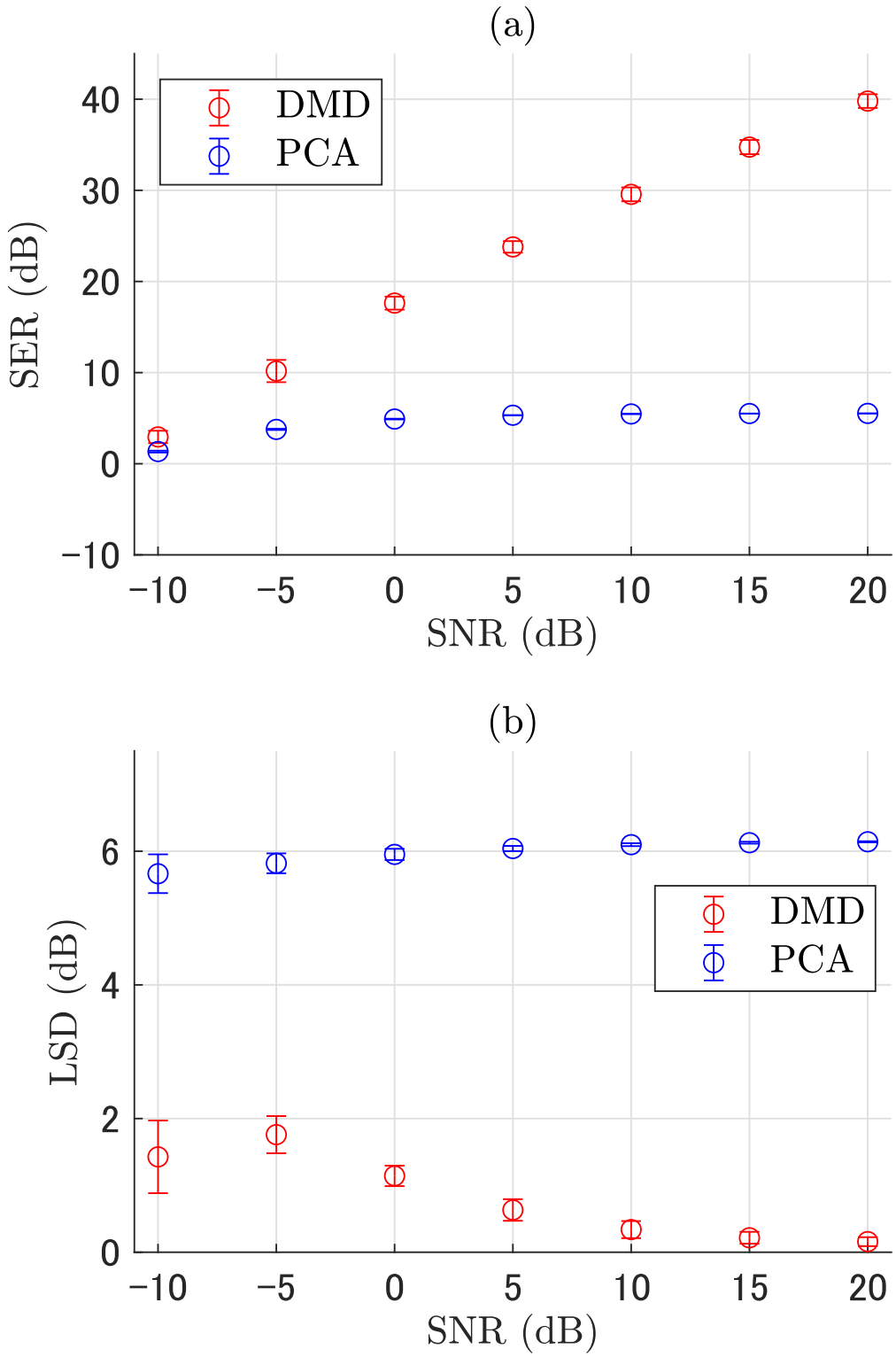

図 4.10: 減衰率 −13.8 の調波複合音と白色雑音との混合信号における，DMD と PCA の音源分離シミュレーションの結果（入力行列の行数 800 の場合）: (a) 原信号と混合信号との SER, (b) 原信号と混合信号との LSD.

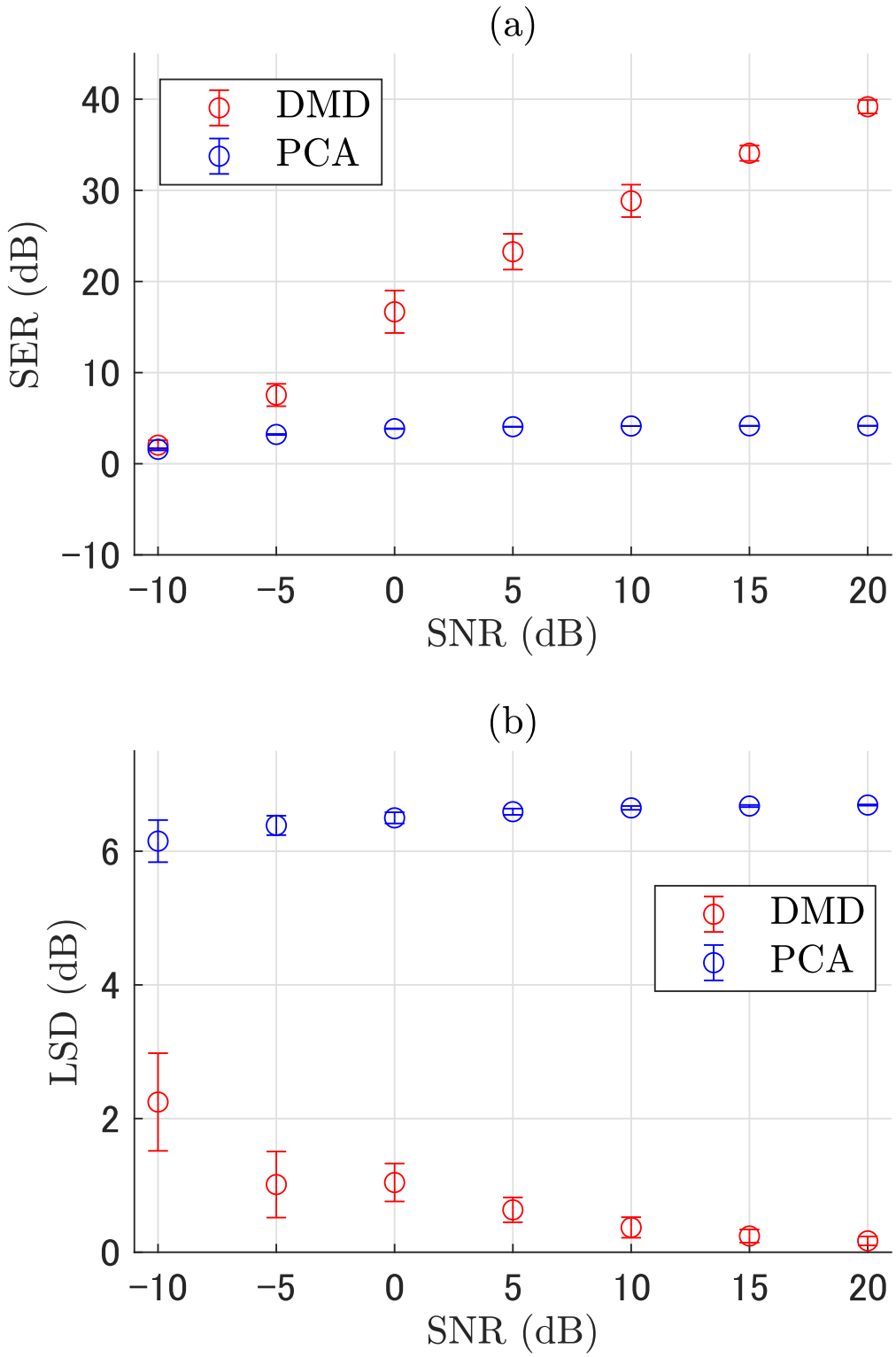

図 4.11: 減衰率 −13.8 の調波複合音と白色雑音との混合信号における，DMD と PCA の音源分離シミュレーションの結果（入力行列の行数 1000 の場合）：(a) 原信号と混合信号との SER, (b) 原信号と混合信号との LSD.

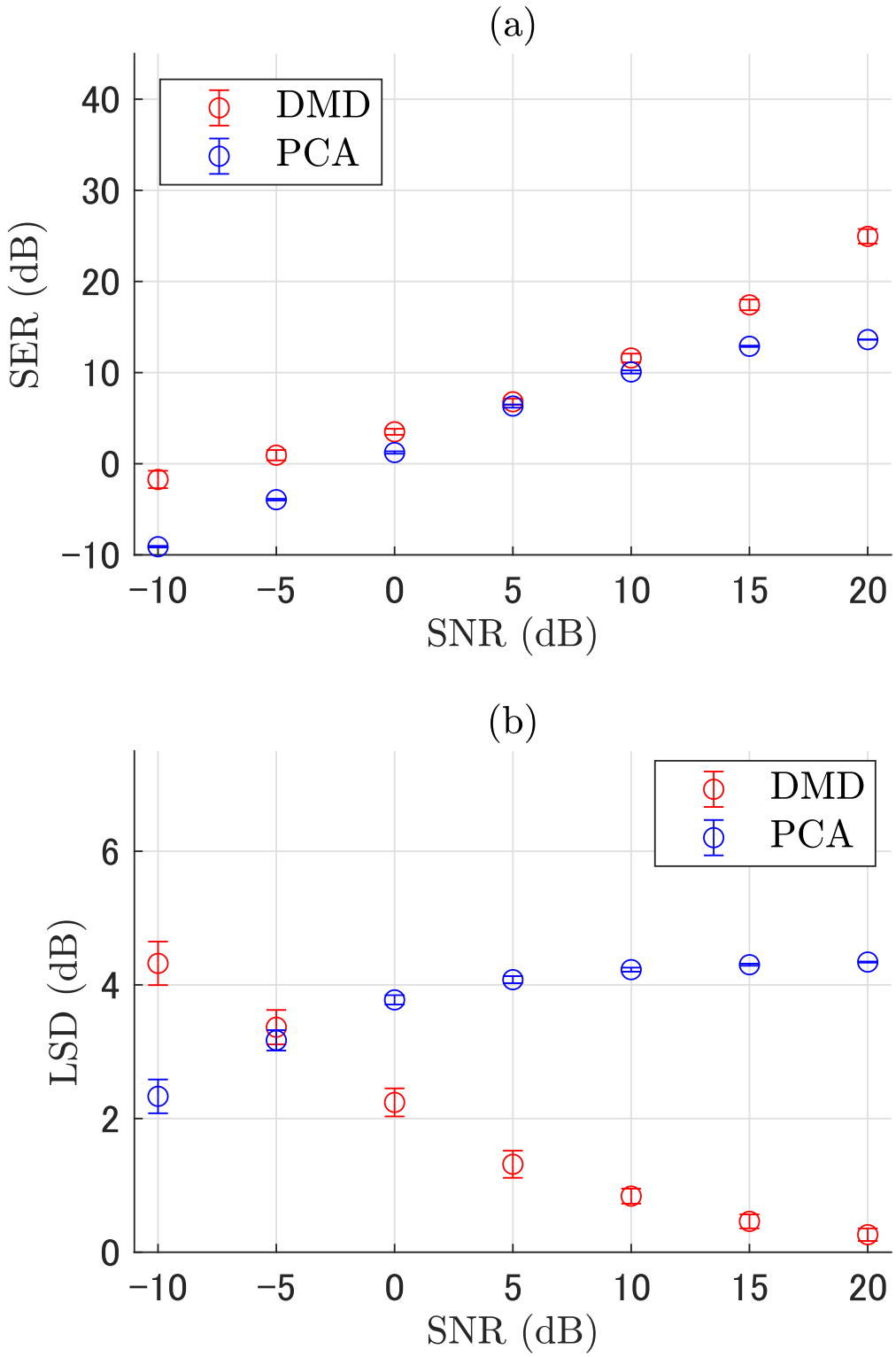

図 4.12: 減衰率 −6.9 の調波複合音とピンク雑音との混合信号における，DMD と PCA の音源分離シミュレーションの結果（入力行列の行数 200 の場合）: (a) 原信号と混合信号との SER, (b) 原信号と混合信号との LSD.

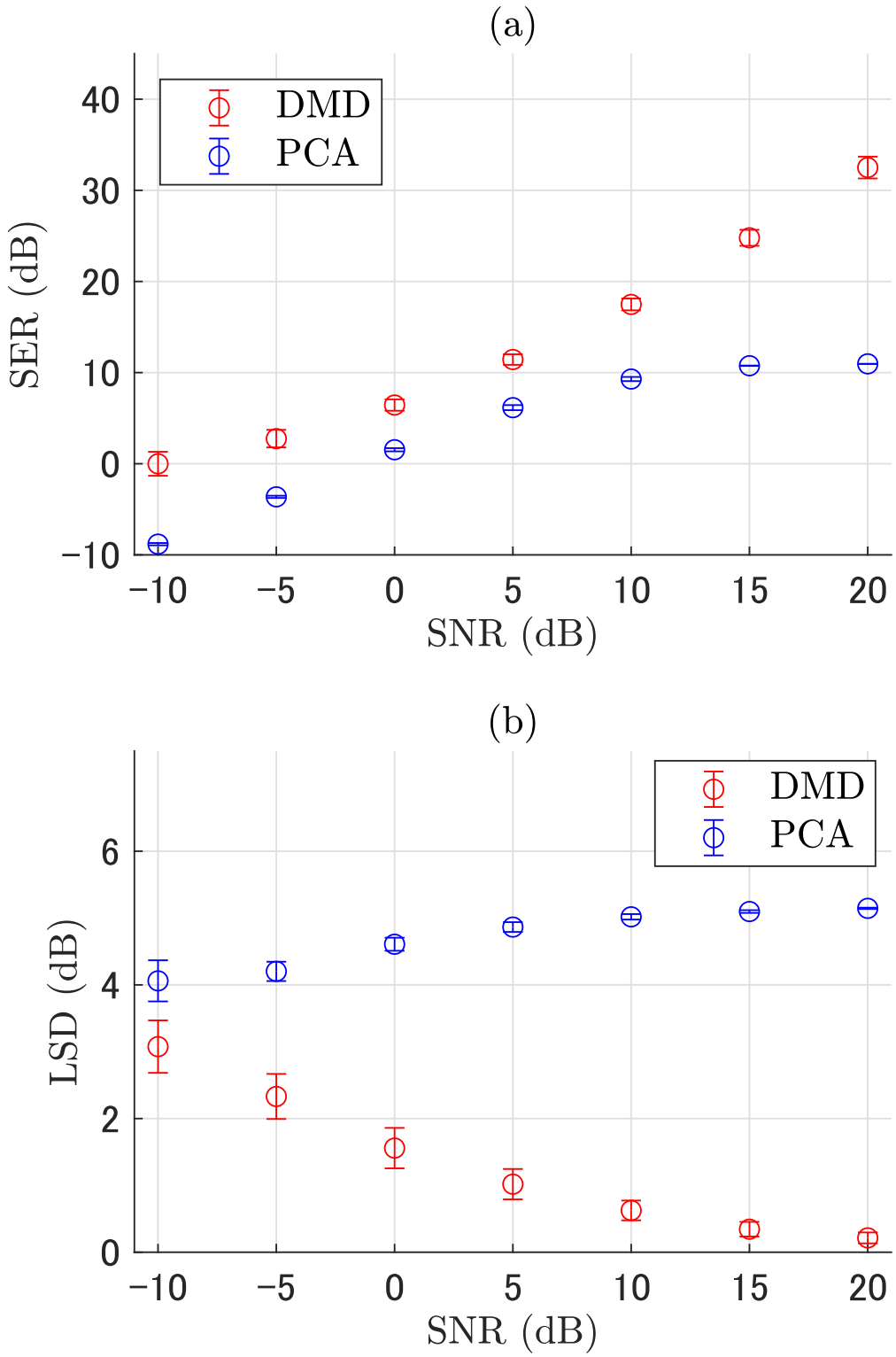

図 4.13: 減衰率 −6.9 の調波複合音とピンク雑音との混合信号における，DMD と PCA の音源分離シミュレーションの結果（入力行列の行数 400 の場合）: (a) 原信号と混合信号との SER, (b) 原信号と混合信号との LSD.

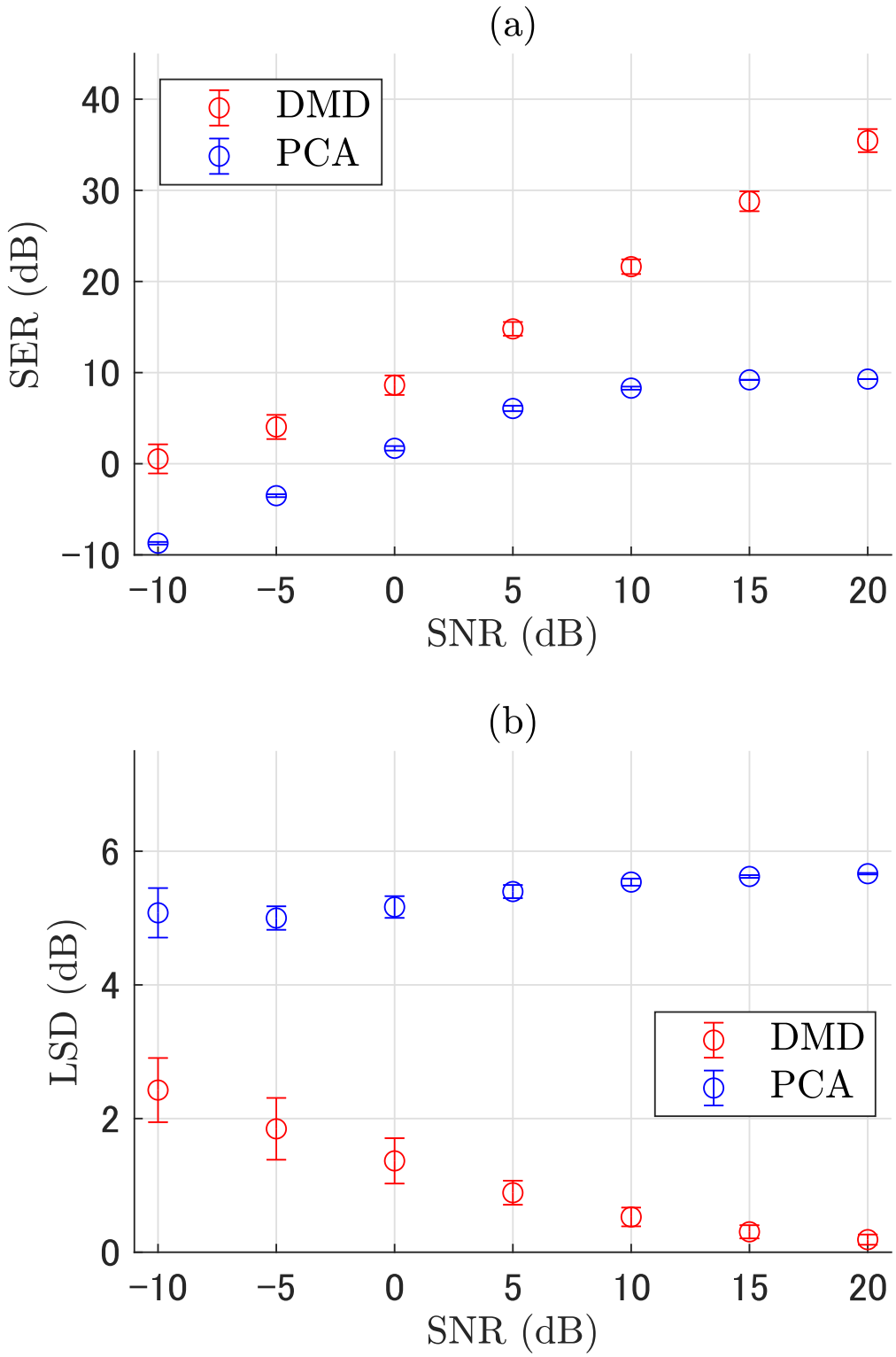

図 4.14: 減衰率 −6.9 の調波複合音とピンク雑音との混合信号における，DMD と PCA の音源分離シミュレーションの結果（入力行列の行数 600 の場合）: (a) 原信号と混合信号との SER, (b) 原信号と混合信号との LSD.

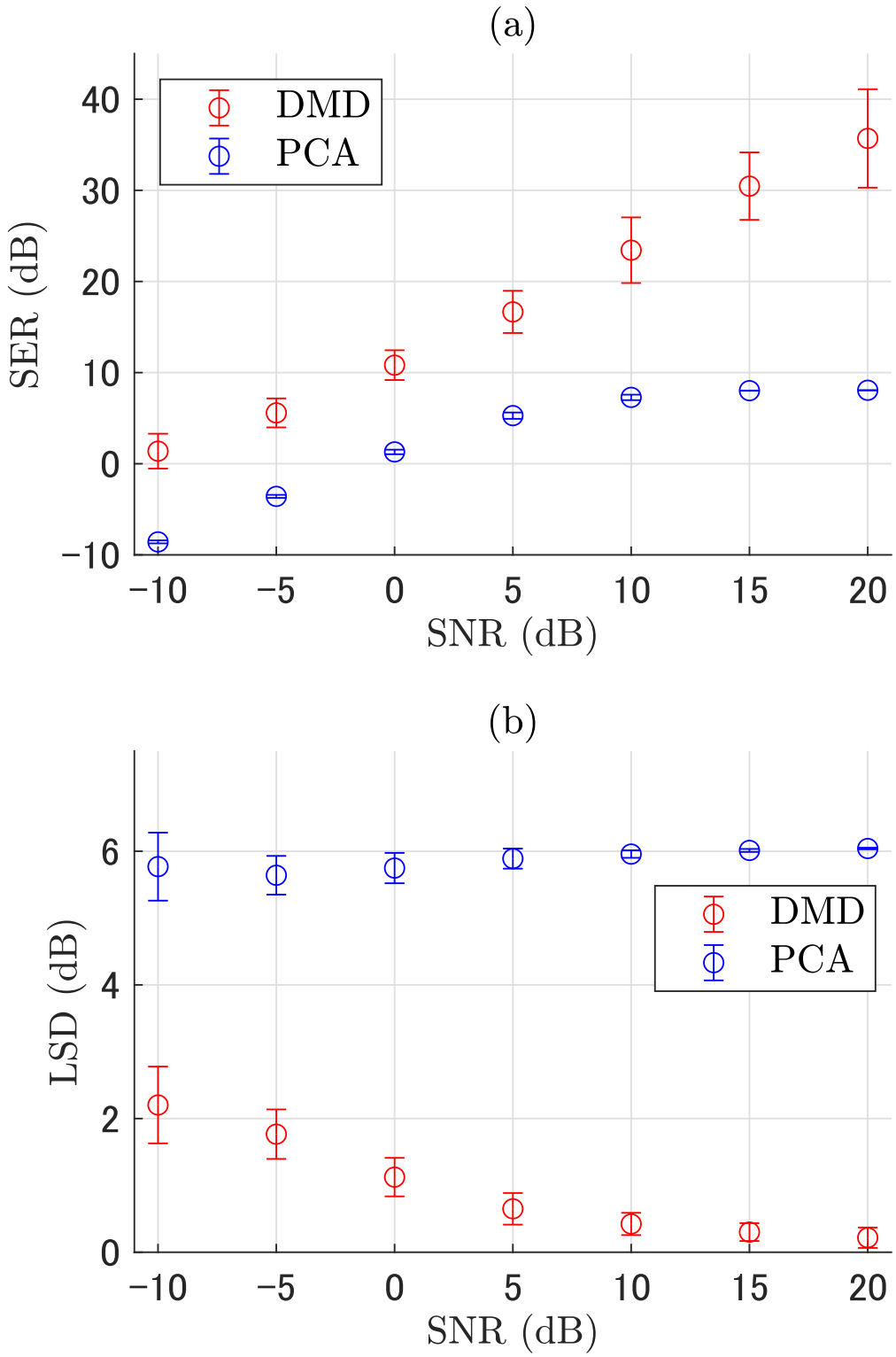

図 4.15: 減衰率 −6.9 の調波複合音とピンク雑音との混合信号における，DMD と PCA の音源分離シミュレーションの結果（入力行列の行数 800 の場合）: (a) 原信号と混合信号との SER, (b) 原信号と混合信号との LSD.

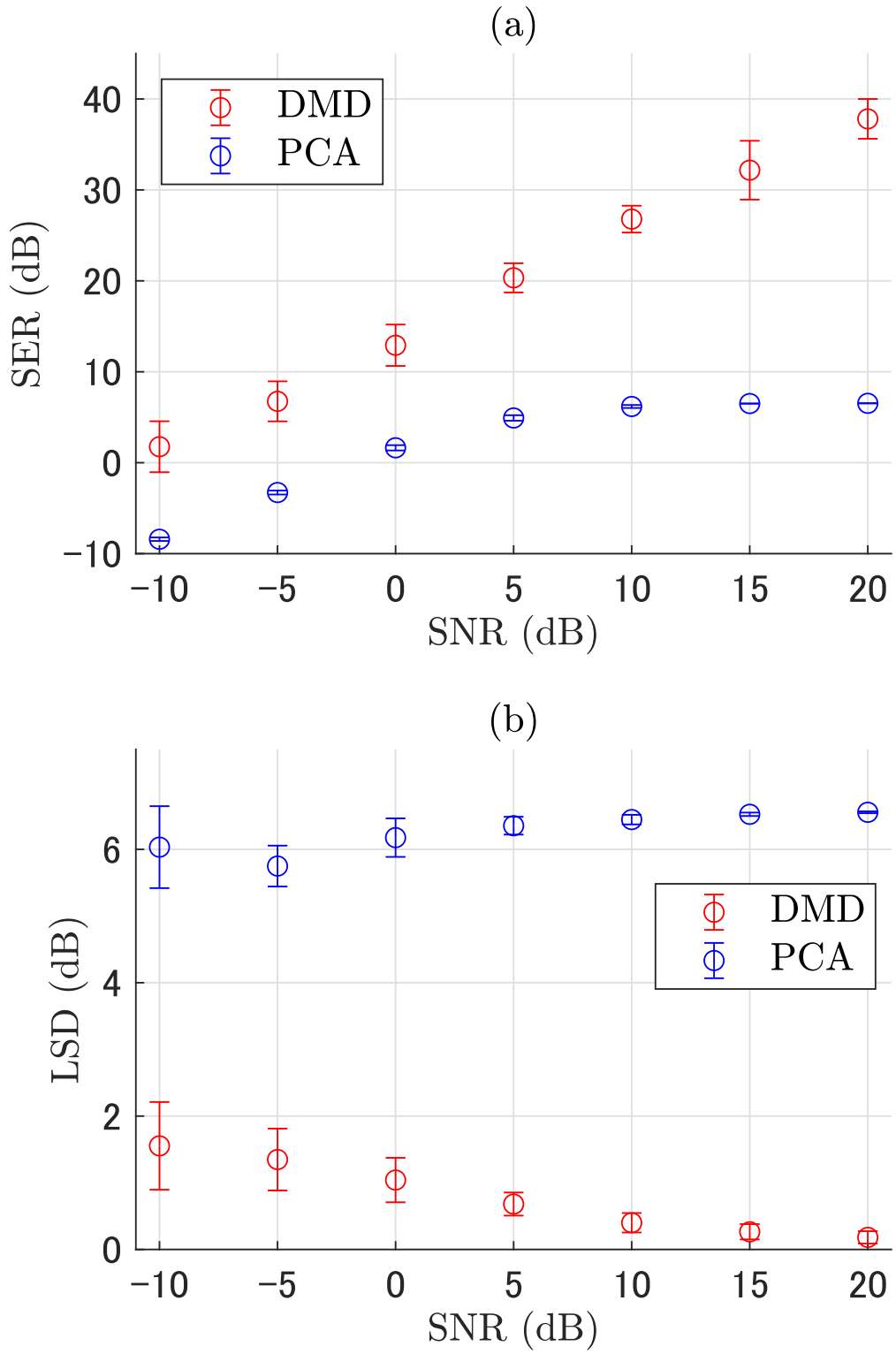

図 4.16: 減衰率 −6.9 の調波複合音とピンク雑音との混合信号における，DMD と PCA の音源分離シミュレーションの結果（入力行列の行数 1000 の場合）：(a) 原信号と混合信号との SER, (b) 原信号と混合信号との LSD.

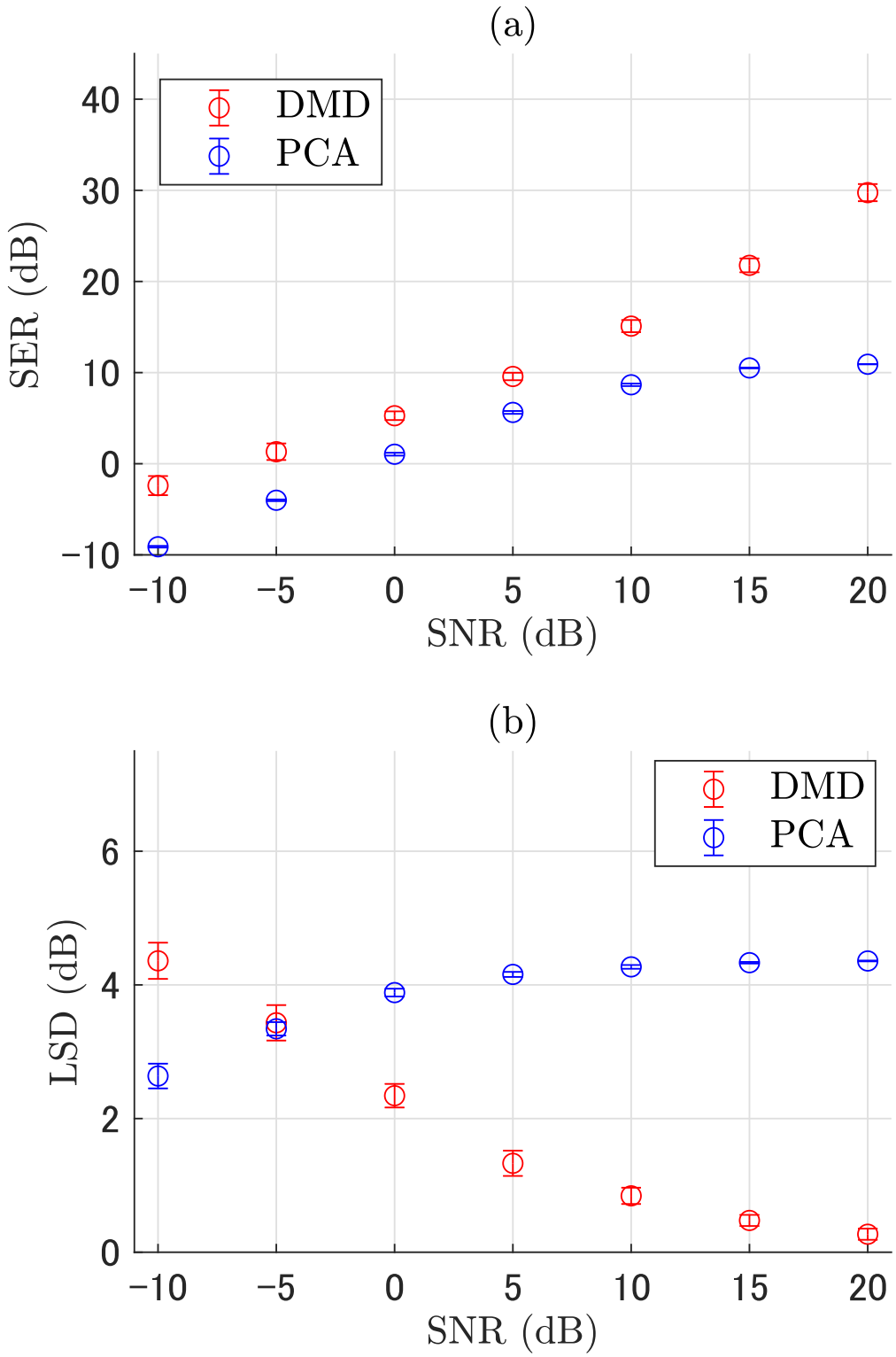

図 4.17: 減衰率 −13.8 の調波複合音とピンク雑音との混合信号における，DMD と PCA の音源分離シミュレーションの結果（入力行列の行数 200 の場合）: (a) 原信号と混合信号との SER, (b) 原信号と混合信号との LSD.

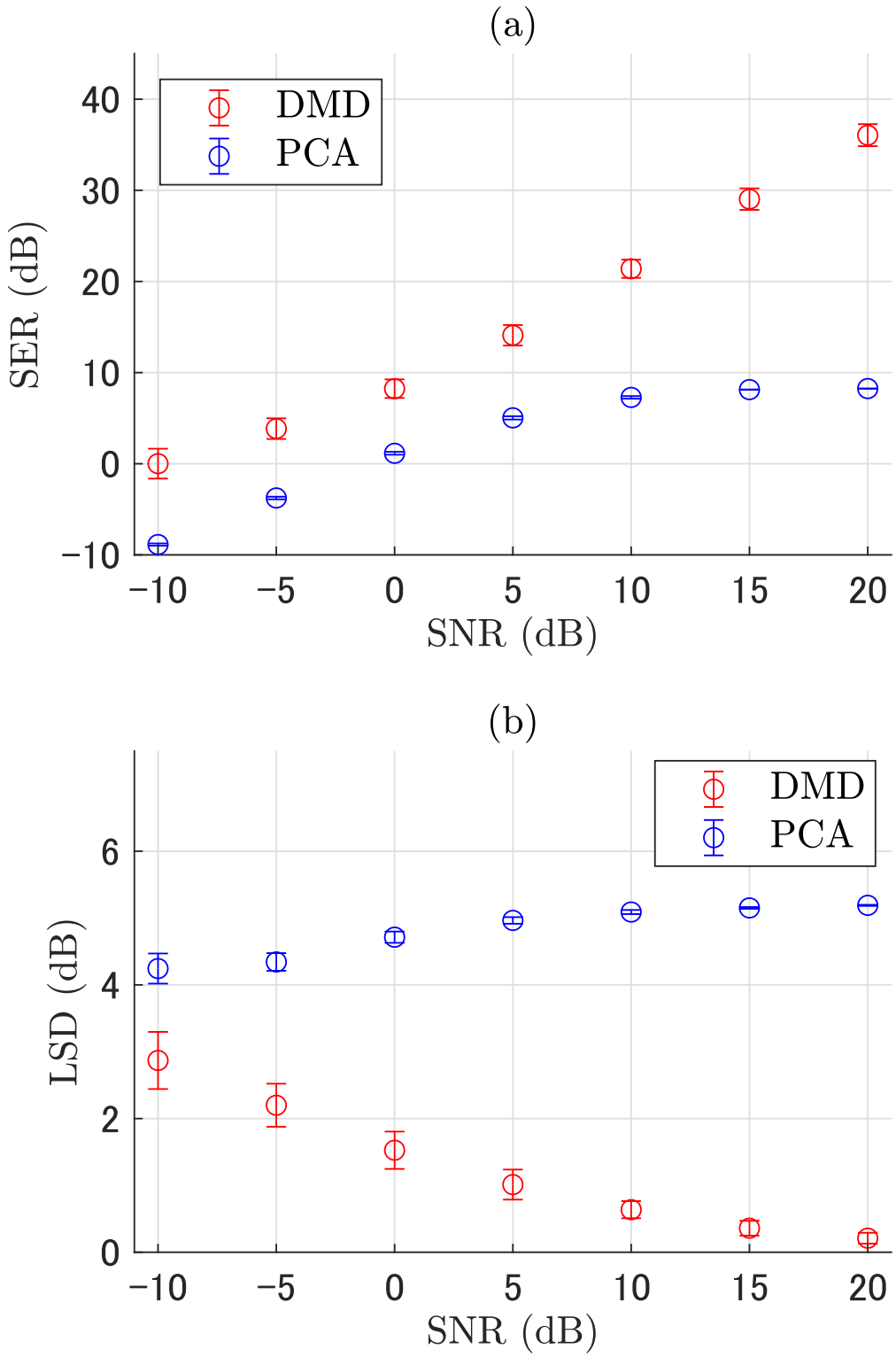

図 4.18: 減衰率 −13.8 の調波複合音とピンク雑音との混合信号における，DMD と PCA の音源分離シミュレーションの結果（入力行列の行数 400 の場合）: (a) 原信号と混合信号との SER, (b) 原信号と混合信号との LSD.

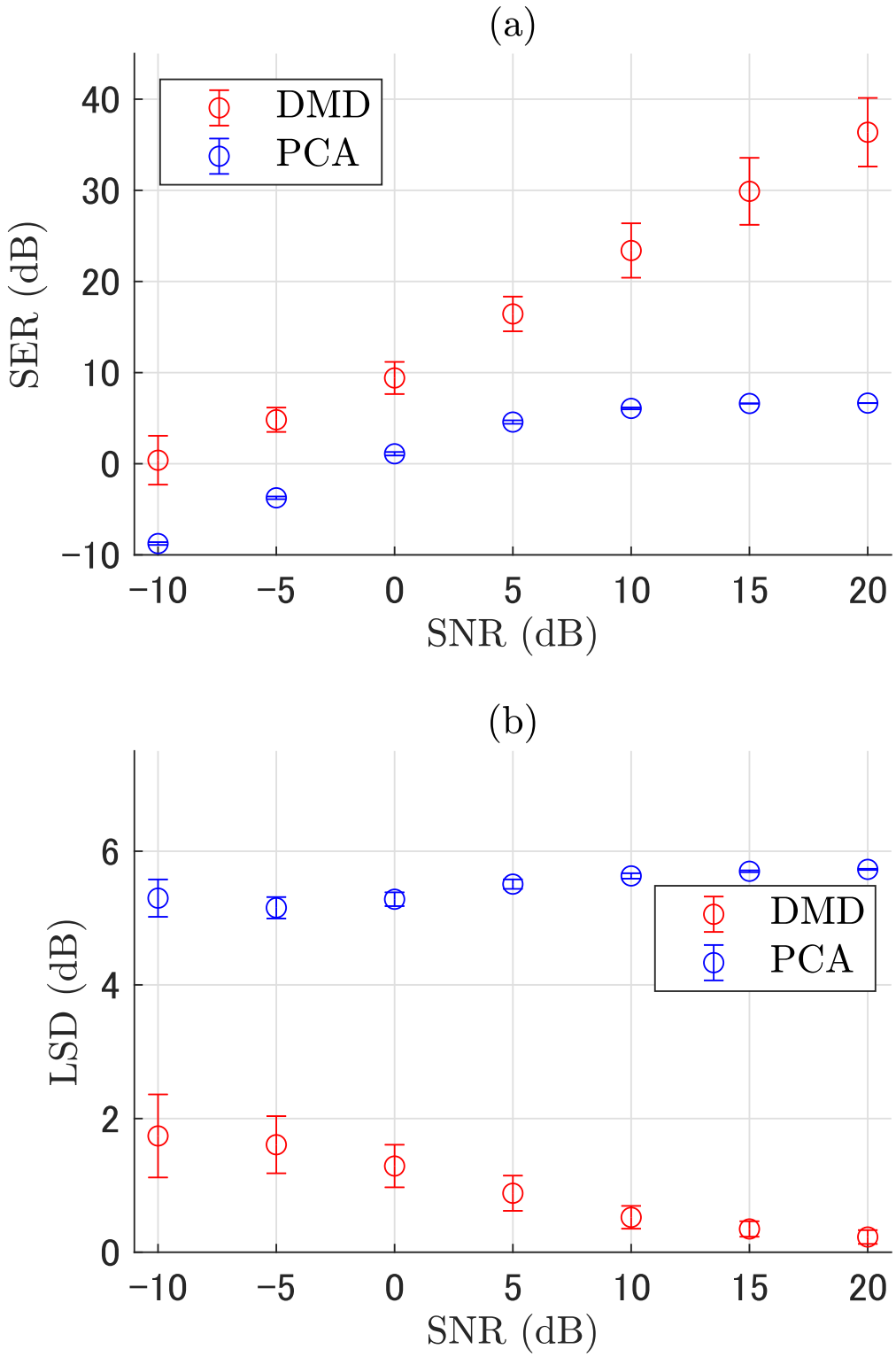

図 4.19: 減衰率 −13.8 の調波複合音とピンク雑音との混合信号における，DMD と PCA の音源分離シミュレーションの結果（入力行列の行数 600 の場合）：(a) 原信号と混合信号との SER, (b) 原信号と混合信号との LSD.

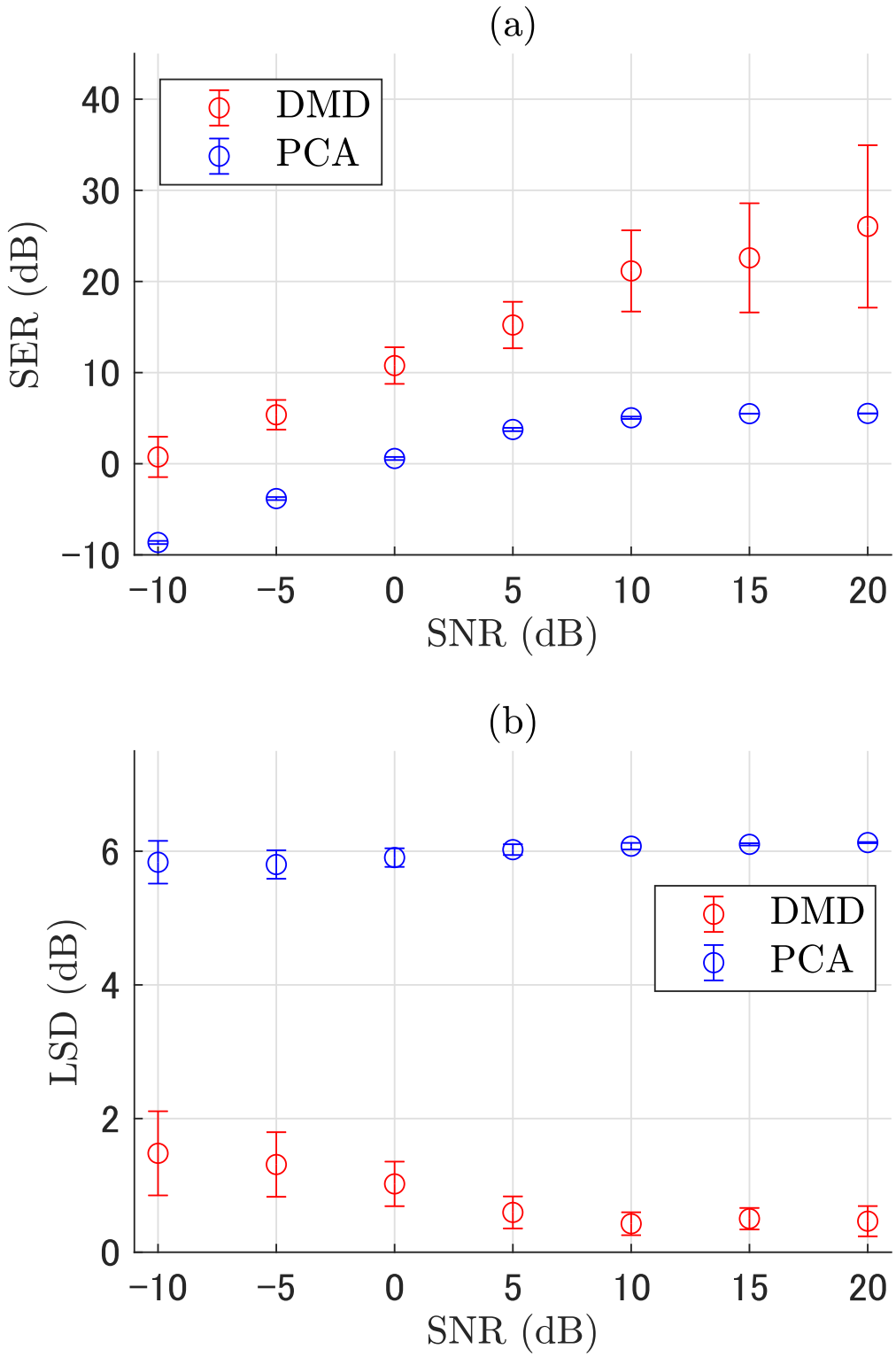

図 4.20: 減衰率 −13.8 の調波複合音とピンク雑音との混合信号における，DMD と PCA の音源分離シミュレーションの結果（入力行列の行数 800 の場合）: (a) 原信号と混合信号との SER, (b) 原信号と混合信号との LSD.

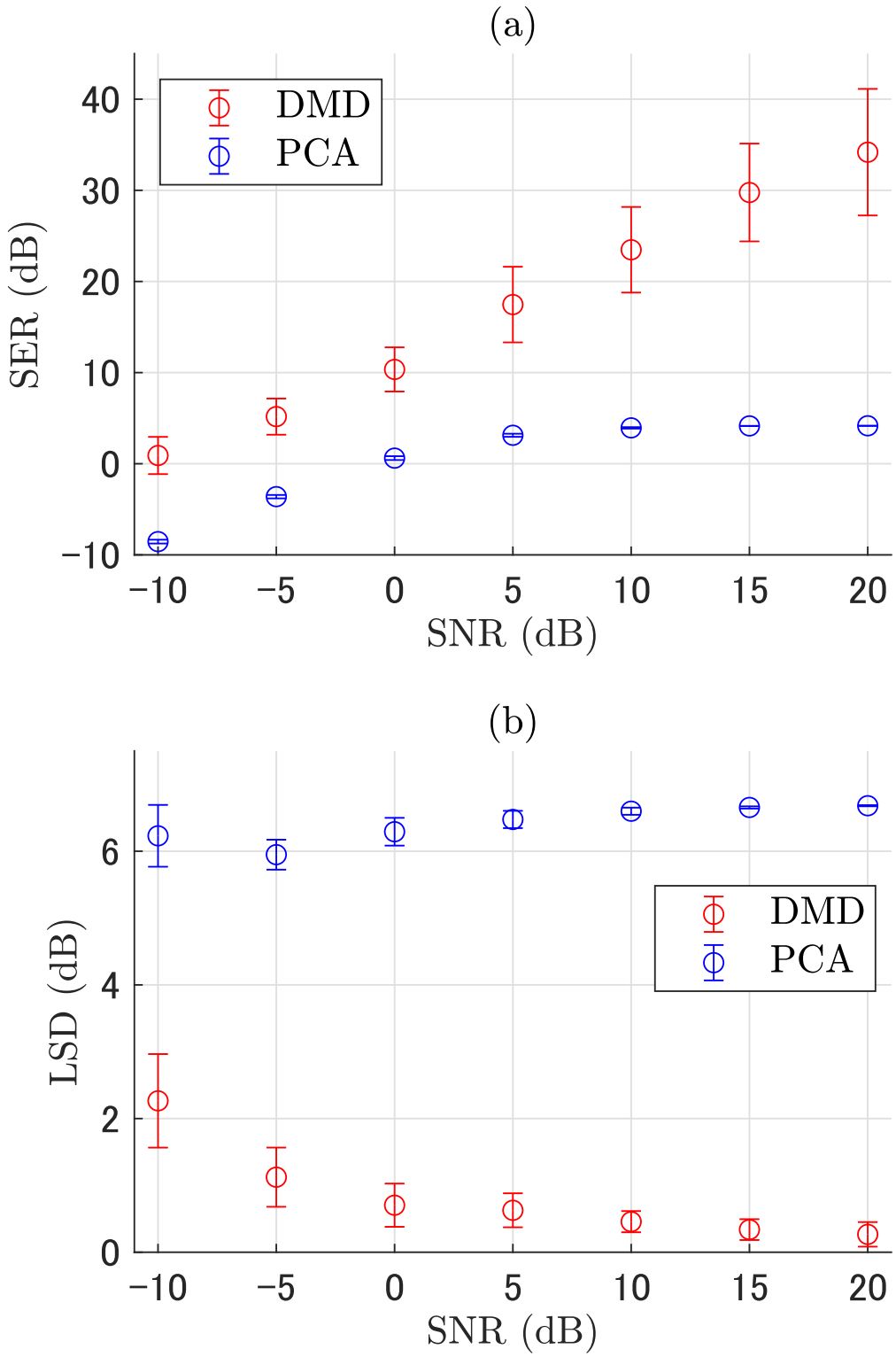

図 4.21: 減衰率 −13.8 の調波複合音とピンク雑音との混合信号における，DMD と PCA の音源分離シミュレーションの結果（入力行列の行数 1000 の場合）：(a) 原信号と混合信号との SER, (b) 原信号と混合信号との LSD.

第5章　考察

今回のシミュレーション結果について，分離性能と時間発展特徴の分布とを参照して考察する．

白色雑音，減衰率 -6.9, 入力行列の行数 $h = 800$，SNR が 20 dB の条件での混合信号の時間発展特徴の分布の一例を図 5.1 に，音響信号・雑音は同条件で SNR が -10 dB での混合信号の時間発展特徴の分布の一例を図 5.2 に示す．図中の赤色のプロット点は分離信号の再合成のために選択された要素である．図 5.1 を見ると，σ_j の値が -6.9 付近の時間発展特徴を選択できていることが読み取れる．これらの時間発展特徴の $\frac{\omega_j}{2\pi}$ は，概ね $\{-2000, -1900, \cdots, -100, 0, 100, \cdots, 1900, 2000\}$ の値を取っており，調波成分の周波数の抽出も精度良く行えていた．この混合信号での評価指標は，SER の平均値が 39.623 dB，LSD の平均値が 0.1469 dB となっており，同条件での PCA より高いことから，DMD の時間発展特徴に基づく音源分離は可能だと考えられる．一方，図 5.2 を見ると，$\frac{\omega_j}{2\pi}$ の値域として $\{-2000, \cdots, 2000\}$ の時間発展特徴が選択されており，調波複合音のもつ周波数成分に該当したモードが選択できたと考えられる．しかし，σ_j の値は -20 から -30 辺りでばらつきが見られる．音響信号に該当するモードと，それ以外のモードとで時間発展特徴の分布は確かに分離傾向が見られるものの，σ_j の値そのものは雑音の影響を受け，負方向に歪められてしまったと考えられる．このシミュレーションでの評価指標は，SER が 3.8232 dB，LSD が 2.6255 dB であり，歪められた σ_j の影響を受けたことで分離性能が低下したと考えられる．このため，低 SNR の条件下では，時間発展特徴によるモードの選択に加え，雑音強度に応じて時間発展特徴の補正も行う必要があると考えられる．LSD は SNR が 20 dB の条件より性能は低下したが，SNR が -10 dB の PCA での LSD 結果よりも良い結果を記録している．このことから，雑音による時間発展特徴の歪みは，周波数成分においては大きく影響しないと考察される．

ピンク雑音，減衰率 -6.9, 入力行列の行数 $h = 800$，SNR が 20 dB の条件での混合信号の時間発展特徴の分布の一例を図 5.3 に，音響信号・雑音は同条件で SNR が -10 dB での混合信号の時間発展特徴の分布の一例を図 5.4 に示す．図中の赤色のプロット点は，前例と同じく分離信号の再合成のために選択された要素である．雑音がピンク雑音の場合でも，高 SNR の条件では音響信号に該当する時間発展特徴をよく抽出できている．対して，低 SNR の条件では分布の分離傾向は見られるものの，σ_j の値は調波複合音の本来の減衰率より負方向に移動しており，歪みが生じている．またピンク雑音の場合，σ_j の歪み方は，$\frac{\omega_j}{2\pi}$ の絶対値が大きい成分は歪

みが小さく，絶対値が小さい成分は歪み方が大きくなる傾向が見られた．ピンク雑音は高い周波数成分より低い周波数成分のパワーが相対的に大きいため，低い周波数の時間発展特徴ほどピンク雑音の影響を大きく受けたものと考えられる．SNR が 20 dB での評価指標は，SER が 35.696 dB，LSD が 0.2174 dB であり，SNR が -10 dB での SER は 1.3804 dB，LSD は 2.2031 dB であった．LSD の結果は PCA よりも DMD の方が良いことから，DMD の分析は PCA と比べ，周波数領域の分析と成分抽出に有利であると考えられる．

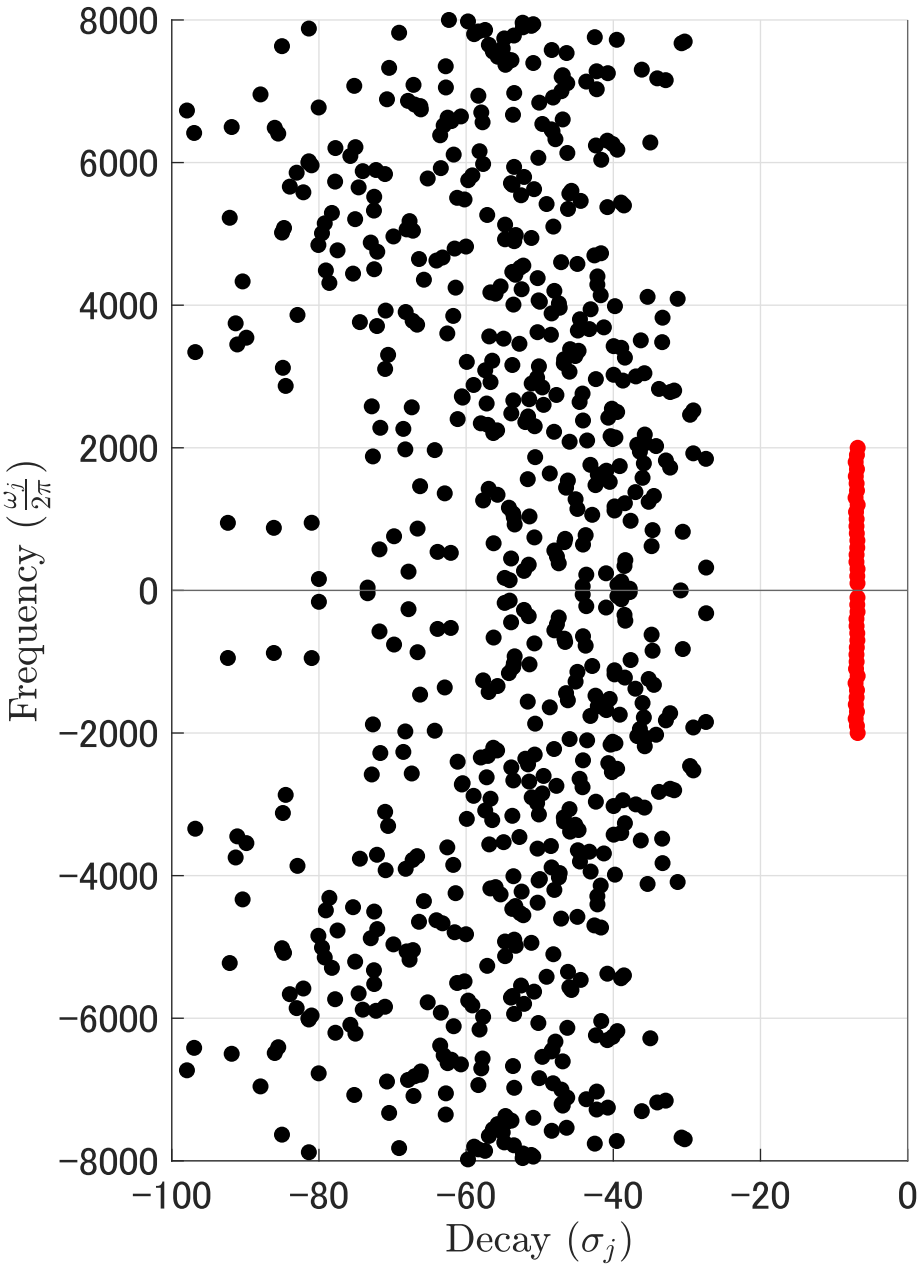

図 5.1: 減衰率 −6.9 の調波複合音と白色雑音との SNR が 20 dB の混合信号における，入力行列の行数 800 の場合に抽出された時間発展特徴の分布図.

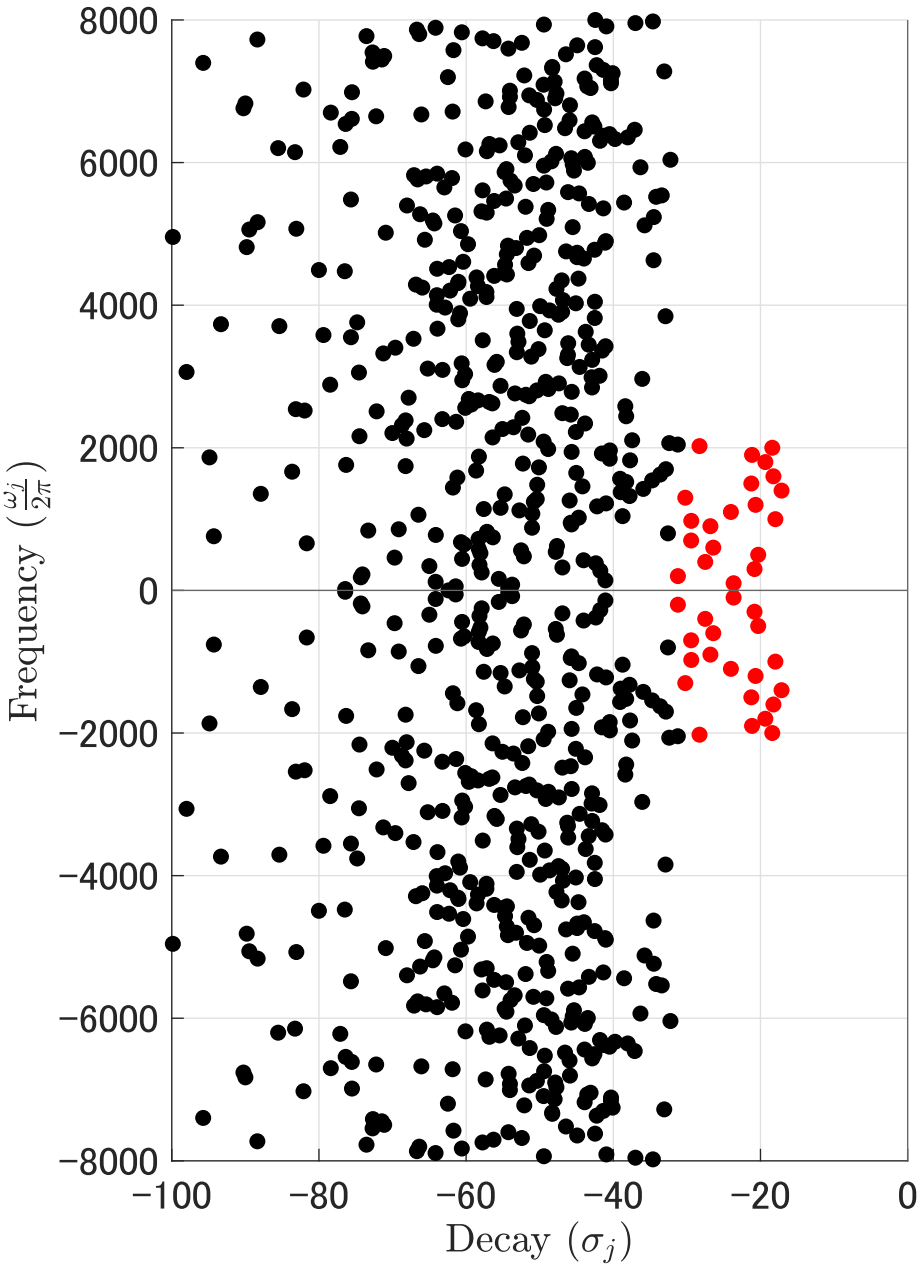

図 5.2: 減衰率 −6.9 の調波複合音と白色雑音との SNR が −10 dB の混合信号における，入力行列の行数 800 の場合に抽出された時間発展特徴の分布図.

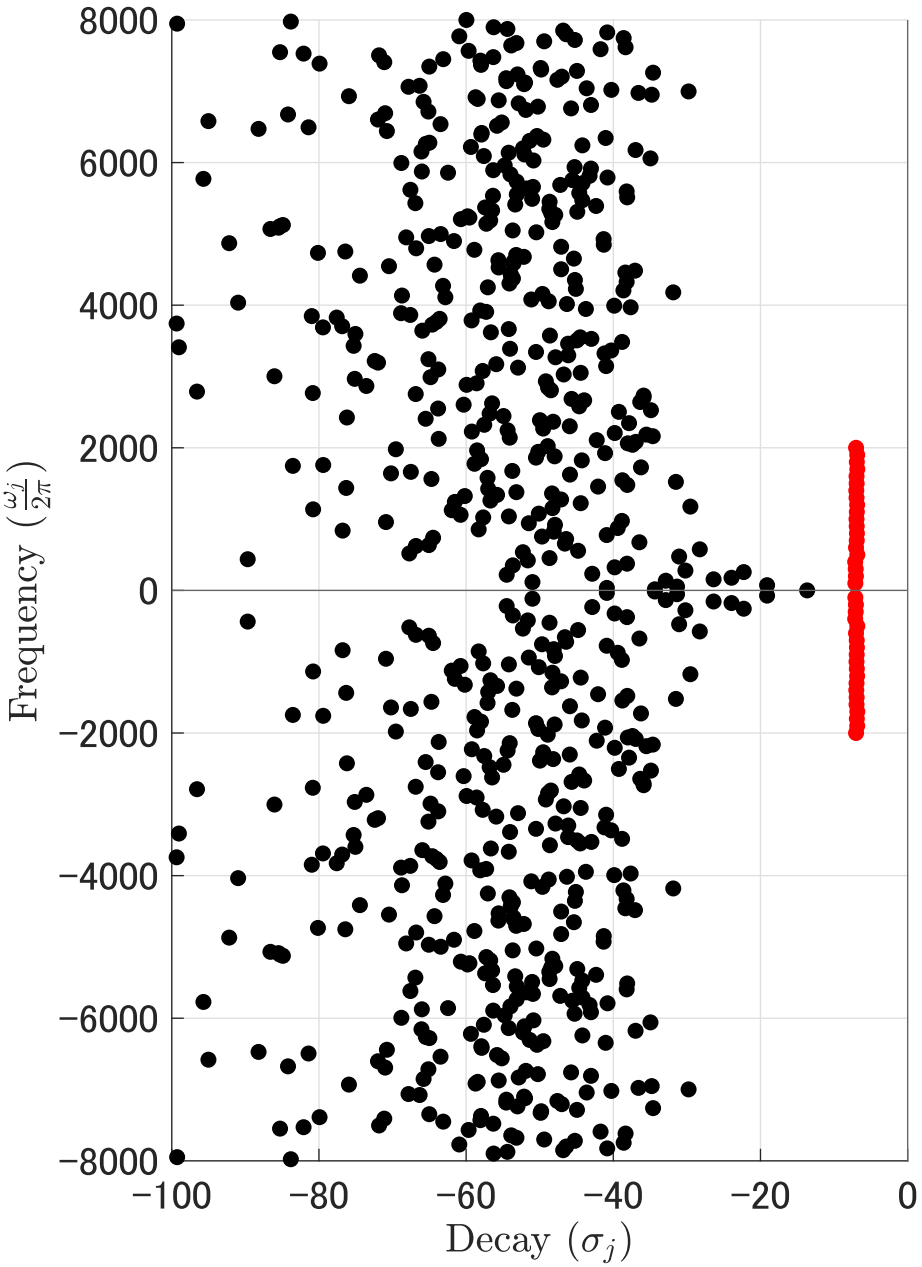

図 5.3: 減衰率 −6.9 の調波複合音とピンク雑音との SNR が 20 dB の混合信号における，入力行列の行数 800 の場合に抽出された時間発展特徴の分布図.

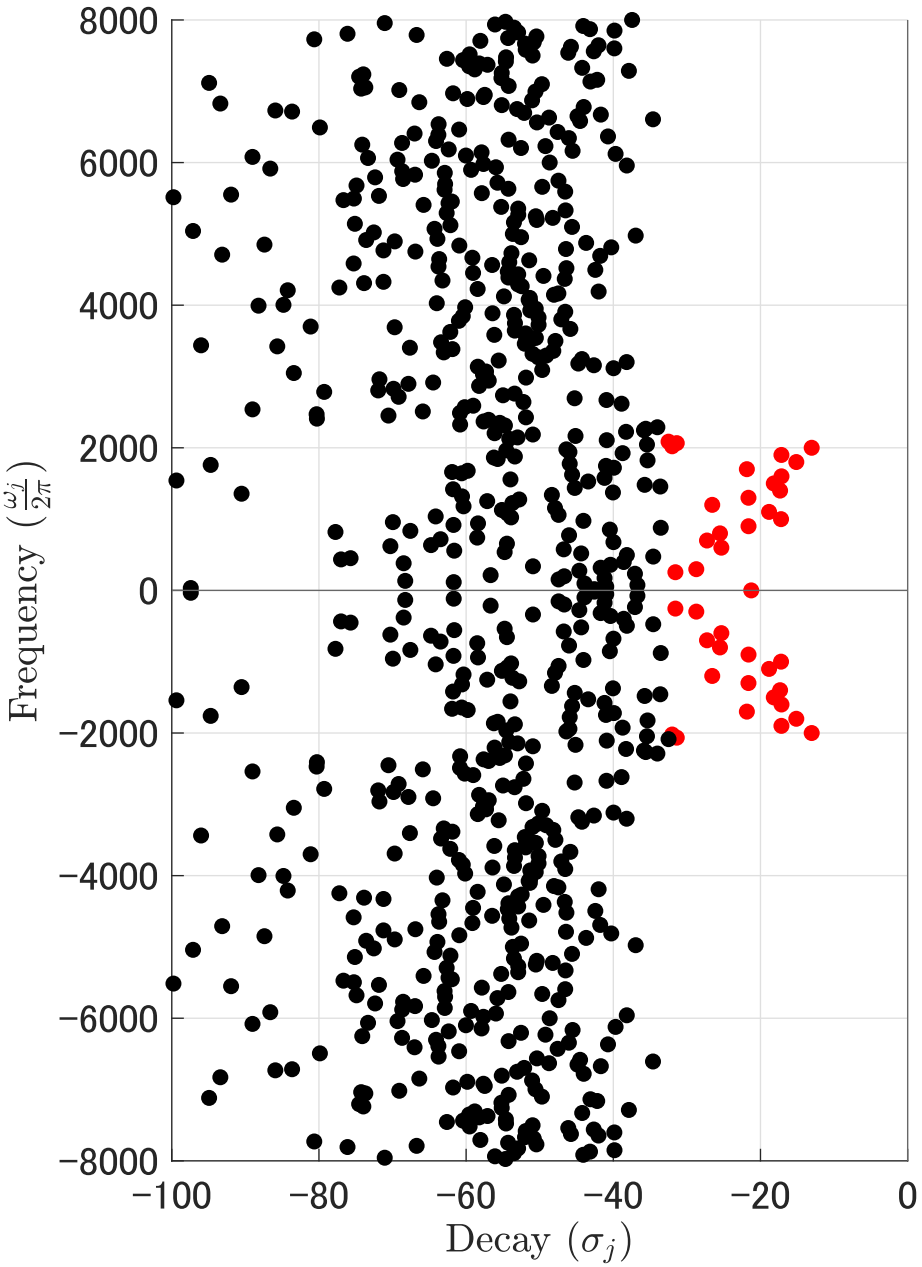

図 5.4: 減衰率 −6.9 の調波複合音とピンク雑音との SNR が −10 dB の混合信号における，入力行列の行数 800 の場合に抽出された時間発展特徴の分布図.

第6章　結論

6.1　本研究で明らかにしたこと

本研究では，DMD による音源分離の可能性を検討するため，先んじて DMD と音響信号との関係性について調査を行った．その結果，以下の知見が得られた．

- DMD によって抽出される時間発展特徴は，周波数と減衰率の 2 つのパラメータを持ち，解析対象の信号は DMD によって固有の周波数・減衰率をもつ成分の線形和として表現される．
- DMD 分析によって得られる時間発展特徴の分布は，音響信号と雑音とで大きく異なる．
- 音響信号と雑音の混合信号を DMD 分析した場合，時間発展特徴は音響信号のものとそれ以外とで分布の様子が分かれる．

以上の知見を利用し，DMD から抽出される時間発展特徴を活用した分離アルゴリズムを考案し，PCA の音源分離法と性能を比較した．その結果，以下のことが明らかになった．

- 一定比率で減衰する調波複合音について，白色雑音，及びピンク雑音を付加した場合，DMD の時間発展特徴の音源分離法は概ね PCA の音源分離法以上の精度で分離が可能である．
- 上記分離性能について，入力行列のサイズが小さく，低 SNR の条件下では，DMD の音源分離と PCA の音源分離とで差が生じない場合もあるため，DMD 分析のために適当な入力行列のサイズ設定が重要である．
- 低 SNR の条件下では，DMD の時間発展特徴に基づく分離操作は可能であるが，減衰率の歪みが生じる．

今回の検討から，適当な入力条件を設定すれば，一定比率で減衰する調波複合音を PCA を上回る性能で分離できることが明らかになった．この点で，DMD による音源分離の可能性はあると結論づける．

6.2 残された課題

本研究を進展させていく上で，以下の点が課題だと考える．

- 時間発展特徴の歪みと雑音の強度との関係の調査
- 周波数が時間的に変動する信号の分析
- 非定常雑音との音源分離の検討

1点目について，時間発展特徴のグルーピングによるモード選択は可能であるものの，時間発展特徴の値そのものは雑音の影響を受けて歪み，雑音のパワーが大きいとその歪みも大きくなることも明らかになった．雑音のパワーやそのスペクトルが，音響信号の時間発展特徴にどのような影響を及ぼすかを明らかにすることができれば，雑音の影響を加味した時間発展特徴の補正が可能となり，分離性能をより向上させることが可能だと考えられる．

2点目について，今回の検討では周波数成分が時間減衰するような信号を対象にし，DMD分析との関係性を明らかにした．しかし，例えばFM音のように，周波数成分が時間経過で高くなったり低くなったりする信号とDMDとの関係性は未知であり，検討の余地がある．周波数成分の局所的な時間変動を捉えるには，短時間フーリエ変換やウェーブレット変換のように，信号をフレームで区切って逐次分析していく方法が代表的であるが，これをDMDにも適用すれば，FM音の検討が可能になると考えられる．実際に，統計的性質が時間変化する時系列データに対しDMDを適用するための提案もされており[44][45]，これらのDMD応用手法を活用した非定常信号の分析も有効と考えられる．これを進めることで，将来的には，音声を対象としたDMDの分析と分離法への応用も検討できるものと考える．

3点目について，今日までの音源分離技術が抱える課題として，非定常信号どうしの分離がある．今回のシミュレーションでは非定常な人工音と定常雑音との分離を検討したが，これを非定常雑音としても分離が可能かどうか，検討する必要があると考える．これに向かい，非定常雑音とDMDの時間発展特徴との関係性を調査することが，今後必要であると考える．

謝辞

主指導教員の鵜木祐史教授には的確なご指導とご助言を数多く賜り，本研究の遂行に不可欠なものでした．深く感謝いたします．また，研究内容は勿論のこと，日々の研究活動についても様々なご助言を賜りました，赤木正人先生，木谷俊介講師，上江洲安史特任助教，大田恭士さん，磯山拓都さんに心から感謝いたします．また，中間審査会，修士論文審査会で的確なコメントとご助言を賜りました，岡田将吾准教授，吉高淳夫准教授に深く感謝いたします．そして，鵜木研究室の皆様には，公私共に大変お世話になりましたこと，感謝いたします．

最後に，学生生活を金銭面・精神面で応援し続けてくれた私の家族に，感謝を申し上げます．ありがとうございました．

参考文献

[1] 鹿野 清宏, 中村 哲, 伊勢 史郎, 音声・音情報のディジタル信号処理, 昭晃堂, 東京, 1997.

[2] 大串 健吾, 音響聴覚心理学, 誠信書房, 東京, 2019.

[3] 電子情報通信学会, 音響信号処理, 知識ベース 知識の森, 2 群, 6 編, 2012.

[4] 浅野 太, 音のアレイ信号処理―音源の定位・追跡と分離―, コロナ社, 東京, 2011.

[5] 牧野 昭二, 荒木 章子, 向井 良, 澤田 宏, “畳込み混合のブラインド音源分離（<特集>独立成分分析とその応用特集号）,” システム／制御／情報, Vol. 48, No. 10, pp. 401–408, 2004.

[6] 澤田 宏, 荒木 章子, 牧野 昭二, “音源分離技術の最新動向,” 電子情報通信学会誌, Vol. 91, No. 4, pp. 292–296, 04 2008.

[7] 戸上 真人, “音源分離技術の基礎と動向,” 電子情報通信学会 基礎・境界ソサイエティ Fundamentals Review, Vol. 16, No. 4, pp. 257–271, 2023.

[8] T. Jalal, T. Jalil, M. Nasser, S. Jinqiu, B. Vaclav, M. Rainer, “An evaluation of noise power spectral density estimation algorithms in adverse acoustic environments,” in: *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 4640–4643, 2011.

[9] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N. Yen, C. C. Tung, H. H. Liu, “The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis,” *Proceedings: Mathematical, Physical and Engineering Sciences*, Vol. 454, No. 1971, pp. 903–995, 1998.

[10] M. K. Molla, K. Hirose, N. Minematsu, “Robust voiced/unvoiced speech classification using empirical mode decomposition and periodic correlation model,” pp. 2530–2533, 2008.

[11] H. Taufiq, M. K. Hasan, “Suppression of residual noise from speech signals using empirical mode decomposition,” *IEEE Signal Processing Letters*, Vol. 16, No. 1, pp. 2–5, 2009.

[12] J. R. Hershey, C. Zhuo, R. J. Le, S. Watanabe, “Deep clustering: Discriminative embeddings for segmentation and separation,” in: *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 31–35, 2016.

[13] L. Yi, M. Nima, “Tasnet: Time-domain audio separation network for real-time, single-channel speech separation,” in: *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 696–700, 2018.

[14] 亀岡 弘和, “深層学習に基づく音源分離,” 日本音響学会誌, Vol. 75, No. 9, pp. 525–531, 2019.

[15] Y. Ephraim, H. L. V. Trees, “A signal subspace approach for speech enhancement,” *IEEE Transactions on Speech and Audio Processing*, Vol. 3, No. 4, pp. 251–266, 1995.

[16] H. Kris, “A review of signal subspace speech enhancement and its application to noise robust speech recognition,” *EURASIP Journal on Advances in Signal Processing*, Vol. 67, No. 12, pp. 1586–1604, 2006.

[17] J. N. Kutz, S. L. Brunton, B. W. Brunton, J. L. Proctor, *Dynamic Mode Decomposition*, Society for Industrial and Applied Mathematics, Philadelphia, PA, 2016.

[18] P. J. Schmid, “Dynamic mode decomposition of numerical and experimental data,” *Journal of Fluid Mechanics*, Vol. 656, pp. 5–28, 2010.

[19] J. H. Tu, C. W. Rowley, D. M. Luchtenburg, S. L. Brunton, J. N. Kutz, “On dynamic mode decomposition: Theory and applications,” *Journal of Computational Dynamics*, Vol. 1, No. 2, pp. 391–421, 2014.

[20] 大道 勇哉, 五十嵐 康彦, “動的モード分解による多次元時系列解析,” 日本神経回路学会誌, Vol. 25, No. 1, pp. 2–9, 2018.

[21] 神谷 俊輔, 大泉 匡史, “動的モード分解の概要と活用法—神経システムの制御問題への応用を目指して—,” 日本神経回路学会誌, Vol. 30, No. 2, pp. 73–83, 2023.

[22] J. Mann, J. N. Kutz, "Dynamic mode decomposition for financial trading strategies," *Quantitative Finance*, Vol. 16, No. 11, pp. 1643–1655, 2016.

[23] J. Grosek, J. N. Kutz, "Dynamic mode decomposition for real-time background/foreground separation in video," arXiv, 1404.7592, 2014.

[24] J. L. Proctor, P. A. Eckhoff, "Discovering dynamic patterns from infectious disease data using dynamic mode decomposition," *International Health*, Vol. 7, No. 2, pp. 139–145, 02 2015.

[25] 土肥 宏太, 武石 直也, 矢入 健久, 堀 浩一, "動的モード分解を用いた音響データの異常検知," 人工知能学会全国大会論文集, Vol. JSAI2018, pp. 1P202–1P202, 2018.

[26] 谷川 理佐子, 矢田部 浩平, 及川 靖広, "動的モード分解を用いた空力音と流れの解析," 日本音響学会講演論文集, pp. 607–608, 2020.

[27] T. Santosh, K. Samaneh, P. Norman, B. Miroslaw, W. David, "Dynamic mode decomposition for univariate time series: Analysing trends and forecasting," working paper or preprint, February 2017.

[28] P. Smaragdis, "Blind separation of convolved mixtures in the frequency domain," *Neurocomputing*, Vol. 22, No. 1, pp. 21–34, 1998.

[29] A. Hiroe, "Solution of permutation problem in frequency domain ica, using multivariate probability density functions," in: *Independent Component Analysis and Blind Signal Separation*, pp. 601–608, 2006.

[30] S. Mogami, H. Sumino, D. Kitamura, N. Takamune, S. Takamichi, H. Saruwatari, N. Ono, "Independent deeply learned matrix analysis for multichannel audio source separation," in: *2018 26th European Signal Processing Conference (EUSIPCO)*, pp. 1557–1561, 2018.

[31] S. Boll, "Suppression of acoustic noise in speech using spectral subtraction," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, Vol. 27, No. 2, pp. 113–120, 1979.

[32] J. S. Lim, A. V. Oppenheim, "Enhancement and bandwidth compression of noisy speech," *Proceedings of the IEEE*, Vol. 67, No. 12, pp. 1586–1604, 1979.

[33] K. Paliwal, A. Basu, "A speech enhancement method based on kalman filtering," in: *ICASSP '87. IEEE International Conference on Acoustics, Speech, and Signal Processing*, Vol. 12, pp. 177–180, 1987.

[34] 吉井 和佳, 糸山 克寿, “スパース性に基づく音楽音響信号の分解（小特集ースパース表現に基づく音響信号処理ー）,” 日本音響学会誌, Vol. 71, No. 11, pp. 607–614, 2015.

[35] V. Tuomas, “Monaural sound source separation by nonnegative matrix factorization with temporal continuity and sparseness criteria,” *IEEE Transactions on Audio, Speech, and Language Processing*, Vol. 15, No. 3, pp. 1066–1074, 2007.

[36] C. Angkana, R. C. Ann, “Singing voice separation for mono-channel music using non-negative matrix factorization,” in: *2008 International Conference on Advanced Technologies for Communications*, pp. 243–246, 2008.

[37] L. Daniel, S. H. Sebastian, “Algorithms for non-negative matrix factorization,” in: T. Leen, T. Dietterich, and V. Tresp, editors, *Advances in Neural Information Processing Systems*, Vol. 13. MIT Press, 2000.

[38] 亀岡 弘和, “非負値行列因子分解の音響信号処理への応用（<小特集>近年の音響信号処理における数理科学の進展）,” 日本音響学会誌, Vol. 68, No. 11, pp. 559–565, 2012.

[39] H. Sawada, N. Ono, H. Kameoka, D. Kitamura, H. Saruwatari, “A review of blind source separation methods: two converging routes to ilrma originating from ica and nmf,” *APSIPA Transactions on Signal and Information Processing*, Vol. 8, No. 1, pp. e12–e12, 2019.

[40] H. Dubey, A. Aazami, V. Gopal, B. Naderi, S. Braun, R. Cutler, A. Ju, M. Zohourian, M. Tang, H. Gamper, M. Golestaneh, R. Aichner, “Icassp 2023 deep noise suppression challenge,” arXiv, 2303.11510, 2023.

[41] R. Schmidt, “Multiple emitter location and signal parameter estimation,” *IEEE Transactions on Antennas and Propagation*, Vol. 34, No. 3, pp. 276–280, 1986.

[42] E. Vincent, H. Sawada, P. Bofill, S. Makino, J. P. Rosca, “First stereo audio source separation evaluation campaign: Data, algorithms and results,” in: *Independent Component Analysis and Signal Separation*, pp. 552–559, 2007.

[43] 福井 勝宏, 島内 末廣, 日岡 裕輔, 中川 朗, 羽田 陽一, 大室 仲, 片岡 章俊, “雑音抑圧のための雑音対信号比率に基づく雑音パワー推定,” *Journal of Signal Processing*, Vol. 18, No. 1, pp. 17–28, 2014.

[44] M. Senka, C. Nelida, M. Igor, “Koopman operator family spectrum for nonautonomous systems - part 1,” arXiv, 1703.07324, 2017.

[45] 武石, 直也, “時変動的モード分解,” 人工知能学会全国大会論文集, Vol. JSAI2019, No. 33, pp. 4P2J302–4P2J302, 2019.