香川高専

卒業研究論文

論文題目

多重解像度時間周波数表現に基づく独立低ランク行列分析

提出年月日	令和 4 年 2 月 25 日
学　　科	電気情報工学科
氏　　名	細谷 泰稚　　印
指導教員（主査）	北村 大地 講師　　印
副　　査	重田 和弘 教授　　印
学　科　長	辻 正敏 教授　　印

香川高等専門学校

Independent low-rank matrix analysis based on multi-resolution time-frequency representations

Taichi Hosotani

Department of Electrical and Computer Engineering
National Institute of Technology, Kagawa College

Abstract

Blind source separation (BSS) is a technique for estimating each original audio source from an observed signal, and independent low-rank matrix analysis (ILRMA) can achieve high performance. Optimization in ILRMA consists of estimating spatial and source models (training of spatial demixing filters and low-rank approximation of a sourcewise time-frequency structure). In conventional ILRMA, the same time-frequency resolutions are used for both models. However, it is reported that the separation performance of ILRMA strongly depends on the resolution of time-frequency representation. Thus, the performance of ILRMA should be improved by utilizing different resolutions of time-frequency representation for the both models. In this thesis, I propose a new ILRMA algorithm that introduces multiple time-frequency representations into spatial and source model. Also, I indicate experimentally that the same resolution of time-frequency representations in the both models does not always provide the best performance.

Keywords: blind source separation, independent low-rank matrix analysis, spectrogram resolutions, window function

（和訳）

ブラインド音源分離（BSS）は，複数の音源が混合した観測信号から混合前の個々の音源を推定する技術であり，独立低ランク行列分析（ILRMA）が高い分離性能を示している．ILRMA における最適化は，空間モデル推定（空間分離フィルタの学習）及び音源モデル推定（各音源の時間周波数構造の低ランク近似）からなる．従来の ILRMA では，各モデルの推定において同一の解像度の時間周波数表現を用いている．しかしながら，ILRMA の分離性能は時間周波数領域の解像度に強く依存することが報告されている．このことから，モデル毎に最適な解像度の時間周波数表現を用いた方が良い分離をもたらすと推測される．そこで，本論文では，ILRMA の空間モデル及び音源モデルの推定に異なる解像度の時間周波数表現を導入した手法を提案する．そして，各モデルの推定に同一の解像度の時間周波数表現を用いることが，必ずしも最大の分離性能を与えるとは限らないことを実験的に示す．

目次

第 1 章

緒言

1.1 本論文の背景

ブラインド音源分離（blind source separation: BSS）[1] は，混合系や音源情報が未知の条件下で，複数の音源が混合した観測信号から混合前の各音源信号を推定する技術である．Fig. 1.1 に BSS の概略図を示す．BSS は補聴器や自動採譜，音声認識等の様々な技術に応用されている．

チャネル数（観測に利用したマイクロホン数）が混合している音源数以上となる観測信号を扱う優決定条件 BSS は，音源間の統計的な独立性に基づいた独立成分分析（independent component analysis: ICA）[2, 3, 4] の登場以降，盛んに研究されている．ICA を周波数毎に適用することで耐残響性を高めた周波数領域独立成分分析（frequency-domain independent component analysis: FDICA）[5] が提案され，FDICA で推定される周波数毎の分離信号の順番を適切に並び替えるパーミュテーション問題の解決法が検討された [6, 7, 8]．2006 年には，FDICA に対して音源の時間周波数構造仮定を導入することで，パーミュテーション問題を回避しながら分離信号を推定する独立ベクトル分析（independent vector analysis: IVA）[9, 10, 11] が登場した．その後，補助関数法 [12, 13, 14] 及び反復射影法（iterative projection: IP）[15] に基づく安定かつ高速な IVA（auxiliary-function-based IVA: AuxIVA）[16] が提案されている．

2016 年には，パーミュテーション問題を回避するために何らかの音源モデルを導入するという IVA の画期的なアイデアを拡張して，非負値行列因子分解（nonnegative matrix factorization: NMF）[12, 17] に基づく低ランク時間周波数構造を ICA の音源モデルに取り入れた独立低ランク行列分析（independent low-rank matrix analysis: ILRMA）[1, 18, 19] が提案された．また，これらの音源モデルを plug-and-play で変更可能な最適化アルゴリズムを採用した BSS [20, 21, 22] も提案されている．さらに，時間周波数領域におけるスペクトログラム無矛盾性 [23, 24] と呼ばれる性質を FDICA 及び IVA に導入した BSS [25] や ILRMA に導入した BSS（consistent ILRMA）[26, 27] も提案されている．

ILRMA の最適化は，時間周波数領域における空間モデル（周波数毎の分離行列）の更新と音源モデル（NMF による低ランク時間周波数構造）の更新からなる．信号の時間周波数表現

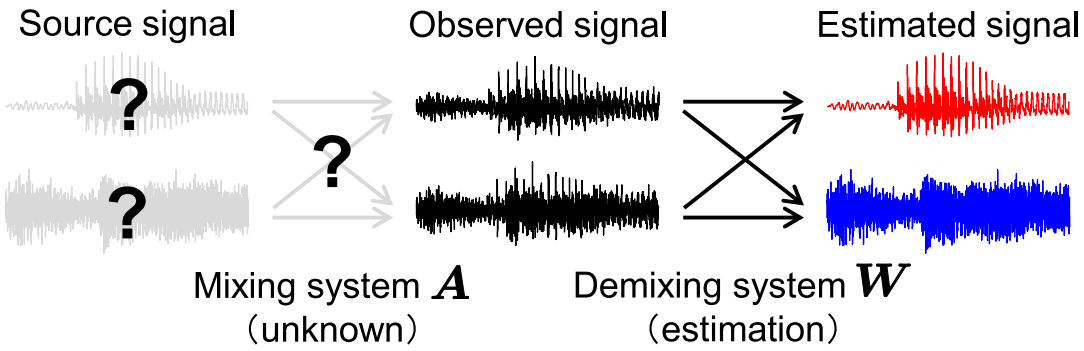


Fig. 1.1. Overview of BSS.

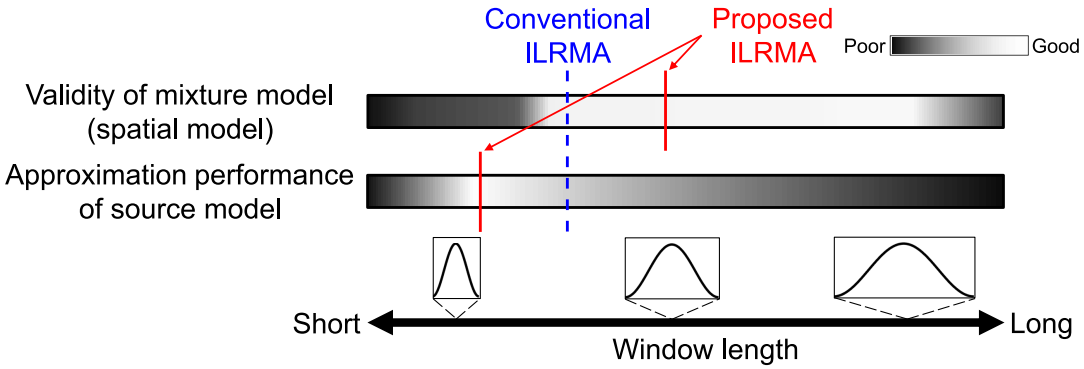


Fig. 1.2. Comparison between conventional and proposed ILRMA in terms of window length used in STFT. Conventional ILRMA uses same window length for both spatial and source models, whereas proposed ILRMA can set window length to optimal values in each of models.

は短時間 Fourier 変換（short-time Fourier transform: STFT）によって得ることができる．過去の実験的な調査 [28] により，ILRMA の分離性能は信号の時間周波数表現を求める際の STFT の窓長に，強く依存していることが分かっている．具体的には，STFT の窓長が極端に短い場合，周波数領域での瞬時混合仮定が成立せず性能が劣化 [29] し，逆に窓長が極端に長い場合，時間フレーム数が減少することによる統計的推定の不安定さから性能が劣化する．従って，STFT の窓長にはトレードオフが存在する．

1.2　本論文の目的

本論文では，1.1 節で述べた STFT の窓長に対する BSS 性能の傾向に基づき，ILRMA の性能向上を目的として，ILRMA における空間モデル及び音源モデルの最適化に，異なる解像度の時間周波数表現を導入したアルゴリズムを提案する．Fig. 1.2 に提案手法と従来手法の違いを表す図を示す．従来の ILRMA では，空間モデル及び音源モデルの最適化において，同一の窓長で STFT して得られる観測信号（すなわち，1 種類の解像度の時間周波数表現）のみ

を用いているため，Fig. 1.2 に示すように，両モデルの窓長が互いに一致する組しか選べないという制約がある．本論文では，空間モデル及び音源モデルにおける最適な窓長をそれぞれ設定できるアルゴリズムを提案する．これはすなわち，異なる複数の解像度（多重解像度）の時間周波数表現に基づく BSS であり，従来の ILRMA を時間周波数解像度という観点から一般化したアルゴリズムに対応している．また提案手法では，STFT による時間周波数解析時にパラメトリックな窓関数である Chebyshev 窓を用いる．これにより，時間周波数表現の（行列としての）サイズを変えることなく見かけ上の窓長を変化させることができ，シンプルな最適化アルゴリズムを得ることができる．実験では，各モデルの最適化における見かけ上の窓長を様々に変化させ，それぞれの窓長に対する音源分離性能の比較を行う．得られた実験結果から，従来手法の consistent ILRMA と提案手法の分離性能を比較し，各モデルの時間周波数表現の違いが分離性能に与える影響について調査する．

1.3 本論文の構成

2 章では，提案手法を理解する上で重要な，STFT や NMF 等の BSS の基礎技術及び ILRMA を含む従来手法について述べる．3 章では，ILRMA の空間モデルと音源モデルの最適化にそれぞれ異なる解像度の時間周波数表現を導入するという提案手法の動機及び詳細について述べる．具体的には，ILRMA の各モデルに異なる解像度の時間周波数表現を導入することがもたらす分離性能への影響，導入の際の問題点，及び問題に対する解決策について説明する．4 章では，提案手法の空間モデル及び音源モデルにおける窓長を変化させることで，従来手法である consistent ILRMA と提案手法の比較実験を行い，得られた結果の傾向について述べる．最後に，5 章で本論文の総括を行い，今後の課題を述べる．

第 2 章

BSS の基礎技術と従来手法

2.1 まえがき

本章では，BSS の基礎技術と従来手法について説明する．時間領域の信号を捉える上で，その信号を時間的に変化するスペクトルとして表現すること，すなわち時間周波数領域で表現することは非常に有効な手段である．特に，音源信号の混合は，時間領域で捉えるよりも時間周波数領域で捉える方が適している．このことから，2.2 節では，時間領域の信号から，その信号の時間周波数領域の表現を得る手法である STFT について取り上げる．また，2.3 節では，時間領域及び周波数領域における BSS の定式化を行う．さらに，2.4 節では，時間領域における音源分離手法である ICA 及び ICA を周波数領域に拡張した FDICA の概要を述べる．本論文で提案する手法の核となる ILRMA は，FDICA に対して，音源の時間周波数構造における低ランク性を導入した手法である．音源の時間周波数構造を低ランク近似する場合には，NMF という手法が有効である．2.5 節では，この NMF の概要を述べ，ILRMA で用いられる NMF の一種である ISNMF の定式化を行う．そして，2.6 節では，ILRMA の原理について述べる．提案手法との比較を行う consistent ILRMA は，ILRMA に対して，自然なスペクトログラムの持つ性質であるスペクトログラム無矛盾性を導入した手法である．そこで，2.7 節では，この consistent ILRMA の原理について述べる．最後に，2.8 節で本章の総括を行う．

2.2 STFT

STFT は，Fig. 2.1 に示すように，時間領域の信号から時間的に変化する音色（スペクトル）としての表現を得る手法である．STFT において，時間領域から周波数領域への変換（解析）時の窓関数の長さ（すなわち，短時間区間信号の長さ）及びシフト長をそれぞれ Q 及び τ とする．このとき，時間領域の信号 $\mathbf{z} = [\, z[1], z[2], \ldots, z[l], \ldots, z[L] \,]^{\mathrm{T}} \in \mathbb{R}^{L}$ の j 番目の短時間区間（時間フレーム）信号は次式で表される．

$$\begin{aligned} \mathbf{z}^{[j]} &= \big[\, z[(j-1)\tau+1], z[(j-1)\tau+2], \ldots, z[(j-1)\tau+Q] \,\big]^{\mathrm{T}} \\ &= \big[\, z^{[j]}[1], z^{[j]}[2], \ldots, z^{[j]}[q], \ldots, z^{[j]}[Q] \,\big]^{\mathrm{T}} \in \mathbb{R}^{Q} \end{aligned} \tag{2.1}$$

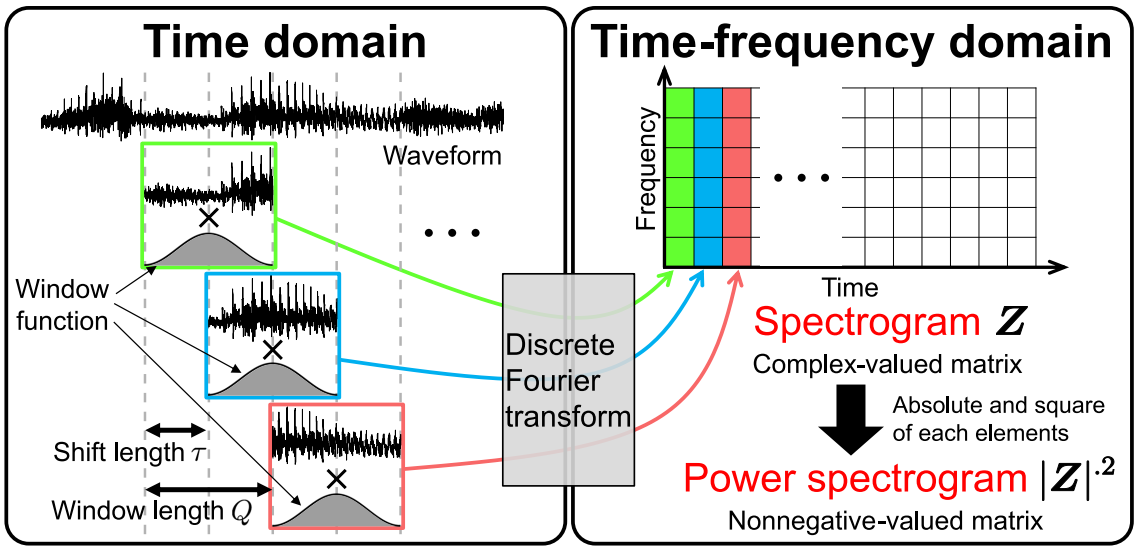


Fig. 2.1. Mechanism of STFT.

ここで，$l = 1, 2, \ldots, L$，$j = 1, 2, \ldots, J$，及び $q = 1, 2, \ldots, Q$ はそれぞれ離散時間のインデクス，時間フレーム，及び時間フレーム内のサンプルのインデクスであり，$\cdot^{\mathrm{T}}$ は転置を表す．なお，本論文では，信号の時間領域及び時間周波数領域における二つの表現の混同を避けるため，信号の時間領域のベクトルや行列をローマン体，信号の時間周波数領域におけるベクトルや行列をイタリック体でそれぞれ示す．また，時間フレーム数 J は次式で与えられる．

$$J = \frac{L}{\tau} \tag{2.2}$$

ただし，J は正の整数となる必要があるため，信号長 L がシフト長 τ で割り切れるように信号をゼロ埋めする．そして，信号 $\mathbf{z}$ の STFT によって得られた複素スペクトログラム $\boldsymbol{Z}$ を次式で表記する．

$$\boldsymbol{Z} = \mathrm{STFT}_{\boldsymbol{\omega}}(\mathbf{z}) \quad \in \mathbb{C}^{I \times J} \tag{2.3}$$

ここで，$\boldsymbol{\omega} = [\omega[1], \omega[2], \ldots, \omega[q], \ldots, \omega[Q]]^{\mathrm{T}} \in \mathbb{R}^{Q}$ は STFT で用いる解析時の窓関数を表す．このとき，複素スペクトログラム $\boldsymbol{Z}$ の (i, j) 成分 z_{ij} は次式で表される．

$$z_{ij} = \sum_{q=1}^{Q} z^{[j]}[q]\omega[q] \exp\left\{-\frac{2\pi\imath(q-1)(i-1)}{F}\right\} \tag{2.4}$$

ここで，F は窓長 Q 以下の正の整数を，$i = 1, 2, \ldots, I$ は周波数ビンのインデクスを，$\imath$ は虚数単位を示している．また，周波数ビン数 I は次式で与えられる．

$$I = \left\lfloor \frac{F}{2} \right\rfloor + 1 \tag{2.5}$$

なお，$\lfloor \cdot \rfloor$ は床関数を表す．本稿では，$F = Q$ として扱う．このとき，周波数ビン数 I は窓長 Q に依存する．

周波数領域から時間領域への変換（合成）時の窓関数を $\widetilde{\boldsymbol{\omega}}$ とおくとき，逆 STFT を $\mathrm{ISTFT}_{\widetilde{\boldsymbol{\omega}}}(\cdot)$ と表記する．本論文では，$\boldsymbol{\omega}$ と $\widetilde{\boldsymbol{\omega}}$ のペアが次式の完全再構成条件を満たすことを仮定する．

$$\mathbf{z} = \mathrm{ISTFT}_{\widetilde{\boldsymbol{\omega}}}(\mathrm{STFT}_{\boldsymbol{\omega}}(\mathbf{z})) \qquad \forall \mathbf{z} \in \mathbb{R}^L \tag{2.6}$$

ここで，$\mathrm{ISTFT}_{\widetilde{\boldsymbol{\omega}}}(\boldsymbol{Z})$ は複素スペクトログラム $\boldsymbol{Z}$ を時間信号 $\mathbf{z}$ に戻す逆 STFT を表す．なお，$\boldsymbol{\omega}$ と $\widetilde{\boldsymbol{\omega}}$ のペアが式 (2.6) を満たすとき，τ の上限は窓関数 $\boldsymbol{\omega}$ の情報（窓関数の各点での値と窓長）によって定まる[*1]．このことから，周波数ビン数 J は窓長 Q によって制限される．

式 (2.4) に示すように，STFT は，一定時間毎に信号を切り出し，それぞれの区間信号に解析窓関数 $\boldsymbol{\omega}$ を乗じて，離散 Fourier 変換（discrete Fourier transform: DFT）を施すという処理からなる．STFT を適用して得られた複素スペクトログラム $\boldsymbol{Z}$ は，行が周波数，列が時間の複素行列として表すことができる．また，音響信号処理では各時間周波数成分の大きさのみを取り扱うことも多い．その場合は，複素スペクトログラム $\boldsymbol{Z}$ の各要素に関して絶対値を取った振幅スペクトログラム $|\boldsymbol{Z}| \in \mathbb{R}_{\geq 0}^{I \times J}$ や，絶対値の 2 乗をとったパワースペクトログラム $|\boldsymbol{Z}|^{\cdot 2} \in \mathbb{R}_{\geq 0}^{I \times J}$ を処理の対象とする．ここで，行列に対する絶対値記号及びドット付き指数乗はそれぞれ要素毎の絶対値及び要素毎の指数乗を表す．例として，Figs. 2.2(a) 及び (b) にそれぞれ音楽信号及び音声信号のパワースペクトログラムを示す．図中の色の変化は，青色に近づくほどパワーが小さく，黄色に近づくほどパワーが大きいことを表している．

2.3　時間領域及び周波数領域における BSS の定式化

N 個の音源信号が M 個のマイクロホンで観測される状況を考える．この状況での信号の残響長を L' とする．n 番目のチャネルの音源信号，m 番目のチャネルの観測信号，及び n 番目のチャネルの分離信号をそれぞれ次式で表す．

$$\mathbf{s}_n = \left[\, s_n[1], \ldots, s_n[l], \ldots, s_n[L] \,\right]^{\mathrm{T}} \in \mathbb{R}^L \tag{2.7}$$

$$\mathbf{x}_m = \left[\, x_m[1], \ldots, x_m[l], \ldots, x_m[L] \,\right]^{\mathrm{T}} \in \mathbb{R}^L \tag{2.8}$$

$$\mathbf{y}_n = \left[\, y_n[1], \ldots, y_n[l], \ldots, y_n[L] \,\right]^{\mathrm{T}} \in \mathbb{R}^L \tag{2.9}$$

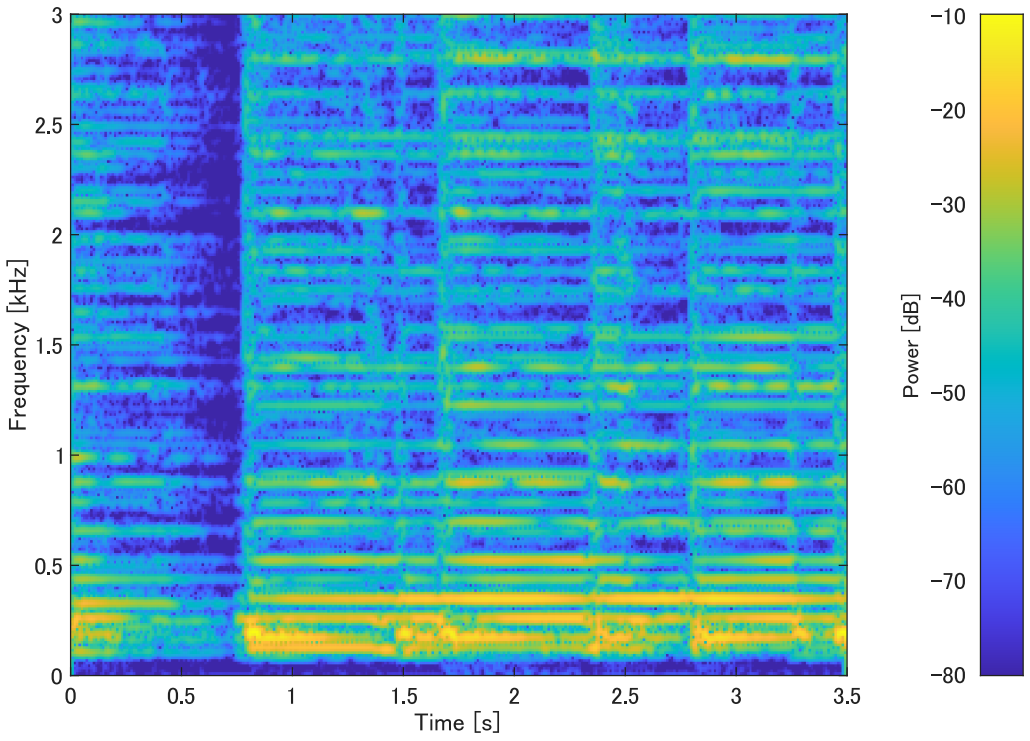
また，時間領域の多チャネルの音源信号，観測信号，及び分離信号をそれぞれ次式で表す．

$$\mathbf{s}[l] = \left[\, s_1[l], \ldots, s_n[l], \ldots, s_N[l] \,\right]^{\mathrm{T}} \in \mathbb{R}^N \tag{2.10}$$

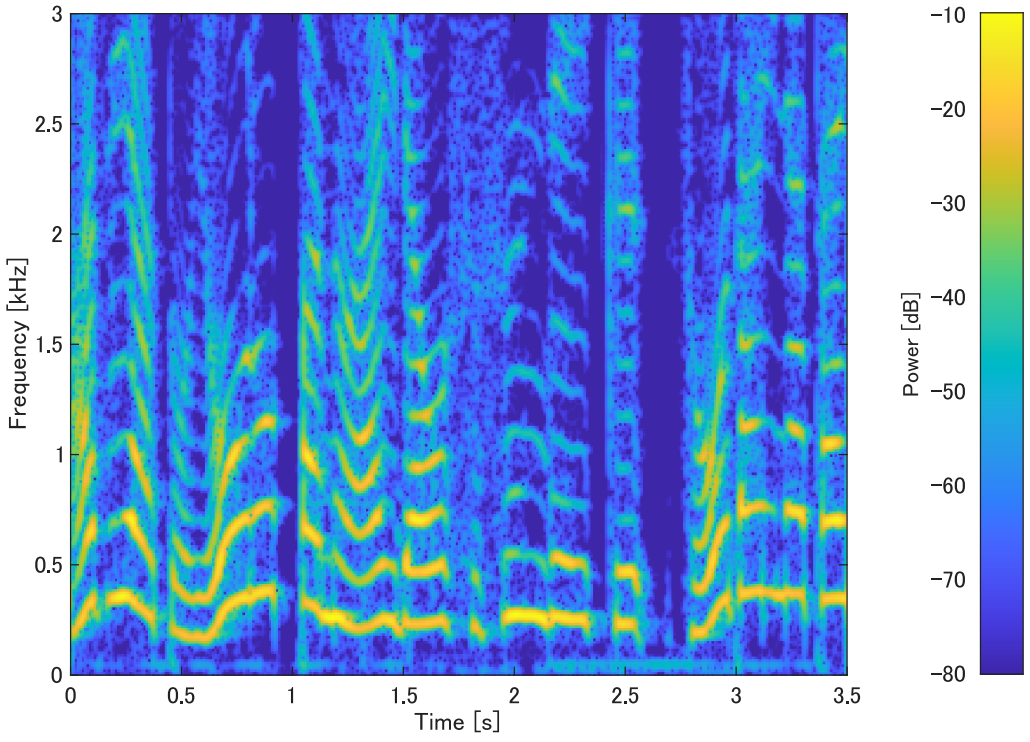
$$\mathbf{x}[l] = \left[\, x_1[l], \ldots, x_m[l], \ldots, x_M[l] \,\right]^{\mathrm{T}} \in \mathbb{R}^M \tag{2.11}$$

$$\mathbf{y}[l] = \left[\, y_1[l], \ldots, y_n[l], \ldots, y_N[l] \,\right]^{\mathrm{T}} \in \mathbb{R}^N \tag{2.12}$$

[*1] STFT を適用する際に情報の欠落が生じていない場合に，時間信号の再構成が可能となる．情報の欠落が生じないようにシフト長 τ を定める必要がある．



(a)



(b)

Fig. 2.2. Power spectrogram of (a) music and (b) speech signals.

ここで，$n = 1, 2, \ldots, N$ 及び $m = 1, 2, \ldots, M$ はそれぞれ音源及びチャネルのインデクスを表す．さらに，各音源信号に対する多チャネルの観測信号及び混合係数をそれぞれ次式で表す．

$$\mathbf{c}_n[l] = \left[\, c_{n1}[l], \ldots, c_{nm}[l], \ldots, c_{nM}[l] \,\right]^{\mathrm{T}} \in \mathbb{R}^M \tag{2.13}$$

$$\mathbf{a}_n[l'] = \left[\, a_{n1}[l'], \ldots, a_{nm}[l'], \ldots, a_{nM}[l'] \,\right]^{\mathrm{T}} \in \mathbb{R}^M \tag{2.14}$$

ただし，$l' = 0, 1, \ldots, L' - 1$ は残響時間のインデクスである．

各チャネルに STFT を適用して得られる音源信号，観測信号，分離信号，及び各音源信号に対する観測信号のスペクトログラムの (i, j) 番目の要素をそれぞれ次式で表す．

$$\boldsymbol{s}_{ij} = [\, s_{ij1}, \ldots, s_{ijn}, \ldots, s_{ijN} \,]^{\mathrm{T}} \in \mathbb{C}^N \tag{2.15}$$

$$\boldsymbol{x}_{ij} = [\, x_{ij1}, \ldots, x_{ijm}, \ldots, x_{ijM} \,]^{\mathrm{T}} \in \mathbb{C}^M \tag{2.16}$$

$$\boldsymbol{y}_{ij} = [\, y_{ij1}, \ldots, y_{ijn}, \ldots, y_{ijN} \,]^{\mathrm{T}} \in \mathbb{C}^N \tag{2.17}$$

$$\boldsymbol{c}_{ijn} = [\, c_{ijn1}, \ldots, c_{ijnm}, \ldots, c_{ijnM} \,]^{\mathrm{T}} \in \mathbb{C}^M \tag{2.18}$$

また，音源信号，観測信号及び分離信号に対して，各チャネルの時間周波数行列（スペクトログラム）の表記を $\boldsymbol{S}_n \in \mathbb{C}^{I \times J}$，$\boldsymbol{X}_m \in \mathbb{C}^{I \times J}$ 及び $\boldsymbol{Y}_n \in \mathbb{C}^{I \times J}$ と定義する．

無響室等の特殊な環境を除き，実際の音響信号の収録環境では音源の混合において残響が発生する．このような音源の混合は，次式で表される畳み込み混合となる．

$$\mathbf{x}[l] = \sum_{n=1}^{N} \mathbf{c}_n[l] = \sum_{n=1}^{N} \sum_{l'=0}^{\min(L', l)-1} \mathbf{a}_n[l'] s_n[l - l'] \tag{2.19}$$

なお，残響が生じない環境では，式 (2.19) で $L' = 1$ とした混合となる．

混合系の残響時間 L' が STFT における窓長 Q よりも十分短い場合，(2.19) は STFT によって，近似的に周波数領域での瞬時混合へ変換される．ILRMA を含む周波数領域 BSS では，この条件を仮定することで，時間領域での畳み込み混合を次式に示す周波数領域での瞬時混合として扱う．

$$\boldsymbol{x}_{ij} = \sum_{n=1}^{N} \boldsymbol{c}_{ijn} \approx \sum_{n=1}^{N} \boldsymbol{a}_{in} s_{ijn} \tag{2.20}$$

ここで，$\boldsymbol{a}_{in} = [\, a_{in1}, \ldots, a_{inm}, \ldots, c_{inM} \,]^{\mathrm{T}} \in \mathbb{C}^M$ は各音源信号についての混合係数に対して，各チャネルに Q 点 DFT を適用して得られるベクトル[*2]であり，次式で与えられる．

$$a_{inm} = \sum_{q=1}^{Q} a_{nm}[q] \exp\left\{ -\frac{2\pi\imath(q-1)(i-1)}{F} \right\} \tag{2.21}$$

なお，上式において，$a_{nm}[q]$ が範囲外の参照となる場合，すなわち $q \geq L'$ となる場合については，$a_{nm}[q] = 0$ として扱う．いま，周波数毎の混合行列を $\boldsymbol{A}_i = [\, \boldsymbol{a}_{i1}, \ldots, \boldsymbol{a}_{in}, \ldots, \boldsymbol{a}_{iN} \,] \in$

[*2] 正確には，式 (2.4) と同様に，Q 点 DFT を適用して得られたベクトルから対称な成分を取り除いたものである．

$\mathbb{C}^{M\times N}$ とすると，式 (2.20) は次式のように書ける．

$$\boldsymbol{x}_{ij} = \boldsymbol{A}_i \boldsymbol{s}_{ij} \tag{2.22}$$

優決定条件 BSS では $M = N$ を仮定でき，BSS は $\boldsymbol{A}_i$ の逆行列を推定する問題となる．この逆行列を $\boldsymbol{W}_i \approx \boldsymbol{A}_i^{-1}$ とすると，分離信号は次式となる．

$$\boldsymbol{y}_{ij} = \boldsymbol{W}_i \boldsymbol{x}_{ij} \tag{2.23}$$

ここで，$\boldsymbol{W}_i = [\,\boldsymbol{w}_{i1}, \ldots, \boldsymbol{w}_{in}, \ldots, \boldsymbol{w}_{iN}\,]^{\mathrm{H}} \in \mathbb{C}^{N\times M}$ は分離行列と呼ばれ，$^{\mathrm{H}}$ はエルミート転置を表す．

2.4 ICA 及び FDICA の概要

統計数理アルゴリズムである ICA は，時間領域での瞬時混合及び信号の統計的性質を仮定した BSS である．ICA では，前述の仮定を用いて，瞬時混合系の逆系を推定することで，音源分離を行っている．

ICA が仮定する統計的性質とは，音源信号間の統計的独立性及び音源信号が従う生成モデルの非ガウス性である．信号のチャネルの順序及びスケール（大きさ）の違いはこれらの性質に影響を与えない．従って，ICA によって推定される分離信号には，以下の任意性が存在する．

1. 分離信号のチャネルの順序には任意性がある
2. 分離信号のスケールには任意性がある

これらの任意性は分離信号に対して Fig. 2.3 のように現れる．上記の任意性 1 より，元々の信号源の順序が入れ替わる可能性がある．また，任意性 2 より，分離信号のスケールが混合前の音源信号のスケールから変化してしまう可能性がある．なお，信号のスケールの任意性に関しては，プロジェクションバック（projection back: PB）法 [30] と呼ばれる補正方法が提案されている．さらに，ICA には上記の問題に加えて，残響の生じた混合信号に対する音源分離性能が著しく悪化するという問題がある．これは，2.3 節で述べたように，残響が生じた音源の混合は瞬時混合ではなく畳み込み混合となり，ICA における時間領域での瞬時混合仮定が成り立たないことに起因する．

残響を含む信号に対する音源分離は，畳み込み混合系の逆系を推定することで達成できる．しかし一般に，時間領域における畳み込み混合系の逆系の推定は困難である．この問題は，2.2 節で述べた STFT を用いることで解決できる．畳み込みは STFT によって積和に変換されるため，時間領域での畳み込み混合を式 (2.22) で表される時間周波数領域での瞬時混合として扱うことが可能となる．このことを利用して，観測信号を STFT して得られたスペクトログラムの各周波数ビンの複素時系列信号 $\boldsymbol{x}_{i1}, \boldsymbol{x}_{i2}, \ldots, \boldsymbol{x}_{ij}, \ldots, \boldsymbol{x}_{iJ}$ に対して，独立な ICA を適用し，分離信号のスペクトログラムを推定する手法が提案されている．この手法は FDICA と呼ばれる．前述した通り，ICA の分離信号にはスケールや順序の任意性がある．FDICA では周波数毎に独立な ICA が適用されるため，各周波数ビンで推定した分離信号のスペクトログ

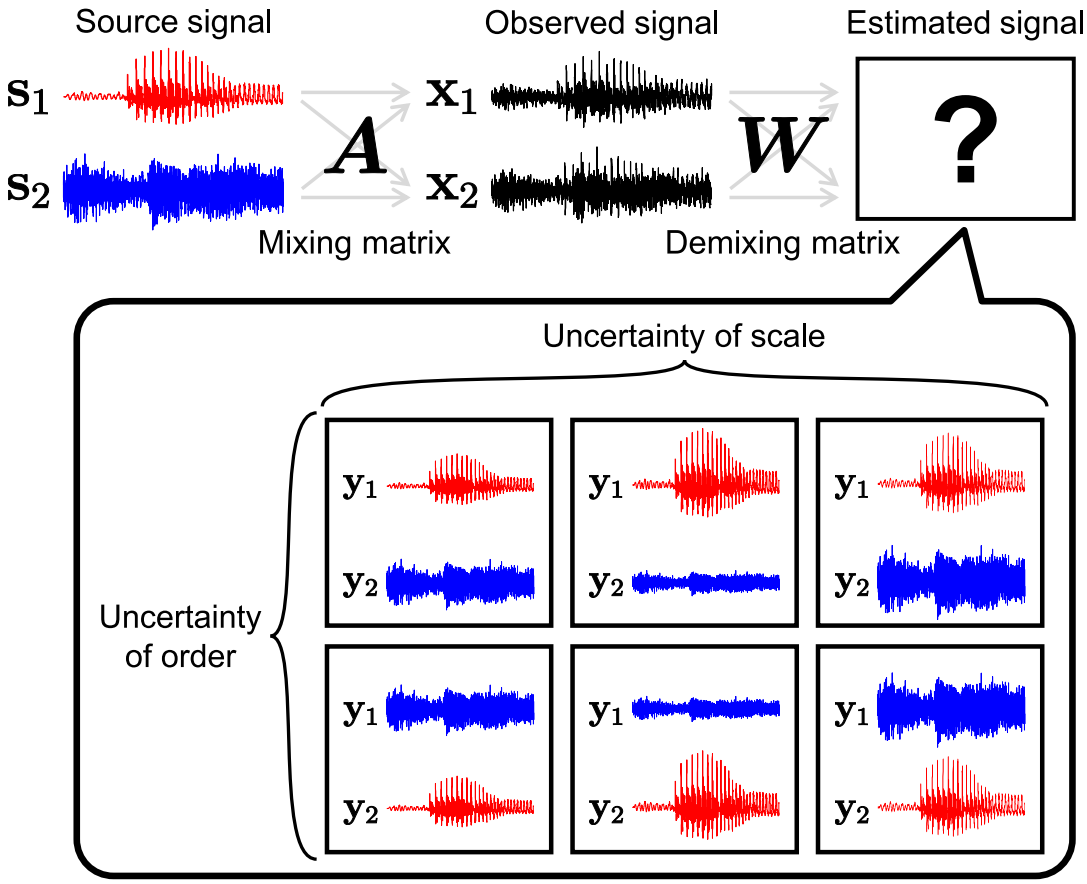


Fig. 2.3. Uncertainty in ICA. ICA cannot determine order and scales of estimated signals.

ラムの並びとスケールがバラバラになるという問題が生じる．周波数毎のスケールの任意性については，前述の PB 法により解析的に復元可能である．それに対し，Fig. 2.4 に示すような周波数毎の順序の任意性を解決することは困難である．このような問題をパーミュテーション問題と呼び，これを解決することが FDICA における大きな課題である．

2.5　NMF

2.5.1　NMF の概要

NMF とは，1999 年に D. D. Lee と H. S. Seung によって提案された非負行列に対する分解アルゴリズムである．行列の分解は，線形方程式を解くための LU 分解や QR 分解，ベクトル空間の概念に基づく固有値分解や特異値分解等が代表的である．それに対して，NMF は非負行列を分解対象としている点において，これらの行列分解手法と大きく異なっている．2.2 節で説明した通り，時間領域の信号に対して，STFT を経て得られる振幅（あるいはパワー）スペクトログラムは非負行列である．従って，これらの行列は NMF の分解対象となる．

NMF は，次式に示すように，非負行列を別の 2 つの非負行列の行列積に分解する数理アル

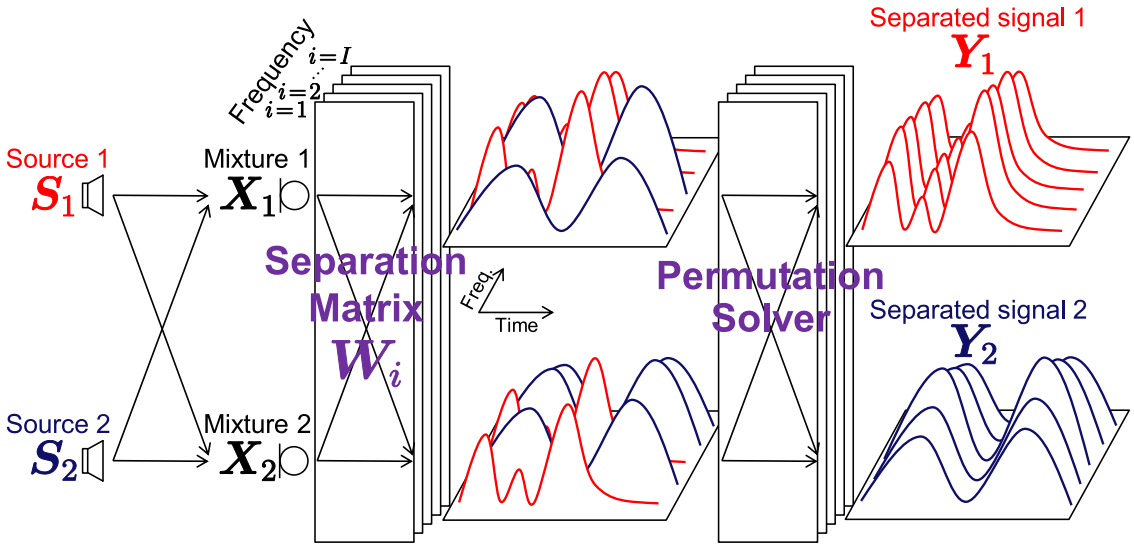


Fig. 2.4. Permutation problem in FDICA.

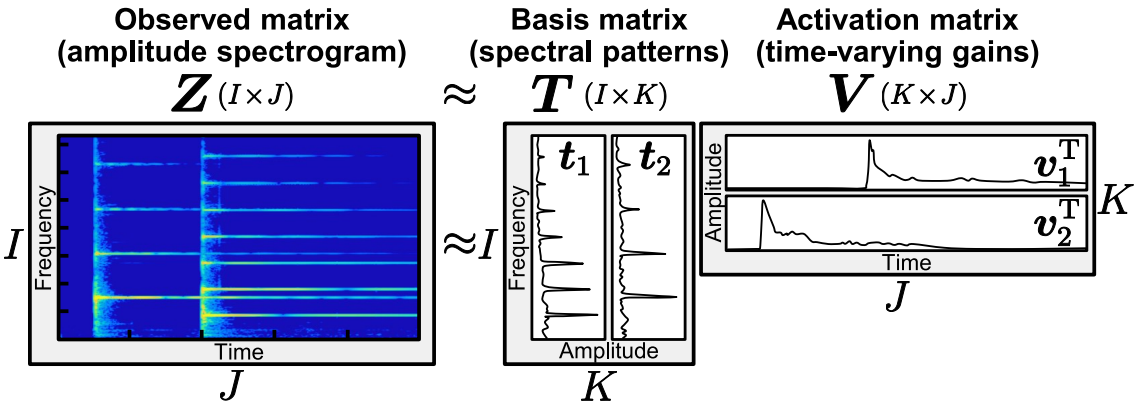


Fig. 2.5. NMF for audio signals, where $K = 2$.

ゴリズムである．

$$\boldsymbol{Z} \approx \boldsymbol{T}\boldsymbol{V} \tag{2.24}$$

ここで，$\boldsymbol{Z} \in \mathbb{R}_{\geq 0}^{I \times J}$ は分解の対象となる非負行列であり，$\boldsymbol{T} = [\,\boldsymbol{t}_1\ \boldsymbol{t}_2\ \cdots\ \boldsymbol{t}_K\,] \in \mathbb{R}_{\geq 0}^{I \times K}$ 及び $\boldsymbol{V} = [\,\boldsymbol{v}_1\ \boldsymbol{v}_2\ \cdots\ \boldsymbol{v}_K\,]^{\mathrm{T}} \in \mathbb{R}_{\geq 0}^{K \times J}$ はそれぞれ基底行列及びアクティベーション行列と呼ばれる非負行列である．$\boldsymbol{T}$ の列ベクトルである $\boldsymbol{t}_k$ は基底ベクトルと呼ばれ，その本数 K は通常 $K \ll \min(I, J)$ となるように設定される．ただし，$k = 1, 2, \ldots, K$ は基底ベクトルのインデクスを示す．行列 $\boldsymbol{T}\boldsymbol{V}$ のランクは基底ベクトルの本数 K が最大であるため，行列 $\boldsymbol{Z}$ のランクが K を上回る場合，$\boldsymbol{Z}$ を $\boldsymbol{T}\boldsymbol{V}$ によって完全に再構成することができない．従って，この場合において，NMF は低ランク近似分解となる．これは，$\boldsymbol{Z}$ 中に頻出する少数（K 個）の潜在的なパターンを基底ベクトルとして抽出できる教師無し学習であると解釈できる．

NMF を音響信号に適用する場合，振幅（あるいはパワー）スペクトログラムを非負観測行列 $\boldsymbol{Z}$ とすることが一般的である．この場合，Fig. 2.5 に示すように，音響信号中の頻出スペク

トルが $\boldsymbol{t}_k$ として得られ，さらに各スペクトルの時間的強度変化が $\boldsymbol{v}_k$ として現れる．このように，NMF は音響信号中のスペクトルパターンを学習できるため，音楽信号解析 [31] や音源分離 [32, 33] 等に頻繁に適用される．なお，NMF では，$\boldsymbol{T}$ 及び $\boldsymbol{V}$ の推定問題を解析的に解くことができないため，次節に示す最適化問題を解く必要がある．

2.5.2　ISNMF における最適化問題及び反復更新式の定式化

NMF では，次式の最適化問題を解くことで非負の変数行列 $\boldsymbol{T}$ 及び $\boldsymbol{V}$ を推定する．

$$\underset{\boldsymbol{T},\boldsymbol{V}}{\text{Minimize}}\ \mathcal{D}(\boldsymbol{Z}|\boldsymbol{T}\boldsymbol{V})\ \ \text{s.t.}\ \ t_{ik}, v_{kj} \geq 0\ \ \forall i,j,k \tag{2.25}$$

ここで，t_{ik} 及び v_{kj} はそれぞれ $\boldsymbol{T}$ 及び $\boldsymbol{V}$ の要素である．また，$\mathcal{D}(\cdot|\cdot)$ は 2 つの行列間の乖離度を測る関数である．そのような関数としては, 二乗 Euclid 距離, 一般化 Kullback–Leibler ダイバージェンス，Itakura–Saito ダイバージェンスが用いられる．本論文では，Itakura–Saito ダイバージェンスに基づく NMF（ISNMF）を取り扱う．

$\boldsymbol{A}$ 及び $\boldsymbol{B}$ を $I \times J$ 型の非負行列とするとき，これらの非負行列に対する Itakura–Saito ダイバージェンスは次式で定義される．

$$\mathcal{D}_{\mathrm{IS}}(\boldsymbol{A}|\boldsymbol{B}) = \sum_{i=1}^{I}\sum_{j=1}^{J}\left(\frac{a_{ij}}{b_{ij}} - \log\frac{a_{ij}}{b_{ij}} - 1\right) \tag{2.26}$$

ただし，a_{ij} 及び b_{ij} はそれぞれ $\boldsymbol{A} \in \mathbb{R}_{\geq 0}^{I\times J}$ 及び $\boldsymbol{B} \in \mathbb{R}_{\geq 0}^{I\times J}$ の要素を表す．式 (2.25) に Itakura–Saito ダイバージェンスを導入することで，ISNMF の最適化問題は次式となる．

$$\begin{aligned}\underset{\boldsymbol{T},\boldsymbol{V}}{\text{Minimize}}\ \mathcal{D}_{\mathrm{IS}}(\boldsymbol{Z}|\boldsymbol{T}\boldsymbol{V}) = &\sum_{i=1}^{I}\sum_{j=1}^{J}\left(\frac{z_{ij}}{\sum_{k=1}^{K} t_{ik}v_{kj}} - \log\frac{z_{ij}}{\sum_{k=1}^{K} t_{ik}v_{kj}} - 1\right)\\ &\text{s.t.}\ \ t_{ik}, v_{kj} \geq 0\ \ \forall i,j,k\end{aligned} \tag{2.27}$$

式 (2.27) で表される最適化問題に補助関数法を適用して得られる反復更新式は次式のようになる．

$$t_{ik}^{[t+1]} = t_{ik}^{[t]}\sqrt{\frac{\sum_{j=1}^{J} z_{ij}\left(\sum_{k'=1}^{K} t_{ik'}^{[t]}v_{k'j}^{[t]}\right)^{-2} v_{kj}^{[t]}}{\sum_{j=1}^{J} v_{kj}^{[t]}\left(\sum_{k'=1}^{K} t_{ik'}^{[t]}v_{k'j}^{[t]}\right)^{-1}}} \tag{2.28}$$

$$v_{kj}^{[t+1]} = v_{kj}^{[t]}\sqrt{\frac{\sum_{i=1}^{I} z_{ij}\left(\sum_{k'=1}^{K} t_{ik'}^{[t]}v_{k'j}^{[t]}\right)^{-2} t_{ik}^{[t]}}{\sum_{i=1}^{I} t_{ik}^{[t]}\left(\sum_{k'=1}^{K} t_{ik'}^{[t]}v_{k'j}^{[t]}\right)^{-1}}} \tag{2.29}$$

ここで，各変数の上付き文字 $\cdot^{[t]}$ は変数更新の反復回数である．式 (2.28) 及び (2.29) は，次の

ように行列形式で表現することもできる．

$$\boldsymbol{T} \leftarrow \boldsymbol{T} \odot \left\{ \frac{\left[\boldsymbol{Z} \odot (\boldsymbol{T}\boldsymbol{V})^{\cdot -2}\right]\boldsymbol{V}^{\mathrm{T}}}{(\boldsymbol{T}\boldsymbol{V})^{\cdot -1}\boldsymbol{V}^{\mathrm{T}}} \right\}^{\cdot \frac{1}{2}} \tag{2.30}$$

$$\boldsymbol{V} \leftarrow \boldsymbol{V} \odot \left\{ \frac{\boldsymbol{T}^{\mathrm{T}}\left[\boldsymbol{Z} \odot (\boldsymbol{T}\boldsymbol{V})^{\cdot -2}\right]}{\boldsymbol{T}^{\mathrm{T}}(\boldsymbol{T}\boldsymbol{V})^{\cdot -1}} \right\}^{\cdot \frac{1}{2}} \tag{2.31}$$

ここで，行列間の演算 $\odot$ 及び行列間の分数はそれぞれ要素毎の積及び商を表す．なお，式 (2.30) 及び (2.31) では反復回数 t の表記は省略し，変数更新を表す演算子 $\leftarrow$ を用いている．

2.6 ILRMA

2.4 節では，FDICA にはパーミュテーション問題という大きな問題が存在することを述べた．それに対して，2006 年に提案された IVA では，FDICA の音源モデルに「同一音源の全周波数成分は連動して生起する傾向にある」という仮定を導入することで，パーミュテーション問題を回避している．このように，FDICA の音源モデルに他の仮定を導入する方法はパーミュテーション問題の回避策として有効である．その一方で，IVA が仮定する全周波数成分の共変性は，周波数成分の部分的な強度変化を表現することができない．つまり，IVA は楽器音信号のような，基本周波数とその倍音のみが強い共変性を持つ音響信号，すなわち，共変性の程度が各時間周波数に依存して変化する音響信号の分離には適していない．この問題を克服した BSS として，ILRMA が提案されている．ILRMA は，FDICA の音源モデルに「音源の時間周波数構造は低ランクで近似できる」という仮定を導入した BSS である．この仮定によって，ILRMA では，各音源の時間周波数領域における共変性をより詳細に取り扱うことができ，明確な倍音構造を有する音響信号に良く適合した分離が可能となる．

ILRMA では，分離信号の各チャネルにおける複素スペクトログラム $\boldsymbol{Y}_n \in \mathbb{C}^{I \times J}$ に対して，次式の複素ガウス分布を生成モデルとして仮定する．

$$\begin{aligned} p(\boldsymbol{Y}_n) &= \prod_{i,j} p(\boldsymbol{y}_{ijn}) \\ &= \prod_{i,j} \frac{1}{\pi r_{ijn}} \exp\left(-\frac{|y_{ijn}|^2}{r_{ijn}}\right) \end{aligned} \tag{2.32}$$

ただし，r_{ijn} は各音源の時間周波数毎の分散であり，パワーの期待値として $r_{ijn} = \mathrm{E}[|y_{ijn}|^2]$ と書ける．なお，$\mathrm{E}[\cdot]$ は期待値を表す．式 (2.32) の仮定の下での，分離行列 $\boldsymbol{W}_i$ に関する観測信号の負対数尤度関数は次式で与えられる．

$$\mathcal{L} = -2J \sum_i \log|\det \boldsymbol{W}_i| + \sum_{i,j,n} \left(\frac{|y_{ijn}|^2}{r_{ijn}} + \log r_{ijn} \right) \tag{2.33}$$

ここで，本節の冒頭で述べた ILRMA の音源モデルに対する仮定を導入すると，分散 r_{ijn} は

NMF によって次式のように低ランク近似される．

$$r_{ijn} = \sum_k t_{ikn} v_{kjn} \tag{2.34}$$

また, r_{ijn}, t_{ikn}, 及び v_{kjn} は非負の値であり, それぞれを要素として持つ分散行列 $\boldsymbol{R}_n \in \mathbb{R}_{>0}^{I\times J}$, 基底行列 $\boldsymbol{T}_n \in \mathbb{R}_{\geq 0}^{I\times K}$，及びアクティベーション行列 $\boldsymbol{V}_n \in \mathbb{R}_{\geq 0}^{K\times J}$ によって，次式のように行列形式で表現できる．

$$\boldsymbol{R}_n = \boldsymbol{T}_n \boldsymbol{V}_n \tag{2.35}$$

式 (2.34) 及び (2.35) は，各音源の時間周波数構造 $\boldsymbol{R}_n$ がランク K の非負行列で近似されることを示している．ここで，式 (2.33) に式 (2.34) を代入することで次式が得られる．

$$\mathcal{L} = -2J \sum_i \log|\det \boldsymbol{W}_i| + \sum_{i,j,n} \left(\frac{|y_{ijn}|^2}{\sum_k t_{ikn} v_{kjn}} + \log \sum_k t_{ikn} v_{kjn} \right) \tag{2.36}$$

ILRMA では，分離行列 $\{\boldsymbol{W}\}_{i=1}^{l}$，基底行列 $\{\boldsymbol{T}\}_{n=1}^{N}$，及びアクティベーション行列 $\{\boldsymbol{V}\}_{n=1}^{N}$ に関して，式 (2.36) の最小化を行っている．式 (2.36) の第 2 項（変数 i，j，及び k に関する総和）の最小化は，ISNMF の最適化問題 (2.27) において，$\boldsymbol{Z}$，$\boldsymbol{T}$，及び $\boldsymbol{V}$ をそれぞれ $|\boldsymbol{Y}_n|^{\cdot 2}$, $\boldsymbol{T}_n$，及び $\boldsymbol{V}_n$ とした式に対応する[*3]．これは，ILRMA における分散の最尤推定が，分離信号の各チャネルにおけるパワースペクトログラム $|\boldsymbol{Y}_n|^{\cdot 2}$ を低ランク近似しながら分離行列 $\boldsymbol{W}_i$ の推定を行うこと，すなわち分離信号のスペクトログラムを低ランクに誘導することに等価であることを示している．また，式 (2.36) の第 1 項（分離行列 $\boldsymbol{W}_i$ の行列式の負対数を取った項）及び第 2 項内部の y_{ijn} に関する部分は IVA における負対数尤度関数に対応する．このことから，ILRMA は IVA と同様に，周波数成分の共変性，及び，FDICA の性質である分離信号間の独立性を考慮しているともいえる[*4]．つまり，ILRMA は，音源の時間周波数構造 $\boldsymbol{R}_n$ を低ランク行列として近似しながら，時間周波数構造の共変性と分離信号間の独立性を加味した分離行列 $\boldsymbol{W}_i$ の推定を行っていると解釈できる．本論文では，分離行列 $\boldsymbol{W}_i$ を空間モデル（あるいは空間分離フィルタ），音源の時間周波数構造 $\boldsymbol{R}_n$ を音源モデルと呼ぶ．ここまでで説明した，ILRMA における音源分離の原理を Fig. 2.6 に示す．

先に述べたことから，ILRMA における音源モデルの推定には式 (2.30) 及び (2.31) で表される ISNMF の更新式を，空間モデルの推定には AuxIVA で提案された反復最適化手法である IP を適用することができる．ここで用いる IP は，初期の IVA での反復最適化手法である自然勾配法 [34, 35] よりも，高速かつ安定な分離行列の最適化手法であることが実験的に示されている．空間モデルの推定では，分離行列 $\boldsymbol{W}_i$ は行毎に更新され，その更新式は IP を用い

[*3] 最小化を考える上で，定数項の差異は問題とならない．

[*4] 既に述べたように，IVA と ILRMA の間で共変性の構造は異なる．IVA では全周波数で一様な強度変化を持つ音源構造を仮定しているが，ILRMA ではより詳細な音源構造を仮定している．

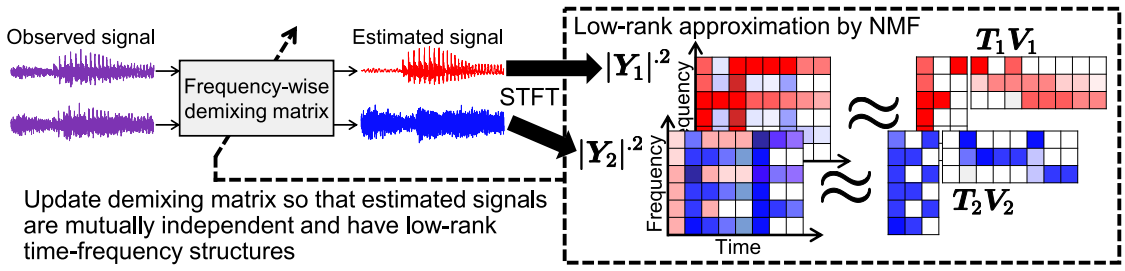


Fig. 2.6. Principle of BSS based on ILRMA.

て次式で与えられる．

$$\boldsymbol{U}_{in} \leftarrow \frac{1}{J} \sum_{j} \frac{1}{[\boldsymbol{T}_n \boldsymbol{V}_n]_{i,j}} \boldsymbol{x}_{ij} \boldsymbol{x}_{ij}^{\mathrm{H}} \tag{2.37}$$

$$\boldsymbol{w}_{in} \leftarrow \left(\boldsymbol{W}_i \boldsymbol{U}_{in}\right)^{-1} \boldsymbol{e}_n \tag{2.38}$$

$$\boldsymbol{w}_{in} \leftarrow \boldsymbol{w}_{in} \left(\boldsymbol{w}_{in}^{\mathrm{H}} \boldsymbol{U}_{in} \boldsymbol{w}_{in}\right)^{-\frac{1}{2}} \tag{2.39}$$

ここで，$[\cdot]_{r,c}$ は行列の (r, c) 番目の要素を表す．なお，$\boldsymbol{e}_n \in \{0, 1\}^N$ は n 番目の要素が 1，他要素が 0 の単位ベクトルである．また，$\boldsymbol{w}_{in}$ は分離行列 $\boldsymbol{W}_i$ の行ベクトルであり，分離ベクトルと呼ばれる．各分離ベクトルを更新した後は，分離信号 $\boldsymbol{Y}_n$ を次式で更新する．

$$y_{ijn} \leftarrow \boldsymbol{w}_{in}^{\mathrm{H}} \boldsymbol{x}_{ij} \tag{2.40}$$

音源モデルの推定では，基底行列 $\boldsymbol{T}_n$ 及びアクティベーション行列 $\boldsymbol{V}_n$ が式 (2.30) 及び (2.31) によって更新され，式中の変数を ILRMA の変数で書き換えた更新式は次式で与えられる．

$$\boldsymbol{T}_n \leftarrow \boldsymbol{T}_n \odot \left\{ \frac{\left[|\boldsymbol{Y}_n|^{\cdot 2} \odot \left(\boldsymbol{T}_n \boldsymbol{V}_n\right)^{\cdot -2}\right] \boldsymbol{V}_n^{\mathrm{T}}}{\left(\boldsymbol{T}_n \boldsymbol{V}_n\right)^{\cdot -1} \boldsymbol{V}_n^{\mathrm{T}}} \right\}^{\cdot \frac{1}{2}} \tag{2.41}$$

$$\boldsymbol{V}_n \leftarrow \boldsymbol{V}_n \odot \left\{ \frac{\boldsymbol{T}_n^{\mathrm{T}} \left[|\boldsymbol{Y}_n|^{\cdot 2} \odot \left(\boldsymbol{T}_n \boldsymbol{V}_n\right)^{\cdot -2}\right]}{\boldsymbol{T}_n^{\mathrm{T}} \left(\boldsymbol{T}_n \boldsymbol{V}_n\right)^{\cdot -1}} \right\}^{\cdot \frac{1}{2}} \tag{2.42}$$

$\boldsymbol{T}_n$ 及び $\boldsymbol{V}_n$ の更新後に，分散行列 $\boldsymbol{R}_n$ の更新を明示的に行うのであれば，それは次式で表される．

$$\boldsymbol{R}_n \leftarrow \boldsymbol{T}_n \boldsymbol{V}_n \tag{2.43}$$

以上の空間モデルと音源モデルの更新式を交互に反復することで，式 (2.33) を最小化できる．また，ILRMA によって推定された分離信号には，FDICA や IVA と同様に，周波数毎の順序とスケールの任意性がある．従って，反復最適化後は，次式の PB 法を適用することで，分離信号のスケールを補正する．

$$\tilde{\boldsymbol{y}}_{ijn} = \boldsymbol{W}_i^{-1} \left(\boldsymbol{e}_n \odot \boldsymbol{y}_{ij}\right) = y_{ijn} \boldsymbol{\lambda}_{in}, \tag{2.44}$$

ここで，$\tilde{\boldsymbol{y}}_{ijn} = [\,\tilde{y}_{ijn1}, \tilde{y}_{ijn2}, \ldots, \tilde{y}_{ijnm}, \ldots, \tilde{y}_{ijnM}\,]^{\mathrm{T}} \in \mathbb{C}^M$ はスケール補正後の分離信号の (i,j) 番目の成分，$\boldsymbol{\lambda}_{in} = [\,\lambda_{in1}, \lambda_{in2}, \ldots, \lambda_{inm}, \ldots, \lambda_{inM}\,]^{\mathrm{T}} \in \mathbb{C}^M$ はスケール補正係数を表す．また，ベクトル間の演算 $\odot$ は，行列に対する演算と同様に，要素毎の積を表す．

2.7　Consistent ILRMA

2.7.1　スペクトログラム無矛盾性

STFT を適用する際の窓関数を乗じるという操作は，その窓関数のスペクトルが時間周波数領域において，（時間領域での各短時間区間信号のスペクトルを重み係数として）周波数方向に畳み込まれることに相当する．従って，時間周波数領域での大きなパワー値は周波数方向に滲み，共起性が生まれる．また，STFT を適用した際の時間フレーム間のオーバーラップにより，時間周波数領域において，時間方向の冗長性が生じる．ここで述べた冗長性とは，互いに交わる 2 つの時間フレームが共有している区間の情報は，両方のフレームに含まれるという性質のことである．これによって，時間周波数領域での大きなパワー値は時間方向にも滲み，共起性が生まれる．結果的に，時間周波数領域では，各時間周波数成分の近傍で一貫した共起性が生じていることが自然である．この共起性はスペクトログラム無矛盾性と呼ばれる．また，時間周波数領域における何らかの信号処理によって，共起性が崩された状態はスペクトログラム矛盾と呼ばれる．Figs. 2.7(a) 及び (b) に，共起性が一貫していない矛盾したスペクトログラム，及び，このスペクトログラムに対応する無矛盾なスペクトログラムをそれぞれ示す．ここで，$\boldsymbol{S}_{\mathrm{art}}$ は，人工的に作られたスペクトログラム，すなわち一貫した共起性が考慮されていない矛盾したスペクトログラムを表す．Fig. 2.9(a) は，中心部分の時間周波数グリッドのみが大きなパワー値を持つ矛盾したスペクトログラムである．また，Fig. 2.9(b) は，Fig. 2.9(a) のスペクトログラムに対して，後述する方法により無矛盾性を担保したスペクトログラムである．

任意のスペクトログラムに対するスペクトログラム無矛盾性の担保は，逆 STFT 及び STFT を続けて適用することで実現できる．この操作を視覚的に表現した図を，Fig. 2.8 に示す．Fig. 2.8 内の時間領域の信号 $\mathbf{s}$ は，STFT によって時間周波数領域の信号 $\boldsymbol{S}$ に写される．このスペクトログラム $\boldsymbol{S}$ に対し，何らかの信号処理を加えることで，新たに得られたスペクトログラムを $\boldsymbol{S}'$ とする．$\boldsymbol{S}'$ が共起性の崩れた矛盾したスペクトログラムとなった場合，$\boldsymbol{S}'$ に直接対応する時間波形は時間領域に存在しない．この矛盾したスペクトログラム $\boldsymbol{S}'$ に対して，逆 STFT を適用すると，時間周波数領域における無矛盾なスペクトログラム $\boldsymbol{S}''$ に射影されたうえで時間領域に変換される．その結果，新たな時間信号 $\mathbf{s}''$ が得られる．従って，時間周波数領域のいかなるスペクトログラムも，一度逆 STFT を適用して時間領域に戻し，再び STFT を適用して時間周波数領域の信号に変換することで，無矛盾なスペクトログラムに変換することができる．

Figs. 2.9(a) 及び (b) に，共起性に一貫性のない矛盾した音楽信号のスペクトログラム，及び，このスペクトログラムに対応する無矛盾なスペクトログラムをそれぞれ示す．Fig. 2.9(a)

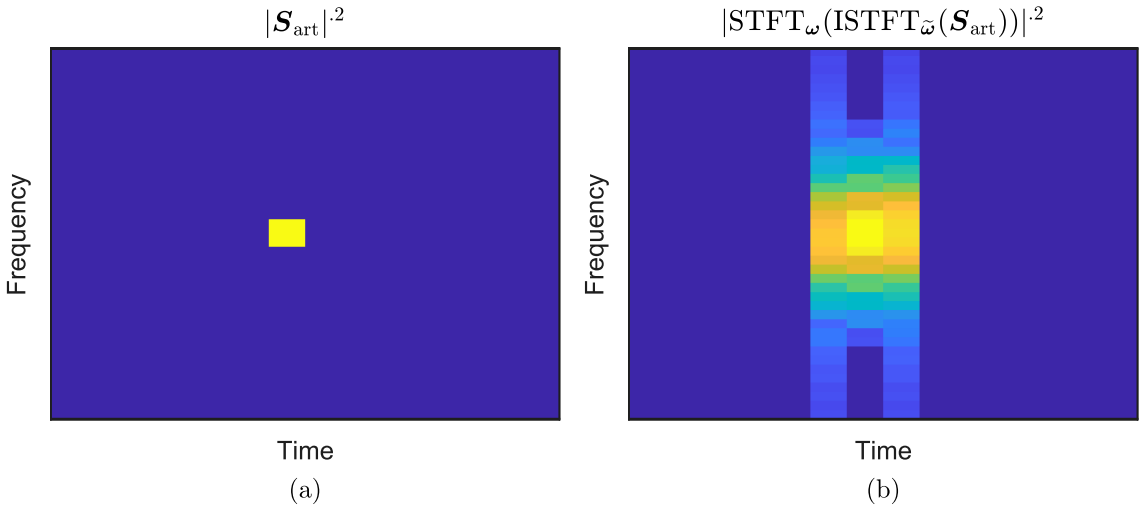


Fig. 2.7. (a) inconsistent power spectrograms $|\boldsymbol{S}_{\mathrm{art}}|^2$ and (b) their consistent version obtained by applying inverse STFT and STFT. Spectrogram of (a) is artificially produced with random phase.

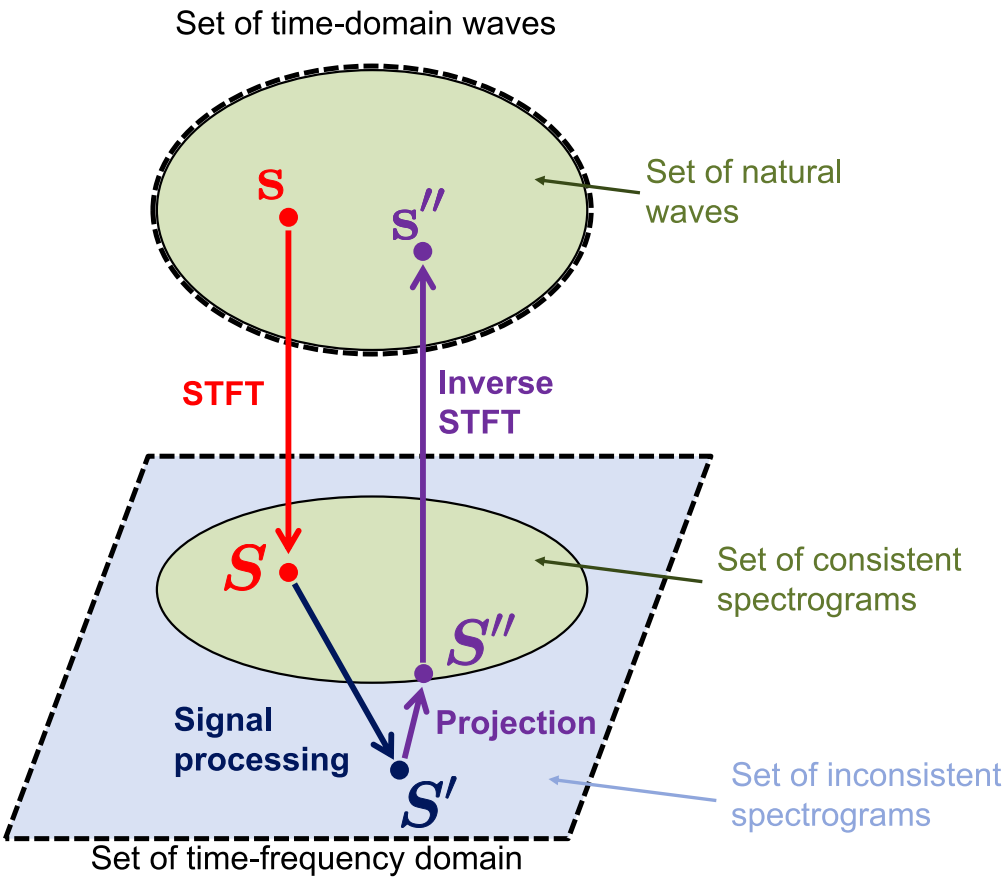


Fig. 2.8. Spectrogram (in)consistency with STFT and inverse STFT.

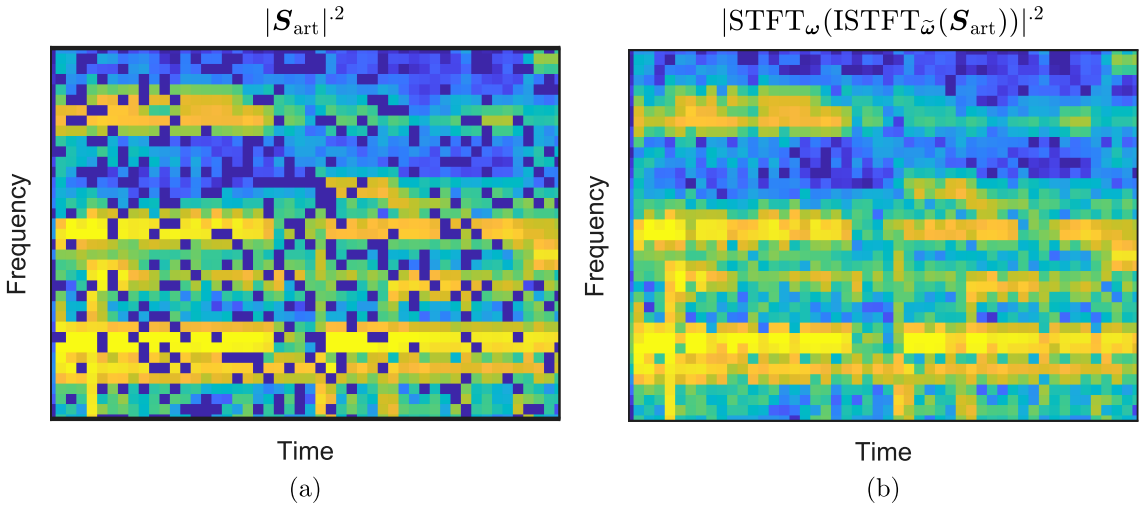


Fig. 2.9. (a) inconsistent power spectrograms $|\boldsymbol{S}_{\mathrm{art}}|^2$ and (b) their consistent version obtained by applying inverse STFT and STFT. Spectrogram of (a) is music and speech signals with random dropout.

は，音楽信号の無矛盾なスペクトログラムに対して，時間周波数グリッドをランダムに選び，それらのパワーを 0 にしたスペクトログラムである．Figs. 2.7 及び 2.9 から，無矛盾なスペクトログラムにおいては，時間と周波数の両方向に信号のパワーの滲みが生じていることが確認できる．これにより，スペクトログラム無矛盾性の担保は，スペクトログラムの時間方向及び周波数方向に対して，スムージングを施す処理であると解釈できる．このことから，スペクトログラム無矛盾性は，スペクトログラムにおける各時間周波数近傍の連動性とも解釈できる．

Fig. 2.8 の通り，逆 STFT は，矛盾したスペクトログラムから無矛盾なスペクトログラムを復元する操作である．すなわち，スペクトログラム $\boldsymbol{Z}$ の無矛盾性は次式によって特徴付けることができる．

$$\mathcal{E}(\boldsymbol{Z}) = \boldsymbol{Z} - \mathrm{STFT}_{\boldsymbol{\omega}}(\mathrm{ISTFT}_{\widetilde{\boldsymbol{\omega}}}(\boldsymbol{Z})) \tag{2.45}$$

このとき，(2.45) のノルム $\|\mathcal{E}(\boldsymbol{Z})\|$ が 0 となるスペクトログラム $\boldsymbol{Z}$ を無矛盾と呼ぶ．

2.7.2　スペクトログラム無矛盾性に基づく ILRMA

前節で述べたスペクトログラム無矛盾性は，BSS の性能向上に寄与することが明らかにされている [25]．これは，スペクトログラム無矛盾性の担保によって与えられる時間周波数領域の共起性が，パーミュテーション問題を緩和する働きを持っていることに起因する．パーミュテーション問題が生じたスペクトログラムに対して，スペクトログラム無矛盾性の担保が与える影響を Fig. 2.10 に示す．Fig. 2.10(a) は，音楽信号のパワースペクトログラム $|\boldsymbol{S}|^{\cdot 2}$ である．Fig. 2.10(b) は，スペクトログラム $\boldsymbol{S}$ に対して，人工的にパーミュテーション問題を引き起こしたもののパワースペクトログラム $|\boldsymbol{S}^{(\mathrm{perm})}|^{\cdot 2}$ である．Fig. 2.10(c) は，$\boldsymbol{S}^{(\mathrm{perm})}$ に対してスペクトログラム無矛盾性を担保したもののパワースペクトログラム

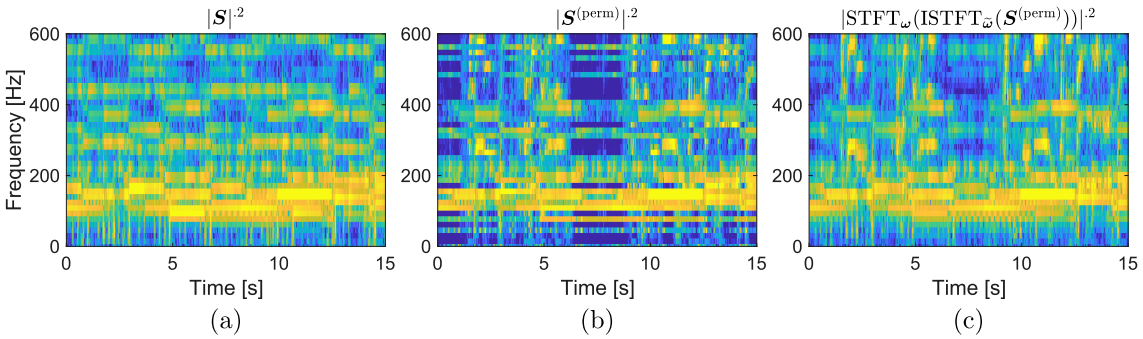


Fig. 2.10. Smoothing effect of spectrogram consistency applied to permutation misaligned music signal. (a) is power spectrogram of original source signal $|\boldsymbol{S}|^{.2}$. (b) is randomly permuted version of (a), which simulates the permutation problem and is denoted as $\boldsymbol{S}_n^{(\mathrm{perm})}$. (c) is the consistent version of (b).

$|\mathrm{STFT}_{\boldsymbol{\omega}}(\mathrm{ISTFT}_{\widetilde{\boldsymbol{\omega}}}(\boldsymbol{S}^{(\mathrm{perm})}))|^{.2}$ である．これらの図から分かるように，スペクトログラムにおける時間周波数領域の共起性は，周波数方向の連続性を強調する．その結果，周波数毎の音源成分が正しく整列された状態に誘導され，パーミュテーション問題を緩和することができる．

スペクトログラム無矛盾性を ILRMA の最適アルゴリズムの毎反復に導入した BSS として，consistent ILRMA が提案されている [26, 27]．ILRMA の空間モデルの反復最適化更新式 (2.37)–(2.40) 及び NMF 音源モデルの反復最適化更新式 (2.41)，(2.42) において，以下の演算を挿入することで，毎回の反復においてスペクトログラム無矛盾性を担保する．

$$\boldsymbol{Y}_n \leftarrow \mathrm{STFT}_{\boldsymbol{\omega}}(\mathrm{ISTFT}_{\widetilde{\boldsymbol{\omega}}}(\boldsymbol{Y}_n)) \tag{2.46}$$

式 (2.46) は，Fig. 2.11 に示すように，分離信号のスペクトログラム $\boldsymbol{Y}_n$ を無矛盾なスペクトログラムの集合へと射影していることに対応する．ここで，赤色，青色，橙色，及び紫色の矢印はそれぞれ STFT の適用，ILRMA における分離行列 $\boldsymbol{W}_i$ の反復更新，Consistent ILRMA におけるスペクトログラム無矛盾性の担保，及び ISTFT の適用（すなわち，無矛盾なスペクトログラム集合への射影及び時間領域への変換）を表す．また，$\mathbf{x}$，$\boldsymbol{X}$，$\boldsymbol{Y}$，及び $\boldsymbol{S}$ はそれぞれ時間領域での観測信号，観測信号 $\mathbf{x}$ のスペクトログラム，分離信号のスペクトログラム，及び音源信号のスペクトログラムを表す．仮に $\boldsymbol{Y}_n$ が無矛盾であれば，式 (2.46) は $\boldsymbol{Y}_n$ に影響を与えない．また，$\boldsymbol{Y}_n$ に矛盾があれば，Figs. 2.7，2.9，及び 2.10 に示すように，式 (2.46) は $\boldsymbol{Y}_n$ の時間方向と周波数方向の両方にスムージングをかける形で作用する．

上記の新しい処理の導入により，最適化の毎反復において，矛盾したスペクトログラムが無矛盾なスペクトログラムに射影される．つまり，Fig. 2.11 に示すように，Consistent ILRMA では，ILRMA よりも真の音源信号に近づきながら音源分離を進めることができる．

2.6 節で述べた通り，ILRMA で推定した分離信号には周波数毎のスケールの任意性がある．この任意性は，スペクトログラム無矛盾性を崩す原因となる．従って，Consistent ILRMA では，周波数毎のスケールの任意性に起因する矛盾の影響を最小限に抑える必要がある．これを目的として，最適化の毎反復で式 (2.46) を行う直前に，式 (2.44) の PB 法を適用する．PB 法

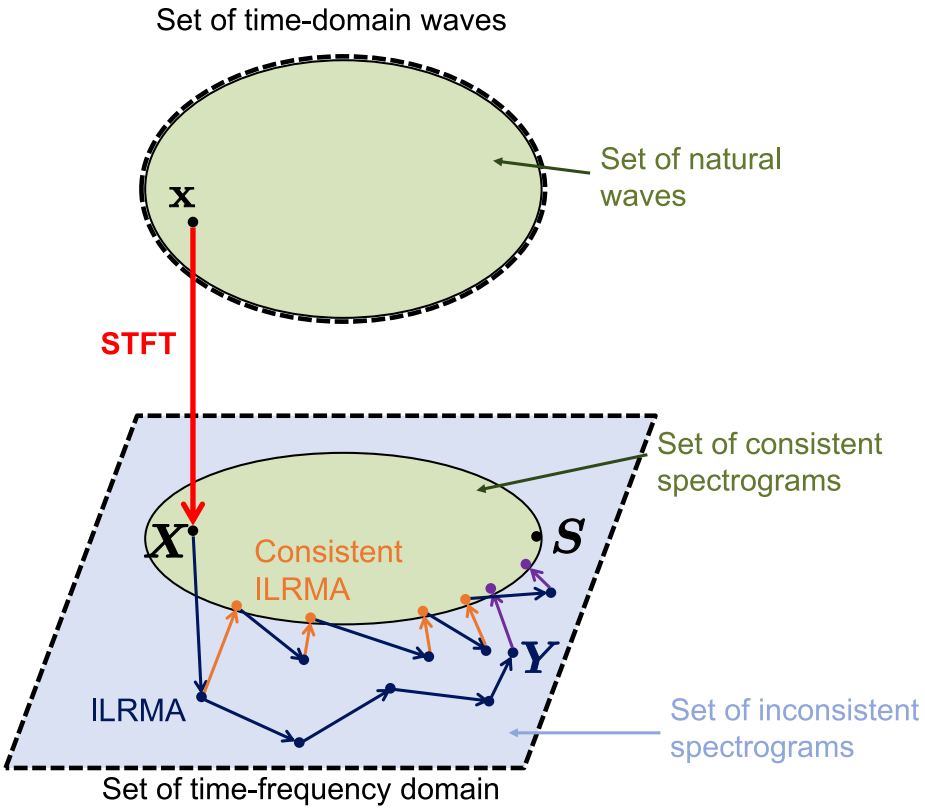


Fig. 2.11. Comparison between ILRMA and consistent ILRMA, where arrows in red show STFT, arrows in dark blue show iterative update of parameters in ILRMA, arrows in orange show ensuring process of spectrogram consistency, and arrows in purple show projection onto set of consistent spectrograms applied in inverse STFT. Separated signal estimated by consistent ILRMA tends to approach to oracle source signal $\boldsymbol{S}$.

の適用によって，目的関数 (2.33) の値が変動する．この変化分を補正するために，他の変数を次のように更新する．

$$\boldsymbol{w}_{in} \leftarrow \lambda_{inm_{\mathrm{ref}}} \boldsymbol{w}_{in} \tag{2.47}$$

$$y_{ijn} \leftarrow \boldsymbol{w}_{in}^{\mathrm{H}} \boldsymbol{x}_{ij} \tag{2.48}$$

$$t_{ikn} \leftarrow |\lambda_{inm_{\mathrm{ref}}}|^2 t_{ikn} \tag{2.49}$$

ここで，m_{ref} は PB 法で用いるリファレンスチャネルのインデクスである．

2.8　本章のまとめ

本章では，BSS の基礎技術と従来手法について説明した．次章では，2.6 節で説明した ILRMA の空間モデル及び音源モデルの更新に，それぞれ異なる解像度の時間周波数表現を導入した方法を提案する．また，提案手法が持つ利点や，提案手法のアイデアを実現する上で生じる問題点とそれに対する解決策について詳しく述べる．

第 3 章

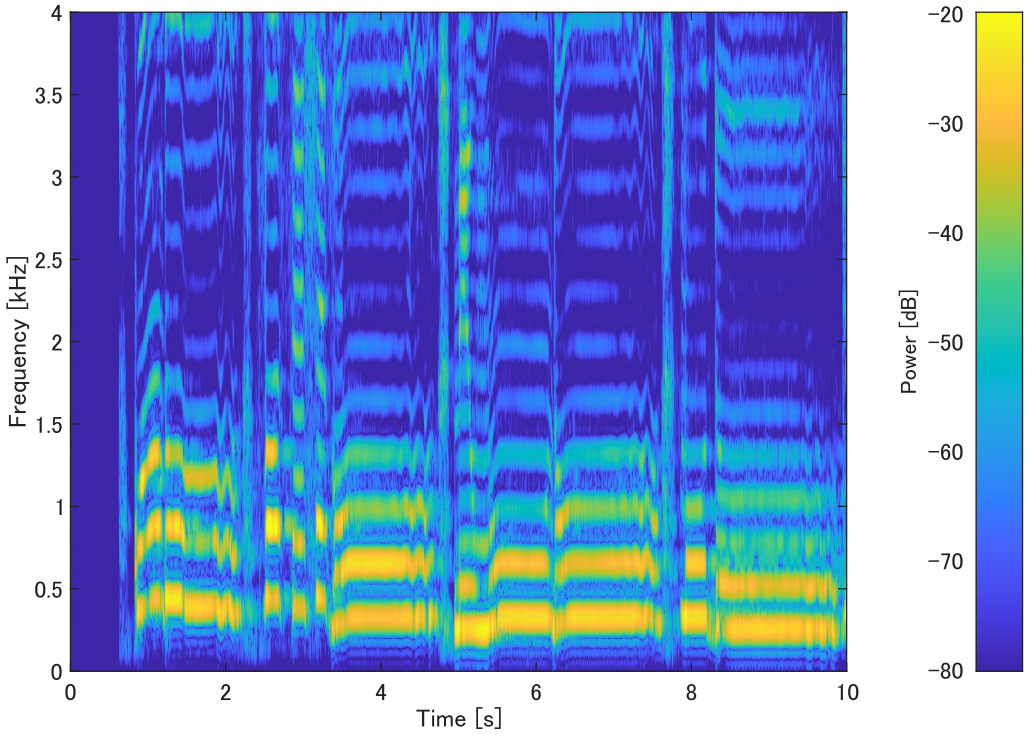
提案手法

3.1 まえがき

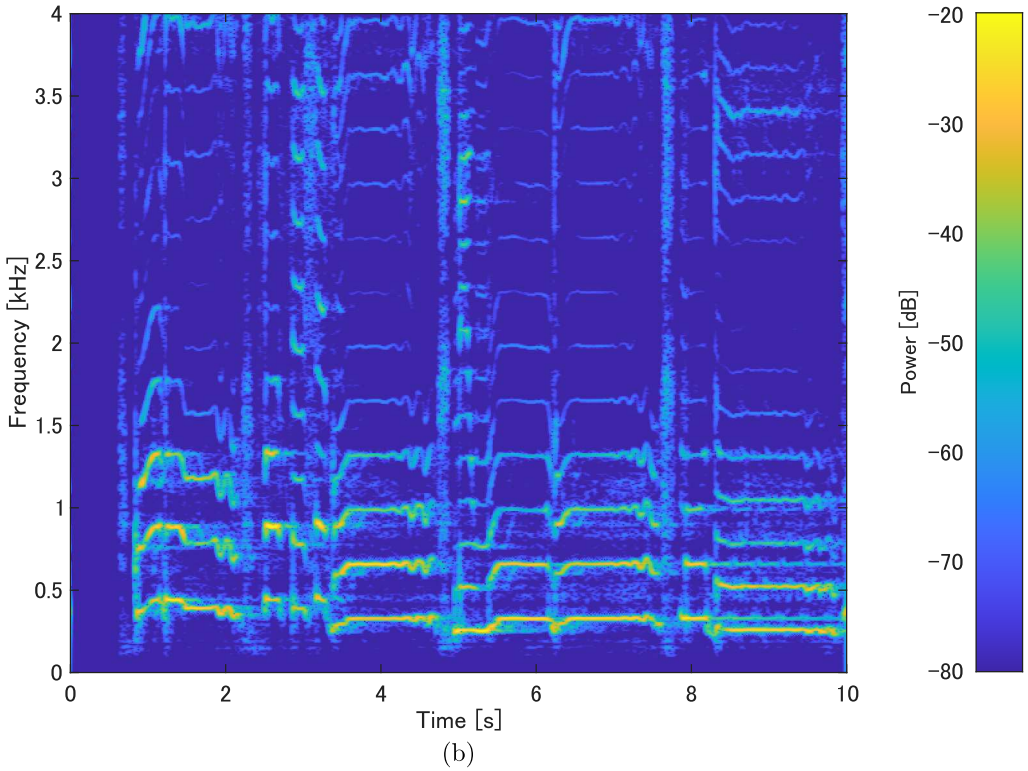
本章では，提案手法である多重解像度時間周波数表現に基づく ILRMA について説明する．3.2 節では，ILRMA の空間モデルと音源モデルの最適化にそれぞれ適した窓長が存在するという予想に基づき，提案手法の動機について詳しく述べる．3.3 節では，提案手法のアイデアを実現する上で生じる問題について述べる．3.4 節では，3.3 節で述べた問題点を回避するために，窓関数の有効幅を変化させるという方法を導入し，その具体的な内容について説明する．3.5 節では，窓関数の有効幅を設定できる Chebyshev 窓の説明を行う．3.6 節では，提案手法における最適化アルゴリズムを示す．最後に，3.7 節で本章の総括を行う．

3.2 動機

2.3 節で述べた通り，STFT における窓長が観測信号の残響時間より短い場合，時間領域の残響による畳み込み混合を時間周波数領域の瞬時混合に変換できず，式 (2.22) の仮定が成り立たない．このことから，式 (2.23) での高精度な BSS は原理的に困難となる．そのため，STFT の窓長は残響時間を十分超える長さ（例えば 256 ms 以上）に設定されることが一般的である [28]．よって，ILRMA では，空間モデルの最適化を行う際に適切な窓長が存在すると予想される．一方で，音声や楽器音等の音響信号を NMF で時間周波数解析する場合は，16〜128 ms 程度の窓長で STFT を適用する場合が多い（例えば，文献 [31]，[33]，[36] 等）．ILRMA においても，各音源のスペクトログラムを NMF でモデル化する関係上，低ランク行列として良く近似できる窓長が存在するはずである．例として，Figs. 3.1(a) 及び (b) に，ボーカルの音響信号を異なる窓長（32 ms 及び 256 ms）で STFT して得られるパワースペクトログラムを示す．この図から，NMF によるスペクトログラムの近似精度は STFT の窓長の長さに大きく依存することが予想される．従って，ILRMA においては音源モデルの最適化の観点からも適切な窓長が存在し，それは前述の空間モデルの窓長の最適値とは異なる可能性がある．



(a)



(b)

Fig. 3.1. Vocal spectrograms using (a) narrow (32 ms) and (b) wide (256 ms) windows.

3.3 異なる時間周波数解像度を扱う上での問題点

前節で述べたことを基に，ILRMA の空間モデルと音源モデルの最適化に異なる窓長の窓関数を導入することを考える．いま，空間モデル及び音源モデルの最適化における STFT の窓関数を，それぞれ $\boldsymbol{\omega}^{(\mathrm{Ex})}$ 及び $\boldsymbol{\omega}^{(\mathrm{In})}$ とする．なお，本節に限り，二つの窓関数 $\boldsymbol{\omega}^{(\mathrm{Ex})}$ と $\boldsymbol{\omega}^{(\mathrm{In})}$ の長さは互いに異なっているものとして扱う．観測信号 $\mathbf{x}_m$ を窓関数 $\boldsymbol{\omega}^{(\mathrm{Ex})}$ 及び $\boldsymbol{\omega}^{(\mathrm{In})}$ で STFT して得られるスペクトログラムを，それぞれ $\boldsymbol{X}_m^{(\mathrm{Ex})}$ 及び $\boldsymbol{X}_m^{(\mathrm{In})}$ と表す．また，分離信号 $\mathbf{y}_n$ を窓関数 $\boldsymbol{\omega}^{(\mathrm{Ex})}$ 及び $\boldsymbol{\omega}^{(\mathrm{In})}$ で STFT して得られるスペクトログラムを，それぞれ $\boldsymbol{Y}_n^{(\mathrm{Ex})}$ 及び $\boldsymbol{Y}_n^{(\mathrm{In})}$ と表す．分離信号の更新式 (2.40) にスペクトログラム $\boldsymbol{X}_m^{(\mathrm{Ex})}$ 及び $\boldsymbol{Y}_n^{(\mathrm{Ex})}$ を導入した式を以下に示す．

$$y_{ijn}^{(\mathrm{Ex})} \leftarrow \boldsymbol{w}_{in}^{\mathrm{H}} \boldsymbol{x}_{ij}^{(\mathrm{Ex})} \tag{3.1}$$

ただし，$y_{ijn}^{(\mathrm{Ex})}$ は，分離信号の n 番目のチャネルを窓関数 $\boldsymbol{\omega}^{(\mathrm{Ex})}$ で STFT して得られる観測信号のスペクトログラムの (i,j) 番目の要素である．また，$\boldsymbol{x}_{ij}^{(\mathrm{Ex})}$ は，観測信号の各チャネルを窓関数 $\boldsymbol{\omega}^{(\mathrm{Ex})}$ で STFT して得られる観測信号のスペクトログラムの (i,j) 番目の要素であり，次式で表される．

$$\boldsymbol{x}_{ij}^{(\mathrm{Ex})} = [\, x_{ij1}^{(\mathrm{Ex})}, \ldots, x_{ijm}^{(\mathrm{Ex})}, \ldots, x_{ijM}^{(\mathrm{Ex})} \,]^{\mathrm{T}} \in \mathbb{C}^M \tag{3.2}$$

音源モデルの反復更新式 (2.41) 及び (2.42) にスペクトログラム $\boldsymbol{Y}_n^{(\mathrm{In})}$ を導入した式を以下に示す．

$$\boldsymbol{T}_n \leftarrow \boldsymbol{T}_n \odot \left\{ \frac{\left[|\boldsymbol{Y}_n^{(\mathrm{In})}|^{.2} \odot (\boldsymbol{T}_n \boldsymbol{V}_n)^{\cdot -2} \right] \boldsymbol{V}_n^{\mathrm{T}}}{(\boldsymbol{T}_n \boldsymbol{V}_n)^{\cdot -1} \boldsymbol{V}_n^{\mathrm{T}}} \right\}^{\cdot \frac{1}{2}} \tag{3.3}$$

$$\boldsymbol{V}_n \leftarrow \boldsymbol{V}_n \odot \left\{ \frac{\boldsymbol{T}_n^{\mathrm{T}} \left[|\boldsymbol{Y}_n^{(\mathrm{In})}|^{.2} \odot (\boldsymbol{T}_n \boldsymbol{V}_n)^{\cdot -2} \right]}{\boldsymbol{T}_n^{\mathrm{T}} (\boldsymbol{T}_n \boldsymbol{V}_n)^{\cdot -1}} \right\}^{\cdot \frac{1}{2}} \tag{3.4}$$

また，空間モデルの反復更新式の 1 つである (2.37) にスペクトログラム $\boldsymbol{X}_m^{(\mathrm{Ex})}$ を導入した式を以下に示す．

$$\boldsymbol{U}_{in} \leftarrow \frac{1}{J} \sum_j \frac{1}{[\boldsymbol{T}_n \boldsymbol{V}_n]_{i,j}} \boldsymbol{x}_{ij}^{(\mathrm{Ex})} \boldsymbol{x}_{ij}^{(\mathrm{Ex})\mathrm{H}} \tag{3.5}$$

2.2 節で述べた通り，スペクトログラムの周波数ビン数は STFT の窓長に依存する[*1]．よって，$\boldsymbol{\omega}^{(\mathrm{Ex})}$ と $\boldsymbol{\omega}^{(\mathrm{In})}$ の長さが異なる状況において，二つのスペクトログラム $\boldsymbol{X}_n^{(\mathrm{Ex})}$ 及び $\boldsymbol{Y}_n^{(\mathrm{In})}$ のサイズ（周波数ビン数）は互いに異なる．つまり，式 (3.1) で得られたスペクトログラム

[*1] スペクトログラムの時間フレーム数は，STFT の窓関数の情報（窓関数の各点での値と窓長）によって制限される．STFT の窓長が時間フレーム数に直接影響することはないため，ここでは周波数ビン数に対する影響のみを考える．

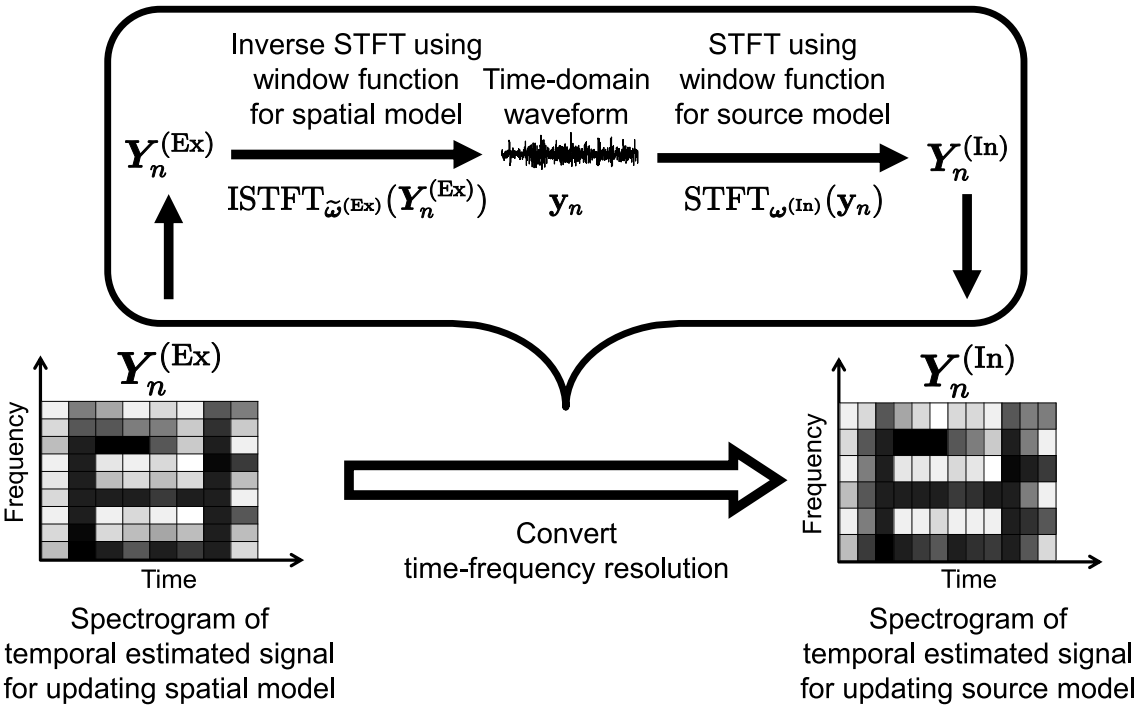


Fig. 3.2. Process flow for converting time-frequency resolution via time-domain waveform.

$\boldsymbol{Y}_n^{(\mathrm{Ex})}$ を式 (3.2) の $\boldsymbol{Y}_n^{(\mathrm{In})}$ として用いることはできない．これは，ILRMA の空間モデル及び音源モデルの最適化更新に，異なる窓長で得られたスペクトログラム $\boldsymbol{Y}_n^{(\mathrm{Ex})}$ 及び $\boldsymbol{Y}_n^{(\mathrm{In})}$ をそのまま用いることはできないことを意味する．従って，ILRMA における各モデルの最適化に異なる窓長の窓関数を導入するためには，二つのスペクトログラム $\boldsymbol{Y}_n^{(\mathrm{Ex})}$ と $\boldsymbol{Y}_n^{(\mathrm{In})}$ とでサイズが異なっているという問題を何らかの方法で解決する必要がある．

スペクトログラムのサイズを一致させるには，様々な方法が考えられる．本論文では，スペクトログラム $\boldsymbol{Y}_n^{(\mathrm{Ex})}$ 及び $\boldsymbol{Y}_n^{(\mathrm{In})}$ のサイズを一致させる方法の一つとして，Fig. 3.2 に示す方法を提案する．これは，スペクトログラムに逆 STFT 及び STFT を施すことで，時間領域を介してサイズを一致させる方法である．

Fig. 3.2 の処理の詳しい説明は次節に回し，ここからは残った問題について説明する．前述の通り，二つのスペクトログラム $\boldsymbol{X}_m^{(\mathrm{Ex})}$ 及び $\boldsymbol{Y}_n^{(\mathrm{In})}$ の周波数ビン数は互いに異なる．ここで，スペクトログラム $\boldsymbol{X}_m^{(\mathrm{Ex})}$ 及び $\boldsymbol{Y}_n^{(\mathrm{In})}$ の周波数ビン数をそれぞれ $I^{(\mathrm{Ex})}$ 及び $I^{(\mathrm{In})}$ とする．式 (3.3) 及び (3.4) から，音源モデルの最適化によって得られる基底行列 $\boldsymbol{T}_n$ は $I^{(\mathrm{In})} \times K$ 型の非負行列，アクティベーション行列 $\boldsymbol{V}_n$ は $K \times J$ 型の非負行列となる．これらの行列を式 (3.5) に用いると，式中の $\boldsymbol{T}_n\boldsymbol{V}_n$ のサイズは $I^{(\mathrm{In})} \times J$ となる．しかし，式 (3.5) で使っているスペクトログラム $\boldsymbol{X}_m^{(\mathrm{Ex})}$ のサイズは $I^{(\mathrm{Ex})} \times J$ である．$I^{(\mathrm{Ex})} \neq I^{(\mathrm{In})}$ であるから，式 (3.5) において，$[\boldsymbol{T}_n\boldsymbol{V}_n]_{i,j}$ と $\boldsymbol{x}_{ij}^{(\mathrm{Ex})}$ の間で周波数ビンのインデクス i のとる範囲が異なる．つまり，ILRMA の空間モデル及び音源モデルの最適化に用いる STFT の窓長が互いに異なっている場合，アルゴリズム内部の計算に不都合が生じ，各モデルの反復更新を進めることができない．

前述の問題を解決するために，Fig. 3.2 と同様の方法を適用し，音源モデルの更新式により得られた分散行列 $\boldsymbol{R}_n = \boldsymbol{T}_n\boldsymbol{V}_n$ とスペクトログラム $\boldsymbol{X}_m^{(\mathrm{Ex})}$ のサイズを一致させることを考える．分散行列 $\boldsymbol{R}_n$ は音源モデルの窓関数 $\boldsymbol{\omega}^{(\mathrm{In})}$ から得られた表現であるため，$\boldsymbol{\omega}^{(\mathrm{In})}$ による時間周波数表現から $\boldsymbol{\omega}^{(\mathrm{Ex})}$ による時間周波数表現への変換，すなわち Fig. 3.2 の逆変換を施す形となる．なお，Fig. 3.2 において，空間モデルと音源モデルの立ち位置を入れ替えることで，逆変換が得られる．ここで，分散行列 $\boldsymbol{R}_n$ はパワースペクトログラムであるため，これに対して逆 STFT を適用する際には，位相を付与しなければならない[*2]．このとき，分散行列 $\boldsymbol{R}_n$ に付与する位相としては，様々なものが考えられる．つまり，Fig. 3.2 の方法をそのまま適用すると，分散行列 $\boldsymbol{R}_n$ に対する位相付与の任意性という問題に直面する．スペクトログラムの周波数ビン数は STFT の窓長に依存するため，$\boldsymbol{\omega}^{(\mathrm{Ex})}$ と $\boldsymbol{\omega}^{(\mathrm{In})}$ の長さを揃えることで，$\boldsymbol{T}_n\boldsymbol{V}_n$ と $\boldsymbol{X}_m^{(\mathrm{Ex})}$ のサイズが一致し，位相付与の任意性を回避できる．しかし，この解決法は提案手法における「ILRMA の空間モデル及び音源モデルの最適化に異なる窓長の時間周波数表現を用いる」という戦略と矛盾する．そこで，提案手法では，空間モデルと音源モデルの最適化更新に異なる時間周波数解像度のスペクトログラムを用いるとともに，Fig. 3.2 内の二つの窓関数の実際の長さを一致させるという方法によって，この二律背反の状態を解決する．その方法の具体的な内容については，次節で説明する．なお，位相付与の任意性に関しては，本論文の趣旨から外れるため，これ以上は踏み込まないこととする．

3.4　多重解像度時間周波数表現に基づく ILRMA

提案手法では，ILRMA の空間モデルと音源モデルにそれぞれ異なる窓長の窓関数を導入し，各変数の最適化を行う．その際，前節で述べた位相付与の任意性を回避するため，各モデルの変数更新において，Fig. 3.3 に示すように，STFT における窓関数の有効幅（細さ）のみを変化させることにより，見かけ上の窓長を変化させるという方法を提案する．この方法では，実際の窓長は変化しないため，二つのスペクトログラム $\boldsymbol{X}_m^{(\mathrm{Ex})}$ と $\boldsymbol{Y}_n^{(\mathrm{In})}$ のサイズが一致する．これによって，二つの行列 $\boldsymbol{T}_n\boldsymbol{V}_n$ と $\boldsymbol{X}_m^{(\mathrm{Ex})}$ のサイズも一致し，式 (3.5) における問題が解消される．このとき，提案手法のアルゴリズムでは，Fig. 3.2 の逆変換を適用する必要がなくなる．つまり，常に Fig. 3.2 に示す一方向の変換のみを反復して適用することになる．なお，窓関数の有効幅を変化させる方法については，次節で説明する．

改めて，空間モデル及び音源モデルの最適化に用いる窓関数をそれぞれ $\boldsymbol{\omega}^{(\mathrm{Ex})} \in \mathbb{R}^Q$ 及び $\boldsymbol{\omega}^{(\mathrm{In})} \in \mathbb{R}^Q$ とする．これらの窓関数は，Fig. 3.3 に示すように，有効幅は異なっているが，実際の窓長は同一である．提案手法では，前者の $\boldsymbol{Y}_n^{(\mathrm{Ex})}$ が空間モデルの最適化に，後者の $\boldsymbol{\omega}^{(\mathrm{In})}$ が音源モデルの最適化に用いられる．これらの解像度が異なる二つの時間周波数表現の変換に

[*2] 逆 STFT の対象は複素スペクトログラムであるため，位相情報を失ったパワースペクトログラムに対しては，逆 STFT をそのまま適用することができない．

は，次式で表される，時間領域を経由した方法を使用する．

$$\boldsymbol{Y}_n^{(\mathrm{In})} \leftarrow \mathrm{STFT}_{\boldsymbol{\omega}^{(\mathrm{In})}}(\mathrm{ISTFT}_{\widetilde{\boldsymbol{\omega}}^{(\mathrm{Ex})}}(\boldsymbol{Y}_n^{(\mathrm{Ex})})) \tag{3.6}$$

この処理の概念は，Fig. 3.2 に示す方法と一致する．式 (3.6) の時間周波数表現の変換は，従来の ILRMA における分離行列の反復更新則 (2.37)–(2.40) と，NMF 音源モデルの反復更新則 (2.41) 及び (2.42) の間に，PB 法と共に挿入される．すなわち，空間モデルの最適化に用いた時間周波数表現 $\boldsymbol{Y}_n^{(\mathrm{Ex})}$ から，音源モデルの最適化に用いる分離信号の時間周波数表現 $\boldsymbol{Y}_n^{(\mathrm{In})}$ へと，時間領域を介して解像度の変換を行っている．Fig. 3.3 の方法を基にした上記の処理により，ILRMA の空間モデル及び音源モデルのそれぞれで，独立した時間周波数表現を利用することが可能となる．

式 (3.6) により，提案手法ではスペクトログラム無矛盾性の担保も行っている．これは，分離信号のスペクトログラム $\boldsymbol{Y}_n$ を異なる解像度の無矛盾なスペクトログラムの集合に射影する操作であるともいえる．特に，空間モデル最適化用の窓関数 $\boldsymbol{\omega}^{(\mathrm{Ex})}$ と音源モデル最適化用の窓関数 $\boldsymbol{\omega}^{(\mathrm{In})}$ が一致する場合，式 (3.6) は consistent ILRMA におけるスペクトログラム無矛盾性を担保する式 (2.46) に一致する．つまり，この場合において，提案手法のアルゴリズムは consistent ILRMA のアルゴリズムに一致する．従って，提案手法は，consistent ILRMA を複数の解像度の時間周波数表現に一般化した手法とも解釈できる．提案手法を視覚的に表現した図を Fig. 3.4 に示す．ここで，図中の赤色，青色，及び紫色の矢印はそれぞれ STFT の適用，ILRMA における分離行列 $\boldsymbol{W}_i$ の反復更新，及び ISTFT の適用を表す．この図から，提案手法では，空間モデル及び音源モデルの最適化に用いる時間周波数表現が二層構造をなしていることが見て取れる．なお，図中の青色の矢印は，音源モデルの変数更新で得られた分散行列 $\boldsymbol{R}_n$ を空間モデルの変数更新に使っていることから生じる表現であり，Fig. 3.2 のような時間周波数表現の変換は行っていない．

3.5　Chebyshev 窓に基づく 2 種類の窓関数の設計

一般的な窓関数の例として Hann 窓を挙げ，その時間波形及び周波数特性を Fig. 3.5 に示す．Fig. 3.5 のように，窓関数の周波数特性にはメインローブ及びサイドローブと呼ばれるピークが存在する．窓関数には，メインローブの幅が狭いこと，そしてサイドローブの大きさが小さいことが求められる．しかし，この二つの特徴の間にはトレードオフが存在する．このことから，トレードオフの限界となる窓関数は何かという問題，すなわちメインローブ幅が与えられた下でのサイドローブピークの最大値が最も小さい窓関数は何かという問題が提起される．窓関数，窓関数の周波数特性，及びメインローブ幅をそれぞれ $\boldsymbol{\omega} = [\,\omega[1], \omega[2], \ldots, \omega[q], \ldots, \omega[Q]\,]^{\mathrm{T}} \in \mathbb{R}^{Q}$，$\boldsymbol{\Omega} = [\,\Omega[1], \Omega[2], \ldots, \Omega[i], \ldots, \Omega[I]\,]^{\mathrm{T}} \in \mathbb{R}^{I}$，及び $2i_c$ とすると，この問題は次式で表される．

$$\underset{\boldsymbol{\omega}}{\mathrm{Minimize}} \ \ \|\mathrm{sidelobes}(\boldsymbol{\Omega})\|_\infty \ \ \text{s.t.} \ \ \sum_{q=1}^{Q} \omega[q] = 1 \tag{3.7}$$

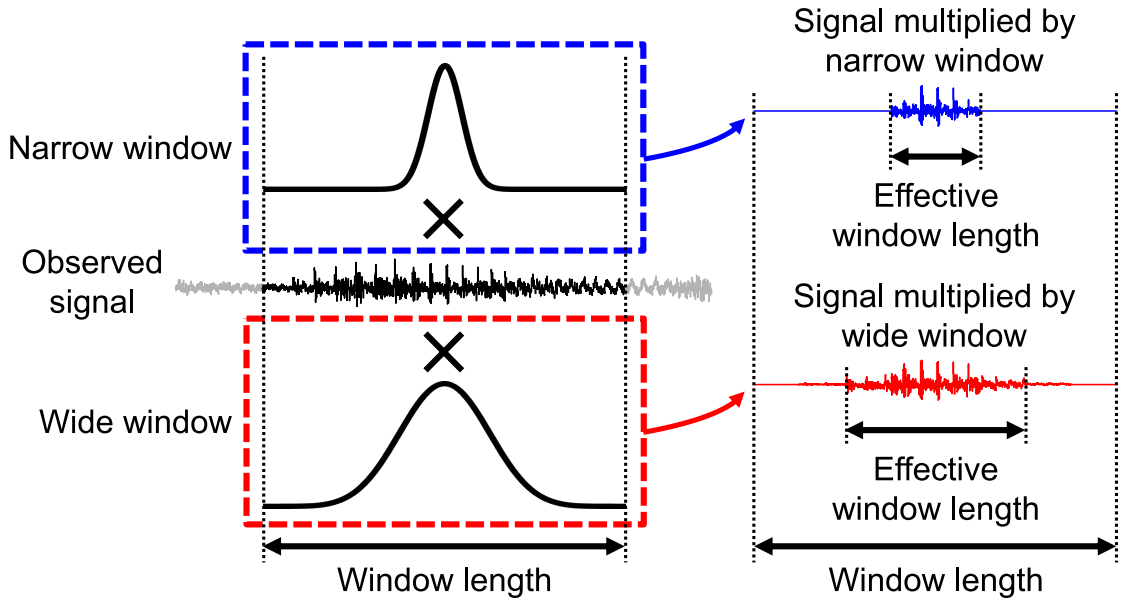


Fig. 3.3. Narrow and wide window functions used in proposed algorithm.

ここで，$\|\cdot\|_\infty$ 及び sidelobes$(\cdot)$ はそれぞれ Chebyshev ノルム及び窓関数の周波数特性におけるサイドローブの集合であり，次式で定義される．

$$\text{sidelobes}(\boldsymbol{\Omega}) = \{\Omega[i] \mid i > i_c\} \tag{3.8}$$

$$\|A\|_\infty = \max_{x \in A} |x| \tag{3.9}$$

ただし，A は複素数値からなる集合である．よって，式 (3.7) は次式のようにも書き表せる．

$$\underset{\boldsymbol{\omega}}{\text{Minimize}} \ \max_{i > i_c} |\Omega[i]| \ \ \text{s.t.} \ \ \sum_{q=1}^{Q} \omega[q] = 1 \tag{3.10}$$

なお，式 (3.7) を満たす窓関数には定数倍の任意性があるため，$\sum_q \omega[q] = 1$ を制約条件として定めている．また，式 (3.7) はサイドローブピークの最大値 $\|\text{sidelobes}(\boldsymbol{\Omega})\|_\infty$ が定められた下での，メインローブ幅が最も狭い窓関数を求める問題と等価であるから，次式のように表すこともできる．

$$\underset{\boldsymbol{\omega}}{\text{Minimize}} \ i_c \ \ \text{s.t.} \ \ \Omega[0] = 1, |\Omega[i]| \leq \alpha \ \ \forall |i| \geq i_c \tag{3.11}$$

ただし，α は与えられたサイドローブピークの内，最大のピーク値である．ここで，式 (3.11) における条件 $\Omega[0] = 1$ は式 (3.7) における条件 $\sum_q \omega[q] = 1$ に対応している．式 (3.11) を満たす窓関数は Dolph-Chebyshev 窓 [37] と呼ばれる．本論文では，Dolph-Chebyshev 窓を単に Chebyshev 窓と呼ぶ．長さ Q の Chebyshev 窓に DFT を施した関数は次式で与えられる．

$$\Omega[i] = \frac{T_{Q-1}\left(\beta \cos \frac{\pi(i-1)}{Q}\right)}{T_{Q-1}\left(\beta\right)} \tag{3.12}$$

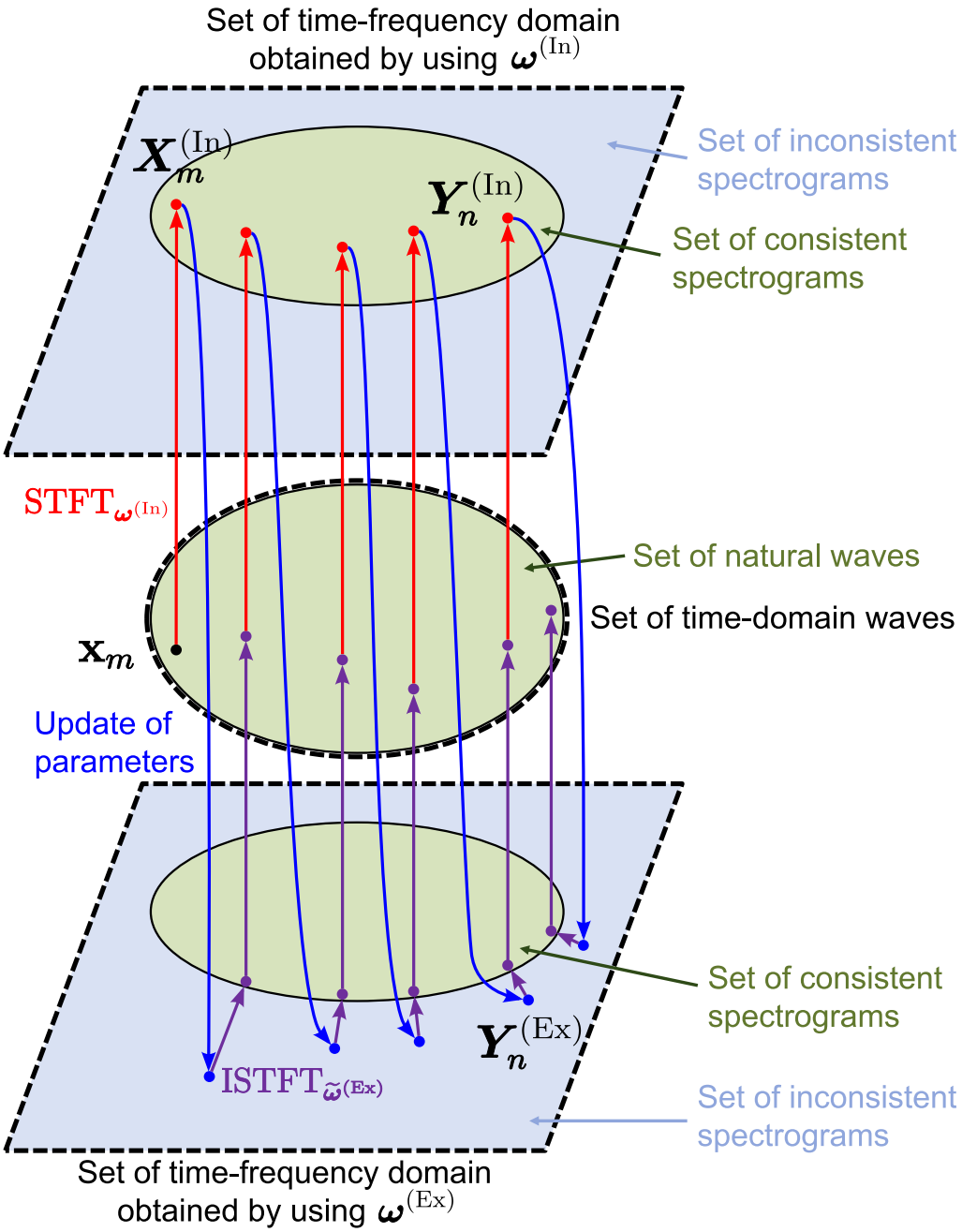


Fig. 3.4. Overview of proposed ILRMA. Red arrows indicate STFT using $\boldsymbol{\omega}^{(\mathrm{In})}$. Purple arrows indicate ISTFT using $\widetilde{\boldsymbol{\omega}}^{(\mathrm{Ex})}$. Blue arrows indicate iterative update of parameters in ILRMA.

ここで，β は次式で表される．

$$\beta = \cosh\left(\frac{1}{Q}\cosh^{-1} 10^{\frac{\alpha}{20}}\right) \tag{3.13}$$

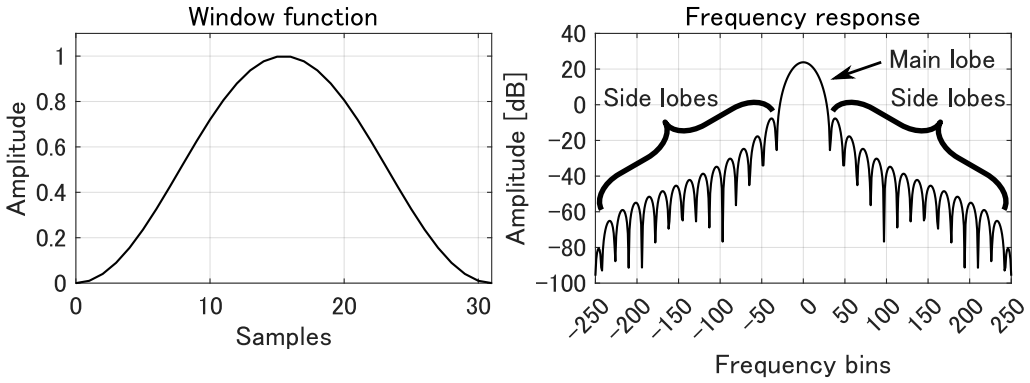


Fig. 3.5. Waveform and frequency response of Hann window.

また，$T_b(\cdot)$ は第一種 Chebyshev 多項式であり，任意の実数 x に対して次式で定義される．

$$T_b(x) = \begin{cases} \cos\left(b\cos^{-1}x\right) & \text{if } -1 \le x \le 1 \\ \cosh\left(b\cosh^{-1}x\right) & \text{if } x > 1 \\ (-1)^b\cosh\left(b\cosh^{-1}(-x)\right) & \text{if } x < -1 \end{cases} \tag{3.14}$$

ただし，b は任意の非負整数である．Chebyshev 窓は，式 (3.12) に逆 DFT を施した関数を正規化して得られる．本論文では，サイドローブピークの最大値 α をサイドローブレベルという窓関数のパラメタで与える．このとき，サイドローブレベルは窓関数の有効幅を左右するパラメタとなる．

Figs. 3.6(a)–(c) に $Q = 64$，$\alpha = 60$ [dB], 200 [dB], 500 [dB] の正規化を行っていない Chebyshev 窓の時間波形及び周波数特性を示す．Figs. 3.6(a) 及び (b) から，メインローブのピークとサイドローブのピークの差がそれぞれ 60 dB と 200 dB であることが確認できる．なお，Fig. 3.6(c) において，メインローブのピークとサイドローブのピークの差が 500 dB となっていないのは，計算機イプシロンに起因する現象である．また，Chebyshev 窓はサイドローブが等リプルであるという性質を持っていることも確認できる．さらに，サイドローブレベルが大きくなるにつれて，Chebyshev 窓の有効幅は細くなっていくことも見て取れる．

3.6 最適化アルゴリズム

提案手法における最適化アルゴリズムの擬似コード[*3]を Algorithm 1 に示す．Algorithm 1 中の 6 及び 7 行目が NMF 音源モデルの更新，8–10 行目が IP に基づく分離行列の更新，11–14 行目が m_{ref} 番目のチャネルへの PB 法，15 行目が異なる解像度の時間周波数表現の変換を表す．

[*3] 記法は algorithms という TeX パッケージに準拠する．

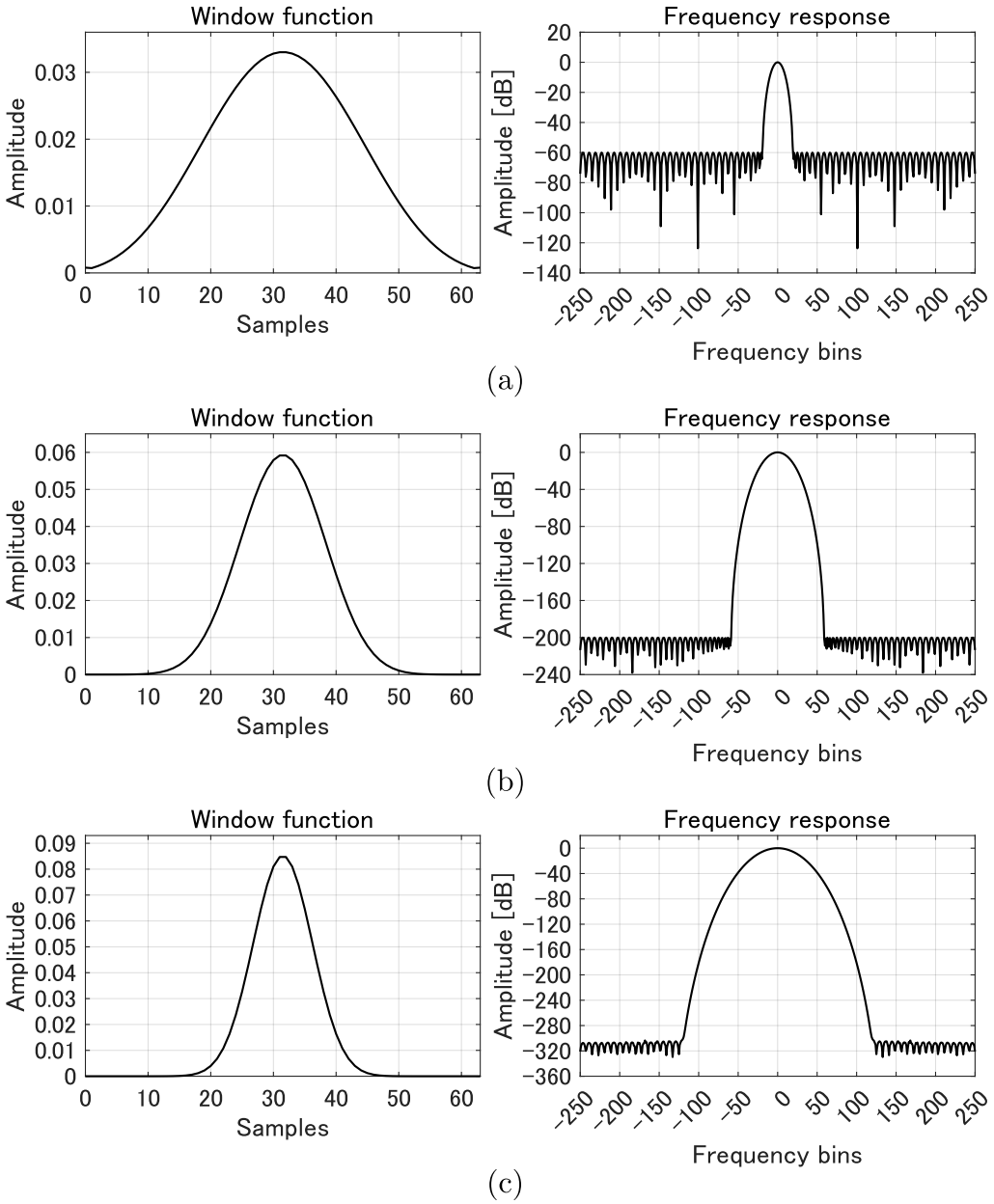


Fig. 3.6. Waveforms and frequency responses of non-normalized Chebyshev window. Side lobe levels are (a) 60 dB, (b) 200 dB, and (c) 500 dB. As value of α increases, window function gets narrower.

3.7　本章のまとめ

本章では，多重解像度時間周波数表現に基づく ILRMA を提案し，その具体的な処理について説明した．その中で，提案手法の利点，実現における問題点，及びその問題への解決策についても述べた．次章では，提案手法の空間モデル及び音源モデルの最適化に用いる窓長を変化

Algorithm 1 Proposed ILRMA

Input: $\{\mathbf{x}[l]\}_{l=1}^{L}$, maxIter
Output: $\{\mathbf{y}[l]\}_{l=1}^{L}$

Initialize $\{\boldsymbol{T}_n\}_{n=1}^{N}, \{\boldsymbol{V}_n\}_{n=1}^{N}, \{\boldsymbol{W}_i\}_{i=1}^{I}$
$\boldsymbol{X}_m^{(\mathrm{In})} = \mathrm{STFT}_{\boldsymbol{\omega}^{(\mathrm{In})}}(\mathbf{x})$
$\boldsymbol{X}_m^{(\mathrm{Ex})} = \mathrm{STFT}_{\boldsymbol{\omega}^{(\mathrm{Ex})}}(\mathbf{x})$
$y_{ijn}^{(\mathrm{In})} = \boldsymbol{w}_{in}^{\mathrm{H}} \boldsymbol{x}_{ij}^{(\mathrm{In})} \ \forall i, j, n$
for iter $= 1, 2, \ldots,$ maxIter **do**
  $\boldsymbol{T}_n \leftarrow \boldsymbol{T}_n \odot \left\{ \frac{\left[|\boldsymbol{Y}_n^{(\mathrm{In})}|^{\cdot 2} \odot (\boldsymbol{T}_n \boldsymbol{V}_n)^{\cdot -2} \right] \boldsymbol{V}_n^{\mathrm{T}}}{(\boldsymbol{T}_n \boldsymbol{V}_n)^{\cdot -1} \boldsymbol{V}_n^{\mathrm{T}}} \right\}^{\cdot \frac{1}{2}} \ \forall n$
  $\boldsymbol{V}_n \leftarrow \boldsymbol{V}_n \odot \left\{ \frac{\boldsymbol{T}_n^{\mathrm{T}} \left[|\boldsymbol{Y}_n^{(\mathrm{In})}|^{\cdot 2} \odot (\boldsymbol{T}_n \boldsymbol{V}_n)^{\cdot -2} \right]}{\boldsymbol{T}_n^{\mathrm{T}} (\boldsymbol{T}_n \boldsymbol{V}_n)^{\cdot -1}} \right\}^{\cdot \frac{1}{2}} \ \forall n$
  $\boldsymbol{U}_{in} \leftarrow \frac{1}{J} \sum_j \frac{1}{[\boldsymbol{T}_n \boldsymbol{V}_n]_{i,j}} \boldsymbol{x}_{ij}^{(\mathrm{Ex})} \boldsymbol{x}_{ij}^{(\mathrm{Ex})\mathrm{H}} \ \forall i, n$
  $\boldsymbol{w}_{in} \leftarrow (\boldsymbol{W}_i \boldsymbol{U}_{in})^{-1} \boldsymbol{e}_n \ \forall i, n$
  $\boldsymbol{w}_{in} \leftarrow \boldsymbol{w}_{in} \left(\boldsymbol{w}_{in}^{\mathrm{H}} \boldsymbol{U}_{in} \boldsymbol{w}_{in} \right)^{-\frac{1}{2}} \ \forall i, n$
  $\lambda_{in} \leftarrow \left[\boldsymbol{W}_i^{-1} \right]_{m_{\mathrm{ref}}, n} \ \forall i, n$
  $\boldsymbol{w}_{in} \leftarrow \lambda_{in} \boldsymbol{w}_{in} \ \forall i, n$
  $y_{ijn}^{(\mathrm{Ex})} \leftarrow \boldsymbol{w}_{in}^{\mathrm{H}} \boldsymbol{x}_{ij}^{(\mathrm{Ex})} \ \forall i, j, n$
  $[\boldsymbol{T}_n]_{i,k} \leftarrow |\lambda_{in}|^2 [\boldsymbol{T}_n]_{i,k} \ \forall i, k, n$
  $\boldsymbol{Y}_n^{(\mathrm{In})} \leftarrow \mathrm{STFT}_{\boldsymbol{\omega}^{(\mathrm{In})}}(\mathrm{ISTFT}_{\widetilde{\boldsymbol{\omega}}^{(\mathrm{Ex})}}(\boldsymbol{Y}_n^{(\mathrm{Ex})})) \ \forall n$
end for
$\mathbf{y} = \mathrm{ISTFT}_{\widetilde{\boldsymbol{\omega}}^{(\mathrm{Ex})}}(\boldsymbol{Y}_n^{(\mathrm{Ex})})$

させ，音源分離の実験を行う．そして，得られた結果における傾向の評価，及び，従来手法である consistent ILRMA と提案手法の分離性能の比較を行う．

第 4 章

実験

4.1 まえがき

本章では，提案手法の空間モデル及び音源モデルの最適化に用いる窓長を，サイドローブレベルというパラメタによって，様々に変化させることで，従来手法である consistent ILRMA と提案手法の比較実験を行う．4.2 節では，実験に使用する音源や提案手法におけるパラメタ設定等の実験条件について述べる．4.3 節では，全実験結果の内，傾向がよく確認できる代表的なデータの実験結果について述べる．4.3 節に掲載しなかった実験結果については付録 A に示す．最後に，4.4 節で本章の総括を行う．

4.2 実験条件

RWCP データベース [38] 収録のインパルス応答 E2A（$T_{60} = 300$ ms）による 2 音源の畳み込み混合を行い，10 曲分の 2 チャネル観測信号を生成した．ここで，用いたインパルス応答の収録条件は Fig. 4.1 に示す通りである．そして，これらの観測信号に対する BSS 性能を従来手法の consistent ILRMA 及び提案手法の 2 手法で比較した．Table 4.1 に実験で用いた音源信号（ドライソース）の詳細を示す．評価指標は音源対歪み比（source-to-distortion ratio: SDR）[39] の改善量（improvement of SDR: SDRi）を用いた．STFT における窓関数 $\boldsymbol{\omega}^{(\mathrm{Ex})}$ 及び $\boldsymbol{\omega}^{(\mathrm{In})}$ には長さ 256 ms（4096 点）の Chebyshev 窓を用い，サイドローブレベルパラメタ α を変化させることにより，窓関数の有効幅を変更した．空間モデル及び音源モデルの最適化における両方の Chebyshev 窓のサイドローブレベルには，Table 4.2 に示す値を用い，これらを総当たり的に変化させた．その他の実験条件は Table 4.2 に示す通りである．なお，STFT の実装には DGTtool [40, 41] を用いた．

4.3 実験結果

各サイドローブレベルに対して，異なる乱数シードを用いて 10 回実験を行った際の平均 SDRi を付録 A に示す．ここで，これらの図における色の濃淡は，暗いほど平均 SDRi が低

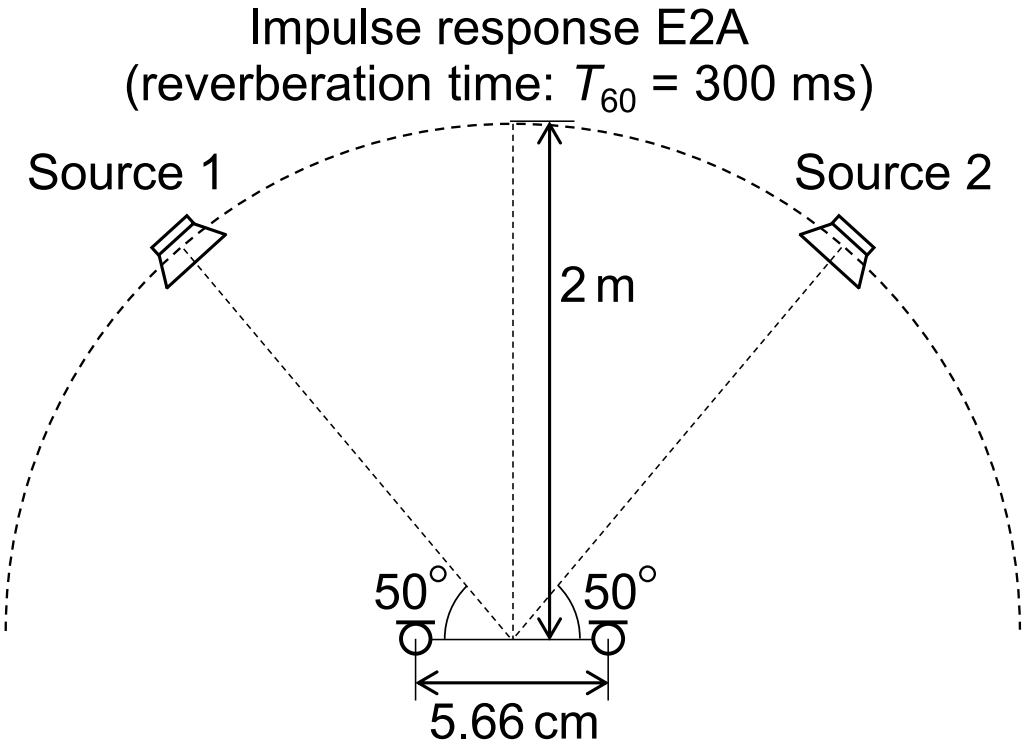


Fig. 4.1. Recording conditions of impulse response E2A.

Table 4.1. Music sources obtained from SiSEC2011 dataset used as dry sources

Song number	Data name	Source (1/2)	Length [s]
1	another_dreamer-the_ones_we_love	drums/guitar	25.6
2	another_dreamer-the_ones_we_love	guitar/vocals	25.6
3	bearlin-roads	acoustic_guit_main/vocals	14.6
4	bearlin-roads	drums/bass	14.6
5	bearlin-roads	piano/acoustic_guit_main	14.6
6	fort_minor-remember_the_name	violins_synth/vocals	24.6
7	fort_minor-remember_the_name	vocals/drums	24.6
8	tamy-que_pena_tanto_faz	guitar/vocals	13.6
9	ultimate_nz_tour	drums/vocals	18.6
10	ultimate_nz_tour	guitar/synth	18.6

く，明るいほど平均 SDRi が高いことを示している．なお，Figs. 4.3–4.6 の太線で囲んだ対角部分は，空間モデルと音源モデルの両方で同一の窓関数を使用していることから，従来の consistent ILRMA の結果に対応する．また，それ以外の部分（非対角部分）は異なる解像度の時間周波数表現を用いる提案手法の結果に対応する．

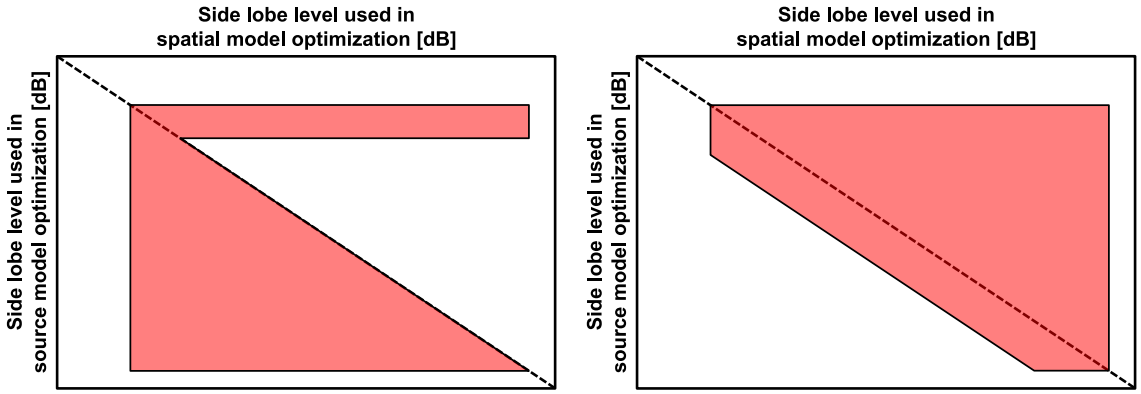
まず，実験結果の全体的な傾向を述べる．行方向及び列方向を音源モデル及び空間モデルのサイドローブレベルとして実験結果を表した図は，平均 SDRi が比較的高い場所の分布によって，Figs. 4.2(a) 及び (b) の 2 種類に大別される．これらの図は，Figs. A.1–A.11 に対応し，図中の赤色の部分は，比較的高い平均 SDRi が観測されたパラメタに相当する場所を示してい

Table 4.2. Experimental conditions

Window shift length	32 ms
Side lobe level of Chebyshev window [dB]	{20, 30, 40, 50, 60, 70, 80, 90, 100, 120, 150, 200, 300, 500, 700, 1000, 1500, 2000, 3000}
Number of NMF bases K	10
Initialization of parameters	$\boldsymbol{W}_i$: identity matrix $\boldsymbol{T}_n$ and $\boldsymbol{V}_n$: uniformly distributed random values in the range $(0, 1)$
Number of iterations for parameter updates	200
Number of trials	10 with different random seeds
Reference channel m_{ref} for projection back technique	1

る．Fig. 4.2(a) は，比較的高い平均 SDRi を示す部分が対角部分の下側に広がっているという性質を表している．また，音源モデルのサイドローブレベルがおよそ 50〜60 dB であるときに，空間モデルのサイドローブレベルが大きくなっても，比較的高い分離性能を示し続けるという特徴も有している．Figs. A.2 及び A.3 より，Fig. 4.2(a) の傾向は Song 1 及び Song 9 の結果に見られる．Fig. 4.2(b) は，対角部分のやや下側から上側にかけて，比較的高い平均 SDRi を示す部分が広がっているという性質を表している．Figs. A.4，A.7，A.8，及び A.11 より，Fig. 4.2(b) の傾向は Song 3，Song 6，Song 7，及び Song 10 の結果に見られる．また，Fig. A.3 では，対角部分の下側と上側両方に平均 SDRi が高い部分が広がっている．よって，Song 2 の結果は Figs. 4.2(a) 及び (b) の傾向を部分的に含んでいるといえる．さらに，Fig. A.6 では，音源モデルのサイドローブレベルが 50 dB，空間モデルのサイドローブレベルが 70〜500 dB という横に細長い領域で，平均 SDRi が高い部分が広がっている．このことから，Song 5 の結果は Fig. 4.2(a) の傾向を部分的に含んでいるといえる．なお，残りの Song 4 及び Song 8 の結果である Figs. A.5 及び A.9 では，平均 SDRi の分布は Figs. 4.2(a) 及び (b) のどちらにも該当しないことが見て取れる．

次に，10 曲の観測信号全ての平均 SDRi，Fig. 4.2(a) の傾向を示す Song 9 の平均 SDRi，及び Fig. 4.2(b) の傾向を示す Song 3 と Song 6 の平均 SDRi を取り上げて，分離性能の評価を行う．10 曲の観測信号全て，Song 3，Song 6，及び Song 9 の平均 SDRi のうち，サイドローブレベルが 40〜1500 [dB] のデータをそれぞれ Figs. 4.3–4.6 に示す．Figs. 4.4–4.6 において，最も高い平均 SDRi を示しているのは，空間モデル用の窓関数 $\boldsymbol{\omega}^{(\mathrm{Ex})}$ 及び音源モデル用の窓関数 $\boldsymbol{\omega}^{(\mathrm{In})}$ のサイドローブレベルがそれぞれ 50 dB と 60 dB，60 dB と 90 dB，及び 70 dB と 120 dB の場合であり，対角成分（consistent ILRMA の結果）から外れた場所に位置していることが分かる．また，Fig. 4.3 より，10 曲の観測信号全ての平均 SDRi においても，空間モデル用の窓関数 $\boldsymbol{\omega}^{(\mathrm{Ex})}$ のサイドローブレベルが 80 dB や 90 dB の場合は，音源



(a) (b)

Fig. 4.2. Rough classification of experimental results by distribution of better SDRi group. Red parts indicate group of relatively better SDRis. (a) is trend of Song 1 and Song 9. (b) is trend of Song 3, Song 6, Song 7, and Song 10.

Average SDRi	Side lobe level used in spatial model optimization [dB]														
Side lobe level used in source model optimization [dB]	40	50	60	70	80	90	100	120	150	200	300	500	700	1000	1500
40	3.25	4.55	1.21	1.24	1.66	1.64	1.80	1.70	1.61	1.69	2.09	2.25	2.00	1.73	0.89
50	3.89	6.86	4.77	4.73	4.76	4.43	4.33	4.53	5.05	5.74	5.72	5.82	5.18	4.62	3.96
60	2.68	6.83	9.17	7.92	7.87	6.88	6.78	6.80	6.99	6.91	6.73	6.34	6.13	5.29	4.22
70	2.02	5.87	6.96	8.42	6.96	6.85	6.50	6.20	5.80	5.41	5.16	5.29	5.30	5.19	4.07
80	1.18	3.28	7.20	8.10	7.58	7.66	7.08	6.63	6.28	5.84	5.30	4.96	4.92	4.73	3.82
90	0.58	1.16	6.61	8.39	7.62	7.61	7.32	6.80	6.47	6.07	5.32	4.79	4.87	4.52	3.66
100	0.22	0.80	5.86	7.72	8.19	7.62	7.38	6.89	6.41	5.82	5.27	4.93	4.86	4.42	3.44
120	-0.45	0.00	4.60	6.68	8.30	8.03	7.38	6.75	6.33	5.70	5.22	5.01	4.80	4.33	3.46
150	-1.32	-0.56	4.08	4.20	6.56	6.23	7.49	6.65	6.21	5.64	5.10	4.84	4.81	4.22	3.16
200	-2.41	-1.92	5.19	4.29	4.93	6.73	7.12	6.15	6.22	5.67	5.14	4.61	4.25	4.14	3.26
300	-1.12	2.08	5.38	5.07	4.95	4.56	5.07	4.08	5.85	5.42	5.10	4.34	4.04	4.18	3.35
500	0.34	2.31	4.46	4.13	4.35	4.19	3.63	2.93	3.53	4.80	4.48	4.25	3.99	3.88	3.24
700	-0.91	3.34	5.33	4.13	4.62	4.50	3.85	3.35	3.41	4.01	3.56	4.40	3.89	3.84	2.92
1000	-0.49	3.70	4.74	4.33	4.32	3.97	3.70	3.45	3.55	2.95	3.07	4.44	3.80	3.65	2.96
1500	1.43	3.79	4.53	3.90	3.96	4.21	3.93	4.10	3.50	2.38	2.32	3.82	3.45	3.48	2.94

Poor Good

Fig. 4.3. Average SDRi using 10 different initializations of all 10 music. Range of side lobe level is from 40 to 1500.

モデル用の窓関数 $\boldsymbol{\omega}^{(\mathrm{In})}$ のサイドローブレベルを 100 dB や 120 dB とした方が，consistent ILRMA よりも高い性能を示すことが読み取れる．以上から，音源モデル及び空間モデルの最適化に同一の窓関数を用いることが常に最良の結果を与えるとは限らないことが実験的に示された．

Average SDRi	Side lobe level used in spatial model optimization [dB]														
Side lobe level used in source model optimization [dB]	40	50	60	70	80	90	100	120	150	200	300	500	700	1000	1500
40	3.17	3.84	1.26	2.50	1.90	1.32	1.01	1.14	1.27	1.29	1.00	1.45	0.21	0.16	-1.02
50	3.49	12.49	3.62	3.66	3.00	1.74	2.11	2.79	2.78	3.86	5.26	5.85	5.25	4.69	2.83
60	-1.10	13.06	12.61	10.14	9.43	9.50	9.19	8.66	8.41	7.01	7.76	5.25	4.61	4.23	3.36
70	2.32	7.98	1.49	11.91	11.11	9.64	9.57	8.80	8.11	7.74	6.73	5.55	5.20	3.61	4.38
80	-1.90	-0.53	6.16	12.78	11.81	10.81	10.34	9.45	9.07	8.60	7.57	6.14	5.23	4.49	3.77
90	-3.11	-2.51	9.62	12.62	11.77	11.35	9.70	9.03	8.67	8.49	7.38	6.50	5.83	4.51	4.28
100	-2.56	-0.35	8.74	12.32	12.02	10.81	10.12	9.51	8.95	7.99	6.89	6.22	5.64	4.40	4.53
120	-3.54	-1.26	10.88	9.76	11.94	10.85	10.63	9.27	8.81	7.77	6.84	6.66	6.31	5.62	4.71
150	-3.86	-2.29	8.20	9.23	11.36	11.42	10.61	9.23	8.73	7.82	6.83	6.41	6.33	6.11	4.58
200	-7.45	-7.61	5.11	5.29	6.89	9.98	10.87	10.36	9.05	8.22	6.75	6.32	6.32	6.64	5.05
300	-3.94	10.08	5.20	5.59	4.86	4.82	6.39	8.89	10.10	8.46	7.17	6.58	6.47	6.58	5.29
500	2.28	11.98	4.38	-1.31	2.83	2.55	2.57	4.05	4.82	8.43	7.69	6.69	6.22	6.11	5.52
700	-4.74	12.43	3.96	3.22	3.33	4.03	2.74	2.65	3.68	4.96	7.61	6.26	5.56	5.48	4.12
1000	-5.30	11.65	3.60	3.39	2.91	3.46	3.06	3.32	3.50	3.34	5.36	7.01	5.86	4.86	4.29
1500	2.78	10.72	2.55	2.67	2.18	2.51	2.11	2.06	2.49	2.89	1.43	5.66	6.40	4.45	4.22

Poor Good

Fig. 4.4. Average SDRi using 10 different initializations of Song 3. Range of side lobe level is from 40 to 1500.

Average SDRi	Side lobe level used in spatial model optimization [dB]														
Side lobe level used in source model optimization [dB]	40	50	60	70	80	90	100	120	150	200	300	500	700	1000	1500
40	-0.23	3.26	0.16	-0.64	-1.76	-1.73	-1.44	-1.57	-1.22	0.04	1.13	0.79	0.43	0.17	-1.21
50	3.05	13.08	3.50	2.26	2.67	1.92	2.42	2.92	3.13	4.42	5.11	6.49	7.27	4.74	4.77
60	5.77	10.95	12.21	8.31	10.03	11.18	10.76	10.72	11.54	10.76	8.69	9.92	9.31	8.46	6.86
70	4.26	12.67	13.33	10.48	11.26	11.36	11.03	11.12	11.13	10.79	10.32	10.12	9.04	9.34	7.19
80	1.48	8.30	13.54	10.80	11.78	12.24	10.56	10.87	10.49	10.56	10.16	9.83	9.72	8.89	7.32
90	0.20	4.17	13.78	12.04	12.29	11.22	11.19	10.35	10.42	9.82	9.25	9.09	9.66	9.32	7.43
100	-1.51	0.52	13.49	12.54	11.14	11.73	11.39	10.27	10.45	9.87	9.29	9.66	9.93	9.41	7.02
120	-0.35	-0.45	11.52	13.20	12.82	11.23	11.56	10.62	10.55	9.89	9.19	9.94	9.46	9.34	7.09
150	-1.70	0.06	5.56	6.66	11.54	11.96	12.43	11.13	11.09	9.61	9.22	9.40	9.25	8.98	6.39
200	-3.82	-1.97	5.94	9.05	3.76	7.28	9.77	11.91	11.51	11.22	9.75	8.91	7.93	8.76	6.38
300	-2.70	-1.64	7.15	7.62	5.71	5.25	3.82	6.28	8.53	9.94	9.92	7.94	7.78	8.18	6.50
500	-1.36	-1.14	7.65	5.67	4.16	2.87	2.20	4.43	4.55	4.61	9.55	8.57	8.06	7.37	7.14
700	-1.76	0.99	4.69	4.52	4.32	5.21	4.69	4.72	3.56	3.46	3.49	9.85	8.30	7.56	5.90
1000	-1.88	-0.02	4.78	3.61	5.87	5.46	4.25	4.08	3.43	2.61	-0.57	7.79	7.90	7.85	5.94
1500	-0.08	-0.90	-0.73	2.26	4.49	4.88	4.88	4.08	4.82	2.07	3.10	6.55	7.26	7.15	5.51

Poor Good

Fig. 4.5. Average SDRi using 10 different initializations of Song 6. Range of side lobe level is from 40 to 1500.

4.4　本章のまとめ

本章では，提案手法の空間モデル及び音源モデルの最適化における窓長を様々に変化させ，音源分離の実験を行った．そして，得られた結果の傾向を評価し，従来手法である consistent ILRMA との比較を行った．実験結果から，ILRMA の空間モデル及び音源モデルの最適化に

Average SDRi	Side lobe level used in spatial model optimization [dB]														
Side lobe level used in source model optimization [dB]	40	50	60	70	80	90	100	120	150	200	300	500	700	1000	1500
40	1.32	8.62	5.02	4.58	4.60	4.61	4.68	4.78	4.94	4.80	5.45	5.79	5.75	5.56	5.03
50	1.93	10.03	10.07	10.10	10.37	10.25	10.21	10.25	10.00	9.87	9.99	9.81	8.85	9.90	9.42
60	2.06	1.77	12.19	10.13	7.60	7.88	8.63	9.02	9.32	10.00	9.59	9.32	10.56	10.63	8.02
70	1.78	1.05	13.64	6.33	5.81	5.54	5.09	4.91	3.86	3.40	3.77	6.51	7.46	9.54	7.03
80	1.58	-0.17	14.25	8.07	5.86	6.80	5.84	4.50	4.61	3.51	3.55	4.87	5.97	6.48	6.99
90	1.35	0.74	9.63	13.17	3.78	5.45	5.57	4.30	3.98	3.30	2.89	4.45	5.01	6.62	6.20
100	-0.12	1.38	9.78	8.21	10.59	4.74	4.92	4.55	3.52	3.17	3.00	4.53	4.76	5.95	6.21
120	0.31	-0.08	13.65	15.32	12.84	11.30	4.49	3.95	2.59	2.56	2.45	4.25	4.74	5.28	6.26
150	-1.96	-0.36	12.88	11.96	6.78	2.90	6.37	3.74	2.44	2.24	2.53	3.29	4.21	4.20	4.51
200	-2.20	0.22	12.75	10.26	10.68	12.53	11.68	3.70	2.07	2.04	2.53	3.29	3.60	3.96	4.46
300	-0.42	5.74	12.31	11.95	12.35	11.43	11.70	9.83	8.39	3.02	2.53	3.37	3.36	3.56	4.26
500	-0.09	4.27	4.15	11.16	3.84	9.17	4.61	5.59	11.12	11.92	1.44	3.40	3.51	3.21	2.67
700	0.53	6.77	13.89	12.79	11.61	11.72	11.13	10.27	10.59	10.99	7.11	4.00	3.42	3.23	2.51
1000	1.08	6.96	13.10	12.43	12.41	11.74	11.64	10.31	10.49	10.39	9.78	2.83	3.31	3.39	3.26
1500	1.65	9.04	12.67	11.61	12.15	12.30	12.05	10.92	8.81	3.20	5.46	5.24	2.01	3.46	3.37

Poor Good

Fig. 4.6. Average SDRi using 10 different initializations of Song 9. Range of side lobe level is from 40 to 1500.

同一の窓関数を用いることが必ずしも最良の結果を与えるとは限らないことが示された．また，音源分離の対象とする観測信号に応じて，分離性能の傾向が変化し，そこには少なくとも 2 つ以上のパターンが存在することを確認した．

第 5 章

結言

本論文では，ILRMA に対して，空間モデル及び音源モデルの最適化にそれぞれ異なる窓長の窓関数を導入する手法を提案した．これは，ILRMA における空間モデルと音源モデルのそれぞれに適切な窓長が存在するのではないかという考えに基づくものである．提案手法では，窓関数の有効幅を変化させるという方法により，スペクトログラムのサイズを変えることなく，見かけ上の窓長を変化させ，空間モデルと音源モデルに異なる時間周波数解像度のスペクトログラムを用いることで生じる問題を解消した．窓関数のサイドローブレベル（サイドローブピークの最大値）とメインローブ幅（有効幅）がトレードオフの関係にあることに注目し，窓関数の有効幅を変化させるために，サイドローブレベルを設定可能な Chebyshev 窓を用いた．実験結果から，提案手法は consistent ILRMA と比較して，分離精度が向上する場合があることが確認された．

最後に，今後の展望を述べる．4.3 節で述べたように，実験結果には，観測信号毎に少なくとも 2 つ以上のパターンが存在していた．このことから，実験結果の傾向と観測信号の構成要素（音源信号の楽器の種類や時間周波数構造等）の関係性を調べることで，各々の観測信号が持つ固有の特徴が提案手法の分離性能に及ぼす影響を明確にできると予想される．これを基に，音源分離に用いる信号の特徴に応じて，空間モデルと音源モデルに最適な窓関数を設定できると考える．また，式 (2.36) で表される ILRMA のコスト関数に，複数の解像度の時間周波数表現を導入することで，提案手法のコスト関数を設計することがおそらく可能である．このコスト関数の設計により，提案手法をより理論的な面から捉えることができると思われる．以上の，観測信号に応じた窓関数の最適化及び提案手法のより理論的な解析によって，提案手法の更なる分離性能の向上が期待できる．

謝辞

本論文は，香川高等専門学校電気情報工学科北村研究室にて行われた研究に基づくものです．

まず，本研究を進めるにあたり，ご多忙のところ熱心にご指導くださいました指導教員の北村大地講師に心より感謝申し上げます．北村大地講師には，論文執筆や研究に関する議論など，細部にわたるまで丁寧にご指導いただきました．研究室に配属されるまで，研究というものに触れたことがなかった私にとっては，大変貴重な経験となりました．また，実験を行う際の計算機サーバや，Notion 上での充実したナレッジ，学生が相談しやすい環境など，研究を行う上で快適な環境作りを徹底して行っていただけたことに対しては，感謝の念に堪えません．

本論の副査である重田和弘教授には，論文の構成や記述に関して大変有益な助言を頂き，大変お世話になりました．ここに厚く御礼申し上げます．

早稲田大学の矢田部浩平講師には，共同研究を通じ，研究に関する議論から，国内学会の原稿の記述に関することまで，多数のご助言をいただきました．心より感謝申し上げます．

北村研究室の先輩方が執筆された論文は本論文を執筆する上で参考にさせていただきました．また，北村研究室同期の川口翔也氏・蓮池郁也氏・溝渕悠朔氏・村田佳斗氏には，ゼミや日頃のディスカッションのほか，1 年に亘る研究室生活を様々な面で支えていただきました．ここに感謝申し上げます．

最後になりますが，現在に至るまで私の学生生活を金銭的に支え，暖かく見守って下さった両親には感謝の念に堪えません．これまで本当にありがとうございました．

参考文献

[1] H. Sawada, N. Ono, H. Kameoka, D. Kitamura, and 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. e12, pp. 1–14, 2019.

[2] P. Comon, "Independent component analysis, a new concept?," *Signal Processing*, vol. 36, no. 3, pp. 287–314, 1994.

[3] A. Hyvärinen, "Independent component analysis: recent advances," *Philosophical Transactions of the Royal Society A*, vol. 371, 2013.

[4] 浅野 太, "音のアレイ信号処理-音源の定位・追跡と分離-," **日本音響学会編 音響テクノロジーシリーズ**, **コロナ社**, 2018

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

[6] N. Murata, S. Ikeda, and A. Ziehe, "An approach to blind source separation based on temporal structure of speech signals," *Neurocomputing*, vol. 41, no. 1-4, pp. 1–24, 2001.

[7] H. Sawada, R. Mukai, S. Araki, and S. Makino, "A robust and precise method for solving the permutation problem of frequency-domain blind source separation," *IEEE Transactions on Speech and Audio Processing*, vol. 12, no. 5, pp. 530–538, Sep. 2004.

[8] H. Saruwatari, T. Kawamura, T. Nishikawa, A. Lee, and K. Shikano, "Blind source separation based on a fast-convergence algorithm combining ICA and beamforming," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 14, no. 2, pp. 666–678, 2006.

[9] A. Hiroe, "Solution of permutation problem in frequency domain ICA using multivariate probability density functions," *Proc. International Conference on Latent Variable Analysis and Signal Separation*, pp.601–608, 2006.

[10] T. Kim, T. Eltoft, and T.-W. Lee, "Independent vector analysis: an extension of ICA to multivariate components," *Proc. International Conference on Latent Variable Analysis and Signal Separation*, pp. 165–172, 2006.

[11] T. Kim, H. T. Attias, S.-Y. Lee, and T.-W. Lee, "Blind source separation exploiting

higher-order frequency dependencies," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 15, no. 1, pp. 70–79, 2007.

[12] D. D. Lee and H. S. Seung, "Algorithms for non-negative matrix factorization" *Proc. International Conference on Neural Information Processing Systems*, pp. 556–562, 2000.

[13] D. R. Hunter and K. Lange, "A tutorial on MM algorithms," *The American Statistician*, vol. 58, no. 1, pp. 30–37, 2004.

[14] M. Nakano, H. Kameoka, J. Le Roux, Y. Kitano, N. Ono, and S. Sagayama, "Convergence-guaranteed multiplicative algorithms for nonnegative matrix factorization with β-divergence," *Proc. IEEE International Workshop on Machine Learning for Signal Processing*, 2010, pp. 283–288.

[15] N. Ono and S. Miyabe, "Auxiliary-function-based independent component analysis for super-Gaussian sources," *Proc. Latent Variable Analysis and Signal Separation*, pp. 165–172, 2010.

[16] N. Ono, "Stable and fast update rules for independent vector analysis based on auxiliary function technique," *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pp. 189–192, 2011.

[17] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, no. 6755, pp. 788–791, 1999.

[18] D. Kitamura, N. Ono, H. Sawada, H. Kameoka, and H. Saruwatari, "Determined blind source separation unifying independent vector analysis and nonnegative matrix factorization," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 24, no. 9, pp. 1626–1641, 2016.

[19] D. Kitamura, N. Ono, H. Sawada, H. Kameoka, and H. Saruwatari, "Determined blind source separation with independent low-rank matrix analysis," in *Audio Source Separation*, S. Makino, Ed., pp. 125–155. Springer, Cham, 2018.

[20] K. Yatabe and D. Kitamura, "Determined blind source separation via proximal splitting algorithm," *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 776–780, 2018.

[21] K. Yatabe and D. Kitamura, "Time-frequency-masking-based determined BSS with application to sparse IVA," *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 715–719, 2019.

[22] K. Yatabe and D. Kitamura, "Determined BSS based on time-frequency masking and its application to harmonic vector analysis," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 29, pp. 1609–1625, 2021.

[23] J. L. Roux, H. Kameoka, N. Ono, and S. Sagayama, "Fast signal reconstruction from magnitude STFT spectrogram based on spectrogram consistency," *Proc. DAFx*, 2010.

[24] J. L. Roux and E. Vincent, “Consistent Wiener filtering for audio source separation,” *IEEE Signal Processing Letters*, vol. 20, no. 3, pp. 217–220, 2013.

[25] K. Yatabe, “Consistent ICA: Determined BSS meets spectrogram consistency,” *IEEE Signal Processing Letters*, vol. 27, pp. 870–874, 2020.

[26] 豊島直，北村大地，矢田部浩平, “スペクトログラム無矛盾性を用いた独立低ランク行列分析,” **日本音響学会** *2020* **年秋季研究発表会講演論文集**, pp. 291–294, 2020.

[27] D. Kitamura and K. Yatabe, “Consistent independent low-rank matrix analysis for determined blind source separation,” *EURASIP Journal on Advances in Signal Processing*, vol. 2020, no. 46, 35 pages, 2020.

[28] D. Kitamura, N. Ono, and H. Saruwatari, “Experimental analysis of optimal window length for independent low-rank matrix analysis,” *Proc. European Signal Processing Conference*, pp. 1210–1214, 2017.

[29] S. Araki, R. Mukai, S. Makino, T. Nishikawa, and H. Saruwatari, “The fundamental limitation of frequency domain blind source separation for convolutive mixtures of speech,” *IEEE Transactions on Speech and Audio Processing*, vol. 11, no. 2, 2003.

[30] K. Matsuoka and S. Nakashima, “Minimal distortion principle for blind source separation,” *Proc. International Conference on Latent Variable Analysis and Signal Separation*, pp. 722–727, 2001.

[31] C. Févotte, N. Bertin, and J.-L. Durrieu, “Nonnegative matrix factorization with the Itakura-Saito divergence. With application to music analysis,” *Neural Computation*, vol. 21, no. 3, pp. 793–830, 2009.

[32] D. FitzGerald, M. Cranitch, and E. Coyle, “On the use of the beta divergence for musical source separation,” *Proc. IET Irish Signals and Systems Conference*, 2009.

[33] D. Kitamura, H. Saruwatari, K. Yagi, K. Shikano, Y. Takahashi, and K. Kondo, “Music signal separation based on supervised nonnegative matrix factorization with orthogonality and maximum-divergence penalties,” *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, vol. E97-A, no. 5, pp. 1113–1118, 2014.

[34] A. J. Bell and T. J. Sejnowski, “An information-maximization approach to blind separation and blind deconvolution,” *Neural Computation*, no. 7, pp. 1129–1159, 1995.

[35] S. Douglas and S. Amari, “Natural-gradient adaptation,” in *Unsupervised adaptive filtering*, Ed. S. Haykin, vol. I, pp. 13–61, Wiley, 2000.

[36] Y. Iwase and D. Kitamura, “Supervised audio source separation based on nonnegative matrix factorization with cosine similarity penalty,” *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, vol. E105-A, no. 6, 2022.

[37] C. L. Dolph, “A Current Distribution for Broadside Arrays Which Optimizes the Relationship between Beam Width and Side-Lobe Level,” *Proceedings of the IRE*, vol. 34, pp. 335–348, 1946.

[38] S. Nakamura, K. Hiyane, F. Asano, T. Nishiura, and T. Yamada, “Acoustical sound database in real environments for sound scene understanding and hands-free speech recognition,” *Proc. International Conference on Language Resources and Evaluation*, pp. 965–968, 2000.

[39] E. Vincent, R. Gribonval, and C. Févotte, “Performance measurement in blind audio source separation,” *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 14, no. 4, pp. 1462–1469, 2006.

[40] K. Yatabe, DGTtool. Zenodo, 2021, doi:10.5281/ZENODO.5010751.

[41] 矢田部浩平, “短時間フーリエ変換および離散ガボール変換の MATLAB 実装について,” **日本音響学会** *2021* **年秋季研究発表会講演論文集**, pp. 253–256, 2021.

発表文献一覧

国内学会

1. **細谷泰稚**, 北村大地, 矢田部浩平, “解像度の異なる複数の時間周波数表現を用いた独立低ランク行列分析,” 日本音響学会 2022 年春季研究発表会講演論文集, 1-1P-2, 2022.

付録 A

4 章の実験に対する全結果

Figs. A.1–A.11 に 4 章の実験により得られた結果を示す．ここで，Fig. A.1 は 10 曲の観測信号全ての平均 SDRi，Figs. A.2–A.11 は Song 1〜Song 10 の平均 SDRi である．なお，4.3 節に掲載した Figs. 4.3–4.6 は，サイドローブレベルが 40〜1500 [dB] のデータである．それに対して，ここで示す Figs. A.1–A.11 は，サイドローブレベルが 20〜3000 [dB] のデータという点において異なっている．つまり，Figs. A.1，A.4，A.6，及び A.10 は，それぞれ Figs. 4.3–4.6 を含んだデータとなっている．

Average SDRi	Side lobe level used in spatial model optimization [dB]																		
Side lobe level used in source model optimization [dB]	20	30	40	50	60	70	80	90	100	120	150	200	300	500	700	1000	1500	2000	3000
20	1.88	2.24	0.99	-0.10	-0.57	0.14	-0.23	-0.09	-0.14	-0.14	-0.25	-0.47	-0.14	-0.26	-0.61	-0.77	-1.42	-2.08	-3.39
30	1.96	1.95	3.43	0.73	-0.24	0.09	0.37	0.55	0.38	0.63	0.53	0.44	0.33	0.16	0.03	-0.17	-0.57	-1.28	-3.13
40	1.55	1.41	3.25	4.55	1.21	1.24	1.66	1.64	1.80	1.70	1.61	1.69	2.09	2.25	2.00	1.73	0.89	-0.10	-2.48
50	1.50	1.55	3.89	6.86	4.77	4.73	4.76	4.43	4.33	4.53	5.05	5.74	5.72	5.82	5.18	4.62	3.96	2.51	-0.72
60	1.44	1.24	2.68	6.83	9.17	7.92	7.87	6.88	6.78	6.80	6.99	6.91	6.73	6.34	6.13	5.29	4.22	2.94	0.35
70	1.39	1.49	2.02	5.87	6.96	8.42	6.96	6.85	6.50	6.20	5.80	5.41	5.16	5.29	5.30	5.19	4.07	2.85	0.28
80	0.81	0.92	1.18	3.28	7.20	8.10	7.58	7.66	7.08	6.63	6.28	5.84	5.30	4.96	4.92	4.73	3.82	2.70	-0.05
90	-0.41	0.75	0.58	1.16	6.61	8.39	7.62	7.61	7.32	6.80	6.47	6.07	5.32	4.79	4.87	4.52	3.66	2.55	-0.10
100	-2.23	-0.35	0.22	0.80	5.86	7.72	8.19	7.62	7.38	6.89	6.41	5.82	5.27	4.93	4.86	4.42	3.44	2.69	0.01
120	-5.39	-1.96	-0.45	0.00	4.60	6.68	8.30	8.03	7.38	6.75	6.33	5.70	5.22	5.01	4.80	4.33	3.46	2.47	0.11
150	-5.65	-1.80	-1.32	-0.56	4.08	4.20	6.56	6.23	7.49	6.65	6.21	5.64	5.10	4.84	4.81	4.22	3.16	2.23	-0.11
200	-6.60	-3.08	-2.41	-1.92	5.19	4.29	4.93	6.73	7.12	6.15	6.22	5.67	5.14	4.61	4.25	4.14	3.26	2.23	0.02
300	-3.67	-1.35	-1.12	2.08	5.38	5.07	4.95	4.56	5.07	4.08	5.85	5.42	5.10	4.34	4.04	4.18	3.35	2.19	0.44
500	-2.10	-0.66	0.34	2.31	4.46	4.13	4.35	4.19	3.63	2.93	3.53	4.80	4.48	4.25	3.99	3.88	3.24	2.22	0.25
700	-3.24	-1.35	-0.91	3.34	5.33	4.13	4.62	4.50	3.85	3.35	3.41	4.01	3.56	4.40	3.89	3.84	2.92	2.28	0.29
1000	-4.10	-1.28	-0.49	3.70	4.74	4.33	4.32	3.97	3.70	3.45	3.55	2.95	3.07	4.44	3.80	3.65	2.96	1.91	-0.33
1500	-2.82	0.01	1.43	3.79	4.53	3.90	3.96	4.21	3.93	4.10	3.50	2.38	2.32	3.82	3.45	3.48	2.94	2.10	-0.41
2000	-2.10	0.18	1.51	3.45	4.23	4.59	4.65	4.01	3.79	3.89	3.56	3.25	2.58	2.33	2.90	2.95	2.93	2.09	-0.13
3000	-2.27	0.34	1.42	3.42	3.48	4.02	4.01	3.91	4.09	3.94	3.68	3.03	2.78	1.52	2.03	2.15	2.95	2.34	0.40

Poor Good

Fig. A.1. Average SDRi using 10 different initializations of all 10 music.

Average SDRi	Side lobe level used in spatial model optimization [dB]																		
Side lobe level used in source model optimization [dB]	20	30	40	50	60	70	80	90	100	120	150	200	300	500	700	1000	1500	2000	3000
20	4.88	3.48	-1.17	-2.13	-3.04	-2.08	-1.80	-1.74	-2.15	-2.39	-2.29	-1.97	-1.56	-1.09	-1.15	-1.51	-2.38	-2.99	-2.78
30	5.05	4.06	-0.31	-2.59	-3.54	-2.83	-2.16	-1.67	-1.69	-1.90	-1.77	-2.21	-1.75	-1.20	-2.16	-1.61	-2.55	-2.95	-3.79
40	4.96	2.87	1.02	3.94	-2.74	-2.95	-3.06	-2.59	-2.14	-1.49	-2.65	-2.28	-2.43	-0.94	-0.59	-0.97	-1.60	-2.56	-4.33
50	5.25	3.97	3.09	5.46	3.40	3.90	4.76	4.21	2.68	3.56	4.55	5.57	3.17	0.86	0.29	-0.11	-0.71	-1.12	-4.32
60	2.59	2.81	1.74	5.95	8.31	6.73	6.62	0.79	-1.16	0.09	2.85	3.67	2.76	2.13	1.80	0.43	-0.14	-1.38	-4.09
70	2.53	2.36	3.03	5.68	2.33	8.95	2.75	1.08	0.47	-0.19	-0.71	-1.86	-1.40	1.34	1.23	0.94	-0.17	-1.38	-3.12
80	2.59	3.18	2.70	5.70	1.65	5.27	2.13	3.35	2.80	2.18	2.12	0.87	-1.27	-1.73	-0.29	0.01	-0.08	-0.64	-3.81
90	0.34	2.04	2.36	4.84	0.86	4.06	5.05	5.18	5.10	4.60	2.97	2.13	0.91	-1.09	-1.44	-1.28	-1.00	-1.29	-3.64
100	-1.36	0.13	1.50	3.84	2.93	5.31	5.16	6.40	5.52	3.92	3.18	2.13	1.60	0.24	0.12	-1.10	-1.13	-1.12	-3.70
120	-6.22	-0.71	0.06	1.69	-1.27	5.94	3.32	6.15	6.38	4.53	3.52	1.36	1.28	0.44	-0.10	-0.97	-1.46	-1.67	-3.78
150	-6.94	-1.24	-0.38	0.19	-0.74	-0.55	5.91	3.67	4.78	3.46	2.09	1.76	0.94	0.45	-0.63	-1.20	-1.22	-1.84	-4.59
200	-6.19	-1.84	-1.45	-1.24	4.35	1.83	5.68	5.11	6.46	2.81	2.01	1.27	0.50	-0.66	-1.56	-0.83	-1.01	-1.82	-4.62
300	-1.88	-0.88	0.87	0.56	3.04	4.70	5.28	4.42	4.85	2.88	4.30	-0.21	0.05	-1.47	-1.79	-0.28	-0.78	-1.71	-4.29
500	1.25	-1.81	0.62	-0.40	3.42	4.27	4.64	4.37	3.96	2.85	3.70	2.77	1.11	-2.17	-2.05	-0.46	-1.03	-1.99	-4.46
700	-3.37	-1.29	-0.74	-0.13	3.73	4.07	3.95	3.62	4.48	4.37	4.07	1.50	1.26	-0.88	-1.75	-0.47	-1.04	-1.92	-4.52
1000	-4.84	-2.26	-1.80	1.69	3.73	4.04	4.66	4.37	3.32	3.05	3.73	3.28	1.86	1.96	-1.20	-0.79	-1.30	-2.05	-4.69
1500	1.45	2.41	3.28	1.51	0.41	3.16	4.24	4.57	3.20	3.44	3.36	1.94	1.06	0.38	0.64	-1.01	-1.46	-1.84	-4.57
2000	1.61	2.68	2.93	1.35	2.95	0.53	3.93	3.43	3.65	3.11	1.88	1.77	1.02	0.08	0.11	-0.90	-1.60	-2.00	-4.49
3000	1.72	2.82	3.34	1.57	0.10	2.63	4.13	3.30	3.23	2.95	2.76	1.74	1.14	0.04	0.00	0.22	-0.50	-1.87	-3.52

Poor Good

Fig. A.2. Average SDRi using 10 different initializations of Song 1.

Average SDRi	Side lobe level used in spatial model optimization [dB]																		
Side lobe level used in source model optimization [dB]	20	30	40	50	60	70	80	90	100	120	150	200	300	500	700	1000	1500	2000	3000
20	0.04	0.69	1.03	-1.76	0.35	1.53	0.65	0.48	0.85	0.70	1.11	1.12	0.65	0.72	0.25	0.30	-0.20	-0.46	-1.49
30	0.07	1.18	-0.17	0.30	-0.89	0.04	0.06	0.15	0.50	1.21	1.57	1.51	1.73	1.14	0.69	0.77	0.26	-0.21	-1.46
40	-0.12	1.19	7.01	2.22	-1.22	-0.13	0.11	0.46	0.61	0.35	0.28	1.27	1.67	3.82	3.00	3.06	1.43	0.17	-2.31
50	-1.36	0.93	7.35	1.24	6.59	0.68	0.83	1.41	0.96	1.81	4.92	7.75	7.20	7.19	5.78	5.29	4.35	2.26	-2.64
60	-1.53	-1.95	6.64	8.17	8.85	8.85	10.04	9.55	9.46	9.60	8.50	7.81	7.42	6.74	6.30	5.54	4.39	2.72	0.14
70	-1.83	-1.82	-0.64	7.88	7.55	9.57	10.08	10.13	9.55	9.31	8.78	8.18	7.67	7.18	6.47	5.91	4.06	2.90	0.45
80	-1.99	-2.33	-1.79	6.40	7.59	8.43	10.15	9.92	9.72	9.05	8.62	8.08	7.52	7.10	6.68	5.78	3.93	2.97	0.21
90	-2.48	-2.33	-2.02	-0.03	7.60	7.45	9.43	9.65	9.39	8.97	8.68	8.16	7.11	6.69	6.52	5.44	4.02	3.05	0.31
100	-5.30	-2.22	-1.93	-1.63	7.67	7.22	8.79	9.42	9.45	9.02	8.72	7.88	7.12	6.64	6.19	5.27	3.99	3.17	0.37
120	-7.65	-3.12	-2.84	-2.29	5.69	7.05	7.98	8.44	9.22	8.77	8.48	7.70	6.96	6.69	6.22	5.50	4.05	2.96	0.62
150	-8.04	-4.85	-4.26	-4.86	3.81	5.74	7.18	7.65	8.19	8.77	8.02	7.42	6.97	6.63	6.09	5.51	4.21	2.97	0.84
200	-3.62	-5.54	-5.76	-4.73	7.81	7.39	8.17	8.99	7.42	7.57	7.77	7.34	7.14	6.87	6.19	5.70	3.94	2.97	0.77
300	-2.24	-1.67	-1.29	4.48	8.69	8.45	9.47	9.50	10.20	8.42	7.68	7.60	7.26	6.12	5.89	5.46	4.59	3.19	0.32
500	-4.86	-2.67	-0.88	4.83	7.05	8.22	7.87	7.65	10.08	7.66	8.56	7.49	6.55	5.98	5.76	5.27	4.69	3.30	0.14
700	-5.86	-2.92	-2.70	4.95	8.29	8.96	9.51	9.91	9.94	9.15	8.89	7.26	5.61	5.93	5.47	5.26	4.49	3.24	0.47
1000	-7.00	-2.41	1.11	5.51	7.81	9.19	9.25	8.57	8.86	8.42	7.87	6.85	6.20	5.93	5.78	5.50	4.83	3.56	0.44
1500	-4.28	1.48	2.07	6.13	7.43	8.43	8.51	8.51	7.57	8.10	7.63	5.93	6.03	5.31	5.43	5.37	4.98	3.54	0.38
2000	-0.11	1.15	2.21	5.80	7.63	8.15	6.85	6.05	8.00	7.31	7.13	6.46	6.11	4.94	4.95	5.29	5.16	3.53	0.33
3000	-0.20	1.00	2.77	5.83	6.44	6.99	6.53	6.51	6.34	6.63	6.77	5.87	5.56	5.71	5.12	4.45	4.33	3.52	0.41

Poor Good

Fig. A.3. Average SDRi using 10 different initializations of Song 2.

Average SDRi	Side lobe level used in spatial model optimization [dB]																			
Side lobe level used in source model optimization [dB]	20	30	40	50	60	70	80	90	100	120	150	200	300	500	700	1000	1500	2000	3000
20	-0.67	1.57	0.26	-2.21	-2.36	-1.13	-1.02	-1.13	-1.49	-1.71	-1.97	-2.45	-2.41	-2.55	-2.84	-3.18	-3.52	-4.02	-4.36
30	-0.48	0.77	3.45	-0.80	-1.33	0.00	-0.51	-0.90	-1.07	-1.36	-1.98	-1.69	-2.14	-2.14	-2.29	-2.55	-2.95	-3.43	-5.31
40	-0.02	0.52	3.17	3.84	1.26	2.50	1.90	1.32	1.01	1.14	1.27	1.29	1.00	1.45	0.21	0.16	-1.02	-2.21	-4.03
50	-0.04	0.27	3.49	12.49	3.62	3.66	3.00	1.74	2.11	2.79	2.78	3.86	5.26	5.85	5.25	4.69	2.83	1.30	-1.40
60	-0.07	-1.32	-1.10	13.06	12.61	10.14	9.43	9.50	9.19	8.66	8.41	7.01	7.76	5.25	4.61	4.23	3.36	3.35	0.05
70	0.08	-0.78	2.32	7.98	1.49	11.91	11.11	9.64	9.57	8.80	8.11	7.74	6.73	5.55	5.20	3.61	4.38	3.23	0.11
80	-0.21	0.09	-1.90	-0.53	6.16	12.78	11.81	10.81	10.34	9.45	9.07	8.60	7.57	6.14	5.23	4.49	3.77	3.79	0.07
90	-1.81	-0.52	-3.11	-2.51	9.62	12.62	11.77	11.35	9.70	9.03	8.67	8.49	7.38	6.50	5.83	4.51	4.28	3.78	0.00
100	-3.84	-0.69	-2.56	-0.35	8.74	12.32	12.02	10.81	10.12	9.51	8.95	7.99	6.89	6.22	5.64	4.40	4.53	3.90	0.14
120	-5.24	-4.07	-3.54	-1.26	10.88	9.76	11.94	10.85	10.63	9.27	8.81	7.77	6.84	6.66	6.31	5.62	4.71	3.55	0.18
150	-8.61	-4.44	-3.86	-2.29	8.20	9.23	11.36	11.42	10.61	9.23	8.73	7.82	6.83	6.41	6.33	6.11	4.58	3.42	0.17
200	-13.21	-8.29	-7.45	-7.61	5.11	5.29	6.89	9.98	10.87	10.36	9.05	8.22	6.75	6.32	6.32	6.64	5.05	3.63	0.64
300	-4.67	-2.35	-3.94	10.08	5.20	5.59	4.86	4.82	6.39	8.89	10.10	8.46	7.17	6.58	6.47	6.58	5.29	3.78	0.71
500	-4.80	-2.31	2.28	11.98	4.38	-1.31	2.83	2.55	2.57	4.05	4.82	8.43	7.69	6.69	6.22	6.11	5.52	3.57	0.36
700	-3.83	-5.39	-4.74	12.43	3.96	3.22	3.33	4.03	2.74	2.65	3.68	4.96	7.61	6.26	5.56	5.48	4.12	3.62	0.41
1000	-5.04	-5.57	-5.30	11.65	3.60	3.39	2.91	3.46	3.06	3.32	3.50	3.34	5.36	7.01	5.86	4.86	4.29	3.33	-0.25
1500	-4.95	-1.02	2.78	10.72	2.55	2.67	2.18	2.51	2.11	2.06	2.49	2.89	1.43	5.66	6.40	4.45	4.22	3.45	0.24
2000	-5.31	-0.68	3.70	7.34	2.32	2.50	1.83	1.85	2.11	1.54	1.70	2.10	2.12	3.23	4.63	4.56	3.73	3.23	1.25
3000	-5.83	-1.25	3.29	5.05	1.57	2.28	1.94	1.71	2.06	1.95	1.90	1.83	2.12	2.33	3.03	3.75	3.81	3.14	2.02

Poor Good

Fig. A.4. Average SDRi using 10 different initializations of Song 3.

Average SDRi	Side lobe level used in spatial model optimization [dB]																			
Side lobe level used in source model optimization [dB]	20	30	40	50	60	70	80	90	100	120	150	200	300	500	700	1000	1500	2000	3000
20	3.86	1.36	-2.43	-3.30	-0.42	-1.60	-2.16	-1.10	-1.13	-0.90	-0.79	-1.14	-0.86	-0.95	-1.27	-1.65	-2.67	-3.29	-4.74
30	3.98	0.44	0.22	-2.62	-0.49	-0.65	-1.70	0.31	0.16	0.20	0.05	0.16	0.36	-0.41	-0.54	-0.50	-1.48	-2.32	-4.14
40	-1.24	-1.60	-0.10	0.21	-0.99	0.06	-1.65	-0.59	-0.29	-0.09	-0.06	0.18	0.20	-1.36	-1.23	-1.63	-2.68	-3.38	-4.58
50	0.03	-2.30	-0.38	1.51	-0.67	-0.61	-1.02	-1.09	-0.90	-1.05	-0.97	-0.80	-1.30	-0.93	-1.30	-0.95	-1.04	-2.63	-4.24
60	2.21	1.83	-1.38	1.60	-0.14	-0.62	-0.20	-0.06	-0.08	-0.29	-0.46	-0.09	-1.32	-1.44	-0.36	-1.62	-0.78	-2.05	-3.36
70	2.26	3.58	1.29	1.84	0.94	0.20	-0.16	-0.34	-0.69	-0.74	-0.60	-1.04	-0.82	-0.81	-0.57	-0.83	-0.98	-2.32	-3.81
80	-0.96	0.51	1.73	1.62	-0.13	1.80	0.56	0.20	-0.16	0.23	-1.00	-1.09	-1.34	-1.06	-1.04	-0.92	-1.94	-2.83	-3.08
90	-2.89	-0.72	0.41	-0.55	-0.18	2.01	1.84	0.11	0.37	0.13	0.03	-0.26	-0.91	-1.30	-0.26	-1.28	-1.68	-2.99	-3.82
100	-7.50	-1.82	0.49	0.45	-0.53	1.95	2.15	0.90	0.48	0.51	-0.66	-1.08	-0.64	-1.34	-1.15	-1.20	-2.06	-2.65	-3.75
120	-4.17	0.92	0.48	0.54	0.61	-0.04	1.69	1.10	0.30	-0.19	-0.46	-0.83	-0.64	-1.12	-1.27	-1.48	-2.11	-3.14	-3.89
150	-1.31	-1.06	-0.83	-0.06	3.97	2.48	-0.73	0.69	0.43	-0.35	-0.28	-1.25	-1.50	-1.63	-1.22	-1.66	-2.33	-3.59	-4.54
200	-2.62	-1.12	-0.49	0.84	2.77	1.63	2.51	-0.61	0.44	-0.04	-0.82	-1.69	-1.48	-1.20	-1.43	-1.85	-2.47	-3.54	-4.17
300	-3.83	-2.19	-2.12	-0.46	2.24	0.72	-0.83	0.35	-0.19	0.44	0.32	-0.72	-1.95	-1.40	-1.90	-1.79	-2.97	-3.82	-4.31
500	-0.86	-1.45	-1.52	0.66	1.02	-1.38	-1.62	-0.51	-1.32	-1.36	-1.87	0.39	-1.43	-1.94	-1.84	-1.98	-2.81	-3.71	-4.43
700	-0.89	-1.91	-1.09	1.52	3.30	-1.81	-0.33	-1.40	-1.80	-1.70	-1.38	-1.01	0.25	-2.32	-2.31	-2.15	-2.64	-3.20	-4.39
1000	-2.76	-2.07	-2.46	2.26	2.38	-0.47	-1.72	-1.39	-1.98	-1.40	-1.30	-0.27	-0.30	-0.95	-2.24	-2.35	-3.09	-3.28	-5.55
1500	-2.83	-2.75	-1.26	1.14	2.46	-0.28	-2.24	-2.25	-1.78	-1.50	-1.05	-1.08	-1.03	0.63	-1.94	-2.69	-2.67	-3.72	-5.64
2000	-4.08	-2.82	-0.78	0.74	1.49	0.57	-0.15	-1.41	-1.90	-0.66	-1.39	-1.03	-1.04	-0.14	0.29	-2.04	-3.06	-3.71	-6.13
3000	-4.25	-2.31	-0.53	-0.36	0.59	-0.09	-0.71	-1.47	-1.37	-2.04	-0.59	-1.41	-1.13	-0.87	-1.46	-1.21	-2.55	-4.04	-5.20

Poor Good

Fig. A.5. Average SDRi using 10 different initializations of Song 4.

Average SDRi	Side lobe level used in spatial model optimization [dB]																		
Side lobe level used in source model optimization [dB]	20	30	40	50	60	70	80	90	100	120	150	200	300	500	700	1000	1500	2000	3000
20	0.68	0.60	2.34	2.76	1.62	4.46	0.96	0.24	-0.84	-0.06	-0.21	-0.10	0.18	0.21	0.83	0.48	-1.09	-1.58	-2.80
30	0.37	0.38	5.80	4.67	1.14	3.32	3.75	3.59	3.29	3.63	2.53	2.12	1.11	0.36	0.24	0.06	-0.01	-1.67	-2.53
40	-0.05	0.05	1.63	6.91	2.34	2.33	2.99	3.35	3.06	2.78	2.36	1.93	1.51	0.92	0.57	0.37	0.33	-0.23	-1.09
50	0.18	1.57	2.85	6.13	3.91	6.86	7.92	4.91	5.26	5.10	5.48	6.06	6.02	7.28	4.90	2.68	3.37	1.70	-2.17
60	0.82	1.73	1.20	4.95	7.53	7.17	6.23	1.53	2.34	3.05	3.32	3.67	5.46	6.90	5.61	4.08	3.07	1.75	-1.51
70	0.63	1.31	1.67	3.96	4.62	7.09	0.28	2.47	2.26	2.62	1.98	1.79	0.68	1.52	1.56	2.74	1.62	0.98	-1.31
80	0.51	0.37	1.63	1.66	4.56	4.20	3.89	5.04	3.68	3.75	3.05	3.46	2.72	1.35	0.26	-0.18	0.57	-1.15	-2.23
90	0.21	0.92	0.88	1.31	2.63	5.21	4.54	4.22	3.66	3.94	3.38	3.79	3.20	1.70	1.34	0.44	-0.51	-2.22	-3.18
100	0.19	0.43	0.19	1.02	1.72	4.99	3.97	4.43	3.89	3.88	3.53	3.33	3.17	1.66	1.17	0.31	-0.63	-2.40	-3.47
120	-9.92	0.78	0.93	2.22	-0.38	0.81	5.38	3.72	3.77	3.50	3.46	3.25	3.08	1.45	0.89	0.16	-1.11	-2.34	-3.82
150	-2.82	-0.05	1.30	2.33	-0.49	-0.05	3.10	1.29	3.85	3.57	3.54	3.35	3.07	1.37	0.73	-0.15	-1.05	-2.01	-3.81
200	-7.09	-2.85	-1.82	-1.16	0.49	0.19	-0.25	0.17	0.37	-0.37	3.45	3.40	3.04	1.72	0.91	-0.13	-0.60	-2.17	-3.53
300	-3.21	-1.05	0.13	1.63	-0.21	-0.06	0.49	-0.17	0.80	0.36	1.88	1.36	2.04	1.74	0.96	0.06	-0.84	-1.88	-2.81
500	-0.99	1.70	2.77	0.61	1.31	0.72	-0.27	0.55	-0.06	0.09	1.23	1.08	-0.11	1.12	0.88	-0.08	-0.84	-1.69	-3.38
700	-2.94	-0.04	0.04	2.17	1.02	1.75	1.22	1.66	0.85	-0.41	0.44	0.95	0.30	0.66	0.63	-0.03	-1.06	-1.85	-3.19
1000	-3.50	0.00	0.75	1.68	1.46	2.10	1.99	1.52	1.98	2.12	1.65	-0.02	0.60	0.11	0.28	0.25	-1.11	-2.39	-3.44
1500	-1.45	0.47	1.90	1.95	1.79	2.15	2.69	3.01	3.46	2.88	2.22	0.21	1.31	0.12	0.69	-0.58	-1.01	-1.82	-3.98
2000	-1.93	1.00	1.79	2.36	2.12	2.10	2.27	2.43	2.27	3.10	3.10	1.28	0.89	-0.07	0.21	0.03	-0.97	-1.61	-4.35
3000	-1.55	1.19	1.22	2.54	2.25	2.54	2.84	2.70	2.83	2.77	3.21	2.19	0.57	-0.03	-0.22	0.21	1.37	-1.34	-2.70

Poor Good

Fig. A.6. Average SDRi using 10 different initializations of Song 5.

Average SDRi	Side lobe level used in spatial model optimization [dB]																		
Side lobe level used in source model optimization [dB]	20	30	40	50	60	70	80	90	100	120	150	200	300	500	700	1000	1500	2000	3000
20	-1.45	-1.34	-0.99	0.29	-1.93	-1.77	-1.59	-1.38	-1.86	-0.70	-1.28	-1.35	-0.96	-0.85	-1.06	-0.90	-1.43	-1.90	-3.49
30	-1.38	-2.63	1.69	-0.07	-1.46	-1.90	-2.05	-1.35	-1.40	-1.11	-0.24	-0.32	-0.07	-0.40	-0.32	-1.09	0.31	-0.73	-4.36
40	-1.24	-0.72	-0.23	3.26	0.16	-0.64	-1.76	-1.73	-1.44	-1.57	-1.22	0.04	1.13	0.79	0.43	0.17	-1.21	-2.22	-4.98
50	-1.30	-1.62	3.05	13.08	3.50	2.26	2.67	1.92	2.42	2.92	3.13	4.42	5.11	6.49	7.27	4.74	4.77	2.20	-0.23
60	-1.36	-0.90	5.77	10.95	12.21	8.31	10.03	11.18	10.76	10.72	11.54	10.76	8.69	9.92	9.31	8.46	6.86	5.51	0.72
70	-1.42	-0.87	4.26	12.67	13.33	10.48	11.26	11.36	11.03	11.12	11.13	10.79	10.32	10.12	9.04	9.34	7.19	6.17	0.77
80	-1.82	-1.33	1.48	8.30	13.54	10.80	11.78	12.24	10.56	10.87	10.49	10.56	10.16	9.83	9.72	8.89	7.32	6.45	0.49
90	-2.65	0.17	0.20	4.17	13.78	12.04	12.29	11.22	11.19	10.35	10.42	9.82	9.25	9.09	9.66	9.32	7.43	6.01	0.63
100	-1.91	-0.89	-1.51	0.52	13.49	12.54	11.14	11.73	11.39	10.27	10.45	9.87	9.29	9.66	9.93	9.41	7.02	6.23	0.13
120	-3.22	-1.61	-0.35	-0.45	11.52	13.20	12.82	11.23	11.56	10.62	10.55	9.89	9.19	9.94	9.46	9.34	7.09	5.91	0.83
150	-6.36	-4.26	-1.70	0.06	5.56	6.66	11.54	11.96	12.43	11.13	11.09	9.61	9.22	9.40	9.25	8.98	6.39	4.80	1.04
200	-6.32	-3.59	-3.82	-1.97	5.94	9.05	3.76	7.28	9.77	11.91	11.51	11.22	9.75	8.91	7.93	8.76	6.38	5.10	0.49
300	-8.62	-2.10	-2.70	-1.64	7.15	7.62	5.71	5.25	3.82	6.28	8.53	9.94	9.92	7.94	7.78	8.18	6.50	4.55	2.60
500	-2.85	-0.96	-1.36	-1.14	7.65	5.67	4.16	2.87	2.20	4.43	4.55	4.61	9.55	8.57	8.06	7.37	7.14	6.04	2.09
700	-4.47	-1.59	-1.76	0.99	4.69	4.52	4.32	5.21	4.69	4.72	3.56	3.46	3.49	9.85	8.30	7.56	5.90	5.69	2.07
1000	-6.20	-2.01	-1.88	-0.02	4.78	3.61	5.87	5.46	4.25	4.08	3.43	2.61	-0.57	7.79	7.90	7.85	5.94	4.83	3.70
1500	-6.34	-1.76	-0.08	-0.90	-0.73	2.26	4.49	4.88	4.88	4.08	4.82	2.07	3.10	6.55	7.26	7.15	5.51	4.53	3.85
2000	-5.23	-1.92	-0.19	0.10	3.88	3.82	2.91	0.25	0.79	5.17	4.71	4.77	2.44	3.73	3.05	6.28	6.00	4.51	3.12
3000	-4.19	-1.97	-0.48	3.31	2.63	4.01	5.79	6.10	5.22	5.38	4.32	3.89	3.66	2.43	-0.40	1.16	6.07	5.71	3.25

Poor Good

Fig. A.7. Average SDRi using 10 different initializations of Song 6.

Average SDRi	Side lobe level used in spatial model optimization [dB]																		
Side lobe level used in source model optimization [dB]	20	30	40	50	60	70	80	90	100	120	150	200	300	500	700	1000	1500	2000	3000
20	2.04	2.02	0.09	0.45	-1.87	-1.47	-1.17	-1.07	-0.96	-1.07	-0.96	-1.22	-1.08	-1.05	-1.09	-1.30	-1.91	-2.95	-4.95
30	2.07	2.05	4.23	1.91	-0.11	-0.92	-0.90	-0.82	-0.89	-0.18	-0.32	-0.13	0.18	0.15	0.73	1.04	0.85	0.52	-1.04
40	2.10	1.22	4.52	7.34	3.50	2.54	2.69	2.76	2.71	2.68	2.71	2.75	2.96	3.14	2.89	2.28	1.71	1.05	-2.48
50	2.02	2.19	5.97	11.98	7.44	7.78	7.65	7.69	7.54	7.29	7.22	6.98	7.23	7.90	7.48	7.40	6.90	5.97	3.69
60	2.01	1.38	6.99	11.93	12.91	12.33	11.99	11.64	11.32	10.91	10.59	10.27	9.49	8.35	7.97	7.54	6.86	6.29	4.54
70	1.84	0.08	2.25	7.86	12.41	12.17	12.08	11.79	11.50	11.28	10.74	9.94	9.60	8.61	7.92	7.39	6.98	6.33	4.59
80	0.37	-0.09	0.59	1.72	11.77	11.85	11.74	11.38	11.11	10.78	10.52	9.84	9.08	8.51	7.86	7.63	6.96	6.26	4.63
90	-1.47	-1.58	0.54	1.10	11.41	11.61	11.37	11.14	11.02	10.71	10.42	9.82	8.89	8.36	8.00	7.55	6.88	6.25	4.65
100	-6.56	-2.68	-2.44	0.24	11.21	11.11	11.35	10.98	10.83	10.58	10.30	9.63	8.89	8.28	7.87	7.54	6.81	6.26	4.71
120	-7.31	-3.81	-3.14	-2.08	1.60	7.66	11.16	10.77	10.42	10.38	10.07	9.71	8.98	8.32	7.93	7.40	6.68	6.23	4.79
150	-7.72	-0.76	-3.83	-0.72	3.44	1.89	3.23	5.69	10.58	9.99	9.75	9.40	8.89	8.23	7.70	7.28	6.61	6.14	4.78
200	-4.03	-1.24	-0.05	-0.20	2.86	0.25	6.51	7.43	8.54	9.60	9.47	9.29	8.72	7.86	7.50	7.32	6.62	6.04	4.65
300	-4.02	-2.91	-1.50	-0.16	4.85	1.85	2.13	0.77	3.19	0.54	5.29	8.34	8.29	7.92	7.72	7.21	6.65	5.98	4.67
500	-1.38	0.16	-0.40	-0.69	3.06	3.67	9.30	6.98	7.54	0.02	0.64	1.81	7.45	7.66	7.22	7.22	6.46	5.86	4.45
700	-2.24	0.08	0.02	1.54	2.65	2.26	3.07	2.61	2.34	1.68	1.51	5.39	2.46	7.04	7.06	7.10	6.01	6.03	4.52
1000	-2.60	0.16	0.66	2.00	3.05	2.04	2.05	2.02	1.60	2.26	1.68	0.52	2.50	6.13	6.86	6.96	6.07	5.42	4.38
1500	-3.99	-1.02	-0.19	3.34	10.67	1.68	1.59	2.48	2.71	2.43	0.67	0.20	1.30	5.90	5.95	6.31	5.83	5.28	3.20
2000	0.03	-0.31	-0.32	2.92	3.15	8.37	8.31	7.03	3.85	2.17	2.34	2.15	2.16	0.50	2.45	5.17	5.59	5.33	3.68
3000	0.69	-0.22	-0.54	2.62	2.85	2.97	1.86	0.53	1.49	2.52	2.39	1.65	1.89	-6.20	0.55	1.29	5.29	5.14	4.25

Poor Good

Fig. A.8. Average SDRi using 10 different initializations of Song 7.

Average SDRi	Side lobe level used in spatial model optimization [dB]																		
Side lobe level used in source model optimization [dB]	20	30	40	50	60	70	80	90	100	120	150	200	300	500	700	1000	1500	2000	3000
20	1.22	3.23	1.08	0.02	-0.29	0.62	0.91	0.69	1.59	1.80	2.08	1.11	2.55	2.47	1.05	0.85	1.13	0.18	-1.52
30	1.59	2.66	4.39	-0.14	-0.87	-0.61	0.23	0.44	0.35	0.83	1.41	1.12	1.67	1.75	2.26	1.31	0.54	0.30	-1.24
40	1.64	1.64	3.97	0.11	0.50	0.34	0.94	1.04	0.96	1.08	0.93	0.56	1.71	1.15	1.17	1.26	1.80	0.90	-0.51
50	1.71	2.13	2.86	3.33	-0.09	0.29	-0.49	1.04	0.42	-0.18	0.50	0.68	1.67	2.13	2.23	2.29	1.55	0.34	-1.92
60	1.49	1.00	0.26	1.63	1.92	1.45	2.15	2.19	2.86	2.43	2.16	2.75	3.95	3.10	2.51	2.77	1.93	0.56	-1.41
70	1.49	2.89	0.02	2.86	2.21	2.38	1.62	2.52	2.34	1.45	1.73	2.45	2.54	1.46	3.19	2.53	2.84	1.60	-0.79
80	1.79	1.17	1.42	4.44	0.55	6.04	3.44	2.97	3.32	2.61	2.49	2.22	3.01	3.17	3.59	5.11	3.14	1.19	-0.75
90	1.61	2.13	1.16	1.77	1.06	3.23	2.39	3.48	3.46	2.94	3.57	3.36	3.26	2.63	3.29	3.87	2.59	1.49	-1.36
100	-0.34	1.19	4.27	1.25	2.56	3.26	3.85	3.18	3.13	3.07	2.82	2.81	2.24	2.83	3.35	3.92	1.33	1.04	-1.39
120	-2.71	0.67	2.31	1.10	2.23	3.78	3.35	2.90	3.03	2.95	2.80	2.72	2.18	2.55	3.10	2.89	2.05	1.20	-0.92
150	-2.14	-0.07	1.58	0.55	3.00	2.99	5.24	3.66	3.74	2.99	3.14	3.06	2.29	3.18	4.86	3.28	1.54	1.70	-1.88
200	-4.72	-0.50	1.98	-0.68	3.12	5.08	2.82	6.81	3.92	2.73	4.23	2.74	2.65	1.95	2.52	2.40	2.11	1.95	-1.91
300	0.41	2.17	1.08	-0.31	3.86	4.69	4.11	4.39	7.12	4.03	3.89	4.06	3.72	2.07	1.59	2.92	2.58	2.79	-0.41
500	-1.45	1.42	0.82	0.83	4.40	4.55	6.23	4.62	3.69	3.37	2.46	3.96	1.98	2.66	2.16	2.78	2.50	1.16	-0.76
700	0.14	0.78	1.54	0.98	4.20	3.92	6.32	4.44	2.72	2.66	2.13	7.12	3.36	3.12	2.32	2.93	2.55	2.08	0.06
1000	-0.83	2.27	2.49	3.17	2.47	3.16	3.99	2.88	2.35	1.87	3.45	2.68	5.55	3.06	1.59	1.55	2.35	0.41	-1.50
1500	-1.54	2.56	2.49	2.95	1.85	3.03	1.99	4.02	3.55	5.96	3.38	5.18	4.07	2.21	0.67	3.30	2.74	2.26	-0.58
2000	-1.35	2.47	2.06	2.41	1.63	2.56	3.23	3.53	3.39	2.58	3.66	2.74	3.24	1.94	6.08	3.15	3.06	2.25	1.33
3000	-2.83	2.65	2.79	2.89	0.66	2.06	2.81	3.34	3.15	3.11	2.99	1.39	3.33	2.63	5.73	2.91	3.37	3.18	0.08

Poor Good

Fig. A.9. Average SDRi using 10 different initializations of Song 8.

Average SDRi	Side lobe level used in spatial model optimization [dB]																		
Side lobe level used in source model optimization [dB]	20	30	40	50	60	70	80	90	100	120	150	200	300	500	700	1000	1500	2000	3000
20	1.13	1.31	-0.81	-0.02	0.69	0.21	1.16	1.45	1.29	0.81	-0.23	-0.37	-0.67	-0.84	-1.40	-1.30	-2.12	-2.63	-4.37
30	0.59	1.61	3.81	3.35	2.79	2.68	3.30	3.22	2.87	2.00	1.66	1.55	0.97	0.31	-0.26	-1.16	-1.90	-2.09	-4.26
40	1.44	0.93	1.32	8.62	5.02	4.58	4.60	4.61	4.68	4.78	4.94	4.80	5.45	5.79	5.75	5.56	5.03	3.98	0.06
50	1.45	1.54	1.93	10.03	10.07	10.10	10.37	10.25	10.21	10.25	10.00	9.87	9.99	9.81	8.85	9.90	9.42	8.60	4.98
60	1.52	1.58	2.06	1.77	12.19	10.13	7.60	7.88	8.63	9.02	9.32	10.00	9.59	9.32	10.56	10.63	8.02	7.40	6.26
70	1.54	1.57	1.78	1.05	13.64	6.33	5.81	5.54	5.09	4.91	3.86	3.40	3.77	6.51	7.46	9.54	7.03	5.82	4.59
80	1.55	1.55	1.58	-0.17	14.25	8.07	5.86	6.80	5.84	4.50	4.61	3.51	3.55	4.87	5.97	6.48	6.99	5.55	3.57
90	1.52	1.08	1.35	0.74	9.63	13.17	3.78	5.45	5.57	4.30	3.98	3.30	2.89	4.45	5.01	6.62	6.20	5.49	4.14
100	1.36	-0.85	-0.12	1.38	9.78	8.21	10.59	4.74	4.92	4.55	3.52	3.17	3.00	4.53	4.76	5.95	6.21	5.97	4.57
120	-8.33	-4.89	0.31	-0.08	13.65	15.32	12.84	11.30	4.49	3.95	2.59	2.56	2.45	4.25	4.74	5.28	6.26	5.56	3.99
150	-3.09	-0.54	-1.96	-0.36	12.88	11.96	6.78	2.90	6.37	3.74	2.44	2.24	2.53	3.29	4.21	4.20	4.51	3.90	4.27
200	-9.70	-2.11	-2.20	0.22	12.75	10.26	10.68	12.53	11.68	3.70	2.07	2.04	2.53	3.29	3.60	3.96	4.46	3.08	4.91
300	-4.28	-1.12	-0.42	5.74	12.31	11.95	12.35	11.43	11.70	9.83	8.39	3.02	2.53	3.37	3.36	3.56	4.26	2.25	5.05
500	-3.37	-0.69	-0.09	4.27	4.15	11.16	3.84	9.17	4.61	5.59	11.12	11.92	1.44	3.40	3.51	3.21	2.67	2.87	4.93
700	-5.72	-0.49	0.53	6.77	13.89	12.79	11.61	11.72	11.13	10.27	10.59	10.99	7.11	4.00	3.42	3.23	2.51	2.46	3.92
1000	-5.48	-0.38	1.08	6.96	13.10	12.43	12.41	11.74	11.64	10.31	10.49	10.39	9.78	2.83	3.31	3.39	3.26	2.64	0.30
1500	-1.69	-0.09	1.65	9.04	12.67	11.61	12.15	12.30	12.05	10.92	8.81	3.20	5.46	5.24	2.01	3.46	3.37	2.85	0.26
2000	-1.21	0.28	1.74	9.58	11.74	11.78	11.87	11.07	10.38	9.17	8.10	8.03	7.87	8.57	4.91	3.28	3.39	3.03	1.20
3000	-2.92	1.19	1.44	8.69	11.07	10.77	9.88	10.89	10.86	10.10	9.45	8.61	8.07	8.11	8.07	7.83	3.19	3.75	2.02

Poor Good

Fig. A.10. Average SDRi using 10 different initializations of Song 9.

Average SDRi	Side lobe level used in spatial model optimization [dB]																		
Side lobe level used in source model optimization [dB]	20	30	40	50	60	70	80	90	100	120	150	200	300	500	700	1000	1500	2000	3000
20	7.11	9.46	10.48	4.86	1.53	2.66	1.73	2.67	3.26	2.11	2.08	1.65	2.78	1.31	0.63	0.50	0.01	-1.13	-3.40
30	7.77	8.96	11.19	3.26	2.36	1.80	3.69	2.48	1.72	3.00	2.44	2.31	1.28	2.04	1.94	2.01	1.20	-0.17	-3.12
40	8.05	8.02	10.16	9.07	4.27	3.82	9.81	7.75	8.87	7.30	7.49	6.40	7.71	7.70	7.75	7.02	5.08	3.55	-0.58
50	7.04	6.88	8.70	3.32	9.88	12.38	11.94	12.26	12.60	12.80	12.90	13.03	12.86	11.66	11.03	10.27	8.18	6.45	1.01
60	6.71	6.28	4.63	8.32	15.30	14.73	14.82	14.64	14.45	13.79	13.64	13.24	13.48	13.10	12.98	10.90	8.63	5.30	2.18
70	6.74	6.61	4.26	6.90	11.06	15.11	14.74	14.34	13.84	13.47	13.03	12.69	12.47	11.48	11.45	10.70	7.72	5.20	1.34
80	6.25	6.10	4.30	3.68	12.06	11.70	14.47	13.92	13.61	12.91	12.78	12.34	12.03	11.43	11.23	10.01	7.51	5.39	0.35
90	3.53	6.28	4.08	0.80	9.65	12.49	13.79	14.27	13.70	13.00	12.59	12.05	11.19	10.83	10.73	9.98	8.37	5.92	1.31
100	2.94	3.92	4.26	1.32	1.01	10.31	12.93	13.64	14.05	13.63	13.23	12.47	11.16	10.55	10.68	9.70	8.38	6.45	2.53
120	0.84	-3.80	1.32	0.65	1.48	3.37	12.52	13.85	13.98	13.68	13.46	12.84	11.86	10.93	10.72	9.60	8.43	6.49	3.08
150	-9.46	-0.72	0.72	-0.41	1.17	1.62	11.96	13.36	13.95	13.99	13.61	13.04	11.76	11.05	10.74	9.85	8.40	6.86	2.59
200	-8.49	-3.71	-3.01	-2.67	6.74	1.96	2.51	9.63	11.74	13.16	13.51	12.90	11.80	11.06	10.54	9.47	8.12	7.05	2.95
300	-4.39	-1.36	-1.30	0.91	6.71	5.19	5.97	4.83	2.78	-0.85	8.16	12.40	11.91	10.49	10.28	9.87	8.17	6.76	2.89
500	-1.74	-0.05	1.18	2.14	8.19	5.71	6.48	3.66	3.01	2.63	0.12	5.53	10.61	10.52	9.95	9.38	8.05	6.80	3.56
700	-3.26	-0.76	-0.21	2.13	7.50	1.61	3.22	3.23	1.43	0.10	0.67	-0.47	4.17	10.32	10.21	9.46	8.35	6.67	3.52
1000	-2.73	-0.48	0.41	2.13	4.97	3.80	1.80	1.05	1.91	0.45	0.97	0.14	-0.24	10.52	9.85	9.22	8.35	6.60	3.30
1500	-2.53	-0.14	1.69	2.03	6.22	4.29	4.00	2.02	1.58	2.68	2.63	3.24	0.51	6.19	7.38	9.09	7.84	6.46	2.79
2000	-3.38	-0.06	1.96	1.88	5.37	5.54	5.43	5.83	5.37	5.39	4.42	4.18	0.97	0.51	2.28	4.69	8.00	6.39	2.78
3000	-3.37	0.28	0.85	2.07	6.64	6.07	5.03	5.49	7.05	6.01	3.58	4.53	2.64	1.00	-0.08	0.90	5.12	6.20	3.38

Poor Good

Fig. A.11. Average SDRi using 10 different initializations of Song 10.