Vol.2018-MUS-119 No.3
Vol.2018-SLP-122 No.3
2018/6/16

自由聴点オーディオの概略とその要素技術

大谷 健登[1,a)]

概要：計算機性能の向上や，信号処理技術の進展により，これまでの受動的な楽曲受聴ではなく，ユーザがインタラクティブに楽曲を編集しながら楽曲を受聴する，能動的音楽鑑賞という枠組みが存在する．その一つである自由聴点オーディオは，収録された音響信号中の音源信号の配置や受聴位置をユーザが自由に操作することで，楽曲の空間的印象を自由に加工することができるシステムである．自由聴点オーディオシステムは，収録された混合信号に含まれる個々の音源信号を取り出すための音源分離技術と，頭部伝達関数などを利用して分離された音源信号に仮想的な方向感を付与する立体音響技術から構成される．本チュートリアル講演では，自由聴点オーディオシステムを理解するために，システムの概略とその実現のために必要な要素技術について概説する．

An Outline of Selective Listening Point Audio System and Technologies Behind the System

Ohtani Kento[1,a)]

1. はじめに

インターネットの発達やスマートフォンなどといった携帯型デバイスの普及により，我々は好きなときにインターネット上の楽曲データをダウンロードし，音楽を楽しむことができるようになった．しかしながら，提供される楽曲データは，アーティストが楽曲を構成する楽器音信号をもとに各楽器音の音量や音色，空間的配置が設定されたステレオ信号であることが一般的である．一般的な音楽プレーヤでは，ユーザは基本的には再生若しくは停止の操作を行うだけでよく，非常に手軽に楽曲を鑑賞することができる．一方で，おおまかな楽曲の印象であれば操作することが可能ではあるが，個々の楽器音信号を個別に操作するなどといった幅広い楽曲の印象操作を行うことは難しい．例えば，多くの音楽プレーヤに搭載されている周波数イコライザ [1] では，楽曲全体の周波数帯域ごとのパワーの操作を行うことで楽曲の印象操作を実現しているが，操作が楽曲全体へと作用するため，それほど柔軟な制御を行うことができない．

このような受動的な楽曲鑑賞を超え，ユーザ自身がインタラクティブに楽曲を編集しながら行う楽曲鑑賞は能動的楽曲鑑賞と呼ばれ，これまでも様々な研究が行われてきた [2], [3], [4]．例えば，特定の楽器パートの音量バランスを個別に調整したり [5]，音色やリズムパターンなどを個別に調整したり [6] することができるようになり，ユーザは従来の受動的な楽曲鑑賞の体験を超えて，楽曲を各自の好みに近づけることでより楽曲鑑賞を楽しむことが可能になる．また，楽器音と受聴者の間の距離を操作することで，特定の楽器の信号を強調したり，ボーカルの信号を抑圧したりすることができ，楽器の練習やカラオケなどといった用途に利用し，楽曲鑑賞を超えた体験を行うことも可能となる．編集後の楽曲を他のユーザと共有することにより，ユーザ自身が新たな楽曲の創作者となることもでき，能動的楽曲鑑賞システムがより手軽な楽曲の創作を行うためのシステムであると考えることもできる．このように，能動的楽曲鑑賞により，従来の受動的な楽曲鑑賞をより豊かなものにすることが可能である．

自由聴点オーディオは音源分離技術や立体音響技術を基にした能動的楽曲鑑賞の一形態であり，収録された音響信

[1] 名古屋大学大学院 情報学研究科
Graduate School of Nagoya University, Nagoya, Aichi 464–8603, Japan
[a)] ohtani.kento@g.sp.m.is.nagoya-u.ac.jp

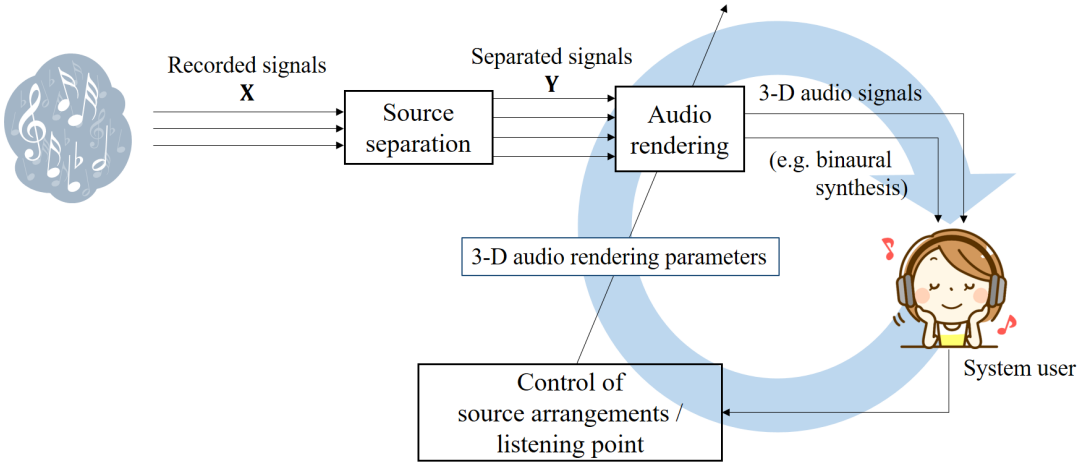

図 1 自由聴点オーディオの概要．1) 音源分離技術により混合音から音源信号を推定する，2) 聴取者は各自の望むように音源配置・受聴位置を操作する，3) 立体音響技術により聴取者の要望に合わせて立体音響信号を提示する．

Fig. 1 Overview of selective listening point audio system. 1) Estimate source signals by using source separation techniques, 2) listeners control the positions of source signals and their listening point, 3) system presents the 3-D audio signals for listener depending on their requests.

号中に存在する音源信号の配置や仮想的な受聴位置を自由に操作することのできるシステムである．これにより，例えば特定の音源信号を強調・抑圧したり，収録時とは異なる音場環境を再現しながら楽曲信号を受聴することができるようになる．本稿では，この自由聴点オーディオの概要と，その背景に存在する要素技術について概説する．

2. 自由聴点オーディオ

自由聴点オーディオ [7]，若しくは受聴位置選択型オーディオ [8] と呼ばれる技術は，収録された音響信号について，音源信号の配置や受聴位置を自由に操作・変更することを目的とした技術である．自由聴点オーディオは図 1 に示すような以下の手続きによって実現される．

(1) 複数の音源により構成された混合信号を収録・合成
(2) 音源分離技術により各音源信号を取得
(3) 聴取者による音源配置・受聴位置の操作
　(聴取者と各音源間の音響伝達関数の決定)
(4) 各分離信号と対応する音響伝達関数の畳み込み
(5) 畳み込み後の音響信号の統合及び呈示

音源分離には，独立成分分析 (Independent Component Analysis: ICA) [9] や非負値行列因子分解 (Non-negative Matrix Factorization: NMF) [10] 等が利用され，立体音響呈示に利用する音響伝達関数には頭部伝達関数 (Head Related Transfer Function: HRTF) [11] 等が利用される．

自由聴点オーディオは，マイクロホンの配置や収録環境によって 3 種類に大別することができる．

(1) 周囲配置型 [7], [8]
　周囲配置型では，聴取者の移動可能範囲である音場を取り囲むようにマイクロホンアレイが配置される．この方式は自由視点テレビ [12] のための音響部分として検討が始まったものである．
(2) 内部配置型
　内部配置型では，60ch の無指向性マイクロホンを収めた直径 8cm ほどの小型正十二面体マイクロホンアレイ [13] を利用して収録を行う．音源分離は ICA により行われ，分離処理の高速化 [14] や距離推定 [15] なども行われている．
(3) ステレオ信号対応型 [16], [17]
　新たな楽曲体験の提案のため，自由聴点オーディオ技術を 2ch ステレオ信号に適用したシステムである．これにより，市販の CD などへの応用が期待される．

2.1 HRTF を利用した自由聴点オーディオの定式化

ここでは，一例として，HRTF を利用した自由聴点オーディオの定式化について述べる．なお，本節では楽曲信号が事前に用意された N 音源からなると考え，$S_n(\tau, k)$ を時間フレーム τ，周波数チャネル k における n 番目の音源信号とする．

信号 $S_n(\tau, k)$ は空間定位関数 $G_n^{(\mathrm{L/R})}(k)$ と掛け合わされることで，立体音響信号 $Y^{(\mathrm{L/R})}(\tau, k)$ となる．ただし，L/R は左右を表しており，$G_n^{(\mathrm{L/R})}(k)$ はそれぞれ，n 番目の音源の位置 $\mathbf{p}_n$ と $\mathbf{p}_0$ に位置する受聴者の左右の耳までの空間定位関数である．空間定位関数 $G_n^{(\mathrm{L/R})}(k)$ は，例えば，HRTF を利用することで 3 次元定位に関する重要な特徴を持つバイノーラル効果関数 (Binaural Effect Function: BEF) を用いて実装される．音源信号 $S_n(\tau, k)$ と空間定位

関数 $G_n^{(\mathrm{L/R})}(k)$ を合成することで，空間的特徴を保持した音源信号である，立体音響信号 $Y^{(\mathrm{L/R})}(\tau,k)$ が合成される．

$$Y^{(\mathrm{L})}(\tau,k)=\sum_{n=1}^{N}G_n^{(\mathrm{L})}(k)S_n(\tau,k), \tag{1}$$

$$Y^{(\mathrm{R})}(\tau,k)=\sum_{n=1}^{N}G_n^{(\mathrm{R})}(k)S_n(\tau,k), \tag{2}$$

ここでは音源が受聴者と同じ高さの二次元平面上に配置される条件について検討するため，$\mathbf{p}_n(n=0,\ldots,N)$ を二次元空間中の位置として定義する．

BEF は [18] のような事前に計測された HRTF を利用して近似的にモデル化されるとする [19]．HRTF は残響の影響を含まない伝達関数であるため，$G_n^{(\mathrm{L/R})}(k)$ は受聴者の位置 $\mathbf{p}_0$ と各音源の位置 $\mathbf{p}_n$ の相対座標のみに依存する．相対座標が相対距離 r_n と二次元平面上での相対角度 θ_n によって表現されるとき，$G_n^{(\mathrm{L/R})}(k)$ は r_n と θ_n の関数として表現することができる．

$$G_n^{(\mathrm{L/R})}(k)=G^{(\mathrm{L/R})}(k,r_n,\theta_n) \tag{3}$$

多くの HRTF のデータベースでは，[18] のように，ダミーヘッドから等距離の場所に異なる角度で離散的にスピーカを配置することで HRTF の計測を行う．任意の距離及び方向に対する空間定位関数 $G^{(\mathrm{L/R})}(k,r_n,\theta_n)$ を生成するために，角度方向に線形補間を行った HRTF [19] に対して，距離減衰を掛け合わせた信号として空間定位関数をモデル化する．空間定位関数モデルは以下のように表現される．

$$G^{(\mathrm{L/R})}(k,r_n,\theta_n)=R(k,r_n,r_m)\bar{H}^{(\mathrm{L/R})}(k,r_m,\theta_n) \tag{4}$$

ただし，$R(k_n,r_n,r_m)$ 及び $\bar{H}^{(\mathrm{L/R})}(k,r_m,\theta_n)$ はそれぞれ距離減衰項と線形補間を行った HRTF を表現している．また，r_m は HRTF の計測における音源スピーカとダミーヘッドとの間の距離を表す．$R(k,r_n,r_m)$ の実装の一つとして，距離に反比例して振幅が減衰すると仮定した．

$$R(k,r_n,r_m)=\frac{r_m}{r_n}\exp(\mathrm{j}\omega_k\frac{(r_m-r_n)}{c}) \tag{5}$$

ただし，ω_k と c はそれぞれ k 番目の周波数チャネルの正規化角速度及び音速を表しており，j は虚数単位である．

3. 自由聴点オーディオを構成する要素技術

3.1 音源分離技術

N 種類の音源信号 $s_1(t),s_2(t),\ldots,s_N(t)$ が存在することを仮定する．このとき，音源信号が何らかの混合過程に従って混合され，O 本のマイクロホンによって O 種類の混合信号 $x_1(t),x_2(t),\ldots,x_O(t)$ として収録されるとする．混合過程を行列 $\mathbf{A}$ で表現できるとすると仮定すると，混合信号は以下の式に従って計算できる．

$$\mathbf{x}(t)=\mathbf{A}\mathbf{s}(t) \tag{6}$$

ただし，$\mathbf{x}(t)$ 及び $\mathbf{s}(t)$ はそれぞれ以下のように観測信号及び音源信号を並べたベクトルである．

$$\mathbf{x}(t)=[x_1(t),x_2(t),\ldots,x_O(t)]^{\mathrm{T}} \tag{7}$$

$$\mathbf{s}(t)=[s_1(t),s_2(t),\ldots,s_N(t)]^{\mathrm{T}} \tag{8}$$

$\mathbf{A}$ は混合行列と呼ばれ，音源信号 $\mathbf{s}(t)$ がどのような過程を経て混合されたかの情報が含まれている．また，O は観測チャネル数，N は音源数を表す．なお，この式ように音源信号の遅延などの影響を考える必要のない混合系を瞬時混合系と呼ぶ．このとき，音源分離問題は，何らかの関数 $\mathcal{M}$ を用いて複数の音源が混合された信号 $\mathbf{x}(t)$ から元の音源信号の推定値 $\hat{s}_n(t)$ を分離・強調する問題である．

$$\hat{s}_n(t)=\mathcal{M}(\mathbf{x}(t)) \tag{9}$$

音源分離技術は数多くの研究がなされているが，収録チャネル数が音源数よりも多い優決定条件下で高精度に音源分離を実現する手法として ICA が存在する．ICA に基づく音源分離 [9], [20] では，混合行列 $\mathbf{A}$ に関する情報を利用することなく，分離行列 $\mathbf{W}=\mathbf{A}^{-1}$ を推定することで，音源信号の推定信号 $\hat{\mathbf{s}}(t)=[\hat{s}_1(t),\hat{s}_2(t),\ldots,\hat{s}_N(t)]^{\mathrm{T}}$ を推定する．

$$\hat{\mathbf{s}}(t)=\mathbf{W}\mathbf{x}(t) \tag{10}$$

ICA では，分離の際に音源信号 $\mathbf{s}(t)$ 及び混合行列 $\mathbf{A}$ について，以下の 3 種類の仮定を置く．

(1) 音源信号は互いに統計的に独立である
(2) 音源信号は非ガウスな分布から生成されている
(3) 混合行列は逆行列を持ち，また混合過程は時不変である

ICA ではこの仮定に基づき，分離後の各信号 $\hat{s}_1(t),\hat{s}_2(t),\ldots,\hat{s}_N(t)$ が互いに独立となるように分離行列 $\mathbf{W}$ を反復更新により推定する．ICA の発展形として，畳み込み混合の問題に対処するために時間周波数領域で ICA を適用する周波数領域 ICA(Frequency-Domain ICA: FD-ICA) [21], [22] が存在する．FD-ICA では，各周波数チャネルの複素時系列信号に対し独立に ICA を適用するため，周波数チャネルごとに音源信号の対応関係が変わるパーミュテーション問題と呼ばれる課題が生じるが，この問題を解決するための手法についても様々に検討されている [23], [24]．パーミュテーション問題に対して，単一の分離行列の最適化手法で対応するため，独立ベクトル分析 (Independent Vector Analysis: IVA) と呼ばれる手法も存在する [25], [26]．IVA では，周波数ごとに独立に適用していた ICA を，周波数チャネル方向にまとめたベクトルに対して適用しパーミュテーション問題を回避する．また，IVA と NMF の低ランク性を併用する独立低

ランク行列分析 (Independent Low-Rank Matrix Analysis: ILRMA) [27] も提案され，優決定条件下であれば，高精度に音源分離が可能であることが示されている．

収録チャネル数が音源数よりも少ない，劣決定条件と呼ばれる状態で音源分離を実現するために，NMF と呼ばれる手法も存在する．NMF に基づく音源分離 [10], [28] では，混合音信号パワースペクトログラム $|\mathbf{X}(\tau,k)|^2$ を楽曲信号中に頻出するスペクトルパターンを並べた基底行列 $\mathbf{B}(\tau,k)$ と各スペクトルパターンの時間的な強度の変動を表すアクティベーション行列 $\mathbf{U}(\tau,k)$ に分解する．

$$|\mathbf{X}(\tau,k)|^2 \approx \mathbf{B}(\tau,k)\mathbf{U}(\tau,k) \tag{11}$$

NMF では，非負値制約を課しながら，距離関数 $D(|\mathbf{X}(\tau,k)|^2||\mathbf{B}(\tau,k)\mathbf{U}(\tau,k))$ を最小化するように $\mathbf{B}(\tau,k),\mathbf{U}(\tau,k)$ を更新する．NMF では，少数の非負基底ベクトルの線形結合によって混合音信号パワースペクトログラムを再現するため，混合音信号パワースペクトログラムの低ランク近似表現を獲得することができる．このとき，NMF では混合信号パワースペクトルが音源信号パワースペクトルの線形結合として表現されることを仮定している．

$$|\mathbf{X}(\tau,k)|^2 = \sum_{n=1}^{N}|S_n(\tau,k)|^2 = \sum_{n=1}^{N} b_n(\tau,k)u_n(\tau,k) \tag{12}$$

ただし，$b_n(\tau,k)$ 及び $u_n(\tau,k)$ はそれぞれ $\mathbf{B}(\tau,k)$，$\mathbf{U}(\tau,k)$ の n 番目の成分を表す．NMF に基づく音源分離では，混合音信号パワースペクトログラムの分解表現 $\mathbf{B}(\tau,k),\mathbf{U}(\tau,k)$ を獲得した後，各音源に対応する成分のみを抽出し，音源信号パワースペクトルの推定値 $|\hat{S}_n(\tau,k)|^2$ を獲得する．その後，以下の式で計算される，音源信号と推定音源信号間の平均二乗誤差を最小化するフィルタであるウィーナーフィルタを適用する．

$$\mathcal{W}_n(\tau,k) = \frac{|\hat{S}_n(\tau,k)|^2}{\sum_{n=1}^{N}|\hat{S}_n(\tau,k)|^2} \tag{13}$$

ただし，$\mathcal{W}_n(\tau,k)$ は音源 n を強調するウィーナーフィルタであり，推定目的音信号 $\hat{S}_n(\tau,k)$ は以下の式 (14) によって計算される．

$$\hat{S}_n(\tau,k) = \mathcal{W}_n(\tau,k)\mathbf{X}(\tau,k) \tag{14}$$

NMF の拡張として，事前に音源信号に対して NMF を適用し基底行列を抽出しておき，それを利用しながら NMF を実施する教師あり NMF(Supervised NMF: SNMF) [29] やパワースペクトル上での加法性の問題に対処した複素 NMF [30] やマルチチャネル NMF [31] などが存在している．

NMF では，音源パワースペクトル間の線形結合を仮定していたが，それを非線形性を加味した混合モデルへ拡張し，振幅/パワースペクトル推定の精度を高めるために，深層ニューラルネットワーク (Deep Neural Network: DNN) を適用する研究も盛んに行われている [32], [33], [34]．これらの手法では，入力として混合音振幅/パワースペクトルを利用し，出力である目的音の振幅/パワースペクトルを教師ありで学習する．その後，ウィーナーフィルタなどを利用することで音源信号の推定を行う．

他にも事前情報を利用した音源分離 (Informed Source Separation: ISS) も存在している [35], [36]．ISS では，観測された混合信号の情報のみならず，音源分離の補助となる情報を同時に利用する事で音源分離精度の向上を狙う．楽曲信号の ISS で利用される情報としては，発音している音源の情報やより直接的に音源のスペクトログラムの情報，また，楽譜情報や対象音源の鼻歌の情報などが存在している．

3.2 立体音響技術

音源信号に空間印象を付与するための立体音響技術についても様々な手法が考えられる．

例えば一般的な楽曲信号のミキシングの際にはパンニング [37] と呼ばれる手法が利用され，左右のチャネルに音圧差を与えることで音源信号に左右方向の空間印象を付与することができる．ただし，単純な音圧の制御にとどまるため，付与することのできる空間印象は非常に限られている．

パンニングのほかにも，4ch [38], 5.1ch [39], 22.2ch オーディオシステム [40] などのサラウンド手法が存在し，映画の音響などで積極的に利用されている．波面合成法 [41] やアンビソニックス [42], [43] といった収録音場を正確に再現するための技術も存在し，パンニングに比べてかなり正確な空間印象を付与することができる．しかしながら，事前の音場の収録や聴取者への呈示の際に，多数のマイクロホンやスピーカを利用した大規模なシステムが必要となることが問題となる．楽曲信号の受聴の場合には，携帯端末などで手軽に受聴する場合も多く，用途によってはこれらの手法の利用は困難となる．

HRTF を利用したバイノーラル再生では，ヘッドホンでの受聴という制限はあるものの，HRTF を音源信号に合成するだけで実現することができ，比較的正確な空間印象を提示することができる．なお，HRTF には個人性の問題が存在し，各個人に適した HRTF を利用しないと正確な空間印象が提示されないという問題が存在する [44]．しかしながら，例えば手軽に空間印象を変化させたいといった場合には，それほど正確な空間印象は必要ないと考えられるため，この問題はそれほど重要ではない．

HRTF を利用した立体音響呈示では，音源信号に対して，式 (1)，(2) のように HRTF を周波数領域で合成するか，HRTF の時間領域表現である頭部インパルス応答 (Head Related Impulse Response: HRIR) を時間領域で畳み込む

図 2　自由聴点オーディオと仮想現実空間との融合．各キャラクターが楽曲中の音源と対応している．

Fig. 2　Application of selective listening point audio system for the virtual reality. Each character corresponds to each audio source in the music.

ことによって，聴取者の左右の耳に呈示する音響信号を生成し，ヘッドホンなどを利用して聴取者の耳に信号を直接呈示する．トランスオーラル再生 TM [45] と呼ばれるスピーカと HRTF を利用した立体音響技術も存在するが，ピンポイントで左右の耳へ個別の信号を提示することは難しいため，一般には音響信号の提示にはヘッドホンが利用される．

インタラクティブな音源配置・受聴位置の操作を実現するためには，ユーザの音源配置操作要求に対しリアルタイムに音源に対応する伝達関数を変化させる必要がある．重畳加算法 [46] と呼ばれる手法を利用することで，リアルタイムに HRIR を音源信号に畳み込むことが可能となり，自由聴点オーディオが実現できる．

4. おわりに

本稿では，自由聴点オーディオの概要とその背景に存在する要素技術である，音源分離技術並びに立体音響技術について概説した．自由聴点オーディオは，1) 音源分離技術により混合信号から各音源信号を分離・強調する，2) 立体音響技術により聴取者の望む空間印象を持った音響信号を提示する，の 2 ステップで構成される．本稿では深く触れてはいないが，例えば [17] のように仮想現実 (Virtual Reality: VR) などと組み合わせることによって，楽曲をただ鑑賞するだけでなく，演奏ステージを VR 空間上に再現するなど更なる楽曲鑑賞の広がりを期待することができる (図 2)．なお，システム自体の評価方法や楽曲以外への応用などについては，今後検討していく必要がある．

参考文献

[1] Swamy, M. and Thyagarajan, K.: Digital bandpass and bandstop filters with variable center frequency and bandwidth, *Proc. of the IEEE*, Vol. 64, No. 11, pp. 1632–1634 (1976).
[2] Goto, M.: Active Music Listening Interfaces Based on Signal Processing, *Proc. of 2007 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '07)*, Vol. 4, pp. IV–1441–IV–1444 (2007).
[3] 後藤真孝：音楽音響信号理解に基づく能動的音楽鑑賞インタフェース，情報処理学会研究報告エンタテインメントコンピューティング（EC），Vol. 2007, No. 37, pp. 59–66 (2007).
[4] 糸山克寿：音楽音響信号の音源分離と能動的音楽鑑賞への応用，*TELECOM FRONTIER*, Vol. 95, pp. 1–5 (2017).
[5] Itoyama, K., Goto, M., Komatani, K., Ogata, T. and Okuno, H.: Instrument equalizer for query-by-example retrieval: improving sound source separation based on integrated harmonic and inharmonic models, *Proc. of the 9th International Society for Music Information Retrieval (ISMIR 2008)*, pp. 133–138 (2008).
[6] Yoshii, K., Goto, M., Komatani, K., Ogata, T. and Okuno, H.: Drumix: an audio player with real-time drum-part rearrangement functions for active music listening, *Trans. of Information Processing Society of Japan*, Vol. 48, No. 3, pp. 1229–1239 (2007).
[7] 福嶋慶繁，丹羽健太，圓道知博，藤井俊彰，谷本正幸，西野隆典，武田一哉：多視点・多聴点データ取得システムを用いた自由視聴点映像生成，電子情報通信学会論文誌. D, 情報・システム， Vol. 91, No. 8, pp. 2039–2041 (2008).
[8] Niwa, K., Nishino, T. and Takeda, K.: Selective Listening Point Audio Based on Blind Signal Separation and Stereophonic Technology, *IEICE Transactions on Information and Systems*, Vol. E92.D, No. 3, pp. 469–476 (2009).
[9] Comon, P.: Independent Component Analysis, a New Concept?, *Signal Process.*, Vol. 36, No. 3, pp. 287–314 (1994).
[10] Smaragdis, P.: Non-negative Matrix Factor Deconvolution; Extraction of Multiple Sound Sources from Monophonic Inputs, *Proc. of the 5th International Conference on Independent Component Analysis and Blind Signal Separation (ICA 2004)*, Lecture Notes in Computer Science, Vol. 3195, pp. 494–499 (2004).
[11] Blauert, J.: *Spatial hearing: the psychophysics of human sound localization*, MIT press (1996).
[12] Toshiaki Fujii, M. T.: Free viewpoint TV system based on ray-space representation (2002).
[13] Ogasawara, M., Nishino, T. and Takeda, K.: Blind Source Separation Using Dodecahedral Microphone Array under Reverberant Conditions, *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, Vol. E94.A, No. 3, pp. 897–906 (2011).
[14] Mizuno, Y., Kondo, K., Nishino, T., Kitaoka, N. and Takeda, K.: Effective Frame Selection for Blind Source Separation Based on Frequency Domain Independent Component Analysis, *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, Vol. E97.A, No. 3, pp. 784–791 (2014).
[15] 丹羽健太， 江崎知，日岡裕輔，西野隆典，武田一哉：空間相関行列の固有値分布に着目した音源別距離推定 (学生論文特集)，電子情報通信学会論文誌. A, 基礎・境界, Vol. 97, No. 2, pp. 68–76 (2014).
[16] 大谷健登，丹羽健太，武田一哉：事前知識を利用した楽曲音源分離技術の空間印象操作系への応用，日本音響学会 2016 年春季研究発表会，pp. 1611–1614 (2016).
[17] Niwa, K., Ohtani, K. and Takeda, K.: Music Staging AI, Demonstrations of 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2017) (2017).

[18] Takeda Lab., Nagoya Univ.: Head Related Transfer Function Database, http://www.sp.m.is.nagoya-u.ac.jp/HRTF/ [Online; accessed Dec. 19, 2017].
[19] Nishino, T., Kajita, S., Takeda, K. and Itakura, F.: Interpolating head related transfer functions in the median plane, *1999 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA 1999)*, pp. 167–170 (1999).
[20] Bell, A. J. and Sejnowski, T. J.: An information-maximization approach to blind separation and blind deconvolution, *NEURAL COMPUTATION*, Vol. 7, No. 6, pp. 1129–1159 (1995).
[21] Sawada, H., Araki, S. and Makino, S.: MLSP 2007 Data Analysis Competition: Frequency-Domain Blind Source Separation for Convolutive Mixtures of Speech/Audio Signals, *2007 IEEE Workshop on Machine Learning for Signal Processing (MLSP 2007)*, pp. 45–50 (2007).
[22] Smaragdis, P.: Blind separation of convolved mixtures in the frequency domain, *Neurocomputing*, Vol. 22, No. 1, pp. 21–34 (1998).
[23] Saruwatari, H., Kurita, S. and Takeda, K.: Blind source separation combining frequency-domain ICA and beamforming, *Proc. of 2001 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '01)*, Vol. 5, pp. 2733–2736 (2001).
[24] Sawada, H., Mukai, R., Araki, S. and Makino, S.: A robust and precise method for solving the permutation problem of frequency-domain blind source separation, *IEEE Trans. on Speech and Audio Processing*, Vol. 12, No. 5, pp. 530–538 (2004).
[25] Hiroe, A.: Solution of Permutation Problem in Frequency Domain ICA, Using Multivariate Probability Density Functions, *Proc. of the 6th International Conference on Independent Component Analysis and Blind Signal Separation (ICA 2006)*, Berlin, Heidelberg, Springer-Verlag, pp. 601–608 (2006).
[26] Kim, T., Lee, I. and Lee, T. W.: Independent Vector Analysis: Definition and Algorithms, *2006 Fortieth Asilomar Conference on Signals, Systems and Computers*, pp. 1393–1396 (2006).
[27] Kitamura, D., Ono, N., Sawada, H., Kameoka, H. and Saruwatari, H.: Determined Blind Source Separation Unifying Independent Vector Analysis and Nonnegative Matrix Factorization, *IEEE/ACM Trans. on Audio, Speech, and Language Processing*, Vol. 24, No. 9, pp. 1626–1641 (2016).
[28] O'Grady, P. D. and Pearlmutter, B. A.: Convolutive Non-Negative Matrix Factorisation with a Sparseness Constraint, *Proc. of 2006 IEEE Workshop on Machine Learning for Signal Processing (MLSP 2006)*, pp. 427–432 (2006).
[29] Kitamura, D., Saruwatari, H., Shikano, K., Kondo, K. and Takahashi, Y.: Music signal separation by supervised nonnegative matrix factorization with basis deformation, *Proc. of 18th International Conference on Digital Signal Processing (DSP 2013)*, pp. 1–6 (2013).
[30] Kameoka, H., Ono, N., Kashino, K. and Sagayama, S.: Complex NMF: A new sparse representation for acoustic signals, *Proc. of 2009 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2009)*, pp. 3437–3440 (2009).
[31] Sawada, H., Kameoka, H., Araki, S. and Ueda, N.: Multichannel Extensions of Non-Negative Matrix Factorization With Complex-Valued Data, *IEEE Trans. on Audio, Speech, and Language Processing*, Vol. 21, No. 5, pp. 971–982 (2013).
[32] Uhlich, S., Giron, F. and Mitsufuji, Y.: Deep neural network based instrument extraction from music, *Proc. of 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2015)*, pp. 2135–2139 (2015).
[33] Smaragdis, P. and Venkataramani, S.: A Neural Network Alternative to Non-Negative Audio Models, *Proc. of 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2017)*, pp. 86–90 (2017).
[34] Niwa, K., Koizumi, Y., Kawase, T., Kobayashi, K. and Hioka, Y.: Supervised source enhancement composed of nonnegative auto-encoders and complementarity subtraction, *Proc. of 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2017)*, pp. 266–270 (2017).
[35] Gorlow, S. and Marchand, S.: Informed Source Separation: Underdetermined Source Signal Recovery from an Instantaneous Stereo Mixture, *Proc. of 2011 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA 2011)*, pp. 309–312 (2011).
[36] Gorlow, S. and Marchand, S.: Informed Audio Source Separation Using Linearly Constrained Spatial Filters, *IEEE Trans. on Acoustics, Speech, and Signal Processing*, Vol. 21, No. 1, pp. 3–13 (2013).
[37] Pulkki, V.: Virtual Sound Source Positioning Using Vector Base Amplitude Panning, *Journal of the Audio Enginerring Society*, Vol. 45, No. 6, pp. 456–466 (1997).
[38] Shiu, Y.-M., Chang, T.-M. and Chang, P.-C.: Realization of surround audio by a quadraphonic headset, *2012 IEEE International Conference on Consumer Electronics (ICCE 2012)*, pp. 13–14 (2012).
[39] Chung, J. F., Liu, D.-J. and Lin, C.-T.: Multiband room effect simulator for 5.1-channel sound system, *2005 IEEE International Symposium on Circuits and Systems*, Vol. 3, pp. 2859–2862 (2005).
[40] Sugimoto, T., Oode, S., Nakayama, Y. and Okubo, H.: Subjective Evaluation of Reproduction Method for Frontal Channels of 22.2 Multichannel Sound over a Direct-View Display, *ITE Trans. on Media Technology and Applications*, Vol. 3, No. 1, pp. 67–75 (2015).
[41] Berkhout, A. J., de Vries, D. and Vogel, P.: Acoustic control by wave field synthesis, *The Journal of the Acoustical Society of America*, Vol. 93, No. 5, pp. 2764–2778 (1993).
[42] Malham, D. G. and Myatt, A.: 3-D Sound Spatialization using Ambisonic Techniques, *Computer Music Journal*, Vol. 19, No. 4, pp. 58–70 (1995).
[43] Poletti, M. A.: Three-Dimensional Surround Sound Systems Based on Spherical Harmonics, *Journal of the Audio Engineering Society*, Vol. 53, No. 11, pp. 1004–1025 (2005).
[44] Morimoto, M. and Ando, Y.: On the simulation of sound localization, *Journal of the Acoustical Society of Japan (E)*, Vol. 1, No. 3, pp. 167–174 (1980).
[45] Song, M. S., Zhang, C., Florencio, D. and Kang, H. G.: An Interactive 3-D Audio System With Loudspeakers, *IEEE Transactions on Multimedia*, Vol. 13, No. 5, pp. 844–855 (2011).
[46] Oppenheim, A. V. and Schafer, R. W.: *Discrete-Time Signal Processing*, Prentice Hall Press, Upper Saddle River, NJ, USA, 3rd edition (2009).