

# 音情景分解に基づく音メディア・ユニバーサル・コミュニケーション・システムの 研究開発 (081707005)

## Research and Development of Sound Media Universal Communication System Based on Auditory Scene Decomposition

### 研究代表者

猿渡洋 奈良先端科学技術大学院大学  
Hiroshi Saruwatari Nara Institute of Science and Technology

### 研究分担者

鹿野清宏<sup>†</sup> 川波弘道<sup>††</sup> 細井裕司<sup>†††</sup>  
Kiyohiro Shikano<sup>†</sup> Hiromichi Kawanami<sup>††</sup> Hiroshi Hoshoi<sup>†††</sup>  
<sup>†</sup>奈良先端科学技術大学院大学 <sup>††</sup>奈良先端科学技術大学院大学 <sup>†††</sup>奈良県立医科大学  
<sup>†</sup>Nara Institute of Science and Technology <sup>††</sup>Nara Institute of Science and Technology  
<sup>†††</sup>Nara Medical University

研究期間 平成 20 年度～平成 22 年度

### 概要

音情景を基本とする音メディア・コミュニケーション・システムを構築するために、まず基本アルゴリズムとして独立成分分析によるブラインド音源分離処理の高精度化・高速化およびそのポスト処理の高品質化を行った。これにより、リアルタイム音情景分解システムを構築した。次に、分解された音情景の拡張・変形・合成および音情景情報の抽出・意味化を行い、音情景拡張・合成システムを構築した。最後に、上記システムの実証的アプリケーションである音情景分解に基づく両耳補聴器システムを構築し、その評価を通じて有効性を確認することができた。

### Abstract

In order to achieve a new sound media communication system on the basis of auditory scene, we first propose an ICA-based blind source separation algorithm with high accuracy and fast convergence, and improve its post-processing. Using these algorithms, we can construct a real-time sound scene decomposition system. Next, the sound-scene augmentation system is developed, which can realize expansion, modification, and synthesis of the decomposed sound scenes and their context. Finally, as the practical test bed application, a binaural hearing aids system based on sound scene decomposition is developed and evaluated, revealing the efficacy of the system.

## 1. まえがき

本研究では、独立な音情景を単位とする柔軟かつ高品質な音メディア・コミュニケーション・システムを提案した。「音情景」とは、各音源信号波形のみを指すのではなく、それを取り巻く環境情報、例えば室内における音の反響具合やその音源の持つ空間的な広がり感などの情報をも含む概念であると定義する。具体的には、「複雑な音情景を統計的な独立成分に分解し、加工・拡張再現する」という総合的音メディアコンテンツ入出力システムの構築を主目的とした。また、このシステムの実証的アプリケーションとして、「音情景分解に基づく両耳補聴器」を想定し、提案法のユニバーサル・コミュニケーション技術としての有用性に関して検討することを最終目標とした。

## 2. 研究内容及び成果

### 2. 1. リアルタイム音情景分解アルゴリズムの開発

本研究では、音情景分解アルゴリズムに関して、単入力・多点出力 (SIMO) 分解型の独立成分分析 (ICA) を採用した。ここでは、ICA の収束速度に関して反復 100 回以下での収束および全体で遅延 100 ms 以下のリアルタイム処理を目標とし、かつポストフィルタ等の改善を通じてその分離精度向上の検討を行った。

まず、ICA の収束速度を向上させるため、反復学習の必要がない解析型二次統計量 ICA を併用した発話検出・初期化手法を提案し、30 回の反復で十分な分離精度が得られることを実証した。これを基に、リアルタイムで動作する

低遅延アルゴリズムの開発に成功した (図 1 参照)。本ソフトウェアは通常の PC 上で稼動するものであり、実際のシステムに導入した際の実測遅延は、56.5 ms であった。つまり、ほぼ人間の知覚に関して問題のないレベル以下の遅延にて瞬時に音源分離を行うことが確認された。

次に、ポストフィルタ等の更なる改善を通じてその分離精度向上の検討を行った。一般に、従来のポスト処理を行う場合、分離精度 (SN 比) は向上するものの、「ミュージカルノイズ」と呼ばれる不快な残留雑音が発生する。本研究開発において、この発生メカニズムを数理的に解析した結果、信号の確率密度の変形具合と音質との間に大きな相関があることを突き止めた。そこで、ICA のポストフィルタのミュージカルノイズ発生量を高次統計量「Kurtosis 比」に基づいて評価し、最適な内部パラメータを設定する枠組みを提案した。これにより、動的な雑音においても、少ないミュージカルノイズにて高い SN 比を達成することが可能となった。また、本ポスト処理の改良に伴う演算量の増加は多くとも 10% 以下であり、基礎となる前述リアルタイム音源抽出システムに導入しても遅延を 70 ms 以下に抑えることが可能であることもあきらかになった。

### 2. 2. 音情景・拡張・情報変換アルゴリズムの開発

より柔軟かつ有用な音情景分解アプリケーションを実現するため、前述の音源分離処理によって抽出された音情景を拡張・情報変換するアルゴリズムの開発を行った。

まず、典型的な音情景の拡張処理システムとして、複合

音の個別音量操作や音定位の変形（音の位置の再配置）を実現する音オブジェクトクラスタリング操作法を提案し、タブレット型 PC 上に GUI とともに実装しそのユーザビリティ評価を行った。その結果、十分な満足感をもってユーザビリティが実現されていることが示された。

音情景の情報変換処理の代表例は、その音情景中から自動的に音声を抽出し情報抽出・内容理解を行うことである。これは、複数の音源が混在する中における目的音声認識処理に相当する。本研究においては、2. 1 節にて述べたリアルタイム音情景抽出処理と音声認識システムとを統合し、目的音声の認識精度に関して評価を行った。実収録された駅雑音下にて発声された音声を認識したところ、従来手法と比べて単語認識率の大幅な改善が確認された。

### 2. 3. 音情景分解・拡張に基づく両耳補聴器の開発

音情景分解アルゴリズムを両耳補聴器へ応用するために、そのアルゴリズム検討および評価を行った。本研究では、ICA による雑音推定と Minimum Mean Squares Error (MMSE) 規範に基づく振幅スペクトル推定によるポスト処理とを組み合わせ、それに基づく両耳補聴器アルゴリズムを提案した。ここでは、(1) 音像定位を保証するため、両耳において同じスペクトルゲインを用いるという制約の元で、(2) MMSE 規範において左右における残留雑音誤差が最小でなければならない、ということが予備検討で判明した。そこで、上記の 2 要素を満たすため、「両耳共通化最適スペクトルゲイン (equi-binaural)」アルゴリズムを開発した。本手法の特徴は、MMSE 規範の元で高い SN 比を実現しつつ、人間の音像定位に関しても補償を行っている点にある。健聴者による評価実験を行ったところ、雑音抑圧性能は従来 ICA や各チャンネル独立に MMSE 振幅スペクトル推定法を適用するものとほぼ同程度であるにもかかわらず、音像定位に関しては大幅な改善が見られた (図 2 参照)。

本研究で提案している両耳補聴器の更なる評価のため、実際の難聴者における語音明瞭度の改善について調べた。普段から両耳補聴器を使用している重度難聴者を対象にして、普段使用している市販両耳補聴器と提案する音源分離型両耳補聴システムを併用した評価を実施した。被験者が 1 名なので統計的に有意な結果とは言えないが、約 8% の明瞭度改善効果が見受けられ、内省意見として「聞いていて疲れない」というコメントを得た。

### 3. むすび

本研究開発では、音情景を基本とする音メディア・コミュニケーション・システムを構築するために、ブラインド音源分離処理を基礎とした音情景抽出・拡張・合成システムを構築した。また、上記システムの実証的アプリケーションとして両耳補聴器システムを構築し、その評価を通じて有効性を確認することができた。

本成果は、国際的にもその独創性が認められるに至り、36 編の査読付き論文 (英文 34、和文 2 編)、61 件の口頭発表、5 件の受賞を得た。人材育成の面においても、多くの修士課程学生 (13 名) と博士課程学生 (3 名) が修了および学位取得に至った。以上より、情報通信技術を駆使したコミュニケーション能力拡張およびライフインノベーション創出に大きく貢献できたと言える。

#### 【誌上発表リスト】

[1] Y. Takahashi, T. Takatani, K. Osako, H. Saruwatari, K. Shikano, "Blind spatial subtraction array for speech enhancement in noisy environment," IEEE Trans. Audio, Speech and Language Processing, Vol.17, No.4, pp.650-664 (2009 年 5 月)

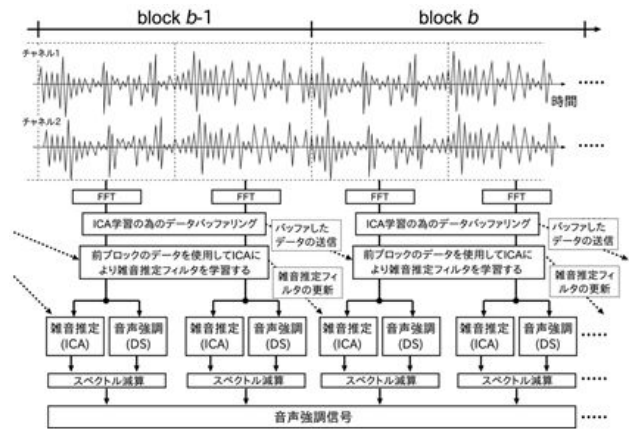


図 1. リアルタイム音情景分解アルゴリズムの実装フロー

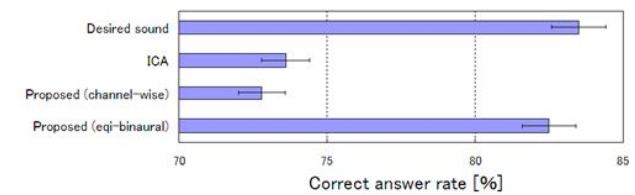


図 2. 両耳補聴器の音像定位に関する主観評価実験結果 (提示された方位の正解率) : Desired sound は原音の定位結果、equi-binaural は提案する両耳共通化最適スペクトルゲインを用いた場合、channel-wise は左右個別に MMSE 振幅スペクトル推定を行った場合、ICA は従来 ICA による結果を示す。提案する両耳共通化アルゴリズムが、ほぼ所望の音像定位感を人間に与えていることが分かる。エラーバーは 95%信頼区間を表す。

[2] H. Saruwatari, Y. Ishikawa, Y. Takahashi, T. Inoue, K. Shikano, K. Kondo, "Musical noise controllable algorithm of channelwise spectral subtraction and adaptive beamforming based on higher-order statistics," IEEE Trans. Audio, Speech and Language Processing, Vol.19, No.6, pp.1457-1466 (2011 年 8 月)

[3] T. Inoue, H. Saruwatari, Y. Takahashi, K. Shikano, K. Kondo, "Theoretical analysis of musical noise in generalized spectral subtraction based on higher-order statistics," IEEE Trans. Audio, Speech and Language Processing, Vol.19, No.6, pp.1770-1779 (2011 年 8 月)

#### 【申請特許リスト】

[1] 井上貴之、高橋祐、猿渡洋、近藤多伸、係数設定装置および雑音抑圧装置、日本、2010 年 2 月 26 日

#### 【受賞リスト】

[1] 高橋祐、IEEE SPS Japan Chapter Student Paper Award, "Blind spatial subtraction array for speech enhancement in noisy environment", 2009 年 11 月

[2] 鎌土記良、日本音響学会学生優秀発表賞, "オーディオオブジェクト操作機能を用いたリアルタイム臨場感コントローラの試作", 2010 年 9 月

[3] 高橋祐、日本音響学会独創研究奨励賞板倉記念, "アレー信号処理と非線形雑音抑圧処理を組み合わせたブラインド音声抽出技術およびその品質定量化理論に関する研究", 2011 年 3 月

#### 【本研究開発課題を掲載したホームページ】

<http://spalab.naist.jp/database/Demo/rtbssa/>