

アドホックマイクロホンアレー

— 複数のモバイル録音機器で行う音響信号処理 —

Ad-hoc Microphone Array

— Acoustic Signal Processing Using Multiple Mobile Recording Devices —

小野順貴 Nobutaka ONO
宮部滋樹 Shigeki MIYABE

Kien LE TRUNG Kien LE TRUNG
牧野昭二 Shoji MAKINO

アブストラクト マイクロホンアレー信号処理は、複数のマイクロホンで取得した多チャンネル信号を処理し、単一マイクロホンでは困難な、音源定位、音源強調、音源分離などを、音源の空間情報を用いることによって行う枠組みである。マイクロホンアレー信号処理においては、チャンネル間の微小な時間差が空間情報の大きな手がかりであり、各チャンネルを正確に同期させるために、従来は多チャンネル A-D 変換器を備えた装置が必要であった。これに対し、我々の身の回りにある、ラップトップ PC、ボイスレコーダ、スマートフォンなどの、同期していない録音機器によりマイクロホンアレー信号処理が可能になれば、その利便性は大きく、適用範囲を格段に広げることができる。本稿では、非同期録音機器を用いたマイクロホンアレー信号処理の新しい展開について、関連研究を概観しつつ、筆者らの取組みを紹介する。

キーワード マイクロホンアレー、非同期、分散、音源定位、音源分離

Abstract Microphone array signal processing is a framework for source localization, source enhancement and source separation with processing multichannel observations achieved using multiple microphones, which are difficult using a single microphone. In microphone array signal processing, the small time difference between channels is a very important cue to obtain spatial information. Therefore, a multichannel A-D converter has been conventionally essential for synchronized observation. On the other hand, if microphone array signal processing can be performed using asynchronous recording devices such as laptop PCs, voice recorders, or smart phones, which are easily available in daily life, it would enhance convenience and increase the number of possible applications markedly. In this review, focusing on a new trend of microphone array signal processing using asynchronously recording devices, we survey existing works and also introduce our approach.

Key words Microphone array, Asynchronous, Distributed, Source localization, Source separation

1. はじめに

音響信号から音源位置を推定したり、雑音の中から目的音源のみを分離、強調したりする技術は、音声認識、高品質通話、雑音源探査、異常音検出、音響信号の編集や加工、ロボット聴覚、補聴器などの幅広い分野で必要とされている技術であり、マイクロホンアレー信号処理は音響信号処理の中でも活発に研究が行われている分野の一つである。2000 年頃から独立成分分析

をはじめとするブラインド信号処理の音響信号処理応用が進展し、最近ではスマートフォン上でも実時間以下の計算時間でブラインド音源分離が可能になっている^{(1)~(3)}(図 1)。2008 年からは国際的な信号分離コンテスト SiSEC (Signal Separation Evaluation Campaign) が 1 年半置きに開催され^{(4), (5)}、2011 年からは Pascal CHiME Challenge^{(6), (7)}という、音源分離技術を活用した実雑音環境下での音声認識コンテストが開催されていることから、多くの研究者の関心を集めていることが分かる。関連する様々なプロジェクトが行われている。例えば、効率的な遠隔会議システムの構築を目的とする CHIL (Computers In the Human Interaction Loop) や AMI (Augmented Multi-party Interaction)、インタラクティブテレビの音声制御を目的とする DICIT (Distant-talking interfaces for Control of Interactive TV)、自律的な異常イベント検出を目的とする DIRAC (Detection and Identification of Rare Audio-visual Cues) などのプロジェクトでは、マイクロホンアレーによる音源定位や音源分離も重要な主要技術の一つとなっており、広範囲における発話の収録や雑音残響環境下での音声強調のために、多素子のマイクロホンアレーや、マイクロホンの分散配置なども試みられている。

小野順貴 正員 国立情報学研究所

E-mail onono@nii.ac.jp

Kien LE TRUNG 国立情報学研究所

E-mail kien@nii.ac.jp

宮部滋樹 筑波大学生命領域学際研究センター

E-mail miyabe@tara.tsukuba.ac.jp

牧野昭二 正員：フェロー 筑波大学生命領域学際研究センター

E-mail maki@tara.tsukuba.ac.jp

Nobutaka ONO, Member (National Institute of Informatics, Tokyo, 101-8430 Japan), Kien LE TRUNG, Nonmember (National Institute of Informatics, Tokyo, 101-8430 Japan), Shigeki MIYABE, Nonmember (Life Science Center of TARA, The University of Tsukuba, Tsukuba-shi, 305-8577 Japan), Shoji MAKINO, Fellow, Member (Life Science Center of TARA, The University of Tsukuba, Tsukuba-shi, 305-8577 Japan).

電子情報通信学会 基礎・境界サイエンス

Fundamentals Review Vol.7 No.4 pp.336-347 2014 年 4 月

©電子情報通信学会 2014

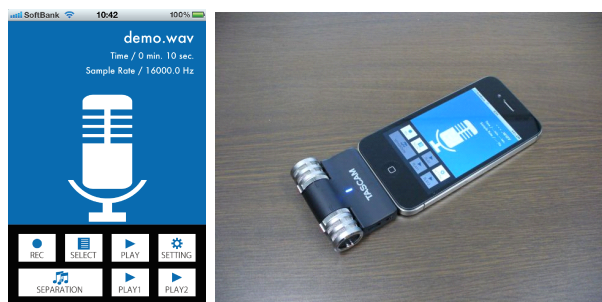


図 1 iPhone 用ブラインド音源分離アプリ (左) とステレオマイクロホン装着した iPhone (右)^{(1)~(3)}

しかしながら、こうした応用システム構築において、マイクロホンアレイシステムには厳密な同期録音が必要不可欠であることが大きな制約条件となる。これは、マイクロホンアレイ信号処理では、各マイクロホンで録音される信号間の微小な時間差（例えば、経路長 3.4 cm に対して 100 μ s）が音源の空間情報の主要な手がかりとなっているためである。半導体技術の発展によりシリコンマイクロホンなどが実用化され、小形のマイクロホンが安価に、かつ標準的に使用できるようになっているにもかかわらず、同期録音のためには結局、各マイクロホンを、厳密に同期がとれた多チャンネル A-D 変換器に接続するしかない。これがマイクロホンアレイの多素子化や分散配置などに対して大きなコストを生じる主要因の一つであり、商用化されているアレイシステムの多くは、小規模（典型的には 2 ~ 4 個）で、高々 20 cm 程度の範囲に集中配置されたものに限定されているのが現状である。

一方、我々の身の回りには、録音機能を持つ機器が多数存在している。音を録音することが目的であるボイスレコーダや、通話を目的としたスマートフォンにとどまらず、ラップトップ PC やタブレット端末の多くも録音機能を有しているし、動画撮影機能を持つデジタルカメラやビデオカメラも録音機器として用いることができる。アドホックマイクロホンアレイは、こうした別々の録音機器による多チャンネル録音を用いて、アドホックに（その場で）マイクロホンアレイを構成してアレイ信号処理を行う枠組みである。録音機器同士の通信を前提とする場合には、ワイヤレスアコースティックセンサネットワークと呼ばれる⁽⁸⁾。こうした枠組みでは、身の回りにある汎用的な録音機器を用いることができるため、特別な機器をあらかじめ設置する必要がなく、また、有線接続が不要であることはシステム構成を大いに簡便にする。また、素子数を増やしたり、広範囲に分散して配置することが容易なため、音源定位や音源分離性能の向上も期待できる。例えば、会議の参加者が各自のモバイル端末で会議を録音し、会議後に録音信号をサーバにアップロードすれば、自動的に各録音信号が同期され、各話者の音声が強調され、音声認識されて、会議の議事録が自動的にメールで送られてくるといった、自動会議議事録作成システム（図 2）は典型的なシナリオの一つである。一方、駐車場の監視システム⁽⁹⁾、交通量のモニタリングシステム⁽¹⁰⁾など、音による監視システムも重要な応用であり、危険な異常音の定位⁽¹¹⁾などにも応用可能と思われる。また、同一場所で撮影された多数の動画像の音響



図 2 アドホックマイクロホンアレイによる会議録音

トラックを用いて動画像同士のアラインメントをとる⁽¹²⁾といった研究も、広義のアドホックマイクロホンアレイ信号処理の一種といえよう。一方、アドホックマイクロホンアレイにおいては、通常マイクロホン位置は未知であり、また各録音チャンネルが同期していないといった問題を解決する必要がある。

アドホックマイクロホンアレイやアコースティックセンサネットワークの研究は、国際的にも近年関心が高まりつつある。最近では、EUSIPCO 2012 において、“Audio processing algorithms for ad-hoc microphone arrays and wireless acoustic sensor networks” という Special Session が企画され、続く EUSIPCO 2013 においても、“Introduction to distributed speech enhancement algorithms for ad hoc microphone arrays & wireless acoustic sensor networks” という Tutorial が行われた^(注1)。Elsevier Signal Processing 誌でも特集号が企画されている。国内においても、例えば日本音響学会の研究発表会では、2012 年頃からアドホックマイクロホンアレイに関する研究発表が増えてきている^{(13)~(21)}ほか、2012 年 5 月には情報処理学会音声言語処理 (SLP) 研究会のワーキンググループとして「音声・音響クラウドWG」が立ち上げられ、携帯情報端末やクラウドが利用できる環境での音声研究の新しいパラダイムが議論されている⁽²²⁾。本稿では、こうしたアドホックマイクロホンアレイの新しい展開について関連研究を概観しつつ、筆者らの取組みを紹介する。

2. アドホックマイクロホンアレイにおける諸問題

アドホックマイクロホンアレイの研究では、想定している条件や応用によって様々な問題が議論されている。まず、幾つかの主要な問題とアプローチを概観する。

(注1) : Sharon Gannot と Alexander Bertrand によるチュートリアル資料が <http://www.eusipco2013.org/tutorials.html> から参照できる。

2.1 チャンネル同期

通常のマイクロホンアレーでは、全てのマイクロホンは同期した A-D 変換器に接続され、同時サンプリングされる。これは、後述するように、マイクロホンアレー信号処理では、チャンネル間の時間差が音源の重要な空間情報となるためである。一方、録音に別々の機器を用いた場合には、各機器の A-D 変換器でデジタル信号に変換され、非同期の多チャンネル信号が得られることになる。非同期の主な要因には、① 録音を開始する時刻が同一でないこと、② サンプリング周波数が同一でないこと、の二つがある。前者は定常的な時間軸シフトを、後者は時間軸の伸縮をもたらす、信号処理には異なる影響を及ぼす。チャンネル同期の方法には、まず、録音機器同士が無線やインターネットなど、別のチャンネルで互いに通信できることを想定して同期をとる手法が考えられる^{(23)~(25)}。一方、音響信号のみを用いて信号同士を同期させることができれば、通信機能を持たない録音機器を使用することができるため、システム構成的な自由度は大きく高まる。このアプローチについての詳細は後述する。

2.2 マイクロホンの位置推定

従来のアレー信号処理に用いられるマイクロホンは、直線、円周のような規則的な配置に並べられ、各マイクロホンの位置は既知であることが前提である。しかしながらアドホックマイクロホンアレーでは、個々の録音機器は有線接続すらされておらず、通常、その位置座標は未知であることが多い。そのため、音源位置を推定したり、特定位置にある音源を強調したりするためには、まずマイクロホン位置を推定することが必要となる。スマートフォンなどを用いる場合には GPS (Global Positioning System) 情報が利用できる場合もあるが、室内環境での測位精度は一般に十分でなく、マイクロホンが観測された音自体を用いて推定する自己校正 (self-calibration) が重要となる。後述するが、幾つかの音源からの到来時間差を手がかりにする場合にはチャンネル同期も必要となり、これらは密接に関連する。

なお、位置の推定は一般にアレー信号処理への応用に限らず重要であり、Indoor Positioning and Indoor Navigation (IPIN)⁽²⁶⁾ という、室内での機器位置推定に関する国際会議が 2010 年以降毎年開催され、音に限らず、電波や光を含めた定位手法が議論されていることも付記しておく。

2.3 アドホックマイクロホンアレーに適した信号処理

チャンネル同期とマイクロホン位置の自己推定は、いわば、アドホックマイクロホンアレーシステムのキャリブレーションである。ひとたび各録音信号が同期され、マイクロホン位置が推定されれば、後は従来のマイクロホンアレー信号処理を適用することができる。しかしながら、必要とされるキャリブレーション、並びにその精度は、適用する後段の信号処理に依存する。逆

に言えば、キャリブレーションの誤差に頑健な信号処理法を検討することも、アドホックマイクロホンアレーにおける問題の一つと言える。

通常のマイクロホンアレー信号処理では、チャンネル間の時間差情報が重要であり、一般には 1ms の同期誤差も許されない場合も多い。しかし、信号の位相情報を用いず、エネルギーのみに基づく自己位置推定法^{(27),(28)}や、時間周波数領域での振幅のみに基づく音源強調^{(17),(29)}などは同期誤差にはかなり頑健であり、得られる振幅スペクトログラムがフレームレベルで同期していればよい。

また、信号処理に必要な同期の性質を考えるための例として、図 3 のように、マイク 1、マイク 2 に対し、音源 A がマイク 1 側にあり、音源 B が正面にある場合を考える。もしマイク 1、マイク 2 が厳密に同期していれば、音源 A の音はマイク 1 に早く到達して時間差が生じ、音源 B の音はほぼ同時にマイク 1、2 に到達するため、観測信号は図 3 (上) のようになるはずである。逆に、音源定位やビームフォーミングなど、マイクロホンの位置情報を用いる信号処理を行うためには、得られた非同期信号を、図 3 (上) のように、マイク位置に整合するように同期をとる必要がある。一方、SN 比最大化ビームフォーマ⁽³⁰⁾や独立成分分析⁽³¹⁾といった音源強調、音源分離手法は、マイクロホンの位置情報を必要としない^(注2)。このような信号処理を行う場合には、マイク位置の推定が不要であるばかりでなく、同期の問題も低減される。すなわち、図 3 (下) のように 2 信号の間に小さな一定の同期誤差が残っていて、観測されるチャンネル間の時間差がマイク位置や音源位置と整合性がとれていなくても、音源からマイクまでの伝達関数は、学習区間の信号から、もしくはブラインドに推定されるため、余り問題にはならない。このような場合には、2 信号の相互相関が最大となるように信号をシフトするだけで十分である。ただし、サンプリング周波数ミスマッチがある場合には、これを補償する必要がある。

2.4 マイクロホン配置

従来のマイクロホンアレーは基本的に集中配置に限られていたが、アドホックマイクロホンアレーでは、別々の録音機器を使用することにより、マイクロホンをかなり柔軟に配置することができる。音源の近く、例えば話者の近くに録音機器を配置できる場合には、それだけで SN 比を向上させることができるし、そうでなくても録音機器を分散させて配置することにより、空間的に広い範囲をカバーできる利点がある。適用する信号処理とともに、最適なマイクロホンの配置を検討することも重要となる。

ステレオ録音機器など、各録音機器が 2 個以上のマイクロホンを有する場合には、通常、それらの位置関係は決まっており、

(注2)：厳密には、例えば周波数領域での独立成分分析には、周波数帯域間のパーミューテーションをそろえる必要があり、音源方向情報を用いてパーミューテーションを解くような場合には、マイク位置情報が必要となる場合もある。しかしながら、近年ではパーミューテーションを解く必要がないブラインド音源分離手法や、位置情報を必要とせず、音源アクティビティの同期性を用いたパーミューテーション解法なども提案されている。

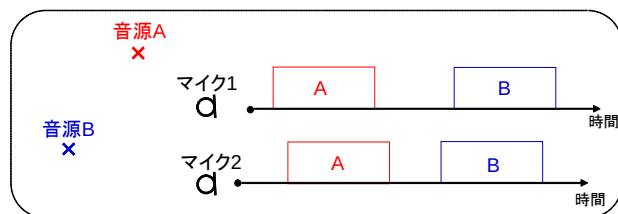
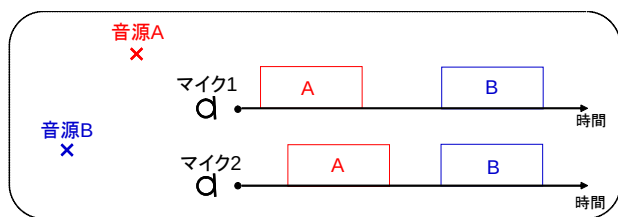


図3 マイクロホン位置に整合した同期 (上), マイクロホン位置に整合していない同期 (下)

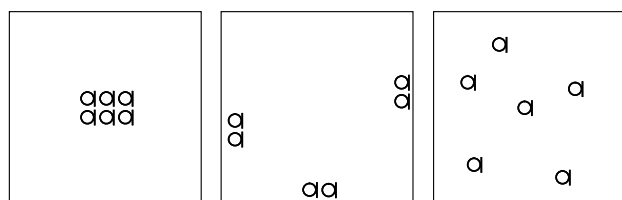


図4 マイクロホン配置の例：集中配置 (左), サブアレー分散配置 (中), 分散配置 (右)

チャンネルも同期しているため、それ自体を小規模な従来型のマイクロホンアレーとみなすことができる (図4)。こうした小規模なサブアレーを分散配置し、まず個々のサブアレー内で処理 (例えば DOA (Direction Of Arriva) 推定や音源強調) を行い、次にサブアレー間で処理結果を共有し、結果を統合したり最も良いものを選択するような階層的な処理方式は、サブアレー間でやりとりする情報量も少なく効率的であり、マイクロホンの自己位置推定^{(32), (33)}やブラインド音源分離^{(34)~(37)}などで、積極的に検討されている。サブアレー内ではマイクロホンは近接して配置されることが多く、近接マイクロホンを効果的に用いる音源定位手法^{(38), (39)}や多チャンネル信号の効率的な圧縮^{(40), (41)}なども関連する話題である。

2.5 分散処理

アドホックマイクロホンアレーは、非同期分散マイクロホンアレーなどとも言われるが、日本語の「分散」、若しくは英語の “distributed” は、空間的位置が広範囲に広がっているだけでなく、処理自体を複数の素子 (ノード) で分けて行う、という意味で用いられることも多い。実際、録音機器間で通信することを考えると、録音信号を全ての機器で共有するためには大きな通信帯域が必要なだけでなく、遠くの録音機器と通信するために大きな送信電力も必要となってしまうため、限られた通信帯域で部分的な情報だけがやりとりできる場合に、分散処理によりどうすれば良い解が得られるか、最適解に近づけられるか、というのは興味深い問題である。本稿では触れないが、ア

ドホックマイクロホンアレー、並びにワイヤレスアコースティックセンサネットワークにおける重要な問題なので、興味のある読者は文献(42)~(46)を参照して頂きたい。

3. マイクロホン位置推定とチャンネル同期

3.1 問題設定

アドホックマイクロホンアレーの典型的な問題の一つとして、マイクロホン位置推定とチャンネル同期の問題を考えてみよう。

いま、 K 個の音源を M 個のマイクロホンで観測するものし、音源、マイクロホンの位置座標をそれぞれ $\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_K, \mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_M$ と表す。三次元空間中では位置座標は三次元ベクトルで表されるが、音響の問題では、音源、マイクロホンは共に、ほぼ同一平面内に分布している場合も多く、その場合には位置座標は二次元ベクトルで表してもよい。よってここでは、 $\mathbf{s}_k, \mathbf{r}_m$ は p 次元のベクトルとし、 $p = 2$ または $p = 3$ とする。また、各音源、各マイクロホンはそれぞれ別の時間原点を持っており、それらを u_k, t_m で表すものとする。簡単のためサンプリングのことは考えない。サンプリング周波数ミスマッチの問題については後述する。このとき、何を未知数として考えるか、何を観測量として考えるかによって、様々な問題のバリエーションが生じる。

未知数

未知数から考えると以下の4種類の問題があり得る。

- 音源定位: $\mathbf{s}_k$ ($k = 1, \dots, K$) の推定
- マイクロホン定位: $\mathbf{r}_m$ ($m = 1, \dots, M$) の推定
- 音源同期: u_k ($k = 1, \dots, K$) の推定
- チャンネル同期: t_m ($m = 1, \dots, M$) の推定

一般には音響信号の観測からは相対的な位置関係のみが推定可能であり、絶対的な位置や向きを決めることはできない。すなわち、音源定位やマイクロホン定位においては、位置原点の選び方の自由度 p 、回転の自由度 $p(p-1)/2$ ^(注3)、あわせて $p(p+1)/2$ の自由度は決められない。同様に時間の絶対的な原点は決められないので、チャンネル同期や音源同期においては、1自由度は決めることができないことに注意する必要がある。

観測量 1: TOA

音源からマイクロホンまで音が伝搬する時間を到来時間 (Time of Arrival; TOA) と呼ぶ。音源 k からマイクロホン m までの到来時間 τ_{km} は、音速を c として、

$$\tau_{km} = \frac{\|\mathbf{s}_k - \mathbf{r}_m\|_2}{c} \quad (1)$$

(注3)：二次元での回転は1種類しかないため、回転の自由度が1であることは分かりやすい。三次元では、地球儀を想像し、まず北極の向きを動かす自由度が緯度、経度の2自由度、更に北極回りの回転の1自由度と考えると、回転の自由度が3であることが分かる。

のように表される。ただし、 $\|\cdot\|_2$ は L_2 ノルムを表す。すなわち、到来時間に音速を乗じると音源-マイクロホン間の距離が得られる。

ある時間区間に k 番目の音源のみが観測でき、またその音源信号 $f_k(t)$ が得られたとする。このとき、マイクロホン m で観測された信号 $g_m(t)$ は、

$$g_m(t) = \alpha_{km} f_k(t - \tau_{km}) + n_m(t) \quad (2)$$

のように表される。ここで、 α_{km} は距離減衰項、 $n_m(t)$ は雑音を表す。よって、 $f_k(t)$ と $g_m(t)$ の相互相関関数は、 $n_m(t)$ が $f_k(t)$ と無相関とすれば、

$$\begin{aligned} \phi_{f_k, g_m}(\tau) &= E[f_k(t)g_m(t + \tau)] \\ &= \alpha_{km} E[f_k(t)f_m(t - \tau_{km} + \tau)] + E[f_k(t)n_m(t)] \\ &= \alpha_{km} \phi_{f_k, f_k}(\tau - \tau_{km}) \end{aligned} \quad (3)$$

と表される。 $\phi_{f_k, f_k}(\tau)$ は $f_k(t)$ の自己相関関数であり、 $f_k(t)$ が広帯域信号の場合には、 $\tau = 0$ に鋭いピークを持つ。よって、 $\phi_{f_k, g_m}(\tau)$ のピーク位置として、 τ_{km} を推定することができる。実際には、ピークの先鋭化や雑音の影響を軽減するために、一般化相互相関⁽⁴⁷⁾が用いられることが多い。

$f_k(t)$ 、 $g_m(t)$ の時間原点が異なる場合には、相互相関関数のピーク位置もこの影響を受ける。音源信号と観測信号の相互相関から求まる見かけの到来時間を $\hat{\tau}_{km}$ で表すと、

$$\hat{\tau}_{km} = \frac{\|\mathbf{s}_k - \mathbf{r}_m\|_2}{c} + u_k - t_m \quad (4)$$

と表される。

観測量 2: TDOA

TOA を求めるには音源信号 $f_m(t)$ が必要であるが、これは通常、未知であることも多い。例えば、音声を対象とした話者位置の推定においても、異常音検出においても、どんな信号波形が到来するか、またそれがいつ発せられたかは通常分からない。このような場合には、音源から二つのマイクロホンまでの到来時間の差として定義される到来時間差 (Time Difference of Arrival; TDOA) が用いられる。音源 k からマイクロホン m 、 n までの到来時間差を τ_{kmn} とすると、

$$\tau_{kmn} = \frac{\|\mathbf{s}_k - \mathbf{r}_m\|_2}{c} - \frac{\|\mathbf{s}_k - \mathbf{r}_n\|_2}{c} \quad (5)$$

と表される。すなわち、到来時間差に音速を乗じると、ある音源から二つのマイクロホンまでの距離差が得られる。

式 (2) と同様に、

$$g_m(t) = \alpha_{km} f_k(t - \tau_{km}) + n_m(t) \quad (6)$$

$$g_n(t) = \alpha_{kn} f_k(t - \tau_{kn}) + n_n(t) \quad (7)$$

より、観測信号 $g_n(t)$ 、 $g_m(t)$ の相互相関関数は (m と n の順番が逆なことに注意)、

$$\begin{aligned} \phi_{g_n, g_m}(\tau) &= E[g_n(t)g_m(t + \tau)] \\ &= \alpha_{kn}\alpha_{km} E[f_k(t - \tau_{kn})f_k(t - \tau_{km} + \tau)] \end{aligned}$$

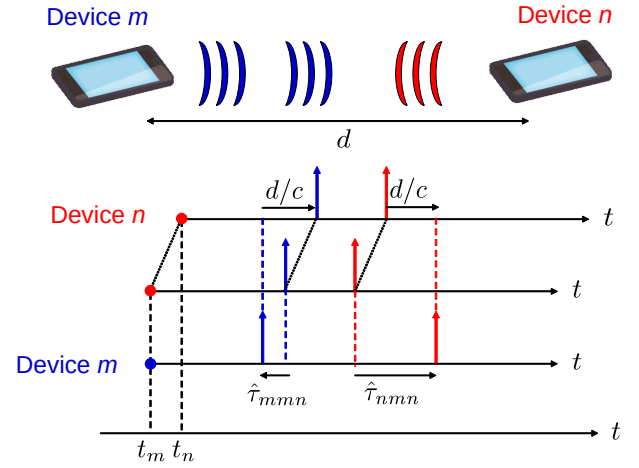


図5 録音機器自体が音を発信する場合の TDOA

$$\begin{aligned} &= \alpha_{kn}\alpha_{km} \phi_{f_k, f_k}(\tau + \tau_{kn} - \tau_{km}) \\ &= \alpha_{kn}\alpha_{km} \phi_{f_k, f_k}(\tau - \tau_{kmn}) \end{aligned} \quad (8)$$

より、 $\phi_{g_n, g_m}(\tau)$ のピーク位置より、 τ_{kmn} が推定できる。これはマイクロホンの観測信号のみから推定可能であり、音源信号を必要としないため、未知の音源を定位したり、未知の音源を手がかりにマイクロホン位置を推定したりする場合には、一般に TDOA が用いられる。

$g_m(t)$ 、 $g_n(t)$ の時間原点が異なる場合には、相互相関関数のピーク位置もこの影響を受ける。 $g_n(t)$ と $g_m(t)$ の相互相関から求まる見かけの到来時間差を $\hat{\tau}_{kmn}$ で表すと、

$$\hat{\tau}_{kmn} = \left(\frac{\|\mathbf{s}_k - \mathbf{r}_m\|_2}{c} - t_m \right) - \left(\frac{\|\mathbf{s}_k - \mathbf{r}_n\|_2}{c} - t_n \right) \quad (9)$$

と表される。

よって、幾つかの問題は、推定された TOA、若しくは TDOA を観測量とし、式 (4) 若しくは式 (9) に基づき未知数を求める問題として、統一的に捉えることができる。例えば、マイクロホン位置 $\mathbf{r}_m$ が既知で、チャンネルが全て同期している ($t_1 = t_2 = \dots = t_M$) 条件で、TDOA から音源位置 $\mathbf{s}_k$ を推定する問題は、通常の到来時間差に基づく音源定位問題^{(48)~(50)}である。マイクロホン位置も未知数とすると、音源とマイクロホン位置の同時推定問題となる⁽⁵¹⁾。

3.2 音の発信を利用した機器位置推定

スマートフォンなど、音を発することができる録音機器を利用できる場合には、音の発信を積極的に利用するアプローチが考えられる。例えば、1 台のスマートフォンに装備されているスピーカとマイクの位置は厳密には異なっているが、これを近似的に等しい ($\mathbf{r}_m \simeq \mathbf{s}_m$) とし、2 台のスマートフォンから互いに音を発信して TDOA を求めたとすると、式 (9) から、

$$\hat{\tau}_{mmn} = -\frac{\|\mathbf{r}_m - \mathbf{r}_n\|_2}{c} - t_m + t_n \quad (10)$$

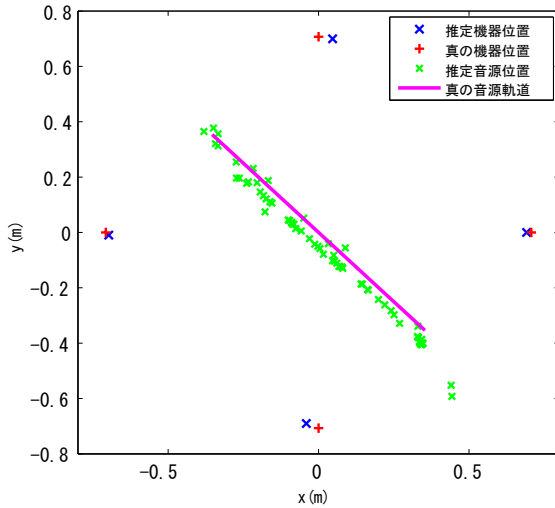
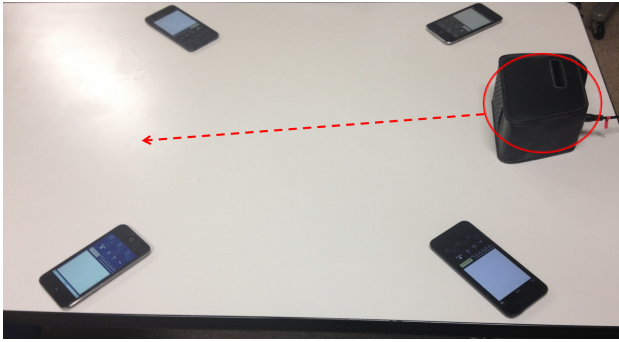


図6 アドホックマイクロホンアレーによる移動音源定位実験の写真 (上) と実験結果 (下)⁽⁵³⁾

$$\hat{\tau}_{nmn} = \frac{\|\mathbf{r}_n - \mathbf{r}_m\|_2}{c} - t_m + t_n \quad (11)$$

であるから (図 5),

$$\|\mathbf{r}_m - \mathbf{r}_n\|_2 = \frac{c}{2}(\hat{\tau}_{nmn} - \hat{\tau}_{mmn}) \quad (12)$$

$$t_m - t_n = -\frac{1}{2}(\hat{\tau}_{mmn} + \hat{\tau}_{nmn}) \quad (13)$$

のように、互いの TOA から距離と時間原点の差が直接的に求まる^{(18), (32), (52)}。各機器間の距離が求まれば、後は多次元尺度法 (MDS: Multi Dimensional Scaling) の適用により相対位置を決めることができる。

我々は更に、音の発信を利用し、位置と時間原点のキャリブレーションだけでなく、サンプリング周波数ミスマッチも併せて補償する手法を提案している⁽⁵³⁾。図 6 に、4 台の iPod touch による移動音源定位の実験結果例を示す。それぞれから発信された TSP (Time-Stretched Pulse) 信号⁽⁵⁴⁾を用いて、各機器の位置、録音開始時刻の推定、サンプリング周波数ミスマッチを推定してキャリブレーションを行った後にスピーカを音源定位したものであり、移動音源であるスピーカの定位がうまく行われていることが分かる。

3.3 ブラインドアライメント

逆に最も難しい場合として、録音機器が同期しておらず音源

も未知という条件下で、各音源に対して観測された TDOA のみから音源位置、マイクロホン位置、各マイクロホンの時間原点を推定することができるかを考えてみよう。我々はこれを、観測信号のみから (ブラインド)、音源、マイクロホン、チャンネルを共通の時空間座標系にそろえる (アライメント) という意味で、ブラインドアライメントと呼んでいる^{(55), (56)}。

式 (5) のように、TDOA は音源から二つのマイクロホンまでの距離差を表している。つまり、 $\hat{\tau}_{kmn}$ が一つ観測されると、音源 k は二つのマイクロホン m, n からの距離差が $c\tau_{kmn}$ である双曲線 (三次元の場合は双曲面) 上にあることを意味する。しかし、マイクロホン位置も未知、マイクロホンの時間原点も未知である場合には、TDOA が一つ推定されただけでは、それが到来時間差によるものなのかマイクロホンの時間原点差によるものなのか分からない。では、全ての音源に対して全てのマイクロホン対で TDOA が得られた場合には、推定は可能になるだろうか？ここではまず、得られる方程式と未知数の関係を調べてみる。

式 (9) においては、 τ_{kmn} が観測量であり、一つの音源につき、 $M-1$ 個の独立な観測量が得られる。一方、未知数は、位置ベクトル $\mathbf{r}_m, \mathbf{s}_k$ と時間原点 t_m であり、あわせて $p(K+M)+M$ 個となる (p は次元数で 2 または 3)。ただし前述のように、位置座標も時間原点も相対的にしか決まらず、 $p(p+1)/2+1$ の自由度は決めることができないので、観測量からこれらの未知数が決定されるためには少なくとも

$$K(M-1) \geq p(K+M)+M - \left(\frac{p(p+1)}{2} + 1 \right) \quad (14)$$

である必要がある。整理すると、

$$(K-p-1)(M-p-1) \geq \frac{p(p+3)}{2} \quad (15)$$

を得る。つまり、時間原点が合っていないチャンネルから得られた TDOA であっても、マイクロホンや音源がたくさんある場合には、観測量の数が未知数を上回ることを意味する。具体的には、例えば、二次元問題 ($p=2$) の場合は、 $M=4$ であれば $K \geq 8$ 、 $M=5$ であれば $K \geq 6$ のような条件が得られる。ただしこれは解が求まるための十分条件であり、必要条件ではないことに注意する必要がある。また、この不等式は多数の音源が必要なことを意味し、非現実的にも思われるかもしれないが、ある音源が移動しながら音を発すれば、短時間区間ごとにみると、別々の位置にある単一音源として扱えるので、キャリブレーションを行うための条件としてはそれほどハードではない。また、例えばサブアレー配置でサブアレー形状が既知の場合には、位置の自由度はより少なくなるし、ステレオ録音機器を用いた場合には、そのステレオ内では同期がとれていて、 $t_1 = t_2, t_3 = t_4, \dots$ が仮定できるなど、想定する条件によっては、未知数が少なくなる場合もあることに注意が必要である。

推定法としては、式 (9) に最小二乗法を適用し、

$$J = \sum_{k,m,n} Q_{kmn}^2 \quad (16)$$

$$Q_{kmn} = (\|\mathbf{s}_k - \mathbf{r}_m\|_2 - ct_m) - (\|\mathbf{s}_k - \mathbf{r}_n\|_2 - ct_n) - c\hat{\tau}_{kmn} \quad (17)$$

を最小化するのが直接的である。この問題は L_2 ノルムが平方根を含んでいるために非線形であり、最小化には勾配法などを適用する必要があるが、筆者らは補助関数法⁽⁵⁷⁾という最適化のアプローチを用いた更新式⁽⁵⁵⁾,⁽⁵⁶⁾を導出している。また、近年ロボットへの応用を想定したオンラインでのキャリブレーション⁽⁵⁸⁾も検討されている。

4. TOA に基づく音源とマイクロホンの同時 定位への双線形アプローチ

TOA は音源とマイクロホンの間の距離情報を与えるため、TDOA より直接的な情報を含んでいるが、TOA を用いたとしても、得られた距離情報から位置座標を求めるのは簡単ではない。式 (1) において、距離は位置座標の平方根を含んで非線形なためである。音源定位の問題では、式 (1) に直接最小二乗法を適用する代わりに、これを变形して、より簡単な形に整理してから最小二乗法を適用することが試みられてきた⁽⁴⁸⁾,⁽⁴⁹⁾。音源とマイクロホンの同時定位問題においては更に、音源同士、マイクロホン同士の距離は直接的には得られない難しさがある(全ての音源、マイクロホン間の距離が得られるのであれば、多次元尺度法 (MDS) が適用できる)。これに対して近年、Crocco らによって興味深いアプローチ⁽⁵⁹⁾が提案されているので、簡単に紹介したい。

ここでは、音源とマイクロホンは同期していて、各音源から各マイクロホンまでの TOA 情報が得られているものとする。TOA 情報は音速 c を乗じれば距離情報と等価であるから、これは、距離行列 $D = (d_{mn})_{M \times N}$ から、 M 個のマイクロホン位置 $\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_M$ と、 N 個の音源位置 $\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_N$ を求める問題となる。また、以下では $\mathbf{r}_m$ と $\mathbf{s}_n$ は三次元ベクトルとするが、二次元においても同様の議論が成り立つ。

距離そのものは平方根演算を含んでいるため、まずその二乗を考えると、

$$d_{mn}^2 = \|\mathbf{r}_m - \mathbf{s}_n\|_2^2 = (\mathbf{r}_m - \mathbf{s}_n)^T (\mathbf{r}_m - \mathbf{s}_n) \quad (18)$$

と書ける。ここで、 T は転置を表す。これは $\mathbf{r}_m, \mathbf{s}_n$ の二乗項を含んでいるが、差をとってこれを消去することを考えると、 $m = 2, \dots, M, n = 2, \dots, N$ 、に対して、

$$d_{11}^2 = (\mathbf{r}_1 - \mathbf{s}_1)^T (\mathbf{r}_1 - \mathbf{s}_1) \quad (19)$$

$$d_{1n}^2 - d_{11}^2 = \mathbf{s}_n^T \mathbf{s}_n - \mathbf{s}_1^T \mathbf{s}_1 - 2\mathbf{r}_1^T (\mathbf{s}_n - \mathbf{s}_1) \quad (20)$$

$$d_{m1}^2 - d_{11}^2 = \mathbf{r}_m^T \mathbf{r}_m - \mathbf{r}_1^T \mathbf{r}_1 - 2(\mathbf{r}_m - \mathbf{r}_1)^T \mathbf{s}_1 \quad (21)$$

のように、 $\mathbf{r}_m, \mathbf{s}_n$ それぞれの二乗項を消去した式を得る。

更に差の差を考えると、

$$d_{mn}^2 - d_{m1}^2 - d_{1n}^2 + d_{11}^2 = -2(\mathbf{r}_m - \mathbf{r}_1)^T (\mathbf{s}_n - \mathbf{s}_1) \quad (22)$$

のように、 $\mathbf{r}_m, \mathbf{s}_n$ 両方の二乗項を消去した式を得る。この式は、 $\mathbf{r}_m, \mathbf{s}_n$ の積で二次式ではあるが、 $\mathbf{r}_m, \mathbf{s}_n$ 、それぞれの変数に対しては線形であり、双線形 (bilinear) と呼ばれる。

よって、 MN 個の式 (18) と等価な表現として、

- 式 (19) (1 個)
- 式 (20) ($N - 1$ 個)
- 式 (21) ($M - 1$ 個)
- 式 (22) ($(M - 1)(N - 1)$ 個)

の式を得る。

特に式 (22) で表された双線形表現が重要であり、新たに行列

$$\mathbf{R}^* = [\mathbf{r}_2 - \mathbf{r}_1, \dots, \mathbf{r}_M - \mathbf{r}_1] \quad (23)$$

$$\mathbf{S}^* = [\mathbf{s}_2 - \mathbf{s}_1, \dots, \mathbf{s}_N - \mathbf{s}_1] \quad (24)$$

並びに、

$$\tilde{d}_{mn} = d_{m+1,n+1}^2 - d_{m+1,1}^2 - d_{1,n+1}^2 + d_{1,1}^2 \quad (25)$$

を要素とする行列

$$\tilde{D} = (\tilde{d}_{mn})_{(M-1) \times (N-1)} \quad (26)$$

を定義すると、式 (22) は

$$\tilde{D} = -2\mathbf{R}^{*T} \mathbf{S}^* \quad (27)$$

と表される。 $\mathbf{R}^*$ は $3 \times (M - 1)$ 、 $\mathbf{S}^*$ は $3 \times (N - 1)$ の行列であり、 $\tilde{D}$ はそれらの積であるから、 $(M - 1) \times (N - 1)$ の行列ではあるが、ランクは 3 以下でなければならないことが分かる。これは $\tilde{D}$ が満たすべき重要な拘束条件となる。

$\tilde{D}$ を特異値分解し、三つ目の特異値で打ち切ることにより、 $\tilde{D} = \mathbf{R}^T \mathbf{V} \mathbf{S}$ のような分解を得ることができる ($\mathbf{V}$ は 3×3 行列) ので、求めたい $\mathbf{R}^*, \mathbf{S}^*$ と特異値分解により得られる $\mathbf{R}, \mathbf{S}$ の関係は

$$\mathbf{R}^* = \mathbf{Q}^{-T} \mathbf{R} \quad (28)$$

$$\mathbf{S}^* = -\frac{1}{2} \mathbf{Q} \mathbf{V} \mathbf{S} \quad (29)$$

のように表すことができる。ここで $\mathbf{Q}$ は、正則な (逆行列をもつ) 3×3 の行列である。よって未知数は、この行列 $\mathbf{Q}$ と $\mathbf{s}_1, \mathbf{r}_1$ に絞られ、まだ使っていない方程式である式 (19)、式 (20)、式 (21) に最小二乗法などを適用し、任意性のある並進と回転の自由度を適当に定めた上で、これらの未知数を決める問題となる。

このアプローチは、近年、TDOA ベースのキャリブレーションなどにも応用されている⁽⁶⁰⁾,⁽⁶¹⁾。

5. サンプリング周波数ミスマッチの補償

5.1 サンプリング周波数ミスマッチの影響

複数の録音機器を用いて音響信号を録音する場合、録音機器の公称サンプリング周波数が同じ場合でも、実際のサンプリング周波数は一般に、機器ごとに僅かに異なっている。

いま、録音機器 1、録音機器 2 のサンプリング周波数を f_1, f_2 で表すと、

$$\varepsilon = \frac{f_2}{f_1} - 1 \quad (30)$$

が、録音機器間の相対的なサンプリング周波数ミスマッチを表す無次元量となる。 ε は、10 ppm (ppm は parts per million で 10^{-6} を表す) の数倍程度に収まることが多い。しかし、こうした僅かなミスマッチであっても、アレー信号処理には深刻な悪影響を及ぼす。

これらの二つの録音機器の観測信号に到来する音波 (連続信号) を $x_1(t)$, $x_2(t)$ として、これらをサンプリングすることを考える。録音機器 1 の録音開始時刻を連続時刻の原点 ($t = 0$) とし、録音機器 2 の録音開始時刻を連続時刻 T_{21} とすると、録音機器 1, 録音機器 2 における離散時刻 n_1 , n_2 は連続時刻 t により

$$n_1 = t f_1 \quad (31)$$

$$n_2 = (t - T_{21}) f_2 = (t - T_{21})(1 + \varepsilon) f_1 \quad (32)$$

と表され、得られる離散信号 $x_1[n_1]$, $x_2[n_2]$ は、

$$x_1[n_1] = x\left(\frac{n_1}{f_1}\right) \quad (33)$$

$$x_2[n_2] = x\left(\frac{n_2}{f_2} + T_{21}\right) = x\left(\frac{n_2}{(1 + \varepsilon)f_1} + T_{21}\right) \quad (34)$$

のように表される。すなわち、ある連続時刻 t に対応する離散時刻 n_1 , n_2 の差は、

$$n_2 - n_1 = \varepsilon n_1 - (1 + \varepsilon) f_1 T_{21} \quad (35)$$

のように表される。第 2 項は録音開始時刻の差に起因する定数項であるが、第 1 項は時間の経過により n_1 が大きくなるのに比例して、拡大していくことになる。これは見かけ上、音源の移動に伴う伝搬時間差の変化と区別がつかない。

例えば、録音機器 1, 2 のサンプリング周波数を 16000 Hz, 16000.5 Hz とし、これらを 20 cm だけ離れた配置で、正面方向の音源信号を 10 秒間録音を行ったとしよう。また簡単のため、 $T_{21} = 0$ とする。このときサンプリング周波数ミスマッチは、31.25 ppm であり、直感的には十分小さいように思われる。しかしながら 10 秒間の音響信号は、録音機器 1 では 160000 サンプル、録音機器 2 では 160005 サンプルに相当する。すなわち、正面方向から到来する音波は、1, 2 に同時に届くが、サンプリング周波数ミスマッチにより、10 秒間の録音信号の最後では 5 サンプルずれることになる。

一方、30 度方向から到来する音波に対する到来時間差は、

$$0.2 \sin 30^\circ / 340.0 \times 16000 \simeq 4.7 \quad (36)$$

で、約 4.7 サンプルに相当する。つまり、31.25 ppm という微小なサンプリング周波数のずれが引き起こす時間差は、10 秒間に音源が正面から 30 度方向に移動したのと区別がつかないことになる。当然のことながら、このままでは、チャンネル間の時間差からは音源位置情報が得られない。また、多くの線形アレー信号処理においては、音源からマイクロホンまでの伝達関数は線形自不変であることが仮定されているため、こうしたチャンネル間の時間差のドリフトは、音源分離などにも深刻な破綻を引き起こす^{(62), (63)}。

これを解決するためには、サンプリング周波数ミスマッチの

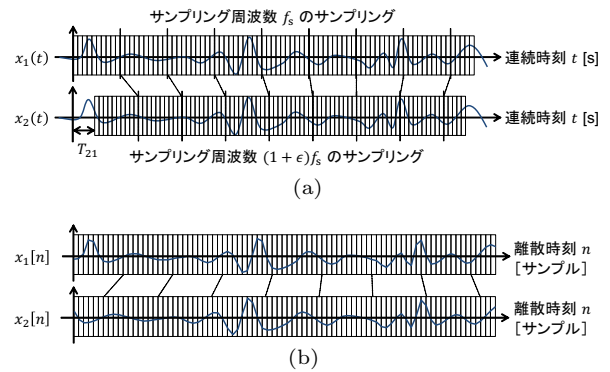


図 7 サンプリング周波数ミスマッチの影響。
(a) 連続波形, (b) 離散波形

推定と補償が必要である。前述のように、サンプリング周波数ミスマッチは離散波形の伸縮を引き起こすため、同一音源から到来する信号を 2 回観測できれば、それらの間隔の比率からサンプリング周波数ミスマッチが推定できる⁽⁶⁴⁾し、ブラインドでの推定も検討されつつある⁽⁶⁵⁾。一方、補償は本来は、sinc 関数補間によるリサンプリングが必要であるが、アレー信号処理の多くは STFT (Short-Time Fourier Transform) 領域で行われるため、アレー信号処理を目的とするならば、良い精度でミスマッチが補償された STFT 表現が得られればよいとも考えられる。

以下では STFT 領域で、サンプリング周波数ミスマッチをブラインドに推定し、補償する我々のアプローチ^{(13), (15), (19), (66), (67)}を紹介する。

5.2 時間周波数領域におけるサンプリング周波数ミスマッチのモデル化

観測信号 $x_1[n_1]$, $x_2[n_2]$ に、フレーム長が L のフレーム分析を適用して得られる、第 m サンプルを中心とするフレーム波形の短時間フーリエ変換 $X_i(k, m)$, $i = 1, 2$, $-L/2 \leq k \leq L/2 - 1$ を以下のように定義する。

$$X_i(k, m) = \sum_{l=0}^{L-1} w(l) x_i\left[l + m - \frac{L}{2}\right] \exp\left(-\frac{2\pi j k l}{L}\right) \quad (37)$$

ここで $w(l)$ は長さ L の再合成可能な窓関数、 k は離散周波数インデックス、 $j = \sqrt{-1}$ を表す。

まず、チャンネル 1 のあるフレームの中心サンプルを m として、 $m - L/2 \leq n_1 \leq m + L/2 - 1$ という 1 フレーム内での時間の対応関係を考える (図 8)。

式 (35) の関係から、サンプル m の近傍の n_1 と n_2 の間には

$$\begin{aligned} (n_2 - m) - (n_1 - m) \\ = \varepsilon(n_1 - m) + \varepsilon m - (1 + \varepsilon) f_1 T_{21} \end{aligned} \quad (38)$$

という関係が成り立つ。 $(n_1 - m)$ はフレーム内時刻に相当することに注意すると、

- 第 1 項の $\varepsilon(n_1 - m)$ はフレーム内時刻に線形に変化する

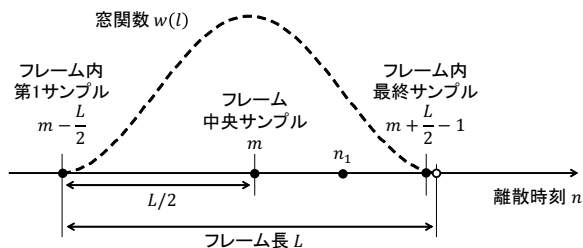


図8 ある分析フレーム内におけるフレーム中心サンプル m と離散時刻 n_1 の関係

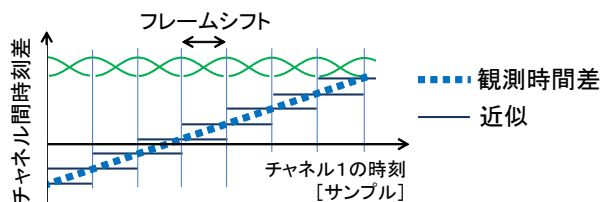


図9 フレーム分析におけるチャンネル間時間差ドリフトの影響とチャンネル内での時間差を無視した近似モデル

る項

- 第2項の εm はフレーム中心時刻 m とともに拡大する項
- 第3項の $(1 + \varepsilon)f_1 T_{21}$ は録音開始時刻の差に起因する項

を表している。

第3項の録音開始時刻の差に起因する項を正確に推定するのは難しいが、2.3で述べたように、ブラインド音源分離などでは、小さなチャンネル間の固定時間差は許容されるため、チャンネル間の相互相関が最大になるようにあらかじめ $x_2[n_2]$ をシフトしておくことで十分小さくできる。よって以下では、このような前処理により $T_{21} \approx 0$ であり、かつ $n_1 = 0$ 、 $n_2 = 0$ が同じ連続時刻を表すように、離散時刻原点を取り直したものとする。

また、いま ε は十分小さく、第1項で表されるフレーム内での変化は無視できると仮定すると、 $m - L/2 \leq n_1 \leq m + L/2 - 1$ において、

$$(n_2 - m) \simeq (n_1 - m) + \varepsilon m \quad (39)$$

のように、サンプリング周波数ミスマッチに起因して発生するチャンネル間の時間差が、フレーム内で一定で、フレーム中心に線形であるというモデルを得る(図9)。

もし全てのフレームの中心サンプル m に対して

$$|\varepsilon m| \ll L \quad (40)$$

であれば、時間差 εm の補償はフレーム内の時間シフトとみなすことができ、これはSTFT領域では線形位相により近似的に補償することができる。例えば前述の例で、録音機器1, 2のサンプリング周波数を16000 Hz, 16000.5 Hzで10秒間録音した信号の場合、10秒間の中心で時間差0になるようにシフトしたとすると、 $|\varepsilon m|$ は最大でも2.5サンプルであるから、 $L = 1024$ サンプルで分析する場合には、式(40)は十分成り立つと考えら

れる。

よって、第2チャンネルの信号 $X_2(k, m)$ に εm の線形位相回転を与えることにより、補償した信号 $\hat{X}_2(k, m)$ は以下のよう

$$\hat{X}_2(k, m) = X_2(k, m) \exp\left(\frac{2\pi j k \varepsilon m}{L}\right) \quad (41)$$

なお、録音が長時間に及んだ場合、式(40)の左辺は増大し、式(40)は満たされなくなるが、その場合には第2チャンネルのフレーム分析におけるフレームシフトを伸縮させる手法を検討している(64), (67)。

5.3 定常音場モデルに基づく最尤法によるミスマッチ推定

前述のように、チャンネル間の時間差のドリフトは、時間変化するチャンネル間位相差を観測信号中の全ての音像に付与するため、時間変化する伝達関数により全ての音源の位置を擬似的に移動させるような効果を生じる。逆に言えば、音源が移動しない場合には、音源が移動して見えているかどうかを調べることで、サンプリング周波数ミスマッチの補償具合を評価することができる。ここでは定常確率モデルに基づく最尤法により、サンプリング周波数ミスマッチを推定する手法について述べる。

全ての音源が移動せず、かつ振幅が定常であると仮定すると、正確な ε の推定を用いてサンプリング周波数のミスマッチを補償した観測信号

$$\hat{\mathbf{X}}(k, m; \varepsilon) = [X_1(k, m), \hat{X}_2(k, m; \varepsilon)]^T \quad (42)$$

は離散周波数 k ごとに定常となる。正確な ε の推定による補償で定常性を回復した観測信号 $\hat{\mathbf{X}}(k, m; \varepsilon)$ が共分散行列 $V(k)$ の零平均多変量複素正規分布に従うと仮定すると、その対数尤う度は

$$J(V, \varepsilon) = \sum_{k, m} \left(-\hat{\mathbf{X}}(k, m; \varepsilon)^H V(k)^{-1} \hat{\mathbf{X}}(k, m; \varepsilon) - \log 2\pi^2 - \log \det V(k) \right) \quad (43)$$

と表せる。ここで H は複素共役転置を表す。 $J(V, \varepsilon)$ は $V(k)$ に関しては、

$$V(k) \leftarrow \frac{1}{|\forall m|} \sum_m \hat{\mathbf{X}}(k, m; \varepsilon) \hat{\mathbf{X}}(k, m; \varepsilon)^H \quad (44)$$

により最大化されるから、これを式(43)に代入すると第1項が定数となるため、対数尤う度関数は定数項を除いて以下のように単純化される。

$$J(\varepsilon) = - \sum_k \log \det \sum_m \hat{\mathbf{X}}(k, m; \varepsilon) \hat{\mathbf{X}}(k, m; \varepsilon)^H \quad (45)$$

サンプリング周波数ミスマッチ ε の補償が不正確であれば、時刻のドリフトによる定常性の低下により対数尤う度 $J(\varepsilon)$ が小さくなるため、 $J(\varepsilon)$ の最大化により ε を推定することができる。この尤う度最大化は解析的に解くことができず、また、例えば

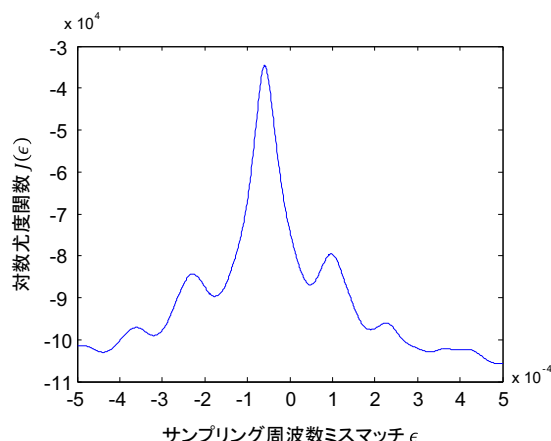


図 10 対数尤度関数 $J(\varepsilon)$ の具体例^{(15), (66)}

図 10 に示すように複数の局所最適解を持つが、我々は局所最適解を避ける機構を導入した黄金比探索を用いて効率的な ε の最尤解の探索アルゴリズムを提案している^{(15), (66)}。

5.4 サンプリング周波数ミスマッチの推定精度と音源分離への寄与

複数の長さのデータに対するサンプリング周波数の補正精度を平均二乗誤差 (RMSE) で評価した実験結果の例^{(15), (66)}を図 11 に示す。最も短い 3 秒の観測信号でも RMSE は元のサンプリング周波数ミスマッチ ε の 10 分の 1 以下に収束し、データが増えるにつれて急速に小さくなる。したがって観測信号の定常性を仮定した尤度がサンプリング周波数のミスマッチの評価尺度として有効であるということが分かる。

次に、サンプリング周波数のミスマッチを提案手法により補償することで、音源分離の性能が回復することを確かめるための分離精度評価を行った例を示す。分離フィルタの学習は、与えられた観測信号全体を使って補助関数型独立ベクトル分析⁽⁶⁸⁾により行った。評価尺度には、信号対ひずみ比 (SDR: Signal-to-Distortion Ratio) を用いた⁽⁶⁹⁾。また、SDR を算出するための参照信号としては、サンプリング周波数の変更を施していないマイクロホンにおける各音源の音像を用いた。

実験結果を図 12 に示す。まず未処理の分離性能が非常に低い値を示していることから、この条件ではサンプリング周波数ミスマッチの補償をしなければ音源分離ができない厳しい条件であるということが分かる。次に、 ε の正しい値を与えた位相補償はサンプリング周波数ミスマッチがない場合よりも SDR が 2 dB 程度低いだけであり、BSS のためのサンプリング周波数ミスマッチ補償に STFT 領域における位相補償が有効であるということが分かる。また、サンプリング周波数ミスマッチをブラインドに推定して補償する提案手法は、正しい ε を与えた場合とほとんど性能差がなく、提案手法の最尤推定は性能限界に近い高い精度でサンプリング周波数ミスマッチ ε を推定できていることを示している。以上より、音源分離のためのサンプリング周波数ミスマッチ補償として提案手法が有効であるということが確認できる。

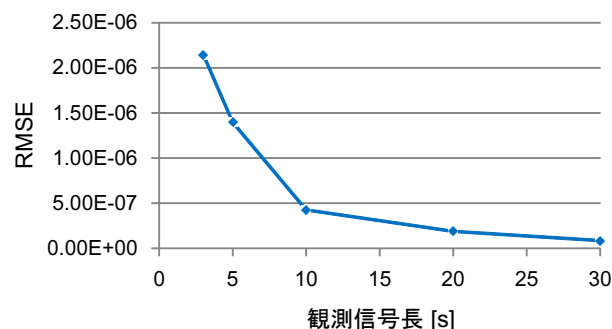


図 11 推定されたサンプリング周波数ミスマッチ ε の平均二乗誤差⁽¹⁵⁾

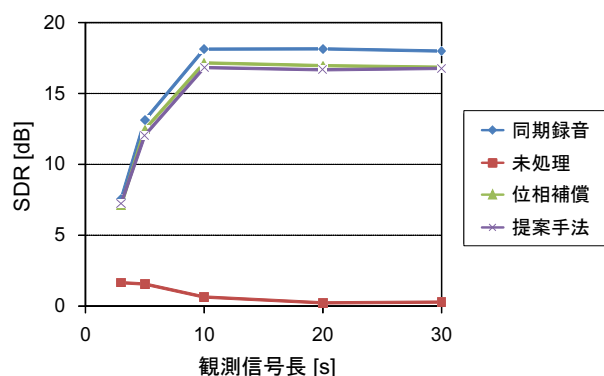


図 12 サンプリング周波数ミスマッチ補償のブラインド音源分離による性能評価^{(15), (66)}

6. おわりに

本稿では、アドホックマイクロホンアレイという新しいマイクロホンアレイ信号処理の枠組みについて紹介した。モバイル端末の普及やネットワークの広帯域化を背景に、こうした研究は今後も必要な方向の一つとなっていくと思われる。本稿では十分紹介しきれなかったトピックもあるが、引用した文献と併せて参照して頂くことで、本稿が読者の一助となれば幸いである。

謝辞 本研究の一部は、科学研究費補助金基盤研究 (B) (課題番号 25280069) の助成を受けて行ったものである。また、図 2 に示したイラストは、東京工業大学の益富恵子氏に描いて頂いた。図 6 に示した音源定位実験は、東京大学大学院の柴田一暁氏、亀岡弘和准教授との共同研究として行った。あわせてここに謝意を表する。

文 献

- (1) N. Ono, "Fast stereo independent vector analysis and its implementation on mobile phone," Proc. IWAENC, Sept. 2012.
- (2) N. Ono, "Blind source separation on iPhone in real environment," Proc. EUSIPCO, Sept. 2013.
- (3) http://www.onn.nii.ac.jp/demo_iphoneBSS.html
- (4) N. Ono, Z. Koldovsky, S. Miyabe, and N. Ito, "The 2013 signal separation evaluation campaign," Proc. MLSP,

- Sept. 2013.
- (5) <http://sisec.wiki.irisa.fr/>
 - (6) E. Vincent, J. Barker, S. Watanabe, J. Le Roux, F. Nesta, and M. Matassoni, "The second CHiME speech separation and recognition challenge: Datasets, tasks and baselines," *Proc. ICASSP*, pp. 126–130, May 2013.
 - (7) http://spandh.dcs.shef.ac.uk/chime_challenge/
 - (8) A. Bertrand, "Applications and trends in wireless acoustic sensor networks: A signal processing perspective," *Proc. Symposium on Communications and Vehicular Technology (SCVT)*, 2011.
 - (9) K. Na, Y. Kim, and H. Cha, "Acoustic sensor network-based parking lot surveillance system," *Proc. European Conference on Wireless Sensor (ECWS)*, pp. 247–262, 2009.
 - (10) 豊田卓矢, 宮部滋樹, 山田武志, 小野順貴, 牧野昭二, "分散型マイクロホンアレイを用いた交通量モニタリング," 2014 信学総大, no. A-10-10, March 2014. (to appear)
 - (11) <http://www.shotspotter.com/>
 - (12) K. Su, M. Naaman, A. Gurjar, M. Patel, and D. Ellis, "Making a scene: Alignment of complete sets of clips based on pairwise audio match," *Proc. ICMR*, Jun. 2012.
 - (13) 宮部滋樹, 小野順貴, 牧野昭二, "非同期録音信号の線形位相補償によるブラインド同期と音源分離への応用," 音講論集, pp. 689–690, Sept. 2012.
 - (14) 町田晃平, 伊藤彰則, "非同期・間欠観測型複数マイクロホンを用いたスペクトル減算による雑音除去の検討," 音講論集, pp. 171–172, March 2013.
 - (15) 宮部滋樹, 小野順貴, 牧野昭二, "非同期録音ブラインド同期のための線形位相補償の効率的な最尤探索," 音講論集, pp. 733–734, March 2013.
 - (16) ソウデン・メレズ, 木下慶介, 中谷智広, "ノード内・ノード間情報の統合に基づく分散マイクアレイ音源分離," 音講論集, pp. 797–798, March 2013.
 - (17) 加古達也, 小林和則, 大室 伸, "非同期分散マイクアレイのための振幅スペクトルビームフォーマの提案," 音講論集, pp. 829–830, March 2013.
 - (18) 柴田一暁, 小野順貴, 亀岡弘和, "音の発信を利用したスマートフォンアレイの機器位置推定," 音講論集, pp. 591–592, Sept. 2013.
 - (19) 宮部滋樹, 小野順貴, 牧野昭二, "非整数サンプルシフトのフレーム分析を用いた非同期録音の同期化," 音講論集, pp. 593–596, Sept. 2013.
 - (20) 木下慶介, 中谷智広, ソウデン・メレズ, "音源信号の距離減衰・局在性を考慮した大規模分散マイク音源分離に関する一検討," 音講論集, pp. 597–600, Sept. 2013.
 - (21) 加古達也, 小林和則, 大室 伸, "非同期分散マイクロホンアレイにおける同時発話区間の複数話者分類," 音講論集, pp. 797–798, Sept. 2013.
 - (22) 秋葉友良, 岩野公司, 緒方 淳, 小川哲司, 小野順貴, 篠崎隆宏, 篠田浩一, 南條浩輝, 西崎博光, 西田昌史, 西村竜一, 原 直, 堀 貴明, "クラウド時代の新しい音声研究パラダイム," 情報学SLP 研報, 2012-SLP-92, no. 4, pp.1-7, July 2012.
 - (23) R. Lienhart, I. Kozintsev, S. Wehr, and M. Yeung, "On the importance of exact synchronization for distributed audio signal processing," *Proc. ICASSP*, pp. IV-840–IV-843, 2003.
 - (24) Y. Chisaki, D. Murakami, and T. Usagawa, "Network-based multi-channel signal processing using the precision time protocol," *Proc. APSIPA*, 2012.
 - (25) J. Schmalenstroeer and R. Haeb-Umbach, "Sampling rate synchronisation in acoustic sensor networks with a pre-trained clock skew error model," *Proc. EUSIPCO*, 2013.
 - (26) <http://ipin-conference.org/>
 - (27) Z. Liu, Z. Zhang, L.-W. He, and P. Chou, "Energy-based sound source localization and gain normalization for ad hoc microphone arrays," *Proc. ICASSP*, pp. II-761–II-764, 2007.
 - (28) M. Chen, Z. Liu, L.-W. He, P. Chou, and Z. Zhang, "Energy-based position estimation of microphone and speakers for ad hoc microphone arrays," *Proc. WASPAA*, pp. 22–25, 2007.
 - (29) 千葉大将, 小野順貴, 宮部滋樹, 山田武志, 牧野昭二, 高橋 祐, "伝達関数ゲイン基底 NMF による分散配置非同期録音における目的音強調の検討," 音講論集, March 2014. (to appear)
 - (30) *Optimum Array Processing*, H. L. Van Trees, Ed., John Wiley & Sons, 2002.
 - (31) A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis*, John Wiley & Sons, 2001.
 - (32) M. Hennecke, T. Plotz, G. A. Fink, J. Schmalenstroeer, and R. Haeb-Umbach, "A hierarchical approach to unsupervised shape calibration of microphone array networks," *Proc. SSP*, pp. 257–260, 2009.
 - (33) F. Jacob, J. Schmalenstroeer, and R. Haeb-Umbach, "Microphone array position self-calibration from reverberant speech input," *Proc. IWAENC*, 2012.
 - (34) F. Nesta and M. Omologo, "Cooperative Wiener-ICA for source localization and separation by distributed microphone arrays," *Proc. ICASSP*, pp. 181–184, 2010.
 - (35) T. Ono, S. Miyabe, N. Ono, and S. Sagayama, "Blind source separation with distributed microphone pairs using permutation correction by intra-pair TDOA clustering," *Proc. IWAENC*, Aug. 2010.
 - (36) M. Souden, K. Kinoshita, M. Delcroix, and T. Nakatani, "Distributed microphone array processing for speech source separation with classifier fusion," *Proc. MLSP*, Sept. 2012.
 - (37) M. Souden, K. Kinoshita, and T. Nakatani, "An integration of source location cues for speech clustering in distributed microphone arrays," *Proc. ICASSP*, pp. 111–115, May 2013.
 - (38) 安藤 繁, 篠田裕之, 小川勝也, 光山 訓, "時空間勾配法に基づく 3 次元音源定位システム," 計測自動制御学会論文集, vol.29, no.5, pp. 520–528, 1993.
 - (39) S. Ando, "An autonomous three-dimensional vision sensor with ears," *IEICE Trans. Inf. and Syst.*, vol. E78-D, no. 12, pp. 1621–1629, Dec. 1995.
 - (40) 鎌本 優, 守谷健弘, 西本卓也, 嵯峨山茂樹, "チャンネル間相関を用いた多チャンネル信号の可逆圧縮符号化," 情処学論, vol. 46, no. 5, pp. 1118–1128, 2005.
 - (41) S. Badar, N. Ono, and L. Daudet, "Microphone multiplexing with diffuse noise model-based principal component analysis," *Proc. WASPAA*, Oct. 2013.
 - (42) A. Bertrand and M. Moonen, "Distributed adaptive node-specific signal estimation in fully connected sensor networks - Part I: Sequential node updating," *IEEE Trans. Signal Process.*, vol. 58, pp. 5277–5291, 2010.
 - (43) A. Bertrand and M. Moonen, "Distributed adaptive node-specific signal estimation in fully connected sensor networks - Part II: Simultaneous & asynchronous node updating," *IEEE Trans. Signal Process.*, vol. 58, pp. 5292–5306, 2010.
 - (44) A. Bertrand and M. Moonen, "Distributed adaptive estimation of node-specific signals in wireless sensor networks with a tree topology," *IEEE Trans. Signal Process.*, vol. 59, no. 5, pp. 2196–2210, 2011.
 - (45) I. Himawan, I. McCowan, and S. Sridharan, "Clustered blind beamforming from ad-hoc microphone arrays," *IEEE Trans. Audio, Speech, and Language Process.*, vol. 19, no. 4, pp. 661–676, 2011.
 - (46) Y. Hioka and W. B. Kleijn, "Distributed blind source separation with an application to audio signals," *Proc. ICASSP*, pp. 233–236, 2011.
 - (47) C. H. Knapp and G. C. Cartar, "The generalized correlation method for estimation of time delay," *IEEE Trans. Acoust. Speech Signal Process.*, vol. 24, no. 4, pp. 320–327, Aug. 1976.
 - (48) R. O. Schmidt, "A new approach to geometry of range difference location," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 8, pp. 821–835, Nov. 1972.
 - (49) P. Stoica and J. Li, "Source localization from range-difference measurements," *IEEE Signal Process. Mag.*,

- vol. 23, no. 69, pp. 63–65, Nov. 2006.
- (50) N. Ono and S. Sagayama, “R-means localization: A simple iterative algorithm for range-difference-based source localization,” *Proc. ICASSP*, pp. 2718–2721, 2010.
 - (51) 小林和則, 古家賢一, 片岡章俊, “位置が未知である複数のマイクロホンを用いたブラインド音源位置推定,” *信学論 (A)*, vol. J86-A, no.6, pp. 619–627, June 2003.
 - (52) V. Raykar, I. Kozintsev, and R. Lienhart, “Position calibration of microphones and loudspeakers in distributed computing platforms,” *IEEE Trans. Speech Audio Process.*, vol. 13, no. 1, pp. 70–83, 2005.
 - (53) 柴田一暁, 小野順貴, 亀岡弘和, “音の発信を利用したキャリブレーションに基づくアドホックマイクロホンアレイによる音源定位,” *音講論集*, March 2014. (to appear)
 - (54) Y. Suzuki, F. Asano, H.-Y. Kim, and Toshio Sone, “An optimum computer-generated pulse signal suitable for the measurement of very long impulse responses,” *J. Acoust. Soc. Am.*, vol. 97, no. 2, pp. 1119–1123, 1995.
 - (55) N. Ono, H. Kohno, N. Ito, and S. Sagayama, “Blind alignment of asynchronously recorded signals for distributed microphone array,” *Proc. WASPAA*, pp.161–164, Oct. 2009.
 - (56) K. Hasegawa, N. Ono, S. Miyabe, and S. Sagayama, “Blind estimation of locations and time offsets for distributed recording devices,” *Proc. LVA/ICA*, pp. 57–64, Sept. 2010.
 - (57) 小野順貴, “補助関数法による最適化アルゴリズムとその音響信号処理への応用,” *音響誌*, vol. 68, no. 11, pp. 566–571, 2012.
 - (58) H. Miura, T. Yoshida, K. Nakamura, and K. Nakadai, “SLAM-based online calibration of asynchronous microphone array for robot audition,” *Proc. IROS*, pp. 524–529, 2011.
 - (59) M. Crocco, A. Del Bue, and V. Murino, “A bilinear approach to the position self-calibration of multiple sensors,” *IEEE Trans. Signal Process.*, vol. 60, no. 2, pp. 660–673, Feb. 2012.
 - (60) F. Jiang, Y. Kuang, and K. Åström, “Time delay estimation for TDOA self-calibration using truncated nuclear norm regularization,” *Proc. ICASSP*, pp. 3885–3889, May 2013.
 - (61) Y. Kuang and K. Åström, “Stratified sensor network self-calibration from tdoa measurements,” *Proc. EU-SIPCO*, 2013.
 - (62) E. R. -Arnuncio, T. S. Wada, and B.-H. Juang, “On dealing with sampling rate mismatches in blind source separation and acoustic echo cancellation,” *Proc. WASPAA*, pp. 34–37, Oct. 2007.
 - (63) Z. Liu, “Sound source separation with distributed microphone arrays in the presence of clock synchronization errors,” *Proc. IWAENC*, 2008.
 - (64) R. Sakanashi, N. Ono, S. Miyabe, T. Yamada, and Shoji Makino, “Speech enhancement with ad-hoc microphone array using single source activity,” *Proc. APSIPA*, Oct. 2013.
 - (65) S. M.-Golan, S. Gannot, and I. Cohen, “Blind sampling rate offset estimation and compensation in wireless acoustic sensor networks with application to beamforming,” *Proc. IWAENC*, 2012.
 - (66) S. Miyabe, N. Ono, and S. Makino, “Blind compensation of inter-channel sampling frequency mismatch with maximum likelihood estimation in STFT domain,” *Proc. ICASSP*, pp.674–678, May 2013.
 - (67) S. Miyabe, N. Ono, and S. Makino, “Optimizing frame analysis with non-integer shift for sampling mismatch compensation of long recording,” *Proc. WASPAA*, Oct. 2013.
 - (68) N. Ono, “Stable and fast update rules for independent vector analysis based on auxiliary function technique,” *Proc. WASPAA*, pp. 189–192, Oct. 2011.
 - (69) E. Vincent, C. Fevotte, and R. Gribonval, “Performance

measurement in blind audio source separation,” *IEEE Trans. Acoust. Speech Signal Process.*, vol. 14, no. 4, pp. 1462–1469, 2006.

(EA 研究会提案, 平成 26 年 1 月 24 日受付

2 月 21 日最終受付)



小野順貴 (正員)

2001 東大大学院工学系研究科計数工学専攻博士課程了。博士 (工学)。同年 4 月から東大大学院情報理工学系研究科助手。2005-04 から, 同大学院講師。2011-04 から国立情報学研究所情報学プリンシプル研究系/総合研究大学院大複合科学研究科准教授。音源分離, 音源定位, 音楽信号処理, などの研究に従事。2012 から, IEEE Transactions on Audio, Speech and Language Processing と Acoustical Science and Technology の Associate Editor。また, IEEE Senior Member のほか, 日本音響学会, 情報処理学会, 計測自動制御学会各会員。2000 日本音響学会佐藤論文賞, 2007 日本音響学会栗屋学術奨励賞, 2004 電気学会センサ・マイクロマシンシンポジウム五十嵐賞, 2008IEEE International Symposium on Industrial Electronics (ISIE) Best paper award など各受賞。

audio, Speech and Language Processing と Acoustical Science and Technology の Associate Editor。また, IEEE Senior Member のほか, 日本音響学会, 情報処理学会, 計測自動制御学会各会員。2000 日本音響学会佐藤論文賞, 2007 日本音響学会栗屋学術奨励賞, 2004 電気学会センサ・マイクロマシンシンポジウム五十嵐賞, 2008IEEE International Symposium on Industrial Electronics (ISIE) Best paper award など各受賞。



Kien LE TRUNG

2012 ドイツ Ernst Moritz Arndt 大博士課程了。2013-01 からベトナム Hu 大講師。2013-12 から国立情報学研究所特任研究員。統計学, 情報理論, 音響信号処理の研究に従事。



宮部滋樹

2007 奈良先端大博士後期課程了。博士 (工学)。2007-10 日本学術振興会特別研究員 (PD)。2008-02 から 1 年間 Georgia Institute of Technology 客員研究員。2009 東大大学院情報理工学系研究科システム情報第一研究室特任研究員, 2010-11 同助教, 2011-04 筑波大生命領域学際研究センターマルチメディア研究室助教。音響信号処理の研究に従事。

2012 日本音響学会栗屋学術奨励賞受賞。



牧野昭二 (正員:フェロー)

1981 東北大大学院修士課程了。同年日本電信電話公社 (現 NTT) 入社。以来, NTT 研究所において, 電気音響変換器, 音響エコーキャンセラ, ブラインド音源分離などの音響信号処理の研究に従事。工博。現在, 筑波大生命領域学際研究センター教授。1995 日本音響学会技術開発賞。1997 本会業績賞。2002 電気通信普及財団 テレコムシステム技術賞。2002・2005

日本音響学会論文賞。2002・2004 本会論文賞。2006 ICA Unsupervised Learning Pioneer Award 各受賞。2007IEEE MLSP Competition Award。IEEE Distinguished Lecturer。IEEE Fellow。