

小特集—マイクロホンアレイの新しい技術展開—

非同期分散マイクロホンアレイに基づく音響信号処理*

小野 順 貴 (国立情報学研究所/総合研究大学院大学)**・宮 部 滋 樹, 牧 野 昭 二 (筑波大学)**

43.60.-c

1. はじめに

マイクロホンアレイは、複数のマイクロホンにより音場の空間的な情報を取得し、単一マイクロホンでは困難な、音源定位、音源強調、音源分離などを行う枠組みである。一般には、用いられるマイクロホンの数が多いほど得られる空間情報が多くなるため、制御できる指向性の自由度が増加し、また、マイクロホンを広範囲に配置することができるほどカバーできる範囲が広がり、定位や分離の性能向上が期待できる。

しかしながら、マイクロホンアレイ信号処理においては、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。よって従来は、各チャンネルを正確に同期させるために、すべてのチャンネルは多チャンネル A/D 変換機に接続され、同一クロックによりサンプリングされる必要があるが、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となっていた。

一方、我々の身の周りには、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。こうし

た録音機器によりマイクロホンアレイ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。こうした別々の録音機器による多チャンネル録音に基づくアレイ信号処理の枠組みは、近年、非同期分散マイクロホンアレイ、アドホックマイクロホンアレイ¹などと呼ばれ、国内外で近年関心が高まりつつある [1]。

本稿では、非同期分散マイクロホンアレイの新しい展開について、関連研究と共に著者らの取り組みを紹介する。なお、本稿では特に、非同期マイクロホンアレイにおけるチャンネル非同期の影響、音源分離や音源強調、並びにその応用について中心に紹介する。著者らの別の解説論文 [2, 3] や、同じ本小特集の中臺先生の解説論文と併せて参照いただきたい。

2. チャンネル非同期がアレイ信号処理に与える影響

一般に非同期分散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、各録音チャンネルは同期していない。いわゆるビームフォーミングや音源定位を行うためには、マイクロホン位置の推定も必要となるが、SN 比最大化ビームフォーマ [4] や独立成分分析 [5] など、音源強調、音源分離手法の中には、マイクロホンの位置情報を必要としない手法もある。そこでここでは主に、チャンネル非同期の影響について考える。

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 (図-1 参照)、信号処理には異なる影響を及ぼす。定常的な時間シフトについては、信号間の相互相関が最大となるように信号をシフトすることで十分な場

* Acoustic signal processing based on asynchronous and distributed microphone array.

** Nobutaka Ono (NII, Tokyo, 101-8430/SOKENDAI) e-mail: onono@nii.ac.jp

*** Shigeki Miyabe and Shoji Makino (Tsukuba University, Tsukuba, 305-8573)

¹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も呼ばれる。

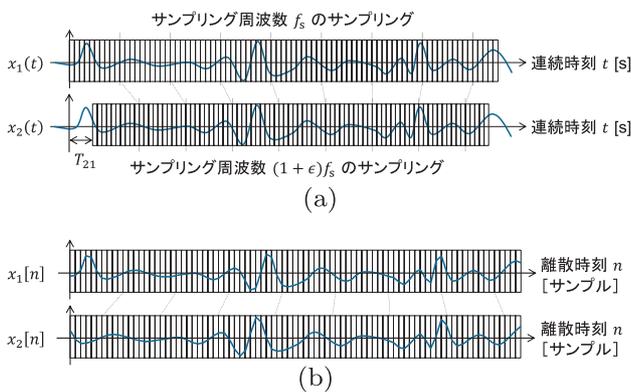


図-1 サンプル周波数ミスマッチの影響
(a) 連続波形, (b) 離散波形

合も多い。音源からマイクまでの伝達関数は、SN比最大化ビームフォーマにおいては学習区間の信号から、独立成分分析においてはブラインドに推定されるため、たとえ信号間に小さな一定のシフト誤差が残っていても、あまり問題にならないためである。しかしながら、サンプル周波数のミスマッチは重大な問題となる。

いま、録音機器 1, 録音機器 2 のサンプル周波数を f_1, f_2 で表すと、

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

が、録音機器間の相対的なサンプル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、こうしたわずかなミスマッチがアレイ信号処理に与える影響は大きい。

例えば、録音機器 1, 2 のサンプル周波数を 16,000 Hz, 16,000.5 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、同時に録音を開始したと考える。このときサンプル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 160,000 サンプル、録音機器 2 では 160,005 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、2 に同時に届くが、サンプル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、

$$0.2 \sin 30^\circ / 340.0 \times 16,000 \simeq 4.7 \quad (2)$$

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ル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、多くの線形アレイ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時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 [6, 7]。

3. サンプル周波数ミスマッチのブラインド補償に基づくブラインド音源分離

サンプル周波数ミスマッチの問題に対する一つの解決法は、ミスマッチ度合いを推定し、補償した後に、従来のアレイ信号処理を適用することである。前述のように、サンプル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ル周波数ミスマッチが推定できる [8] し、観測信号のみからの、サンプル周波数ミスマッチのブラインド推定と補償も検討されつつある [7, 9]。

前述のように、サンプル周波数ミスマッチはチャンネル間の時間差のドリフトを引き起こし、これはすべての音源の位置を疑似的に移動させるような効果を生じる。逆に言えば、音源が移動しないことを仮定できる場合には、チャンネル間時間差ができるだけ定常になるように（音源位置ができるだけ動かないように）することで、サンプル周波数ミスマッチを補償するアプローチが考えられる。我々は近年、STFT (Short-Time Fourier Transform) 領域において、定常な確率モデルに基づく最尤法により、サンプル周波数ミスマッチを推定し、補償する手法を提案している。ここでは手法の詳細は文献 [10–14] を参照いただくこととし、幾つかの実験結果を紹介する。

まず、複数話者の混合音声を録音した、複数の長さのデータに対するサンプル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 [11, 12] を図-2 に示す。最も短い 3 秒の観測信号でも RMSE は元のサンプル周波数ミスマッ

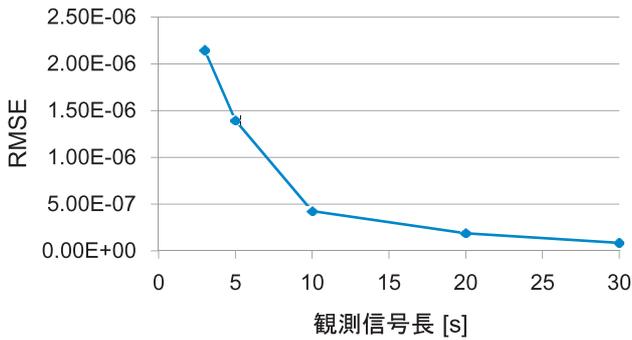


図-2 推定されたサンプリング周波数ミスマッチ ε の平均二乗誤差 [11]

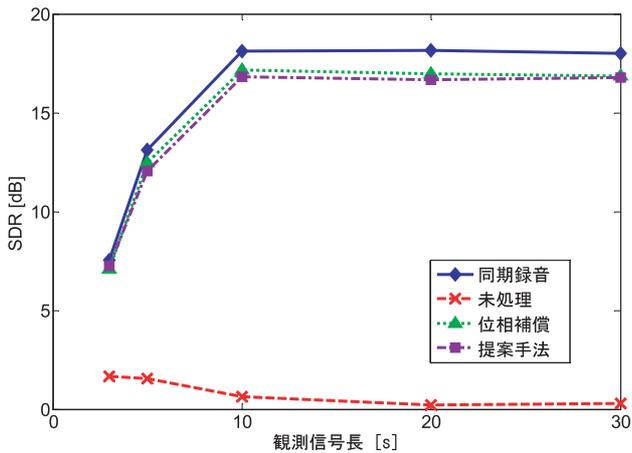


図-3 サンプリング周波数ミスマッチ補償のブラインド音源分離による性能評価 [11, 12]

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。この実験より、観測信号の定常性を仮定した手法の有効性が確認できる。

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 [15] により行った。分離性能評価には BSS Toolbox [16] を用い、評価尺度としては SDR (Signal to Distortion Ratio) を用いた。

実験結果を図-3 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、非同期チャンネルの録音であっても、音源が動かないと仮定できる場合には、こうした手法によりブラインドにチャンネルを同期化し、ブラインド音源分離が適用できると考えられる。また、我々は、符号化録音に対しても、本手法が破綻せず動作することを実験により確認している [17]。

4. 伝達関数ゲイン基底 NMF を用いた非同期録音に対する教師あり音源分離

非同期分散マイクロホンアレイでは、アレイ素子として別々の録音機器を用いることができるため、関心のある個々の音源の近くにマイクロホン配置することができる場合も多い。こうした場合には、チャンネル間の時間差だけでなく、チャンネル間の振幅比も、音源を区別する重要な手がかりとなる。そこで我々は、STFT 領域での振幅情報のみを用い、非同期録音に頑健な音源分離手法についても検討している。

通常、マイクロホンアレイにおける混合過程は、STFT 領域上での複素スペクトルに対する線形混合として

$$\mathbf{X}(\tau, \omega) = A(\omega)\mathbf{S}(\tau, \omega) \quad (3)$$

のように表現される。ここで、 $\mathbf{X}(\tau, \omega)$ 、 $\mathbf{S}(\tau, \omega)$ は多チャンネル観測信号の STFT 表現、 $A(\omega)$ は混合行列を表す。しかしながら前述のように、サンプリング周波数ミスマッチの影響により、チャンネル間の時間差がドリフトし、みかけ上の混合過程が時間と共に変化するような場合には、このような線形時不変の混合モデルは成り立たない。

ここでは代わりに、振幅スペクトルに対して近似的に線形混合が成立すると仮定する。すなわち、

$$\bar{\mathbf{X}}(\tau, \omega) \simeq \bar{A}(\omega)\bar{\mathbf{S}}(\tau, \omega) \quad (4)$$

ただし、 $\bar{\mathbf{X}}(\tau, \omega)$ 、 $\bar{\mathbf{S}}(\tau, \omega)$ は、 $\mathbf{X}(\tau, \omega)$ 、 $\mathbf{S}(\tau, \omega)$ の要素ごとに絶対値をとったベクトル (振幅スペクトルベクトル)、 $\bar{A}(\omega)$ は、各音源から各マイクロホンまでの伝達関数のゲインを要素とする混合

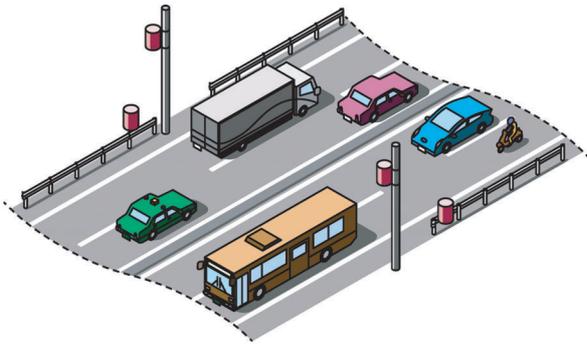


図-7 アドホックマイクロホンアレイによる交通量モニタリング

を必要とせず、身の回りにある汎用的な録音機器を用いることができ、また、配線を考慮することなくシステムを構築できる柔軟性が大きく、様々な応用可能性があると考えられる。モバイル端末の録音を用いた会議事録の自動作成などは典型的なシナリオの一つであるし、危険な異常音の位置推定 [21] などにも応用可能と思われる。

ここでは音による監視システムの一つとして、交通量モニタリングへの応用を紹介しよう。道路の交通量を把握することは渋滞緩和や道路整備などのために重要であり、定点観測のためには、ループ型センサのほか、超音波センサ、赤外線センサ、画像処理などが用いられている [22] が、これらの設置や維持にかかるコストは少くない。カメラによって撮影された動画に対して画像処理を行う手法も試みられているが、動画はデータが膨大になる他、暗いところでは精度がでない問題がある。これに対し図-7のように、道路の両端に録音機器を設置するだけアドホックマイクロホンアレイを構成し、音響信号から道路の交通量を推定することができるになれば、設置のコストも大幅に削減でき、昼夜問わず交通量のモニタリングが容易になることが期待できる。

ここでは図-8に示すように、筑波大学近辺の2車線道路において、道路の両側のほぼ同じ位置に設置したICレコーダにより車両の走行音を観測し、通過車両数の推定を試みた我々の実験例 [23] を紹介する。

車両がほぼ定速走行している場合には、車両がマイクロホンに一番接近した時刻で録音信号のパワーが極大になり、パワーのピーク検出により通過車両数が推定できることが期待される。しかしながら実際には、周囲の雑音の影響や、走行音の細

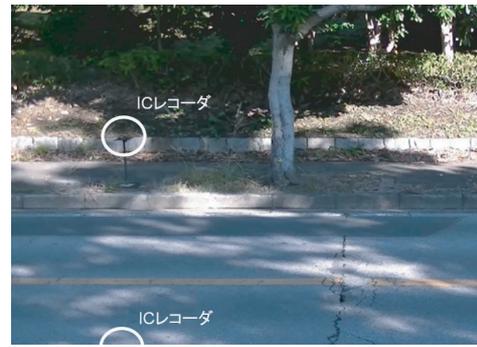


図-8 予備収録を行った筑波大学キャンパス内の道路

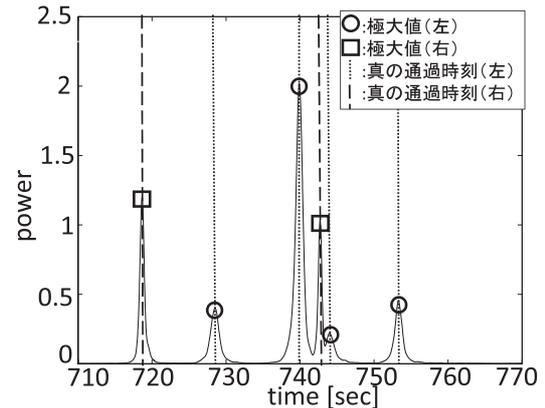


図-9 各走行方向に対する極大値と真の通過時刻 [23]

かな変動により、車両通過時刻に対応しないピークが現れるなどの問題が生じる。よって我々は、1) ウィーナフィルタによる雑音抑圧、2) パワーの平滑化、3) 閾値処理によるピーク検出を試みた。

また、車両の走行車線推定も試みた²。車線によって両側のマイクロホンからの距離が異なることから、チャンネル間で時間差が生じると考えられる。ただしこの時間差を検出するためには、チャンネルが同期している必要がある。ここでは、録音の最初と最後に、道路の中央付近で同期用の音響信号を発生し、これを用いて録音時刻のオフセットとサンプリング周波数ミスマッチを補償して同期させ [8]、その後、各チャンネルのパワーの極大値の時間差を検出した。また、走行音の録音と共に車両の走行を動画として撮影し、録画後に人手で車両の走行方向と通過時刻のラベリングを行い、比較を行った。走行車両に対するパワーの極大値と真の通過時刻の比較の例を図-9に、また各走行方向の真の交通車両数と推定交通車両数を表-2に示す。この結果から、パワー検出に基づく

² 2車線道路ではこれにより、車両の走行方向が推定できる。

表-2 各走行方向の真の交通車両数と推定交通車両数 [23]

	右	左
真の交通車両数	26	38
推定交通車両数	27	40
真の通過時刻と一致した推定交通車両数	22	35

単純な手法でも、比較的高い精度で走行車両の進行方向の推定ができていことが分かる。

今後は、ビームフォーミングや音源分離手法と組み合わせることで、精度向上と共に、走行車両が多く、混雑している道路への適用可能性を検証していく予定である。

6. おわりに

本稿では、非同期分散マイクロホンアレイという新しいマイクロホンアレイを用いた音響信号処理について紹介した。引用した文献と併せて参照していただくことで、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、4 章の内容は、筑波大学大学院の千葉大将氏、山田武志准教授、ヤマハ株式会社の高橋祐氏との共同研究、5 章の内容は、筑波大学大学院の豊田卓矢氏との共同研究として行ったものであり、図-7 に示したイラストは、アトリエ・サルバドールの益富恵子氏に描いていただいたものである。あわせてここに謝意を表す。

文 献

- [1] A. Bertrand, "Applications and trends in wireless acoustic sensor networks: A signal processing perspective," *Proc. Symp. Communications and Vehicular Technology (SCVT)* (2011).
- [2] 小野順貴, Trung-Kien Le, 宮部滋樹, 牧野昭二, "アドホックマイクロホンアレイ—複数のモバイル録音機器で行う音響信号処理—," 信学会 *Fundam. Rev.*, 7, 336–347 (2014).
- [3] https://www.jstage.jst.go.jp/article/essfr/7/4/7_336/_pdf
- [4] H.L. Van Trees, Ed., *Optimum Array Processing* (John Wiley & Sons, New York, 2002).
- [5] A. Hyvärinen, J. Karhunen and E. Oja, *Independent*

- Component Analysis* (John Wiley & Sons, New York, 2001).
- [6] E. Robledo-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 (2007).
- [7] Z. Liu, "Sound source separation with distributed microphone arrays in the presence of clock synchronization errors," *Proc. IWAENC* (2008).
- [8] R. Sakanashi, N. Ono, S. Miyabe, T. Yamada and S. Makino, "Speech enhancement with ad-hoc microphone array using single source activity," *Proc. AP-SIPA*, Oct. (2013).
- [9] S. Markovich-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).
- [10] 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 (2012.9).
- [11] 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 (2013.3).
- [12] 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 (2013).
- [13] 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 (2013.9).
- [14] S. Miyabe, N. Ono and S. Makino, "Optimizing frame analysis with non-integer shift for sampling mismatch compensation of long recording," *Proc. WASPAA* (2013).
- [15] N. Ono, "Stable and fast update rules for independent vector analysis based on auxiliary function technique," *Proc. WASPAA*, pp. 189–192 (2011).
- [16] E. Vincent, C. Fevotte and R. Gribonval, "Performance measurement in blind audio source separation," *IEEE Trans. Acoust. Speech Signal Process.*, 14, 1462–1469 (2006).
- [17] 宮部滋樹, 小野順貴, 牧野昭二, 高橋 祐, "非同期マイクロホンアレイの符号化録音におけるビットレートと同期性能の関係," 音講論集, pp. 725–726 (2014.3).
- [18] 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 (2013.3).
- [19] 戸上真人, 川口洋平, 小窪浩明, 大淵康成, "音源のチャンネル間振幅差を基底ベクトルとする音源分離," 音講論集, pp. 803–804 (2010.3).
- [20] 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, pp. 757–760 (2014.3).
- [21] <http://www.shotspotter.com/> (参照 2014-05-27).
- [22] 飯田恭敬, 北村隆一, 交通工学 (オーム社, 東京, 2008).
- [23] 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, p. 151 (2014.3).