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*

☆柴田一暁 (東大院・情報理工), 小野順貴 (NII/総研大), 亀岡弘和 (東大院・情報理工)

1 はじめに

マイクロホンアレイ信号処理の新しい枠組みとして、我々の身の回りにある様々な録音機器をアレイ信号処理に用いる、アドホックマイクロホンアレイが研究されている [1]。この手法は、多チャネル AD 変換器のような特別な機器を必要せず、また録音機器の個数を増やすことにより簡単にチャネル数を増やすことができるといった利点があるが、一般に通常のアレイ信号処理技術を適用するためには機器位置を推定したり [2]、チャネル間時間同期をとる必要がある [3] という課題が存在する。当研究室では近年アドホックマイクロホンアレイの研究に取り組んでおり [4, 5, 6]、音源定位とこれらのキャリブレーションを同時に行う手法が提案されているが、最適化関数が複雑であることや多数の音源が必要となることが課題であった。本研究はこれらの研究と異なり、近年目覚しく普及しているスマートフォンのような端末を録音機器とする条件において複数の機器位置を推定し、それらの録音信号同士を同期させる簡便で音源に依存しない手法を提案する。音を発信する機能を持つデバイスによる機器位置推定手法が先行研究 [7] で議論されているが、本稿ではこれを元に、機器間のサンプリング周波数ミスマッチを含む時間同期を同時に行う手法を提案し、実環境での実験結果を示す。また、その応用としてキャリブレーション結果に基づく音源定位手法を提案し、実験結果とともに述べる。

2 本研究の目的

アドホックマイクロホンアレイでは録音機器の位置が未知であり、録音開始時刻も機器ごとに異なり未知である。また、サンプリング周波数も機器ごとに少しずつずれが存在している。到来時刻や位相差を用いた音源分離などでは時間同期がとれている必要があり、音源定位や位置情報を用いたビームフォーミングなどを行うためには各録音機器の位置を推定する必要がある。但し機器位置の推定に関しては、通常のアレイ信号処理では録音機器同士の相対的な位置関係が得られればよく、これは録音機器同士の距離から多次元尺度法 (MDS)[8] により求めることができる。よって本研究では、録音機器同士の距離推定および時間同期（録音開始時刻の推定とサンプリング周波数ミスマッチの補償）の問題を考える。

3 音の発信を利用した自己定位・時間同期の手法

3.1 本研究のアプローチ

通常のアレイ信号処理では到来する音響信号は未知であるため、音源位置、もしくはマイクロホン位置の主要な手がかりは、マイクロホンで録音された信号から求まる到来時間差 (Time Difference Of Arrival; TDOA) であり、この情報からは 2 つマイクロホンと音源の距離差がわかるだけで、音源とマイクロホンの距離、もしくはマイクロホン同士の距離は直接的には求まらない。しかしながら到来する音響信号が既知であり、かつ音響信号が生じた時刻がわかっていれば（つまり音源とマイクロホンで時計が同期していれば）音の到来時間 (Time Of Arrival; TOA) から、音源とマイクロホンの距離を直接求めることができる。

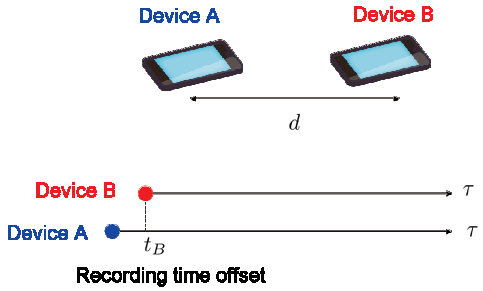

Fig. 1 機器間の距離が d, 録音開始時刻差が t_B である機器ペアの図

録音機器がスマートフォンのような端末である場合、録音機器自身が音を発信することができるが、録音機器同士の時計は同期していないため、到来時間は直接的には求まらない。しかしながら、2 つの録音機器が互いに音を発信することにより、これらの時計の時刻原点が異なっていても、距離を測定する手法が提案されている [7]。本研究では、これを元に録音開始時刻の推定を同時に行い、さらに機器による音の発信を利用したサンプリング周波数の補償を組み合わせ、より広い条件で適用可能な時間同期を行う。

3.2 音の相互発信を利用した機器間距離・録音開始時刻推定

簡単のために、各機器のスピーカーとマイクロホンの位置の差は無視できると仮定する。機器 A と B がともに音を発信する場合を考える。このとき、各チャネルにおける 2 つの音の TOA の差は録音開始時刻によらない。これらの TOA の差の差をとると、2 機器間を音が伝達するのにかかる時間の 2 倍となる。これの $\frac{1}{2}$ 倍に音速を乗じることで機器間距離が得られる。

また、各機器に 2 つの音が到達する時刻の和は、どちらも各音が発信された時刻の和に 2 機器間の距離の伝達時間を加えたものとなりグローバルな時間軸において等しくなるはずである。したがって、各チャネルにおける 2 つの音の TOA の平均時刻の差をとることによって録音開始時刻の差が求められる。

*Source Localization of Ad hoc Microphone Arrays Based on Calibration Using Sound Emissions. by Kazuaki SHIBATA (The University of Tokyo), Nobutaka ONO (National Institute of Informatics / The Graduate University for Advanced Studies), and Hirokazu KAMEOKA (The University of Tokyo)

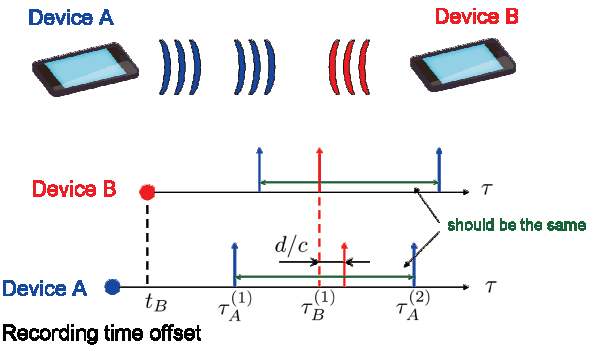

Fig. 2　2 機器によるキャリブレーションのための音の発信

3.3　サンプリング周波数ミスマッチの補償

録音機器ごとのサンプリング周波数にはミスマッチが存在しており、これを補償することもチャネル間時間同期の重要な要素である。機器 A,B のサンプリング周波数をそれぞれ f_A, f_B とすると、その相対誤差は $\varepsilon = \frac{f_B}{f_A} - 1$ で表せる。本研究では録音機器による音の発信を利用できることを考慮し、坂梨らのサンプリング周波数ミスマッチの補償方法 [9] を適用する。この手法は、ある音源が 2 回音を発信したと仮定すると、音源とマイクロホン位置が動かないのであれば、これらの 2 つの音が観測される間隔はいずれのマイクロホンでも等しくなることを利用する手法である。本研究では、一つの録音機器から 2 回音を発信することとし、各チャネルにおけるこれらの到来時刻の差からサンプリング周波数の相対誤差を計算し、補正を行う。

3.4　アルゴリズム

ここまでで述べた機器間距離と録音開始時刻の同時推定のためには各機器が最低でも 1 回ずつ音を出す必要があり、坂梨の手法によりサンプリング周波数ミスマッチを補償するためにはある機器が 2 回音を出す必要がある。このため、本研究では機器 A が 2 回、機器 B が 1 回音を発信することとする。ここで $n_{X\to Y}^{(k)}$ は機器 X 由来の k 回目の音が機器 Y で観測された時刻（サンプル数）であり、また $\tau_X^{(k)}$ は機器 X 由来の音が発信された時刻を、機器 A の録音開始時刻を時間原点として表したものとする。このとき以下の関係が成り立つ。

$$
\begin{aligned}
n_{A\to A}^{(1)} &= \tau_A^{(1)} f_A && (1)\\
n_{B\to A}^{(1)} &= (\tau_B^{(1)} + d/c) f_A && (2)\\
n_{A\to A}^{(2)} &= \tau_A^{(2)} f_A && (3)\\
n_{A\to B}^{(1)} &= (\tau_A^{(1)} + d/c - t_B) f_B && (4)\\
n_{B\to B}^{(1)} &= (\tau_B^{(1)} - t_B) f_B && (5)\\
n_{A\to B}^{(2)} &= (\tau_A^{(2)} + d/c - t_B) f_B && (6)
\end{aligned}
$$

1 回目と 2 回目の A 由来の音について、時間間隔と各チャネルで観測される時刻の関係は以下で表される。

$$
\begin{aligned}
n_{A\to A}^{(2)} - n_{A\to A}^{(1)} &= (\tau_A^{(2)} - \tau_A^{(1)}) f_A && (7)\\
n_{A\to B}^{(2)} - n_{A\to B}^{(1)} &= (\tau_A^{(2)} - \tau_A^{(1)}) f_B && (8)
\end{aligned}
$$

この 2 式から、サンプリング周波数の相対誤差は以下で表せる。

$$
\varepsilon = \frac{n_{A\to B}^{(2)} - n_{A\to B}^{(1)}}{n_{A\to A}^{(2)} - n_{A\to A}^{(1)}} - 1 \qquad (9)
$$

求められた ε から f_B を $(1+\varepsilon)f_A$ と置き換えられる。これを用いて式 1, 2, 4, 5 は

$$
\begin{aligned}
\tau_A^{(1)} &= \frac{1}{f_A} n_{A\to A}^{(1)} && (10)\\
\tau_B^{(1)} &= \frac{1}{f_A} n_{B\to A}^{(1)} - d/c && (11)\\
\tau_A^{(1)} &= \frac{1}{(1+\varepsilon)f_A} n_{A\to B}^{(1)} - d/c + t_B && (12)\\
\tau_B^{(1)} &= \frac{1}{(1+\varepsilon)f_A} n_{B\to B}^{(1)} + t_B && (13)
\end{aligned}
$$

と書き換えられる。これより、距離 d および録音開始時刻 t_B は

$$
d = \frac{c}{2f_A}\left\{(n_{B\to A}^{(1)} - n_{A\to A}^{(1)}) - (1+\varepsilon)(n_{B\to B}^{(1)} - n_{A\to B}^{(1)})\right\} \qquad (14)
$$

$$
t_B = \frac{1}{2f_A}\left\{(n_{B\to A}^{(1)} + n_{A\to A}^{(1)}) - (1+\varepsilon)(n_{B\to B}^{(1)} + n_{A\to B}^{(1)})\right\} \qquad (15)
$$

により求められる。

以上で 2 機器の場合について述べたが、機器数が 3 以上の場合も、機器 A が 2 回、それ以外の機器が 1 回音を出すことで同様の推定が可能である。

3.5　到来時刻の推定

ここまでで述べた手法を用いて推定を行うためには高い精度で TOA を測定する必要がある。しかし実環境においては TOA を正確に推定するのは容易ではない。そこで我々は以前に TSP 信号 (Time Stretched Pulse) によるインパルス応答測定手法 [11] に基づく TOA 推定を拡張し閾値処理を組み合わせた手法を示した [12]。これにより実環境において機器の特性などの影響でインパルス応答の反射波成分が直接波よりも大きくなる場合にも正確な TOA の推定が可能となる。

4　キャリブレーションに基づく音源定位

前章では音の発信を利用した機器位置・時間同期のキャリブレーション手法を提案した。本章では、このキャリブレーションの応用としての未知音源の定位について述べる。

録音機器の配置によらず適用できる音源定位の強力な手法に、到来時間差 (TDOA: Time Difference Of Arrival) を用いるものが存在する。先行研究においてこの考え方に基づいて音源位置を推定するアルゴリズムが考案されており、本稿ではこれをキャリブレーションを行った機器に適用することで音源定位を行う。

対象となる音源は単一と仮定し、音源から到来する音信号を前章までの手法でキャリブレーションを行っ

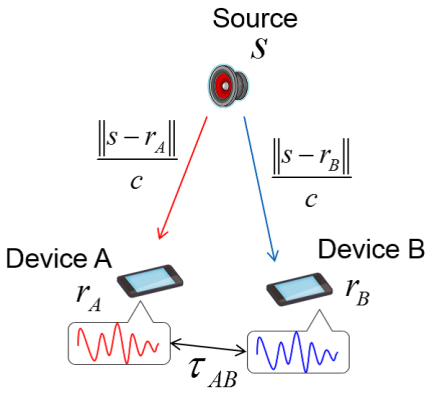

Fig. 3　観測される到来時間差と距離に基づく伝達時間

た L 個の機器で録音するものとする。録音信号から到来時間差を求めることができれば、これに音速を乗じることで音源-機器間の距離の差となる。これを観測量とし、以下の目的関数

$$J(s) = \sum_{m=1}^{L}\sum_{n=1}^{L}(||s-r_m|| - ||s-r_n|| - d_{mn})^2 \quad (16)$$

を最小化する音源位置 s を求める。

この目的関数は音源位置のパラメータに対して非線形であるため、閉形式で解くことはできない。しかしこの形式の目的関数は補助関数法による最適化が提案されている [13]。この最適化手法は簡潔な更新式が得られ、目的関数の収束が保証されるという長所を持つ。以下で補助関数法による最適化のアルゴリズムを示す。

目的関数を直接最小化する代わりに、以下の補助関数を導入する。

$$Q(s,\eta) = 2L\sum_{m=1}^{L}\left|\left|s - r_m + (\bar{r} + \bar{d_m})e_m\right|\right|^2 + C(\eta) \quad (17)$$

この補助関数は $J(s) = \min_\eta Q(s,\eta)$ の関係を満たす。ここで $\eta = \bar{r}, e_1, \cdots, e_L$ は補助変数、$C(\eta)$ は最適化に無関係な項であり、d_m は以下の式で求められる定数である。

$$\bar{d_m} = \frac{1}{L}\sum_{n=1}^{L} d_{mn} \quad (18)$$

以下の通りのパラメータ s と補助変数 η の最小化を交互に反復することにより最適化が行われる。

$$\bar{r} = \frac{1}{L}\sum_{m=1}^{L}|s - r_m| \quad (19)$$

$$e_m = \frac{s-r_m}{|s-r_m|} \quad (20)$$

$$s = \frac{1}{L}\sum_{m=1}^{L}(r_m + (\bar{r} + \bar{d_m})e_m) \quad (21)$$

この反復更新式は音源・マイクロホン間の距離の平均演算と音源推定距離の平均演算の繰り返しとなっている。

なお、この推定のためには未知音源による信号の TDOA を得る必要があるが、その代表的な手法はチャネル間相互相関関数のピーク時刻をとるものである。しかし通常の相互相関を用いた推定では残響や定常雑音の存在する環境で精度が低下する。そこで本研究ではコヒーレンスによる周波数重み付けを行った一般化相互相関 [10] を用いて推定を行った。これは相互相関に直接音成分の大きな周波数成分を強調するフィルタをかけ精度を向上させるものである。相互相関は 2 つの信号 $x(t), y(t)$ のクロススペクトル $S_{xy}(\omega)$ の逆フーリエ変換で表されるが、以下のフィルタ $U(\omega)$ をかけた、$U(\omega)S_{xy}(\omega)$ を逆フーリエ変換したものが一般化相互相関となる。

$$U(\omega) = \frac{1}{|E[X(\omega)Y^*(\omega)]|}\frac{|\gamma(\omega)|^2}{1-|\gamma(\omega)|^2} \quad (22)$$

ただしこの式の $X(\omega), Y(\omega)$ は周波数領域で表された信号で、$\gamma(\omega)$ は以下の式で与えられるコヒーレンスである。

$$\gamma(\omega) = \frac{E[X(\omega)Y^*(\omega)]}{\sqrt{|E[X(\omega)]|^2}\sqrt{|E[Y(\omega)]|^2}} \quad (23)$$

これは $X(\omega), Y(\omega)$ の各周波数成分の相関係数でありパワーによらない量である。フィルタ $U(\omega)$ の係数のうち $\frac{1}{|E[X(\omega)Y^*(\omega)]|}$ は白色化を行う項である。$\frac{|\gamma(\omega)|^2}{1-|\gamma(\omega)|^2}$ はコヒーレンスによる重み付けで、直接波が支配的な帯域では位相差が一定でコヒーレンスが大きいのに対し残響やノイズが支配的な帯域ではコヒーレンスは 0 に近づくため、これは直接波成分の強調を意味する。

5　実環境実験

5.1　時間同期の評価

実環境実験により本研究の時間同期について評価する。しかしトライアル毎の録音開始時刻や機器毎のサンプリング周波数ミスマッチは未知であるため、実際に時間同期の精度を直接評価するのは困難である。そこで本研究では、間接的にこれを評価する。録音する 2 つの機器 A, B の他に 2 機器のおおむね中間に新たに用意した機器 C を配置し、機器 A, B が音を出した後に機器 C も TSP を出した。これにより求められた機器 C 由来の音の TOA を、A, B の音によりサンプリング周波数ミスマッチを測定し録音開始時刻を求めることで補正ができ、これにより真の TDOA が得られる。機器 C の位置は変化していないので、時間同期が正確に行われればこの TDOA はトライアルによって変化しないはずである。表 1 で機器 A, B 間の距離と 6 回のトライアルで得られた TDOA の標準偏差を示す。

0.1ms の TDOA の差は、距離 3.4cm に相当するため高い精度で同期できていると考えられる。

5.2　機器位置推定の評価

本研究の機器位置推定について評価するため実環境実験を行った。距離推定の精度を含めた詳細な実験

Table 1　機器間距離別時間同期済 TDOA の標準偏差

機器間距離 [m]	TDOA の標準偏差 [ms]
0.5	0.049
1.0	0.065
1.5	0.062
2.0	0.120
2.5	0.137
3.0	0.112

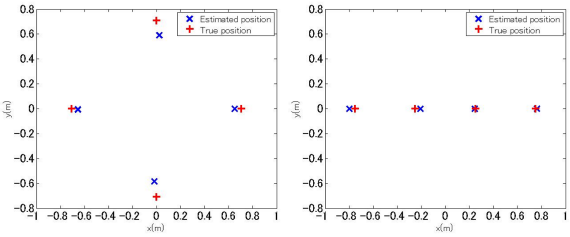

Fig. 4　機器位置推定結果と真の位置

結果を我々は以前に示している [12]。本稿では、推定結果の例を図で示すのみとする。図 4 では 2 種類の配置について推定を行った結果を示しているが、位置推定はおおむね正確に行われており、誤差は 10cm 以下となっている。

5.3　音源定位

音源定位についても、実環境において精度良く推定できることを確認するため評価実験を行った。録音は第 4 世代と第 5 世代の iPod touch 2 台ずつの計 4 台の機器を用い、最初にこれらの録音機器が音を出し、その後未知音源が音を出すという順で行った。録音データに対して、まず録音機器による音を用いて機器位置の推定および時間同期を行い、その後同期されたチャネル間で TDOA を推定しこれを観測量として定位した。未知音源は音声の録音データを用い、スピーカーを動かすことで移動音源とした。TDOA 推定は前章で述べた通りに一般化相互相関を用いて行った。ここでコヒーレンスを計算するときの期待値は相関の計算時に参照した 15 フレーム全ての平均を用いた。また、音源位置推定の反復回数は 1000 回とした。

図 6 で、正解・推定機器位置および時刻ごとの推定音源位置と真の軌道を示す。音源位置は多くの時刻ではおおむね正しく推定されている。ただし一般に音源の無音区間や複数の音が混合している時間区間では、単一音源の TDOA が得られず、ここでの仮定を満たさない。よって本実験では、最小化された目的関数の値自体を用いて、こうした時間区間を除いている。

6　まとめと今後の展望

本稿では録音機器による音の発信を利用した自己定位と時間同期の手法およびこれらのキャリブレーションの応用として未知音源の定位の手法を提案した。また実環境での時間同期の精度を評価し、さらに音源定位が実環境においても高い精度で行えることを示した。今後は機器位置推定の精度をさらに高めるため機器のマイク・スピーカー位置を考慮したモデルを考え、また音源定位以外のアレイ信号処理技術に本稿で示したキャリブレーションを行ったアドホックマイクロホンアレイが利用可能かを考察することを検討している。

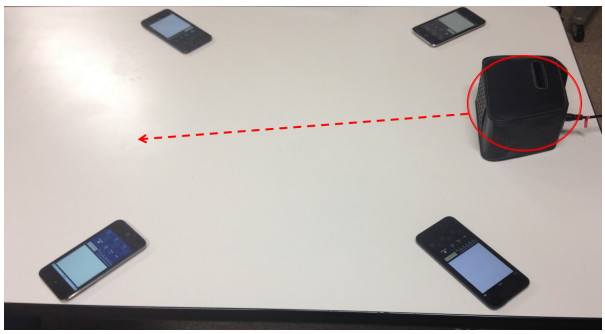

Fig. 5　実験における機器配置の写真

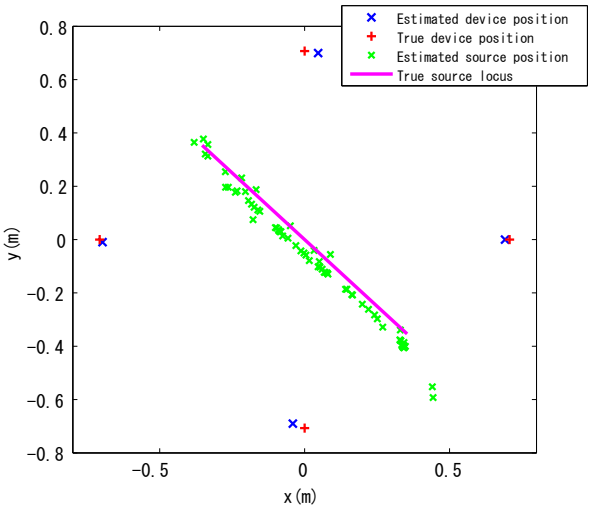

Fig. 6　移動音源の軌跡と時刻毎の推定音源位置

参考文献

[1] Chen, et al., Proc. WASPAA, 22-25, 2007.
[2] McCowan, et al., IEEE Trans. ASLP, 666-670, 2008.
[3] Miyabe, et al., Proc. ICASSP, 674-678, 2013.
[4] Ono, et al., Proc. WASPAA, 161-164, 2009.
[5] Hasegawa, et al., Proc. ICA/LVA, 57-64, 2010.
[6] Ono, et al. Proc. IWAENC, 2010.
[7] Hennecke, et al., Proc. HSCMA, 127-132, 2011.
[8] Torgerson, *Psychometrika*, vol. 17, 401-419, 1952.
[9] 坂梨他, 信学技報, vol.112, no.347, 17-22, 2012.
[10] Knapp et al., IEEE Trans. ASSP, 320-327, 1976.
[11] Aoshima, J. Acoust. Soc. Am., vol. 69, no.5, 1484-1488, 1981.
[12] 柴田他, 音講論, 1-1-8, 528-529, 2013.
[13] Ono, et al., Proc. ICASSP, 2718-2721, 2010.