

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*

☆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、多チャンネル 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ンネル数を増やすことができ、様々な応用が期待されている。しかし一般に録音機器位置は未知であり、各チャンネルも同期していないため、通常のアレイ信号処理を適用するためには、機器位置を推定したり、チャンネルを同期する必要がある [2, 3, 4]。本稿では、近年目覚しく普及しているスマートフォンをアレイ信号処理に用いるために、複数のスマートフォン位置を推定し、それらの録音信号同士を同期させる手法について論じる。スマートフォンは、音源としても機能させることができることに着目した手法が先行研究 [5, 6] で議論されているが、本稿ではこれを元に、残響を含む実環境で頑健に動作し、また機器間のサンプリング周波数ミスマッチにも対応した手法を提案し、実環境での実験結果を示す。

2 スマートフォンアレイの機器位置推定

2.1 問題設定

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るため、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。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) 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の問題を考える。

2.2 機器間距離推定アルゴリズム

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るならば (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。

スマートフォンアレイを考えた場合、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も、距離を測定する手法が提案されている [6]。簡単のために、各機器のスピーカとマイクロホンの位置の差は無視できると仮定する。このとき、機器 a から発信した音が機器 b に到来した時刻 (機器 b の時計で測定) を τ_b^a と表すと、到来時刻同

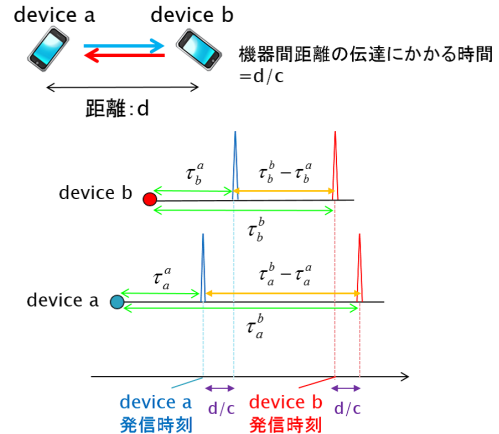


Fig. 1 音の相互発信を利用した到来時間の推定

士の差の関係から $\frac{2d_{ab}}{c} = -\tau_a^a + \tau_a^b + \tau_b^b - \tau_b^a$ の式が得られる (d_{ab} は機器間距離、 c は音速)。この式に各到来時刻と音速を代入することで、機器間距離を求めることができる。

しかしながら、実環境、特に残響が存在する環境下で到来時刻を求めるのは簡単ではない。また、機器間のサンプリング周波数のミスマッチも大きな誤差要因となる。次節以降、これらの問題をクリアする機器間距離推定法について議論する。

2.3 残響環境下での到来時刻推定

音の到来時刻を推定する最も単純な方法は、音源信号としてインパルスを用い、受信側では直接音として受信信号の最大ピークをとる時刻を求めることである。しかしながら、インパルス音は高い SN 比を確保するのが難しいため、ここではインパルス応答測定によく用いられる TSP 信号を用いる。TSP 信号は時間反転信号を畳み込むとインパルスになるという性質をもつ。この性質を利用し、録音信号に時間反転信号を畳み込むと、音源から録音信号までのインパルス応答に相当する信号 (ただしここでは時間同期がとれていない) が得られる。到来時刻を推定するためには、インパルス応答のうち直接音に対応する応答部を検出すればよく、一般には最大ピークの時刻を検出すればよいと考えられるが、実環境では図 2 の例のように反射音に由来するピークの方が大きくなることも起こりうる。

そこで、直接波は必ず最初に到来する性質を利用し、従来のように最大ピークをとる手法ではなく、ある閾値を超えた最も早いピークを検出することで直接波の到来時刻を推定する手法を提案する。閾値の決め方はいろいろな方法が考えられるが、本研究では、応答のスケールに依存しない簡便な方法として、応答全体の最大ピークの $\frac{1}{2}$ 倍を閾値として用いた。

2.4 サンプリング周波数ミスマッチの補償

録音機器同士のサンプリング周波数ミスマッチも、到来時間差の推定に深刻な影響を与える。本研究では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 [7] を適用する。

* Self Localization of Smartphone Arrays Using Sound Transmission. 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)

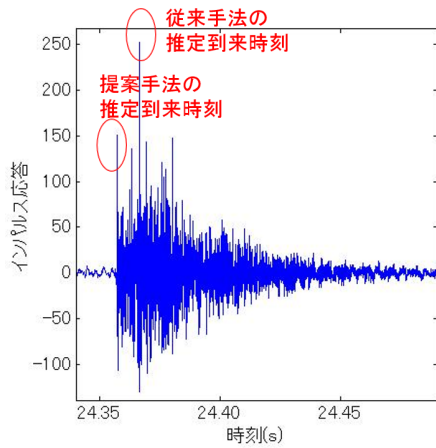


Fig. 2 他機器由来の音の推定インパルス応答

この手法は、ある音源が2回音を発したと仮定すると、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、各録音機器から2回ずつ音を発信することとし、これらの到来時刻を用いて周波数誤差を補正する。具体的には、各録音信号ごとに到来時刻を推定する。その後、同一の録音信号由来の到来時刻同士の差をそれぞれで求める。各機器の周波数を f_a, f_b と表したとき、機器 a を基準とした場合の機器 b の周波数誤差は $\epsilon = \frac{f_b}{f_a} - 1$ となる。実際の音の発信時刻の差を T とし、各録音信号で観測時刻のサンプル数差を n_a, n_b とすると $n_a = f_a T, n_b = f_b T$ の関係がある。よって周波数誤差は $\epsilon = \frac{n_b}{n_a} - 1$ で得られる。これを利用し、機器 b の到来時刻 t_b を機器 a の基準に補正した値 $t_b^0 = \frac{1}{1+\epsilon} t_b$ を用いて機器間の距離を求める。

3 実環境実験

3.1 サンプリング周波数ミスマッチの検証

機器のサンプリング周波数ミスマッチがどの程度であるかを確認するため実験を行った。1m離れた2つの機器がそれぞれ2回 TSP 信号を発信し、到来時刻の間隔を録音信号ごとに測定し、到来時刻間隔のサンプル数と一方の機器を基準とした周波数誤差を示す。録音機器は iPod touch 4 を2台用いた場合と、iPod touch 4 と 5 の2種類を用いた場合について示している。音源信号として帯域 0~24000Hz、信号長約 2.7 秒の TSP 信号を用い、各機器の設定上のサンプリング周波数は 48000Hz とした。また、機器間距離を求める際に用いる音速は約 20 °C の室温環境に対応して 343m/s とした。

表1で示しているように、機器が異なるとサンプリング周波数ミスマッチは大きくなる結果が得られた。しかし、ミスマッチの補償後の距離推定結果は 0.97m であり、真値に近い値が得られた。

Table 1 到来時刻間隔サンプル数と推定周波数誤差

	機器1間隔	機器2間隔	周波数誤差
同機種	376720	376725	13ppm
他機種	389139	388970	434ppm

3.2 残響環境下での機器間距離推定

室内環境における残響の影響および直接波到来時刻推定の精度を確認するため実環境実験を行った。2mから3mまで20cm刻みで機器間距離を変化させ、

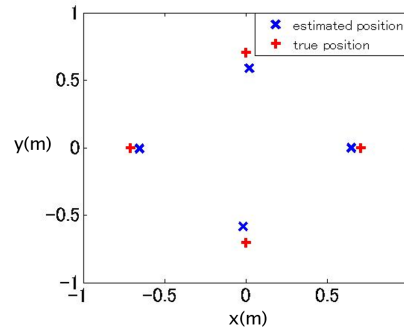


Fig. 3 位置推定結果

それぞれについて5回ずつ録音し、録音信号に従来手法および前章で述べた手法を用いて機器間の距離を測定した結果の平均を示す。どちらの手法についてもサンプリング周波数ミスマッチの補償は行った。機器は iPod touch 4 を2台用いた。音源信号、音速についての条件は前節の実験と同じである。

Table 2 従来手法と提案手法の推定距離平均 (m)

真の距離	2.0	2.2	2.4	2.6	2.8	3.0
従来手法	2.38	3.55	3.90	4.81	4.51	4.83
提案手法	1.94	2.12	2.34	2.54	2.73	2.94

従来手法では機器間距離が長くなるにつれ正解よりも長く推定されることがわかる。これは距離が長いほど直接波の減衰により反射波が相対的に大きくなり、直接波よりも後に到来する反射波の成分がインパルス応答の最大ピークとして現れやすくなるからだと考えられる。それに対して、提案手法は長距離測定においても真値に近い値が得られた。

また、iPod touch 4 が2台、5 が2台の計4台の機器を用いて提案手法で得た機器間距離に MDS を適用して位置推定を行った結果の例を図3に示す。全体的に機器同士の距離が小さく推定されているが、大まかな形状は正解に近いものが得られた。

4 まとめと今後の展望

本稿では音の発信機能を持つ録音機器による機器位置推定について述べ、残響が存在する環境での推定手法およびサンプリング周波数ミスマッチの補償を提案した。また提案手法の実環境実験における結果を示した。提案手法を用いた実験結果はおおむね正解に近いが、全体的に正解よりも機器間距離がやや短い値に推定されている。この原因として最も大きなものは機器のマイクとスピーカーの位置の違いであると考えられ、今後はこの違いを考慮したモデルにより正確な推定を試みるとともに、本研究をもとに話者の定位を行うことを検討している。

参考文献

- [1] Chen, et al., IEEE Workshop. ASPAA, 22-25, 2007.
- [2] McCowan, et al., IEEE Trans. ASLP, 666-670, 2008.
- [3] Hasegawa, et al., Proc. ICA/LVA, 57-64, 2010.
- [4] Miyabe, et al., Proc. ICASSP, 674-678, 2013.
- [5] Peng, et al., Proc. ACM Sensys, 1-14, 2007.
- [6] Hennecke, et al., Proc. HSCMA, 127-132, 2011.
- [7] 坂梨他, 信学技報, vol.112, no.347, 17-22, 2012.