Vol.2010-SLP-82 No.3
2010/7/23

音源分離法 SAFIA を用いたロボット動作雑音中の話者方向判定

川野恵右[†]　春木智貴[†]　川端豪[†]

モーター音や機械雑音への対処はロボットの音声処理を考える上で重要である．これらの直接雑音は人間の音声よりも大きな音量でマイクに入力されることもあり，例えば音声方向の判定に大きな悪影響を及ぼす．SAFIA は雑音混入音声中から特定の音声ストリームを分離する強力な手法として知られている．本報告では，SAFIA を音声の到来方向の判定に適用することを考える．ロボットの肩および胸にマイクを三角形に配置し，周波数帯域ごとの到達位相差を SAFIA で取捨選択し方向を判定する．実験の結果，SAFIA による音源分離を利用することによって，S/N 比-9dB の条件で方向判定誤りの約 85%を削減出来ることが分かった．

Speaker-direction Detection under Mechanical Noises using SAFIA Sound Segregation Method

Keisuke Kawano[†], Tomoki Haruki[†] and Takeshi Kawabata[†]

Reduction of motor and mechanical noises is an important issue for robot audition. Such direct noises are often larger than human voices and fatally disturb the speaker-direction detection. SAFIA is known as a powerful method to segregate speech streams from noisy sounds. This paper adopts the SAFIA to speaker-direction detection fields. Three microphones, located at robot shoulders and a chest, make a triangle. The inter-channel phase differences among these three channels indicate the speaker direction. Experiments show that the SAFIA method reduces 85% of direction-detection errors for S/N -9dB.

1. まえがき

近年のロボット研究の発展に伴い，ロボットと人間が共存する生活像が思い浮かべられるようになってきている．そのために，ロボットと人間のコミュニケーション手段が重要であり，その一つの形態として音声による対話がある．ロボットとの音声対話を実現するには，もちろん人間の対話内容を理解することが必要であり，このためには音声認識技術が必要である．さらに人が話しかけた方向を判定し，正しく行動するために，精度のよい方向判定技術が求められる．これによって，話しかけられた方向に振り向く，話しかけられた人に近づくなどの動作を行うことが出来る．しかし，音声認識，方向判定を行うにあたって，ロボット自身の発する機械音が，音声認識，話者方向判定の精度を大きく下げているという問題があり，雑音に強いシステムを構築する必要がある．

方向判定の手法として，アレーマイク[1][2]による方法があるが，マイクが多数必要となり，コストが増加してしまう．そこで少数マイクを用いた手法の必要性がでてくる．

藤原らはロボットに装着したマイクに加え，話者の口元に近接マイクを使用する手法を提案した[3]．この手法では，近接マイクで得られた雑音の入っていない音声の信号成分を，ロボットが認識した雑音混入音声から探し出すことにより雑音中でも高精度に方向判定を行うことを可能にした．しかし，この手法では近接マイクの利用が必須になる．

そこで，沼波らは音声キューを用いた近接マイク省略の手法を提案した[4]．これは，近接マイクで得た音声を探す代わりとして，一般的な音声モデルである音声キューを作り，記憶させておくことにより，近接マイクを用いず方向判定を行うことが可能になった．

これらの手法は，雑音の混入した音声中から音声ストリームを探し出す手法だと考えられる．一方，別のアプローチ方法として雑音を取り除く手法もある．スペクトルサブトラクション[5]や SAFIA[6]がこれにあたる．

本報告では，少数マイクによる話者方向判定の枠組みにおいて，雑音除去によって方向判定の精度向上を試みる．具体的には，雑音混入音声中から雑音スペクトルを減算するスペクトルサブトラクション法，到達位相差により音源分離する SAFIA の有効性を調べる．図 1 に少数マイクを用いて話者の方向を判定するロボットのイメージ写真を示す．

[†] 関西学院大学 理工学研究科
School of Science and Technology, Kwansei Gakuin University

Vol.2010-SLP-82 No.3
2010/7/23

図 1　話者方向判定のためのマイク配置

2. 雑音除去の手法

2.1 スペクトルサブトラクション

スペクトルサブトラクションとは，雑音混入音声のスペクトルから，推定される雑音のスペクトルを減算することによって，雑音を除去する手法である．スペクトルサブトラクションでは，前もって雑音のみを録音し，その平均パワースペクトルを求めておく．次に，入力信号(雑音混入音声)の各時刻に対するパワースペクトルを計算する．雑音の平均パワースペクトルを入力信号のパワースペクトルから減算することにより，雑音の成分を周波数領域で雑音混入音声から取り除く，というものである．処理のイメージを図 2 に示す．

減算する雑音スペクトルの重み（減算係数）は，小さすぎると雑音除去が不十分になり，逆に大きくしすぎると，除去してはいけない音声成分を過度の減算により壊してしまう可能性がある．よってなるべく適切な値を求めるために，雑音のない音声，雑音の混入した音声，雑音除去後の音声成分のスペクトログラムをそれぞれ観察することで減算する値を決定した．

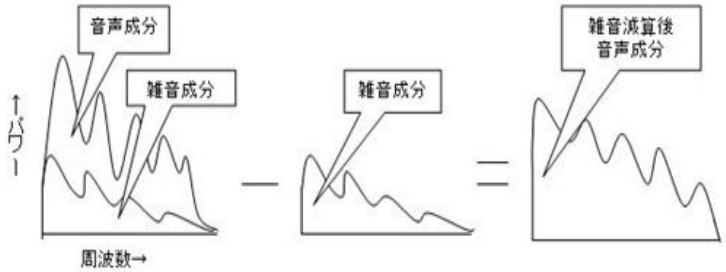

図 2　スペクトルサブトラクションのイメージ図

2.2 SAFIA

青木らは，複数のマイクを用いて特定の音源信号を取り出す手法として SAFIA（sound source Segregation based on estimating incident Angle of each Frequency component of Input signals Acquired by multiple microphones）を提案した[6]．SAFIA における信号の流れを図 3 に示す．目的音源 S_1，雑音源 S_2，マイク 1，マイク 2 がそれぞれ配置されており，目的音源はマイク 2 よりもマイク 1 の近くに配置されているとする．また，目的音と雑音は各々異なる調波構造を持っていると仮定する．マイク 1 の入力信号を $x_1(n)$，マイク 2 の入力信号を $x_2(n)$ とする．各々に離散フーリエ変換を施し，周波数 ω における周波数成分を $X_1(\omega)$ と $X_2(\omega)$ とする．

次に，式（1），式（2）に従って到達位相差 $\Delta\phi(\omega)$ 及び到達レベル差 $\Delta A(\omega)$ を求める．

$$\Delta\phi(\omega) = \arg(X_1(\omega)) - \arg(X_2(\omega)) \quad (1)$$

$$\Delta A(\omega) = 20\log_{10}\left(\frac{|X_1(\omega)|}{|X_2(\omega)|}\right) \quad (2)$$

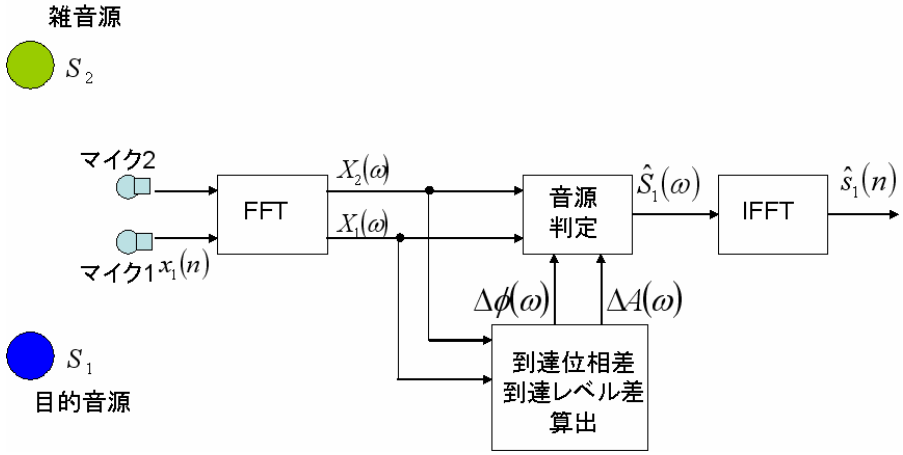

図 3　SAFIA における信号の流れ

目的音および雑音の調波構造がスパースであれば，多くの周波数成分において目的音と雑音いずれかのみが主成分となる．このため，各周波数では主成分となる単独の音源に基づくマイク間位相差及びレベル差が観測される．これらの値から，下記のように各周波数成分が目的音あるいは雑音のどちらに属するかを判定することが出来る．

図 3 のような配置においては，$X_1(\omega)$に含まれる目的音成分のレベルは，$X_2(\omega)$に含まれるものよりも大きく、位相も進んでいる．これにより判定部では，$\Delta\phi(\omega)$と$\Delta A(\omega)$が正である帯域は，目的音の周波数成分であると判定出来る．$\Delta A(\omega)$から判定する判定式を式（3）に，$\Delta A(\omega)$から判定する判定式を式（4）に示す．音源判定部分において，これらの判定により，目的音成分推定値 $\hat{S}_1(\omega)$と雑音成分推定値 $\hat{S}_2(\omega)$が算出される．

$$\left.\begin{aligned}\hat{S}_1(\omega)=X_1(\omega),\quad \hat{S}_2(\omega)=0,\quad (\Delta A(\omega)\geq 0)\\ \hat{S}_1(\omega)=0,\quad \hat{S}_2(\omega)=X_2(\omega),\quad (\Delta A(\omega)\leq 0)\end{aligned}\right\} \quad (3)$$

$$\left.\begin{aligned}\hat{S}_1(\omega)=X_1(\omega),\quad \hat{S}_2(\omega)=0,\quad (\Delta \phi(\omega)\geq 0)\\ \hat{S}_1(\omega)=0,\quad \hat{S}_2(\omega)=X_2(\omega),\quad (\Delta \phi(\omega)\leq 0)\end{aligned}\right\} \quad (4)$$

式（3），式（4）によって重み付けられた目的音成分推定値の各周波数成分 $\hat{S}_1(\omega)$に逆フーリエ変換を施し，時間領域の目的信号 $\hat{s}_1(n)$を復元する．このようにして SAFIA は音源の位置に基づく特徴量を用いて，特定領域内にある音源のみを抽出する．

判定に到達位相差と到達レベル差のどちらを用いるかは，マイク配置など入力系の特性に応じて選択する．本研究において用いた図 6 のマイク配置に対して事前検討を行い，今回は到達位相差を用いて方向判定を行う．

3. 方向判定実験

3.1 実験条件

実験には市販のトイロボットを使用した．このロボットの左肩と右肩と胸それぞれに無指向性のマイクを装着する（図 1）．図 4 にマイクの位置関係を示す．暗騒音が 36dBA である防音室において，音韻の出現頻度を考慮した 50 単語をロボットに対して 12 方向（図 5）（0°，30°，60°，90°，120°，150°，180°，210°，240°，270°，300°，330°）から 20 代前半の男性 1 名が発声した．この時，発話者とロボットとの距離は 1m程度，音声を発声する際に，ロボットを静止させ，駆動雑音が入らない状態での録音を全ての角度について行った．同じ条件で，ロボットの腕を駆動させ，駆動雑音のみの録音を行った．本報告では，広い範囲の S/N 比での検討を行うために，計算機中で音声データと雑音データを加算し，これを実験対象とする．雑音は，S/N 比 5dB～-10dB の範囲で可変し，雑音の増加により方向判定の精度がどのように変化するかを観測する．

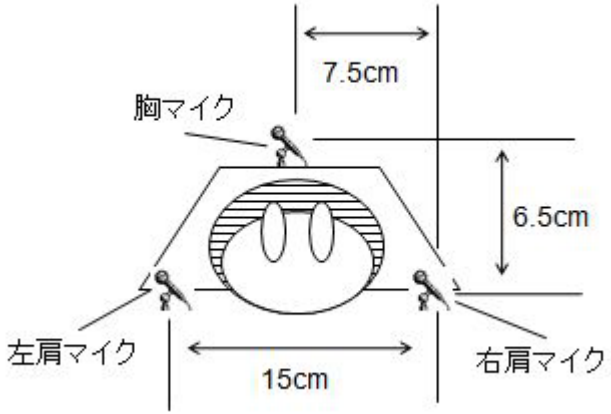

図 4　ロボットに取り付けたマイクの位置関係

Vol.2010-SLP-82 No.3
2010/7/23

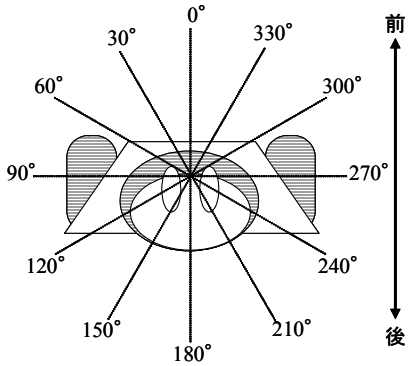

図 5　実験における音声の入射角度

3.2　マイク間時間差による話者方向判定

人間は，両耳に入る音声の時間差を音声の到来方向の判定に利用していると考えられている[7]．本実験では 3 つのマイクを使用するが，まずは 2 マイクでの仕組みを説明する．図 6 に示すように，2 つのマイクに入力された音声信号の相互相関関数を計算し，ピークを求める．相互相関関数は，時間差を持つ音声信号の類似度が最も高くなったときにピークをもつので，このピークがマイク間の時間差となる．マイク m の入力を $x_m(n)$（$n = 0,1,...,N$）とすると，マイク m-n 間の時間差であるピーク $p_{m,n}$ は式（5）によって求められる．

$$p_{m,n} = \arg\max_{\tau}\left\{\frac{1}{N-\tau}\sum_{i=0}^{N-\tau} x_m(i)\cdot x_n(i+\tau)\right\} \quad (5)$$

式（5）で学習データによって角度 θ におけるマイク m-n 間の時間差を求め、平均と分散を $\mu_{m,n}(\theta)$，$\sigma^2{}_{m,n}(\theta)$ とする．

判定したい音声の時間差を式（5）により求め $p_{m,n}$ とすると到来方向 θ_o は式（6）の分散正規化距離の最小化により求められる．

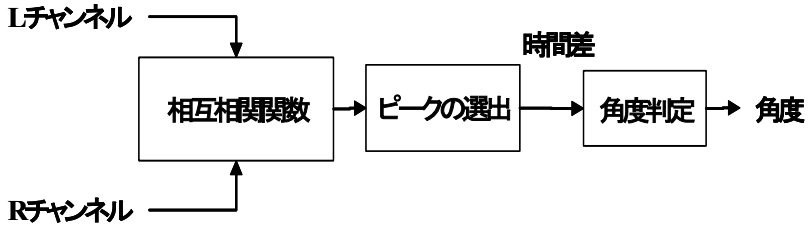

図 6　マイク間時間差による方向判定

$$\theta_o = \arg\min_{\theta}\sum_{m\neq n}\frac{\left(p_{m,n}-\mu_{m,n}\right)^2}{\sigma_{m,n}(\theta)^2} \quad (6)$$

3.3　スペクトルサブトラクションの導入

2 チャンネルによるスペクトルサブトラクションによる方向判定について図 7 に示す．まず，ロボットのマイクに入力される雑音混入音声に対してスペクトルサブトラクションにより雑音除去を行う．次に，雑音除去後の入力音声に対して，左，右，胸の 3 つのマイクから 2 マイクずつ（左肩，右肩），（左肩，胸），（右肩，胸）の計 3 つの組み合わせについて，チャンネル間の時間差をクロススペクトルに式（5）を適用することにより求め，式（6）から角度判定を行う．

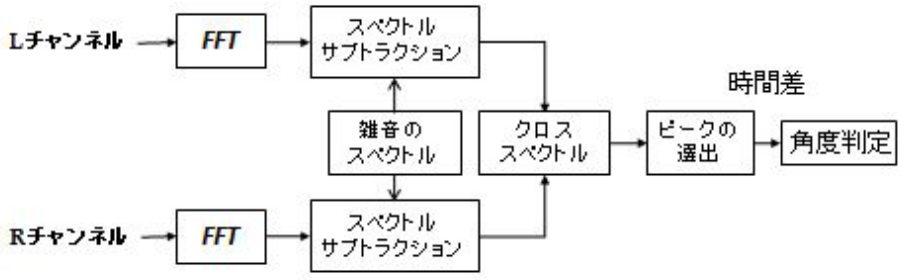

図 7　スペクトルサブトラクションによる角度判定

3.4 SAFIA の導入

SAFIA の典型的枠組みでは，目的音と雑音方向を既知とし，目的音と雑音の位置が対称となるように 2 つのマイクを設置し，各周波数の判別の閾値を 0 に設定していたが，本報告では目的音の方向が既知ではなく，またモーターやギアなど雑音源が複数存在するため，目的音と雑音が対称な位置関係にならない．そこで，目的音の到達位相差をあらかじめ学習させておき，実測音声の到達位相差との差異が小さければ目的音声，差異が大きければ雑音が混入していると判別する．マイクは 3 つを用いて各辺での到達位相差を組み合わせて方向判定を行う．

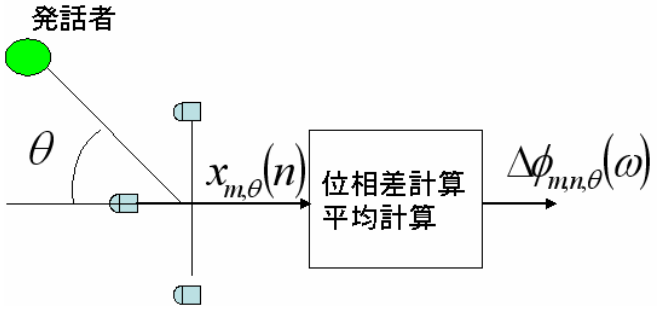

図 8　ある角度の音源に対する位相差と平均の事前測定

角度 θ からのマイク m への入力を $x_{m,\theta}$，フーリエ変換を行ったものを $X_{m,\theta}(\omega)$ とする．

マイク m とマイク n の到達位相差 $\Delta\phi_{m,n}(\omega)$ は式（7）で求める．

$$\Delta\phi_{m,n}(\omega)=\arg(X_{m,\theta}(\omega))-\arg(X_{n,\theta}(\omega)) \quad (7)$$

また，パワーが小さいときは，$\Delta\phi_{m,n}(\omega)$ は安定しないので計算しない．

角度 θ における学習データで $\Delta\phi_{m,n}(\omega)$ の平均を求め，それを $\Delta\phi_{\theta,m,n}(\omega)$ とする．

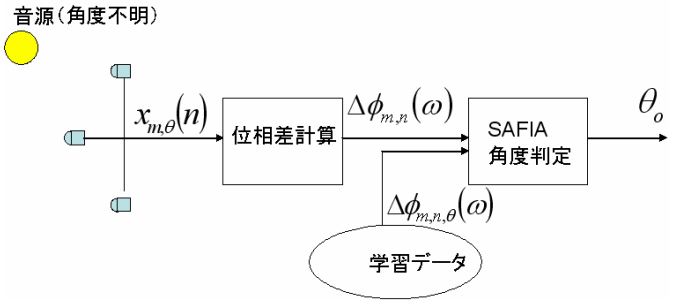

図 9　位相差による角度判定

そして，入力信号の到達位相差 $\Delta\phi_{m,n}(\omega)$ に対して $\Delta\phi_{\theta,m,n}(\omega)$ との二乗距離 $d_{\theta,m,n}$ を求め，最も距離の近くなる角度 θ を到来方向 θ_o とする．距離計算の際に，差が閾値より大きい場合には雑音の影響を受けている帯域と判断し計算しない．

$$\left.\begin{aligned} d_{\theta,m,n}(\omega)&=\left(\Delta\phi_{m,n}(\omega)-\Delta\phi_{\theta,m,n}(\omega)\right)^2, \quad \left(\left(\Delta\phi_{m,n}(\omega)-\Delta\phi_{\theta,m,n}(\omega)\right)^2\le\alpha\right)\\ d_{\theta,m,n}(\omega)&=0, \quad \left(\left(\Delta\phi_{m,n}(\omega)-\Delta\phi_{\theta,m,n}(\omega)\right)^2>\alpha\right)\end{aligned}\right\} \quad (8)$$

これを計算回数で正規化して

$$d_\theta=\frac{1}{C_\theta}\sum_{\omega}\sum_{m\neq n}d_{\theta,m,n}(\omega) \quad (9)$$

C_θ は計算が有効になった回数.

式（10）のように最も距離の少なかった角度を到来方向 θ_o とする．

$$\theta_o=\arg\min_{\theta} d_\theta \quad (10)$$

4. 実験結果と考察

表 1 に雑音除去手法による方向判定精度の違いをまとめた．横方向に S/N 比を変化させており，∞，4.8dB，-2.0dB，-5.0dB，-7.1dB，-9.1dB の 6 種類で検討した．方向の判定には図 6 に示すマイク間の時間差による手法を用いるが，組み合わせる雑音除去の手法として「なし」「SS（スペクトルサブトラクション）」「SAFIA」の 3 種類を縦方向に配置している．

まず，雑音のない場合（∞）を見ると，当然のことであるがどの方法でも 100%近い方向判定精度が達成できている．雑音のレベルが上がるにつれ雑音除去を行わない場合の精度は著しく劣化していることが分かる．次に，スペクトルサブトラクションを用いた場合の結果を見ると，残念なことにほとんど雑音除去の効果が方向判定精度に反映されていないことが分かる．スペクトルサブトラクション後のクロススペクトルを観察してみると，信号の位相が壊れていて雑音除去なしの場合と変わらないことが分かった．これは，スペクトルを減算する際に位相情報を壊しているためだと思われる．一方 SAFIA を用いた場合は，雑音の影響はあるものの，他の手法に比べて精度低下は緩やかである．これは，雑音の比重が大きくなっても，雑音の入ってない周波数帯域が存在し，その帯域をもとに方向判定を行っている結果だと考えられる．雑音が大きくなってくると，雑音に妨害される周波数帯域が広くなり，情報が少なくなるため精度が劣化するが，時間差による方向判定と違い，方向判定に使う情報がすべて雑音に妨害されないので，他の手法よりも精度が落ちないと思われる．表 2 に SAFIA による方向判定の混同表を示す．これらの結果より，提案法による方向判定は雑音の影響が少なく，優れていることが分かった．

表 1　雑音除去手法による方向判定精度の違い

SN比 / 雑音除去手法	∞	4.8dB	-2.0dB	-5.0dB	-7.1dB	-9.1dB
なし	99.8%	74.3%	35.3%	21.2%	14.6%	13.8%
SS		74.5%	34.5%	23.0%	12.8%	12.7%
SAFIA	100%	99.8%	99.5%	98.1%	93.1%	87.1%

表 2　SAFIA を用いた方向判定結果

IN \ OUT	0°	30°	60°	90°	120°	150°	180°	210°	240°	270°	300°	330°
0°	48	0	0	1	0	0	0	1	0	0	0	0
30°	1	48	0	0	0	0	0	0	0	1	0	0
60°	0	1	48	0	0	1	0	0	0	0	0	0
90°	0	1	0	49	0	0	0	0	0	0	0	0
120°	2	0	5	0	32	1	3	3	1	2	0	1
150°	4	0	2	0	0	38	0	1	1	0	4	0
180°	2	2	1	1	0	0	39	0	2	3	0	0
210°	2	1	2	0	0	1	1	42	0	0	1	0
240°	8	1	0	0	1	2	1	0	32	0	5	0
270°	0	0	0	1	0	0	0	0	0	49	0	0
300°	0	0	0	0	0	0	0	0	0	0	50	0
330°	0	0	1	0	0	0	1	0	0	0	0	48

5. あとがき

少数マイクによる話者方向判定の枠組みにおいて，雑音除去によって方向判定の精度向上を試みた．具体的には，雑音混入音声中から雑音スペクトルを減算するスペクトルサブトラクション法, 到達位相差により音源分離する SAFIA の有効性を検討した．比較実験の結果，スペクトルサブトラクションでは，精度向上は見られなかったが，SAFIA を用いる場合は，多少の精度劣化はあるものの，全体的に雑音に強いことが分かった．

参考文献

1) 菊池他，信学技報, DSP, (1999) 23-28
2) 浅野, "ロボットにおける音源位置推定"，日本音響学会誌, 63(1), (2007) 41-46
3) 藤原他，信学技報, 13-18 (2006-06)
1) 沼波，信学技報, SP, 2008 (134) 49-53
2) S.F. Boll, IEEE Trans. Acoust., Speech and Signal Processing., vol.ASSP-27, No.2, (1979) 113-120
3) Mariko Aoki et al, Acoust. Sci. & Tech. 22(2) 149-157 (2001)
4) Jeffress, L., J. Comp. Physiol Psychol. 41 (1948) 35-39