

非校正マイク群を用いた音源位置区別のための尺度

A Scale for Distinguishing the Positions of Sound Source with Non-Calibrated Microphones

小林智行 亀田能成 大田友一
Tomoyuki KOBAYASHI Yoshinari KAMEDA Yuichi OHTA

筑波大学 大学院 システム情報工学研究科
Graduate School of System and Information Engineering, University of Tsukuba

1. はじめに

生活環境中の人間の行動を音像により観測することを考える場合、行動に伴い発生する音の種類は声、物音、作業により発生する副次音など多様であり、音の大きさ、長さ、スペクトル分布など、その性質についても大きく異なっている。

また、生活環境中に多数のマイクロフォンをセンサとして配置し、人間行動の観測を行うことを考えると、マイク全てを均一に配置し、その全てについて校正を行うことは、用いるセンサの数が増えるにつれ現実的ではなくなる。

そこで、本稿では非均一に配置されたマイク群を非校正のまま利用し、そのもとで音源位置を区別するための尺度について検討を行う。

異なる位置での発音を区別していくことを考えた場合、同一位置での発音に対して特徴ベクトルが一意に求まる事が音源位置区別のために望ましい性質である。

2. 非校正マイク群を用いた音源位置の区別

本研究では、マイク複数台を一群のマイク群とし、音源から 2 台のマイクへの音の到達時間差 (Delay Of Arrival : DOA) の組を特徴ベクトルとして用いる。

音速が一定かつマイクが固定された状態であるならば、DOA は音源の位置にのみ依存する。マイク 1 ペアから求まる DOA からは、音源位置がマイクを焦点とする二葉双曲面の片葉上のどこかであることを推定出来る。このことから、3 ペアの DOA 組からは 3 次元的に一意な音源位置の推定を行うことが出来る。よって、3 台以上のマイク群を用いることで、同一位置での発音に対する特徴ベクトルは一意に定まると考えられる。

2.1 マイク間の到達時間差の推定

本稿では、白色化相互相関法 (Cross-Power Spectrum 法 : CSP 法) [1] を用いて 2 台のマイク i, j 間で DOA の推定を行う。CSP 法では、2 台のマイクの入力波形 $x_i(n), x_j(n)$ に対し、周波数領域でのマッチングを行い、類似度 $CSP_{i,j}(k)$ が最大となる遅延量 τ を推定された DOA とする。

CSP 法について以下に示す。

$$CSP_{i,j}(k) = IDFT \left[\frac{DFT [x_i(n)]^* DFT [x_j(n)]}{|DFT [x_i(n)]| |DFT [x_j(n)]|} \right]$$

$$\tau = \arg \max(CSP_{i,j}(k))$$

マイク間の位置関係が与えられるならば、推定された DOA から音源の方向を求めることが出来る。しかし、本研究ではマイクを均一に配置せず、また、マイクに対して校正を行わないため、上式によって推定された DOA のみを利用して音源位置の区別を行うことになる。

2.2 音源位置区別のための尺度の導入

観測により得られた特徴ベクトルから、音源位置の区別を行うための尺度について以下に述べる。

N 台のマイクを使用した場合、 $N C_2$ 組の DOA を得ることが出来る。このとき、特徴ベクトルは $N C_2$ 次元のベクトルになる。

観測により複数の特徴ベクトルが得られたとき、特徴ベクトル間の類似度の評価を行い、これを音源位置区別尺度とする。

本稿で用いた音源位置区別尺度 d について以下に示す。ここで a_{ij} は、ある位置で音が発生した時 (これを音像 A とする) のマイク i, j 間の DOA、 b_{ij} は別のある位置で音が発生した時 (これを音像 B とする) のマイク i, j 間の DOA とする。

$$d(A, B) = \sqrt{\sum_{i, j | 1 \leq i < j \leq N C_2} (a_{ij} - b_{ij})^2}$$

この音源位置区別尺度は音源の種類に依らず、2 つの音源位置が同じであれば必ず零になるという性質を持つ。他方で、本稿で提案した尺度については、シミュレーションによりある程度音源位置の区別を行うことが出来ることが確認されている。

現在の定義での検討課題は、新しい音像 C に対して、例え C が B に対して異なる位置に存在しても、 $d(A, B) = d(A, C)$ というように、異なる位置関係に対して同じ値を返してしまう可能性があることである。

そこで、この音源位置区別尺度が二音像間のユークリッド距離に対してどの程度比例するかについて、シミュレーションと実際の実験環境の両方において検証を行っていく予定である。

参考文献

- [1] 西浦, 山田, 中村, 鹿野: “マイクロホンアレーを用いた CSP 法に基づく複数音位置推定”, 電子情報通信学会論文誌, Vol.J83-D-II, No.8, 2000.