

1 チャンネルマイクロホンを用いた到来方向推定の音声信号への拡張

A-5 Direction-of-Arrival Estimation of Speech Signals Using a Single Channel Microphone

竹内 祐太[†] 村上 隆啓[†]

Yuta TAKEUCHI[†] Takahiro MURAKAMI[†]

[†] 明治大学理工学部電気電子生命学科

[†] Department of Electronics and Bioinformatics, Meiji University

1. はじめに

音響信号の到来方向推定は、2 チャンネル以上のマイクロホンを用いる手法が一般的である。しかし、多チャンネルのマイクロホンを用いる場合、推定に用いるシステムが大きくなり、またマイクロホンを複数用いるためコストがかかる。これらの問題を解決するために、本研究では 1 チャンネルのみのマイクロホンを用いて音響信号の到来方向を推定する。先行研究では、Swept-Sine 信号のように特殊な周波数スペクトルを持った音響信号について、1 チャンネルマイクロホンでの到来方向推定に成功している[1]。本研究では、これを一般的な音声信号に拡張する。

2. 1 チャンネルマイクロホンによる到来方向推定

1 チャンネルマイクロホンを用いた到来方向の推定では、マイクロホンの近くに設けた壁を利用する[1]。近くに壁がある場合、マイクロホンで観測される信号は、音源からの直接波と壁からの反射波が重なった信号となる。このとき、音響信号に含まれる周波数成分のうち以下の式で与えられる周波数 f を持つ成分は、直接波と反射波の位相差が半波長になり振幅が小さくなるため、この周波数は観測信号の周波数スペクトル上に現れるノッチ周波数となる[2]。

$$f = \frac{c}{4d \cos \theta} \quad (1)$$

ここで、 c は音速、 d は壁とマイクロホン間の距離、 θ は到来方向である。式(1)より、音響信号の到来方向 θ はノッチ周波数 f を用いて次式で表される。

$$\theta = \cos^{-1} \left(\frac{c}{4fd} \right) \quad (2)$$

この関係を利用して、観測信号の周波数スペクトルからノッチ周波数を検出することで、音響信号の到来方向を推定することができる[1]。文献[1]では、Swept-Sine 信号のようにパワースペクトルが一定となる音響信号の場合について、ノッチ周波数を検出して到来方向を推定することに成功している。一方、より一般的な音声信号を用いて到来方向を推定する場合、音声信号のパワースペクトルは一定ではないため、文献[1]の方法ではスペクトルからノッチ周波数が正しく推定されないため到来方向の推定精度が劣化する。そこで本研究では、文献[1]の方法を一般的な音声信号の場合に拡張する。

本研究では、一般的な音声信号では非定常性が強いいため時間平均すると白色雑音に近づき、周波数スペクトルも一定に近づくことを利用する。提案法では、観測信号を時間平均するために、比較的長い時間観測した信号を STFT (短時間フーリエ変換) によって時間周波数領域に変換してから、(1)得られた各フレームのパワースペクトルの平均を求め、(2)得られた各フレームの振幅スペクトルの平均を求め、(3)上記(1)において、STFT を行うときに Zero-Padding DFT (離散時間フーリエ変換) を用いる、(4)上記(2)において、STFT を行うときに Zero-Padding DFT を用いる、の 4 種類の方法を用いる。これらの方法によって得られたスペクトルから、文献[1]の方法によってノッチ周波数を検出し、式(2)によって到来方向を推定する。

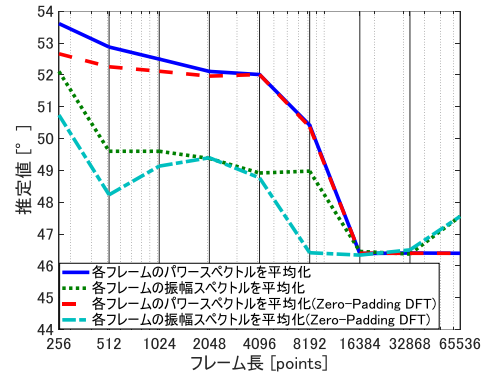


図 1 到来方向の推定値

3. 実験

提案法の有効性を確かめるために、実環境において 1 チャンネルマイクロホンの近傍に壁を設けたシステムを用いて様々な到来方向についてのインパルス応答を測定し、これらに音源信号を畳み込んで生成した観測信号に提案法を適用して到来方向の推定精度を調べた。音源信号は、男性が発話した 35 [s] の音声信号を用いた。サンプリング周波数を 48000 [Hz]、STFT におけるフレーム長を 256~65536 [points]、STFT におけるフレームシフトをフレーム長/4 [points]、STFT における Zero-Padding DFT の長さを 65536 [points]、窓関数をハニング窓、音源とマイクロホンの距離を 1 [m]、設置した壁とマイクロホンの距離を 1.5 [cm] とした。

図 1 に到来方向が 45 [°] の場合の 4 種類の方法による推定結果を示す。図 1 から、(4)の方法で STFT におけるフレーム長を 16384 [points] としたときに最も推定精度が良いことが分かった。他の到来方向や音源信号の場合でも、同様の結果が得られた。

4. まとめ

本研究では、1 チャンネルマイクロホンを用いた到来方向推定法の、一般的な音声信号への拡張を行った。提案法では、音声信号の非定常性を利用して、観測信号のスペクトルを時間方向で平均化することで到来方向に由来するノッチ周波数の検出を可能とした。実環境における実験によって、サンプリング周波数が 48000 [Hz] の場合、Zero-Padding DFT を用いた STFT によって得られた振幅スペクトルを平均化する方法で、STFT におけるフレーム長を 16384 [points]、STFT における Zero-Padding DFT の長さを 65536 [points] としたときに最も精度が良くなること分かった。

参考文献

- [1] 宇津貴之, 村上隆啓, “1 チャンネルマイクロホンによる音響信号の到来方向推定法”, 第 31 回信号処理シンポジウム講演論文集, pp.201-206, 2016.
- [2] V. C. Raykar, R. Duraiswami and B. Yegnanarayana, “Extracting the frequencies of the pinna spectral notches in measured head related impulse responses”, The Journal of the Acoustical Society of America, Vol. 118, pp. 364-374, 2005.