フルランク空間相関行列モデルに基づく拡散性雑音除去

Diffuse Noise Reduction using a Full-rank Spatial Covariance Model

礒佳樹 [1]，荒木章子 [2]，牧野昭二 [1]，中谷智広 [2]，澤田宏 [2]，山田武志 [1]，宮部滋樹 [1]，中村篤 [2]
Keiju Iso, Shoko Araki, Shoji Makino, Tomohiro Nakatani, Hiroshi Sawada, Takeshi Yamada, Shigeki Miyabe, Atsushi Nakamura
[1]筑波大学，[2]NTT 研究所
University of Tsukuba， NTT Communication Science Laboratories, NTT Corporation

1．はじめに

音源分離において，観測信号中に含まれる残響や雑音は，分離性能を低下させる．観測信号のみを用いて分離を行う方法をブラインド音源分離と呼ぶが，この手法は多くが無響や無雑音を仮定しているもので，残響や雑音環境下での性能は必ずしも高くない．本研究では，残響下の音源分離と拡散性雑音の除去を同時に行うことを目的として，まずは雑音除去の検討を行う．

本稿では，Duong らによる音源の空間特性を表すフルランク相関行列を用いた音源分離手法[1]を，雑音除去に拡張することを検討する．この手法は，空間相関行列の適切なモデル化[1]と適切な初期値を用いることで，残響環境下でも高い分離性能が得られることが知られている[2]．本稿では，雑音の空間特性を考慮しながら音源分離のためのパラメータ推定を行うことで，音源分離と同じ要領で拡散性雑音の除去ができることを報告する．

2．提案法

まず，観測信号 $\mathbf{x}$ は 1 人が発話した音源信号 s に，残響及び拡散性雑音が付加されたものを 2ch のマイクで観測したものとし，観測信号 $\mathbf{x}$ 及びその相関行列を以下のようにする．

$$\mathbf{x}(n,f)=s(n,f)\mathbf{h}(f)+\mathbf{N}(n,f) \tag{1}$$

$$\mathbf{R}_x(n,f)=\sigma_s(n,f)\mathbf{R}_s(f)+\sigma_n(n,f)\mathbf{R}_n(f) \tag{2}$$

ここで信号は時間周波数領域上で表し，n はフレーム番号，f は周波数帯を表す．$\sigma_s,\mathbf{R}_s$ はそれぞれ音声信号の分散及び空間相関行列であり，$\sigma_n,\mathbf{R}_n$ はそれぞれ拡散性雑音の分散及び空間相関行列である．ここで $\sigma_n,\mathbf{R}_n$ について拡散性雑音を想定し信号のモデル化を行う．

拡散性雑音が共分散行列 $\mathbf{R}_{\mathbf{c}_n}$ を持つ零平均の複素ガウス分布による確率変数であると仮定し，その共分散行列を以下のようにおく．

$$\mathbf{R}_{\mathbf{c}_n}(n,f)=\sigma_n(n,f)\mathbf{R}_n(f) \tag{3}$$

これらのモデルパラメータ及び推定信号を EM アルゴリズムによって更新するが，その初期値に関し，σ_n は観測信号の先頭フレームの無音区間の分散，$\mathbf{R}_n(f)$ は sinc 関数を用いた相関行列を用いて以下のようにする．

$$\sigma_n^{\text{init}}=\mathbf{x}(1,f) \tag{4}$$

$$\mathbf{R}_n(f)=\begin{bmatrix}1 & \operatorname{sinc}\left(\frac{2\pi fd}{c}\right)\\ \operatorname{sinc}\left(\frac{2\pi fd}{c}\right) & 1\end{bmatrix} \tag{5}$$

ここで $\mathbf{x}$ は観測信号，d は既知のマイク間距離であるとし，c は音速である．

これらの初期設定を用いて，文献[2]に基づいて EM アルゴリズムを用いてパラメータ $\sigma_s,\mathbf{R}_s,\sigma_n$ 及び推定信号を更新する．ただし，$\mathbf{R}_n(f)$ は(5)で固定する．

3．実験

提案法の有効性を検証するために，1 人が発話した音声信号に拡散性雑音を付加した 2ch 観測信号に対して雑音除去を行い，得られた推定信号を客観尺度によって評価した．雑音付加前の残響を含んだ音声信号に対して，雑音付加後の音声信号（観測信号），提案法で雑音除去を行った音声信号，観測信号に対してウィーナーフィルタにより雑音除去を行った音声信号のそれぞれの性能を比較した．なお，客観評価値には以下の式で得られる SN 比を使用する．

$$\mathrm{SNR}=10\log_{10}\frac{\sum_n\|\mathbf{c}(n)\|^2}{\sum_n\|\mathbf{c}(n)-\hat{\mathbf{c}}(n)\|^2} \tag{6}$$

ここで $\mathbf{c}(n)$ は残響を含んだ音声信号，$\hat{\mathbf{c}}(n)$ は観測信号に対する雑音除去によって推定された信号である．単位は dB であり，この尺度の数値が高いほど性能が良い．

実験条件として，音声データは約 8 秒，マイク間距離は 4cm で既知とする．また，サンプリング周波数は 8 kHz であり，FFT のサイズは 1024 点とし，オーバーラップを 256 点に設定した．提案法の反復回数は 40 回である．2 種類の音声及び 2 種類の残響時間に対する結果を表 1，表 2 に示す．

表 1：残響時間 130[ms]での SNR

アルゴリズム	音声 1	音声 2
観測信号	20.0	7.9
提案法	25.3	13.5
ウィーナーフィルタ	23.6	12.1

表 2：残響時間 250[ms]での SNR

アルゴリズム	音声 1	音声 2
観測信号	20.9	9.3
提案法	25.2	14.3
ウィーナーフィルタ	24.3	13.3

実験結果より，提案法の雑音除去が有効であることが分かった．

4．おわりに

本稿は，提案法による雑音除去の有効性を示した．今後は残響下の音源分離と拡散性雑音の除去を同時に行うことを目指す．

[1] N. Q. K. Duong, E. Vincent and R. Gribonval, "Under-determined reverberant audio source separation using a full-rank spatial covariance model", IEEE Trans. on ASLP, vol. 18, no. 7, pp. 1830-1840, Sept. 2010.
[2] K. Iso, S. Araki, S. Makino, T. Nakatani, H. Sawada, T. Yamada, and A. Nakamura, "Blind source separation of mixed speech in a high reverberation environment, " in Proc. HSCMA, pp.36-39, 2011.