尤度最大化に基づくエコー推定を用いたマルチスピーカ音響エコーキャンセラの検討*

○古賀健太郎, 滝口哲也, 有木康雄 (神戸大)

1 はじめに

カーナビの音声操作など，音楽が鳴っている環境での音声認識では，音楽雑音がマイクで観測されるため観測信号の SN が悪くなり，音声認識率が低下する．そこで，観測信号の SN を改善する音響エコーキャンセラが必要になる．

車内の場合，音楽雑音のチャンネル数は 2ch 以上で，マルチスピーカから出力される．先行研究 [1] では，2ch の参照信号を 1ch にしてエコー推定を行っているが，複数のエコーパスをまとめて扱うエコー推定ではキャンセル結果が十分に収束しないことがある．そこで，[2] において，マルチスピーカからマイクまでの各エコーパスを独立に推定するような，尤度最大化に基づくマルチスピーカ音響エコーキャンセラの検討を行った．

[2] では車室内で人の配置が異なる場合のみを想定していた．本稿では，車室内以外の，物の配置が異なる環境と温度が異なる環境の場合でシミュレーション実験を行い，SN が改善されることを示す．

2 マルチスピーカ音響エコーキャンセラのモデル

マルチスピーカ（スピーカ数：4）からの音楽雑音が 1ch マイクで観測される音響エコーキャンセラのモデルを Fig. 1 に示す．

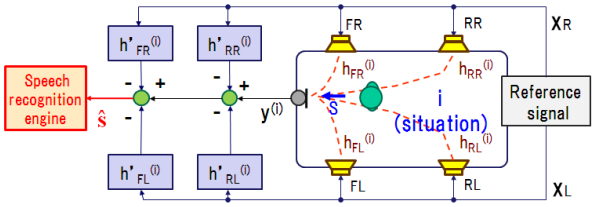

Fig. 1 マルチスピーカ音響エコーキャンセラの構成

環境 i において，マイクの時間領域における観測信号 $y^{(i)}$ は

$$y^{(i)} = s + N^{(i)} \tag{1}$$

と書ける．s は話者の音声である．$N^{(i)}$ は音響エコーで，

$$N^{(i)} = \sum x_L(h_{FL}^{(i)} + h_{RL}^{(i)}) + \sum x_R(h_{FR}^{(i)} + h_{RR}^{(i)}) \tag{2}$$

と書ける．x_L，x_R は 2ch の参照信号，$h_{FL}^{(i)}$，$h_{FR}^{(i)}$，$h_{RL}^{(i)}$，$h_{RR}^{(i)}$ はそれぞれの環境 i における各スピーカ $(FL,\ FR,\ RL,\ RR)$ からマイクまでの伝達特性である．このとき，推定すべき音響エコー $N'^{(i)}$ は

$$N'^{(i)} = \sum x_L(h'^{(i)}_{FL} + h'^{(i)}_{RL}) + \sum x_R(h'^{(i)}_{FR} + h'^{(i)}_{RR}) \tag{3}$$

と書ける．よって，話者のクリーン音声 $\hat{s}^{(i)}$ は，

$$\hat{s}^{(i)} = y^{(i)} - N'^{(i)} \tag{4}$$

となる．$\hat{s}^{(i)}$ において，目的とする音声 s を残すため，推定誤差を最小にするように $N'^{(i)}$ は推定されるべきである．

3 尤度最大化に基づくエコー推定を用いたマルチスピーカ音響エコーキャンセラ

文献 [1] や適応フィルタによるエコー推定では，音響エコー推定結果が十分に収束しない．そこで，マルチスピーカからマイクまでの各エコーパスを独立に推定することを考え，環境 i のエコーパスに対応した固定フィルタをあらかじめ準備する．環境 i は変化するので，推定したい数だけの環境 i $(i = 1, 2, \ldots, N)$ でインパルス応答を測定し [3]，各スピーカ-マイク間のエコーパスに対応した固定フィルタを $h'_{FL}{}^{(i)}$，$h'_{FR}{}^{(i)}$，$h'_{RL}{}^{(i)}$，$h'_{RR}{}^{(i)}$ $(i = 1, 2, \ldots, N)$ とする．

参照信号 x_L，x_R に，実環境 i $(i = 1, 2, \ldots, N)$ を推定した固定フィルタをそれぞれ畳み込み，観測信号 $y^{(i)}$ からキャンセルして，クリーン音声候補 $\hat{s}^{(1)}, \hat{s}^{(2)}, \ldots, \hat{s}^{(N)}$ を算出する．

クリーン音声候補の中から，尤度最大のクリーン音声候補を選択し，クリーン音声とする．$\hat{s}^{(1)}, \hat{s}^{(2)}, \ldots, \hat{s}^{(N)}$ から，音声の MFCC 特徴量 $\hat{S}^{(1)}_{MFCC}, \hat{S}^{(2)}_{MFCC}, \ldots, \hat{S}^{(N)}_{MFCC}$ を計算する．MFCC は音声データに対し FFT を行い，結果のパワー成分の対数を取った値を離散コサイン変換したものである．

MFCC 特徴 o の尤度 $P(o)$ は式 5 の通り，W 個の重みつき正規分布の和として求められる．正規分布の w 番目の平均は μ_w，分散は σ_w である．また，λ_w は，$\sum_1^W \lambda_w = 1$ となる重み係数である．

$$P(o) = \sum_{w=1}^{W} \lambda_w N(o; \mu_w, \sigma_w) \tag{5}$$

各話者の音響モデルを $\psi = \{\lambda, \mu, \sigma\}$ としたとき

$$\hat{i} = \arg\max_i P(\hat{S}^{(i)}_{MFCC} | \psi) \tag{6}$$

となる $\hat{i}$ を計算し，このときの $\hat{s}^{(\hat{i})}$ が尤度最大のクリーン音声候補である．全体の構成図を Fig. 2 に示す．

4 評価実験

物の配置が異なる環境と温度が異なる環境で，シミュレーション実験により学習同定法と提案手法の SN 改善効果を比較する．

4.1 実験条件

4.1.1 推定環境

物の配置がそれぞれ異なる環境 (8 通り) を Fig. 3 に示す．物の配置は同じで，温度がそれぞれ異なる環境 (3 通り) を Fig. 4 に示す．

4.1.2 観測信号とアルゴリズム

観測信号は，話者数が 5，話者ごとの文章の数が 20，周波数 16kHz である．音楽雑音は参照信号にインパルス応答を畳み込んだシミュレーション信号である．

アルゴリズムのうち，学習同定法の適応フィルタのタップ長は 1200，また，2ch 参照信号を足し合わせて 1ch にした参照信号を適応フィルタの入力とし

*Multi-loudspeaker acoustic canceller based on echo estimation with maximum likelihood, by KOGA Kentaro, TAKIGUCHI Tetsuya, ARIKI Yasuo (Kobe University)

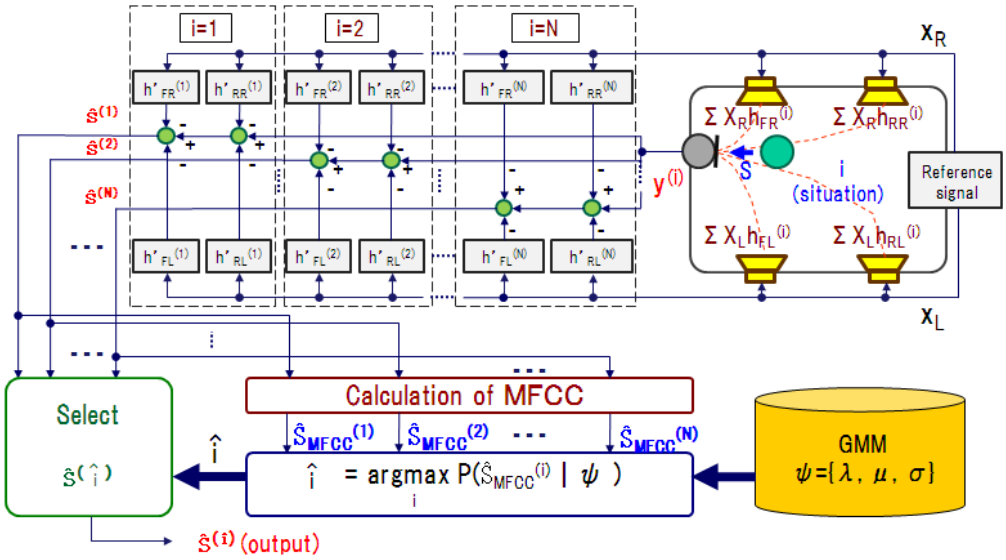

Fig. 2　尤度最大化基準を用いた車室内音響エコーキャンセラの構成図

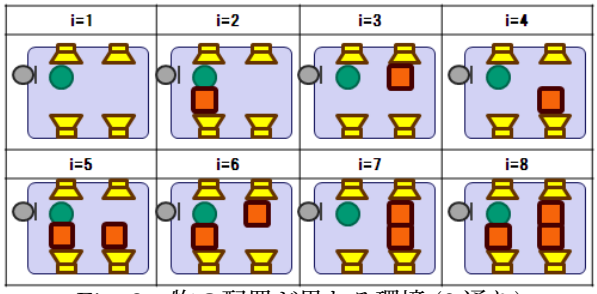

Fig. 3　物の配置が異なる環境 (8 通り)

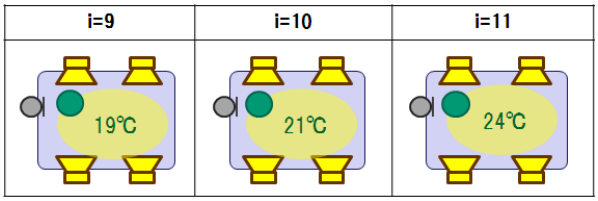

Fig. 4　温度が異なる環境 (3 通り)

ている．提案手法の固定フィルタのタップ長は 1200，GMM 学習に用いた話者数は 1(特定話者)，文章数は 20，混合数は 32，MFCC の次元数は 16，特徴抽出時のフレーム幅は 32，シフト幅は 8(ms) としている．

観測信号の環境と固定フィルタで推定する環境は，物の配置が異なる場合は $i = 1$〜8，温度が異なる場合は $i = 9$〜11 としている．

4.2　実験結果

実験結果を Fig. 5 に示す．

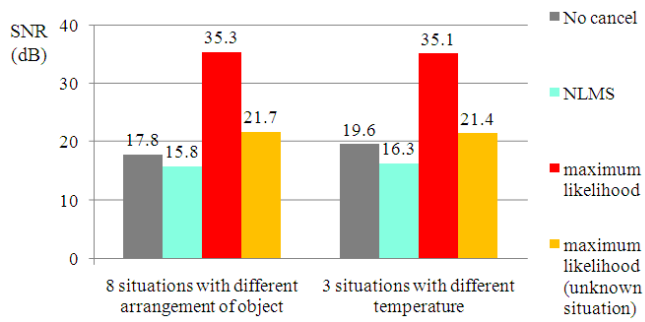

Fig. 5　実験結果

物の配置が異なる環境ではベースライン 17.8(dB) に対し，学習同定法によるキャンセラを用いた場合が 15.8(dB)，尤度最大化基準に基づくキャンセラを用いた場合が 35.3(dB)，温度が異なる環境ではベースライン 19.6(dB) に対し，学習同定法によるキャンセラを用いた場合が 16.3(dB)，尤度最大化基準に基づくキャンセラを用いた場合が 35.1(dB) と，いずれも提案手法の方が SN が改善される結果となった．

環境 o で観測した信号 $y^{(o)}$ に対して，同じ環境 o を推定した固定フィルタ $h'^{(o)}$ を用いてキャンセルしたクリーン音声候補 $\hat{s}^{(o)}$ を選択できる割合を，正しい環境の選択率と定義する．これは今回の実験では物の配置が異なる場合，温度が異なる場合の両方で，全ての話者で 100% となった．尤度最大化基準に基づく音響エコーキャンセラにおいて、観測信号の実環境と固定フィルタで推定した環境が一致していれば、キャンセルしない信号や従来の適応フィルタでキャンセルした信号と比べて、キャンセル後の信号の SN が向上するといえる．

参考として，提案手法で環境 o で観測した信号 $y^{(o)}$ に対して，同じ環境 o を推定した固定フィルタ $h'^{(o)}$ を用いなかった場合 (未知環境の場合) の SN を，Fig. 5 に併せて示す．物の配置が異なる環境では 21.7(dB)，温度が異なる環境 21.4(dB) と，同じ環境 o を推定した固定フィルタ $h'^{(o)}$ を用いている場合と比べて，SN 改善効果が小さい．

5　おわりに

本稿では，尤度最大化に基づく音響エコーキャンセラによって，車室内以外の，物の配置や温度が異なる環境においても従来の適応フィルタよりも観測信号の SN を改善できることをシミュレーション実験により示した．今後は，環境が未知の場合の SN 改善について検討していく．

参考文献

[1] S. Miyabe, T. Takatani, Y. Mori, H. Saruwatari, K. Shikano, and Y. Tatekura, "Double-Talk Free Spoken Dialogue Interface Combining Sound Field Control With Semi-Blind Source Separation", ICASSP 2006.
[2] 古賀，2008 春季研究発表会，3-P-6
[3] 佐藤，日本音響学会誌 58 巻 10 号，pp.669-676，2002.