

尤度最大化基準を用いたエコー推定に基づく 車室内音響エコーキャンセラの検討*

○古賀健太郎 (富士通テン, 神戸大), 滝口哲也, 有木康雄 (神戸大)

1 はじめに

車内向け音声認識システムでは, 車内で音楽がスピーカから出力されている状況下で音声認識を行うと, 音声認識マイクに認識対象の音声と音響エコーが混入し, 音声認識の妨げとなる. そこで, 音響エコーキャンセラによって音響エコーを除去し, SN を改善することによって認識率を確保する試みがなされている.

車室内の音響エコーはマルチスピーカから出力され, 単一マイクで観測される. このとき, 音響エコーの基となる参照信号は通常 2ch 以上から構成される. 先行研究 [1] では車室内の音響エコーを, 2ch の参照信号の和を入力としたフィルタで推定している. しかし, 車の内装を構成するシートや窓ガラス等によって車内の音の反射は複雑化しているため, 1ch にまとめた参照信号による音響エコー推定は正確では無く, キャンセル結果が十分に収束されないと考えられる. 本稿では, 車室内マルチスピーカ環境において, 2ch の参照信号を独立に用いて音響エコーを推定し, キャンセルする方法について検討する.

2 車室内音響エコーキャンセラのモデル

マルチスピーカから出力され, 単一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.

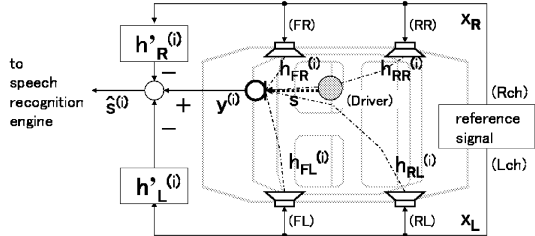


Fig. 1 車室内音響エコーキャンセラの構成

2ch の参照信号をそれぞれ x_L , x_R , 車内環境 i における各スピーカからマイクまでの伝達特性を $h_{FL}^{(i)}$, $h_{FR}^{(i)}$, $h_{RL}^{(i)}$, $h_{RR}^{(i)}$. ドライバの音声を s とすると, マイクの観測信号 $y^{(i)}$ は $y^{(i)} = s + N^{(i)}$ と書ける. ここで $N^{(i)}$ は音響エコーで, $N^{(i)} = \sum x_L(h_{FL}^{(i)} + h_{RL}^{(i)}) + \sum x_R(h_{FR}^{(i)} + h_{RR}^{(i)})$ となる. 音響エコーキャンセラにおいて, 推定すべき音響エコーを $N'^{(i)}$, FL, RL スピーカに対応する伝達特性フィルタを $h_L'^{(i)}$, FR, RR スピーカに対応する伝達特性フィルタを $h_R'^{(i)}$ とした場合, $N'^{(i)} = \sum x_L h_L'^{(i)} + \sum x_R h_R'^{(i)}$ となる. 求める結果 $\hat{s}^{(i)}$ は

$$\hat{s}^{(i)} = y^{(i)} - N'^{(i)} \quad (1)$$

となる. $\hat{s}^{(i)}$ において, 音声 s を残し音響エコーの推定誤差を小さくするために音響エコー推定 $N'^{(i)}$ の最適化が必要となるが, 学習同定法等の適応フィルタを用いると, 適応の基となるキャンセル結果が 1ch の

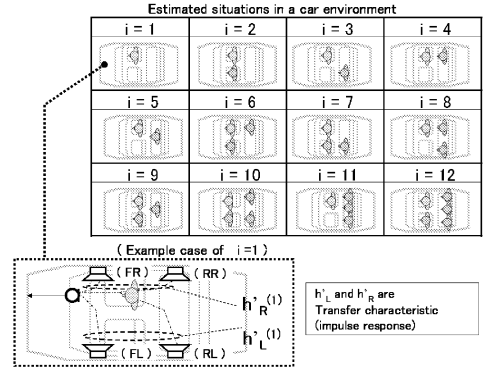


Fig. 2 12通りの伝達特性

ため全てのフィルタが同じ伝達特性に収束してしまう. そこで観測信号 $y^{(i)}$ に対し, 音響エコー推定結果 $N'^{(i)}$ を最適化するための伝達特性 $h_L'^{(i)}$, $h_R'^{(i)}$ を選択する仕組みが必要となる.

3 尤度最大化基準を用いた車室内音響エコーキャンセラ

車環境で想定できる全ての伝達特性に対して音響エコーを作成し, 観測信号から減算を行う. 実環境と想定環境の伝達特性が一致する場合, 音響エコーをキャンセルした結果にはクリーン音声のみ存在していると考えられるが, 実環境と想定環境の伝達誤差にミスマッチが存在する場合, キャンセル後の信号にはクリーン音声とエコー誤差信号が存在する. よって, キャンセル後の信号に対して, 音響モデルを用いて尤度計算を行い, 尤度最大化基準により最適な想定環境を選択することを考える.

3.1 伝達特性

伝達特性は車内のインパルス応答測定を元にして計算する [2]. 車内ではさまざまな人の配置, 物の配置が考えられるため, いろいろな状況を設定し, 状況ごとに伝達特性を計算する.

今回は Fig. 2 に示すように人員配置がそれぞれ異なる 12通りの車内状況を想定した. Fig. 2 中の状況 i において伝達特性を計算した結果のうち, 長さが T の部分を $h_L'^{(i)}$ (Lch スピーカ対応), $h_R'^{(i)}$ ($i = 1, 2, \dots, 12$) (Rch スピーカ対応) とする. 運転手不在の状況, 後部座席に 2 名いて片側に偏る状況は, 音声操作を行う車内状況として考えにくいと判断し, 除外している.

3.2 音声尤度の計算

音声 $\hat{s}^{(i)}$ に対し, MFCC (Mel Frequency Cepstrum Coefficient) を計算し, これを $\hat{S}_M^{(i)}$ とする. 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.

求めた音響特徴量 MFCC を用いて, 各話者毎の GMM (Gaussian Mixture Model) を学習する. 話者

* A study on acoustic echo canceler based on echo estimation with maximum likelihood in a car environment, by Kentaro Koga (Fujitsu TEN and Kobe University), Tetsuya Takiguchi, Yasuo Ariki (Kobe University)

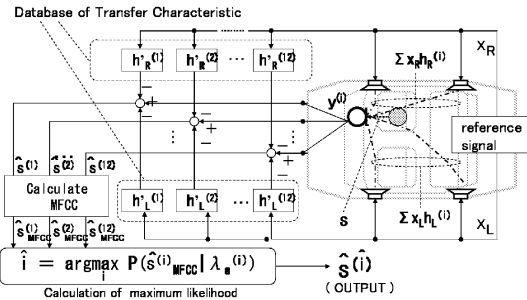


Fig. 3 尤度最大化基準を用いた車室内音響エコーキャンセラの構成図

のMFCC特徴を o とすると、音声尤度 $P(o)$ は式(2)の通り、 W 個の重みつき正規分布の和として求められる。 W 個の正規分布のうち w 番目の平均は μ_w 、分散は σ_w である。また、 λ_w は $\sum_{w=1}^W \lambda_w dw = 1$ となる重み係数である。

$$P(o) = \sum_{w=1}^W \lambda_w N(o; \mu, \sigma_w) \quad (2)$$

3.3 尤度最大化基準を用いた車室内音響エコーキャンセラ

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をFig. 3に示す。

ある車内環境 i にて観測されたマイクの観測信号を $y^{(i)} = s + \sum x_L h^{(i)} + \sum x_R h^{(i)} (= s + N^{(i)})$ とする。 $y^{(i)}$ に対し、あらゆる環境を想定し12通りの全ての伝達特性の組み合わせから音響エコー $N^{(1)}, N^{(2)}, \dots, N^{(12)}$ を作成する。次に、これらの想定された音響エコーを用いて観測信号からキャンセルした結果 $\hat{s}^{(1)}, \hat{s}^{(2)}, \dots, \hat{s}^{(12)}$ をそれぞれ算出する。

キャンセル結果 $\hat{s}^{(1)}, \hat{s}^{(2)}, \dots, \hat{s}^{(12)}$ から音声のMFCC特徴量 $\hat{S}_M^{(1)}, \hat{S}_M^{(2)}, \dots, \hat{S}_M^{(12)}$ を算出する。各話者の音響モデルを λ_s としたとき

$$\hat{i} = \arg \max_i P(\hat{s}^{(i)} | \lambda_s) \quad (3)$$

となる $\hat{i}$ を算出する。このときの $\hat{s}^{(\hat{i})}$ が音声尤度を最大とするキャンセル結果となる。

4 シミュレーション実験

参照信号 x に伝達特性を畳み込み、マイクの観測信号をシミュレーションした信号 $y^{(i)} (i = 1, 2, \dots, 12)$ を用いて、実験を行う。

畳み込んだ伝達特性は3.1で計算したインパルス信号をインパルス長 $N_I (N_I > T)$ 全て用いたものである。音響エコーキャンセラのフィルタと同じ伝達特性(長さ T)を用いているため、入力 $y^{(o)}$ に対し、尤度最大化基準によって $\hat{s}^{(o)}$ を選択、出力することができれば最もSNを改善できることになる。

シミュレーション観測信号に対し、学習同定法によるキャンセラと尤度最大化基準を用いたキャンセラによるキャンセル効果のSN値平均をFig. 4に示す。参考として、元の観測信号 $y^{(i)}$ のSN平均、尤度最大化キャンセラによる理想値(全ての $y^{(o)} (o = 1, 2, \dots, 12)$ に対し、出力が $\hat{s}^{(o)}$)の場合のSN平均を併記している。音響エコー推定のための参照信号は、学習同定法では元の2ch参照信号の和として、提案手法では3章の手法に基づき参照信号ごとに独立にエコー推定

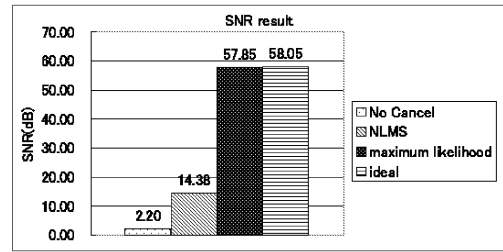


Fig. 4 実験結果

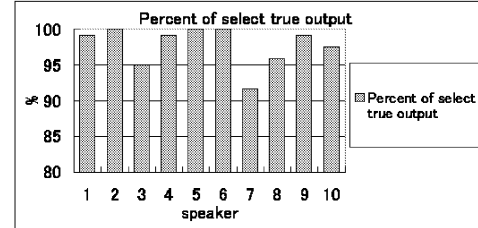


Fig. 5 伝達特性の選択率

を行っている。評価データの条件はTable. 1、アルゴリズムに関する条件はTable. 2の通りである。

Fig. 4に示す通り、尤度最大化を用いた音響エコーキャンセラによるキャンセル効果は学習同定法と比較して43.47(dB)の改善が見られる。このときの正しい結果の選択率(入力 $y^{(o)}$ に対し、 $\hat{s}^{(o)}$ を出力)はFig. 5の通りであり、尤度最大化基準により全ての話者に対して90%以上の確率で理想の解を選択できている。選択率が100%には至らない部分が理想の場合との0.20(dB)の差になっている。

Table 1 評価データの条件

話者	10名
文章	10文
信号の周波数	16kHz
インパルス長 N_I	1365
インパルスを測定した車種	will サイファ

Table 2 アルゴリズムに関する条件

フィルタのタップ長 T	1200
GMM学習に用いた話者数	1名(特定話者)
GMM学習の文章	50文
GMMの混合数	32
MFCC次元数	16
MFCC特徴抽出のフレーム幅	32ms
MFCC特徴抽出のシフト幅	8ms

5 おわりに

今回、2chの参照信号を音響エコー推定に使い、尤度最大化基準によりキャンセル結果を選定する音響エコーキャンセラを提案した。また、シミュレーション実験により、1chの参照信号を用いた適応フィルタよりも観測信号のSNを改善できることを示した。

今後は想定する車内状況のバリエーション増加、キャンセル結果を選択するアルゴリズムの効率化、実環境で収録した観測信号を用いた場合のキャンセル効果確認、車の種類を変えた場合のキャンセラ効果確認を行いたい。

参考文献

- [1] Abderrahman, pp.838-841, Interspeech 2007.
- [2] 佐藤, 日本音響学会誌 58巻 10号, pp.669-676, 2002.