社団法人　情報処理学会　研究報告
IPSJ SIG Technical Reports

2005－ＳＬＰ－59（26）
2005／12／22

AURORA2/CENSREC3による雑音抑圧手法の評価

李衛鋒†　　チャン・フィ・ダット†　　武田一哉†

† 名古屋大学情報科学研究科
〒464-8603 名古屋市千種区不老町 1
E-mail: †lee@sp.m.is.nagoya-u.ac.jp, ††hdtran@i2r.a-star.edu.sg, †††takeda@is.nagoya-u.ac.jp

キーワード　雑音抑圧，音声認識，音声コーパス，音響モデル，重回帰

Comparative evaluation of noise reduction methods using AURORA2/CENSREC3

Weifeng LI†, Tran HUY DAT†, and Kazuya TAKEDA†

† Graduate School of Information Science, Nagoya University
Furo-cho1, Chikusa-ku, Nagoya 464-8603, JAPAN
E-mail: †lee@sp.m.is.nagoya-u.ac.jp, ††hdtran@i2r.a-star.edu.sg, †††takeda@is.nagoya-u.ac.jp

1. スペクトルフィルタに基づく雑音抑圧法のAURORA2Jによる比較評価

本節では雑音重畳音声スペクトルと元音声のスペクトルの同時確率分布 $f_{\mathbf{S},\mathbf{X}}(\mathbf{s},\mathbf{x})$ を，雑音と音声それぞれの分布 $f_{\mathbf{S}}(\mathbf{s})$ と $f_{\mathbf{N}}(\mathbf{n})$ の畳み込みにより計算する手法を取り上げ，AURORA2Jにより評価する。

1.1 最大事後確率に基づく音声スペクトル推定

雑音重畳音声と元音声の振幅スペクトル，$\mathbf{X}$ と $\mathbf{S}$ の同時密度関数 $f_{\mathbf{X},\mathbf{S}}(\mathbf{x},\mathbf{s})$ が与えられれば，雑音重畳音声が与えられた下で元音声の最大事後確率推定 $\hat{\mathbf{s}}$ が，

$$\hat{\mathbf{s}} = \arg\max_{\mathbf{s}} f_{\mathbf{S}|\mathbf{X}}(\mathbf{s}|\mathbf{x}) = \arg\max_{\mathbf{s}} \frac{f_{\mathbf{S},\mathbf{X}}(\mathbf{s},\mathbf{x})}{f_{\mathbf{X}}(\mathbf{x})}$$
$$= \arg\max_{\mathbf{s}} f_{\mathbf{S},\mathbf{X}}(\mathbf{s},\mathbf{x})$$

により計算される。雑音スペクトルと音声スペクトルの密度関数 $f_{\mathbf{N}}(\mathbf{n})$ と $f_{\mathbf{S}}(\mathbf{s})$ が与えられれば，

$$f_{\mathbf{S},\mathbf{X}}(\mathbf{s},\mathbf{x}) = f_{\mathbf{X}|\mathbf{S}}(\mathbf{x}|\mathbf{s}) f_{\mathbf{S}}(\mathbf{s}) = f_{\mathbf{N}}(\mathbf{x}-\mathbf{s}) f_{\mathbf{S}}(\mathbf{s})$$

により同時確率が計算される。

Ephraim らは，雑音の振幅スペクトル $|\mathbf{N}|$ の分散が σ_N^2 で与えられ，複素スペクトル $\mathbf{N}$ の実部 N_r と虚部 N_i それぞれが平均零，分散 $\sigma_N^2/2$ の独立なガウス分布に従うと仮定することで，雑音混入音声の振幅スペクトル $|\mathbf{X}|$ の条件付密度関数は，$I_0(z)$ を変形ベッセル関数として，

$$f_{|\mathbf{X}||\mathbf{S}}(|\mathbf{x}||\mathbf{s}) = \frac{|\mathbf{x}|}{\pi\sigma_N^2} \exp\left(-\frac{|\mathbf{x}|^2+|\mathbf{s}|^2}{\sigma_N^2}\right) I_0\left(\frac{2|\mathbf{x}||\mathbf{s}|}{\sigma_N^2}\right)$$

により与えられることを示した。さらに Ephraim らは，上式を用いるとともに，元音声の事前分布にも雑音と同様のガウス分布関数を用いることで，最小二乗推定誤差基準を満たす音声スペクトル $|\hat{\mathbf{S}}|$ を与えた [1]。

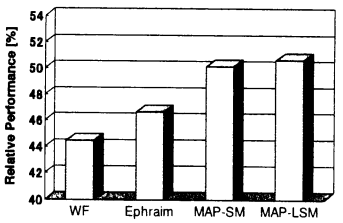

図 1　AURORA2J によるスペクトルフィルタ法の評価。WF: ウイナーフィルタ，Ephraim: Ephraim [1] の方法，MAP-SM，MPA-LSM:一般化ガンマ分布を事前確率に用い，振幅スペクトル，対数振幅スペクトルの領域で最大事後確率基準による推定を行った結果。

1.2 一般化と実装

筆者らは，元音声の事前分布に正規化されたパワースペクトル $\mathbf{S}^2/\sigma_{\mathbf{S}}^2$ の一般化ガンマ分布を用いることで，分布を逐次的

に更新するとともに，$I_0(z)$ の近似により，事後確率の最大化問題が2次方程式の求解に帰着できることを示した。さらに，パワースペクトル $|\mathbf{S}|^2$ や対数パワースペクトル $\log|\mathbf{S}|$ が，振幅スペクトルの非線形変換であることに着目し，その密度関数を，振幅スペクトルの密度関数から計算することで，パワースペクトルや対数パワースペクトルの領域での最大事後確率化を計算する方法を導いた [2]。

上記2つ一般化により，振幅スペクトル，対数パワースペクトルの領域それぞれで，元信号のスペクトルを最大事後確率で与える $|\hat{\mathbf{S}}|$ は，ゲイン関数 $G = |\mathbf{S}|/|\mathbf{X}|$ について，以下の2次方程式をそれぞれ，$A = 3, 1$ として解くことで得られる。

$$G^2 - \frac{G}{1 + a/\xi} + \frac{4a - A}{4\gamma(1 + a/\xi)} = 0$$

但し ξ と γ は，$\xi \triangleq \sigma_x^2/\sigma_N^2$，$\gamma \triangleq |\mathbf{X}|^2/\sigma_N^2$ で与えられる，事前 SNR，事後 SNR である。

1.3 評　価

上記の方法を AURORA2J コーパスを用いて評価した結果を図1に示す。Ephraim らによる最大事後確率に基づくスペクトル推定は，ウイナーフィルタの性能を 2%以上上回ってる。さらに，音声の事前分布と確率領域を一般化することで，50%を超える性能改善率が得られた。

2. 回帰による雑音抑圧法の CENSREC3 による評価

CENSREC3 の評価・学習データは，名古屋大学 CIAIR で収集した車内音声コーパスの一部である。当該コーパスは，CENSREC3 で利用されているバイザー位置に設置された遠隔マイクロホンの他，ヘッドセットの接話マイクと車内の5箇所に分散配置されたマイクロホンで受音された音声を含んでいる。本節では，ヘッドセットマイクで収録された音声を参照信号として，遠隔マイク音声のスペクトルを回帰する雑音抑圧法の評価結果を報告する [3], [4]。ただし本節で使用している評価データは，CENSREC3 の評価データのうち6名のみである。

2.1 同時録音コーパス

遠隔マイクと接話マイクで収録された音声を雑音重畳音声 $\mathbf{X}$，元音声と $\mathbf{S}$ と考えれば，上記のコーパスはいわゆる「同時録音コーパス」(parallel recording) であり，当該コーパスを利用することで，$f_{\mathbf{S},\mathbf{X}}(\mathbf{s},\mathbf{x})$ を直接推定することが可能である。重回帰分析は，上記の同時確率分布が与えられた下で，二乗誤差を最小とする写像を学習することと考えられる。

$$g(\mathbf{X})_{\mathrm{MMSE}} = \arg\min_g E\left[\{\mathbf{s} - g(\mathbf{X})\}^2\right]$$
$$\approx \arg\min_g \frac{1}{K}\sum_{k=0}^{K-1}\{\mathbf{s}_k - g(\mathbf{x}_k)\}^2$$

以下の実験では回帰システム $g(\cdot)$ として，3層のパーセプトロンを用いた。

2.2 実装と実験

CENSREC3 を用いた実験結果を図2に示す。同時録音を用いる回帰法は，単一入力の Ephraim の方法に比べ高い認識性能が得られる。分散・遠隔マイクを用いた方法は，適応的にマイクアレイの指向性を制御する複数マイク方式と比べてもより高い性能が得られており，回帰の計算量が適応アレイの計算量に比べて大幅に少ないことから，同時録音が可能であれば回帰法が有効な方法であることが分かる。また，推定雑音を利用する重回帰法の性能も，分布の重ね合わせに基づく方法に比べて優れた性能を実現している。

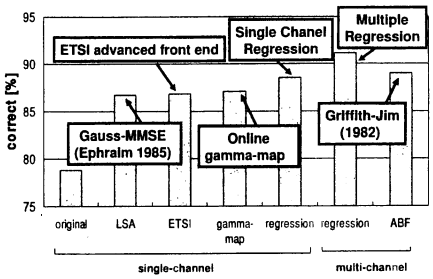

図 2　CENSREC3 による評価結果（15 走行条件下での 50 単語の平均識別率）。左から，バイザー位置で受音された遠隔音声 (original)，Ephraim らの方法 (LSA [1])，ウイナーフィルタの2段処理 (ETSI [5])，推定ノイズと雑音音声の非線形重回帰 (regression)，(これらは，バイザー位置のマイクで受音された音声のみを用いる。) 5本の分散マイクで収録された遠隔音声を入力として非線形重回帰 (regression)，バイザー位置に設置された4本のマイクを用いたリニアアレイによる適応ビームフォーマ (ABF [6])

謝辞 本研究は文科省リーディングプロジェクト「e-Society 基盤ソフトウエアの総合開発」及び，科学研究費補助金（基盤研究 A 15200014）の一部として行われた。

文　献

[1] Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean-square error log-spectral amplitude estimator," IEEE Trans. Acoustics, Speech and Signal Processing, vol.ASSP-33, no.2, pp.443-445, 1985.

[2] T.H.Dat, K.Takeda, F.Itakura, "Generalized Gamma modeling of speech and its online estimation for speech enhancement," Proc. ICASSP2005, SPTM-L8.4 2005

[3] W.Li, K.Itou, K.Takeda, and F.Itakura, "Optimizing regression for in-car speech recognition using multiple distributed microphones," Proc. ICSLP, pp.2689-2692, 2004.

[4] W.Li, K.Itou, K.Takeda, and F.Itakura, "Two-stage noise spectra estimation and regression based in-car speech recognition using single distant microphone," Proc. ICASSP2005, pp.I-533-536, 2005.

[5] "Speech processing, transmission and quality aspects (STQ); distributed speech recognition; advanced front-end feature extraction algorithm; compression algorithm," ETSI ES 202 050 v1.1.1, 2002.

[6] J. Griffiths and C. Jim, "An Alternative Approach to Linearly Constrained Adaptive Beamforming," IEEE Transactions on Antennas and Propagation, AP-30(1), pp. 27-34, Jan. 1982.