月4 2013 冬学期 [4830-1032]

第7回

音声音響信号処理

（統計的手法による音声強調）

亀岡弘和

東京大学大学院情報理工学系研究科
日本電信電話株式会社
NTTコミュニケーション科学基礎研究所

講義内容(キーワード)

- 信号処理、符号化、標準化の実用システム例の紹介
- 情報通信の基本(誤り検出、訂正符号、変調、IP)
- 符号化技術の基本(量子化、予測、変換、圧縮)
- 音声分析･合成･認識･強調、音楽信号処理
- 統計的信号処理の基礎(スペクトル、ガウス過程、最尤推定)
- ガウス性確率変数の基本性質
- 時間周波数分析(短時間フーリエ変換、ウェーブレット変換)
- ウィナーフィルタとカルマンフィルタ
- 音声生成過程のモデル(ソースフィルタ理論と藤崎モデル)
- 自己回帰モデルと線形予測分析
- 独立成分分析によるブラインド音源分離
- 非負値行列因子分解によるスペクトログラムの分解表現
- スペクトル間擬距離
- 最適化アルゴリズム(EMアルゴリズム、補助関数法)

講義スケジュール

10/ 7	守谷先生担当
10/15 (火)	守谷先生担当
10/21	守谷先生担当
10/28	（休講）
11/ 5 (火)	線形予測分析と自己回帰モデル
11/11	時間周波数解析
11/18	（休講）
11/25	非負値行列因子分解
12/ 2	**統計的手法による音声強調**
12/ 9	
12/16	吉井和佳 氏 (産業技術総合研究所)
1/15 (水)	
1/20	戸田智基 准教授 (奈良先端科学技術大学院大学)

成績評価

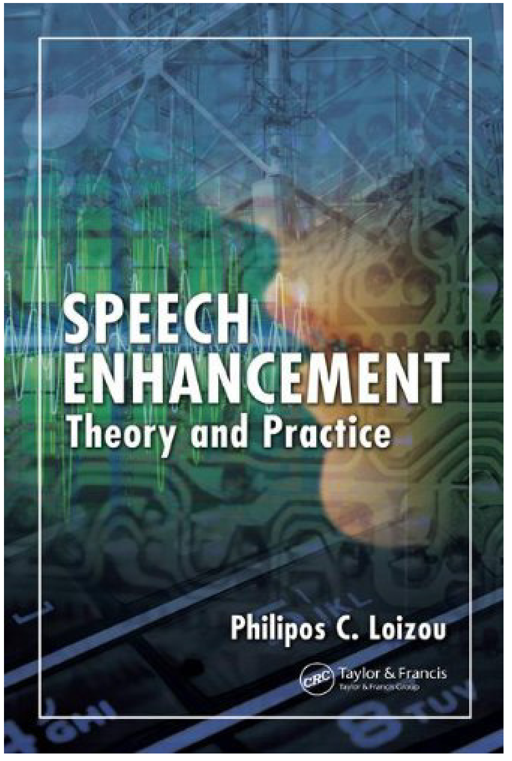
- レポート課題
  - 本講義に関連する論文を1つ選び、発表資料形式（パワーポイント等）にまとめて学期末に提出してください。提出先は最終講義にてお知らせします。
  - 「どの程度本質を理解しているか」「要点が分かりやすく記述されているか」「なぜその論文を重要と考えたか」を評価の規準にして採点します。
  - 毎回の講義後にその回の講義に関連する論文を1つ挙げる予定です。それらの中から選んでも良いですし、自分で自由に探してきてもOKです。
- 講義の感想
  - レポートとともに講義に対する感想文も一緒に提出して下さい。

※講義資料は講義用ホームページにアップしていく予定。

本日の話題

- 音声強調
  - 雑音や残響が重畳された観測信号から音声信号を強調したり抽出したりする技術
  - 音声アプリケーション（携帯電話，音声認識，航空通信，テレ会議システム，補聴器利用者による音声通信）において特に有用

- 参考文献
  - P. C. Loizou, *Speech Enhancement: Theory and practice*, CRC Press, 2007.



目次

- 音声強調問題の設定
- 解法の分類
  - 推定対象
    - 音声の複素スペクトル／振幅スペクトル
  - 推定量
    - 最尤(ML)推定量／最小平均二乗誤差(MMSE)推定量

> - 解法1: 複素スペクトルのMMSE推定量 [Wiener1949]
> - 解法2: 振幅スペクトルのML推定量 [McAulay1980]
> - 解法3: 振幅スペクトルのMMSE推定量 [Ephraim1984]

- 事前SN比（または音声パワースペクトル密度）の推定方法
  - 最尤法（パワー減算(PS)法）

目次

- 音声強調問題の設定
- 解法の分類
  - 推定対象
    - 音声の複素スペクトル／振幅スペクトル
  - 推定量
    - 最尤(ML)推定量／最小平均二乗誤差(MMSE)推定量

- 解法1: 複素スペクトルのMMSE推定量 [Wiener1949]
- 解法2: 振幅スペクトルのML推定量 [McAulay1980]
- 解法3: 振幅スペクトルのMMSE推定量 [Ephraim1984]

- 事前SN比（または音声パワースペクトル密度）の推定方法
  - 最尤法（パワー減算(PS)法）

問題設定

- **問題:**
  - $Y(\omega),\ X(\omega),\ N(\omega)$: 観測信号，音声信号，雑音信号の離散Fourier変換（複素スペクトル）
  - 雑音重畳音声

$$Y(\omega) = X(\omega) + N(\omega)$$

から音声に関係するパラメータ θ を推定したい

- **仮定:**
  - 音声 $X(\omega)$ と雑音 $N(\omega)$ は無相関
  - 音声 $X(\omega)$ は平均0の複素正規分布 $\mathcal{N}(0, \lambda_X(\omega))$ に従う
  - 雑音 $N(\omega)$ は平均0の複素正規分布 $\mathcal{N}(0, \lambda_N(\omega))$ に従う
  - 雑音パワースペクトル密度 $\lambda_N(\omega)$ は既知（例えば無音声区間から推定済みという状況を想定）

目次

- 音声強調問題の設定
- 解法の分類
  - 推定対象
    - 音声の複素スペクトル／振幅スペクトル
  - 推定量
    - 最尤(ML)推定量／最小平均二乗誤差(MMSE)推定量

• 解法1: 複素スペクトルのMMSE推定量 [Wiener1949]
• 解法2: 振幅スペクトルのML推定量 [McAulay1980]
• 解法3: 振幅スペクトルのMMSE推定量 [Ephraim1984]

- 事前SN比（または音声パワースペクトル密度）の推定方法
  - 最尤法（パワー減算(PS)法）

目次

- 音声強調問題の設定
- 解法の分類
  - 推定対象
    - 音声の複素スペクトル／振幅スペクトル
  - 推定量
    - 最尤(ML)推定量／最小平均二乗誤差(MMSE)推定量

> - 解法1: 複素スペクトルのMMSE推定量 [Wiener1949]
> - 解法2: 振幅スペクトルのML推定量 [McAulay1980]
> - 解法3: 振幅スペクトルのMMSE推定量 [Ephraim1984]

- 事前SN比（または音声パワースペクトル密度）の推定方法
  - 最尤法（パワー減算(PS)法）

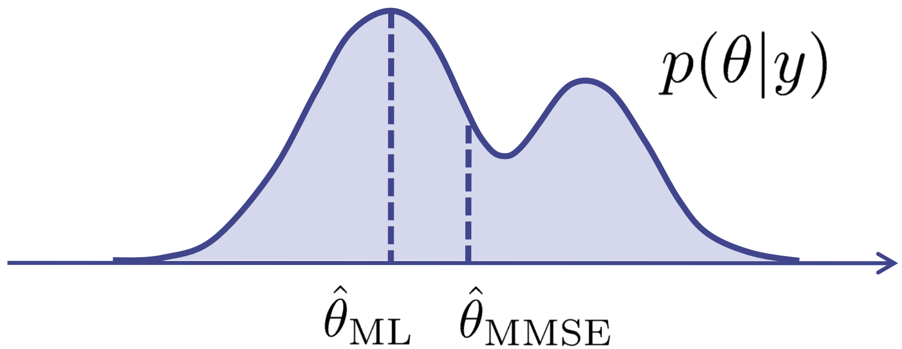
解法の分類

- θ の推定量
  - 最尤(Maximum Likelihood)推定量

$$\hat{\theta}_{\mathrm{ML}} = \underset{\theta}{\operatorname{argmax}}\, p(Y|\theta)$$

  - 最小平均二乗誤差(Minimum Mean Squared Error)推定量

$$\hat{\theta}_{\mathrm{MMSE}} = \underset{\hat{\theta}}{\operatorname{argmin}}\, \mathbb{E}\big[(\hat{\theta}-\theta)^2|Y\big] = \mathbb{E}\big[\theta|Y\big]$$



- 何を θ とするか？
  - 音声スペクトル $X(\omega)$
  - 音声スペクトル $X(\omega)$ の振幅 $A(\omega) := |X(\omega)|$

解法の分類

推定量 \ θ	$X(\omega)$	$A(\omega) := \lvert X(\omega)\rvert$
ML	—	[R.J. McAulay and M.L. Malpass (1980)]
MMSE	[N. Wiener (1949)]	[Y. Ephraim and D. Malah (1984)]

N. Wiener (1949), Extrapolation, interpolation and smoothing of stationary time series with engineering applications, Campbridge, MA: MIT Press.
R.J. McAulay and M.L. Malpass, “Speech enhancement using a soft-decision noise suppression filter,” IEEE Trans. on Acoust. Speech Signal Process., Vol. 28, No. 2, pp. 137-145, 1980.
Y. Ephraim and D. Malah, “Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator,” IEEE Trans. Acoust. Speech Signal Process., Vol. 32, No. 6, pp. 1109-1121, 1984.

目次

- 音声強調問題の設定
- 解法の分類
  - 推定対象
    - 音声の複素スペクトル／振幅スペクトル
  - 推定量
    - 最尤(ML)推定量／最小平均二乗誤差(MMSE)推定量

> - 解法1: 複素スペクトルのMMSE推定量 [Wiener1949]
> - 解法2: 振幅スペクトルのML推定量 [McAulay1980]
> - 解法3: 振幅スペクトルのMMSE推定量 [Ephraim1984]

- 事前SN比（または音声パワースペクトル密度）の推定方法
  - 最尤法（パワー減算(PS)法）

目次

- 音声強調問題の設定
- 解法の分類
  - 推定対象
    - 音声の複素スペクトル／振幅スペクトル
  - 推定量
    - 最尤(ML)推定量／最小平均二乗誤差(MMSE)推定量

- 解法1: 複素スペクトルのMMSE推定量 [Wiener1949]
- 解法2: 振幅スペクトルのML推定量 [McAulay1980]
- 解法3: 振幅スペクトルのMMSE推定量 [Ephraim1984]

- 事前SN比（または音声パワースペクトル密度）の推定方法
  - 最尤法（パワー減算(PS)法）

解法の分類

推定量 \ θ	$X(\omega)$	$A(\omega) := \lvert X(\omega)\rvert$
ML	—	[R.J. McAulay and M.L. Malpass (1980)]
MMSE	[N. Wiener (1949)]	[Y. Ephraim and D. Malah (1984)]

N. Wiener (1949), Extrapolation, interpolation and smoothing of stationary time series with engineering applications, Campbridge, MA: MIT Press.
R.J. McAulay and M.L. Malpass, “Speech enhancement using a soft-decision noise suppression filter,” IEEE Trans. on Acoust. Speech Signal Process., Vol. 28, No. 2, pp. 137-145, 1980.
Y. Ephraim and D. Malah, “Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator,” IEEE Trans. Acoust. Speech Signal Process., Vol. 32, No. 6, pp. 1109-1121, 1984.

解法の分類

推定量 \ θ	$X(\omega)$	$A(\omega) := \lvert X(\omega) \rvert$
ML	—	[R.J. McAulay and M.L. Malpass (1980)]
MMSE	[N. Wiener (1949)]	[Y. Ephraim and D. Malah (1984)]

N. Wiener (1949), Extrapolation, interpolation and smoothing of stationary time series with engineering applications, Campbridge, MA: MIT Press.

R.J. McAulay and M.L. Malpass, “Speech enhancement using a soft-decision noise suppression filter,” IEEE Trans. on Acoust. Speech Signal Process., Vol. 28, No. 2, pp. 137-145, 1980.

Y. Ephraim and D. Malah, “Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator,” IEEE Trans. Acoust. Speech Signal Process., Vol. 32, No. 6, pp. 1109-1121, 1984.

$X(\omega)$ のMMSE推定量 (1/2)

- 問題設定：
  - $Y(\omega) = X(\omega) + N(\omega)$
  - $N(\omega) \sim \mathcal{N}_{\mathbb{C}}(N(\omega); 0, \lambda_N(\omega)), \quad X(\omega) \sim \mathcal{N}_{\mathbb{C}}(X(\omega); 0, \lambda_X(\omega))$
- 求めたいのは $\hat{X}_{\mathrm{MMSE}}(\omega) = \mathbb{E}[X(\omega)|Y(\omega)]$

多変量Gauss分布の性質

$$p(y, x) = \mathcal{N}\left(\begin{bmatrix} y \\ x \end{bmatrix}; \begin{bmatrix} \bar{y} \\ \bar{x} \end{bmatrix}, \begin{bmatrix} \Sigma_{yy} & \Sigma_{yx} \\ \Sigma_{xy} & \Sigma_{xx} \end{bmatrix}\right) \longrightarrow \mathbb{E}[x|y] = \bar{x} + \Sigma_{xy}\Sigma_{yy}^{+}(y - \bar{y})$$

$$p(Y(\omega), X(\omega)) = \mathcal{N}_{\mathbb{C}}\left(\begin{bmatrix} Y(\omega) \\ X(\omega) \end{bmatrix}; \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \lambda_X(\omega) + \lambda_N(\omega) & \lambda_X(\omega) \\ \lambda_X(\omega) & \lambda_X(\omega) \end{bmatrix}\right)$$

$$\longrightarrow \mathbb{E}[X(\omega)|Y(\omega)] = \frac{\lambda_X(\omega)}{\lambda_X(\omega) + \lambda_N(\omega)} Y(\omega)$$ （Wienerフィルタ）

$X(\omega)$ のMMSE推定量 (2/2)

- WienerフィルタはX $X(\omega)$ のMMSE推定量

$$\hat{X}_{\mathrm{MMSE}}(\omega) = \underline{G_{\mathrm{Wiener}}(\omega)} Y(\omega)$$

$$\begin{aligned} G_{\mathrm{Wiener}}(\omega) &= \frac{\lambda_X(\omega)}{\lambda_X(\omega) + \lambda_N(\omega)} \\ &= \frac{\xi(\omega)}{\xi(\omega) + 1} \end{aligned}$$

$$\xi(\omega) = \frac{\lambda_X(\omega)}{\lambda_N(\omega)}$$

事前SN比

解法の分類

推定量 \ θ	$X(\omega)$	$A(\omega) := \lvert X(\omega) \rvert$
ML	—	[R.J. McAulay and M.L. Malpass (1980)]
MMSE	[N. Wiener (1949)]	[Y. Ephraim and D. Malah (1984)]

N. Wiener (1949), Extrapolation, interpolation and smoothing of stationary time series with engineering applications, Campbridge, MA: MIT Press.
R.J. McAulay and M.L. Malpass, “Speech enhancement using a soft-decision noise suppression filter,” IEEE Trans. on Acoust. Speech Signal Process., Vol. 28, No. 2, pp. 137-145, 1980.
Y. Ephraim and D. Malah, “Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator,” IEEE Trans. Acoust. Speech Signal Process., Vol. 32, No. 6, pp. 1109-1121, 1984.

解法の分類

推定量 \ θ	$X(\omega)$	$A(\omega) := \lvert X(\omega) \rvert$
ML	—	[R.J. McAulay and M.L. Malpass (1980)]
MMSE	[N. Wiener (1949)]	[Y. Ephraim and D. Malah (1984)]

N. Wiener (1949), Extrapolation, interpolation and smoothing of stationary time series with engineering applications, Campbridge, MA: MIT Press.
R.J. McAulay and M.L. Malpass, "Speech enhancement using a soft-decision noise suppression filter," IEEE Trans. on Acoust. Speech Signal Process., Vol. 28, No. 2, pp. 137-145, 1980.
Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator," IEEE Trans. Acoust. Speech Signal Process., Vol. 32, No. 6, pp. 1109-1121, 1984.

$A(\omega)$ のML推定量 (1/4)

- 問題設定：
  - $Y(\omega) = A(\omega)e^{\mathrm{j}\phi(\omega)} + N(\omega)$
  - $N(\omega) \sim \mathcal{N}_{\mathbb{C}}(N(\omega); 0, \lambda_N(\omega)), \quad \phi(\omega) \sim \mathrm{Uniform}[0, 2\pi)$
- 求めたいのは $\hat{A}_{\mathrm{ML}}(\omega) = \underset{A(\omega)}{\mathrm{argmax}}\, \underline{p(Y(\omega)|A(\omega))}$

$$
\begin{aligned}
&\underline{p(Y(\omega)|A(\omega), \phi(\omega))} \\
&\quad = \mathcal{N}_{\mathbb{C}}(Y(\omega); A(\omega)e^{\mathrm{j}\phi(\omega)}, \lambda_N(\omega)) \\
&\quad = \frac{1}{\pi\lambda_N(\omega)} \exp\left(-\frac{|Y(\omega) - A(\omega)e^{\mathrm{j}\phi(\omega)}|^2}{\lambda_N(\omega)}\right) \\
&\quad = \frac{1}{\pi\lambda_N(\omega)} \exp\left(-\frac{|Y(\omega)|^2 - 2A(\omega)\mathrm{Re}\{e^{-\mathrm{j}\phi(\omega)}Y(\omega)\} + A(\omega)^2}{\lambda_N(\omega)}\right)
\end{aligned}
$$

$$
\underline{p(Y(\omega)|A(\omega))} = \frac{1}{2\pi}\int_0^{2\pi} \underline{p(Y(\omega)|A(\omega), \phi(\omega))}\mathrm{d}\phi(\omega)
$$

$A(\omega)$ のML推定量 (2/4)

- $\phi(\omega)$ の周辺化：

$$
\begin{aligned}
&\frac{1}{2\pi}\int \underline{p(Y(\omega)|A(\omega),\phi(\omega))}\mathrm{d}\phi(\omega)\\
&=\frac{1}{\pi\lambda_N(\omega)}\int_0^{2\pi}\exp\left(-\frac{|Y(\omega)|^2-2A(\omega)\mathrm{Re}\{e^{-\mathrm{j}\phi(\omega)}Y(\omega)\}+A(\omega)^2}{\lambda_N(\omega)}\right)\mathrm{d}\phi(\omega)\\
&=\frac{1}{\pi\lambda_N(\omega)}\exp\left(-\frac{|Y(\omega)|^2+A(\omega)^2}{\lambda_N(\omega)}\right)\\
&\qquad\cdot\underline{\frac{1}{2\pi}\int_0^{2\pi}\exp\left(\frac{2A(\omega)\mathrm{Re}\{e^{-\mathrm{j}\phi(\omega)}Y(\omega)\}}{\lambda_N(\omega)}\right)\mathrm{d}\phi(\omega)}
\end{aligned}
$$

$$
\rightarrow \underline{I_0\left(\frac{2A(\omega)|Y(\omega)|}{\lambda_N(\omega)}\right)}
$$

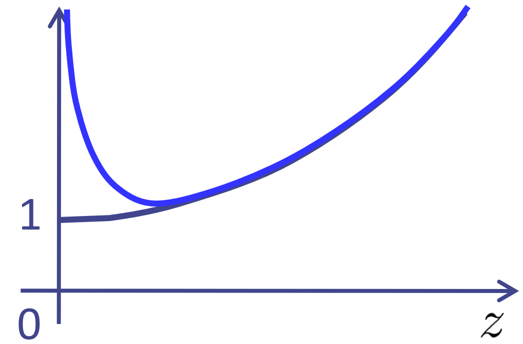
第1種変形Bessel関数：

$$
I_0(|z|)=\frac{1}{2\pi}\int_0^{2\pi}\exp\left(\mathrm{Re}\{ze^{-\mathrm{j}\phi}\}\right)\mathrm{d}\phi
$$

$A(\omega)$ のML推定量 (3/4)

第一種変形Bessel関数の近似：

$$I_0(|z|) \simeq \frac{1}{\sqrt{2\pi|z|}} \exp(|z|)$$



$$I_0\left(\frac{2A(\omega)|Y(\omega)|}{\lambda_N(\omega)}\right) \simeq \frac{1}{\sqrt{2\pi\frac{2A(\omega)|Y(\omega)|}{\lambda_N(\omega)}}} \exp\left(\frac{2A(\omega)|Y(\omega)|}{\lambda_N(\omega)}\right)$$

- $p(Y(\omega)|A(\omega))$

$$\simeq \frac{1}{\pi\lambda_N(\omega)} \frac{1}{\sqrt{2\pi\frac{2A(\omega)|Y(\omega)|}{\lambda_N(\omega)}}} \exp\left(-\frac{|Y(\omega)|^2 + A(\omega)^2 - 2A(\omega)|Y(\omega)|}{\lambda_N(\omega)}\right)$$

- 求めたかったのは $\hat{A}_{\mathrm{ML}}(\omega) = \underset{A(\omega)}{\operatorname{argmax}}\, p(Y(\omega)|A(\omega))$

$$\frac{\partial \log p(Y(\omega)|A(\omega))}{\partial A(\omega)} = 0 \quad \longrightarrow \quad \hat{A}_{\mathrm{ML}}(\omega) = \frac{1}{2}\left(|Y(\omega)| + \sqrt{|Y(\omega)|^2 - \lambda_N(\omega)}\right)$$

$A(\omega)$ のML推定量 (4/4)

- 最尤の$A(\omega)$ と観測位相$Y(\omega)/|Y(\omega)|$ により$X(\omega)$ の推定値を得る

$$\hat{X}_{\mathrm{McAulay}}(\omega) = \boxed{\hat{A}_{\mathrm{ML}}(\omega)\frac{Y(\omega)}{|Y(\omega)|}} = \underline{G_{\mathrm{McAulay}}(\omega)}Y(\omega)$$

$$\underline{G_{\mathrm{McAulay}}(\omega)} = \frac{\hat{A}_{\mathrm{ML}}(\omega)}{|Y(\omega)|}$$

$$= \frac{1}{2} + \frac{1}{2}\sqrt{\frac{|Y(\omega)|^2 - \lambda_N(\omega)}{|Y(\omega)|^2}}$$

$$= \frac{1}{2} + \frac{1}{2}\sqrt{\frac{\gamma(\omega) - 1}{\gamma(\omega)}}$$

$$\gamma(\omega) = \frac{|Y(\omega)|^2}{\lambda_N(\omega)}$$

事後SN比

解法の分類

推定量 \ θ	$X(\omega)$	$A(\omega) := \lvert X(\omega)\rvert$
ML	—	[R.J. McAulay and M.L. Malpass (1980)]
MMSE	[N. Wiener (1949)]	[Y. Ephraim and D. Malah (1984)]

N. Wiener (1949), Extrapolation, interpolation and smoothing of stationary time series with engineering applications, Campbridge, MA: MIT Press.
R.J. McAulay and M.L. Malpass, "Speech enhancement using a soft-decision noise suppression filter," IEEE Trans. on Acoust. Speech Signal Process., Vol. 28, No. 2, pp. 137-145, 1980.
Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator," IEEE Trans. Acoust. Speech Signal Process., Vol. 32, No. 6, pp. 1109-1121, 1984.

解法の分類

推定量 \ θ	$X(\omega)$	$A(\omega) := \|X(\omega)\|$
ML	—	[R.J. McAulay and M.L. Malpass (1980)]
MMSE	[N. Wiener (1949)]	[Y. Ephraim and D. Malah (1984)]

N. Wiener (1949), Extrapolation, interpolation and smoothing of stationary time series with engineering applications, Campbridge, MA: MIT Press.
R.J. McAulay and M.L. Malpass, "Speech enhancement using a soft-decision noise suppression filter," IEEE Trans. on Acoust. Speech Signal Process., Vol. 28, No. 2, pp. 137-145, 1980.
Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator," IEEE Trans. Acoust. Speech Signal Process., Vol. 32, No. 6, pp. 1109-1121, 1984.

$A(\omega)$ のMMSE推定量 (1/4)

- 問題設定:
  - $Y(\omega) = X(\omega) + N(\omega), \quad X(\omega) = A(\omega)e^{\mathrm{j}\phi(\omega)}$
  - $N(\omega) \sim \mathcal{N}_{\mathbb{C}}(N(\omega); 0, \lambda_N(\omega)), \quad X(\omega) \sim \mathcal{N}_{\mathbb{C}}(X(\omega); 0, \lambda_X(\omega))$

- 求めたいのは $\hat{A}_{\mathrm{MMSE}}(\omega) = \mathbb{E}[A(\omega)|Y(\omega)]$

$$p(A(\omega)|Y(\omega)) = \frac{p(Y(\omega)|A(\omega))p(A(\omega))}{p(Y(\omega))}$$
$$= \frac{p(Y(\omega)|A(\omega))p(A(\omega))}{\int_0^\infty p(Y(\omega)|A(\omega))p(A(\omega))\mathrm{d}A(\omega)}$$

$$\hat{A}_{\mathrm{MMSE}}(\omega) = \int_0^\infty A(\omega)p(A(\omega)|Y(\omega))\mathrm{d}A(\omega)$$
$$= \frac{\int_0^\infty A(\omega)\underline{p(Y(\omega)|A(\omega))p(A(\omega))}\mathrm{d}A(\omega)}{\int_0^\infty \underline{p(Y(\omega)|A(\omega))p(A(\omega))}\mathrm{d}A(\omega)}$$

$A(\omega)$ のMMSE推定量 (2/4)

- $$\underline{p(Y(\omega)|A(\omega))p(A(\omega))} = \int_0^{2\pi} \underline{p(Y(\omega)|A(\omega),\phi(\omega))}\,\underline{p(A(\omega),\phi(\omega))}\mathrm{d}\phi$$

$$\underline{p(Y(\omega)|A(\omega),\phi(\omega))} = \mathcal{N}_{\mathbb{C}}(Y(\omega); A(\omega)e^{\mathrm{j}\phi(\omega)}, \lambda_N(\omega))$$

$$\underline{p(A(\omega),\phi(\omega))} = \mathcal{N}_{\mathbb{C}}(A(\omega)e^{\mathrm{j}\phi(\omega)}; 0, \lambda_X(\omega))$$

$$\hat{A}_{\mathrm{MMSE}}(\omega) =$$

$$\frac{\int_0^{\infty}\int_0^{2\pi} A(\omega)\mathcal{N}_{\mathbb{C}}(Y(\omega); A(\omega)e^{\mathrm{j}\phi(\omega)}, \lambda_N(\omega))\mathcal{N}_{\mathbb{C}}(A(\omega)e^{\mathrm{j}\phi(\omega)}; 0, \lambda_X(\omega))\mathrm{d}\phi(\omega)\mathrm{d}A(\omega)}{\int_0^{\infty}\int_0^{2\pi} \mathcal{N}_{\mathbb{C}}(Y(\omega); A(\omega)e^{\mathrm{j}\phi(\omega)}, \lambda_N(\omega))\mathcal{N}_{\mathbb{C}}(A(\omega)e^{\mathrm{j}\phi(\omega)}; 0, \lambda_X(\omega))\mathrm{d}\phi(\omega)\mathrm{d}A(\omega)}$$

$$= \sqrt{\frac{\lambda_X(\omega)\lambda_N(\omega)}{\lambda_X(\omega)+\lambda_N(\omega)}}\Gamma(1.5)\Phi(-0.5, 1; -\underline{\nu(\omega)})$$

Γ : ガンマ関数

Φ : 合流超幾何関数
(confluent hypergeometric function)

$$\underline{\nu(\omega)} = \frac{\underline{\xi(\omega)}}{1+\underline{\xi(\omega)}}\underline{\gamma(\omega)}$$

事後SN比

事前SN比

$A(\omega)$ のMMSE推定量 (3/4)

- $$\sqrt{\frac{\lambda_X(\omega)\lambda_N(\omega)}{\lambda_X(\omega)+\lambda_N(\omega)}} = \sqrt{\frac{\lambda_X(\omega)}{1+\xi(\omega)}} = \sqrt{\frac{\xi(\omega)|Y(\omega)|^2}{(1+\xi(\omega))\gamma(\omega)}\frac{\gamma(\omega)}{\gamma(\omega)}} = \sqrt{\frac{\xi(\omega)\gamma(\omega)}{(1+\xi(\omega))}\frac{1}{\gamma(\omega)^2}}Y(\omega) = \frac{\sqrt{\nu(\omega)}}{\gamma(\omega)}Y(\omega)$$

$$\xi(\omega) = \frac{\lambda_X(\omega)}{\lambda_N(\omega)}$$

$$\gamma(\omega) = \frac{|Y(\omega)|^2}{\lambda_N(\omega)}$$

$$\lambda_X(\omega) = \xi(\omega)\lambda_N(\omega) = \xi(\omega)\frac{|Y(\omega)|^2}{\gamma(\omega)}$$

$$\nu(\omega) = \frac{\xi(\omega)}{1+\xi(\omega)}\gamma(\omega)$$

- $\Phi(-0.5, 1; -z) = e^{-z/2}[(1+z)I_0(z/2) + zI_1(z/2)]$

↓

$$\hat{A}_{\mathrm{MMSE}}(\omega) = \frac{\sqrt{\pi}}{2}\frac{\sqrt{\nu(\omega)}}{\gamma(\omega)}\exp\left(-\frac{\nu(\omega)}{2}\right)\left[(1+\nu(\omega))I_0\left(\frac{\nu(\omega)}{2}\right)+\nu(\omega)I_1\left(\frac{\nu(\omega)}{2}\right)\right]Y(\omega)$$

$A(\omega)$ のMMSE推定量 (4/4)

- $A(\omega)$ のMMSE推定量におけるスペクトルゲイン

$$\hat{A}_{\mathrm{MMSE}}(\omega) = \underline{G_{\mathrm{Ephraim}}(\omega)} Y(\omega)$$

$$\underline{G_{\mathrm{Ephraim}}(\omega)} = \frac{\sqrt{\pi}}{2} \frac{\sqrt{\nu(\omega)}}{\gamma(\omega)} \exp\left(-\frac{\nu(\omega)}{2}\right) \left[(1+\nu(\omega)) I_0\left(\frac{\nu(\omega)}{2}\right) + \nu(\omega) I_1\left(\frac{\nu(\omega)}{2}\right)\right]$$

目次

- 音声強調問題の設定
- 解法の分類
  - 推定対象
    - 音声の複素スペクトル／振幅スペクトル
  - 推定量
    - 最尤(ML)推定量／最小平均二乗誤差(MMSE)推定量

> - 解法1: 複素スペクトルのMMSE推定量 [Wiener1949]
> - 解法2: 振幅スペクトルのML推定量 [McAulay1980]
> - 解法3: 振幅スペクトルのMMSE推定量 [Ephraim1984]

- 事前SN比（または音声パワースペクトル密度）の推定方法
  - 最尤法（パワー減算(PS)法）

目次

- 音声強調問題の設定
- 解法の分類
  - 推定対象
    - 音声の複素スペクトル／振幅スペクトル
  - 推定量
    - 最尤(ML)推定量／最小平均二乗誤差(MMSE)推定量

- 解法1: 複素スペクトルのMMSE推定量 [Wiener1949]
- 解法2: 振幅スペクトルのML推定量 [McAulay1980]
- 解法3: 振幅スペクトルのMMSE推定量 [Ephraim1984]

- 事前SN比（または音声パワースペクトル密度）の推定方法
  - 最尤法（パワー減算(PS)法）

事前SN比と事後SN比

- どのスペクトルゲイン関数も $\xi(\omega),\ \gamma(\omega)$ によって決まる
- 事前SN比：

$$\xi(\omega) = \frac{\lambda_X(\omega)}{\lambda_N(\omega)}$$

（目的信号のパワースペクトル密度が必要）

- 事後SN比：

$$\gamma(\omega) = \frac{|Y(\omega)|^2}{\lambda_N(\omega)}$$

（観測パワースペクトルのみでOK）

- 事前SN比が未知のときは？？

→事前SN比を推定する必要あり

事前SN比の推定法

- 最尤法 (Power Subtraction法)
  - $Y(\omega) = X(\omega) + N(\omega)$,
    $N(\omega) \sim \mathcal{N}_{\mathbb{C}}(N(\omega); 0, \underbrace{\lambda_N(\omega)}_{\text{定数}}), \quad X(\omega) \sim \mathcal{N}_{\mathbb{C}}(X(\omega); 0, \lambda_X(\omega))$
    のもとで
    $$\hat{\lambda}_X(\omega) = \underset{\lambda_X(\omega)}{\operatorname{argmax}}\, p(Y(\omega)|\lambda_X(\omega))$$
    を音声パワースペクトル密度の推定値とする単純なやり方
  - 仮定より
    $$p(Y(\omega)|\lambda_X(\omega)) = \mathcal{N}_{\mathbb{C}}(0, \lambda_X(\omega) + \lambda_N(\omega))$$
  - $\dfrac{\partial \log p(Y(\omega)|\lambda_X(\omega))}{\partial \lambda_X(\omega)} = 0 \longrightarrow \hat{\lambda}_X(\omega) = |Y(\omega)|^2 - \lambda_N(\omega)$
  - 事前SN比の推定値は･･･ $\hat{\xi}(\omega) = \dfrac{\hat{\lambda}_X(\omega)}{\lambda_N(\omega)} = \underbrace{\gamma(\omega)}_{\text{事後SN比}} - 1$

PS法による振幅スペクトル推定値を用いた $X(\omega)$ の推定

- PS法による推定振幅 $\sqrt{|Y(\omega)|^2-\lambda_N(\omega)}$ と観測位相 $Y(\omega)/|Y(\omega)|$ により $X(\omega)$ の推定値を得ることも可能

$$\hat{X}_{\mathrm{PS}}(\omega)=\boxed{\sqrt{|Y(\omega)|^2-\lambda_N(\omega)}\frac{Y(\omega)}{|Y(\omega)|}}=\underline{G_{\mathrm{PS}}(\omega)}Y(\omega)$$

$$\underline{G_{\mathrm{PS}}(\omega)}=\frac{\sqrt{|Y(\omega)|^2-\lambda_N(\omega)}}{|Y(\omega)|}$$

$$=\sqrt{\frac{\gamma(\omega)-1}{\gamma(\omega)}}$$

$$\gamma(\omega)=\frac{|Y(\omega)|^2}{\lambda_N(\omega)}$$

事後SN比

- PS法による音声パワースペクトル密度推定値を代入したWienerフィルタと比較してみると・・・

$$G_{\mathrm{Wiener}}(\omega)=\frac{\lambda_X(\omega)}{\lambda_X(\omega)+\lambda_N(\omega)}=\frac{\xi(\omega)}{\xi(\omega)+1}=\frac{\nu(\omega)-1}{\nu(\omega)}$$

$$\xi(\omega)=\nu(\omega)-1$$

まとめ

- 音声強調問題の設定
- 解法の分類
  - 推定対象
    - 音声の複素スペクトル／振幅スペクトル
  - 推定量
    - 最尤(ML)推定量／最小平均二乗誤差(MMSE)推定量

> - 解法1: 複素スペクトルのMMSE推定量 [Wiener1949]
> - 解法2: 振幅スペクトルのML推定量 [McAulay1980]
> - 解法3: 振幅スペクトルのMMSE推定量 [Ephraim1984]

- 事前SN比（または音声パワースペクトル密度）の推定方法
  - 最尤法（パワー減算(PS)法）

レポート課題の対象論文

- Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator," IEEE Trans. Acoust. Speech Signal Process., Vol. 32, No. 6, pp. 1109-1121, 1984