Joint Source-Channel Decoding of Speech Spectrum Parameters over Erasure Channels Using Gaussian Mixture Models

Anand D. Subramaniam. William R. Gardner and Bhaskar D. Rao

Reporter:黃重翔

Professor:陳嘉平

ABSTRACT

A joint source-channel decoding scheme that improves the performance of conventional channel decoders over erasure channels by exploiting the cross-correlation between successive speech frames is presented. Speech spectrum parameters are quantized using the scheme presented in [1]. The joint probability density function (PDF) of the spectrum parameters of successive speech frames is modelled using a Gaussian mixture model (GMM). This model is then used to process the channel decoder output over erasure channels. The performance of two decoding strategies, namely, Maximum Likelihood decoding (ML) and Minimum Mean Squared Error decoding (MMSE) is shown to provide significantly better performance than prediction based schemes.

Outline

- Introduction
- Source Model
- Encoding
- Decoding
- Experimental Results
- Computational Complexity
- Conclusion

Introduction

- 本論文提出一聯合機率密度函數聯合信號源與通道的編碼方法，其中在連續語音框架之間的頻譜參數被用於協助通道解碼器。
- 文中提出兩種策略：最大相似解碼法（Maximum Likelihood）與最小均方差（Minimum mean square error）解碼法。

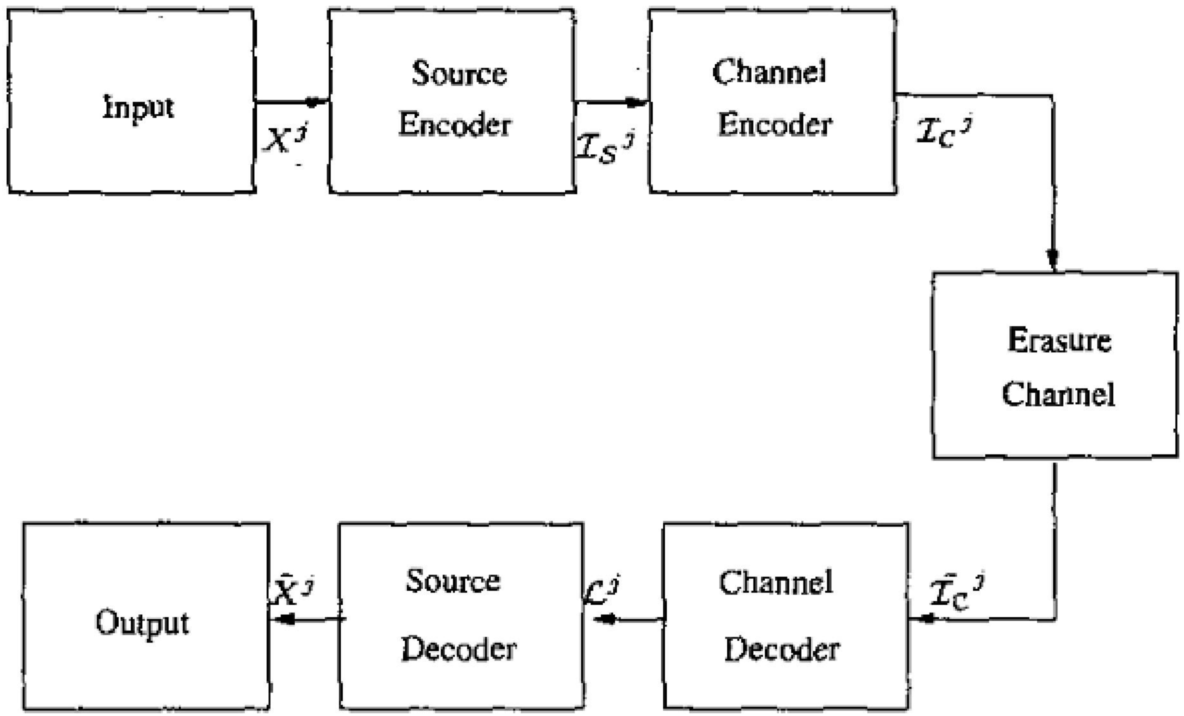
Source Model

- 在連續語音框架之間的頻譜參數用GMM進行模組化

$$
\begin{aligned}
f_{\mathbf{X},\mathbf{Y}}(\mathbf{X},\mathbf{Y}) &= \sum_{i=1}^{m} \alpha_i \mathrm{N}_i\left(\mu_i, C_i\right) \\
\mu_i &= \begin{pmatrix} \mu_i^{\mathbf{X}} \\ \mu_i^{\mathbf{Y}} \end{pmatrix} \\
C_i &= \begin{pmatrix} C_i^{\mathbf{XX}} & C_i^{\mathbf{XY}} \\ C_i^{\mathbf{YX}} & C_i^{\mathbf{YY}} \end{pmatrix}
\end{aligned}
$$

$$
f_{\mathbf{X}}(\mathbf{X}) = \sum_{i=1}^{m} \alpha_i \mathrm{N}_i\left(\mu_i^{\mathbf{X}}, C_i^{\mathbf{XX}}\right)
$$

Overall Communication Scheme



Encoding

- Source encoder: 經過碼率相關的無記憶性量化固定的取得一k-bit的index $\mathcal{I}_S$

- Channel encoder: 每個輸入k-bit的 $\mathcal{I}_S$ 都可以產生n-bit的通道編碼 $\mathcal{I}_C$，它的消除更正能力爲 e_{code} =n-k

Decoding

- 定義「過剩消除」量爲 $e_{ex} = np - e_{code}$
- Channel decoder:使e^j爲在第j^{th}個語音框架中的通道產生的消除個數
  - 若 $e^j \leq e_{code}$則通道輸出可被通道解碼器完整解碼
  - 若 $e^j > e_{code}$ 則通道解碼器輸出一list $\mathcal{L}_j$，包含 $N_j = 2^{e^j_{ex}}$ 個輸入，且 $e^j_{ex} = e^j - e_{code}$ ，此list包含轉換過的來源字集作爲輸入之一

Decoding

- Source decoder: $e^j > e_{code}$ 時才會作用。作用時解碼器會嘗試利用聯合連續框架的頻譜參數的統計值來運算通道解碼器輸出 $\mathcal{L}_j$

Prediction

- 當通道解碼器不能完整解碼通道輸出時，我們根據編碼後前一框架的參數預估目前框架的頻譜參數，利用兩種方法：線性預測（Linear prediction）與GMM預測（GMM-P）。

Linear Prediction

- LP法中利用前一框架的頻譜參數進行目前框架頻譜參數的理想化線性預測估計

$$
\begin{aligned}
\tilde{X}^{j}_{LP} &= \mathbf{A} \cdot (\tilde{X}^{j-1}_{LP} - \mu^{\mathbf{X}}) + \mu^{\mathbf{X}} \\
\mathbf{A} &= C^{\mathbf{XY}}(C^{\mathbf{YY}})^{-1}
\end{aligned}
$$

GMM Prediction

- GMM-P中，基於前一框架決定目前框架的條件性平均值

$$f_{\mathbf{X}|\mathbf{Y}}(\mathbf{X} \mid \mathbf{Y}) = \sum_{i=1}^{m} \beta_i(\mathbf{Y}) \mathrm{N}_i\left(m_i(\mathbf{Y}), \Sigma_i\right) \quad (8)$$

$$m_i(\mathbf{Y}) = \mu_i^{\mathbf{X}} + C_i^{\mathbf{XY}} \left(C_i^{\mathbf{YY}}\right)^{-1} \left(\mathbf{Y} - \mu_i^{\mathbf{Y}}\right)$$

$$\Sigma_i = C_i^{\mathbf{XX}} - C_i^{\mathbf{XY}} (C_i^{\mathbf{YY}})^{-1} C_i^{\mathbf{YX}}$$

$$\beta_i(\mathbf{Y}) = \frac{\alpha_i f_i(\mathbf{Y})}{\sum_{j=1}^{m} \alpha_j f_j(\mathbf{Y})}$$

$$f_i(\mathbf{Y}) = \mathrm{N}\left(\mu_i^{\mathbf{Y}}, C_i^{\mathbf{YY}}\right)$$

GMM Prediction

- 設 $\tilde{X}^{j}_{PRED}$ 為解碼預測估計值，可用條件性密度取得

$$\begin{aligned} \tilde{X}^{j}_{GMMP} &= \mathrm{E}\left(\mathbf{X} \mid \mathbf{Y} = \tilde{X}^{j-1}_{GMMP}\right) \qquad (9) \\ &= \sum_{i=1}^{m} \beta_i(\tilde{X}^{j-1}_{GMMP}) \cdot m_i(\tilde{X}^{j-1}_{GMMP}) \end{aligned}$$

而由於最小化MSE會使估計值比LP的還要好。

Maximum Likelihood

- 此方法藉由找出發生最大機率於前一編碼框架和通道編碼器輸出的來源字集

$$\tilde{X}^{j}_{ML} = \arg\max_{X \in \mathcal{C}_j} P(\mathbf{X} = X \mid \mathbf{Y} = \tilde{X}^{j-1}_{ML})$$

- 爲了計算來源字集X，必須整合X的Voronoi區域之條件性密度，其值爲V(X)，並算出近似機率

$$P(X \mid \tilde{X}^{j-1}_{ML}) = f_{\mathbf{X}|\mathbf{Y}}\left(\mathbf{X} = X \mid \mathbf{Y} = \tilde{X}^{j-1}_{ML}\right) \cdot V(X)$$

Maximum Likelihood

- V(X)可用 $V(X) = \frac{1}{2^k \cdot \Lambda(X)}$ 求得
- $\Lambda(X)$ 爲無記憶性量化器的字集密度，值爲

$$\begin{aligned} \Lambda(X) &= \frac{\sum_{i=1}^{m} g_i(X)^{1/3} (\alpha_i \lambda_i)^{d/(d+2)} \lambda_i^{-d/3}}{\sum_{j=1}^{m} (\alpha_j \lambda_j)^{d/(d+2)}} \\ g_i(X) &= \mathrm{N}\left(\mu_i^{\mathbf{X}}, C_i^{\mathbf{XX}}\right) \end{aligned}$$

λ_i是$C_i^{\mathbf{XX}}$之奇異值的幾何平均值。

Minimum Mean Square Error

- 本方法由前一框架與通道編碼器輸出提供的條件性平均進行估算

$$
\begin{aligned}
\tilde{X}^{j}_{MMSE} &= \mathrm{E}\left(\mathbf{X} \mid \mathbf{Y} = \tilde{X}^{j-1}_{MMSE}, \mathcal{L}_j\right) \\
&= \frac{\sum_{X \in L_j} P(X \mid \tilde{X}^{j-1}_{MMSE}, L_j) \cdot X}{\sum_{X \in \mathcal{L}_j} P(X \mid \tilde{X}^{j-1}_{MMSE}, L_j)}
\end{aligned}
$$

另外特別使用來源模型計算每一可能的字碼，然後計算條件性平均。

Experimental Results

Table 1. Performance Results

e_{ex}	L-P	GMM-P	M-L	MMSE
0.0	1.0749	1.0749	1.0749	1.0749
0.2	1.5792	1.5682	1.1463	1.1485
0.4	2.1755	2.1617	1.2370	1.2311
0.6	2.6915	2.6609	1.3244	1.3123
0.8	3.1068	3.0978	1.3700	1.3697
1.0	3.7007	3.6991	1.4603	1.4451

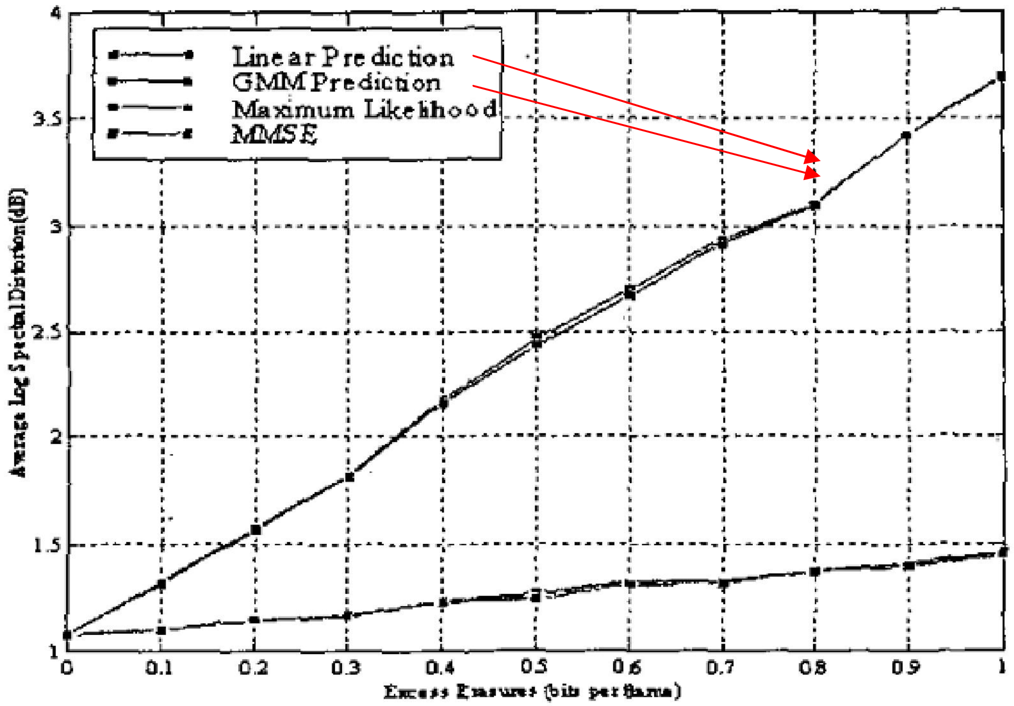


Fig. 2: Performance

Computational Complexity

- N_{exp} 爲計算一個冪值所需的flops

Table 2. Computational Complexity

Scheme	N_{tot} (flops per frame)
LP	$2d^2 + d$
GMM-P	$m(N_{exp} + 2d^2 + d + 3) - 2$
M-L	$2^{e^j_{ex}}[(2m+1)N_{exp} + m(2d^2 + d) + 5m - 2]$
MMSE	$2^{e^j_{ex}}[(2m+1)N_{exp} + m(2d^2 + d) + 5m + d]$

Conclusion

- 建構於GMM預測的來源通道解碼方法比LP法好，然而整體而言只有在字碼相當短的時候會有好的效果。

Bit Allocation (Fixed Rate)

- 在一在編碼本中的字集總數是固定的：

$$2^{b_{tot}} = \sum_{i=1}^{m} 2^{b_i}$$

- 設 α_i 爲估計cluster i的發生機率，則量化器框架的平均失真總和之上界爲

$$D_{tot} = \sum_{i=1}^{m} \alpha_i D_i(b_i)$$

$$2^{b_i} = 2^{b_{tot}} \frac{(\alpha_i \lambda_i)^{d/(d+2)}}{\sum_{j=1}^{m} (\alpha_j \lambda_j)^{d/(d+2)}}$$

Bit Allocation (Fixed Rate)

- 為了進行固定比率的位元定址必須達成

$$\min_{b_i} \ D_{tot} = \sum_{i=1}^{m} \alpha_i D_i(b_i), \quad \text{subject to} \quad 2^{b_{tot}} = \sum_{i=1}^{m} 2^{b_i}$$

- 對 $D_i(b_i)$ 用Gaussian進行高解析表示式

$$\begin{aligned} D_i(b_i) &= \frac{\sqrt{3}\pi}{2} p c_i 2^{-2b_i/p} \\ c_i &= \left[\prod_{k=1}^{p} \lambda_{i,k}\right]^{\frac{1}{p}}, 1 \leq i \leq m \\ \Lambda_i &= \mathrm{diag}(\lambda_{i,1}, \lambda_{i,2}, \ldots, \lambda_{i,p}) \\ \Sigma_i &= Q_i \Lambda_i Q_i^T \end{aligned}$$

Bit Allocation (Fixed Rate)

- 由此可以得到位元定址的公式

$$2^{b_i} = 2^{b_{tot}} \frac{(\alpha_i c_i)^{p/p+2}}{\sum_{i=1}^{m} (\alpha_i c_i)^{p/p+2}}, \quad 1 \leq i \leq m$$

$$2^{b_i} = 2^{b_{tot}} \frac{(\alpha_i \lambda_i)^{d/(d+2)}}{\sum_{j=1}^{m} (\alpha_j \lambda_j)^{d/(d+2)}}$$