

一种基于 GMM 和多项式拟合的 语音编码改进算法

王蓉蓉¹, 李 平², 曾毓敏¹, 韦 怡¹

(1. 南京师范大学物理科学与技术学院, 江苏 南京 210023)

(2. 泰州职业技术学院信息技术学院, 江苏 泰州 225300)

[摘要] 提出一种基于高斯混合模型和多项式拟合的语音编码改进算法. 在 GMM 模型对短时语音谱包络进行参数化的基础上, 将一定数量的语音帧划分为一个片段, 利用谱特征的相关性对片段内的 GMM 参数进行多项式拟合联合编码, 从而使得参数进一步减少. 仿真结果表明, 本文算法码率对比基于 GMM 的语音编码器有显著降低.

[关键词] 语音编码, GMM, 多项式拟合, 范特蒙矩阵

[中图分类号] TN912.3 [文献标志码] A [文章编号] 1672-1292(2017)02-0063-07

An Improved Speech Coding Algorithm Based on GMM and Polynomial Fitting

Wang Rongrong¹, Li Ping², Zeng Yumin¹, Wei Yi¹

(1. School of Physical Science and Technology, Nanjing Normal University, Nanjing 210023, China)

(2. College of Information Technology, Taizhou Polytechnic College, Taizhou 225300, China)

Abstract: A vocoder is proposed basing on polynomial fitting and Gaussian Mixture Model (pGMM). In the vocoder, several frames are collected into a segment after using GMM model to parameterize the short-time speech spectrum envelope. The polynomial trajectory is used to fit the parameters of GMM in a segment according to the correlation between neighboring frames, thus reducing the number of parameters. The results show that the bit rate of pGMM vocoder is further reduced in contrast with the vocoder based on GMM.

Key words: speech coding, GMM, polynomial fitting, Vandermonde matrix

语音编码在语音通信及人类信息交流中占有举足轻重的地位. 低速率语音编码 (4.8 kb/s ~ 1.2 kb/s) 和极低速率语音编码 (低于 1.2 kb/s) 是现代语音编码技术研究和发展的一个重要方向. 典型的低码率语音编码器, 有基于线性预测编码 (LPC) 的语音编码器、基于码激励线性预测 (CELP) 的语音编码器和基于多带激励 (MBE) 的语音编码器等^[1].

在语音编码中, 有一种利用语音段的时间相关关系的方法, 即分段编码. 这种方法可应用于任何基于帧的语音编码器, 可比原始编码器获得更大的压缩率^[2]. Laurent Firin^[3] 提出了在“长期”的基础上编码 LPC 语音编码器的 LSF 参数的方法. 该方法中语音被分段成浊音/清音段, 利用 LSF 帧间相关性将 LSF 向量时间轨迹的离散余弦模型应用于每个片段. 张楠^[4] 在 2.4 kb/s 混合激励线性预测 (MELP) 语音编码方案的基础上, 提出改进的超帧结构分类方法和有效的基音多帧联合矢量量化方法, 从而进一步提高编码效率. 除此之外高斯混合模型也成为在数据处理中比较高效的一种方法. 李平^[5] 提出了一种新颖的基于高斯混合模型 (GMM) 的甚低码率语音编码系统, 该系统利用 GMM 对短时语音谱包络进行拟合模仿的方法来对语音进行参数化表示. 该编码器在编码码率降低到 2.35 kb/s 时, 仍可获得音质令人满意的解码语音. GMM 在处理数据时默认每一帧之间没有相关性, 因此基于 GMM 的低码率语音编码器在编码时就默

收稿日期: 2016-09-18.

基金项目: 江苏省科技支撑计划 (BE2014139)、江苏省自然科学基金 (BK2010546).

通讯联系人: 曾毓敏, 教授, 研究方向: 语音信号处理和图像处理. E-mail: zengyumin@njnu.edu.cn

认了帧与帧之间的独立性^[6]. 但实际上,语音帧之间存在一定的关联. 鉴于上述缺点,为了很好地利用相邻帧间的相关性,本文对文献[5]提出的基于高斯混合模型的编码器进行改进,将固定帧数的语音信号集合到一个片段中,并对片段内各帧的高斯均值和方差参数进行多项式拟合,使得参数由高斯混合模型参数变为相应的拟合参数,参数数量得以减少,进而编码器在码率上有进一步的降低.

1 基于 GMM 的语音编码器

GMM 模型被广泛应用于数据分类、图像识别以及语音识别^[7]等方面. 同时,GMM 模型也可被应用于语音编码中. 文献[5]提出的基于 GMM 谱包络拟合的编码器流程图如图 1 所示.

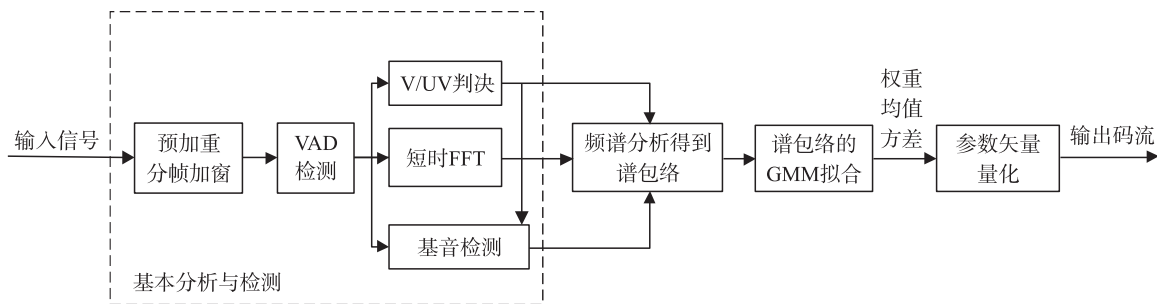


图 1 基于 GMM 模型的语音编码框图
Fig. 1 Speech coding block diagram based on GMM model

文献[5]采用的是 SEEVOC 法得到语音谱包络. 在得到语音谱包络后用 GMM 模型对谱包络进行参数化表示. GMM 函数是概率密度分布,其概率值范围在(0~1),短时语音谱包络可以看成是一个函数分布,但其具有能量,幅值分布不完全在(0~1)范围内,此时可以拿每帧的语音谱包络幅值除以这一帧的语音谱包络幅值和,就能够产生与 GMM 相匹配的分布函数,并保留语音谱幅度信息,即可用 GMM 参数对上述分布函数进行表示.

GMM 参数由权重参量 ω 、均值参量 μ 、方差参量 σ^2 组成. 一个 M 阶高斯混合模型的概率密度函数是由 M 个高斯概率密度函数加权求和得到的^[8],可表示为:

$$p(x|\Phi) = \sum_{i=1}^M \omega_i f_i(x|\varphi_i), \quad (1)$$

式中, M 是高斯混合模型的阶数, ω_i 是混合权重,满足 $\sum_{i=1}^M \omega_i = 1$. $f_i(x|\varphi_i)$ 是子分布,每个子分布可表示为:

$$f_i(x) = \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left\{-\frac{(x_k - \mu_i)^2}{2\sigma_i^2}\right\}, \quad (2)$$

式中, μ_i 是均值参数, σ_i^2 是方差参数. 常常采用 EM 算法估计参数 Φ ^[9]:

$$\omega_i = \frac{1}{N} \sum_{k=1}^N P(x_k) P(i|x_k, \varphi_i), \quad (3)$$

$$\mu_i = \frac{\sum_{k=1}^N P(x_k) P(i|x_k, \varphi_i) x_k}{\sum_{k=1}^N P(x_k) P(i|x_k, \varphi_i)}, \quad (4)$$

$$\sigma_i^2 = \frac{\sum_{k=1}^N P(x_k) P(i|x_k, \varphi_i) (x_k - \mu_i)^2}{\sum_{k=1}^N P(x_k) P(i|x_k, \varphi_i)}. \quad (5)$$

得到高斯混合模型参数后,将参数进行矢量量化后用于传输或储存.

表 1 给出了 GMM 编码器一帧语音编码所需要的比特数.

表 1 GMM 编码器一帧语音编码所需比特数
Table 1 The number of bits required for one frame of GMM vocoder

编码参数	比特数/bit
VAD 检测	1
U/V 判决	1
基音周期	10
谱幅度和值	8
GMM 参数	27
一帧所需比特数	47

(1)VAD 检测

1 为语音,0 为静音,1 bit 量化.

(2)U/V 判决

1 为浊音,0 为清音,1 bit 量化.

(3)基音周期

基音周期整数部分用 8 bits 量化,分数部分 2 bits 量化精确到 1/4,一帧需要 10 bits.

(4)谱幅度和值

由于相邻语音帧的能量幅度变化不大,因此谱幅度和值用差值量化方法,差值整数部分 4 bits 量化,分数部分 3 bits 量化,谱幅度和值增加或减少标记用 1 bit 量化,因此谱幅度和值 8 bits 量化.

(5)GMM 参数

采用 512 码本矢量量化,传输码本标号,512 码本需要 9 bits,权重、均值以及方差 3 种参数共需 27 bits.

因此一帧 GMM 编码器语音编码需要 47 bits.

该编码器的语音分帧帧长为 20 ms,因此,编码速率为 $\frac{47 \text{ bits}}{20 \text{ ms}} = 2.35 \text{ kb/s}$.

文献[5]所构建的基于 GMM 的语音编码器,实现了 2.35 kb/s 的甚低编码码率,解码语音具有良好的自然度和可懂度.

2 基于 GMM 和多项式拟合的编码改进算法(pGMM)

文献[5]编码器使用 GMM 参数对语音谱包络进行参数化表示,但其帧与帧之间的均值参数 μ_i 和方差参数 σ_i^2 是互不相关的. 实际上由于人的发音器官的肌肉运动速度较慢,在一段短时间内语音信号保持相对稳定一致的特征,即语音信号具有短时平稳性^[10]. 经过分帧后,在暂稳时间里的帧之间具有相关性,语音的特征参数也应具有一定的相关性. 本文就利用这种相关性进一步改进文献[5]的算法.

本文 pGMM 算法的编码框图如图 2 所示. 与文献[5]基于 GMM 的编码器的不同之处在于,本文算法在用 GMM 对谱包络进行拟合表示得到高斯混合模型参数之后,每 T 帧划分为一个片段,分别用多项式拟合片段内的均值参数和方差参数,用得到的拟合参数 B 和 C 取代均值参数和方差参数进行参数矢量量化. 本文算法仅对均值和方差进行多项式拟合而不对权重进行拟合的原因在于,经过拟合后解码得到的权重值可能为负值,而权重实际上不可能为负,此外解码返回的值很难满足各权重相加总和为 1 这个条件. 在本文改进算法中,仅对均值和方差参数进行多项式拟合.

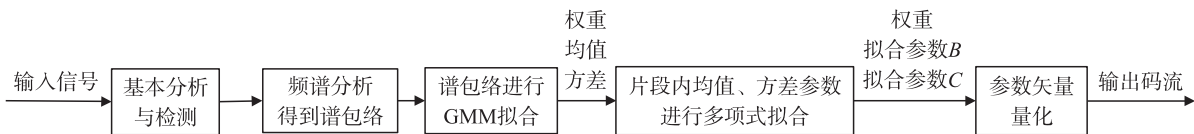


图 2 基于 pGMM 模型的语音编码框图

Fig. 2 Speech coding block diagram based on pGMM model

多项式函数是形式比较简单的函数,形如 $y = a_0 + a_1x + a_2x^2 + \dots + a_nx^n$,通过最小二乘法可求出系数 $a_0, a_1, a_2, \dots, a_n$ ^[11],在此不多赘述. 本文算法的改进之处在于用多项式拟合片段内的高斯混合模型参数. 以高斯混合模型的均值参数为例:高斯混合模型阶数为 M 阶的一帧语音,其均值分量为 $[\mu_1, \mu_2, \dots, \mu_M]$. 假设一个片段有 T 帧,则对于 M 阶高斯混合模型,就有

$$U = \begin{bmatrix} \mu_{11} & \mu_{12} & \cdots & \mu_{1M} \\ \mu_{21} & \mu_{22} & \cdots & \mu_{2M} \\ \vdots & \vdots & \vdots & \vdots \\ \mu_{T1} & \mu_{T2} & \cdots & \mu_{TM} \end{bmatrix}, \quad (6)$$

式中,第 T 行代表第 T 帧中的 M 个均值分量,第 M 列代表一个片段中所有帧的第 M 个高斯分量. 利用暂稳时间内语音信号参数的相关性,也即式(6)每一列的相关性,用多项式对每一列分别进行拟合. 以第 M 个均值向量也即第 M 列为例,用多项式进行拟合,帧数 t 为自变量,均值 μ 为因变量, b 为多项式拟合系数:

第 1 帧的第 M 个高斯均值参数 $\mu_{1M}:\mu_{1M}=1^0\times b_{1M}+1^1\times b_{2M}+\cdots+1^R\times b_{(R+1)M}$,

第 2 帧的第 M 个高斯均值参数 $\mu_{2M}:\mu_{2M}=2^0\times b_{1M}+2^1\times b_{2M}+\cdots+2^R\times b_{(R+1)M}$,

\vdots

第 t 帧的第 M 个高斯均值参数 $\mu_{tM}:\mu_{tM}=t^0\times b_{1M}+t^1\times b_{2M}+\cdots+t^R\times b_{(R+1)M}$,

\vdots

第 T 帧第 M 个高斯的均值参数 $\mu_{TM}:\mu_{TM}=T^0\times b_{1M}+T^1\times b_{2M}+\cdots+T^R\times b_{(R+1)M}$.

将上述函数表达式改写为矩阵模式:

$$\begin{bmatrix} \mu_{1M} \\ \mu_{2M} \\ \vdots \\ \mu_{TM} \end{bmatrix} = \begin{bmatrix} 1^0 & 1^1 & \cdots & 1^R \\ 2^0 & 2^1 & \cdots & 2^R \\ \vdots & \vdots & \vdots & \vdots \\ T^0 & T^1 & \cdots & T^R \end{bmatrix} \times \begin{bmatrix} b_{1M} \\ b_{2M} \\ \vdots \\ b_{(R+1)M} \end{bmatrix}. \quad (7)$$

同理,对于整个语音片段内的所有均值参数,有:

$$\begin{bmatrix} \mu_{11} & \mu_{12} & \cdots & \mu_{1M} \\ \mu_{21} & \mu_{22} & \cdots & \mu_{2M} \\ \vdots & \vdots & \vdots & \vdots \\ \mu_{T1} & \mu_{T2} & \cdots & \mu_{TM} \end{bmatrix} = \begin{bmatrix} 1^0 & 1^1 & \cdots & 1^R \\ 2^0 & 2^1 & \cdots & 2^R \\ \vdots & \vdots & \vdots & \vdots \\ T^0 & T^1 & \cdots & T^R \end{bmatrix} \times \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1M} \\ b_{21} & b_{22} & \cdots & b_{2M} \\ \vdots & \vdots & \vdots & \vdots \\ b_{(R+1)1} & b_{(R+1)2} & \cdots & b_{(R+1)M} \end{bmatrix}, \quad (8)$$

令

$$\mathbf{Z} = \begin{bmatrix} 1^0 & 1^1 & \cdots & 1^R \\ 2^0 & 2^1 & \cdots & 2^R \\ \vdots & \vdots & \vdots & \vdots \\ T^0 & T^1 & \cdots & T^R \end{bmatrix}, \quad (9)$$

令

$$\mathbf{B} = \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1M} \\ b_{21} & b_{22} & \cdots & b_{2M} \\ \vdots & \vdots & \vdots & \vdots \\ b_{(R+1)1} & b_{(R+1)2} & \cdots & b_{(R+1)M} \end{bmatrix}, \quad (10)$$

则式(8)可以写成:

$$\mathbf{U} = \mathbf{Z} \times \mathbf{B}. \quad (11)$$

R 和 T 的值是设定的已知值,因此矩阵 \mathbf{Z} 是确定的. \mathbf{U} 也是由上一步谱包络进行 GMM 参数化表示的均值组成的,只有拟合参数矩阵 \mathbf{B} 是未知矩阵. 经过推导可得:

$$\mathbf{B} = \text{inv}(\mathbf{Z}' \times \mathbf{Z}) \times \mathbf{Z}' \times \mathbf{U}. \quad (12)$$

由上述算法分析可以看出,设计矩阵 \mathbf{Z} 是一个范德蒙矩阵的转置矩阵,是一个大小为 $T \times (R+1)$ 的矩阵, \mathbf{B} 是一个大小为 $(R+1) \times M$ 的矩阵, \mathbf{U} 是一个大小为 $T \times M$ 的矩阵. 得到拟合矩阵 \mathbf{B} 后,用 \mathbf{B} 代替原来的 GMM 参数中的均值矩阵 \mathbf{U} 经过矢量化后用于传输或存储等. 拟合矩阵 \mathbf{B} 相比于均值矩阵 \mathbf{U} ,列数均为 M ,但是行数不同,拟合矩阵 \mathbf{B} 的行数为 $(R+1)$,均值矩阵 \mathbf{U} 的行数为 T , $(R+1) < T$,因此拟合矩阵 \mathbf{B} 相比于均值 \mathbf{U} ,所需的比特数小于 \mathbf{U} 所需的比特数,因而降低编码码率是可行的.

对于方差参量的处理与上述均值参量的处理相同,用拟合矩阵 \mathbf{C} 代替.

当拟合帧数 $T=8$,拟合阶数 $R=2$ 时,基于 pGMM 模型的语音编码器中一个语音片段编码所需的比特分配状况如表 2 所示.

(1)VAD 检测

用 1、0 分别代表语音、静音,1 bit 量化,8 帧构成一个片段,则该片段需要 8 bits.

表 2 pGMM 编码器一个语音片段编码所需比特数

Table 2 The number of bits required for one segment of pGMM vocoder

编码参数	所需比特/bit
VAD 检测	1×8
U/V 判决	1×8
基音周期	10×8
谱幅度和值	8×8
pGMM 参数	72+27+27
片段所需比特数	286

(2)U/V 判决

用 1、0 分别代表浊音、清音,1 bit 量化,共需 8 bits.

(3)基音周期

基音分别用 8 bits、2 bits 量化整数、分数部分,一帧需要 10 bits,片段共需 80 bits.

(4)谱幅度和值

由于相邻语音帧的能量幅度变化不大,因此谱幅度和值用差值量化方法,差值整数部分 4 bits 量化,分数部分 3 bits 量化,谱幅度和值增加或减少标记用 1 bit 量化,因此谱幅度和值 8 bits 量化,一个片段共需 64 bits.

(5)pGMM 参数

进行矢量量化,512 码本用 9 bits. 由于 pGMM 模型中拟合的为均值和方差参数,权重参数并未进行拟合,因此权重、均值和方差分别为 72 bits、27 bits 和 27 bits 量化,共需 126 bits.

因此,一个语音片段 pGMM 编码器语音编码需要 286 bits.

在语音帧长为 20 ms 的情况下,编码速率为 $\frac{286 \text{ bits}}{20 \text{ ms} \times 8} = 1.7875 \text{ kb/s}$. 实际语音中,通常有静默期存在^[12]. 假设静音帧占整个语音的 40%,则该编码器码率为 $1.7875 \times 0.6 = 1.0725 \text{ kb/s}$.

可以看出,本文所给出的 pGMM 编码器码率比文献[5]的 GMM 编码器有很大的降低.

图 3 给出的是基于 pGMM 模型的语音编码器的解码框图. 解码端将接收到的码本角标用查表的方法从码本中得到相应的语音参数,恢复语音信号的短时谱包络,再分别对清音、浊音和静音合成. 基于浊音的准周期,选用正弦信号合成. 清音直接对恢复的谱包络进行修正,乘以一个 $0 \sim 2\pi$ 的随机相位,然后经过 IFFT 得到. 系统生成一个均值为 0,方差为 1 的类似白噪声的随机序列,再与静音帧的平均幅度值相乘得到静音信号. 为了增强语音帧之间的平滑性,采用线性加权处理解码后语音帧的叠加问题,最终得到合成语音.

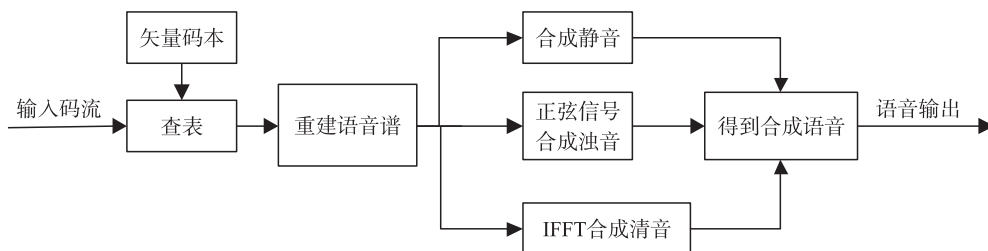


图 3 基于 pGMM 模型的语音编码器解码框图

Fig. 3 Decoding block diagram of speech vocoder based on pGMM model

经过综合考量不同阶数 T 与拟合阶数 R 对应的 pGMM 编码器码率与语音质量之后,本文 pGMM 编码器选择片段帧数 $T=8$,多项式拟合阶数 $R=2$. 图 4 给出了一段语音的原始语音波形图以及基于 pGMM 模型的语音编码器解码语音波形图. 通过对比编码前后的波形图可以看出,解码语音的波形趋势和信号幅度都与原始语音信号相似,保留了原始语音信号的主要特征.

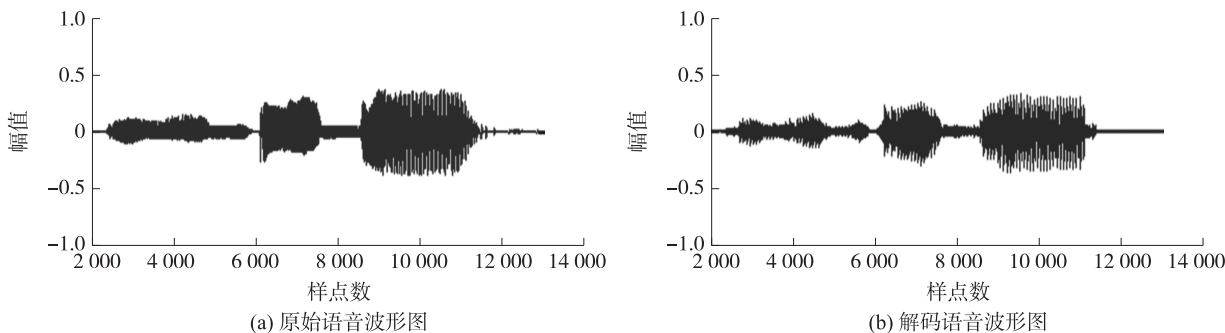


图 4 基于 pGMM 模型的语音编码器解码语音波形图

Fig. 4 Decode waveform of speech vocoder based on pGMM model

3 实验及结果分析

实验平台参数为:Lenovo M8500t 台式计算机,Intel 酷睿 i5 3.3 GHz CPU,4 GB RAM,Windows 7 操作系统,仿真使用软件 MATLAB 2013 进行.

实验选用 8 段语音作为测试用例(4 男 4 女),采样频率为 8 kHz,16 bits 量化.加窗分帧使用汉明窗,帧长为 20 ms,采用 8 阶高斯分量的 GMM 模型对每帧的语音谱包络进行参数化表示.

实验 1 阶数 R 对编码器性能的影响

平均分信噪比^[13]是从时域的角度来考察重建语音对输入语音的失真程度.分段信噪比采用分段的方法分别计算每一段语音信号的信噪比,这里的信噪比实际上是信号与量化误差或噪声的功率比.倒谱距离测度则是从频域来评价输入语音和量化语音的频谱失真度.信噪比越大说明解码语音效果越好,倒谱距离越小说明解码语音与原始语音越接近.

表 3 给出的是有固定 8 帧组成一个片段,当拟合阶数 R 不同时编码器解码语音的码率、分段信噪比值以及倒谱距离值.由表 3 可以看出,片段帧数 T 固定时,拟合阶数 R 越大,编码器码率越高.同时,分段信噪比随着 R 的增加而增大,倒谱距离随着 R 的增加而减小.说明当帧数一定时,阶数越高,编码器码率越高,语音质量越好.

实验 2 帧数 T 对编码器性能的影响

表 4 给出的是拟合阶数 R 固定为 2 阶,片段内的帧数不同时编码器解码语音的码率、分段信噪比值以及倒谱距离值.由表 4 可以看出,拟合阶数 R 固定时,片段帧数 T 越少,编码器码率越高.同时,分段信噪比随着 T 的增加而减小,倒谱距离随着 T 的增加而增大.说明当拟合阶数一定时,帧数越多,编码器码率越低,语音质量越差.

表 3 相同帧数不同多项式拟合阶数的 pGMM 编码器性能对比

Table 3 Performance comparison of pGMM encoder with the same frame number and different polynomial fitting degree

T	R	码率/($\text{kb}\cdot\text{s}^{-1}$)	分段信噪比	倒谱距离
8	1	1.005	-0.902	8.955
8	2	1.073	-0.899	8.575
8	3	1.140	-0.787	8.177
8	4	1.208	-0.720	7.316
8	5	1.275	-0.699	6.300
8	6	1.343	-0.656	5.086
8	7	1.410	-0.516	2.741

表 4 相同多项式拟合阶数不同帧数的 pGMM 编码器性能对比

Table 4 Performance comparison of pGMM encoder with the same polynomial fitting degree and different frame number

T	R	码率/($\text{kb}\cdot\text{s}^{-1}$)	分段信噪比	倒谱距离
4	2	1.275	-0.607	6.730
6	2	1.140	-0.831	8.129
8	2	1.073	-0.839	8.575
10	2	1.032	-0.895	8.850
12	2	1.005	-0.905	9.014

实验 3 与其他编码器的性能比较

将本文提出的 pGMM 编码器(取拟合片段帧数为 8,拟合阶数为 2)与文献[5]提出的 GMM 编码器以及文献[14]提出的极低码率编码器的码率、PESQ 得分以及时延等进行比较,结果如表 5 所示.

从表 5 可以看出,在编码速率方面,本文编码器低于文献[5]编码器,但高于文献[14]编码器;在合成语音质量方面,本文编码器介于文献[5]与文献[14]编码器之间,PESQ 得分比文献[5]编码器的 PESQ 得分低 0.29,比文献[14]编码器的 PESQ 得分高 0.52;在算法延时方面,文献[14]编码器的时延最大,本文编码器次之,文献[5]编码器时延最小.

4 结语

高质量低速率一直是语音编码追求的目标.本文提出的 pGMM 算法将固定帧数的语音集合到一个片段中,并用多项式对该片段内的混合高斯模型参数进行拟合,将参数由高斯混合模型参数变成相应的多项式拟合参数,使得语音表示参数减少,编码比特数下降.实验结果表明,该编码方法在相同阶数不同帧数时,帧数越多编码器码率越低,解码语音质量越差.相同帧数不同阶数时,阶数越高编码器码率越高,解码

表 5 3 种编码器性能比较

Table 5 Performance comparison of three vocoder

编码器	码率/($\text{kb}\cdot\text{s}^{-1}$)	PESQ 得分	时延/ms
文[5]编码器	2.35	2.26	20
文[14]编码器	0.21	1.45	243
本文编码器	1.07	1.97	160

语音质量越好. 当多项式拟合阶数为 2 阶, 片段帧数为 8 帧时, 编码器的码率为 1.072 5 kb/s. 本文算法可以在大幅减少数据量的同时保持语音信号的主要特征.

[参考文献] (References)

- [1] 吴家安. 现代语音编码技术[M]. 北京:科学出版社,2008:287.
WU J A. Modern speech coding technology[M]. Beijing:Science Press,2008:287.(in Chinese)
- [2] DUSAN S, FLANAGAN J L, KARVE A, et al. Speech compression by polynomial approximation[J]. IEEE transactions on audio, speech, and language processing, 2007, 15(2): 387-395.
- [3] LAURENT F. Adaptive long-term coding of LSF parameters trajectories for large-delay/very-to ultra-low bit-rate speech coding[J/OL]. EURASIP journal on audio, speech, and music processing, 2010 [2016-08-10]. <http://hal.archives-ouvertes.fr/hal-00534492./DOI/10.1155/2010/597039>.
- [4] 张楠, 韩笑蕾, 张洋. 基于 MELP 算法的超帧结构状态统计及改进算法研究[J]. 电脑知识与技术, 2012, 8(16): 3 982-3 986.
ZHANG N, HAN X L, ZHANG Y. Studing on improved algorithms based on the statistics of super-frame structure state of the MELP algorithm[J]. Computer knowledge and technology, 2012, 8(16): 3 982-3 986.(in Chinese)
- [5] 李平, 曾毓敏. 基于 GMM 的甚低码率语音编码器[J]. 光电子技术, 2007, 27(3): 110-114.
LI P, ZENG Y M. A very low bit-rate vocoder based on GMM[J]. Photoelectronic technique, 2007, 27(2): 109-114.(in Chinese)
- [6] GHALEHJEH S H, ROSE R C. Linear regression based acoustic adaptation for the subspace gaussian mixture model[J]. IEEE transactions on audio, speech, and language processing, 2014, 22(9): 1 391-1 402.
- [7] HWANG H, TSAO Y, WANG H. Incorporating global variance in the training phase of GMM-based voice conversion[C]// Proceedings of APSIPA. Kaohsiung: IEEE Xplore, 2013: 1-6.
- [8] ÖZBEK Y, JOHNSON M H, DEMIREKLER M. Estimation of articulatory trajectories based on Gaussian Mixture Model (GMM) with audio-visual information fusion and dynamic kalmansmoothing[J]. IEEE transactions on audio, speech, and language processing, 2001, 19(5): 1 180-1 195.
- [9] SCHWARTZ B, GANNOT S, HABETS E A P. Online speech dereverberation using Kalman filter and EM algorithm[J]. IEEE/ACM transactions on audio, speech, and language processing, 2015, 23(2): 394-406.
- [10] 孙林慧, 杨震. 基于压缩感知的分布式语音压缩与重构[J]. 信号处理, 2010, 26(6): 824-829.
SUN L H, YANG Z. Distributed speech compression and reconstruction based on compressed sensing theory[J]. Signal processing, 2010, 26(6): 824-829.(in Chinese)
- [11] 刘慧婷, 张旻, 程家兴. 基于多项式拟合算法的 EMD 端点问题的处理[J]. 计算机工程与应用, 2004(16): 84-86.
LIU H T, ZHANG M, CHENG J X. Dealing with the end issue of EMD based on polynomial fitting algorithm[J]. Computer engineering and applications, 2004(16): 84-86.(in Chinese)
- [12] FORTUNE S A, HOPGOOD J R. Speech classification for enhancing single channel blind dereverberation[C]// Signal Processing Conference, 2008 16th European. Washington: IEEE, 2008: 1-5.
- [13] 邓峰, 鲍枫, 鲍长春. 基于 MPEG-AAC 编码器的压缩域音频增强方法[J]. 电子学报, 2014, 42(6): 1 410-1 418.
DENG F, BAO F, BAO C C. Audio enhancement in compressed domain based on MPEG-AAC codec[J]. Acta electronica sinica, 2014, 42(6): 1 410-1 418.(in Chinese)
- [14] CERNAK M, GARNER P N, LAZARIDIS A, et al. Incremental syllable-context phonetic vocoding[J]. IEEE/ACM transactions on audio, speech, and language processing, 2015, 23(6): 1 019-1 030.

[责任编辑: 严海琳]