

基于贝叶斯阴阳机的 2kb/s NMF-WI 语音编码算法

郭莉莉, 鲍长春

(北京工业大学电子信息与控制工程学院 语音与音频信号处理研究室, 北京 100124)

摘要: 本文提出了一种改进型的基于非负矩阵分解(Nonnegative Matrix Factorization, NMF)的特征波形(Characteristic Waveform, CW)分解算法,一方面应用惩罚次胜者竞争学习算法(Rival Penalized Competitive Learning, RPCL)和贝叶斯阴阳机(Bayesian Ying Yang, BYY)和谐学习算法,来计算 NMF 分解阶数,在没有明显降低语音质量的前提下,降低了编码器的复杂度;另一方面根据 CW 的能量与编码矩阵的能量间的变化关系,提出了相位谱的混合自回归合成方法,提高了语音的自然度.最后,开发出一套改进型 2kb/s NMF-WI 低复杂度语音编码方法,采用基于 K-L 散度的 NMF 迭代算法和收敛速度更快的基矢量 Mel 刻度分带初始化方法,按照基音周期的统计分布将特征波形分为 6 类,在 CW 分解模块,复杂度下降了 10MOPS,语音质量提高,与采用 4bit 散布矢量量化相位谱的 2.16kb/s NMF-WI 语音编码器的语音质量相当.

关键词: 语音编码; 波形内插; 特征波形; 非负矩阵分解

中图分类号: TN912.3 **文献标识码:** A **文章编号:** 0372-2112(2009)05-1146-08

2kb/s Bayesian Ying Yang Waveform Interpolative Speech Coding Based on Non-Negative Matrix Factorization

GUO Li-li, BAO Chang-chun

(Speech and Audio Signal Processing Laboratory, School of Electronic Information and Control Engineering, Beijing University of Technology, Beijing 100124, China)

Abstract: An improved characteristic waveform decomposition based on nonnegative matrix factorization was proposed. Two methods based on Bayesian Ying Yang(BYY) harmony learning and rival penalized competitive learning(RPCL) to compute factorization rank of nonnegative matrix factorization(NMF) were proposed. Computational complexity is decreased and speech quality is not decreased obviously. Mixed autoregressive model for construction of WI phase was proposed according to the energy of CW and coding matrix, which improves the naturalness. In the end, a low complexity NMF-WI speech coding at 2kb/s was developed. NMF based on Kullback-Leibler divergence and Mel scale band partitioning initialization used for basis vectors were proposed, and CWs were classified into six based on pitch distribution. In CW factorization, computational complexity dropped by 10 MOPS. Speech quality is increased, and equivalent to 2.16kb/s NMF-WI using 4bit phase VQ.

Key words: speech coding; waveform interpolation; characteristic waveform; non-negative matrix factorization

1 引言

波形内插语音编码模型作为当今最具潜力的低速率语音编码方案之一,因其良好的性能,越来越受到人们的重视^[1].在 WI 语音编码模型^[2-5]中, CW 分解在模型中起关键作用.最初的 WI 模型是通过 FIR 滤波将特征波形分解为类周期成分的慢渐变波形和类噪声成分的快渐变波形.但 FIR 滤波增加了一帧延时,同时分解精度也很难保证.目前也有人用基音同步小波变换和奇异值分解^[6,7]的方法来分解特征波形,但它们各自都有相应的局限性.文献^[8]提出了应用 NMF 分解 CW,并得到了基于 NMF 的波形内插 2kb/s 语音编码模型,但是分

解精度不高,运算复杂度有待降低.本文从模型的一些关键技术和相位合成的角度入手,改进了此编码模型,提高了运算精度,降低了计算复杂度.

2 非负矩阵分解的基本原理

非负矩阵分解的基本思想是:对于任意给定的一个非负矩阵 $V \in R^{n \times m}$, NMF 算法能够找到两个非负矩阵 $W \in R^{n \times r}$ 和 $H \in R^{r \times m}$,使这两个矩阵的乘积近似等于给定的矩阵 $V \approx WH$.由于分解前后的矩阵中仅包含非负的元素,因此,原矩阵中的列向量可以解释为对左矩阵中所有列向量的加权和,而权重系数为右矩阵中对应的列向量中的元素,分解得到的左矩阵 W 称为基矩阵

(Basis Matrix), 右矩阵 H 称为编码矩阵(Coding Matrix).

为了评价非负矩阵分解的性能, 定义一个评价 NMF 分解好坏的测度, 也叫代价函数(也称“目标函数”), Lee 和 Seung 提出了两种代价函数, 如式(1)(2)所示^[10]:

代价函数 1: 欧氏距离:

$$\|V - WH\|^2 = \sum_j (V_j - (WH)_{ij})^2 \quad (1)$$

代价函数 2: $K-L$ 散度:

$$D(V \| WH) = \sum_j \left[V_j \log \frac{V_j}{(WH)_{ij}} - V_j + (WH)_{ij} \right] \quad (2)$$

当且仅当 $V = WH$ 时, 式(1)(2)为零.

相应地, Lee 和 Seung 得到了如下的 NMF 迭代规则, 并证明了其收敛性^[10].

规则 1: 基于 $\|V - WH\|^2$ 的迭代规则(简称 ENMF):

$$H_{a\mu} \leftarrow H_{a\mu} \frac{(WV)_{a\mu}}{(W^T WH)_{a\mu}} \quad W_{ka} \leftarrow W_{ka} \frac{(VH^T)_{ka}}{(WHH^T)_{ka}} \quad (3)$$

规则 2: 基于 $D(V \| WH)$ 的迭代规则(简称 KNMF):

$$H_{a\mu} \leftarrow H_{a\mu} \frac{\sum_k W_{ka} V_{ik} / (WH)_{i\mu}}{\sum_k W_{ka}} \quad (4)$$

$$W_{ia} \leftarrow W_{ia} \frac{\sum_{\mu} H_{a\mu} V_{i\mu} / (WH)_{i\mu}}{\sum_{\mu} H_{a\mu}}$$

在 WI 语音编码器中, 残差信号通过傅氏级数表示得到非负的幅度谱, 特征波形沿相位轴展开、沿时间轴渐变, 一维残差信号转变成非负的二维特征波形幅度谱矩阵, 假设处理 $n \times m$ 维的 CW 幅度谱矩阵, 用 $V_{n \times m}$ 表示, 该幅度谱矩阵的各个元素都是非负值, 表示为 $V \succeq 0$, 对矩阵 $V_{n \times m}$ 进行线性分解, 用下式表示:

$$V_{n \times m} \approx W_{n \times r} H_{r \times m} \quad (5)$$

其中 $W_{n \times r}$ 为基矩阵, $H_{r \times m}$ 为编码矩阵, r 为分解阶数, 非负矩阵分解要求基矩阵和编码矩阵也都非负, 即 $W \succeq 0, H \succeq 0$. 文献[16]中给出了两种标准 NMF 算法和局部非负矩阵分解(LNMF)在绝对误差和计算复杂度方面的比较, KNMF 的收敛速度最快, 计算复杂度最低. 所以, 最终选择 KNMF 作为 CW 幅度谱分解算法.

3 NMF 分解 CW 的关键技术

为了改进原 NMF-WI 编码器的性能, 本节对一些关键技术进行阐述.

3.1 特征波形的分类

在 WI 语音编码器中, 基音周期设定在 20~120 之间变化. 由于 CW 的长度范围很大, 为了提高 NMF 分解

精度, 需要首先对特征波形进行分类, 本文采用文献[11]提出的根据基音周期的统计分布规律, 将 CW 非均匀分为 6 类的方法. 与文献[8]相比, 在相同的初始化条件下, 此方法没有增加算法的绝对误差, 且平均迭代次数从原来的 10 次, 下降为 7 次, 计算复杂度大大减少.

3.2 分解阶数的确定

在实际中我们仅仅知道所要分解的矩阵, 其真实的分解阶数隐藏于这些数据中, 并且无法从分解式确定出分解阶数. 目前, 人们还没有给出任何可行的非负矩阵分解阶数或模型选取准则^[17]. 若选取阶数较小, 其模型趋于简单, 但分解精度则会偏低. 反之, 若选取阶数较大, 其分解的精度提高, 但模型趋于复杂. 如何选择最优的分解阶数是非负矩阵分解中一个关键问题. 本节从聚类和自动模型选择两个角度入手, 提出了基于惩罚次胜者竞争学习(RPCL)算法和基于贝叶斯阴阳机(BYY)和谐学习算法的非负矩阵分解的分解阶数选取算法. 在非负矩阵分解中, 基矩阵的每一列都在某种程度上表示一个局部特征, 每个 CW 都由这些局部特征相加组成, 所以, 我们可以把这些表征局部特征的基矢量看成是 CW 矢量集的聚类, 那么求解 NMF 的分解阶数的问题就可以转化为求解聚类个数的问题. RPCL 是一种十分有效的聚类算法, 算法在聚类时能够自动地确定数据中的类别个数, 据此, 本文提出了一种基于 RPCL 算法的分解阶数选取算法.

RPCL 算法的基本思想是: 对于每一个输入的样本点, 不仅获胜单元的权向量得到学习(修正), 而且它的对手(非获胜者, 失败者)也要进行负学习. RPCL 算法将获胜单元吸引过来的同时, 将次胜者(失败者)推开. 也就是说, 每个类别只将一个权矢量吸引至它的类中心, 而组织较近的权矢量向它靠近, 同时, 以一定的惩罚力度使次胜者(失败者)远离类中心^[12].

设 V 为包含 N 个 CW 矢量的输入数据集, 记为: $V = \{v_1, v_2, \dots, v_k, \dots, v_N\}$, V 中第 i 个元素 v_i 是一个 p 维矢量 $v_i = (v_{i1}, v_{i2}, \dots, v_{ip})$. 初始化 RPCL 算法中有 m 个节点, 相应的基矢量为 ω_i , 其中 $(i = 1, \dots, m)$, 一般 $m > r$, r 为实际分解阶数, 第 i 个基矢量 ω_i 是一个 p 维矢量 $\omega_i = (\omega_{i1}, \omega_{i2}, \dots, \omega_{ip})$, 每个节点代表一个预先设置的类别, 节点的基矢量表示该类的中心. RPCL 算法过程可以描述如下:

第一步: 随机初始化基矢量 ω_i , $(i = 1, \dots, m)$ 并设迭代次数 $t = 1$;

第二步: 把数据集 V 中的每一个样本 v 分别对基矢量 ω_i 求

$$j_c \|v - \omega_c\|^2 = \min_j \|v - \omega_j\|^2 \quad (6)$$

$$j_d \|v - \omega_d\|^2 = \min_{j \neq c} \|v - \omega_j\|^2$$

其中 c 表示获胜单元, d 表示次胜(失败)单元, 标志位

$$\mu_i = \begin{cases} 1, & \text{if } i = c \\ -1, & \text{if } i = d \\ 0, & \text{其他} \end{cases} \quad (7)$$

或

$$\mu_i = \begin{cases} 1, & \text{if } i = c \\ -1, & \text{其他} \end{cases} \quad (8)$$

其中, v_c 和 v_d 分别为获胜基矢量和次胜(失败)基矢量的获胜频率, 基矢量 ω_j 的获胜频率 $y_j = \frac{n_j}{\sum_{i=1}^m \mu_i}$, n_j 是 μ_i

= 1 的次数, 基矢量 ω_i 获胜次数.

第三步: 修改基矢量 ω_i

$$\Delta \omega_i(t) = \begin{cases} \eta_c(v - \omega_i(t)), & \mu_c = 1 \\ -\eta_d(v - \omega_i(t)), & \mu_d = -1 \\ 0, & \text{其他} \end{cases} \quad (9)$$

或者

$$\Delta \omega_i(t) = \begin{cases} \eta_c(v - \omega_i(t)), & \mu_c = 1 \\ -\eta_d(v - \omega_i(t)), & \text{其他} \end{cases} \quad (10)$$

其中 $0 \leq \eta_c, \eta_d \leq 1$ 分别是竞争获胜者及次胜者(失败者)单元的学习率, 为了使算法收敛, 一般 $\eta_d \ll \eta_c$.

第四步: 迭代次数增加, $t = t + 1$, 若 $t < T$ 则转到第二步, 否则算法结束, 其中, T 是算法开始确定的权矢量修改的最大次数.

上述两种 RPCL 算法, 一种对竞争优胜者进行学习, 只对次胜者进行负学习, 其他的权矢量保持不变, 采用的公式为(6)(8)(9), 称其为 RPCL1 算法; 另一种对竞争优胜者进行学习, 对失败者全部进行负学习, 采用的公式为(7)(8)(10), 称其为 RPCL2 算法. 两种 RPCL 算法中惩罚次胜者的力度 η_d 应该远远小于奖励获胜者的力度 η_c , 否则算法不稳定. 但是如果惩罚次胜者的力度太小, 那么算法的效率就大大降低, 甚至不能排除多余的聚类中心, 因此, 根据实际情况需正确选取学习率. RPCL 算法是一个启发式算法, Ma, Wang&Xu^[13]对算法的特性进行了研究, 建立了一种价值函数理论, 并在此基础上证明了 RPCL 算法的收敛性.

图 1 给出了两种算法在处理不同长度 CW 时, 得到的分解阶数统计直方图. 从图中可以看出, 对于每一类 CW, RPCL1 算法得到的聚类个数不大于 RPCL2 算法, 这是因为在 RPCL2 算法中, 处理每个数据点时所有的基矢量都要移动, 不但算法复杂, 而且也使基矢量减弱了各自的属性特征; 相反, 在 RPCL1 算法中, 处理每个数据点时只需要移动距离该数据点最近的和次近的聚类矢量, 算法简单, 也可以很好的保持聚类矢量的固有属性特征. 而且, 随着 CW 长度的增加, 聚类的个数也增加, 即样本维数越大, 需要表征样本集的基矢量越多. 最后给出了混合 CW 的聚类结果, 若各类 CW 选取不同

分解阶数, 将导致很大的计算复杂度和空间复杂度, 编码器不可实现, 综合考虑, 以混合 CW 得到的分解阶数作为各类 CW 进行 NMF 的分解阶数. 在 RPCL1 算法中, 以 10210 个混合长度的 CW 为样本, 设初始聚类个数为 20, 经过充分迭代后, 得到空聚类的个数为 7, 非空聚类个数为 13. 所以, CW 进行 NMF 的分解阶数为 13.

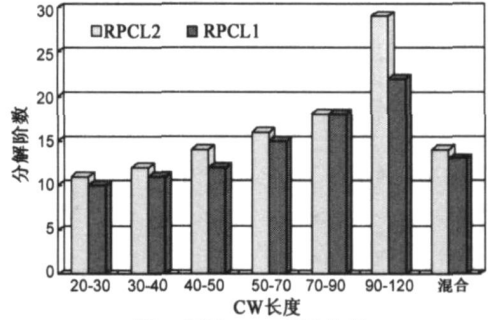


图1 两种RPCL算法比较

但是, RPCL 算法的鲁棒性较差, 受样本数据影响较大, 如果样本的类别不是很明显, 聚类的结果往往不稳定, 所以, 为了提高聚类的精度, 下面又提出了一种基于 BYY 的分解阶数选取算法.

NMF 的分解阶数反映了非负矩阵分解的规模或深度, 对应一个分解阶数就是一个非负矩阵分解模型, 因此, NMF 的分解阶数选择问题可以转化为一个模型选择问题. 贝叶斯阴阳机和谐学习算法是一种新颖的自动模型选择方法, 它是对联合概率密度分布的一种估计算法. 一个 BYY 系统通过两类贝叶斯分解的联合密度函数 $p(v, w) = p(v)p(w|v)$ 和 $q(v, w) = q(v|w) \cdot q(w)$ 来描述样本值 $v \in V \subset R^n$ 和与之对应的隐含信息 $w \in W \subset R^m$, 这两个联合密度函数分别称为阳机和阴机. NMF 算法将输入矩阵分解为基矩阵的线性组合, 对输入矩阵进行降维, 即从样本值中求解隐含信息. 在本文 v 表示 CW 矢量样本空间 V 的一个样本, 它在类空间 W 中的代表矢量就是 w , w 为隐含信息, 即 v 所属类中心. 给定一组 CW 矢量集 $D_c = \{v_j\}_{j=1}^N$, 一个 BYY 系统学习的目的是通过一个和谐学习原则来确定 $p(w|v), p(v), q(v|w), q(w)$, 此和谐学习原则通过最大化下面的函数来实现^[14]:

$$H(p || q) = \int p(w|v)p(v) \ln[q(v|w)q(w)] dv dw - \ln z_q \quad (11)$$

其中 z_q 为一个定值. 若 $p(w|v)$ 和 $q(v|w)$ 都是参数, 则称此 BYY 系统有 BF 方向的结构(BF 结构), 在 CW 矢量组成的有限混合模型中, 可以使用下面的 BYY 系统的 BI-结构. 令 $q(j) = \alpha_j$ 且 $\alpha_j \geq 0, \sum_{j=1}^k \alpha_j = 1$, 忽略定值 z_q

(设 $z_q = 1$), CW 矢量样本满足经验分布 $p_0(v) = \frac{1}{N} \sum_{i=1}^N$

$\delta(v - v_i)$. BI 结构可用参数形式表示:

$$p(w = j | v) = \frac{\alpha_j q(v | \theta_j)}{q(v, \Theta_k)}, q(v, \Theta_k) = \sum_{j=1}^k \alpha_j q(v | \theta_j) \quad (12)$$

$q(v | \theta) = q(v | w = j)$, 且 θ 包含所有参数, $\Theta_k = \{ \alpha_j, \theta_j \}_{j=1}^k$. 把式 (12) 带入式 (11) 得到:

$$H(p || q) = J(\Theta_k) = \frac{1}{N} \sum_{i=1}^N \sum_{j=1}^k \frac{\alpha_j q(v_i | \theta_j)}{\sum_{i=1}^k \alpha_i q(v_i | \theta_i)} \ln [\alpha_j q(v_i | \theta_j)] \quad (13)$$

由于 CW 足够多, 其分布可以看成高斯分布, v 满足分布 $q(v | \theta) = q(v | m_j, \Sigma_j)$

$$= \frac{1}{(2\pi)^{n/2} |\Sigma_j|^{1/2}} e^{-\frac{1}{2}(v - m_j)^T \Sigma_j^{-1} (v - m_j)} \quad (14)$$

其中 m_j 为均值矢量, Σ_j 为方差矩阵, 此时, BYY 系统的 BF 结构包括 CW 矢量集 D_v 的概率分布的高斯混合模型

$q(v, \Theta_k) = \sum_{j=1}^k \alpha_j q(v | m_j, \Sigma_j)$. $\alpha_j = e^{\beta_j} / \sum_{i=1}^k e^{\beta_i}$, $j = 1, \dots, k$, $-\infty < \beta_1 \dots \beta_k < \infty$, $\Sigma_j = B_j B_j^T$, B_j 是一个非奇异方阵. 推导得到以下迭代规则^[15]:

$$\begin{aligned} \Delta \beta_j &= \eta \frac{\alpha_j}{N} \sum_{i=1}^N \sum_{l=1}^N h(il | v_i) U(il | v_i) (\delta_j - \alpha_j) \\ \Delta m_j &= \eta \frac{\alpha_j}{N} \sum_{i=1}^N h(jl | v_i) U(jl | v_i) \Sigma_j^{-1} (v_i - m_j) \\ \Delta B_j &= \eta \frac{\alpha_j}{N} \sum_{i=1}^N h(jl | v_i) U(jl | v_i) [\Sigma_j^{-1} (v_i - m_j) (v_i - m_j)^T \Sigma_j^{-1} - \Sigma_j^{-1}] B_j \end{aligned} \quad (15)$$

$$\begin{aligned} U(il | v_i) &= \sum_{l=1}^k [\delta_l - p(l | v_i)] \ln \alpha_l p(v_i | \theta_l) + 1 \\ h(il | v_i) &= \frac{p(v_i | \theta_i)}{\sum_{l=1}^k \alpha_l p(v_i | \theta_l)} \end{aligned} \quad (16)$$

$p(il | v_i) = \alpha_l h(il | v_i)$

为了使基于 BYY 的分解阶数选取算法收敛更快, 结果更准确, 以 RPCL 算法得到的分解阶数和聚类中心为初始值, 通过基于贝叶斯阴阳机和諧学习原则的自动模型选择学习算法, 迭代得到聚类中心的分布函数的均值和方差. 如果某个聚类中心包含的样本个数很少, 甚至趋近于 0, 则舍弃此类别, 得到最终的分解阶数. 将不同长度的 CW 进行实验, 结果与 RPCL1 算法稍有不同, 偏差一般在 3 之内. 将混合长度的 CW 作为样本进行实验, 初始化聚类个数 13, 聚类中心是 RPCL1 算法的结果, 经过迭代计算, 得到的聚类个数是 13. 最终选取的分解阶数为 13.

表 1 不同分解阶数的算法比较

分解阶数	绝对误差	计算复杂度
(均匀分带初始化) 16	2.543648	$7 * (2mn + mn + m)$
(Mel 分带初始化) 13	2.579308	$4.8 * (2mn + mn + m)$

将本文提出的算法与文献[8]进行比较, 如表 1 所示, 从表中可以看出, 即使分解阶数减少, 但是绝对误差变化不大, 仅仅增加了 0.03, 而平均迭代次数从原来的 7 次下降到现在的 4.8 次, 计算复杂度大大降低.

3.3 分带初始化

在 NMF 中, 基矩阵和编码矩阵必须进行初始化, 本节讨论的初始化都是针对基矩阵 W 的初始化. 初始化的好坏直接影响 NMF 分解的效果, 好的初始化算法可以提高分解精度, 加快收敛速度, 降低复杂度.

本文比较了三种初始化算法, 如图 2 所示, 以第 6 类特征波形为例, 迭代次数为 1000 次, 保证收敛. 从图

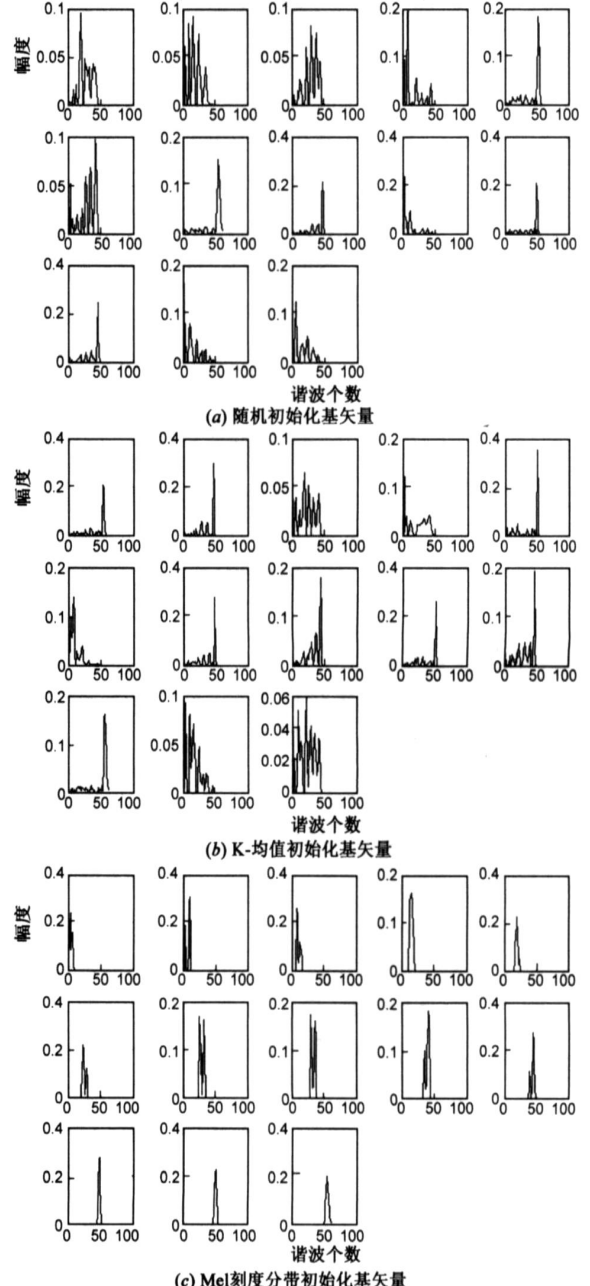


图 2 K-均值和 Mel 刻度分带初始化的基矢量比较

中可以看出, 随机初始化虽然简单, 但是迭代后的基矢量没有很好的表征语音特征波形幅度谱特性; 基矢量经过基于 K -均值初始化后, 基矢量间的分布规律性差, 重叠严重, 没有很好的去除基矢量间的冗余; 虽然训练后得到的基矢量稀疏性变强, 但是仍不能很好地反映“基”的概念; Mel 刻度分带初始化的基矢量更加稀疏, 基矢量之间有效频带的重叠较小, 彼此冗余较小, 作为“基”的概念更加明显. 表 2 给出了三种初始化算法的比较, 均采用 KNMF 算法, CW 分为 6 类. 从表中可以看出: Mel 刻度分带初始化方法绝对误差最小, 平均迭代次数最少, 计算复杂度最低.

表 2 初始化算法比较

算法	绝对误差	计算复杂度
Mel 分带初始化	2.529989	$5 * (2mn + mn + m)$
K -均值	2.624936	$6.2 * (2rmn + mn + rm)$
随机初始化	2.579308	$7 * (2mn + mn + m)$

3.4 CW 非负矩阵分解的复杂度

在 CW 的 NMF 分解过程中, 非负矩阵分解的初始化、CW 的分类、分解阶数的选取、基矩阵的训练都是为了得到更具有代表性的基矢量, 其计算复杂度不计入分解复杂度. 本文采用的方法是 CW 分为 6 类、分解阶数为 13, 基矢量的 Mel 刻度分带初始化的 KNMF 方法, 文献[8, 9]中采用的方法是 CW 分为 9 类、分解阶数为 16, 基矢量均匀分带初始化的 ENMF 方法, 原始 WI 算法采用的 CW 分解方法是 FIR 低通滤波, 三种方法的计算复杂度比较, 如表 3 所示.

表 3 不同分解方法的计算复杂度比较 单位: WMOPS

分解算法	KNMF	ENMF	FIR 低通滤波
计算复杂度	8.376	18.878	9.697

从表中看出, KNMF 计算复杂度最低, 比 ENMF 方法降低了 56%, 比 FIR 滤波降低了 1.32 个 MOPS, 而且, NMF-WI 编码器去掉了对齐操作, 计算复杂度大大降低.

4 编码矩阵的量化与相位的合成

4.1 编码矩阵的分裂式矢量化

在 WI 编码模型中, 提取的特征波形矩阵最大维数是 61×10 , 经过非负矩阵分解后得到的编码矩阵的维数为 13×10 , 然后将此编码矩阵进行量化.

基于 Mel 刻度的分带初始化方法训练出来的基矢量按照从低频到高频的顺序排列, 而待分解的 CW 矢量由这些基矢量加性组合而成, 组合系数是编码矩阵对应的列. 从这个意义上讲, 编码矩阵的低维用来组合 CW 的低频信息, 高维用来组合 CW 的高频信息. 我们知道, 对语音质量而言, 语音信号的低频信息比高频信息重要, 在低速率语音编码中, 应尽量保证低频信息不要丢失, 所以, 我们采用较精确量化编码矩阵的低维

粗糙量化高维的方法. 具体方法如下: 首先将编码矩阵进行降维, 采用的方法是抽取, 即每隔一个 CW 抽取一个, 将 13×10 的编码矩阵变为 13×5 , 然后分裂为三个较小的矩阵 4×5 , 4×5 , 5×5 分别进行量化.

4.2 相位谱的混合自回归合成

由于在低速率语音编码中, 比特数有限, 很多低速率编码器采用不传输相位信息在合成端直接合成的方法. 在原始 WI 模型中, CW 通过 FIR 滤波, 被分解为表示类周期成分的慢渐变波形 (Slowly Evolving Waveform, SEW) 和表示非周期成分和噪声成分的快渐变波形 (Rapidly Evolving Waveform, REW), SEW 采用固定相位合成, REW 采用随机相位合成, 最后得到了较好的合成语音质量. 因此, 在一定编码速率下, 实际相位中固定相位和随机相位的比例取决于原 CW 的清浊程度, 本文引入了周期因子来描述 CW 的清浊程度. CW 的能量与其浊音性近似呈正比, 即能量大的 CW 的浊音性较强, 反之, 能量相对小的 CW 的浊音性较弱, 因此, 可以用 CW 的能量来表征其清浊程度, 在此基础上, 本文提出了一种相位谱的混合自回归合成方法.

周期因子是从矩阵能量的角度引入的, 如果假设 v_j 和 h_j 分别是矩阵 W 和 H 所对应的列向量, 则式(5)还可以表示为: $v_j = Wh_j$, 则原始 CW 幅度谱矢量的能量为其 Frobenius 范数的平方:

$$\|v_j\|_F^2 = \|Wh_j\|_F^2 \leq \|W\|_F^2 + \|h_j\|_F^2 \quad (17)$$

周期因子定义为: $\beta_{peri} = \frac{\|h_j\|_F^2}{G}$, G 为随机选取的 10000 个 CW 的能量的平均值或最后一子帧的能量.

将周期因子归一化得:

$$\beta_{peri} = \frac{\|h_j\|_F^2}{\|h_j\|_F^2 + G} \quad (18)$$

在 NMF 过程中 $\|W\|_F^2 = 1$, 则 v_j 与 h_j 的能量变化趋势基本一致, 即 CW 的能量可以用对应的编码矩阵的能量来近似表示. 也可以通过实验来分析编码矩阵的能量分布与 CW 的能量之间的关系. 以第 6 类 CW 为例, 图 3 给出了通常情况下编码矢量的能量与 CW 的能

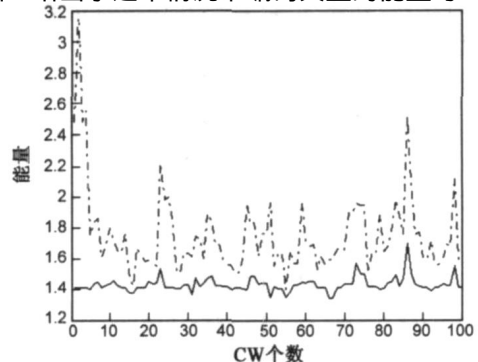


图 3 编码矢量与 CW 的能量分布比较

量的分布比较图. 其中, 虚线表示编码矢量的能量, 实线表示对应特征波形的能量. 可以发现, 编码矢量的能量与 CW 的能量变化趋势基本一致. 综上所述, CW 的浊音性与编码矢量的能量分布情况密切相关, 因此, 周期因子可用编码矢量的能量来定义.

周期因子利用编码矩阵的能量分布规律描述了 CW 的清浊程度, 在低速率语音编码中, 解码端可以用周期因子控制随机相位和固定相位的混合比例, 合成最终的相位信息. 为了保持相位谱的连续性, 将前一子帧的相位谱反馈, 对当前子帧进行加权, 得到当前子帧的相位谱:

$$\Psi_m(k) = \rho \Psi_{m-1}(k) + (1 - \rho) (\beta_{peri} Q_m(k) + (1 - \beta_{peri}) R_m(k)) \quad (19)$$

其中 $R_m(k)$ 为随机矢量, 取值在 $[-\pi, \pi]$ 内按均匀分布随机产生(各帧相位不同)的相位, $Q_m(k)$ 是固定矢量, 用老年男子语音中提取的相位, ρ 为反馈系数, 混合自回归相位合成的原理框图如图 4 所示.

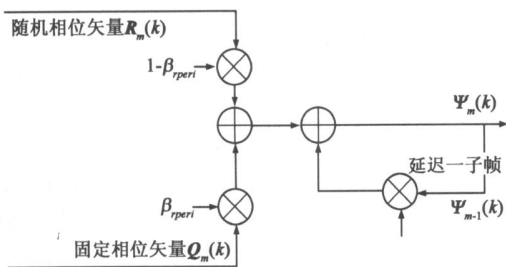


图4 混合自回归合成原理框图

1999 年 Oded Gottesman 提出了分析合成(AbS) 散布相位矢量量化^[16]方案, 它结合了感觉加权, 在 WI 模型中的 SEW 相位合成中得到了较好的结果. 在 NMF-WI 模型中, 我们将散布相位矢量量化应用于整个特征波形的相位合成. 每帧提取 10 个特征波形, 每个 CW 都有一个相位矢量, 由于比特数有限, 不能量化所有 CW 的相位, 每帧提取一个相位矢量, 在最后一个采样点处, 在合成端, 采用三次多项式插值的方法, 插出中间各子帧的相位. 实验发现采用 4~ 6 比特量化 SEW 相位谱, 效果已经较令人满意^[17]. 所以, 本文采用 4bit 量化相位.

将混合自回归相位合成应用到 NMF-WI 编码模型中, 以女声语音“我到无锡市”为例, 与固定相位合成、4bit 相位散布矢量量化后的语音波形对比如图 5. 如图可知, 本文提出的混合自回归相位合成方法得到的语音比固定相位合成的语音自然度高, 虽然引入了一些噪声, 但刺耳的蜂鸣声减弱了. 总体来讲, 合成相位的误差都比较大, 矢量量化相位的结果要比直接合成的效果好, 浊音段比较清晰. 从听音结果看, 混合自回归相位合成方法与矢量量化相位方法得到的语音自然度相当, 前者稍有一些噪声, 这与周期因子、反馈因子的选取不无关系.

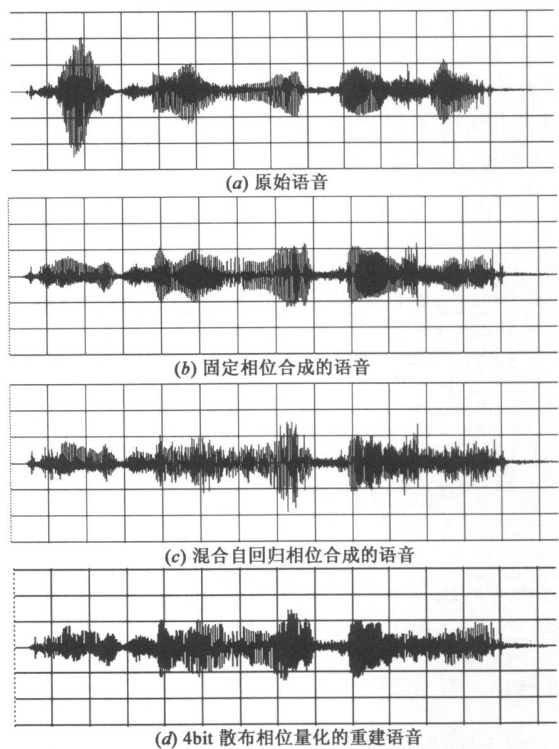


图5 不同相位合成的语音图对比

5 NMF-WI 编码方案的性能评价

3.4 两节主要介绍了基于非负矩阵分解的特征波形的分解和量化, 现在介绍本编码器其它参数的量化方法: LSF 参数采用预测分裂式矢量量化方法 (PSVQ)^[11]; 基音周期采用 7 比特的均匀标量量化^[18]; 由于 CW 功率的动态范围很宽, 于是先转换为对数功率, 然后采用插值量化的方法^[19, 20]. 表 4 给出了改进型 2kb/s NMF-WI 编码器的比特分配情况.

表 4 2kb/s NMF-WI 编码器的比特分配

量化参数	更新速率(Hz)	比特分配
LSF	40	20
基音周期	40	7
CW 功率	40	7
低频矩阵	40	7
中频矩阵	40	5
高频矩阵	40	4

综上, 本文基于特征波形的非负矩阵分解技术, 对传统的 NMF-WI 语音编码器进行了改进, 改进后的 2kb/s NMF-WI 编码器框图示于图 6. 其中根据非负矩阵分解的特点, 编码端去除了特征波形对齐模块. 加上虚线框的相位量化模块, 则得到 2.16kb/s NMF-WI 编码器.

本文提出 2kb/s NMF-WI 编码器与加入相位量化的 2.16kb/s NMF-WI 编码器和文献[8, 9] 中提出的编码器一起进行了 MOS 分测试, 并进行了性能对比. 测试所采用的语音数据由 16 句标准的汉语语音测试句组成, 包

括 8 句男声语音和 8 句女声语音, 采样频率均为 8kHz. 测试小组由 5 名年龄在 20~30 岁之间的青年人员组成. 表 5 是所有语音的 MOS 分测试结果. 通过测试结果得知, 本文提出的 2kb/s NMF-WI 算法的语音质量优于原编码算法, 与 2.16kb/s NMF-WI 相位量化算法相当.

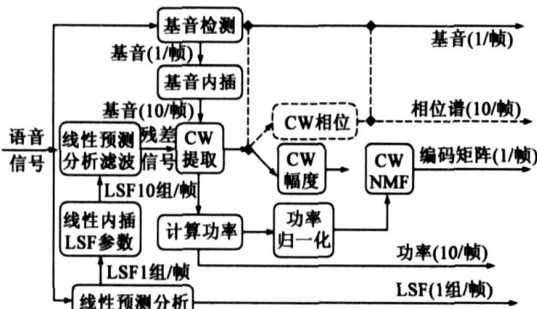


图6 改进型NMF-WI编码器

表5 MOS分测试结果

算法	原 2kbps NMF WI	改进 2kbps NMF WI	2.16kbps NMF WI
MOS 分	3.004375	3.033125	3.0475

6 结论

综上所述, 本文基于文献[8], 提出了一种改进型基于非负矩阵分解的特征波形分解与量化模型. 本文主要从两方面进行改进, 一方面提高语音质量, 在原 NMF-WI 编码器^[8]中, 采用相同的固定随机相位, 相位的相关性较强, 合成语音引入了较强的蜂鸣声, 语音的音质很低. 本文将相位谱分解为随机相位和固定相位的加权和, 根据 CW 的能量与编码矩阵的能量间的变化关系, 引入周期因子控制加权系数, 提出了一种相位谱的混合自回归合成方法, 有效改善了合成语音有很强的蜂鸣声的缺点, 提高了语音的自然度. 另一方面降低复杂度, 从聚类维数和模型选择两个角度入手, 提出了基于惩罚次胜者竞争学习算法(RPCL)和基于贝叶斯阴阳机(BYY)和谐学习算法, 两种计算 NMF 分解阶数的方法, 增强了原来实验方案的理论基础, 没有明显降低语音质量, 编码器的复杂度下降.

本文比较的三个 NMF-WI 编码器, 缺陷依然在于重建语音不够自然, 有些蜂鸣声. 所以, 下一步工作应重点考虑如何对 CW 相位谱进行高效合理化.

参考文献:

- [1] 鲍长春. 数字语音编码原理[M]. 西安: 西安电子科技大学出版社, 2007. 220-262.
- [2] Kleijn W. B, Haagen J. Waveform Interpolation for Coding and Synthesis. Speech coding and Synthesis[M]. Elsevier Science, 1995, 175-207.
- [3] Changchun Bao, Jing Li, Fengyan Qi. High quality waveform interpolation speech coding at 2kb/s[J]. Chinese Journal of Electronics, 2007, 16(2): 257-262.
- [4] Fengyan QI, Changchun BAO. A simplified algorithm for representation and alignment of the characteristic waveform in WI speech coding[A]. 8th International Conference on Signal Processing(ICSP'06) [C]. Guilin, 2006: 646-649.
- [5] Li Jing, Bao Changchun. Quantization of SEW and REW magnitude for 2kb/s waveform interpolation speech coding[A]. The 4th International Symposium on Chinese Spoken Language Processing(ISCSP2004) [C]. Hong Kong, 2004. 12, 141-144.
- [6] 王贵平, 鲍长春, 张鹏. 基于奇异值分解的低速率波形内插语音编码算法[J]. 电子学报, 2006, 36(1): 135-140. Guiping Wang, Changchun Bao, Peng Zhang. Low bit rate waveform interpolation speech coding based on SVD[J]. Acta Electronica Sinica, 2006, 36(1): 135-140. (in Chinese)
- [7] 张鹏, 鲍长春. 基于 SVD 的低复杂度语音特征波形分解方法[J]. 信号处理, 2005, 21(4A): 160-163. Changchun Bao, Peng Zhang. The decomposition of speech characterization waveforms with low complexity based on SVD[J]. signal processing, 2005, 21(4A): 160-163. (in Chinese)
- [8] 张鹏, 鲍长春, 郭莉莉. 基于非负矩阵分解的 2kb/s 波形内插语音编码算法[J]. 电子学报, 2008, 36(4): 632-638. Peng Zhang, Changchun Bao, Lili Guo. 2kb/s waveform interpolation speech coding based on nonnegative matrix factorization[J]. Acta Electronica Sinica, 2008, 36(4): 632-638. (in Chinese)
- [9] Peng Zhang, Changchun BAO. A novel 2kb/s waveform interpolation speech coder based on non-negative matrix factorization[A]. Interspeech [C]. Antwerp, Belgium, 2007. 1661-1664.
- [10] D D Lee, H S Seung. Learning the parts of objects by non-negative matrix factorization[J]. Nature, 1999, 401: 788-791.
- [11] 郭莉莉, 鲍长春. 基于非负矩阵分解的语音特征波形分解方法[J]. 信号处理, 2007, 23(4A): 107-110. Lili Guo, Changchun Bao. Decomposition of speech characteristic waveform based on NMF[J]. Signal Processing, 2007, 23(4A): 107-110. (in Chinese)
- [12] Xu L, Krzyzak A, Oja E. Rival penalized competitive learning for clustering analysis, RBF Net and Curve Detection[J]. IEEE Transaction Neural Networks, 1993, (4): 636-648.
- [13] Jinwen Ma, Taijun Wang. A cost function approach to rival penalized[J]. IEEE Transaction on Systems, August 2006, 36(4): 722-737.
- [14] Jinwen MA, et al. Conjugate and natural gradient rules for BYY harmony learning on gaussian mixture with automated model selection[J]. International Journal of Pattern Recognition and Artificial Intelligence, 2005, 19(5): 701-713.
- [15] Jinwen MA. Automated Model Selection(AMS) on finite mix

ture: a new perspective for data modeling[J]. Chinese Journal of Engineering Mathematics. 2007(8) , 24(4) : 571- 583.

- [16] Gottesman. Dispersion phase vector quantization for enhancement of waveform interpolative coder[A]. Proceedings on IEEE International Conference on Acoustics, Speech, and Signal Processing[C]. Arizona, 1999. 269- 272.
- [17] 陈悦, 鲍长春. 一种用于 WI 语音编码的相位预测式矢量量化方法[J]. 电子信息学报, 2007, 29(11): 2672- 2675. Yue Chen. Changchun Bao. A predictive phase vector quantization method in WI speech coding[J]. Journal of Electronics & Information Technology, 2007, 29(11): 2672- 2675. (in Chinese)
- [18] 罗亚飞, 鲍长春. 基于 DCT 分带谱熵与信号分解的高精度基音检测算法[J]. 电子学报, 2007, 35(1) : 13- 22. Yafei Luo. Changchun Bao. Super resolution pitch detection based on band partitioning spectral entropy and signal decomposition in DCT domain[J]. Acta Electronica Sinica, 2007, 35(1) : 13- 22. (in Chinese)
- [19] Li Jing, Bao Changchun. A 2kb/s enhanced waveform interpolation speech coder[A]. 2004 7th International Conference on Signal Processing[C]. Beijing: Publishing House of Electronics Industry, 2004, 8: I- 598- I- 601.

- [20] 李靓, 鲍长春. 波形内插语音编码器中快渐变波形幅度的量化[J]. 信号处理, 2004, 20(6) : 545- 548. Jing Li, Changchun Bao. Quantization of REW magnitude in waveform interpolation speech coding[J]. Signal processing, 2004, 20(6) : 545- 548. (in Chinese)

作者简介:



郭莉莉 女, 1981 年生于河北唐山, 硕士研究生, 研究方向为语音信号处理、窄带语音编码。



鲍长春 男, 1965 年生于内蒙古赤峰, 博士, 教授、博士生导师, 国际语音通信学会(ISCA) 会员, 中国电子学会理事, 信号处理学会委员, 《通信学报》编委会副主任委员, 《信号处理》和《数据采集与处理》编委. 主要研究领域为语音与音频信号处理及编码等.
E mail: chcbao@bjut. edu. cn

电子学报

2009 年第 5 期 Acta Electronica Sinica No. 5 2009

(总第 313 期) (Monthly) (Series No. 313)

主管单位	中国科学技术协会	China Association for Science and Technology
主办单位	中国电子学会	Published by the Chinese Institute of Electronics, Beijing
编辑	《电子学报》编辑委员会	Edited by Editorial Board of Acta Electronica Sinica
主编	王守觉	Chief Editor: WANG Shou jue
总编辑	刘力	Director: LIU Li
通信处	北京 1 6 5 信箱 (邮政编码 100036)	Add: Editorial Office of Acta Electronica Sinica (P O Box 165, Beijing 100036, China)
电话	(010) 68279116, 68285082	Tel: 86- 10- 68279116, 68285082
传真	(010) 68173796	Fax: 86- 10- 68173796

Home page: <http://www.ejournal.org.cn>; <http://dzxu.chinajournal.net.cn>

Email: new@ejournal.org.cn; wanghui@ejournal.org.cn

排版印刷 北京墨禹天成印刷有限公司 Printed by Moyutiancheng Co. Ltd., Beijing, China

国内总发行 北京市报刊发行局 Distributed by

Domestic: Beijing Baokan Faxingju, China

国外总发行 中国国际图书贸易总公司 Foreign: China International Book Trading Corporation

国内订购处 全国各邮电局 Subscription Office —All Local Post Offices in China

国内统一刊号: CN11- 2087/TN 邮发代号(国内/ 国外) : 2- 891/M436 国内定价 ¥ 40. 00

© 1994-2010 China Academic Journal Electronic Publishing House. All rights reserved. <http://www.cnki.net>