

基于非负矩阵分解的 2kb/s 波形内插语音编码算法

张 鹏, 鲍长春, 郭莉莉

(北京工业大学电子信息与控制工程学院, 北京 100022)

摘 要: 在波形内插(Waveform Interpolation, WI)语音编码器中,如何低延时、高精度并且低复杂度的分解和量化特征波形(Characteristic Waveform, CW)一直是该编码模型的研究热点和难点.本文提出用非负矩阵分解(Nonnegative Matrix Factorization, NMF)方法来分解语音特征波形.该分解方法仅需要当前帧的语音信号,不会给编码器带来额外的延时;为了提高分解精度,本文在CW分解之前先对CW按照其子帧的最大基音周期进行分类,然后按不同类别进行分解.另外,本文结合耳蜗模型提出了NMF的基矢量分带初始化算法,将CW的分解精度提高到与二阶奇异值分解相当的水平;为了降低WI编码器的计算复杂度,本文去除了传统WI编码器中的特征波形对齐模块,同时将NMF的分解阶数设定为16以折中CW分解的计算复杂度和分解精度.最后,本文基于矩阵量化技术,对非负矩阵分解后的编码矩阵采用分裂式矩阵量化方案来量化.主观A/B测试表明,本文提出的2kb/s NMF2WI编码器的合成语音质量接近于2.4kb/s SVD2WI编码器.MOS分测试表明,本文提出的2kb/s NMF2WI编码器的合成语音质量稍差于2.4kb/s MELP编码器.

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: TN91213 文献标识码: A 文章编号: 0372-2112 (2008) 04-0632-07

2kb/s Waveform Interpolation Speech Coding Based on Nonnegative Matrix Factorization

ZHANG Peng, BAO Changchun, GUO Lili

(School of Electronic Information and Control Engineering, Beijing University of Technology, Beijing 100022, China)

Abstract: In WI coding scheme, how to decompose and quantize the characteristic waveforms with low delay, low complexity and high precision have always been a hot research topic. The characteristic waveform decomposition based on nonnegative matrix factorization is proposed in this paper. This CW decomposition method doesn't bring any additional delay to WI coder; In order to improve decomposition precision, the CW is firstly classified according to the maximum pitch of its subframes before being decomposed. Besides, band-partitioning initialization constraints are set to basis vectors before NMF is carried out, and this has made the CW decomposition precision of NMF-based method be comparable with that of 2 ranks of SVD; In order to reduce the computational complexity of WI coder, the CW alignment procedure is removed in our NMF2WI coder, and the factorization rank of NMF is set to 16 as a tradeoff between computational complexity and decomposition precision. In the end, the low dimensional coding matrix is quantized by splitting matrix quantization scheme. The subjective A/B listening tests show that the proposed 2kb/s NMF2WI coder can give smooth speech with quality close to 2.4kb/s SVD2based WI coder. Mean Opinion Score test results indicate that the performance of proposed coder is a little worse than that of 2.4kbps MELP coder.

Key words: speech coding; waveform interpolation; characteristic waveform; nonnegative 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,以其良好的性能,越来越受到人们的重视.波形内插语音编码技术最先是由瑞典皇家理工学院的Kleijn教授在1991年提出的,称为原型波形内插(Prototype Waveform Interpolation2PWI)语音编码算

法^[1,2]. PWI在310~410kb/s可获得好的重建语音质量,但它仅适用于浊音语音,清音语音还需用其他方法(如CELP)来编码.为了进一步降低编码速率,Kleijn教授又提出了特征波形分解语音编码算法,称为特征波形内插(Characteristic Waveform Interpolation2CWI)语音编码算法^[3,4].该算法将语音信号表示为在残差域渐变的特征

波形,用简单的非因果低通滤波器将特征波形分解为慢渐变波形(Slowly Evolving Waveform2SEW)和快渐变波形(Rapidly Evolving Waveform2REW),基于主观听觉,对这两种波形分别编码.该方法避免了传统声码器清浊判决带来的影响,适用于所有的语音段^[5,6].

在波形内插语音编码中,如何低延时、低复杂度、高精度的分解与量化特征波形一直是该编码方案的研究热点和难点.迄今已有多种成功的特征波形分解方案:简单线性相位低通滤波,小波变换分解,奇异值分解,时间分解等.目前,线性相位低通滤波是实现特征波形分解的最简单有效的方法之一,即沿着时间轴,将二维特征波形表面分解成 SEW 和 REW. SEW 表示了语音信号的类周期成分,REW 表示了语音信号的类噪声成分.为了更加精细的分解特征波形,Whisper 实验室的 Chong 等利用基音同步小波变换(Pitch Synchronous Wavelet Transform2PSWT)^[7]分解 CW.这种方法得到了更多波形渐变的细节,对不同感知特性的波形表面进行变维矢量量化,但是编码器总延时比较大,很难应用于实时通信.另外,该实验室的 Lukasiak 等研究者提出了将奇异值分解(Singular Value Decomposition2SVD)理论用于 CW 的分解^[8].此方法对当前帧的特征波形表面进行分解,不需要额外的延时.但因为奇异值分解的计算复杂度较高,并且需要分配较多的比特来量化分解后大量的待量化参数,因此限制了其在低速率语音编码中的应用.为此,国内有学者基于 SVD 提出了低复杂度的 2.4kb/s SVD2WI 编码器^[9~11],该编码器的合成语音质量好于 2.4kb/s MELP 编码器的质量^[9~11].最后,Whisper 实验室的 Ritz 等人采用时间分解^[12]方法利用事件函数和事件目标函数模拟谱参数的渐变,极大地降低了谱参数的冗余度.但是,如何将时间分解方法更好地运用于 WI 编码需要进一步的研究.

上述四种分解方案的共同之处在于将语音信号的特征波形分解为类周期成分和类噪声成分,所以他们都都可以看作是对特征波形的基于/整体0的描述.本文将利用非负矩阵分解^[13]来分解特征波形,这是一种对特征波形基于/部分0而非基于/整体0的分解方式.在语音识别领域,国际上已有学者^[14]利用 NMF 来提取语音特征参数,并取得很好的实验结果,而将 NMF 应用于语音编码尚无研究先例.本文用 NMF 来分解 CW 的基本思路是通过大量的 CW 实验样本训练出语音频谱的/部分0)即语音频谱的谐波结构,然后对于待编码的某帧 CW,通过线性组合这些已经训练好的各次谐波来重建其频谱.不难发现,这种分解方法在编码端只需对刚刚提到的线性组合系数进行编码,而训练出来的频谱/部分0(或称语音频谱的谐波结构)直接保存在解码端即可.这种 CW 分解方法相对于简单线性相位低通滤

波和小波变换来说,优势在于去掉了额外延时;相对于奇异值分解来说,优势在于降低了计算复杂度同时又不降低分解精度.

本文在第 2 节介绍了非负矩阵分解的基本原理,第 3 节介绍了语音信号频谱的/部分0的概念,第 4 节论述了特征波形非负矩阵分解的详细步骤及具体方法,第 5 节讨论了编码矩阵的分裂式矩阵量化,第六节给出了 NMF2WI 编码方案的性能评价,最后是本文的结论.

2 非负矩阵分解的基本原理

NMF 的基本算法可以简单描述为:对于任意给定的一个非负矩阵 $V \in R^{n \times m}$,NMF 通过有限次迭代总能够寻找到一个非负矩阵 $W \in R^{n \times r}$ 和一个非负矩阵 $H \in R^{r \times m}$,使得满足 $V \approx WH$,从而将一个非负的矩阵分解为左右两个非负矩阵的乘积.这里的 r 就是非负矩阵分解的分解阶数,在满足分解精度的情况下,我们总是期望它越小越好.由于 NMF 对分解后的两个矩阵都作了非负限制,这意味着在合成原始矩阵时只存在加法的组合,所以就导致了基于/部分0而非基于/整体0的表征方式.即原矩阵 V 的某个列向量可以解释为对左矩阵 W (又名/基矩阵0)中所有的储存着/部分0的稀疏列向量(又名/基向量0)的加权和,而权重系数为右矩阵 H (又名/编码矩阵0)中对应列向量中的元素.

为了找到满足 $V \approx WH$ 的非负矩阵 W, H , 我们首先选取 V 和 WH 之间欧氏距离的平方作为目标函数,如式(1)所示.其中, i, j 依次表示相应矩阵的行、列下标.

$$\|V - WH\|^2 = \sum_{i=1}^60 \sum_{j=1}^{10} (V_{ij} - (WH)_{ij})^2 \quad (1)$$

Lee 和 Seung 在文献[15]中证明了式(1)定义的目标函数在如下的 NMF 算法迭代过程中逐渐变小并趋于收敛.这里的 W 和 H 都初始化为随机的非负矩阵.式(2)中的 a, L 依次表示相应矩阵的行、列下标,式(3)中的 i, a 依次表示相应矩阵的行、列下标,符号/ z 0 表示把/ z 0 右边的矩阵元素赋给/ z 0 左边对应位置的矩阵元素,即完成了该矩阵各个元素的/更新0操作.

$$H_{aL} z H_{aL} \frac{(W^T V)_{aL}}{(W^T W)_{aL}} \quad (2)$$

$$W_{ia} z W_{ia} \frac{(VH^T)_{ia}}{(WH^T)_{ia}} \quad (3)$$

3 语音频谱的/部分0

对于一般意义上的人脸图像来说,它的/部分0很显然指的是不同版本的嘴、鼻子、眼睛、眉毛等.那么,对于语音信号的频谱来说,其/部分0指的是什么呢?文献[16]指出,人的内耳是由耳蜗和耳蜗神经组成的.

其中耳蜗被基膜分割成一系列充满了流体的腔室. 声波引起这些腔室中的流体进行不同频率的振动, 这些振动又引起基膜的不同位置按照其对应的频率而振动. 所以, 基膜可以视为是不同位置(或不同部分)响应不同频率的一种组织结构, 从信号处理角度来说, 基膜也可以被模型化为一组具有不同中心频率的带通滤波器^[16]. 从上面的分析中我们可以认为, 人耳对语音信号的听觉感知在频域中是基于/部分0的, 而语音信号频谱的/部分0就是频率轴上类似于脉冲结构的各次谐波. 而 WI 编码器中的特征波形在频域中可以通过组合这些/频谱部分0, 即各次谐波, 来重建其频谱.

4 特征波形的非负矩阵分解

在 WI 编码器中, 输入的每一帧语音首先经过 LP 分析滤波器得到一段残差信号, 然后在该帧残差信号上按基音周期的大小提取 10 个特征波形. 为了提高运算精度, 与 CW 有关的运算一般都在频域进行. 于是, 本文选用了离散时间傅氏级数(Discrete Time Fourier Series2 DIFS)作为 CW 的频域表达. 实验表明人耳对一般的相位信息感觉不是很敏感, 于是在编码端可以将各子帧的 DIFS 系数利用极坐标形式去除其 DIFS 系数中的相位信息, 即:

$$|CW_k(n)| = \sqrt{\text{DIFS}_k \cdot A_k^2(n) + \text{DIFS}_k \cdot B_k^2(n)}$$

$$k = 1, 2, \dots, P(n)/2 \quad (4)$$

然后只对经过功率归一化之后的 CW 幅度谱进行量化与编码, 而 CW 相位谱在解码端通过固定相位来重建.

4.1 特征波形的去对齐

不难发现, 上述的处理过程将定义在时间轴上的一维的语音残差信号转化为了 10 个特征波形的幅度频谱, 也就是转化成了一个非负的矩阵. 于是 NMF 算法可以分解这样的一个非负矩阵. 传统的 CW 分解方法, 如简单线性相位低通滤波、小波变换、奇异值分解, 需要在分解特征波形之前进行 CW 的对齐, 以使得特征波形矩阵呈现出非常强的周期性. 与这些传统的分解方法不同, 非负矩阵分解是基于/部分0的分解, 其目标是找到语音谱类脉冲的各次谐波. 所以本文基于 NMF 来分解 CW 时, 去掉了编码端的特征波形对齐操作, 仅此一项就把 WI 编码器的计算量减少了 251 968 MOPS^[17, 18].

4.2 特征波形的分类

在我们的 WI 编码器中, 语音信号基音周期(样点数)的取值范围为 20 至 120, 所以转化到频域后, 特征波形的维数将在 10 至 60 之间变化. 为了提高特征波形的分解精度, 我们先按照每帧语音的基音周期来对特征波形分类, 然后对每一类特征波形都训练出其所对应的基矩阵. 具体的特征波形分类方法为: 首先取该帧特

征波形 10 个子帧的最大基音周期, 然后前 8 类以步长为 10 来划分, 即把最大子帧基音周期落入 20 至 30 之间的特征波形分为第 1 类; 最大子帧基音周期落入 30 至 40 之间的 CW 分为第 2 类, 依此类推. 最后把最大子帧基音周期落入 100 至 120 之间的 CW 分为第 9 类. 这样我们把特征波形总共分为 9 类, 然后分别对这 9 类特征波形选取充分多的实验样本通过 NMF 算法训练出属于该类特征波形的基矩阵 W, 最终我们将得到 9 个不同的基矩阵 W. 显然, 特征波形分类后训练得到的这 9 个基矩阵, 对于分解随基音周期不断变化的特征波形来说, 更具有适应性.

4.3 分解阶数的选取

在用 NMF 算法训练基矩阵之前, 我们需要首先确定 NMF 算法的一个关键参数, 即分解阶数 r. Lee 和 Se2 ung 在文献[13]中建议分解阶数 r 应该按照 $(n + m) r < nm$ 的原则来选取, 其中 n, m 依次是待分解矩阵的行、列维数. 在我们的 WI 编码模型中, 为了满足这个约束, r 应该不大于 8, 而实验表明即使 r 取 8 也不能保证把 CW 幅度谱的重建误差控制在一个可以容忍的限度. 虽然提高分解阶数可以降低 CW 的重建误差, 但是也不能无限度提高 r 值, 因为 r 值的提高会带来两个弊端: (1) NMF 的计算复杂度会随之线性增长; (2) 编码矩阵维数增加并增大其量化难度. 所以, 本文选取分解阶数 r 为 16 以折中上述矛盾. 下面的 4.3 节将介绍 r 为 16 时, NMF 能得到可与 2 阶奇异值分解相比较的 CW 分解精度.

4.4 基矢量的分带初始化

在第 3 节中, 我们讨论了语音频谱的/部分0这一概念, 即语音频谱的/部分0就是频率轴上一系列类脉冲的谐波结构, 所以我们总是期望 NMF 训练得到的基矢量尽可能稀疏并且呈现出类脉冲的结构. 在标准的非负矩阵分解算法中, 基矢量的所有元素都初始化为随机的正数. 与此不同, 本文对基矢量作了分带初始化限制. 具体来说就是基矢量在一个特定的频带内(以下称其为/有效频带0)初始化为随机的正数, 而其他的频带都初始化为 0 值. 另外, 各基矢量的有效频带的带宽都基本相等, 并且相邻基矢量的有效频带相互叠接. 以第 8 类特征波形为例, 16 个基矢量按照如下的方式初始化: 第 1 个基矢量的有效频带覆盖第 1 个到第 6 个样点; 第 2 个基矢量的有效频带覆盖第 4 个到第 9 个样点, 即与前一个基矢量重叠 3 个样点, 并向前推进 3 个样点; 余下的基矢量都按照同样的方式来初始化. 由于非负矩阵分解采用乘法迭代算法, 所以初始化为零值的元素在迭代收敛时依然为 0. 基矢量初始化完毕后, 本文采用式(2)、式(3)所示的迭代算法来训练基矢量, 直至式(1)所示的代价函数收敛.

以第 9 类特征波形为例, 由两种不同的基矢量初始化方法, 经过 NMF 算法训练得到的最终基矢量示于图 1 和图 2. 其中, 每幅图都包含 16 个子图, 依次表示相应的 16 个基矢量. 图 1 是全频带随机初始化后, 训练得到的基矢量; 图 2 是本文提出的分带初始化后, 训练得到的基矢量. 对比二图不难发现, 本文提出的算法所训练出的基矢量, 更具稀疏性, 更能展示语音频谱的谐波结构.

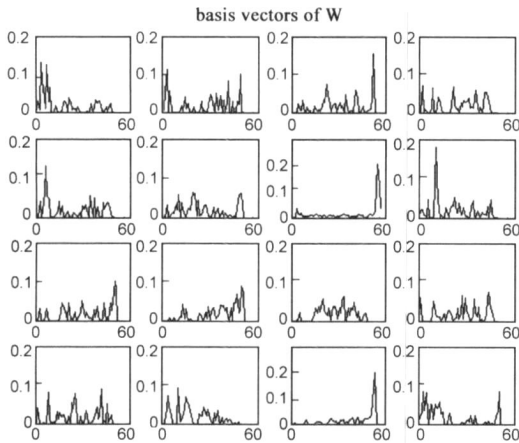


图 1 全频带随机初始化后训练得到的基矢量
横轴: 谐波个数 纵轴: 幅度

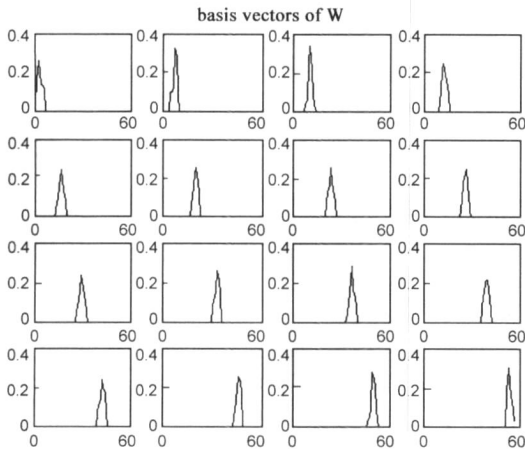


图 2 分带限制初始化后训练得到的基矢量
横轴: 谐波个数 纵轴: 幅度

对 9 类特征波形, 都按照上述的方法分别训练出其

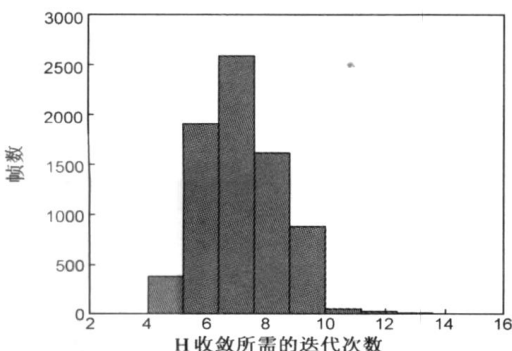


图 3 编码矩阵收敛所需的迭代次数统计直方图

对应的基矢量, 并保存在 W1 编码器的编码端和解码端. 训练完毕后, 当分解新的一帧特征波形时, 首先将此帧 CW 按照其子帧的最大基音周期进行分类(与训练基矢量时的 CW 分类方式相同), 然后取出该类特征波形所对应的已训练好的基矩阵. 保持该基矩阵不变, 并对其编码矩阵采用式(2)所示的算法进行迭代, 直至收敛. 图 3 所示的统计直方图显示, 当迭代次数取 10 时, 编码矩阵即可收敛.

4.1.5 CW 分解的实验结果

测试数据取自 100000 帧语音, 对其特征波形的幅度谱分别进行奇异值分解和非负矩阵分解, 然后重建幅度谱. 选取原始幅度谱与重建幅度谱之间欧式距离的平方作为误差测度. 实验结果为: 由二阶奇异值及其左、右奇异矢量重建的平均误差为 118292, 由一阶奇异值及其左、右奇异矢量重建的平均误差为 214853, 由本文基于标准 NMF 提出的基矢量分带初始化 NMF 算法所重建的平均误差为 118596, 由传统的 NMF 算法重建的平均误差为 211513. 对比以上数据不难发现, 在频域本文提出的分解方法得到了与二阶奇异值分解几乎同等的 CW 重建精度.

任取 4 帧语音数据, 其中既包括浊音也包括清音. 其特征波形示于图 4(a), 由二阶奇异值及其左、右奇异矢量重建的 CW 表面示于图 4(b), 由本文基于标准 NMF 提出的基矢量分带初始化 NMF 算法所重建的 CW 表面示于图 4(c), 由一阶奇异值及其左、右奇异矢量重建的 CW 表面示于图 4(d). 其中, 图 4(d)中虚线椭圆框起的部分与原始波形相比, 发生了很大的失真. 这说明仅用一阶奇异值重建 CW, 将会降低 CW 的合成精度. 对比图 4(a), 4(b), 4(c) 不难发现, 在时域本文提出的分解方法也得到了与二阶奇异值分解几乎同等的 CW 重建精度.

相比于奇异值分解, NMF 在实现了同等分解精度的同时, 还具有更低的计算复杂度: 奇异值分解的计算复杂度为 $O(mn^2)$, 而非负矩阵分解的计算复杂度为 $O(mn)$, 这里 m, n 分别为 CW 幅度谱矩阵的行维数和列维数.

5 编码矩阵的分裂式矩阵量化

特征波形作非负矩阵分解后, 原高维的 CW 幅度谱可以由低维的编码矩阵来描述, 这就降低了参数量化的难度, 达到了数据压缩的目的.

由 4.2 节可知, 基矢量分带初始化限制下的非负矩阵分解算法所训练出来的 W 矩阵的列矢量, 是频率轴上从低频到高频的一系列类脉冲结构的谐波, 而 V 矩阵的列向量可以解释为对这些谐波的加权组合, 加权系数为 H 矩阵对应列中的元素. 从上述基矢量线性组

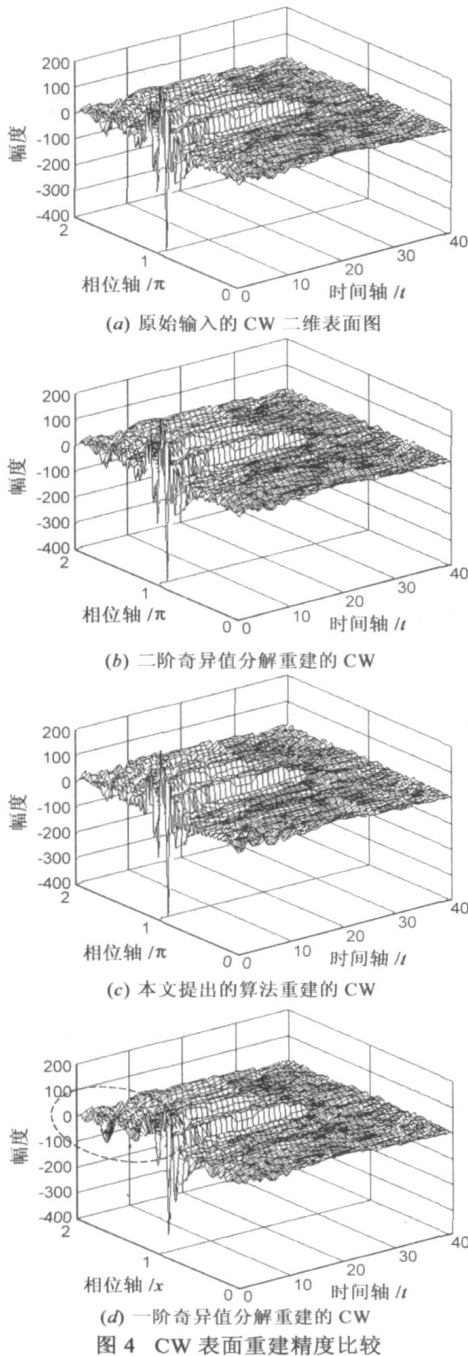


图4 CW表面重建精度比较

合的角度不难发现, H 矩阵中低维的元素是用来组合低频谐波的权重系数, H 矩阵中高维的元素是用来组合高频谐波的权重系数. 所以, 本文将 H 矩阵分裂成 3 个子矩阵: 前 5@10 维的用来加权低频谐波的子矩阵称作低频矩阵, 中间的 5@10 维的用来加权中频谐波的子矩阵称作中频矩阵, 最后的 6@10 维的用来加权高频谐波的子矩阵称作高频矩阵. 然后根据人耳的听觉系统对低频信息感知更加敏感的特点^[1], 可以对低频矩阵分配较多的比特数以精确量化, 而对于高频矩阵则可以粗糙量化.

在对上述 3 个矩阵量化之前, 本文先对其列向量下采样到 5 维, 然后进行矩阵量化^[19] (Matrix Quantization MQ). 对于在编码端没有量化的另外 5 维, 在解码端可以通过线性插值的方式来恢复. 综上, 编码矩阵 H 的量化方案示于图 5.

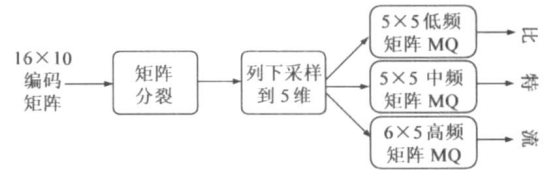


图5 编码矩阵的矩阵量化方案

6 NMF2WI 编码方案的性能评价

前面 4、5 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他参数的量化方法: LSF 参数采用了预测分裂式矢量量化方法 (PSVQ)^[1] 进行量化; 基音周期采用了 7 比特的均匀标量量化; 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^[1,6]. 表 1 为我们的 2kb/s NMF2WI 编码器的比特分配情况.

表1 2kb/s NMF2WI 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 WI 语音编码器进行了改进, 改进后的 NMF2WI 编码器框图示于图 6. 其中为了降低计算复杂度, 编码端去除了特征波形对齐模块.

表2 主观 A/B 听力测试

测试语音	偏爱 2kb/s NMF2WI	偏爱 2.4kb/s SVD2WI	无偏爱
女声语音	18.75%	15.625%	65.625
男声语音	9.375%	28.125%	62.5%
所有语音	14.0625%	21.875%	64.0625%

表3 MOS 分测试结果

测试语音	2kb/s NMF2WI	2.4kb/s MELP
女生语音	3.08	3.10
男声语音	3.04	3.26
所有语音	3.06	3.18

在低速率语音编码中, 由于相位信息已被忽略, 波

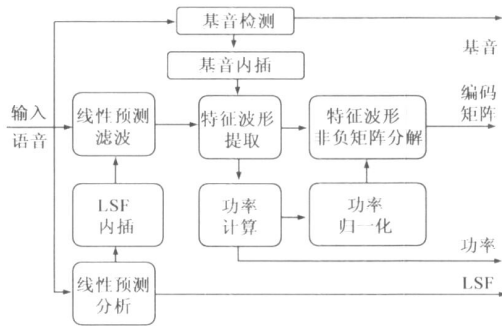


图 6 NMF-WI 编码器框图

形无法匹配,所以以信噪比为基础的客观测度不再适用,通常对编码器进行主观评价^[9]。本文的 2kb/s NMF2WI,属于低速率语音编解码器,所以采用了主观语音质量 A/B 测试。本文对 2kb/s NMF2WI 和 2.4kb/s SVD2WI^[9-11]两个语音编码器进行了语音质量主观对比 A/B 测试。测试语音包括 8 句标准男声语音、8 句标准女声语音,采样频率均为 8kHz,由 12 名男女不同的听音人进行了测试,表 2 为 A/B 测试结果。实验结果表明 2kb/s NMF2WI 合成的语音质量接近于 2.4kb/s SVD2WI 的语音质量。另外,本文对 2kb/s NMF2WI 和 2.4kb/s MELP 两个语音编码器进行了 MOS 分测试,测试语料与主观 A/B 测试的语料相同,听音人由 18 名男女基本各半的青年人组成,MOS 分测试结果示于表 3。测试结果表明,2kb/s NMF2WI 合成的语音质量稍差于 2.4kb/s MELP 的语音质量。

我们的 2kb/s NMF2WI 编码器的主要缺陷在于合成语音中有轻微的蜂鸣声。这是由于编码端抛弃了 CW 相位信息,然后在解码端由固定相位恢复所导致的。

7 结论

综上所述,本文基于文献[13],提出了一种基于非负矩阵分解的特征波形分解与量化模型。同时,在本文提出的基矢量分带初始化限制下,NMF 算法在频域里找到了 CW 幅度谱的/部分0,实现了对 CW 幅度谱的基于/部分0的表征。用 NMF 来分解特征波形的主要优势在于:(1)NMF2WI 不需要额外延时;(2)NMF2WI 不需要特征波形的对齐操作;(3)相对于奇异值分解来说,NMF 在降低了 CW 分解的计算复杂度的同时,保证了 CW 的重建精度不降低。

特征波形经非负矩阵分解后,原高维的 CW 矩阵可以由低维的编码矩阵来描述。实验表明,数据压缩后的 CW 更易于参数量化。不过实验也表明,如果编码端完全抛弃 CW 的相位信息,并在解码端由固定相位来恢复的话,重建的语音将会因为帧间相关性过强而产生轻微的蜂鸣声。所以,下一步考虑在 NMF2WI 框架下对 CW 相位谱进行合理建模并高效量化,以提高 NMF2WI 合成

语音的自然度。

参考文献:

- [1] 鲍长春. 低比特率数字语音编码基础[M]. 北京: 北京工业大学出版社, 2001. 132- 234.
- [2] Kleijn W B. Continuous representation in linear predictive coding[A]. IEEE ICASSP. 91[C]. Toronto, 1991. 201- 204.
- [3] Kleijn W B, Haagen J. Waveform Interpolation for Coding and Synthesis. Speech coding and Synthesis[M]. Holland: Elsevier Science, 1995. 175- 207.
- [4] Kleijn W B, Haagen J. Transformation and decomposition of the speech signal for coding[J]. IEEE Signal Processing Letters, 1994, 1(9): 136- 139.
- [5] 张海. 基于 WI 的低速率语音编码算法研究与实现[D]. 北京: 北京工业大学电子信息与控制工程学院, 2001. 19- 72.
- [6] 朱娜娜. 2kbps 波形内插语音编码算法的研究[D]. 北京: 北京工业大学电子信息与控制工程学院, 2003. 10- 70.
- [7] N R Chong, I S Burnett, J F Chicharo. A new waveform interpolation coding scheme based on pitch synchronous wavelet transform decomposition[J]. IEEE Transactions on Speech and Audio Processing, 2000, 8(3): 345- 348.
- [8] J Lukasiak, I S Burnett. Scalable decomposition of speech waveforms[A]. 2002 IEEE Speech Coding Workshop Proceedings[C]. Tsukuba City, Ibaraki, Japan, 2002. 135- 137.
- [9] 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, 34(1): 135- 140. WANG Gu2ping, BAO Chang2chun, ZHANG Peng. Low bit rates waveform interpolation speech coding based on singular value decomposition[J]. Acta Electronica Sinica, 2006, 34(1), 135- 140. (in Chinese)
- [10] 王贵平. 基于奇异值分解的低速率波形内插语音编码算法的研究[D]. 北京: 北京工业大学电子信息与控制工程学院, 2005. 48- 78.
- [11] Guiping WANG, Chang2chun BAO. Low complexity decomposition for the characteristic waveform of speech signal[A]. ISCSLP2004[C]. Hong Kong: IEEE Press, 2004. 145- 149.
- [12] C H R2z, I S Burnett, J Lukasiak. Very low rate speech coding using temporal decomposition and waveform interpolation [A]. IEEE Workshop on Speech Coding Proceedings[C]. Wisconsin: IEEE Press, 2000. 29- 31.
- [13] D D Lee, H S Seung. Learning the parts of objects by non2negative matrix factorization[J]. Nature, 1999, 401(6755): 788- 791.
- [14] Sven Behnke. Discovering hierarchical speech features using convolutional non2negative matrix factorization[A]. International Joint Conference on Neural Network 2003[C]. Portland, United States: IEEE Press, 2003. 2758- 2763.
- [15] D D Lee, H S Seung. Algorithms for non2negative matrix fa

- torization [J]. In Advance in Neural Information Processing, 2000, 13: 556- 562.
- [16] L R Rabiner, B H Juang. Fundamentals of Speech Recognition [M]. New Jersey: Prentice2Hall, 1993. 131- 139.
- [17] 李靛. 高质量的 2kb/s 波形内插语音编码算法研究 [D]. 北京: 北京工业大学电子信息与控制工程学院, 2005. 13 - 99.
- [18] 齐峰岩, 鲍长春. 波形内插语音编码中特征波形表达和对齐快速算法 [J]. 北京工业大学学报, 2006, 32(6) : 514 - 519.
- Qi Feng2yan, Bao Chang2chun. Fast algorithms on characteristic waveform representation and alignment in waveform interpolation speech coding [J]. Journal of Beijing University of Technology, 2006, 32(6) : 514- 519. (in China)
- [19] C Tsao, R M Gray. Matrix quantizer design for LPC speech using the generalized Lloyd algorithm [J]. IEEE Transactions on Acoustics, Speech and Signal Processing, 1985, 33(3) : 537 - 545.

作者简介:



张 鹏 男, 1980 年 8 月出生于河北省保定市, 现为北京工业大学硕士研究生. 主要研究方向为低比特率数字语音编码、语音信号分解等. E2mail: pengzhang@emails.bjut.edu.cn



郭莉莉 女, 1981 年 2 月出生于河北省唐山市, 现为北京工业大学硕士研究生. 主要研究方向为低比特率数字语音编码、语音信号分解等.



鲍长春 男, 1965 年 6 月出生于内蒙古赤峰市, 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, 5 通信学报、5 数据采集与处理、5 信号处理学报编委. 主要研究领域为数字信号处理与语音编码.