基于二维非负矩阵分解的1kb/s WI语音编码算法

薛二娟,鲍长春,李如玮
(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: 本文针对波形内插(WI)语音编码模型和参数量化等技术进行了研究,并最终提出了一种基于二维非负矩阵分解的1kb/s波形内插(2DNMF-WI)语音编码算法.文中采用二维非负矩阵分解(2D-NMF)方法来分解语音特征波形(CW),该分解方法在行和列两个方向上同时压缩CW幅度谱矩阵的维数,使得CW幅度谱矩阵降维后得到的编码矩阵维数较小,易于量化.此外,在甚低速率语音编码中,由于没有足够的比特数来描述编码参数,往往很难得到高质量的合成语音.本算法采用两帧联合编码、帧间后向预测三级矢量量化、离散余弦变换(DCT)和分裂式矩阵量化等技术来降低编码速率和改善音质.非正式主观听觉测试显示,1kb/s 2DNMF-WI编码器合成语音的质量稍差于2kb/s的NMF-WI语音编码算法.

关键词: 语音编码;波形内插;特征波形;二维非负矩阵分解;两帧联合
中图分类号: TN912.3 **文献标识码:** A **文章编号:** 0372-2112 (2010) 07-1574-06

1kb/s Waveform Interpolative Speech Coding Based on Two-Dimensional Nonnegative Matrix Factorization

XUE Er-juan, BAO Chang-chun, LI Ru-wei
(*Speech and Audio Signal Processing Lab, School of Electronic Information and Control Engineering, Beijing University of Technology, Beijing* 100124, *China*)

Abstract: This paper is focused on the model of waveform interpolation (WI) and its parameters quantization, then a waveform interpolation speech coding algorithm based on two-dimensional nonnegative matrix factorization at 1kb/s is presented. This method makes the dimensions of CW magnitude matrix much lower in columns and rows, so it is convenient for quantizing the coding matrix. In addition, speech coders at very low bit rates can hardly get good performance, for there are no sufficient bits to express these coding parameters. Then two-frame joint, inter-frame backward prediction three-stage vector quantization, discrete cosine transform (DCT) and split matrix quantization techniques are promoted in this paper, in order to reduce the speech coding bit rates as well as to improve the quality of the speech. The results of informal subjective listening test show that the performance of 1kb/s 2DNMF-WI coder is a little worse than that of 2kb/s NMF-WI coder.

Key words: speech coding; waveform interpolation; characteristic waveform; two-dimensional nonnegative matrix factorization; two-frame joint

1 引言

WI语音编码基于线性预测技术,是一种非常优秀的低速率语音编码算法.传统WI编码[1-3]采用FIR低通滤波将残差域渐变的特征波形分解为慢渐变波形(Slowly Evolving Waveform, SEW)和快渐变波形(Rapidly Evolving Waveform, REW),然后对这两种波形分别编码,但是此方法增加了一帧延时,同时分解精度难以保证,分解后的SEW和REW不能完全分离,仍然具有较强相关性,影响量化效果.目前,也有研究者提出采用基音同步小波变换[4]和奇异值分解[5~7]的方法来分解特征波形,但它们各自有其相应的局限性,因此限制了它们在低速率语音编码中的应用.文献[8,16]提出将非负矩阵分解应用于语音编码领域,这是一种对特征波形基于"部分"而非基于"整体"的分解方法,符合人类大脑对事物的感知过程[10].在此基础之上,本文采用新的二维非负矩阵分解[11]CW的方法,它是根据CW幅度谱矩阵列方向的频率变化特性[8,9]以及行方向的波形慢渐变特性[1],对其进行列非负矩阵分解和行非负矩阵分解,从而在行和列两个方向上同时压缩CW的维数,使得CW幅

收稿日期:2009-05-16;修回日期:2009-08-28
基金项目:北京市教委科技发展计划(No.KM200710005001);国家自然科学基金(No.60872063);北京市自然科学基金(No.4042009);北京市属高校人才强教计划

度谱矩阵降维后得到的编码矩阵维数较小,易于量化.

为了满足短波窄带数字保密通信的需要,将语音编码速率降低到 1kb/s 甚至以下,即发展甚低速率语音编码,具有十分重要的理论研究意义和实际应用价值.在甚低速率语音编码中,可用于对传输参数进行编码的比特数很少,如何用有限的比特来获得高质量的合成语音,人们进行了长期的研究.目前速率在 2 ~ 4kb/s 范围内的低速率语音编码算法比较成熟,因此,在已有算法基础之上加长原算法中分析帧的长度或直接减少所传送参数的编码比特数[12]是比较快捷的降低语音编码速率的方法,但是这些方法都将导致语音质量不同程度的下降.近年来,也有一些学者[13,14]提出利用参数帧内、帧间的相关性,将多帧参数联合编码来压缩冗余,达到降低编码速率的目的.多帧参数联合编码已经成功应用于以 MELP[13,14]为模型的甚低速率语音编码方案中,得到了具有较高可懂度和自然度的合成语音.本文即是借鉴此思想,以 2DNMF-WI 为模型,提出一种 1kb/s 的甚低速率波形内插语音编码算法.

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,NMF 算法通过有限次迭代总能够找到两个非负矩阵,使这两个矩阵的乘积近似等于给定的待分解矩阵.NMF 可以用下面的公式表示:已知非负矩阵 V,寻找适当的非负矩阵 W 和 H,使其满足:

$$V_{n\times m}\approx W_{n\times r}H_{r\times m} \tag{1}$$

其中,V 为给定的 n 维数据向量的集合 $V\in R^{n\times m}$,m 为数据样本个数,$W_{n\times r}$ 为基矩阵,它的列向量称为基矢量,$H_{r\times m}$ 为编码矩阵,要求 $W\geqslant 0$,$H\geqslant 0$,r 称为分解阶数,代表基矢量的个数,为了达到数据压缩的目的,D.D.Lee 和 H.S.Seung 在文献[10]建议 r 的选取满足 $(m+n)r<mn$,这样分解后的矩阵 W 和 H 的维数都将小于原矩阵 V.

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,也叫目标函数[10],如式(2)所示:

$$\|V-WH\|^2=\sum_{ij}(V_{ij}-(WH)_{ij})^2 \tag{2}$$

在 W 和 H 矩阵非负的条件下,最小化 $\|V-WH\|^2$,得到迭代规则[15]如式(3)所示.

$$H_{a\mu}\leftarrow H_{a\mu}\frac{(W^{\mathrm{T}}V)_{a\mu}}{(W^{\mathrm{T}}WH)_{a\mu}}\qquad W_{ia}\leftarrow W_{ia}\frac{(VH^{\mathrm{T}})_{ia}}{(WHH^{\mathrm{T}})_{ia}} \tag{3}$$

3　二维非负矩阵分解

3.1　二维非负矩阵分解算法

在图像处理领域,张道强[11]等人提出了二维非负矩阵分解算法.算法详述如下:已知待分解矩阵 $X=[A_1,A_2,\cdots,A_m]$,它是由 m 个非负子矩阵 A_k 组合而成,其中 $A_k\in R^{p\times q}$,$k=1,2,\cdots,m$.对非负矩阵 X 执行二维非负矩阵分解,包含以下两个过程:列非负矩阵分解和行非负矩阵分解,下面对它们分别进行介绍.

3.1.1　列非负矩阵分解

所谓列非负矩阵分解就是寻找一个大小为 $p\times d$ 的非负矩阵 L 和一个大小为 $d\times qm$ 的非负矩阵 H,使其满足:

$$X\approx LH \tag{4}$$

其中,d 为列非负矩阵分解的阶数,分解的左矩阵 L 称为列基矩阵,这是因为 X 矩阵的每一列与原始子矩阵 A_k 的每一列相对应,故称 L 矩阵为列基矩阵[21].分解的右矩阵 H 称为系数矩阵,为方便起见,将 H 矩阵表示为 m 个子系数矩阵 H_k 的组合,即 $H=[H_1,H_2,\cdots,H_m]$,其中,$H_k\in R^{d\times q}$,$k=1,2,\cdots,m$.此时子矩阵 A_k 可以表示成如下形式:

$$A_k\approx LH_k,\qquad k=1,2,\cdots,m \tag{5}$$

本文中 L 和 H 的获取采用基于欧氏距离最小化 $\|X-LH\|^2$ 的非负矩阵分解算法.依据式(3)所示规则不断的迭代,最终可以做到使得矩阵 L 和 H 收敛到局部最优的分解情况.

3.1.2　行非负矩阵分解

由子系数矩阵 H_k 的转置构造一个新的 $q\times dm$ 大小的非负矩阵 $H'=[H_1^{\mathrm{T}},H_2^{\mathrm{T}},\cdots,H_m^{\mathrm{T}}]$.类似地,找到一个大小为 $q\times g$ 的非负矩阵 R 和一个大小为 $g\times dm$ 的非负矩阵 C,使得

$$H'\approx RC \tag{6}$$

其中,g 为行非负矩阵分解的阶数,矩阵 R 和 C 分别为行非负矩阵分解的基矩阵和系数矩阵.将 C 分为 m 个 $g\times d$ 大小的子矩阵 C_k,$k=1,2,\cdots,m$,此时系数矩阵 C 可表示为 $C=[C_1,C_2,\cdots,C_m]$.H_k^{T} 矩阵的每一行与原始子矩阵 A_k 的每一列信息相对应,为了与 2D-NMF 的列非负矩阵分解区别开来,将此次分解过程称为行非负矩阵分解,左矩阵 R 称为行基矩阵[21].此时 H_k^{T} 可以表示成如下形式:

$$H_k^{\mathrm{T}}\approx RC_k,\qquad k=1,2,\cdots,m \tag{7}$$

矩阵 R 和 C 的获取方法与列非负矩阵分解中的 L 和 H 相同,具体迭代过程不再赘述.

我们获得了 $p\times d$ 大小的列基矩阵 L 和 $q\times g$ 大小的行基矩阵 R,将式(7)代入式(5),得到

$$A_k\approx LC_k^{\mathrm{T}}R^{\mathrm{T}},k=1,2,\cdots,m \tag{8}$$

令 $D_k=C_k^{\mathrm{T}},k=1,2,\cdots,m$,

并且 $D=[D_1,D_2,\cdots,D_m]$,那么整理可得:

$$A_k\approx LD_kR^{\mathrm{T}},\qquad k=1,2,\cdots,m \tag{9}$$

$$X \approx LDR^{\mathrm{T}} \tag{10}$$

3.2 CW 的二维非负矩阵分解

在 WI 语音编码算法中，残差信号通过傅氏级数得到非负的幅度谱，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表面. 对于 CW 幅度谱，为了满足甚低速率语音编码的要求，我们将相邻两帧 CW 联合在一起组成一超帧非负 CW 幅度谱矩阵，用 $X_{p\times qm}$ 表示. 已知特征波形的提取速率为 10 个/帧，并且 CW 幅度谱矩阵的每一个列向量表示一个 CW，因此 $q=10$，$m=2$. 而变量 p 表示的是特征波形的维数，它随着基音周期长度的变化而改变.

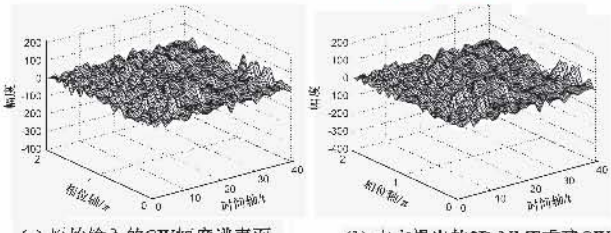
对矩阵 X 进行线性分解用下式表示：

$$\begin{aligned} X_{p\times 10\times 2} &\approx L_{p\times d}H_{d\times 10\times 2} \\ H'_{10\times d\times 2} &\approx R_{10\times g}C_{g\times d\times 2} \end{aligned} \tag{11}$$

其中，L 为列基矩阵，R 为行基矩阵，d 和 g 分别为 CW 列非负矩阵分解以及行非负矩阵分解的分解阶数，实验表明当 $d=16$，$g=6$ 时分解精度最佳. 令 $D=C^{\mathrm{T}}_{g\times d\times 2}$，那么 D 为编码矩阵. 通过对 CW 幅度谱矩阵进行二维非负矩阵分解，原始高维的 CW 幅度谱最终可以由一个 16×12 的编码矩阵来表达，达到了数据压缩的目的，有利于在甚低比特率下对其进行量化编码.

3.3 CW 分解的实验结果

我们任取 4 帧语音数据，将本文提出的 2D-NMF 算法重建的 CW 幅度谱表面与原始 CW 幅度谱表面示于图 1.



(a) 原始输入的CW幅度谱表面　　(b) 本文提出的2D-NMF重建CW

图1 CW幅度谱表面对比图

对比图形，发现 2D-NMF 算法重建的 CW 幅度谱表面与原始 CW 幅度谱表面达到很好的匹配效果，可见，本文提出的 2D-NMF 算法具有良好的 CW 重建精度.

3.4 不同 CW 分解方法的性能比较

本文采用 2D-NMF 算法分解 CW，若 p 和 q 分别为待分解 CW 幅度谱矩阵的行和列维数，d 和 g 分别为列非负矩阵分解、行非负矩阵分解的分解阶数，即分解所得编码矩阵 D 的行和列维数，本文中，$d=16$，$g=6$，并且满足 $d\leqslant p$，$g<q$. 文献[8]中采用 NMF 分解 CW，且分解阶数 $r=16$，p 和 r 为 CW 经 NMF 分解得到的编码矩阵 H 的行和列维数. 上述两种分解方法，编码矩阵与 CW 幅度谱矩阵间的压缩比比较，如表 1 所示.

由于 CW 的维数随基音周期长度的变化而改变，因此我们以 61×10 的 CW 幅度谱矩阵为例，原 CW 幅度谱矩阵经过 2D-NMF 分解，可以由一个大小为 16×6 的编码矩阵来表达，压缩比为 6.3542；而通过 NMF 分解 CW 幅度谱矩阵，则是得到一个大小为 16×10 的编码矩阵，压缩比为 3.8125，通过比较，发现 2D-NMF 算法相比 NMF 算法较大幅度地压缩了 CW 幅度谱矩阵的维数，方便了 CW 的传输和量化.

表 1 不同分解方法的性能比较

分解算法	NMF	2DNMF
压缩比	$\frac{pq}{pr}$	$\frac{pq}{gd}$

4 编码参数的量化

在特征波形内插语音编码模型中，需要量化的参数有线谱频率(LSF)、基音周期、功率和编码矩阵. 而这些参数均存在冗余，编码时若将连续几帧语音参数联合起来组成一个超帧进行编码，可以进一步压缩帧间冗余信息，达到降低参数量化比特数的目的. 结合语音信号的短时慢渐变特性，通常是将 2～4 帧语音信号组成一个超帧，编码生成一帧参数. 然而，联合帧数越多，则意味着算法延时越大，而且随着联合帧数的增加，量化码本存储量和最佳码字搜索的运算量都将急剧增加. 综合考虑各种因素的影响，在我们的 WI 编码算法中将同时量化连续两帧的语音信号.

4.1 线谱频率的量化

基于线性预测系数有较宽动态范围以及在合成滤波器中不稳定性的考虑，本算法将线性预测系数参数转化为 LSF 参数，LSF 的有序性和慢渐变表明 LSF 参数帧内、帧间的相关性[1]. 因此，对其采用两帧联合编码和帧间后向预测技术来减少冗余.

将两帧 LSF 参数组成一个 20 维的矢量，采用帧间后向预测方法去除冗余，设 $\boldsymbol{\omega}^{(n)}=\{\boldsymbol{\omega}^{(n)}_{1,i},\boldsymbol{\omega}^{(n)}_{2,i}\}$，$i=1,2,\cdots,10$ 为当前超帧 LSF 矢量，其中 n 代表超帧标号，$\omega^{(n)}_{1,i}$ 和 $\omega^{(n)}_{2,i}$ 分别为当前超帧中的第一帧和第二帧的 LSF 矢量，i 表示每一帧的第 i 个 LSF 频率. 为了减小计算和量化对象的动态范围[1]，将 LSF 的平均值 $\boldsymbol{\omega}_i$，$i=1,2,\cdots,10$ 从 $\boldsymbol{\omega}^{(n)}$ 中减去，得到第 n 超帧无偏的 LSF 矢量：

$$\begin{aligned} \vec{\boldsymbol{\omega}}^{(n)} &= \{\boldsymbol{\omega}^{(n)}_{1,i}-\boldsymbol{\omega}_i,\boldsymbol{\omega}^{(n)}_{2,i}-\boldsymbol{\omega}_i\} \\ &= \{\vec{\boldsymbol{\omega}}^{(n)}_{1,i},\vec{\boldsymbol{\omega}}^{(n)}_{2,i}\}, \quad i=1,2,\cdots,10 \end{aligned} \tag{12}$$

令 $\hat{\boldsymbol{\omega}}^{(n-1)}_{2,i}$，$i=1,2,\cdots,10$ 为前一超帧中第二帧量化后去除均值的 LSF 矢量，那么帧间后向预测过程可以表示为：

$$\tilde{\boldsymbol{\omega}}^{(n)}=\begin{cases}\alpha_{1,i}\hat{\boldsymbol{\omega}}^{(n-1)}_{2,i} \\ \alpha_{2,i}\hat{\boldsymbol{\omega}}^{(n-1)}_{2,i}\end{cases} \tag{13}$$

用前一超帧中第二帧量化后去除均值的 LSF 矢量来预测当前超帧中的第一及第二帧的 LSF 矢量. 其中 $\alpha_{1,i}$ 和

$\alpha_{2,i}$为预测系数,通过使平方预测误差最小估计预测系数 $\alpha_{1,i}$和$\alpha_{2,i}$.

帧间后向预测去除冗余后,第一种量化方案是对 LSF 参数采用两级分裂矢量量化:用 10 比特对 20 维 LSF 参数进行第一级矢量量化;第二级矢量量化时,将 20 维的 LSF 参数分裂为前 10 维和后 10 维,分别用 7 比特码本和 7 比特码本进行量化.第二种量化方案是对 LSF 参数采用三级矢量量化,对每级 20 维 LSF 矢量分别以 9,8,7 比特来编码.

表 2　两种量化方案量化性能比较

量化方案	ASD(dB)	0~2dB	2~4dB	>4dB
两级分裂矢量量化	1.55	77.7%	21.6%	0.7%
三级矢量量化	1.47	81.8%	17.9%	0.3%

最后,通过选取标准语音库中一段 66 分钟的测试语音,测试 LSF 参数在两种不同的量化方案下引入的量化误差,测试结果见表 2[17],其中,ASD 表示平均谱失真,对比发现 24bit 的帧间预测三级矢量量化优于两级分裂矢量量化.

4.2　基音周期的量化

在低速率语音编码中准确的基音估计至关重要.为了获得较为精确的基音参数,本文基于归一化互相关函数检测基音周期[1,17],并且规定当检测到的语音为清音时,令基音周期 $P=80$(个样点).由于基音频率在 60~500Hz 范围内变化,因此我们设定基音周期的取值范围为 20~120(个样点),结合甚低速率语音编码节省比特数的要求,本文对编码端检测的基音周期值采用两帧联合矢量量化的方法,即将两帧基音周期联合起来组成一个两维矢量 $\boldsymbol{P}=[p_1,p_2]$,为了进一步压缩基音周期的动态范围,取其对数 $\boldsymbol{P}_{\lg}=[\lg(p_1),\lg(p_2)]$,并分配 6 比特码本量化,码书依照 LBG 算法[1]进行设计.

图 2 给出了测试语音"大家都说普通话"的波形图及基音周期轮廓曲线[17].其中,实线所绘制曲线表示 WI 编码端检测到的基音周期,虚线为 WI 解码端基音周期轮廓图.通过观察发现,解码端的基音周期轨迹与编码端基音周期轨迹取得了较好的拟合效果.

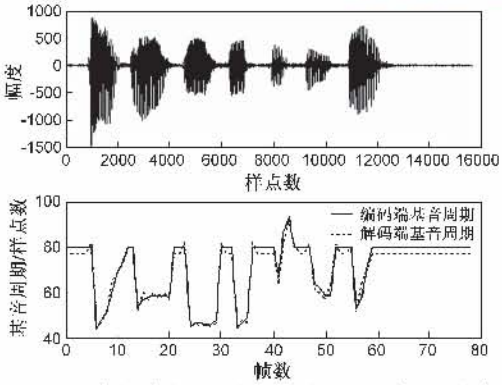


图2　"大家都说普通话"的语音波形及基音周期轮廓

此外,有研究表明[18]:当基音分辨率不大于 3 个样点时,女性合成语音的质量较高,而对于男性语音,当基音分辨率不大于 5 个样点时,合成语音的质量较高.我们选取标准语音库中一段 60 分钟的测试语音,测试两帧联合矢量量化的基音与原始基音周期两者之间的分辨率[17],统计数据见表 3.其中,Δ 代表基音分辨率,Proportion 代表不同分辨率所占的比重.实验结果表明,编码前后基音周期分辨率不大于 3 个样点的比例为 95.7305%,重建语音质量较好,无明显的"金属声"和"蜂鸣声".

表 3　编码前后基音分辨率比例统计

Δ	Proportion	Δ	Proportion
0	13.746%		
1	22.5079%		
2	9.9682%	≤3	95.7305%
3	49.5082%		
4	1.8415%		
>5	2.428%	>3	4.2695%

4.3　功率的量化

在 2DNMF-WI 编码方案中用 7 比特量化功率,由于编码时每帧需要计算 10 个 CW 功率,两帧功率参数联合组成一个 20 维矢量,编码维数和量化比特数间的矛盾十分突出,为了提高量化效率,首先考虑对功率矢量作降维处理.K-L 变换和 DCT 变换是数据压缩常用的变换方法,但是由于 K-L 变换需要事先知道信源的协方差矩阵并求出特征值,因此存在计算量大,难以实时处理的难题.DCT 变换与 K-L 变换相比,因其具有快速算法,所以计算复杂度适中,并且 DCT 变换与 K-L 变换的压缩性能及量化误差接近,常常被认为是性能接近于 K-L 变换的准最佳变换[19,20].DCT 正变换表达式如下:

$$X(k)=\sqrt{\frac{2}{N}}c(k)\sum_{n=0}^{N-1}x(n)\cos\frac{(2n+1)k\pi}{2N},\quad k=0,1,2,\cdots,N-1 \tag{14}$$

其中,

$$c(k)=\begin{cases}1/\sqrt{2}, & k=0\\ 1, & k=1,2,\cdots N-1\end{cases} \tag{15}$$

为了确定 DCT 系数矢量的维数,我们采用不同维数未量化的 DCT 系数重建输入的功率谱.图 3 给出了不同维数的 DCT 系数重建的功率谱.图中 DCT-12、DCT-10 以及 DCT-5 分别表示用 12、10 和 5 个 DCT 系数重建的功率幅度谱.从图中可以看出,随着 DCT 系数维数的增加,重建的功率更加接近于原始输入的功率.当 DCT 系数的维数为 10 时,重建的功率幅度谱接近于 DCT-12 的功率谱,并且明显好于 DCT-5 重建的功率谱.因此,本文将 DCT 系数的维数设为 10.

对于对数域的 20 维功率矢量做 DCT 变换，变换后的 DCT 系数按能量递减的方式排序，并且能量主要集中于低频系数，由上述分析，只取重排后的前 10 维 DCT 系数进行 7 比特矢量量化，码书设计采用 LBG 算法.

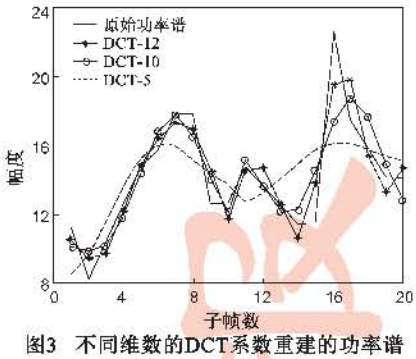


图3 不同维数的DCT系数重建的功率谱

4.4 编码矩阵的量化

为了方便在甚低比特速率下对 CW 进行量化编码，在对 CW 编码前就要对 CW 幅度谱矩阵进行二维非负矩阵分解，它是在行和列两个方向同时压缩 CW 幅度谱矩阵的维数，得到一个低维的 16×6 的编码矩阵，此时的编码矩阵与 NMF 分解 CW 得到的编码矩阵相比维数更小，因此无需对其进行下采样降维处理[8]，在一定意义上保证了编码矩阵的精度. 16×6 维的编码矩阵经过两帧联合维数为 16×12，进而根据编码矩阵行向量之间的独立性，将编码矩阵自上而下分裂为 3 个子矩阵，分别表示为 $\boldsymbol{H}_{5\times12}$、$\boldsymbol{H}_{5\times12}$、$\boldsymbol{H}_{6\times12}$，对这 3 个子矩阵分别以 6,4,3 比特来编码.

本文对基于 2DNMF 分解 CW 幅度谱，采用 13 比特分裂式矩阵量化编码矩阵和基于 NMF[8]分解 CW 幅度谱，对得到的编码矩阵下采样处理并分配 13 比特分裂式矩阵量化两种方案进行了语音质量主观对比 A/B 测试. 测试语音包括 6 句男声语音、6 句女声语音，采样频率均为 8kHz，表 4 为 A/B 测试结果. 实验结果表明基于 2DNMF 分解 CW 幅度谱，采用直接分裂式矩阵量化编码矩阵的方案好于基于 NMF 分解，并对编码矩阵下采样降维处理后进行分裂式矩阵量化的方案.

表 4 主观 A/B 听力测试

测试语音	偏爱基于 NMF 下采样、分裂式矩阵量化	偏爱基于 2D-NMF 分裂式矩阵量化	无偏爱
女声语音	27.08%	33.33%	39.58%
男声语音	22.92%	31.25%	45.83%
所有语音	25%	32.29%	32.71%

5 2DNMF-WI 编码方案的性能评价

2DNMF-WI 语音编码方案中的语音分析帧长为 25ms，编码算法中所有的量化参数均采用两帧联合编码方法，此时超帧帧长为 50ms. 表 5 给出了 1kb/s 2DNMF-WI 算法的比特分配方案.

本文提出的 1kb/s 2DNMF-WI 编码器与 2kb/s NMF-WI 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 12 名年龄在 20～30 岁之间的青年人员组成. 表 6 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 1kb/s 2DNMF-WI 算法的语音质量稍差于 2kb/s NMF-WI 算法.

表 5 1kb/s 2DNMF-WI 编码器的比特分配

参数	比特/超帧	传输速率(比特/秒)
LSF	24	480
基音	6	120
功率	7	140
编码矩阵	13	260
总和	50	1000

表 6 MOS 分测试结果

算法	2kb/s NMF-WI	1kb/s 2DNMF-WI
MOS 分	2.98	2.66

6 结论

综上所述，本文基于文献[8,11]，提出了一种基于二维非负矩阵分解的 1kb/s 甚低速率波形内插语音编码模型. 特征波形经二维非负矩阵分解，原高维的 CW 幅度谱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 更易于量化. 同时，本文根据语音信号的慢渐变特性，提出两帧参数联合编码的方法：对于表征谱包络信息的 LSF 参数，结合其帧内、帧间的相关性，采用帧间后向预测三级矢量量化方案；对两帧联合的基音周期分配 6 比特矢量量化；结合 DCT 降维技术，对降维后的功率进行矢量量化；而对于 CW 二维非负矩阵分解得到的编码矩阵则是采用分裂式矩阵量化方法. 实验结果表明，1kb/s 2DNMF-WI 语音编码算法合成语音的质量稍差于 2kb/s NMF-WI 语音编码算法，本文提出的 1kb/s 2DNMF-WI 算法在自然度方面仍有进一步完善的空间.

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] W B Kleijn, Haagen J. Waveform Interpolation for Coding and Synthesis. Speech coding and Synthesis[M]. Holland: Elsevier Science, 1995. 175－207.

[3] W B Kleijn, J Haagen. Transformation and decomposition of the speech signal for coding[J]. IEEE signal processing letters, 1994, 1(9): 136－139.

[4] N R Chong, I S Burnett, J F Chicharo. Use of pitch synchronous wavelet transform as a new decomposition method for WI[A]. Proceeding of IEEE International Conferance on Acoustics, Speech, Signal Processing[C]. Seattle, Wash, USA:

IEEE,1998.513 – 516.
[5] J Lukasiak ,I S Burnett. Scalable decomposition of speech waveforms[A]. 2002 IEEE Speech Coding Workshop Proceedings[C]. Tsukuba City,Ibaraki,Japan:IEEE,2002.135 – 137.
[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,et al.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:ICSA,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] Zhang Daoqiang, Chen Songcan, Zhou Zhi-hua. Two-dimensional non-negative matrix factorization for face representation and recognition[A]. Proceeding of the ICCV'05 Workshop on Analysis and Modeling of Faces and Gestures[C]. Beijing, China:IEEE,2005.350 – 363.
[12] T Wang, K Koishida. A 1200bps speech coder based on MELP[A]. Proc. ICASSP2000[C]. Istanbul, Turkey: IEEE, 2000.1375 – 1378.
[13] 宾清原,李双田.一种基于 MELP 的高质量的 0.6kb/s 语音编码算法[J].电声技术,2004,36 – 40.
[14] 陈亮,张雄伟.一种 600bps 甚低速率声码器的研究[J].信号处理,2002,18(5):403 – 409.
Chen Liang, Zhang Xiongwei. The study of a new 600bps very low bit-rate vocoder[J]. Signal Processing, 2002, 18(5):403 – 409.(in Chinese)
[15] D D Lee ,H S Seung. Algorithms for nonnegative matrix factorization[A]. Proceedings on Neural Information Processing Systems[C]. Denver, CO, USA: MIT Press, 2000.556 – 562.
[16] 郭莉莉,鲍长春.基于贝叶斯阴阳机的 2kb/s NMF-WI 语音编码算法[J].电子学报,2009,37(5):1146 – 1153.
Guo Li-li, Bao Chang-chun. 2kb/s bayesian ying-yang waveform interpolative speech coding based on non-negative matrix factorization[J]. Acta Electronica Sinica,2009,37(5):1146 – 1153.(in Chinese)
[17] Er-juan Xue, Chang-chun Bao. 1kb/s waveform interpolation speech coding based on non-negative matrix factorization[A]. 2008 9 th International Conference on Signal Processing Proceedings[C]. Beijing:IEEE,2008,1:526 – 529.
[18] Thomas Eriksson, Hong-Goo Kang. Pitch Quantization in low bit-rate speech coding[A]. IEEE ICASSP[C]. Phoenix, AZ, USA:IEEE,.1999,.(1).489 – 492.
[19] Sikora T,Makai B. Shape-adaptive DCT for generic encoding of video[J]. IEEE Trans on Circuits system. 1995,5(2):59 – 62.
[20] 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.
Luo Yafei, Bao Changchun. 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)
[21] 高宏娟,潘晨.基于非负矩阵分解的人脸识别算法的改进[J].计算机技术与发展,2007,17(7):63 – 66.

作者简介:

薛二娟　女,1982 年生于河北石家庄,北京工业大学电子信息与控制工程学院硕士研究生,研究方向为语音信号处理、窄带语音编码.
E-mail:xueerjuan@emails.bjut.edu.cn

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:chchbao@bjut.edu.cn

李如玮　女,1972 年生于四川眉山,在读博士,副教授,主要研究方向:数字语音信号处理和小波变换.
E-mail:liruwei@bjut.edu.cn