面向窄带通信的极低速率语音编码算法研究

刘 斌[1] 陶建华[1] 莫福源[2]
(1. 中国科学院自动化研究所 模式识别国家重点实验室，北京 100190；
2. 中国科学院声学研究所 北京 100190)

摘 要：提出了一种面向窄带通信的极低速率参数语音编码算法。在2.4kbps MELP标准的基础上结合听觉感知，对线谱对参数进行联合矢量量化、对基音周期进行内插和非线性量化、对能量参数进行高效压缩，可以使语音数据在0.5kbps下匀速传输；线谱对参数的预测残差用于矢量量化，这是一种提高合成语音的音质的有效方法。实验结果表明，采用本文提出的语音编码算法可以使语音数据在极低码率下有效的传输，解码端合成的语音具有较高的可懂度。

关键词：联合矢量量化；非线性量化；预测残差；听觉感知
中图分类号：TN912 **文献标识码**：B **文章编号**：1003-0530(2013)09-1134-08

Research on Speech Coding Algorithm at Very Low Bit Rate for Narrow-Band Communication

LIU Bin[1] TAO Jian-hua[1] MO Fu-yuan[2]
(1. National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing 100190;
2. Institute of Acoustics, Chinese Academy of Sciences, Beijing 100190)

Abstract: A speech coding algorithm applied to narrowband transmission at very low bit rate is proposed. Based on the MELP standard working at 2.4kbps, joint vector quantization is applied to linear spectrum pair's parameters, interpolation and nonlinear quantization are applied to pitch, gain parameters are compressed efficiently and auditory perception is considered, speech data could be coded and transmitted smoothly at 0.5kbps The predicative residuals of linear spectrum pair's parameters are also used for vector quantization; it's a suitable way of improving synthetic speech quality. The experimental results reveal that the proposed algorithm can ensure speech effective transmission at very low bit rate while the intelligibility of received speech is acceptable.

Key words: joint vector quantization; nonlinear quantization; predicative residuals; auditory perception

1 引言

随着计算机科学和信号处理技术的不断发展，语音编码技术在最近几十年有了长足的进步，各种语音编码标准不断涌现，在高码率下合成高质量的语音已非难题。但随着码率的降低，特别是降至1kbps以下时，音质下降严重，有时甚至完全无法理解。针对这一问题，Tokuda等提出采用“识别-合成”的方法实现极低速率语音编码[1]，编码端对语音信号进行音素识别，解码端根据音素序列和各音素时长通过统计参数模型合成语音；虽然该方法可在极低码率下合成出高质量语音，但是音素识别准确率直接影响系统性能，同时，说话人特征也很难保留。用变长声码器[2]虽然可大幅度降低码率，但是无法保证语音数据在低码率时匀速传输。目前极低速率语音编码的主流研究方法是用语音信号

收稿日期：2013-06-09；修回日期：2013-08-07
基金项目：国家自然科学基金(61273288, 61233009, 61203258, 61011140075, 90820303)；中国新加坡数字媒体研究院资助项目(CSIDM)

的短时平稳性和相邻帧之间的相关性,采用多帧数据联合编码。Tian Wang 等用3 帧联合的超帧结构,高效量化不同语音参数,提出一种 1.2kbps 的语音编码方法[3];Chamberlain 对谱参数和能量相关的增益参数联合量化、用均值加形状的方法量化基音周期,实现了 0.6kbps 的语音编码算法[4]。Xia Zou 等利用基音周期和增益的相关性,将它们组成联合矢量进行量化,提出一种 0.6kbps 的语音编码算法[5]。对于语音信号,清音和浊音的特征参数存在着较大差异,Chaogang Wu 等通过分模式模式对语音参数进行量化,提高了解码端的语音音质[6]。E. Unver 等融合了分清浊模式量化和联合多帧量化,实现了一种 0.8kbps 的正弦激励语音编码算法,在低码率下保证了语音的音质[7]。近年来,压缩感知理论在语音信号处理中得到了应用[8],肖强等利用联合帧线谱对参数的稀疏性将压缩感知理论应用到低速率语音编码,实现了对谱参数的高效压缩,这种方法虽然可以有效的降低码率,但是重构语音的音质较差[9]。Laura 等将梅尔倒谱系数(Mel Frequency Cepstrum Coefficient, MFCC)作为特征参数设计了一种低速率下的语音编码算法,相比于文献 4 中的方法合成语音的音质略有提升,但是解码端参数重构算法的复杂度较高,制约了这种方法在实际中的应用[10]。联合矢量量化是一种降低码率的有效方法,它可保证在 0.6kbps 码率下合成相对可懂的语音,但是当码率进一步降低时,解码端的语音音质会严重下降。

极低速率语音编码多用参数编码,目前最为成熟的参数语音编码方法是 2.4kbps 的 MELP 标准[11],MELP 标准具有良好的编码性能,它在传统的二元激励线性预测 (Linear Predictive Coding, LPC)声码器模型的基础上从 5 个方面进行了合理的改进:脉冲与噪声混合多带激励源;周期与非周期激励法相结合;自适应谱增强;脉冲扩展滤波器;残差谱的傅立叶幅度建模。因此,在相同编码速率的语音质量方面,MELP 编码器要优于 LPC-10 编码器。本文在 MELP 标准的基础上进行改进,实现 0.5kbps 速率下语音的匀速传输。

文章第 2 部分重点阐述了作者提出的 0.5kbps 码率下的语音编码算法,文章第 3 部分通过实验对主观评测结果和客观评测结果进行分析,文章第 4 部分对本文提出的方法进行总结,并对后续可改进之处进行展望。

2　适用于窄带信道传输的极低码率语音编码算法

2.1　MELP 标准在极低码率下进一步压缩的途径

在 MELP 标准中,谱参数通过多级矢量量化实现压缩,这种方法可以有效的减少谱参数的量化误差,但是逐帧进行矢量量化不仅数据压缩的空间小,而且忽略了相邻语音帧之间的相关性。针对这一问题,可以对多帧数据进行联合矢量量化,在有效的对谱参数进行压缩的同时充分考虑了相邻语音帧之间的相关性。对于清音帧基音周期没有物理意义,因此可以根据相邻帧的清浊组合采用不同的编码方式,减少比特分配数。由于能量对听觉感知的影响最弱,如果对能量相关的增益参数进行高效压缩对听感的影响很小。对于极低速率语音编码,首先需要确保合成语音的可懂度,因此各频带的清浊状态、非周期脉冲标志和傅里叶级数的幅值远没有谱参数和基音周期重要,在这种条件下可以不对上述参数分配比特。本文设计的算法只需对谱参数、基音周期和增益进行编码,可以保证在 0.5kbps下合成的语音具有较高的可懂度。

2.2　极低码率下语音编码算法的整体设计

通过电磁波传输的语音信号通常要求具有低延迟性,对于移动蜂窝通信系统,允许最大延迟不超过 100ms。本文以 4 帧为一组进行联合编码,延迟 90ms 可以满足这一的需求,这种算法也可应用于延迟要求较低的水声通信。语音编码的原理框图如图 1 所示。

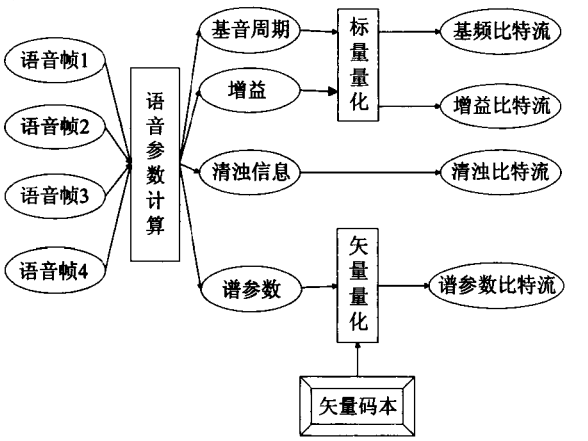

图 1　0.5kbps 下的语音编码流程

Fig. 1　The flowchart of speech coding at 0.5kbps

图1中,4个连续子帧组成一个联合帧,每个子帧时长22.5ms,采用MELP标准中的方法计算各子帧的基音周期、谱参数和增益(本文提到的增益为MELP标准中反映整帧能量的增益1),同时对各子帧的清浊状态进行判定。前两个子帧和后两个子帧的谱参数利用相应的矢量码本分别进行联合矢量量化。对偶数子帧的增益进行标量量化。基音周期根据联合帧中各子帧的清浊组合选择对听感影响较大的关键子帧进行标量量化,0.5kbps下语音编码的比特分配如表1所示。

表1 0.5kbps下语音编码比特分配数(4帧合计)

Tab.1 Bit allocation at 0.5kbps (one super-frame)

参数类型	比特数
谱参数	11+11
基音周期	12
增益	8
清浊模式	3
合计	45

2.3 谱参数的矢量量化

在MELP标准中,对谱参数进行多级矢量量化。该方法可有效地降低量化误差,但需分配相对较多的比特数,在极低码率语音编码中无法直接应用;实际上,相邻语音帧的谱参数之间存在相关性,清音和浊音、稳态帧与过渡帧之间的声道特性存在着明显的差异,在矢量量化时若能充分考虑这些特性,则可在一定程度上提高合成语音音质。

2.3.1 联合矢量生成

对于任意一段时长90ms的联合帧,它的4个子帧的线谱对参数分别为 $lsf_1 \sim lsf_4$,第 i 个子帧的线谱对参数可表示为

$$lsf_i=[lsf_{i1},lsf_{i2}\cdots lsf_{i10}] \tag{1}$$

本文用两帧联合的方式对谱参数进行矢量量化,因此对于包含4个子帧的联合帧可用两个联合矢量 $lsf_A=[lsf_1,lsf_2]$ 和 $lsf_B=[lsf_3,lsf_4]$ 表示谱参数。

2.3.2 谱参数预测残差计算

为了减少谱参数的动态范围,在对谱参数进行联合矢量量化前,首先去除谱参数的均值,线谱对参数的均值通过对全部训练数据集的线谱对参数计算均值得到。其均值表示为

$$lsf_{dc}=[lsf_{dc1},lsf_{dc2}\cdots lsf_{dc10}] \tag{2}$$

则联合帧中第 i 个子帧的线谱对参数去均值后可表示为

$$lsf'_i=lsf_i-lsf_{dc} \tag{3}$$

利用前一个联合矢量末帧的谱参数对当前联合矢量中各子帧的谱参数进行线性预测。因此需要确定预测系数矢量,使预测残差在统计意义上最小。本文用去均值后的前一联合矢量的末帧 lsf'_{pre} 对当前联合矢量的各子帧谱参数进行预测。设 α_m 表示联合矢量中第 m 个子帧的预测系数,$relsf_m$ 表示对该子帧谱参数的预测残差。则

$$relsf_m=lsf'_m-\alpha_m\cdot lsf'_{pre} \tag{4}$$

采用最小均方误差准则计算最优预测系数为

$$\alpha_{mi}=\frac{\sum_{n=1}^{N} lsf'_{mi}lsf'_{prei}}{\sum_{n-1}^{N}(lsf'_{prei})^2} \tag{5}$$

其中,i 表示第 i 阶去均值的线谱对参数;根据联合矢量中各子帧清浊模式的不同组合选择不同的预测系数,因此 N 表示训练数据中相应清浊组合模式下语音帧的数量。经过上述计算后,对于联合帧可得两个表示线谱对预测残差的联合矢量 $relsf_A$ 和 $relsf_B$。

$$relsf_A=[relsf_1,relsf_2] \tag{6}$$

$$relsf_B=[relsf_3,relsf_4] \tag{7}$$

2.3.3 矢量码本的训练

由于清音和浊音的谱参数存在着较大的差异,在联合矢量量化时,根据清浊组合的不同分别进行码本训练。通过训练可以得到“浊音-浊音”、“浊音-清音”、“清音-浊音”、“清音-清音”共4个矢量码本。

矢量码本的训练音库要覆盖不同说话人、多种说话风格以及各种录音环境的语音,以保证码本的泛化能力。对训练音库中的语音采用2.3.2节的方法计算线谱对参数的预测残差,并根据语音帧的清浊状态对预测残差样本进行分组,生成4类谱参数训练数据分别进行后续训练。

对于任意两个矢量码本,在训练过程中需要结合听觉感知计算它们的计权距离。设 $relsf_s$ 和 $relsf_t$ 为任意两个表示线谱对参数预测残差的20维矢量,则它们的距离可表示为

$$d^2(relsf_s,relsf_t)=\sum_{i=1}^{20}\omega_i\nu_i(relsf_{si}-relsf_{ti})^2 \tag{8}$$

其中

$$\omega_i=P(lsf_{si})^{0.3} \qquad 0<i<21 \tag{9}$$

$$v_i=\begin{cases}1 & 0<i<9 \text{ 或 } 10<i<19\\ 0.64 & i=9 \text{ 或 } i=19\\ 0.16 & i=10 \text{ 或 } i=20\end{cases} \quad (10)$$

式(9)中的 $P(lsf_{si})$ 表示逆预测滤波器的功率谱在频率点 lsf_{si} 处的取值[11]。对于任意模式下的训练数据,首先用文献[12]的方法确定初始码本[12],可在保证初始码本质量的同时提高了训练效率。以这些初始码本为质心采用 LBG 算法[13]对该模式下的所有训练数据进行迭代训练,直到收敛。训练完成后可得到不同清浊组合模式下的谱参数矢量码本,每种模式下各包含 2048 个矢量。

MELP 标准中的多级矢量量化虽然可以节省码本存储空间,但是在每一级矢量量化过程中需要保留 8 个备选码本用于下一级码本检索,相比于本文提出的通过 2048 个矢量进行单级矢量量化,算法的时间复杂度并没有得到优化;本文提出的谱参数量化方法虽然提高了算法的空间复杂度,但是并不影响算法的软硬件实现。

2.3.4　分模式预测型矢量量化

对于输入的包含两个子帧的线谱对参数联合矢量,采用 2.3.2 节的方法计算线谱对参数的预测残差,然后根据联合矢量中各子帧的清浊模式组合,选择相应的谱参数码本集,通过式(8)~(10),对谱参数预测残差进行联合矢量量化。需要说明的是,对于输入语音信号的第一个联合矢量,在去除直流分量后,直接通过相应模式下的谱参数码本集进行联合矢量量化。

通过联合矢量量化可以充分考虑相邻语音帧之间的相关性;采用分模式矢量量化,将不同的码本进行分组训练可以有效的提高量化精度;在矢量量化过程中将线谱对参数的预测残差作为特征参数,这种预测型矢量量化可以有效的缩小训练数据的动态范围,降低谱参数的量化误差,从而有助于在低码率下提高合成语音的音质。

2.4　基音周期的量化

对于联合帧中的各子帧的基音周期,需要将其变换到对数域,然后在对数域进行线性量化。MELP 标准中通过 7 比特量化每个子帧的基音周期。由于语音信号的基音周期分布得不均匀,对 7 比特量级的基音周期通过合理的非线性压缩可以将其映射到 6 比特,采用这种方式几乎可以达到 7 比特的量化性能。本文根据基音周期的统计分布特性,将 99 量级的基音周期压缩至 63 量级,使得任意一帧语音的基音周期能够用 6 比特量化。对基音周期进行非线性压缩时合并的量级如表 2 所示。

表 2　基音周期非线性量化映射关系

Tab. 2　Nonlinear quantization mapping relationship of pitch

合并的量级	合并后的量化值	合并的量级	合并后的量化值
0~8	1.33752	72~74	1.96694
9~12	1.40137	75~79	2.00342
13~14	1.41962	80~84	2.04903
15~16	1.44698	85~90	2.09464
17~18	1.46523	90~98	2.15850

本文在进行基音周期量化时,根据联合帧中 4 个子帧清浊组合的不同采用不同的比特分配方案。在对基音周期进行编码时充分考虑了听觉感知的特性,稳定的浊音段尽可能细得量化,而清浊之间的过渡段量化粒度相对较粗,对于基音周期没有物理意义的清音帧不进行量化,通过内插和差分量化等方式削减比特数;相比于文献 4 中均值加形状的基音周期量化方法,可以有效的解决相邻联合帧边界处基音周期不平滑对听感的影响。

表 3　0.5kbps 下语音编码基音周期比特分配(4 帧合计)

Tab. 3　Pitch quantization bit allocation (one super-frame)

清浊组合	清浊模式编码	基音周期比特流 1	基音周期比特流 2
uuuu	000	置 0	置 0
uuuv	001	置 0	第 4 帧码值
uuvu	010	置 0	第 3 帧码值
uuvv	000	第 4 帧码值	第 3 帧差分码值
uvuu	011	置 0	第 2 帧码值
uvuv	001	第 4 帧码值	置 0
uvvu	010	第 2 帧码值	置 0
uvvv	001	第 3 帧码值	第 4 帧码值
vuuu	100	置 0	第 1 帧码值
vuuv	010	第 1 帧码值	第 4 帧码值
vuvu	011	第 1 帧码值	置 0
vuvv	011	第 1 帧码值	第 4 帧码值
vvuu	100	第 1 帧码值	第 2 帧差分码值
vvuv	101	第 1 帧码值	第 4 帧码值
vvvu	110	第 1 帧码值	第 2 帧码值
vvvv	111	第 2 帧码值	第 4 帧码值

注:u 表示清音帧,v 表示浊音帧

基音周期和清浊模式的比特分配如表3所示，其中基音周期比特流1和基音周期比特流2分别包含6比特数据，清浊组合表示联合帧中从第1个子帧到第4个子帧的清浊状态，基音周期的量级范围是1～63，基音周期的差分量级范围是1～15。对于任意一个子帧，如果对其分配6比特，表示对该子帧的基音周期通过7比特线性量化后，再经非线性量化压缩至6比特，如果对其分配4比特，表示利用相邻浊音帧的基音周期对当前子帧的基音周期进行差分编码。

2.5 增益的量化

对于联合帧中各子帧的增益，通过MELP标准可分别计算反映前半帧能量（增益0）和反映整帧能量（增益1）的两个增益参数，由于人耳对能量误差不敏感，本文只对增益1进行编码。考虑到语音信号具有短时平稳的特性，在对联合帧进行语音编码时，只对4个子帧中的偶数帧的增益进行4比特线性量化，其余的增益参数可以在解码时通过相邻帧内插的方式确定。在比特数分配有限的条件下，文献4中多帧增益联合量化的方法容易将清音帧的能量误放大从而影响语音音质，采用本文的方法可以有效的避免这一问题。经过上述压缩后，4个子帧的全部增益参数可以用8比特表示。

2.6 语音参数的解码

语音解码模块通过对谱参数比特流、基音周期比特流、增益比特流和清浊比特流进行解析，重构语音参数。接收端首先对清浊状态进行判定，确定联合帧中各子帧的清浊组合，在此基础上重构其余语音参数。

谱参数需要根据联合矢量中各子帧清浊组合模式的不同，选择所对应的矢量码本重构联合矢量的预测残差，然后根据相应模式下的预测系数和线谱对参数的直流分量分别计算前两个子帧和后两个子帧的线谱对参数。通过式(11)重构线谱对参数。

$$\hat{lsf}_i = lsf_{dc} + \alpha_m \cdot (\hat{lsf}_{pre} - lsf_{dc}) + \hat{relsf}_i \qquad (11)$$

式(11)中，$\hat{lsf}_i$ 表示重构的线谱对参数，lsf_{dc} 表示线谱对参数的直流分量，α_m 表示谱参数预测系数，$\hat{lsf}_{pre}$ 表示前一联合矢量末帧线谱对参数的重构值，$\hat{relsf}_i$ 表示谱参数预测残差的量化值。

对于增益参数，先通过增益比特流对偶数子帧的增益进行解码，重构它们的增益值。其余增益参数通过相邻两帧内插确定。在进行内插计算时，相邻帧的内插系数可取0.5。

基音周期解码需要根据联合帧中各子帧清浊组合的不同选择相应的解码方式。清音帧的基音周期为0，无需对其解码；如果某一浊音帧被分配6比特编码，则先将其非线性解压至7比特，然后重构它的基音周期；如果某一浊音帧被分配4比特编码，则将当前子帧的码值和相邻浊音帧的码值叠加后再重构基音周期；如果某一浊音帧未分配比特编码，则采用内插的方式计算基音周期。

本文提出的语音编码算法只考虑周期脉冲和高斯白噪声两种激励，傅里叶幅值统一取单位长度，因此无需对各频带的清浊状态、非周期脉冲标志和傅里叶级数的幅值进行编解码。

2.7 参数语音合成

在完成语音参数解码后，通过各子帧的语音参数在极低码率下生成语音。激励源采用二元激励，清音采用高斯白噪声作为激励信号，浊音采用周期脉冲作为激励信号。通过自适应谱增强模块有效的弥补LPC滤波器无零点的缺点，增强合成语音共振峰的结构；以每个基音同步周期为单元，对增益进行调节；通过脉冲扩散滤波器，将激励信号的能量在一个基音周期中进行扩散，从而减少在合成语音中的刺耳的成分，最后合成经过0.5kbps码率下压缩后的语音。

3 实验及结果分析

3.1 实验数据和实验方法

实验数据：共包括10个8k采样率、16位线性量化的测试样本，覆盖了不同说话人、不同录音环境下采集的语音。其中测试样本1～4为男性朗读风格语音，测试样本5～8为女性朗读风格语音，测试样本9和10为男女对话的口语语音。

在0.5kbps码率下，连续4帧组成一个联合帧进行语音编码，根据表1对各种语音参数分配比特数。共包括5组对比实验，用于对比本文的方法和其他对谱参数、基音周期和增益的量化方法。

实验1用于评估采用线谱对参数的预测残差进行矢量量化后的改进效果；对比本文方法，实验1的对比方法选择原始线谱对参数作为特征参数根据表1的比特分配方案分清浊模式训练码本，对测试

语音的谱参数进行联合矢量量化并采用本文提出的方法对基音周期和增益进行量化。

实验 2 用于评估谱参数单级矢量量化对语音音质的改进效果;基音周期和能量的量化方式与实验 1 相同,对比本文方法,实验 2 的对比方法采用两级矢量量化(第一级分配 7 比特,第二级分配 4 比特)得到谱参数码本,利用线谱对参数的预测残差作为特征参数分清浊模式训练码本对测试语音的谱参数进行联合矢量量化。

实验 3 分别采用均值加形状和本文提出的内插非线性量化对基音周期进行量化;对比本文方法,实验 3 的对比方法中基音周期的均值用 7 比特量化,基音周期的形状用 5 比特量化,其余参数按照本文提出的方法进行量化。

实验 4 分别采用多帧增益联合量化和本文使用的内插方式进行增益量化;对比本文方法,实验 4 的对比方法中通过 8 比特增益码本对联合帧中 4 个子帧的增益分清浊模式进行联合量化,其余参数根据本文提出的方法进行量化。

实验 5 用于评估对基音周期和增益量级进行压缩前后的语音音质;对比本文方法,实验 5 的对比方法中基音周期采用 99 量级进行量化、增益采用 32 量级进行量化,谱参数根据本文提出的方法进行量化;对比方法的码率约为 0.544kbps。

3.2　客观评测结果

ITU-T 的 P.862 标准[14] 提出了 PESQ 方法,它是目前基于心理声学模型的客观评价算法中获得广泛使用一个方法。本文采用 PESQ 方法作为客观评测标准,对测试语音采用本文提出的方法和五组实验中的对比方法进行 PESQ 评估,客观评测结果如表 4 所示。

3.3　主观评测结果

人耳的听觉感受是评估语音编码结果的重要因素,将本文提出的方法分别和 3.1 节中五组实验的对比方法进行 ABX 主观测试,参与测试的人员均是从事语音信号处理研究的科研人员。在各组测试中,听者通过比较本文方法和各种对比方法编解码后合成的语音,根据主观听感评估各测试样本通过哪种方法生成的语音效果更佳。由于语音是在极低码率下生成的,因此在主观评测时首先考虑可懂度,其次考虑清晰度。图 2 ~ 图 6 为主观测试实验结果,用于表述全部测试人员对所有测试样本采用不同方法进行对比的测试结果。

表 4　PESQ 客观评测得分

Tab.4　PESQ objective evaluation scores

音库编号	本文方法	实验 1 对比方法	实验 2 对比方法	实验 3 对比方法	实验 4 对比方法	实验 5 对比方法
1	2.347	2.323	2.337	2.142	2.356	2.233
2	2.565	2.445	2.481	2.511	2.596	2.597
3	2.786	2.705	2.778	2.761	2.726	2.780
4	2.603	2.585	2.582	2.507	2.590	2.623
5	1.659	1.656	1.614	1.670	1.598	1.636
6	2.329	2.317	2.323	2.270	2.283	2.387
7	2.558	2.530	2.545	2.646	2.429	2.566
8	2.225	2.193	2.211	2.256	2.167	2.334
9	2.407	2.279	2.363	2.373	2.370	2.405
10	2.156	2.020	2.134	2.051	2.167	2.185
均值	2.364	2.305	2.337	2.319	2.328	2.375

第一组主观测试对应于 3.1 节中的实验 1,验证采用线谱对参数的预测残差进行矢量量化后的改进效果,实验结果如图 2 所示。

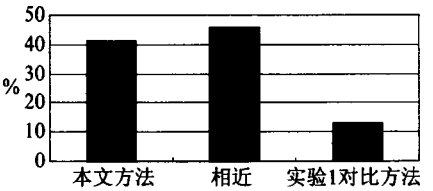

图 2　利用线谱对参数的预测残差进行矢量量化的改进效果

Fig.2　Evaluating effect of LSF predictive residual vector quantization

第二组主观测试对应于 3.1 节中的实验 2,用于评估谱参数的单级矢量量化和多级矢量量化,实验结果如图 3 所示。

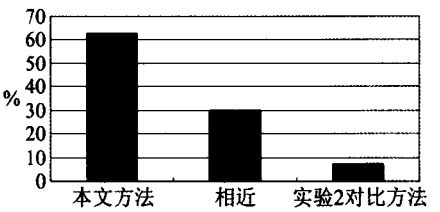

图 3　谱参数不同矢量量化方式比较

Fig.3　Comparison of different vector quantization for spectrum

第三组主观测试对应于 3.1 节中的实验 3,对比采用均值加形状量化基音周期和本文使用的通过内插、非线性量化方式量化基音周期。实验结果

如图4所示。

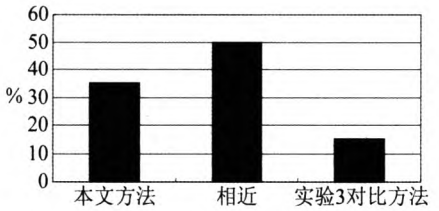

图4 基音周期量化方法的比较

Fig. 4 Comparison of pitch quantization method

第四组主观测试对应于3.1节中的实验4，对比采用多帧增益联合量化和本文采用的通过内插方式进行增益量化。实验结果如图5所示。

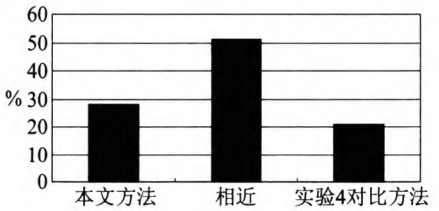

图5 增益量化方法的比较

Fig. 5 Comparison of gain quantization method

第五组主观测试对应于3.1节中的实验5，用于评估对基音周期和增益量级进行压缩前后的语音音质。实验结果如图6所示。

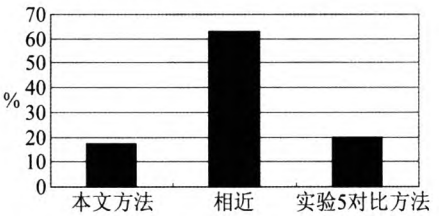

图6 基音周期和能量压缩前后对比

Fig. 6 Evaluating effect of pitch compression and gain compression

3.4 实验结果分析

主观评测结果和客观评测结果均表明，利用线谱对参数的预测残差进行矢量量化有助于音质的提高，相比于直接对线谱对参数进行联合矢量量化，全部测试样本的PESQ得分更高，主观评测更倾向于这种方法；这是由于通过分模式预测型矢量量化缩小了样本的动态范围，从而降低了量化误差。相比于两级矢量量化，在比特数相同的条件下，通过单级矢量量化得到的谱参数码本对测试样本进行编解码后合成语音的音质更好，这种方法的客观评测得分更高，ABX主观测试也更倾向于这种方法。

采用内插和非线性量化方式对基音周期进行量化在分配相同比特数的条件下优于均值加形状的量化方法。测试的10个语音样本中有7个样本PESQ得分更高，从ABX主观测试结果分析，采用本文的方法对基音周期量化效果更好；这是由于通过均值加形状对基音周期量化时对相邻联合帧边界处的基音周期处理得不理想。采用内插方式对增益进行量化在比特分配有限的条件下优于多帧增益联合量化的方法。通过内插方式对增益量化的客观评测得分更高，ABX主观测试也更倾向于这种方法。这是由于在比特分配较少的情况下，多帧增益联合量化容易将清音帧的能量误放大从而降低合成语音的清晰度。通过主观评测和客观评测分析，对基音周期和增益量级进行压缩前后的语音音质比较接近，这是由于人耳对能量的失真和基音周期的失真没有谱参数的失真敏感。

本文提出的语音编码算法在MELP标准的基础上，对语音信号采用4帧一组进行联合语音编码，该算法延迟90ms，可以满足蜂窝移动通信的要求，同时也可以将其应用到水声通信领域。本文对基音周期和能量进行内插和标量量化，相对于MELP标准，并没有提高算法的复杂度；文中对谱参数采用单级码本进行矢量量化，通过2.3.3节分析，虽然增加了码本的存储空间，但是并不影响算法的工程实现且时间复杂度并未明显增加。本文的算法继承了MELP标准中分析端的高通滤波和噪声抑制模块以及合成端的增益抑制模块，在一定程度上保证了算法的音源韧性，同时继承了MELP标准中的语音传输方式，可以保证数据传输时的帧同步。

4 结论

针对极低速率语音编码的客观要求，本文在MELP标准基础上对语音信号采用4帧一组进行联合语音编码；通过对谱参数进行两帧联合矢量量化并利用谱参数的预测残差作为特征参数进行矢量量化，从而在极低码率下有效的保证了合成语音的音质；通过对基音周期进行内插和非线性量化、采用内插的方式对能量相关的增益参数进行高效压缩，从而保证对听感影响相对较小的前提下有效的降低了码率，实现了一种可以在0.5kbps下匀速传输的语音编码算法。

对于参数语音编码,不同类型的参数之间存在着相关性,如果在已知谱参数和基音周期的条件下,通过不同语音参数之间的相关性对极低码率下没有考虑的各频带清浊状态、非周期脉冲标志等参数进行预测,合成语音的音质从理论上分析会进一步提升。本文提出的方法对带噪语音的处理能力相对较弱,如果能够将该方法与语音增强有机结合,可能会带来更广泛的应用。

参考文献

[1] Tokuda K, Masuko T, Hiroi J, et al. A very low bit rate speech coder using HMM-based speech recognition/synthesis techniques[C]// Acoustics, Speech and Signal Processing, 1998. Proceedings of the 1998 IEEE International Conference on. IEEE, 1998, 2: 609-612.

[2] Crosmer J, Barnwell III T. A low bit rate segment vocoder based on line spectrum pairs[C]// Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP'85. IEEE, 1985, 10: 240-243.

[3] Wang T, Koishida K, Cuperman V, et al. A 1200 bps speech coder based on MELP[C]// Acoustics, Speech, and Signal Processing, 2000. ICASSP'00. Proceedings. 2000 IEEE International Conference on. IEEE, 2000, 3: 1375-1378.

[4] Guilmin G, Capman F, Ravera B, et al. New NATO STANAG narrow band voice coder at 600 bits/s[C]// Acoustics, Speech and Signal Processing, 2006. ICASSP 2006 Proceedings. 2006 IEEE International Conference on. IEEE, 2006

[5] Zou X, Wen C, Zhang X, et al. An improved 600bps speech coding based on joint quantization of pitch and gain shape[C]// Communication Technology (ICCT), 2010 12th IEEE International Conference on. IEEE, 2010: 1303-1306.

[6] Wu C, Jiang H, Li B. An improved MELP speech coder[C]// Information Technology and Computer Science, 2009. ITCS 2009. International Conference on. IEEE, 2009, 2: 130-133.

[7] Unver E, Villette S, Kondoz A. Joint quantisation strategies for low bit-rate sinusoidal coding[J]. Signal Processing, IET, 2010, 4(5): 548-559.

[8] 季云云, 杨震. 基于主分量分析的语音信号压缩感知[J]. 信号处理, 2011, 27(7): 1057-1062.
JI Y, YANG Z. PCA-Based Compressed Speech Signal Sensing[J]. Signal Processing, 2011, 27(7): 1057-1062. (in Chinese)

[9] 肖强, 陈亮, 朱涛, 等. 基于压缩感知的线谱对参数降维量化算法[J]. 信号处理, 2011, 27(4): 563-568.
XIAO Q, CHEN L, ZHU T, et al. Dimension Reduction Quantization of LSP Parameters Based on Compressed Sensing[J]. Signal Processing, 2011, 27(4): 563-568. (in Chinese)

[10] Boucheron L E, Leon P L D, Sandoval S. Hybrid scalar/vector quantization of mel-frequency cepstral coefficients for low bit-rate coding of speech[C]// Data Compression Conference (DCC), 2011. IEEE, 2011: 103-112.

[11] ITU-T. Federal information processing standards publication (MELP), specifications for the analog to digital conversion of voice by 2400 bit/second mixed excitation linear prediction, Draft June 1997.

[12] 肖东,莫福源,陈庚,等. 混合激励线性预测语音编码标准中线谱频率量化的研究[J]. 应用声学,2012,3: 109-117.
XIAO D, MO F, CHEN G, et al. Studies of the line spectral frequency vector quantization in mixed excitation linear prediction[J]. Applied Acoustics, 2012, 3: 109-117. (in Chinese)

[13] Yanxia L, Jiawei Y, Ye L. One effective method to design LBG initial codebook[C]// Intelligent Computation Technology and Automation (ICICTA), 2011 International Conference on. IEEE, 2011, 2: 628-631.

[14] Rix A W, Beerends J G, Hollier M P, et al. Perceptual evaluation of speech quality (PESQ)-a new method for speech quality assessment of telephone networks and codecs[C]// Acoustics, Speech, and Signal Processing, 2001. Proceedings. (ICASSP'01). 2001 IEEE International Conference on. IEEE, 2001, 2: 749-752.

作者简介

刘　斌　男,1984 年 12 月生,内蒙古人。中国科学院自动化研究所博士研究生。研究方向为低速率语音编码、语音信号处理等。E-mail: liubin@nlpr.ia.ac.cn

陶建华　男,1972 年 6 月生,江苏人。中国科学院自动化研究所研究员,博士研究生导师。研究方向为语音信号处理、语音合成、语音交互技术等。E-mail: jhtao@nlpr.ia.ac.cn

莫福源　男,1942 年生,江苏人。中国科学院声学研究所研究员,博士研究生导师。研究方向为语音编码、水声信号处理等。E-mail: mofuyuan@aliyun.com