文章编号: 1003-0530(2020)06-0894-08

基于 G. 722. 1 的分布式语音编码

何莹男 陈 喆 殷福亮
(大连理工大学信息与通信工程学院,辽宁大连 116023)

摘 要: 在语音通信网络中,为获得良好的语音通信质量,抗丢包技术不可或缺。为此,本文基于 ITU G. 722. 1 语音编码器,提出一种分布式语音编码方法。该方法在 G. 722. 1 编码器的基础上,构建一个互补编码器; 然后在编码端,对同一帧语音分别用 G. 722. 1 编码器和其互补编码器进行语音编码,并发送编码结果; 在解码端,在接收到其中任一语音码流时,用 G. 722. 1 解码器进行解码,其语音质量不低于 G. 722. 1 编码器的解码结果,而在接收到两个语音码流时,用 G. 722. 1 解码器先分别对两个语音码流进行解码,然后对解码结果进行联合处理,其最终的语音质量有明显提升,即有一定编码增益。仿真实验结果表明,本文分布式语音编码方法的抗丢包效果明显,相对于原始编解码器其语音质量进一步提升。

关键词: 语音编解码器; 分布式处理; 抗丢包; 感知加权滤波器
中图分类号: TN301 **文献标识码**: A **DOI**: 10. 16798/j. issn. 1003-0530. 2020. 06. 011

引用格式: 何莹男,陈喆,殷福亮. 基于 G. 722. 1 的分布式语音编码[J]. 信号处理,2020,36(6): 894-901. DOI: 10. 16798/j. issn. 1003-0530. 2020. 06. 011.
Reference format: He Yingnan, Chen Zhe, Yin Fuliang. Distributed Speech Coding Based on G. 722. 1 Codec[J]. Journal of Signal Processing, 2020, 36(6): 894-901. DOI: 10. 16798/j. issn. 1003-0530. 2020. 06. 011.

Distributed Speech Coding Based on G. 722. 1 Codec

He Yingnan Chen Zhe Yin Fuliang
(College of Information & Communication Engineering, Dalian University of Technology, Dalian, Liaoning 116023, China)

Abstract: In a voice communication network, packet loss concealment technology is crucial for ideal voice communication. A distributed speech coding method based on the ITU G. 722. 1 speech encoder is proposed in this paper. The main process can be divided into several stages. First, A complementary encoder is built based on the the G. 722. 1 encoder. Then, the encoder sends the encoding results of both G. 722. 1 encoder and the complementary encoder for the same frame of speech. When the decoder receives one of the encoding results, it is decoded by the G. 722. 1 decoder, the speech quality is higher than the decoding result of the G. 722. 1 encoder. At the same time, when both encoding results are received, the G. 722. 1 decoder is used to decode the two voice code streams respectively, and then the decoding results are weighted, the final voice quality is significantly improved. In other words, there is a certain coding gain. The simulation experiment results show that the distributed speech coding algorithm proposed in this paper can improve the encoding quality significantly compared to the traditional coding techniques.

Key words: speech codec; distributed processing; packet loss concealment; perceptual weighted filter

1 引言

在现代通信技术中,语音实时通信技术占据重要地位。如何在复杂的网络环境下高质量的传输语音一直是人们的研究热点。随着网络技术的迅猛发展,基于 IP 的语音传输[1] 技术已经广泛使用在多媒

收稿日期: 2020-01-20; **修回日期**: 2020-03-18
基金项目: 国家自然科学基金(61771091, 61871066); 国家高技术研究发展计划(863 计划)(2015AA016306); 辽宁省自然科学基金(20170540159); 中央高校基本科研专项资金(DUT17LAB04)

(C)1994-2020 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

体通信和实时传输中。VoIP 采用用户数据报协议(User Datagram Protocol, UDP)快速地、一次性地传输语音数据,由于 UDP 不能保证数据包能有序并全部到达接收端,当网络临时拥塞出现时,就会发生数据包无法实时到达接收端(即发生丢包)的情况。当丢包率超过 10%,语音通信质量就会明显下降。

目前,VoIP 的抗丢包技术按照应用位置分主要有两类: 解码端、信道。现有音频编码器内置的抗丢包技术大多是在解码端加入一定的丢包处理措施,例如 ITU G. 711 附录一[2]采用带有基音周期检测的波形替代法,通过在丢包帧插入时间上最近的基因周期信号来替代丢包帧的信号,并通过重叠相加法(OverLap Add, OLA)来保证帧间的平滑过渡。ITU G. 723. 1[3]、G. 729[4]、3GPP AMR[5]采用传输状态插值法,通过对丢包帧两端的线性预测系数进行插值,使用原先帧的周期激励,最终重构出丢包帧的音频信号。解码端抗丢包处理的补偿能力有限,仅对丢包率 5% 以下的情况有效。相对于解码端被动的处理,为进一步提高抗丢包能力,可在编码端使用信道编码技术增加冗余来主动应对丢包状况。一种方法是将相同的包发送两遍,这种方法缺点是浪费带宽,并且在同时接收到两个包时造成信息的冗余,可通过解码端反馈信道状态信息给编码端来决定是否需要重复发送相同的包,但反馈回来的包也可能丢失,并且造成较大延时; 另一种方法[6]是在编码端编码两个包,但编码副本包时使用低码率的编码技术,这种方法减小一定带宽浪费,但存在的问题在于丢包时的音频质量会下降。

为了更好地解决上述丢包问题,人们提出了分布式语音编码。分布式语音编码又称多描述语音编码,它通常对同一帧语音信号采用两种或多种编码,每种编码结果独立传输,接收端,在接收到任一码流时,可独立解码; 在接收到两种或多种码流时,进行联合解码获得更好的语音质量[7]。

G. 722. 1[8]是一种常用的语音宽带编码器,可对带宽为 7 kHz 的语音或音乐进行低复杂度的编解码,其速率为 32 kbit/s 或 24 kbit/s,多应用于视频会议系统中,但其抗丢包措施很弱,导致音频通信质量下降。

针对上述问题,本文提出了一种基于 G. 722. 1 的分布式语音编码方法。该方法在 G. 722. 1 编码器(记为编码器 1)的基础上,构建一个互补编码器(记为编码器 2)。在编码端,对同一帧语音分别用 G. 722. 1 编码器和其互补编码器进行编码,并发送编码结果; 在解码端,在接收到其中任一语音码流时,用 G. 722. 1 解码器进行解码,其语音质量不低于 G. 722. 1 编码器的解码结果,而在接收到两个语音码流时,用 G. 722. 1 解码器分别对两个语音码流进行解码,然后对解码结果进行联合处理,其最终的语音质量有明显提升,即有一定编码增益。该方法抗丢包能力强,语音质量良好。仿真实验验证了本文方法的有效性。

2　G. 722. 1 编码器简介

G. 722. 1 编码器的有效编码带宽为 7 kHz,输入信号采样率 16 kHz,帧长为 20 ms(320 个采样点),连续两帧(640 个采样点)进行一次重叠调制变换(Modulated Lapped Transform, MLT),每次变换产生 320 个 MLT 系数的帧。设当前要进行 MLT 的 640 个语音样本为 $x(n)$,MLT 系数为

$$\mathrm{mlt}(m)=\frac{1}{\sqrt{160}}\sum_{n=0}^{639}\sin\left[\frac{\pi(2n+1)}{1280}\right]\times\cos\left(\frac{\pi}{320}nm+\frac{\pi}{640}n-\frac{319\pi}{640}m-\frac{319\pi}{1280}\right)x(n)\quad(1)$$

其中 $m=0,1,\ldots,319$ 为 MLT 系数的索引。

MLT 是一个严格抽样、无损的线性变换过程,它可以分解为一次加窗、重叠和加法运算,然后进行 320 点的 IV 型离散余弦变换(Discrete Cosine Transform, DCT)[7]。所以,式(1)得到的 320 个 MLT 系数对应频率范围 0 到 8 kHz。因编码器支持的有效带宽为 7 kHz,故后 1/8 的 MLT 系数(共 40 个)不参与编码,只编码前 7/8 的 MLT 系数(共 280 个)。将这 280 个 MLT 系数等分为 14 个频带,即每个频带包括 20 个 MLT 系数,频带宽度为 500 Hz。

编码按频带进行。为提高编码效率,分别对 MLT 频带均值和 MLT 系数进行编码。因为改进的分布式编码器主要涉及编码频带均值的部分,下面对该部分进行详细说明,对编码 MLT 系数部分简要阐述,最后简要介绍 G. 722. 1 的解码部分。

MLT 频带均值编码包括频带均值的计算、量化和编码。待编码的 280 个 MLT 系数分成 20 个一组

(C)1994-2020 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

的 R 个频带，即 $R=14$，设 r 为频带索引，$0\leqslant r<R$，则频带 r 包括 $20r$ 至 $20r+19$ 个 MLT 系数。MLT 系数的均方根(Root Mean Square，RMS)值 $\mathrm{rms}(r)$ 定义为

$$\mathrm{rms}(r)=\sqrt{\frac{1}{20}\sum_{n=0}^{19}\mathrm{mlt}^2(20r+n)} \tag{2}$$

其中 n 表示频带 r 中第 n 个 MLT 系数，$0\leqslant n<20$。然后，将 $\mathrm{rms}(r)$ 量化。量化器输出指数 $\mathrm{rms\_index}(r)$ 为

$$\mathrm{rms\_index}(r)=\mathrm{floor}(2\log_2(\mathrm{rms}(r))-1.5) \tag{3}$$

其中 $\mathrm{floor}(\cdot)$ 表示向下取整。量化指数的集合为

$$\begin{cases}-8\leqslant\mathrm{rms\_index}(r)\leqslant 31, & 0<r<R\\ 0\leqslant\mathrm{rms\_index}(r)\leqslant 31, & r=0\end{cases} \tag{4}$$

例如，若 $\mathrm{rms}(r)=310$，则 $\mathrm{rms\_index}(r)=15$。

在量化完成后，对指数 $\mathrm{rms\_index}(r)$ 进行编码。$\mathrm{rms\_index}(0)$ 使用 5 个比特直接编码，$\mathrm{rms\_index}(0)=0$ 值保留，不使用。其余 13 个频带的指数逐个与前一个频带做差，使用霍夫曼编码[3]算法对差值进行编码，以便传输。

下面简要介绍编码 MLT 系数部分。G.722.1 的编码速率固定，所以一帧语音编码出的比特数目也需固定。用总比特数目减去编码频带均值已使用的比特数目和一些固定占用的比特数目可以得到编码 MLT 系数的可用比特数目。在编码 MLT 系数时，首先经过一个分类模块，它的输入是量化后的频带均值 $\mathrm{rms\_index}(r)$ 和当前帧的可用比特数目，输出是 16 组分类。每组分类对应一组对 R 个频带的类别赋值，共有 8 种类别，每种类别规定了一组用于一个频带预定的量化和编码参数。每个频带使用对应的类别参数进行 MLT 系数的编码。MLT 系数的符号和幅度分开编码，符号的编码紧跟幅度编码后，使用 0 或 1 表示。编码幅度时，首先用 $\mathrm{rms}(r)$ 的量化值将 MLT 系数幅度归一化，然后将归一化后的标量组合成矢量，并使用霍夫曼编码进行矢量量化。对分类模块产生的 16 组分类都进行编码，选择编码后比特数目与当前帧剩余可用比特最接近的一组分类用于传输。整个编码比特流的组成如图 1 所示，它分为频带均值比特、分类控制比特和 MLT 系数比特三部分，由于霍夫曼编码的编码长度与输入信号有关，其中第一和第三部分的比特数目可变，但要保证总的编码速率是定值。

第一个比特 最后一个比特

频带均值比特(可变)	4比特分类控制	MLT系数比特(可变)

图 1 G.722.1 编码比特流组成

Fig. 1 G.722.1 bitstream structure

G.722.1 的解码是编码的逆过程。对每一帧，首先解码出 $\mathrm{rms\_index}(0)$(解码前 5 个比特)。然后，对其余频带进行霍夫曼解码和重建。接着，对 4 比特的分类控制进行解码，确定编码器所使用的分类(16 组分类中的一组)。帧内其余比特表示量化的 MLT 系数，按每一频带的类别信息进行解码即可。为获得与编码器相同的类别赋值，解码器中的分类模块与编码器完全一致。解码器若获知某一帧码流受损，则重复前一帧解码出的 MLT 系数；若前一帧码流也已受损，则解码器把所有当前帧的 MLT 系数置为零。最后，对解码出的 280 个 MLT 系数补上 40 个零，并进行 320 点的重叠调制反变换(Inverse Modulated Lapped Transform，IMLT)，以产生 320 个语音时域样值。IMLT 过程是先进行 IV 型 DCT 变换，然后进行加窗、重叠处理。

IV 型 DCT 为

$$u(n)=\sum_{m=0}^{319}\sqrt{\frac{1}{160}}\cos\left(\frac{\pi}{320}mn+\frac{m+n}{2}+\frac{1}{4}\right)\times\mathrm{mlt}(m),\quad 0\leqslant m,n<320 \tag{5}$$

其中，$u(n)$ 表示当前帧 IV 型 DCT，$\mathrm{mlt}(m)$ 表示解码出的 280 个 MLT 系数补上 40 个零的结果。

加窗、重叠处理的具体方法为

$$\begin{cases}y(n)=w(n)\,u(159-n)+\\ \quad w(319-n)\,\mathrm{u\_old}(n), & 0\leqslant n\leqslant 159\\ y(n+160)=w(160+n)\,u(n)-\\ \quad w(159-n)\,\mathrm{u\_old}(159-n), & 0\leqslant n\leqslant 159\end{cases} \tag{6}$$

其中 u_old 表示前一帧 DCT 输出的一半，窗函数

$$w(n)=\sin\left(\frac{\pi}{640}n+\frac{\pi}{1280}\right),\quad 0\leqslant n<320 \tag{7}$$

$u(n)$ 中未使用的后一半存储为 u_old，供下一帧使用，即

$$\mathrm{u\_old}(n)=u(n+160),\quad 0\leqslant n<160 \tag{8}$$

3 基于 G.722.1 的分布式编码器

基于 G.722.1 的分布式编码器如图 2 所示，它

(C)1994-2020 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

包括位于发送端的编码器 1、编码器 2 模块和位于接收端的解码器模块。在发送端,对原始信号使用两个编码器进行编码,编码器 1 为 G. 722. 1 的 ITU-T 提供的原始编码器,编码器 2 为新构建的编码器,称为互补编码器。将两个编码器编码后的语音码流数据分别打包为 packet1 和 packet2,并通过网络上的不同路由进行传送。在接收端,根据实际网络拥塞导致的丢包情况,分情况处理,这仅在 G. 722. 1 的原始解码器基础上进行少量修改即可。

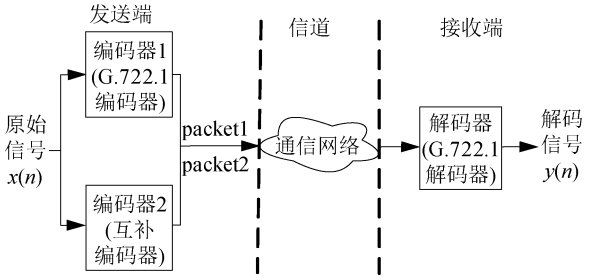

图 2　分布式语音编码器

Fig. 2　Distributed speech encoder

图 2 中编码器 2 的内部框图如图 3 所示,它分为编码参数调节、部分解码、感知误差计算三个模块。执行流程为: 首先,原始信号在初始参数下进行编码,然后将编码后的信号进行部分解码; 将当前部分解码结果与编码器 1 的部分解码结果共同输入到感知误差计算模块进行感知误差的计算; 将计算得到的误差反馈给编码器用于参数调节。如此操作,进行多次反馈,得到最优参数; 最后,使用这组最优参数作为编码器 2 的编码参数进行编码,得到编码器 2 的最终输出。下面对编码器 2 的各模块进行说明,并阐述不同情况下的解码处理方式。

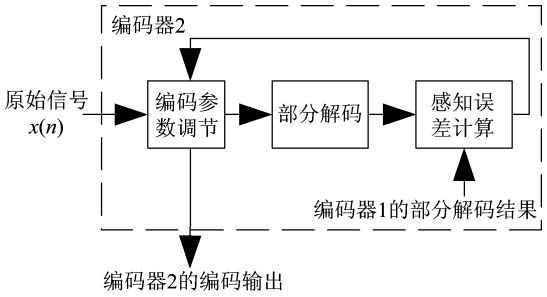

图 3　编码器 2 的内部框图

Fig. 3　Internal block diagram of encoder 2

为保证编码器 2 的编码结果可用 G. 722. 1 原始解码器进行解码,编码器 2 应在原始编码器 1 的基础上,进行参数调节。由式(1)、(5)、(6)、(7)可知,MLT 变换是无损变换,Huffman 编解码是无损压缩,若没有量化误差且信道没有误码,则可以将前 280 个 MLT 系数无损地传输到解码端,可以完成 0 ~7 kHz 频带信号的无损恢复。对于 G. 722. 1 编码器的任何改进,最终都会反映在使解码出的 280 个 MLT 系数更准确。由第 2 节的编码器原理可知,对 MLT 系数的频带均值进行量化时,量化重建值集合如式(3)、(4)所示,其量化间隔大,量化精度低,因此编码器 2 需要对编码器 1 的量化结果进行调整,使编码器 1 和编码器 2 联合解码的 MLT 系数更准确,且单独解码的编码器 2 的 MLT 系数的误差小于编码器 1。本文用反馈结构确定最优的调整方法来保证量化调整的有效性。

编码参数调节模块的具体调整方法如下: 原始编码器的频带均值量化结果为 $\mathrm{rms\_index}(r)$,$0 \leqslant r < R$。设集合 $C = \{\mathrm{rms\_index}(r) - 1, \mathrm{rms\_index}(r) + 1\}$,用集合 C 中任一元素替代 $\mathrm{rms\_index}(r)$ 来完成调整操作。编码器共有 R 个频带,每一频带 r 有两种可能取值。一次调整为对 R 个频带,将每个频带 r 初步量化的 $\mathrm{rms\_index}(r)$ 替换为集合 C 中的任一元素。对全部 R 个频带进行调整,共有约为 $2^R - 1$ 种调整方法。对于的任一频带 r,有 $-8 \leqslant \mathrm{rms\_index}(r) \leqslant 31$,$r \neq 0$,因此对于 $\mathrm{rms\_index}(r) = -8$(对于频带 0,$\mathrm{rms\_index}(0) = 0$),不选取集合 C 的第一个元素,对于 $\mathrm{rms\_index}(r) = 31$,不选取集合 C 的第二个元素。设初始调整方法为 P_0,P_0 为 $2^R - 1$ 种调整方法中的随机一种,最优调整方法为 P_m。

原始信号 $x(n)$ 在经过编码参数调节模块后,进入部分解码模块。部分解码模块完成 G. 722. 1 解码器的部分操作,具体操作与第 2 节相同,MLT 为无损变换,只需解码出 MLT 系数即可用于后续误差的计算。需要指出的是,G. 722. 1 编解码器除了 IMLT 时需要上一帧的部分信息之外,帧与帧之间没有编解码状态的耦合,这是反馈结构的实现基础。

图 3 中编码器 2 部分解码模块的输出与编码器 1 的部分解码结果一起输入到感知误差计算模块进行误差的计算。信噪比(Signal to Noise Ratio, SNR)、总谐波失真(Total Harmonic Distortion, THD)等指标只反映了信号本身数值上的差异,不能反映主观感受。研究表明[10-11],人耳对不同频率的信号有不同的敏

(C)1994-2020 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

感度,对不同频率段所能忍耐的误差也不同。此外,人耳的主观感受与信号幅值的对数成比例,对信号的相位不敏感。为此,本文采用感知误差来作为编码参数调节模块的反馈量。

下面阐述感知误差的计算方法。设当前部分解码模块输出的 MLT 系数为 enc2_mlt(n),编码器 1 的部分解码输出的 MLT 系数为 enc1_mlt(n),当前联合解码的结果记为 com_mlt(n),其计算公式为

$$\text{com\_mlt}(n)=\frac{1}{2}\text{enc1\_mlt}(n)+\frac{1}{2}\text{enc2\_mlt}(n),\quad 0\leqslant n<280 \tag{9}$$

相对误差定义为

$$\tilde{e}(n)=\frac{|\text{mlt}(n)-\text{com\_mlt}(n)|^2}{\text{mlt}^2(n)} \tag{10}$$

感知误差为

$$e_{\text{com}}=\sum_{n=0}^{279}w(n)\log_{10}\tilde{e}(n) \tag{11}$$

其中 $0\leqslant n<280$,mlt(n) 为编码信号经过 MLT 变换的理想值,$w(n)$ 为感知加权权重,它是频率的函数。将式(10)中的 com_mlt(n) 替换为 enc1_mlt(n) 和 enc2_mlt(n),对应的误差记为 e_{enc1} 和 e_{enc2}。

用等响度曲线可计算出 $w(n)$ 为

$$w(n)=\frac{1}{E[25n]+10},\quad n=0,1,\ldots,280 \tag{12}$$

其中 E 是等响度曲线[12]中听阈曲线上对应 $25n$ Hz 频率的 dB 值,如表 1 所示。

于是,在约 2^R-1 次的调整方法解空间中,寻找最优调整方法 P_m,整个编码流程转化为如下优化问题

$$\begin{cases}\min\ e_{\text{com}}\\ \text{s.t.}\ \ e_{\text{enc2}}\leqslant e_{\text{enc1}}\end{cases} \tag{13}$$

具体求解过程可通过在解空间中优化搜索实现。

表 1 等响度表

Tab. 1 Table of equal loudness

$E[\cdot]$/dB	对应频率范围/kHz
45.0	0.024~0.047
30.0	0.048~0.070
25.0	0.071~0.093
20.0	0.094~0.117
15.0	0.118~0.141
13.0	0.142~0.164
12.0	0.165~0.188
11.5	0.189~0.211
10.0	0.212~0.234
10.0	0.235~0.258
10.0	0.259~0.281
9.0	0.282~0.305
9.0	0.306~0.328
8.0	0.329~0.352
8.0	0.353~0.375
8.0	0.399~0.422
8.0	0.423~0.445
7.0	0.446~0.469
6.0	0.470~0.492
6.0	0.493~0.516
5.0	0.517~0.539
4.0	0.540~0.563
4.0	0.564~0.586
3.0	0.587~0.609
2.0	0.610~0.633
2.0	0.634~0.656
2.0	0.657~0.680
1.0	0.681~0.984
1.5	0.985~1.500
0.5	1.501~2.016
-3.0	2.017~3.023
-6.0	3.024~4.031
-2.0	4.032~5.039
2.0	5.040~6.047
6.0	6.048~7.055

分布式编码完成后,解码端需要根据实际丢包情况进行解码。对每一帧数据,编码端编码的两个包记为 packet1 和 packet2,解码时,根据数据包的丢失情况,存在 4 种情形:(1) packet1、packet2 都不丢;

(C)1994-2020 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

(2) packet1不丢而 packet2 丢; (3) packet1 丢而 packet2 不丢; (4) packet1、packet2 都丢。设改进的分布式编解码最终输出的 280 个 MLT 系数记为 newmlt(n),对于上述每种情况的处理方法如表 2 所示。

表 2　MLT 系数的替换公式

Tab. 2　Substitution formula of MLT coefficients

接收状态	解码端 MLT 系数替换公式
(1)	$\text{newmlt}(n)=\frac{1}{2}\text{mlt1}(n)+\frac{1}{2}\text{mlt2}(n)$
(2)	$\text{newmlt}(n)=\text{mlt1}(n)$
(3)	$\text{newmlt}(n)=\text{mlt2}(n)$
(4)	$\text{newmlt}(n)=\text{newmlt\_his}(n)$

无论丢包情况如何,两组码流都按已有的规则正常解码。对于 00,这种情况利用与编码端合成 com_mlt(n) 一样的加权原则,合成一组新的 MLT 系数; 对于 01,使用未丢包的编码器 1 对应的 MLT 系数 mlt1(n); 对于 10,使用未丢包的编码器 2 对应的 MLT 系数 mlt2(n); 对于 11,使用与原始解码器相同的规则,丢弃一帧使用前一帧的 MLT 系数代替,连续丢两帧或以上将 MLT 系数置零。这样就可组合出一组新的 MLT 系数,以用于下一步的重叠调制反变换(IMLT),得到重构的信号。

4　仿真实验与结果讨论

为验证本文方法的有效性,对分布式编码器解码出的语音质量进行主观评价和客观评价,并与原始编码器、相同包发送两遍的抗丢包方法进行比较(下文称对比方法)。

在进行语音主观评价实验时,使用三种方法对同一段语音进行编码,然后选取 12 位试听者对解码出的语音进行试听评分,语音质量主观评价采用 MOS 评分法,评分标准如表 3 所示。语音客观评价采用感知语音质量评估 PESQ (Perceptual Evaluation of Speech Quality)[13]、短时客观可懂度 STOI (Short-Time Objective Intelligibility)[14] 和信号失真比 SDR (Source to Distortion Ratio)[15] 三个指标。在进行客观评价时,对于宽带 PESQ 的测量,用 12 个人的 48 段 8 ~ 10 s 的语音作为输入,使用三种方法进行编解码,对输出结果进行宽带 PESQ 测试取均值。对于 STOI 和 SDR 的测量,用 8 段 8 ~ 10 s 的语音作为输入信号,以编码前的语音作为参考,对三种方法的解码结果计算 STOI 和 SDR 值,并对结果取平均。所有主观和客观实验均分别在 32 kbps 和 24 kbps 两种码率下进行,丢包率分别设定为 0%、1%、3%、5%、10%、20%、30%,丢包方式为随机丢包。实验中,ORG 表示原始方法; OTR 表示对比方法; OUR 表示本文提出的方法。

表 3　MOS 语音评分标准

Tab. 3　Standard of Mean Opinion Score

得分	质量级别	失真级别
5	优	不察觉
4	良	刚有察觉
3	中	有察觉且稍有厌恶
2	差	明显察觉且可厌但是可以接受
1	劣	不可忍受

主观实验结果如表 4 所示。从表 4 可以看出,在不同码率和丢包率下,本文方法在三种方法中 MOS 评分最高,即语音的质量最好。与原始方法相比,本文方法明显改善语音质量; 与对比方法相比,在丢包率≤5% 时,本文方法语音质量略有提升,在丢包率为 10%、20% 和 30% 时,本文方法语音质量的改善较大。在码率为 24 kbps 和 32 kbps 下,呈现类似的实验结果。

客观实验结果如表 5、表 6 和表 7 所示。从表 5 可以看出,与原始方法相比,本文方法 PESQ 值明显提高,在不丢包时,改善值>0. 1; 在丢包率为 10% 和 20% 时,改善值>1。与对比方法相比,本文方法在丢包率≤5% 时,改善值>0. 1,随着丢包率的提高,改善值下降,在丢包率为 30% 时,与对比方法接近一致。三种方法的 STOI 值如表 6 所示,与原始方法相比,本文方法的 STOI 值在不同的码率和丢包率下提升明显。与对比方法相比,在丢包率≤10% 时,本文方法 STOI 值更高; 在丢包率为 20% 和 30% 时,本文方法 STOI 值未必优于对比方法,两种方法彼此接近,STOI 值都保持在 0. 950 以上。表 7 给出三种方法 SDR 值的测试结果。与原始方法相比,本文方法的 SDR 值在不同的码率和丢包率下提升明显。与对比方法相比,在丢包率≤5% 时,本文方法 SDR 值更高; 在丢包率为 10%、20% 和 30% 时,本文方法

(C)1994-2020 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

SDR 值未必优于对比方法，两种方法彼此接近。在实际网络环境中，绝大部分情况下丢包率≤10%，因此本文方法较对比方法有更好的实际表现。

表4 三种方法的 MOS 评分结果

Tab. 4 MOS score results for three methods

丢包率/%	24 kbps			32 kbps		
	ORG	OTR	OUR	ORG	OTR	OUR
0	4.78	4.78	4.83	4.82	4.82	4.83
1	4.35	4.72	4.79	4.57	4.75	4.83
3	4.18	4.62	4.70	4.18	4.62	4.67
5	3.72	4.38	4.47	3.77	4.22	4.30
10	2.91	3.70	3.82	3.32	3.78	3.99
20	2.43	3.02	3.26	2.73	3.20	3.44
30	2.01	2.80	3.01	2.05	2.52	2.65

表5 PESQ 测试结果

Tab. 5 PESQ test results

丢包率/%	24 kbps			32 kbps		
	ORG	OTR	OUR	ORG	OTR	OUR
0	3.607	3.607	3.789	3.679	3.679	3.853
1	3.393	3.608	3.784	3.432	3.68	3.868
3	3.017	3.598	3.766	3.076	3.65	3.822
5	2.774	3.563	3.698	2.795	3.627	3.754
10	2.348	3.387	3.471	2.365	3.425	3.572
20	1.753	2.892	2.942	1.777	2.898	2.964
30	1.461	2.382	2.393	1.464	2.414	2.404

表6 STOI 测试结果

Tab. 6 STOI test results

丢包率/%	24 kbps			32 kbps		
	ORG	OTR	OUR	ORG	OTR	OUR
0	0.989	0.989	0.991	0.992	0.992	0.994
1	0.987	0.989	0.991	0.988	0.991	0.993
3	0.979	0.989	0.991	0.988	0.992	0.994
5	0.971	0.989	0.990	0.974	0.990	0.992
10	0.953	0.986	0.987	0.956	0.988	0.989
20	0.898	0.977	0.973	0.914	0.980	0.976
30	0.864	0.955	0.958	0.848	0.963	0.959

表7 SDR 测试结果

Tab. 7 SDR test results

丢包率/%	24 kbps			32 kbps		
	ORG	OTR	OUR	ORG	OTR	OUR
0	19.599	19.603	20.826	19.952	19.952	21.177
1	16.206	19.402	20.786	16.722	19.952	21.107
3	11.294	19.004	19.839	12.610	19.767	20.259
5	10.464	19.061	19.616	9.266	18.274	19.737
10	6.448	16.022	15.131	6.961	15.051	15.449
20	4.068	10.967	9.919	3.725	10.966	10.101
30	2.328	7.586	7.648	1.719	8.885	7.500

最后，分析本文方法的延时与计算复杂度。实际上，本文基于 G.722.1 的分布式语音编码器没有引入算法上的额外延时，但增加了编码器的计算复杂度。本文方法计算复杂度的增加取决于解式(13)的优化方法的选取及解空间的大小。由于语音能量主要集中在 300 Hz ~ 3400 Hz，因此仅用频带数 $R=7$ 对语音频带进行调整，以减少计算复杂度。在本文实验中，采用线性搜索，当 $R=7$ 时，所需运算次数小于 710 WMOPS，可实现实时编解码，此时，其语音 PESQ 值相比于 $R=14$ 时仅下降小于 2%，但计算量却明显减少。

5 结论

本文基于 ITU G.722.1 编码器，提出一种分布式的抗丢包语音编码方法。该方法在原有编码器基础上，构建一个互补编码器，它通过反馈的方式保证语音编码的感知质量；编码端对一帧语音编码出两个包，每个包都可单独解码，解码端在接收到任何一个包时，解码的语音质量不低于 G.722.1 编码器，在同时接收到两个包时语音质量明显提升。主观和客观的实验结果表明，相对于原始编码器，本文的分布式编码器在不丢包时，一定程度上改善原有编码器的语音质量；在丢包时，仍能保持较好的语音质量。本文工作对于改进网络通信中的语音传输质量具有一定参考意义。

参考文献

[1] Martin O, Gustavo C A, Ciro L B, et al. Comparison be-

(C)1994-2020 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

tween the real and theoretical values of the technical parameters of the VoIP codecs [C] // IEEE Conference on Communications and Computing. Barranquilla, Colombia, Colombia: IEEE, 2019: 1-6.

[2] ITU-T Recommendation G. 711, Appendix I: A high quality low-complexity algorithm for packet loss concealment with G. 711 [S]. 1999, 09.

[3] ITU-T Recommendation G. 723. 1, Dual rate speech coder for multimedia communications transmitting at 5. 3 and 6. 3 kbit/s [S]. 2006, 05.

[4] ITU-T Recommendation G. 729, Coding of speech at 8 kbit/s using conjugate structrue algebraic-code-excited linear prediction(CS-ACELP) [S]. 2012, 06.

[5] 3GPP (3rd Generation Partnership Project) TS 26. 090, AMR speech codec; Transcoding functions [S]. 1999, 12.

[6] Tarek G, Fatiha M. An enhanced interleaving frame loss concealment method for voice over IP network services [C] // European Signal Processing Conference (EUSIPCO). Rome, Italy: IEEE, 2018: 1302-1306.

[7] Hocine C, Fatiha M, Philippe M. Multiple description coding technique to improve the robustness of ACELP based coders AMR-WB [J]. Speech Communication, 2019, 108: 33-40.

[8] ITU-T Recommendation G. 722. 1, Low-complexity coding at 24 and 32 kbit/s for hands-free operations in systems with low frame loss [S]. 2005, 05.

[9] **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.
Bao Changchun. Digital speech coding principle [M]. Xi'an: Xidian University Press, 2007. (in Chinese)

[10] **梁瑞宇, 赵力, 王青云. 语音信号处理 [M]. 北京: 机械工业出版社,** 2018.
Liang Ruiyu, Zhao Li, Wang Qingyun. Speech signal processing [M]. Beijing: Machinery Industry Press, 2018. (in Chinese)

[11] Colm S, Naomi H, Damien K, et al. Objective assessment of perceptual audio quality using ViSQOLAudio [J]. IEEE Transactions on Broadcasting, 2017, 63(4): 693-705.

[12] **呼德, 陈喆, 殷福亮. 一种自动等响度数字混音算法 [J]. 信号处理,** 2017, 33(3): 437-443.
Hu De, Chen Zhe, Yin Fuliang. A digital audio mixing algorithm with equal-loudness [J]. Journal of Signal Processing, 2017, 33(3): 437-443. (in Chinese)

[13] ITU-T Recommendation P. 862, Perceptual evaluation of speech quality(PESQ): An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs [S]. 2005, 11.

[14] Cees H T, Richard C H, Richard H, et al. An algorithm for intelligibility prediction of time-frequency weighted noisy speech [J]. IEEE Transactions on Audio, Speech, and Language Processing, 2011, 19(7): 2125-2136.

[15] Emmanuel V, Remi G, Cedric F. Performance measurement in blind audio source separation [J]. IEEE Transactions on Audio, Speech, and Language Processing, 2006, 14(4): 1462-1469.

作者简介

何莹男　男, 1995 **年生, 辽宁阜新人。大连理工大学信息与通信工程专业研究生, 主要研究方向为语音信号处理。**
E-mail: 824655412@ qq. com

陈　喆　男, 1975 **年生, 大连理工大学教授, 博士生导师,** IEEE **高级会员, 研究方向为语音处理、阵列信号处理和宽带无线通信术。**
E-mail: zhechen@ dlut. edu. cn

殷福亮　男, 1962 **年生, 大连理工大学教授, 博士生导师, 研究方向为语音处理、图像处理、阵列信号处理和宽带无线通信术。**
E-mail: flyin@ dlut. edu. cn

(C)1994-2020 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net