

文章编号:1001-893X(2010)09-0065-04

改进的移动音频带宽扩展算法^{*}

张 聪

(武汉工业学院 计算机与信息工程系,武汉 430023)

摘 要:由于 3G 移动信道环境严酷,带宽资源有限,为了实现低码率、高质量的音频信号传输和重建,在研究现有语音/音频带宽扩展技术的基础上,提出了一种基于 EAAC+ 编码标准的改进移动音频编码算法。该算法通过检测并去除短暂态信号带来的正弦分量,从而降低计算的复杂度,实现传输信道带宽的扩展,并保证重建音频信号的音质。

关键词:3G 通信;移动信道;音频编码;带宽扩展

中图分类号:TN919.1 **文献标识码:**A doi:10.3969/j.issn.1001-893x.2010.09.015

Improved Mobile Audio Bandwidth Extension Algorithm

ZHANG Cong

(Department of Computer and Information Engineering, Wuhan Polytechnic University, Wuhan 430023, China)

Abstract: The environment of 3G mobile channel is complex and the wireless transmission frequency band resources are limited. In order to achieve the transmission and reconstruction of low bit rate and high-quality audio signal, a modified EAAC+ mobile audio coding algorithm based on the existing technology of speech or audio bandwidth extension research is presented in this paper. The algorithm can reduce the computational complexity, achieve the bandwidth extension of transmission channel, and ensure the tone quality of reconstructed audio signal by detecting and removing the sinusoidal component of short transient signal.

Key words: 3G communication; mobile channel; audio coding; bandwidth extension

1 引 言

3G 网络在国内正式运行以及各种新的多媒体移动接入技术的相继涌现,为移动多媒体应用带来了广阔的空间和新的机遇。但是,由于移动通信环境的复杂性、无线传输频带资源的有限性以及传输信道的多变性,使得传统音频编解码技术遇到了新的困难,因此,面向 3G 网络的移动音频编解码技术就成为了移动多媒体应用领域一个新的重点研究方向。为了适应移动环境苛刻的应用需求,本文力争

在提高编码效率、降低运算量两方面克服理论缺陷,突破技术瓶颈,深入研究移动音频编解码中的一些核心关键技术,尤其是带宽扩展技术。

2 现有带宽扩展技术

2.1 语音带宽扩展技术

语音带宽扩展技术的飞速发展是在 20 世纪 80 年代末、90 年代初,为了提高话音效果,人们开始研究能否利用 300 Hz ~ 3.4 kHz 窄带语音信号合成 3.4 ~ 8 kHz 高频成分以重建 8 kHz 的宽带信号。主要方

^{*} 收稿日期:2010-05-06;修回日期:2010-06-18
基金项目:国家自然科学基金重点项目(60832002);湖北省教育厅科研重点项目(D20101704)
Foundation Item: The Key Project of National Natural Science Foundation of China(No.60832002); The Key Project of Hubei Provincial Department of Education Science Foundation(No.D20101704)

法是基于语音生成的源滤波器模型算法,利用窄带语音信号估计高频带激励信号,然后将其通过一个模拟宽带信号频谱包络的合成滤波器滤波得到重建的高频语音信号。其关键是如何产生高频带激励信号以及模拟宽带信号频谱包络的合成滤波器^[1]。

AMR-WB+ 标准是 3GPP 组织推荐的移动语音压缩编码算法,其关键算法是对来自同一激励源的低频信号分量与高频信号分量分别采用 ACELP/TCX 混合编码模式和带宽扩展编码方式。实现原理是从语音的低频信号分量中提取时域空间的激励源特征,再将该激励源特征与高频信号分量在时域空间合成,得到重建的高频信号。AMR-WB+ 标准的缺点是对高频信号所使用的带宽扩展算法中编码比特数保持固定,不能根据信号的类型与特征进行自适应调整,且整个计算复杂度较高^[2]。

2.2 音频带宽扩展技术

目前,感知音频编码算法的带宽扩展技术大都采用 BWE 算法,其原理大致是:在解码端利用信号的低频成分以及预估参数(有些技术未使用)通过相关算法重构信号的高频成分。此类扩展技术可以分为盲目式和非盲目式两类:著名的盲目式带宽扩展技术有对数频域线性外推法及八度音截取等方法,非盲目式带宽扩展技术有 PlusV 法及 BBSM 法等。

20 世纪 90 年代末出现的频谱带宽复制 (Spectral Bandwidth Replication, SBR) 技术^[3],通过在频域上精确重建解码后的信号高频成分实现带宽扩展,既增加了音频信号和语音信号的带宽,也改善了低比特率下信号的编码质量,已经被 3GPP 标准化组织提出的 EAAC+ 标准所采纳,是一种在移动环境下同时具有较高压缩效率和较高音质的成功的带宽扩展方案,其原理如图 1 所示^[4]。

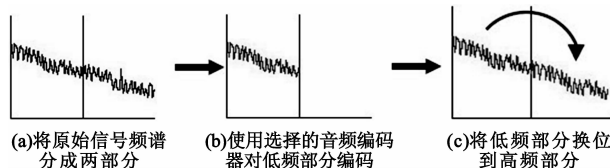


图 1 SBR 带宽扩展方式
Fig.1 SBR bandwidth extension mode

3 改进的音频带宽扩展技术

本文提出的音频带宽扩展技术是结合现有非盲目式带宽扩展技术 BBSM(Bark Band Spectral Model)

法,对 EAAC+ 标准中的 SBR 技术进行改进。

3.1 BBSM 法^[5]

本算法的基本思想是利用音频信号的低频成分与高频成分之间的相关性作为带宽扩展的依据。相关性的衡量通过下述公式计算相关系数来判定:

$$corr = \frac{\sum_{i=0}^{N-1} x[i] \times y[i]}{\sqrt{\left(\sum_{i=0}^{N-1} x^2[i]\right) \times \left(\sum_{i=0}^{N-1} y^2[i]\right)}} \quad (1)$$

式中, $x[i]$ 是低频频带信号, $y[i]$ 是高频频带信号, N 是计算相关性的样值序列的区间长度。

其实施步骤如下:在编码时,通过相关性检测模块分析高频信号与低频信号的关联性。假如相关性较低,则高频成分被看成是白噪声,解码时就不用低频信号,只使用白噪声以达到重建高频成分。如果高频信号与低频信号的相关性较高,则在编码时先将信号的低频成分频谱映射到 Bark 带频谱,并提取每个 Bark 带的能量,同时,信号的高频成分频谱与低频成分频谱相对应,也提取 Bark 带的能量,然后计算低频成分与对应高频成分的 Bark 频谱频带的能量比系数,再将该能量比系数作为参数传到解码器端。在解码时,根据编码器端传送过来的能量比系数,将低频成分的 Bark 频谱频带的能量复制到高频成分所对应的 Bark 频谱频带,并进行能量调整,实现信号高频成分的重建工作。

3.2 EAAC+ 频谱带宽复制算法^[6]

EAAC+ 标准中 SBR 算法利用语音信号和音频信号高频成分与低频成分之间的关联性,用信号的低频成分来复制高频成分。其具体步骤如下:首先,重采样后的输入信号送入 64 子带 QMF 滤波器组,由其对应输入信号子带进行滤波。输入信号经过 QMF 滤波后输出 32 个 QMF 子带,每个子带具有 64 个频率通道,它们被存储在矩阵 X 中,矩阵 X 定义为

$$X(k, l + qmfWriteOffset), 0 \leq k < 64; \\ 0 \leq l < numTimeSlots \cdot RATE \quad (2)$$

然后,选择合适的时间解析度,实现包络分割。通过瞬态检测和帧结构的判断就可以完成包络分割。瞬态检测器从矩阵 X 的子带样点 0 开始执行;接着,帧分离器检测当前帧的时隙包络,从矩阵 X 的子带样点 0 开始执行;音调检测器从矩阵 X 的子带样点 $qmfWriteOffset$ 开始执行,其得到的控制数据和当前时频网格被作为附加控制参数单元。在这个模块

中,自适应噪声级、逆滤波估计和附加正弦估计被确定;最后,形成的包络数据被量化和哈夫曼编码,并打包为一个 SBR 帧送到比特流复用器中。

可以看出,SBR 方法存在较多弊病:低频信号子带波形与高频信号子带波形差距较大,仅仅简单的复制并不能替换高频信号,反而会导致复制过程中信号严重失真;由于复制时只是简单地按照子带序号依次提取进行复制,会出现重建的高频子带正弦峰值与原始信号的高频子带正弦峰值差异较大;从复制的过程来看,实际上在高频区低频子带被连续复制了若干次,这样无疑大大增加了算法的计算复杂度。

3.3 改进方案

针对上述问题,本文结合 BBSM 音频带宽扩展算法,提出了一种基于 EAAC + 的改进带宽扩展算法,其实施步骤如下:

(1)按照 BBSM 算法对信号进行变换

将输入的原始音频信号分成低频和高频两部分,对于低频部分,仍然按照 SBR 算法执行;对于高频部分,按照 BBSM 算法,将整个高频频带再分为高频区的低频子带和高频子带,减少后续传输的码率。

(2)检测并去除正弦分量

SBR 算法中提取的参数较多,包括包络估计、时频分辨率的重建、分析滤波器系数、正弦分量、频谱平滑因子等,其特点是重建信号的效果较好,但是缺点很明显,就是运算复杂度较高。实际上,高频信号中除了谐波以外,还包含较多的正弦信号,但是正弦信号对重建的高频信号音质影响较小,编码却不小。因此,为了降低算法的编码复杂度,本文在改进算法中将其检测并去除。具体做法如下:求出输入信号频谱能量的大小,然后找到高频子带中能量最大的位置,并且计算这个子带与左右其它子带能量的平均值,公式如下:

$$average = \frac{sum(localEnergies(j+i))}{2 \times (RANGE - 1)} \tag{3}$$

式中, $i = -RANGE, -(RANGE - 1), \dots, -3, -2, +2, +3, \dots, (RANGE - 1)$, $RANGE$ 是所取子带的范围, j 是子带中能量最大者的位置。

根据式(3)判断正弦分量并去除。首先经过大量试验将正弦阈值设定为 15,然后判断子带 j 的能量值是否大于 j 左右其它子带能量的平均值乘上正弦阈值,如果大于,则存在正弦分量;否则就不存在。找到正弦分量 j 的位置后将其序号记录下来,同时将 j 附

近位置子带信号的频谱幅度也记录下来,同时经过大量试验将正弦分量能量值定为小于 0.25,只有这些条件全部满足,才确定为正弦分量;否则就不是。

(3)确定音调噪声比

为了完成高频重建,需要确定使用低频信号来复制还是使用白噪声来重建。设定条件是:

$$high_sin_num > low_sin_num + TTN_Ratio \tag{4}$$

$$high_sin_num = 0 \tag{4}$$

式中, $high_sin_num$ 和 low_sin_num 分别表示在高频频带和低频频带中找到的正弦分量数量; TTN_Ratio 是判断两个相邻频带之间音调噪声比是否一致,一般取常数。如果信号中没有正弦分量或者信号高频成分中正弦分量数量为零,在解码器端重建高频信号时只使用白噪声;否则,将低频信号复制到高频频带。

(4)能量计算

计算高频中每一子带与低频子带的能量比作为解码时调整高频参数的依据。按照下式进行计算:

$$band_energy = \sqrt{\frac{\sum_{i=0}^k y_high_no_sin(i)^2}{length(y_high_no_sin)}} \tag{5}$$

$$band_energy_ratio = \frac{\sum_{i=0}^k y_low_no_sin(i)^2}{\sum_{i=0}^k y_high_no_sin(i)^2} \tag{6}$$

如果只有白噪声,不需要计算式(5),将其设为 1,只计算原始高频频带的平均能量值。在解码时,通过白噪声与式(5)相乘调整参数大小。如果不是白噪声,则在低频信号复制到高频频带时,以式(6)的值与式(5)相乘来调整参数大小。

在解码器端,将每个高频子带中记录的正弦分量位置、大小、白噪声的设定值以及子带能量比等参数做为重建高频频带的依据,再结合经过核心编解码器的低频频带,合成全频带的信号频谱,恢复出完整的语音或音频信号。

4 性能评价

为了验证提出的改进的 EAAC + 带宽扩展算法,本文通过一系列的主客观测试来全面评价算法的编码质量。在主客观听力实验中选取了 MPEG 音频标准中 12 个标准音频测试,包含了不同类型的、具有代表性的谐波和非谐波语音信号和音频信号。

4.1 客观测试

为了进一步测试新算法的性能,本算法与 AMR

-WB+ 的带宽扩展算法进行了实验比较。为了保证实验的一致性,选择 ITU-T P.862^[7] 建议中的语音编码客观软件评测工具——PESQ 对两者的算法性能进行了比较(PESQ 值为客观评测得到的 MOS 分)。由于 PESQ 主要用于评测语音信号,因而测试选取的序列为 MPEG 标准和 AMR-WB+ 标准推荐的标准测试序列中的语音信号。由于 PESQ 软件只适合对 16 kHz 语音信号进行客观评分,原始测试信号的采样率为 48 kHz。选择序列后,我们使用在专业音频软件 Audition 3.0 将信号的采样率全部下采样到 16 kHz 后再用 PESQ 软件进行评测。根据测试结果来看,文中设计的带宽扩展算法的 PESQ 值与 AMR-WB+ 标准中带宽扩展算法的 PESQ 值近似,说明两者的性能基本相当。

4.2 主观测试

根据国内 AVS 组织的音频小组提出的主观听力测试规范^[8],每次主观听力测试的材料由 Ref/A/B 组成,其中 Ref 是原始未编码的信号,A/B 均是编解码后的信号。在测试中,如果 A 是本文文所实现的解码结果,那么 B 就是参考解码器的解码结果,反之亦然。在测试中,对比的参考编码器为 AMR-WB+ 编解码器。

如图 2 所示,根据 12 个测试序列总的平均值来看,本文提出的改进的带宽扩展算法与 AMR-WB+ 带宽扩展算法的性能基本相当,主观评测结果与客观 PESQ 评测结果基本相一致。

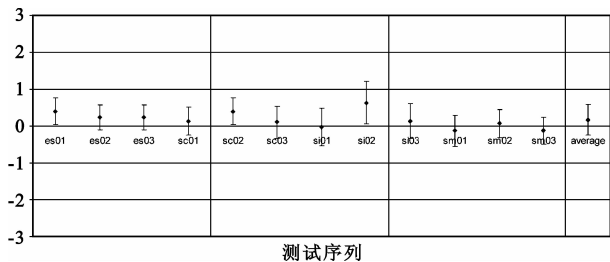


图 2 主观听力测试结果
Fig.2 Subjective listening test results

5 结束语

本文针对 EAAC+ 编码标准的带宽扩展算法提出的改进方案,根据测试结果判断,达到了与 AMR

-WB+ 编码标准基本上相当的音质。对于研究中出现的一些特殊情况,例如:有些原始信号中的低频成分和 高频成分具有峰值,而这些峰值在信号重建的过程中会丢失,因此,如何更精确地重建原始信号是后续研究值得关注的方向。

参考文献:

[1] Julien Epps. Wideband Extension of Narrowband Speech for Enhancement and Coding[D]. New South Wales, Australia: The University of New South Wales, 2000.

[2] 3GPP Specification Series TS26.290, Audio codec processing functions, Extended Adaptive Multi-Rate Wideband AMR-WB+ codec, Transcoding functions[S].

[3] Martin Diets, Lars Liljeryd, Kristofer Kjörling, et al. Spectral Band Replication, a novel approach in audio coding[C]//Proceedings of the 112th AES Convention. Munich, Germany: AES, 2002: 1-8.

[4] Arttu Laaksonen. Bandwidth extension in high-quality audio coding[D]. Helsinki: Helsinki University of Technology, 2005.

[5] Seo J, Jang D, Kang K. A Simple Method for Reproducing High Frequency Components at Low-Bit Rate Audio Coding[C]//Proceedings of the 113th AES Convention. LOS Angeles, 2002: 1-8.

[6] 3GPP TS 26.404, Enhanced aacPlus general audio codec; Enhanced aacPlus encoder SBR part[S].

[7] ITU-T 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].

[8] 阎建新, 金尚煜. AVS 音频内部主观听力测试参考规范[M]. [S.l.]: AVS 音频工作组, 2005.

YAN Jian-xin, JIN Shang-yu. AVS Audio Inner Subjective Listening Test Reference Criterion[M]. [S.l.]: AVS Audio Workgroup, 2005. (in Chinese)

作者简介:

张 聪(1968-),男,上海人,教授、硕士生导师,1999 年获工学硕士学位,主要研究方向为多媒体信息处理和网络通信。

ZHANG Cong was born in Shanghai, in 1968. He received the M.S. degree in 1999. He is now a professor and also the instructor of graduate students. His research interests include multimedia information processing and network communication.

Email: hb_wh_zc@126.com