一种基于子帧联合编码的 600b/s 低速语音编码算法 [1]

陈 亮 张雄伟

(解放军理工大学通信工程学院电子信息工程系 南京 210007)

摘 要 为了适应无线通信等甚低速语音通信应用，文中提出一种基于子帧联合编码的 600b/s 语音编码算法。该算法的激励源采用混合激励模型，声道参数使用帧内预测多级矢量量化进行高效量化，在参数编解码时提出了子帧分类联合的思想，并在编码端使用语音增强抑制背景噪声，解码端使用后滤波处理来改善语音质量，这些方面较传统 LPC 算法有了明显改进。同时，选用 TI 公司的 TMS320VC5416 DSP 芯片实时实现了该算法。非正式主观试听结果表明，该算法在可懂度、清晰度等方面与传统的 2.4kb/sLPC 算法相当，而速率仅为 LPC 算法的 1/4，是甚低速率的一种良好的编码方案。

关键词 语音编码，子帧联合编码，混合激励，线性预测

中图号 TN912.3, TN911.3

1 引 言

由于数字化语音在可靠性、抗干扰和保密等方面优于模拟语音，因此已在有线、无线和卫星通信等领域发挥了重要作用。近年来，随着计算机网络的飞速发展，中低速语音编码以其低速率和良好的编码质量在多媒体通信中越来越受到重视。90 年代以来，研究热点主要集中在 2.4kb/s 以上的中低速率，2.4kb/s 以下低速率并未有成熟的标准或算法推出。但是仍有不少场合需要用到低于 2.4kb/s 的语音编码技术，例如无线短波信道恶劣，可靠通信的速率很低，要在短波信道上实现数字语音通信就必须进一步降低语音编码速率。因此，研究 1.2kb/s 以下甚低速率语音编码算法在理论和实际应用上都具有重要意义。

本文在吸收混合激励线性预测 (MELP) 算法 [1] 特点的基础上，综合线性预测编码 (LPC)[2] 和多带激励 (MBE)[3] 算法的优点，建立一种新的子帧联合编码的模型结构。引入分带 LPC 模型，并灵活运用混合激励、非周期脉冲、自适应谱增强和脉冲散布滤波等技术，在 DSP 芯片 TMS320VC5416 上实时实现了 600b/s 语音编码算法，非正规试听表明合成语音的可懂度和清晰度与传统的 2.4kb/s LPC 算法相当。

本文的安排如下：第 2 节介绍 600b/s 语音编码算法的基本特点；第 3 、4 节分别介绍了该算法的编码、解码方案；第 5 节比较了本算法与 2.4kb/s LPC 算法的性能；最后为本文的总结。

2 600b/s 编码算法基本特点

本算法虽然速率很低，但仍然采用了一套完整的方法确保对激励源的准确提取。和 LPC 算法区别在于：算法将语音不仅仅划分为清音和浊音，还增加了第 3 类语音——抖动浊音，且各类语音的激励源模型不同。表 1 对比了 LPC 算法与本算法的语音分类以及相应的激励模型。第 3 类语音解决了 LPC 对过渡音和较弱浊音不能正确分类的难题，改善了语音合成的清晰度和自然度。

除此之外，算法还使用分带混合激励、非周期脉冲、残差谐波谱、自适应谱增强和脉冲散布滤波等关键技术改善激励源。

[1] 2001-09-18 收到，2002-05-09 改回
国家高等学校骨干教师资助计划资助

表 1 两种算法的语音分类与激励信号模型

LPC 算法		600b/s 算法	
语音分类	激励信号	语音分类	激励信号
清音	白噪声	清音	白噪声
浊音	周期脉冲	抖动浊音	白噪声和非周期脉冲的混合
		浊音	白噪声和周期脉冲的混合

2.1 分带混合激励

分带的思想来源于 MBE 算法，分带可以从频域上对激励信号进行更加精细的刻划，合成激励更准确。本算法将 [0,4kHz] 的语音频带分成 5 个固定的频段 ([0,500Hz] [500Hz,1kHz] [1kHz,2kHz] [2kHz,3kHz] [3kHz,4kHz]) 处理。分带滤波器由 5 个带通滤波器相加得到，频率响应见图 1 。由于低频对语音的影响更大，同时为了便于基音提取，算法对低频段的划分更精细。

混合激励的合成质量取决于分带滤波器的频响和各子带混合比例的计算。理想的激励信号应具有平坦的功率谱，同时在各个频带上脉冲和噪声加权相加后功率和应保持常数。因此，当各子带的加权值全为 1 时，整个激励应是一个无畸变的脉冲。考虑 FIR 滤波器具有良好的线性相位，使用汉明窗设计 32 阶线性相位 FIR 滤波器作为合成滤波器。子带清浊音混合比例由输入语音在各子带内的脉冲成分和噪声成分的相对功率决定。

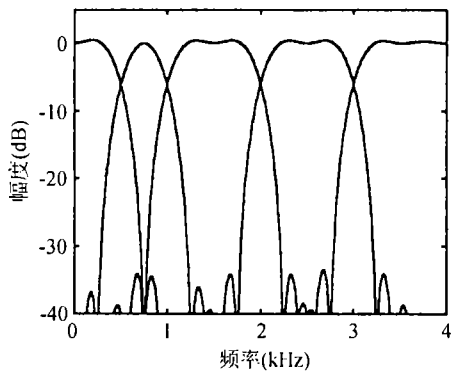

图 1 分带合成滤波器的频响

2.2 使用非周期脉冲

当女性说话的基音较高或存在噪声时，合成语音可能出现类似电流的噪声。通过减弱周期性可以去除这些噪声。本算法使用非周期脉冲来减弱周期性。由于解码端按基音周期来合成语音，若本帧为抖动浊音帧，则通过使基音周期 T 在 $[0.75T, 1.25T]$ 之间随机变动来减弱周期性。这种办法可以很好地模拟不稳定的声门脉冲。

2.3 残差谐波谱的处理

LPC 残差信号含有大量的语音信息，传统 LPC 算法在生成激励脉冲时只反映了它的周期性，并没有注意其幅频特征。本文借鉴原型波形内插 (PWI) 算法 [4] 的思想，对残差信号最低 10 阶谐波进行矢量量化， 10 阶以上谐波幅度则认为是平坦谱。在解码端，残差谱按基音周期进行离散傅里叶反变换 (IDFT) 就可以得到周期脉冲激励序列。与固定脉冲序列相比，该方法提供了更多的灵活性，在很大程度上提高了合成语音的自然度、清晰度和抗背景噪声能力，大大改善 LPC 合成语音嘶哑和合成音重等弱点。

2.4 自适应谱增强技术

由于 LPC 合成滤波器的极点形状与自然语音的共振峰存在偏差，导致共振峰之间合成语音谱的波谷不如原始语音谱的波谷尖锐，从而合成语音听起来发闷。为了使合成语音与原始语

音在共振区有更好的匹配，引入了自适应谱增强技术。自适应谱增强通过让激励信号经自适应谱增强滤波来实现。自适应谱增强滤波器由阶数为线性预测阶数的零极点滤波器与 1 阶零点滤波器级联而成，其目的是突出激励谱中共振峰频率处的谱幅度，相应提高整个短时谱在共振峰处的信噪比。

2.5 脉冲散布滤波

在语音合成以后加入脉冲散布滤波，使分带合成的语音与原始语音在非共振区有更好的匹配。滤波器系数是通过将典型男性周期脉冲谱强制变为平坦谱，再进行傅里叶反变换得到的。脉冲散布滤波减弱了某些频带的周期性，降低了基音周期为典型周期时的峰峰值，合成语音的蜂鸣效果降低，更为连贯自然。脉冲散布滤波器的频响曲线如图 2 所示。

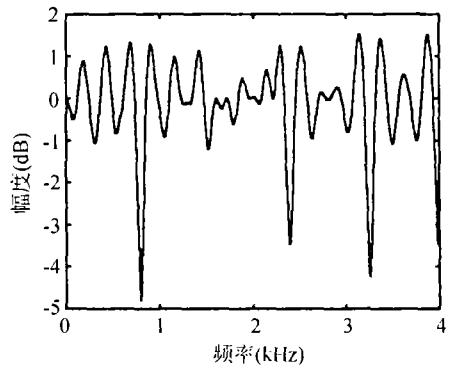

图 2　脉冲散布滤波器的频响曲线

3 600b/s 编码算法

本算法与传统的语音编码算法的区别主要在于以下几个方面：通过语音增强抑制背景噪声，高效 LSP 量化以及基于发声听觉特性的子帧联合编解码。以下对这几个方面逐一描述。

3.1 语音增强

在背景噪声近似为宽带平稳加性高斯白噪声时，以启发式方法为基础，采用平滑谱相减技术进行自适应噪声抑制 [5]，该方法比传统谱相减法有更好的抑制背景噪声效果，且残留的音乐噪声进一步降低。首先对长度为 40ms 的输入信号进行 FFT 变换，得到该段包含的谱信息。然后用语音检测器判断该段为噪声还是带噪语音，并通过检测出的噪声谱信息和语音帧相关性来不断修正噪声模型。如果检测出语音，将带噪语音谱与估计的噪声谱相减，经 IFFT 得到增强语音。增强语音作为输入信号进入编码器开始编码。

3.2 LSP 量化

在基于线性预测的语音编码算法中，线性预测系数的量化精度对语音合成质量具有举足轻重的影响。本文首先将 LPC 系数转化为线谱对 (LSP) 系数，然后采用帧内预测多级矢量量化技术 [6] 对 LSP 进行矢量量化。使用帧内预测量化器后，计算及量化精度都有了一定提高，使用 18bit 量化性能与 MELP 无帧内预测 25bit 量化性能基本相当。衡量线性预测系数量化精度使用谱畸变指标，为了达到“透明”传输，对 LPC 参数的量化要求满足以下 3 个要求：

(1) 平均谱畸变要小于 1dB 。

(2) 谱畸变大于 2dB 的帧的比例小于 2% 。

(3) 没有谱畸变大于 4dB 的帧。

选取一段典型测试音 (男声和女声“他去无锡市，我到黑龙江”)，测试计算 LPC 参数及量化引入的噪声。将 MELP 算法中无帧内预测的 LSP 量化方法和本文帧内预测 LSP 量化方法进

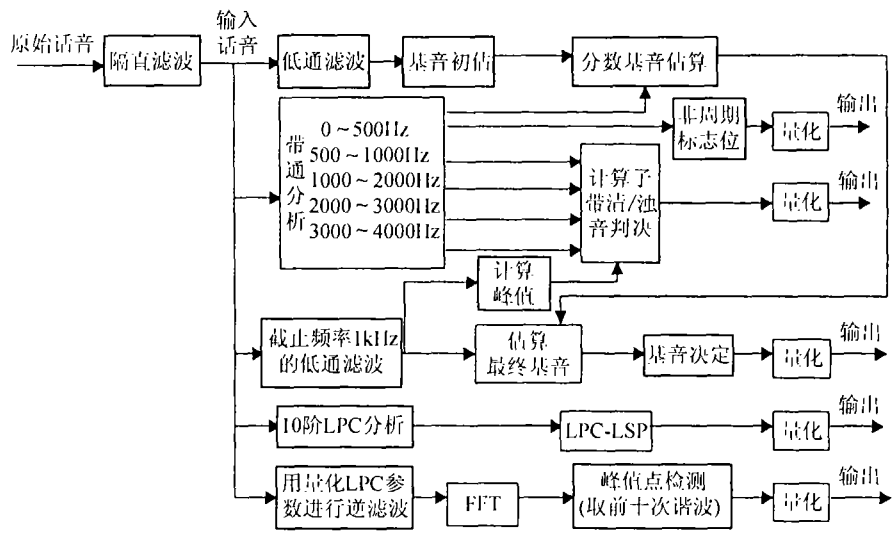

图 3 子帧编码方框图

行比较，结果见表 2，其中谱畸变用 N 表示，单位分贝。从表中可以看出，采用 18bit LSP 量化可以很好地代替 25bit 无帧内预测量化方案。

表 2 两种量化方法的性能比较

量化算法		谱畸变比例 (%)				平均谱畸变 (dB)
		N <1dB	1dB< N <2dB	2dB< N <4dB	N >4dB	
MELP	19bit	86.9	11.3	1.6	0.2	0.45
(无帧内预测)	25bit	96.2	3.3	0.5	0	0.23
本文算法	18bit	95.9	4.1	0	0	0.25
(帧内预测)	26bit	97.0	3.0	0	0	0.18

3.3 子帧联合编解码器

3.3.1 *汉语发声特性* 一般语音都是由清音和浊音连接形成的。浊音短时频谱相对稳定，而清音频谱随时间变化剧烈。通过对大量语音信号的分析，发现语音能量变化最大的地方在于清音和浊音的过渡部分，且清音大都持续时间短，能量弱，但对合成语音的清晰度和可懂度有着直接的影响，编解码时需尽量保持清音的合成效果。

根据汉语的发声特性，本文提出基于子帧联合的编解码方案。其中编码器每 40ms 处理一个子帧，每 4 个子帧编码结束后，根据前后子帧出现的语音信号类型将子帧参数进行有效封装。

3.3.2 *子帧编码器结构* 图 3 给出了每个子帧编码的方框图。

原始语音经隔直流滤波后作以下处理：(1) 低通滤波后用归一化互相关法粗估基音，并根据 [0,500Hz] 子带信号估算分数基音；(2) 带通分析。计算 5 个子带语音强度，确定各子带的清浊音判决，其中 [0,500Hz] 子带强度用于设定非周期标志位；(3)1kHz 截止频率的低通滤波，残差信号结合上一子帧的基音和当前子帧的分数基音确定最终的基音周期；(4)LPC 分析，并将 LPC 系数转换为 LSP 参数量化输出；(5) 量化后的 LSP 参数转换为 LPC 参数进行逆滤波，残差信号补 0 至 512 点进行 FFT 变换，根据频谱峰点检测找出前 10 次谐波对应的傅里叶系数量化输出。这样，各子帧量化后的参数如表 3 所示。

表 3　子帧参数比特分配

参数	浊音帧	清音帧
LSP 参数	18	18
增益 (Gain)	5	5
全局清浊音判决 (U/V)	1	1
基音周期 (Pitch)	5	–
非周期标志位 (Jitter)	1	–
残差谐波谱 (Mag)	8	–
分带清 / 浊音判决 (Bpvc)	4	–
子帧比特数	42	24

3.3.3 编码器结构　帧编码器的作用是将 4 个子帧的参数联合编码。由于每帧只能传输 96bit，因此应尽量保护对模型敏感的参数 (如基音周期等)，随时间缓变的参数在解码端由分类内插规则合成。因清音对语音清晰度和可懂度有直接影响，帧编码器要保持清音参数，必要时舍弃次要比特，使得合成语音获得较高的质量。

(1) 帧结构　打包参数的帧结构分为 3 个部分：同步头、帧模式和帧参数。同步头占用帧结构第 1 比特，该位相邻帧 0, 1 交替以获得连续同步。帧模式占用 4bit，它反映当前各子帧的清浊音模式，并指示后面的参数结构，不同的帧模式对应着不同的帧参数。帧结构如图 4 所示。

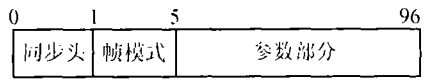

图 4　帧结构示意图

(2) 帧模式　设上一帧的最后一个子帧为 0 子帧，当前帧的 4 个子帧分别为 1、2、3、4 子帧，则帧模式由 0~4 子帧的全局清浊音判决共同决定。例如，若 0~4 子帧的全局清浊音判决分别为 0, 0, 1, 1, 1 (0 表示清音子帧，1 表示为浊音子帧)，则帧模式为 00111。其中 0 子帧的全局清浊音判决可从上一帧的帧模式获得，故在帧封装时只需用 4bit 来表示帧模式即可，如上例的帧模式最终表示为 0111。帧模式共 2^5 种状态，但其中一些状态很难出现，如浊音的持续时间通常在 80ms 以上，帧模式 00100 几乎不可能出现。为了减少封装复杂度，去除帧模式中很少出现的状态，将 32 种状态减少为 16 种，并按属性划分为 3 类：

(a) 2, 4 子帧均为浊音帧：　00111, 01111, 11111;

(b) 2, 4 子帧均为清音帧：　00000, 10000, 11000;

(c) 2, 4 子帧一清一浊：　00001, 00011, 00110, 01100, 01110, 10001, 10011, 11001, 11100, 11110。如果帧模式出现其他情况，则该帧结构按与上面 3 类中最接近的模式进行封装。

(3) 帧参数　帧参数的选择与帧模式紧密联系。根据表 3，帧模式反映了 4 个子帧的全局清浊音判决，因此对于每个子帧，完全描述清音帧参数需 23bit，完全描述浊音帧参数需 41bit。因为帧参数不可能将子帧所有参数封装，所以按照汉语发声特性保留清音和过渡帧 (全局清浊音 0, 1 交替时) 参数，连续浊音帧对模型不敏感的参数不必封装，在解码端由前后子帧参数内插得到。

对 (a) 类，封装 2, 4 子帧的所有参数共 2×41=82bit，余 9bit 由 1, 3 子帧参数选择重要部分封装。

对 (b) 类，封装 1, 2, 4 子帧的 LSP 和增益共 3×23=69bit，余 22bit 选择剩下的重要参数封装。

对 (c) 类，封装 2, 4 子帧的所有参数共 23+41=64bit，余 27bit 由 1, 3 子帧参数选择重要参数封装。

4 600b/s 解码算法

4.1 子帧解码

解码端接收到一帧参数并检测到同步头后，按照与帧模式相对应的编码规则进行解码。解码端保持了上帧的最后一子帧 (即本帧的 0 子帧) 参数用于本帧参数的解码与内插，保证语音信号的渐变性。

解码时先根据帧模式解出对应 4 个子帧的相应参数，对未加以传输的参数，根据人的发声听觉特性和前后子帧的清浊音关系，由一个自适应插值因子线性内插恢复。例如对于帧模式 00111，拆封时解码出 2, 4 子帧的全部参数、1 子帧的增益和 3 子帧的分带清浊音判决，尚有 1 子帧的 LSP 参数和 3 子帧的 LSP, Gain, Pitch, Jitter 和 Mag 参数未知。根据人的发声听觉特性，1 子帧 (清音子帧) 的 LSP 参数由 0, 2 子帧的 LSP 参数线性内插得到，由于 0 子帧和 2 子帧分别为清音和浊音，故内插因子 $\gamma = 0.25$，即

$$\mathrm{LSP}_1 = (1-\gamma)\mathrm{LSP}_0 + \gamma \mathrm{LSP}_2 \tag{1}$$

这样处理使得 1 子帧的 LSP 参数保持了较好的清音特性。同样，3 子帧的 LSP, Gain, Pitch 和 Mag 参数按照同样的方法内插。由于后 3 个子帧皆为浊音，故此时内插因子 $\gamma = 0.5$。同时，3 子帧的 Jitter 参数与 2 子帧的 Jitter 参数保持一致。

4.2 子帧语音合成

解码端按子帧合成语音，在激励信号的合成方式及后处理上与传统 LPC 算法有较大区别，合成流程图如图 5 所示。

图中合成滤波器的作用相当于依据清浊音判决对子带激励信号在频域加权求和。合成激励源时，脉冲合成滤波器系数为浊音频带的带通滤波器系数之和，噪声合成滤波器系数为清音频带的带通滤波器系数之和，滤波器系数每个基音周期更新一次。脉冲和噪声源各自滤波后相加得到混合激励。随后混合激励信号先经过自适应谱增强滤波器改善共振峰的形状，然后使用直接形式的 LPC 滤波器进行语音合成，其滤波器系数由插值后的 LSP 参数得到，最终经增益调整和脉冲散布滤波输出合成语音。

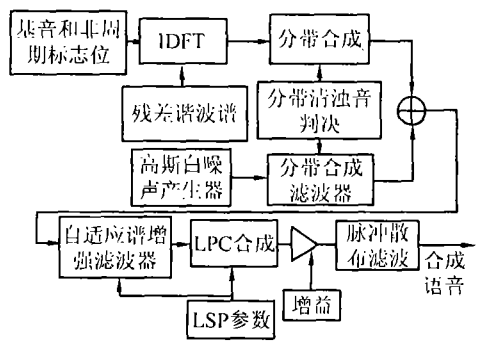

图 5 子帧语音合成框图

5 性能比较

实验中选取一段典型测试音 (男声和女声 “他去无锡市，我到黑龙江”)，分别采用 2.4kb/s LPC 算法和本文 600b/s 算法编码并合成。图 6 给出了一段语音 (清浊过渡音) 的原始波形及分别采用两种算法合成的时域波形和频谱。从图中可以发现，本算法对于 LPC 难以处理的过渡音 (如 LPC 将图中语音判为清音) 仍然有较好的合成效果，波形无论在时域上还是在频域上都更接近于原始语音。

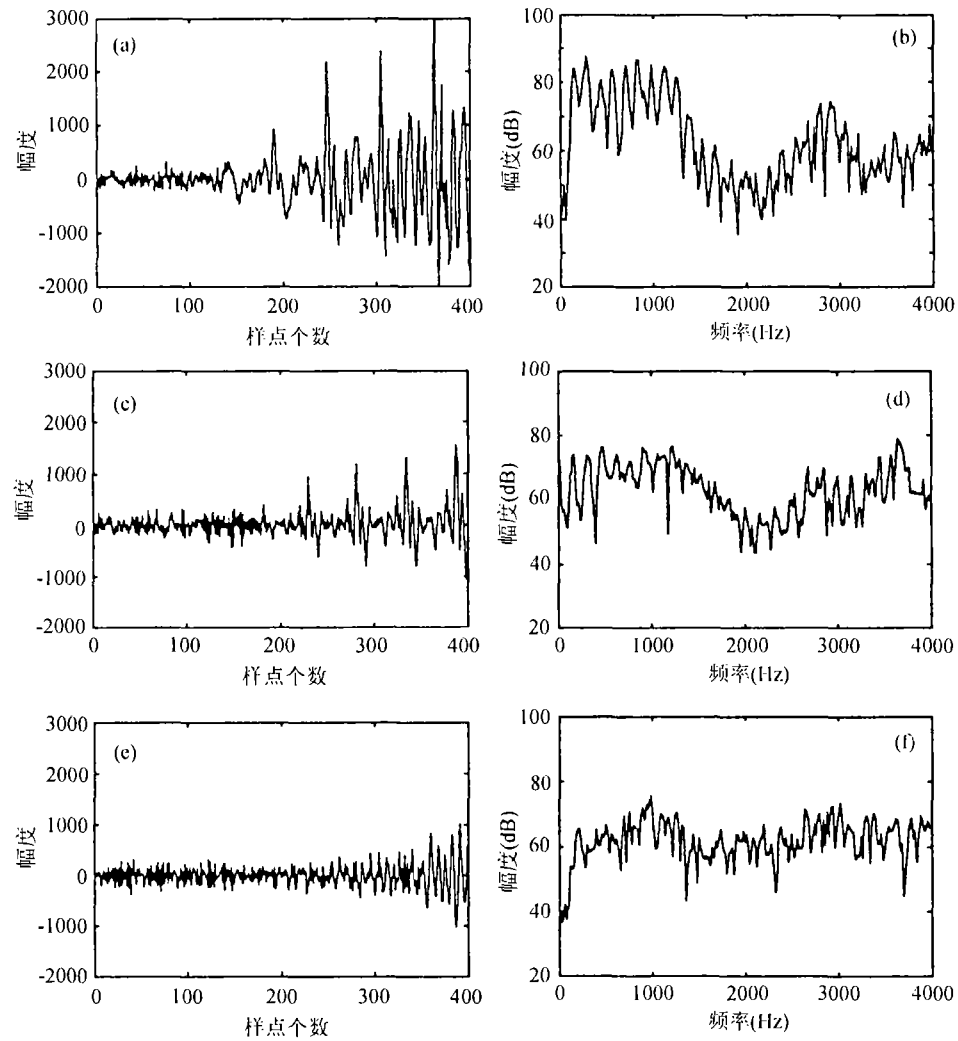

图 6　原始语音、600b/s 算法合成语音和 2.4kb/s LPC 合成语音的时域波形及频谱比较

(a) 原始语音时域波形　(b) 原始语音频谱

(c) 600b/s 算法合成语音时域波形　(d) 600b/s 算法合成语音频谱

(e) 2.4kb/s LPC 合成语音时域波形　(f) 2.4kb/s LPC 合成语音频谱

6 结 束 语

本文提出了一种基于子帧联合编码的 600b/s 甚低速语音编码算法，在模型结构、LSP 量化及编解码等方面较传统 LPC 算法有了较大改进，且速率只有 LPC 算法的 1/4 。在硬件设计时，选用 TI 公司的 TMS320VC5416 DSP 芯片实时实现了该算法声码器。非正式主观测试结果表明，合成语音可懂度和清晰度与 LPC 算法基本相当，而自然度高于 LPC 算法，保持了较高的质量。该声码器可实用于无线通信等低速率语音编码场合，是甚低速率下一种良好的语音编码方案。

参 考 文 献

[1] L. M. Supplee, A. V. McCree, MELP: the new federal standard at 2400 bit/s, Proc ICASSP'97, Munnich, Germany, 1997, 1591–1594.

[2] T. Tremain, The government standard linear predictive coding algorithm: LPC-10, Speech Technology Magazine, April, 1982, 40–49.

[3] D. W. Griffin, J. S. Lim, Multi-band excitation vocoder, IEEE Trans. on ASSP, 1988, 36(8), 1223–1235.

[4] W. B. Kleijn, Encoding speech using prototype waveforms, IEEE Trans. on Speech and Audio Processing, 1993, 1(4), 386–399.

[5] L. Arslan, A. McCree, V. Viswanathan, New methods for adaptive noise suppression, in Proc. IEEE ICASSP, Detroit, Michigan, May 1995, 1, 812–815.

[6] 陈 亮，陈 敏，张雄伟，LSP 参数的快速计算及其高效量化研究，解放军理工大学学报，2(5), 24–27.

A 600b/s SPEECH CODING ALGORITHM BASED ON SUB-FRAME JOINT CODING

Chen Liang Zhang Xiongwei

(*Dept. of Electron. & Info. Eng., Institute of Comm. Eng., PLAUST, Nanjing* 210007, *China*)

Abstract In order to meet the applications of wireless communication, a new speech algorithm based on sub-frame joint coding is proposed in this paper. This new algorithm improves several aspects such as model structure, LSP quantization and coding. This algorithm has been implemented on TMS320VC5416 DSP processor. Informal listening testing results show that the intelligibility and naturalness performances of the 600b/s speech coding algorithm are close to those of the 2.4kb/s LPC algorithm, while the rate is only 1/4 of LPC algorithm.

Key words Speech coding, Sub-frame joint coding, Mixed excitation, Linear prediction

陈 亮： 男，1974 年生，博士生，讲师，主要研究领域为多媒体信号处理、语音编码、数字通信等.

张雄伟： 男，1965 年生，教授，博士生导师，电子信息工程系主任，中国通信学会理事，中国通信学会青年工作委员会委员，中国电子学会高级会员，江苏省电子信息专业委员会副主任委员. 主要研究领域为语音信号处理、网络信息处理、数字通信等.