

◇ 研究报告 ◇

## 定点脉冲线性预测编码方法研究\*

马震<sup>†</sup> 吴殿红

(滨州学院信息工程系 滨州 256600)

**摘要** 在多脉冲线性预测编码的基础上,本文提出了位置无关脉冲搜索算法。该算法不需要搜索脉冲位置,而是根据给定的脉冲位置一次性解出脉冲幅度矢量。这就保证了得到的脉冲组合在最小二乘意义下是最优的,为改进合成语音质量提供了理论基础。进而在激励脉冲与位置无关的理论基础上,提出了定点脉冲线性预测编码方法。对所提出的算法在 MATLAB 下进行了仿真,仿真结果发现位置无关脉冲搜索算法得到的合成语音质量优于序贯法,编码时间也要比序贯法短。定点脉冲线性预测编码方法可以在 2.7 kbps 的编码速率下获得与 G.729 相近的合成语音。

**关键词** 语音编码,多脉冲线性预测编码,定点脉冲线性预测编码

**中图分类号:** TN912.3 **文献标识码:** A **文章编号:** 1000-310X(2016)02-0137-07

**DOI:** 10.11684/j.issn.1000-310X.2016.02.007

## Research on fixed-location pulse linear prediction coding

MA Zhen WU Dianhong

(Department of Information Engineering, Binzhou University, Binzhou 256600, China)

**Abstract** In this paper, a location-independent pulse search algorithm based on multi-pulse linear prediction coding (MP-LPC) is presented. This algorithm can solve all the amplitudes of the pulses at a time according to given pulse locations which need not be solved using analysis-by-synthesis procedure. This can ensure that the solved pulses are optimal in a least square sense, which provides the theoretical foundation for improving the quality of synthesized speech. And then based on the location-independent pulse search algorithm, fixed-location pulse linear prediction coding (FLP-LPC) method is presented. The presented algorithm is simulated in MATLAB, the results show that the synthesized speech using the pulses solved by location-independent pulse search algorithm has better speech quality and shorter coding time than the sequential method. FLP-LPC can obtain the synthesized speech closed to G.729 at the coding rate of 2.7 kbps.

**Key words** Speech coding, Multi-pulse linear prediction coding, Fixed-location pulse linear prediction coding

2015-06-28 收稿; 2015-08-10 定稿

\*国家自然科学基金项目(30870666), 山东自然科学基金项目(ZR2014FL005), 山东教育厅科技发展项目(J08LJ52), 滨州学院科研基金项目(BZXYG1004, BZXYG1007)

作者简介: 马震(1980-), 男, 山东滨州人, 硕士研究生, 研究方向: 语音信号处理、生物医学信号处理。

<sup>†</sup>通讯作者 E-mail: 13954380080@139.com

## 1 引言

作为人类交流信息的最常用方式之一,语音的压缩编码一直受到关注<sup>[1]</sup>。移动通信系统中,随着用户的增加,大幅度增加的话务量导致每话路分配的带宽非常有限。另外,卫星通信系统中,信道非常昂贵,如何充分利用有限的带宽来传输更多的语音也是一个关键问题。保密通信、短波及超短波通信<sup>[2]</sup>也迫切需要降低编码速率。

合成语音的质量和编码速率是一对永恒的矛盾。参数编码只关注并提取声道模型<sup>[3]</sup>与激励的参数,其编码速率一般4 kbps以下,其中MELP算法可以在2.4 kbps的编码速率下获得清晰度和自然度可接受的合成语音。典型的参数编码有二元激励线性预测算法、混合激励线性预测算法(Mixed excitation linear prediction, MELP)<sup>[2]</sup>、波形插值算法(Waveform interpolation, WI)<sup>[4]</sup>、正弦变换编码算法(Sinusoidal transform coding, STC)<sup>[5]</sup>以及多带激励编码算法(Multi-band excitation, MBE)<sup>[6]</sup>等。混合编码可以在4~16 kbps编码速率范围内,保持良好的合成语音质量,一直被应用于不同领域。典型算法有多脉冲线性预测编码(Multi-pulse linear prediction coding, MP-LPC)<sup>[7]</sup>、码激励线性预测编码(Code excited linear prediction, CELP)<sup>[8]</sup>等。

MP-LPC能够在较低的编码速率上获得较好的语音质量,但其脉冲搜索算法复杂度较高。本文在传统的MP-LPC的基础上提出了位置无关脉冲搜索算法,这种算法不需要搜索脉冲的位置,脉冲的位置可以随机生成,通过对这些位置处的脉冲幅度进行优化而保证合成语音和原始语音在最小二乘意义下最接近。在此基础上,提出了对于不同语音帧都采用位置固定的脉冲的定点脉冲线性预测编码方法。位置无关脉冲搜索算法可以有效的降低编码时间,同时可以提高合成语音质量。而定点脉冲线性预测编码可以在提高语音质量的前提下,降低编码速率。

## 2 多脉冲线性预测编码算法

设帧长为 $N$ 的一帧内,采用 $M$ 个脉冲作为激励信号,激励信号可以表示为 $p(n) = \sum_{k=1}^M g_k * \delta(n - n_k)$ ,这里 $g_k$ 和 $n_k$ 分别表示第 $k$ 个

脉冲的幅度和位置。多脉冲编码的关键是确定 $g_k$ 和 $n_k$ ,使得合成语音和原始语音在感觉上差别最小。

合成语音可以表示为

$$\hat{s}(n) = \hat{s}_0(n) + \sum_{k=1}^M g_k * h(n - n_k), \quad (1)$$

其中, $\hat{s}_0(n)$ 为线性预测(LP)综合滤波器的零输入响应,是之前帧对当前帧的影响; $h(n)$ 为LP综合滤波器的单位冲激响应。

合成语音 $\hat{s}(n)$ 与 $s(n)$ 之间的误差为

$$\begin{aligned} e_s(n) &= s(n) - \hat{s}(n) \\ &= \bar{e}_n(n) - \sum_{k=1}^M g_k * h(n - n_k), \end{aligned} \quad (2)$$

其中, $\bar{e}_n(n) = s(n) - \hat{s}_0(n)$ 为原始语音减去之前帧的影响,也就是当前帧激励通过LP综合滤波器产生的等效语音。 $e_s(n)$ 通过感觉加权滤波器可得

$$\begin{aligned} e(n) &= [\bar{e}_n(n) - \sum_{k=1}^M g_k * h(n - n_k)] * w(n) \\ &= \bar{e}_w(n) - \sum_{k=1}^M g_k * h_w(n - n_k), \end{aligned} \quad (3)$$

其中, $h_w(n)$ 是加权综合滤波器的单位冲激响应。感觉加权均方误差 $E_m$ 为

$$\begin{aligned} E_m &= \sum_{n=1}^N e^2(n) \\ &= \sum_{n=1}^N [\bar{e}_w(n) - \sum_{k=1}^M g_k * h_w(n - n_k)]^2. \end{aligned} \quad (4)$$

多脉冲编码的中心思想就是通过确定 $g_k$ 和 $n_k$ 来使得 $E_m$ 最小。通过将 $E_m$ 对 $g_k$ 和 $n_k$ 分别求偏导数,并使之等于0,可以得到 $M$ 个线性方程和 $M$ 个非线性方程。如果要同时解这 $2M$ 个方程,过程是非常复杂的。所以多脉冲算法采用序贯搜索的方法,每次运算确定一个脉冲的位置和幅度,经过 $M$ 次可以确定 $M$ 个脉冲的位置和幅度。第 $j$ 个脉冲的 $n_j$ 为使下式取最大值所对应的位置:

$$R_{eh}^2(n_j)/R_{hh}(n_j, n_j), \quad (5)$$

之后再根据确定的 $n_j$ 来计算 $g_j$ :

$$g_j = R_{eh}(n_j)/R_{hh}(n_j, n_j), \quad (6)$$

其中,

$$R_{eh}(n_j) = \sum_{n=1}^N \bar{e}_w(n) * h_w(n - n_j), \quad (7)$$

$$R_{hh}(n_k, n_j) = \sum_{n=1}^N h_w(n - n_k) * h_w(n - n_j). \quad (8)$$

### 3 位置无关 (Location-independent) 脉冲搜索算法

从上面可知, 多脉冲算法的思想就是通过确定  $g_k$  和  $n_k, k = 1, \dots, M$ , 力求

$$\|e(n)\| = \left\| \bar{e}_w(n) - \sum_{k=1}^M g_k * h_w(n - n_k) \right\| = 0.$$

也就是存在  $g_k$  和  $n_k$  使得

$$\sum_{k=1}^M g_k * h_w(n - n_k) = \bar{e}_w(n). \quad (9)$$

以上  $M$  个等式写作矩阵的形式为

$$\mathbf{H}\boldsymbol{\beta} = \mathbf{E}, \quad (10)$$

其中,

$$\mathbf{H}(n_1, n_2, \dots, n_M) = \begin{bmatrix} h_w(1 - n_1) & \cdots & h_w(1 - n_M) \\ \vdots & & \vdots \\ h_w(N - n_1) & \cdots & h_w(N - n_M) \end{bmatrix}_{N \times M}, \quad (11)$$

$$\boldsymbol{\beta} = [g_1, g_2, \dots, g_M]^T, \quad (12)$$

$$\mathbf{E} = [\bar{e}_w(1), \bar{e}_w(2), \dots, \bar{e}_w(N)]^T. \quad (13)$$

所以, 当  $n_1, n_2, \dots, n_M$  确定之后,  $\mathbf{H}$  也随之确定, 公式 (3) 可以看作一个线性系统。当  $N = M$ , 且  $\mathbf{H}$  非奇异时, 线性系统 (10) 有唯一解; 而在多数情况下  $N > M$ , 该系统为超定的, 且不一定相容。求  $g_k$  也就等价于求公式 (10) 的极小范数最小二乘解:

$$\|\mathbf{H}\hat{\boldsymbol{\beta}} - \mathbf{E}\| = \min_{\boldsymbol{\beta}} \|\mathbf{H}\boldsymbol{\beta} - \mathbf{E}\|, \quad (14)$$

式 (14) 的极小范数最小二乘解可以通过求的 Moore-Penrose 广义逆得到<sup>[9]</sup>

$$\hat{\boldsymbol{\beta}} = \mathbf{H}^+ \mathbf{E}, \quad (15)$$

Moore-Penrose 广义逆可以通过正交投影 (Orthogonal projection)、奇异值分解 (Singular value decomposition, SVD) 等方法求得<sup>[9]</sup>。

可见, 脉冲的位置可以随机的确定, 这并不影响线性系统 (10) 存在极小范数最小二乘解。所以本文的多脉冲编码算法步骤如下:

第 1 步设定 (随机产生或指定均可) 各脉冲位置  $n_i, i = 1, \dots, M$ ;

第 2 步计算单位冲激响应矩阵  $\mathbf{H}$ ;

第 3 步计算脉冲幅度矢量  $\boldsymbol{\beta}: \boldsymbol{\beta} = \mathbf{H}^+ \mathbf{E}$ 。

从上面的推导可得, 脉冲位置并不会影响如式 (10) 所描述的线性系统存在极小范数最小二乘解, 也就是存在位于这些位置上的脉冲使得合成语音和原始语音在最小二乘意义下最接近。而脉冲位置的传输会导致编码速率增加, 从而占用大量的带宽。基于上面的讨论, 对于不同的语音帧, 可以采用位置相同 (固定的), 但是幅度不同的脉冲串作为激励信号。这样, 脉冲位置就不需要传输, 而只需要传输脉冲幅度和 LPC 系数, 我们称之为定点脉冲线性预测编码。

## 4 实验结果

### 4.1 位置无关脉冲搜索算法的效果

为了验证脉冲位置的不同是否会影响到最终的合成语音质量, 采用来自于五个不同说话者、同为“波澜壮阔”一词的语音对本文提出的位置无关脉冲搜索算法进行检验。语音的采样率均为 8000 Hz, 每 160 个样点分为一帧, 共 813 帧。每帧语音提取 18 个脉冲, 采用本文算法对同一帧语音提取脉冲 50 次, 每次提取之前, 脉冲位置随机确定, 进而采用本文提出的算法计算得到各脉冲幅度。

其中一帧语音的脉冲提取与语音合成结果如图 1 所示, 其中图 1(a) 和图 1(b) 为原始语音帧的残差信号和语音波形。为了与序贯法进行比较, 首先采用序贯搜索算法得到的脉冲位置, 而各脉冲的幅度则分别采用序贯法和位置无关脉冲搜索算法进行计算。图 1(c)、1(d) 和图 1(e)、1(f) 分别是序贯法和本文算法得到的激励信号 (位置相同, 但幅度不同) 及合成语音波形。两个脉冲激励都可以获得与原始语音非常接近的合成语音, 信噪比分别为 17.1471 dB 和 19.9547 dB, 本文算法优于序贯法。图 1(g)~图 1(p) 中分别列举了五个不同位置的脉冲激励所产生的合成语音, 信噪比均优于序贯法得到的合成语音。可见, 位于不同位置的脉冲, 应用本文算法所计算的幅度, 均可以得到较好的合成语音。对于所有语音帧 40650 次试验, 平均信噪比可以达到 19.2937 dB。本文算法在合成语音质量方面优于序贯法是因为, 序贯法每次确定一个脉冲, 在每次

搜索中都可以保证新增加的脉冲是最优的,但是不能保证所有的脉冲组合是最优的。而本文提出的算法可以保证得到的脉冲组合在最小二乘意义下是

最优的,这在理论上就保证了合成语音的质量。平均编码时间序贯法为0.1224 s,本文算法的时间为0.0019 s,只是序贯法的0.0155倍。

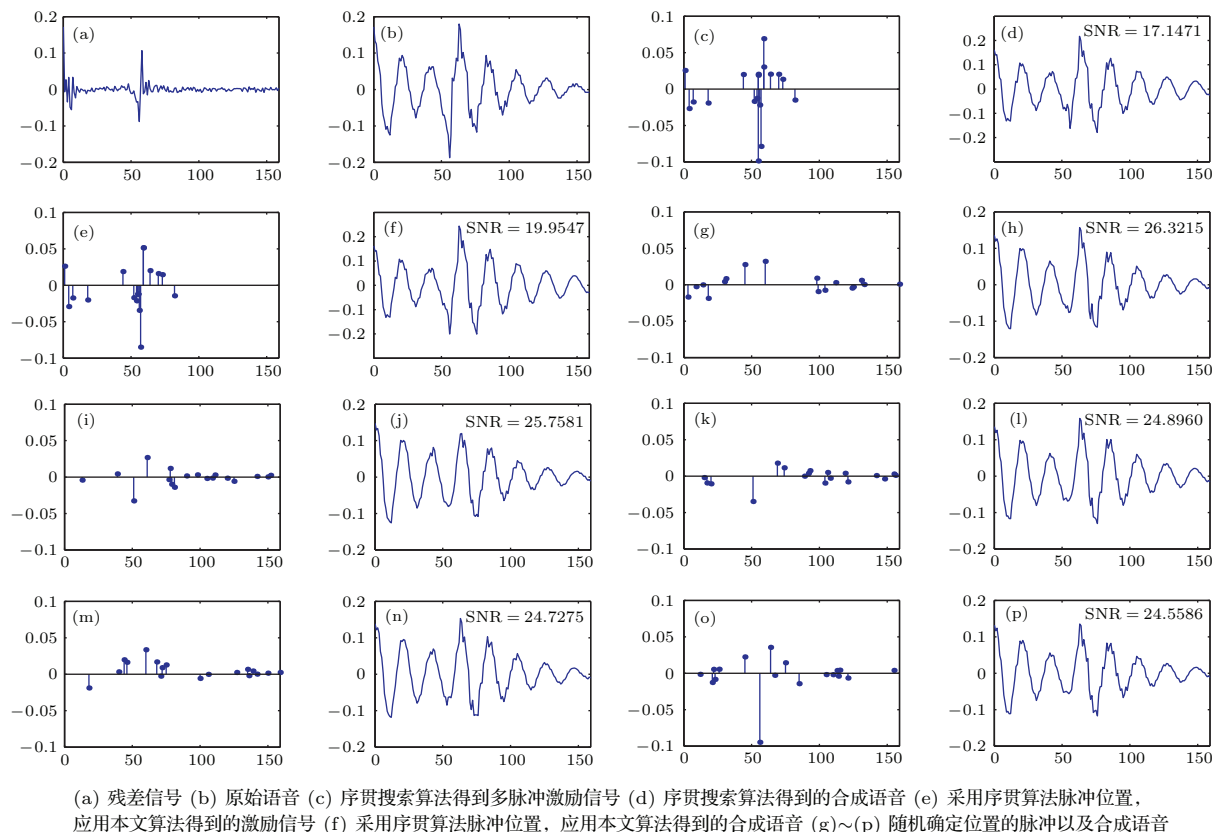


图1 一帧语音的脉冲提取与语音合成结果

Fig. 1 Excitation signals and the synthesized speech by using them

对上述语音,每帧语音用本文算法提取脉冲50次,脉冲个数分别为16、18、20、22、24、28、32,脉冲位置随机确定,各脉冲幅度用本文提出的算法计算,结果如图2所示。从图2(a)可以看出,随着脉冲数的增加,SNR的分布没有明显的变化,各个盒图的盒高、中值等都比较接近。50次实验所有帧SNR的均值会随脉冲数的增加有略微增加,但不明显;而其标准差会随脉冲数的增加而降低,这说明采用更多的脉冲会使得求脉冲幅度的过程更加稳定,受脉冲位置的影响更小。文献[10]中也提及,对于特定的语音,脉冲数增加到某个值,继续增加并不会对合成语音的质量有明显的改进。而从本文的方法来看,对于特定语音帧,当脉冲个数增加到某个值,其单位冲激响应矩阵 $\mathbf{H}$ 的秩并不会继续增加,对脉冲幅度的求解并无明显的帮助。图2(b)给出了平均编码时间随脉冲数增加的变化,从图2中可以看出,随

着脉冲个数的增加,平均编码时间会逐渐增加。当脉冲数增加到32个,平均编码时间为0.0028 s,也远远低于序贯法的0.1224 s。图3给出了同一帧,脉冲数不同的激励信号及对应的合成语音。实验结果显示各合成语音都与原始语音比较接近,具有较好的信噪比。

## 4.2 定点脉冲线性预测编码效果

### 4.2.1 未量化的合成语音质量比较

本文所提出的方法中,激励脉冲的位置不必通过算法确定,而是可以随机产生。这样,就可以对于每帧语音都采用位置相同,但是幅度不同的脉冲作为激励信号。分别用序贯法和本文提出的算法来处理一段语音,共2.5960 s,20768个样点。每160点分为一帧,每帧提取18个脉冲。不同算法得到的合成语音波形如图4所示。用本文算法得到的合成语音

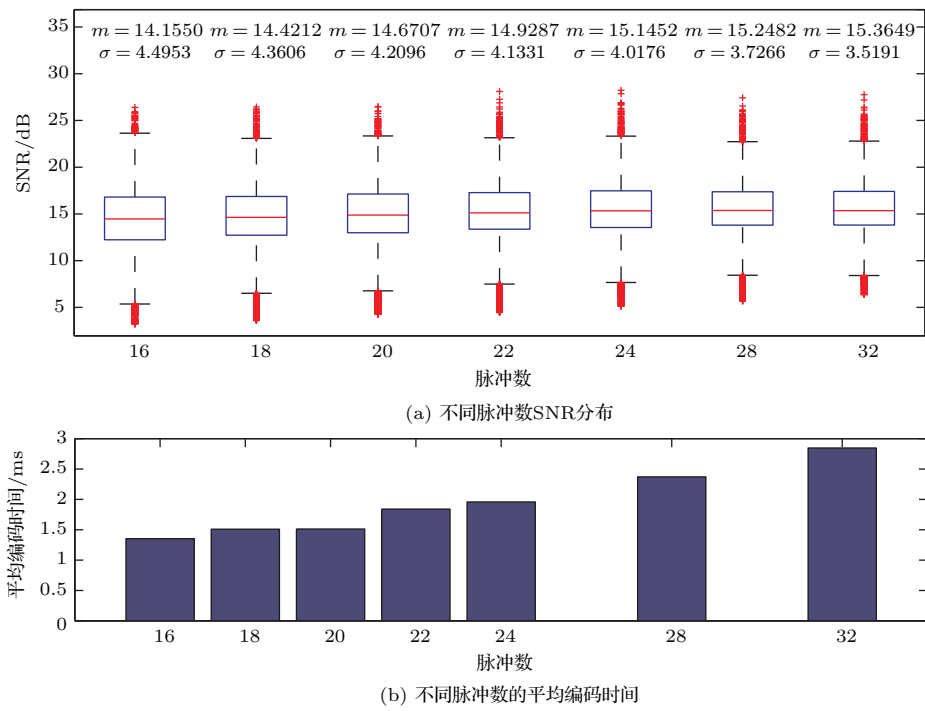


图2 平均编码时间随脉冲数增加的变化

Fig. 2 SNR and average coding time for different pulse numbers

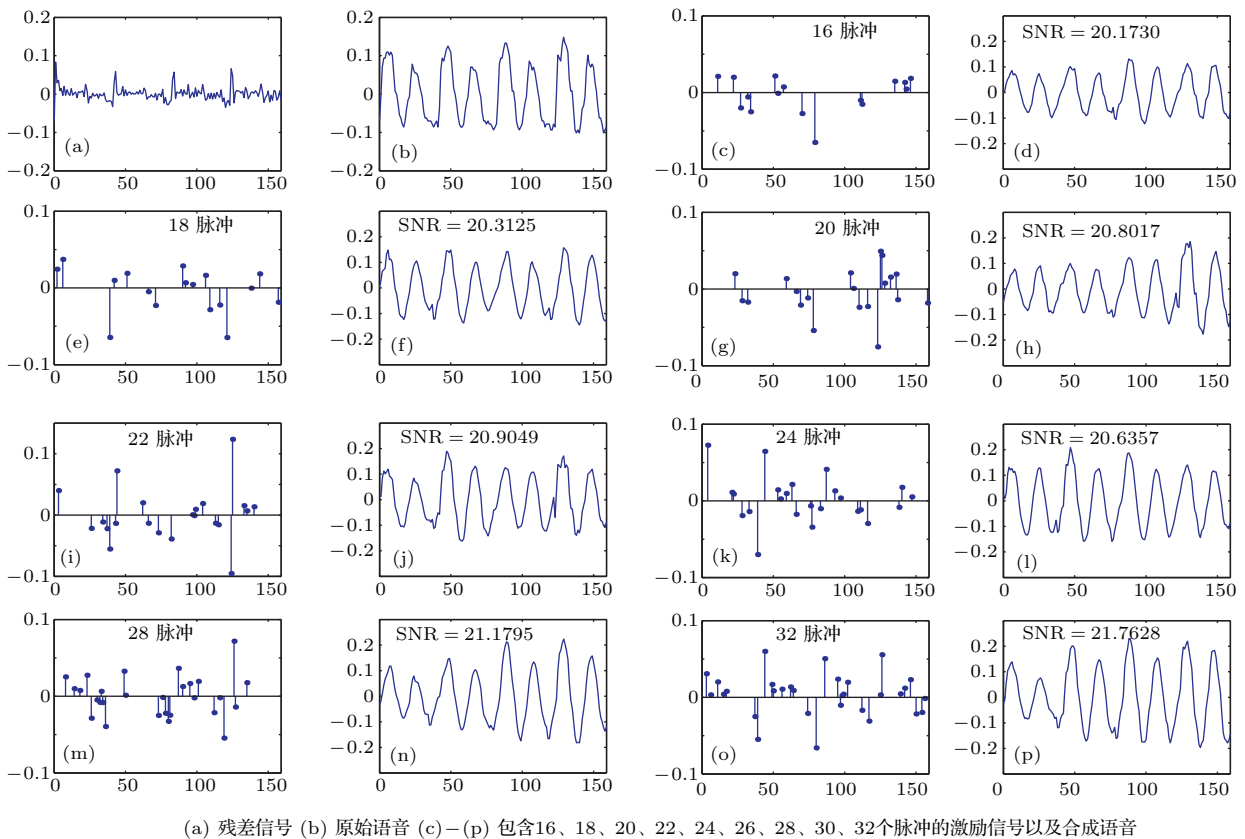


图3 同一帧, 脉冲数不同的激励信号及对应的合成语音

Fig. 3 Excitation signals with different numbers of pulses and synthesized speech by using them for the same speech frame

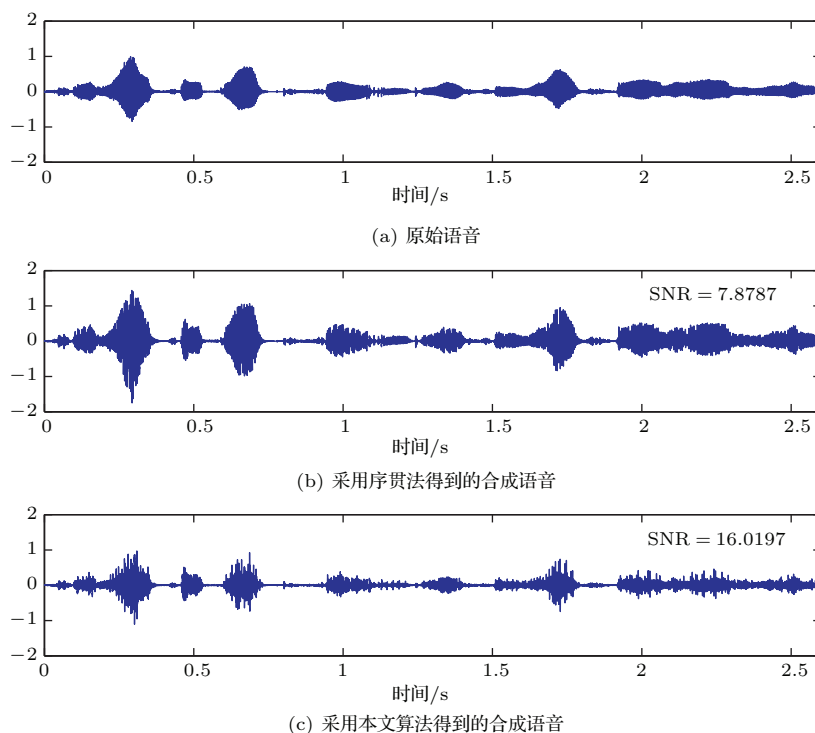


图4 不同算法得到的合成语音波形图

Fig. 4 Original speech and synthesized speech with different methods

信噪比为16.0197 dB, 合成语音具有良好的自然度和可懂度; 而传统的序贯搜索方法所得到的合成语音信噪比为7.8787 dB。对于该段语音, 按照ITU-T P.862建议书提供的PESQ(Perceptual evaluation of speech quality)方法进行客观MOS评价, 定点脉冲线性预测方法得到的合成语音可以取得4.049的PESQ\_MOS得分, 优于MP-LPC的3.478, 在清晰度和自然度方面都优于MP-LPC。

#### 4.2.2 量化方案及量化后合成语音质量比较

从上面的实验结果发现, 对于20 ms的语音帧, 16个脉冲已经足以获得质量良好的合成语音。所以在FLP-LPC的编码方案中, 采用16个位置上均匀分布(两个脉冲之间的间隔为10样点)的脉冲作为激励信号。脉冲幅度首先进行归一化, 然后对增益和归一化后的脉冲幅度矢量进行编码。需要编码的参数有LSF系数、首个非零脉冲出现的位置、增益、各脉冲的幅度。LSF系数和归一化后的脉冲幅度采用多级矢量量化的方法, 增益用4个比特来编码, 首个非零脉冲位置(0~9)用4个比特来标识。所以具体的比特分配如表1所示。

测试数据的采样频率为8000 Hz, 实验数据包含两个男声、两个女声, 共20组样本, 录音环境存

在背景噪声, 如开门声、汽车噪声等。分别采用本文编码方案及G.723.1、G.729对各组样本进行编解码, 并进行了PESQ\_MOS评价。实验结果如表2所示。从表2中可以看出, FLP-LPC可以在2.7 kbps的编码速率下获得与G.729类似的合成语音, 略优于G.723.1。

表1 FLP-LPC 比特分配

Table 1 FLP-LPC bit allocation

参数	bits 数
LSF 系数	18
首个非零脉冲出现的位置	4
增益	4
脉冲幅度矢量	28
总计	54

表2 FLP-LPC 与 G.723.1、G.729 的比较

Table 2 Performance comparison of FLP-LPC with G.723.1 and G.729

编码方法/标准	编码速率 (kbps)	PESQ_MOS
FLP-LPC	2.7	3.724
G.723.1	5.3	3.583
G.729	8	3.730



## 5 结论

针对MP-LPC中存在的问题,本文提出了位置无关脉冲搜索算法。理论研究发现脉冲的位置并不会影响最优幅度矢量在最小二乘意义下存在,所以提出的算法并不对脉冲的位置进行搜索。经过对不同语音帧的多种脉冲位置组合的实验发现,位置不同的脉冲作为激励都可以产生质量很好的合成语音,这些合成语音均优于传统MP-LPC算法得到的合成语音。在提高了合成语音质量的基础上,该算法并未增加编码时间,其编码时间约占MP-LPC编码时间的1.5%。为了研究脉冲数对合成语音质量的影响,对同一语音,提取不同个数的脉冲。结果发现,16个脉冲的激励信号足以满足20 ms语音帧的编码需求。

为了进一步降低编码速率,提出了定点脉冲线性预测编码。定点脉冲线性预测编码中各帧的位置都采用固定的组合,每次只是通过位置无关脉冲搜索算法来得到脉冲的幅度矢量。这样就不需要对脉冲位置进行编码,可以有效的降低编码速率。虽然编码速率降低,但是合成语音质量并未受到影响,该算法的合成语音信噪比远高于传统MP-LPC,主观听音效果也优于MP-LPC。进而,提出了脉冲均匀分布的FLP-LPC编码方案。研究发现FLP-LPC可以在2.7 kbps的编码速率下取得与G.729接近的合成语音质量,略优于G.723.1。FLP-LPC的合成语音质量高、编码时间短,同时也降低了编码速率,唯一不足是矩阵的伪逆运算需要较大的内存。

## 参 考 文 献

[1] 武朋辉, 杨百龙, 时磊, 等. 基于离散余弦变换的语音压缩采

- 样和编码算法[J]. 应用声学, 2015, 34(1): 17-23.
- WU Penghui, YANG Bailong, SHI Lei, et al. Speech compressive sensing and codec algorithm based on discrete cosine transform[J]. J. Applied Acoustics, 2015, 34(1): 17-23.
- [2] 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, 31(2): 109-117.
- XIAO Dong, MO Fuyuan, CHEN Geng, et al. Studies of the line spectral frequency vector quantization in mixed excitation linear prediction[J]. J. Applied Acoustics, 2012, 31(2): 109-117.
- [3] 马震, 陈延萍. 最小方差无失真响应浊音谱建模方法研究[J]. 应用声学, 2008, 27(4): 326-332.
- MA Zhen, CHEN Yanping. Modeling of sonant based on minimum variance distortionless response spectrum[J]. J. Applied Acoustics, 2008, 27(4): 326-332.
- [4] KLEIJN W B. Encoding speech using prototype waveforms[J]. IEEE Transactions on Speech and Audio Processing, 1993, 1(4): 386-399.
- [5] MCAULAY R J, QUATIERI T F. Speech analysis/synthesis based on a sinusoidal representation[J]. IEEE Trans Acoustic, Speech, Signal Processing, 1986, 34(4): 744-754.
- [6] GRIFFIN D W, LIM J S. Multi-band excitation vocoder[J]. IEEE Trans. Acoustic, Speech, Signal Processing, 1987, 36(8): 1223-1235.
- [7] ATAL B S, HANAUER S L. Speech analysis and synthesis by linear prediction of the speech wave[J]. The Journal of the Acoustical Society of America, 1971, 50(2): 637-655.
- [8] SCHROEDER M R, ATAL B S. Code-excited linear predictive(CELP): High quality speech at very low bit rates[C]. IEEE Int. Conf. ASSP, Tampa, 1985: 937-940.
- [9] SERRE D. Matrices: Theory and applications[M]. New York: Springer-Verlag, 2002: 147-158.
- [10] 马震, 蒋保臣, 陶立慧. MPLPC激励脉冲提取算法研究[J]. 计算机应用与软件, 2007, 24(11): 183-185.
- MA Zhen, JIANG Baochen, TAO Lihui. Research on MPLPC excited-pulse abstract algorithm[J]. Computer Applications and Software, 2007, 24(11): 183-185.