第 36 卷 第 1 期
2017 年 1 月

应用声学
Journal of Applied Acoustics

Vol. 36, No.1
January, 2017

◇ 研究报告 ◇

个数可变脉冲线性预测编码研究*

马 震†

(滨州学院信息工程系 滨州 256600)

摘要 不同语音帧的激励信号复杂性不同，所以采用相同个数的脉冲作为激励信号并不合理。针对这一点，提出了个数可变脉冲线性预测编码算法。该算法不固定脉冲个数，而是根据激励信号的复杂度而确定。个数可变脉冲线性预测编码的目的是用尽量少的脉冲数来满足误差约束，这可以看作一个稀疏表示问题。进而，给出了具体的脉冲搜索算法以及个数可变脉冲线性预测编码方案。实验结果发现增加脉冲可以减少误差，但是前面搜索出的脉冲对误差的贡献要大于后搜索出的脉冲。与 G.723.1 和 G.729 比较发现，个数可变脉冲线性预测编码可以在约 4.2 kbps 的编码速率下获得优于 G.723.1 的合成语音，但略差于 G.729。本文算法的编码时间较长，是下一步需要解决的问题。

关键词 语音编码，个数可变脉冲线性预测编码，稀疏表示，正交匹配追踪
中图分类号: TN912.3 **文献标识码:** A **文章编号:** 1000-310X(2017)01-0048-06
DOI: 10.11684/j.issn.1000-310X.2017.01.008

Number-variable pulse linear prediction coding

MA Zhen

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

Abstract It is not reasonable to use same number of pulses for all frames with different excitation complexity in linear prediction coding (LPC). For this, number-variable pulse linear prediction coding (NVP-LPC) is presented in this paper. In this algorithm, the number of pulses is not fixed but determined according to the complexity of speech signal. NVP-LPC aims at satisfying the error constraint at the cost of the least pulses, which can be seen as a sparse representation problem. Moreover, the detailed pulse search algorithm and coding scheme are presented. The results show that adding pulses can reduce the error, and the pulses determined earlier make more contributions to reduce the error than latter pulses. Comparing with G.723.1 and G.729, results show NVP-LPC can obtain the synthesized speech superior to G.723.1 but slightly worse than G.729 at the coding rate of about 4.2 kbps. The future work focuses on reducing the coding time.

Key words Speech coding, Number-variable pulse linear prediction coding, Sparse representation, Orthogonal matching pursuit

2016-04-14 收稿; 2016-07-22 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, 博士, 研究方向: 语音信号处理、生物医学信号处理。

†通讯作者 E-mail: 13954380080@139.com

1 引言

话音业务是移动通信系统的最基本也是最重要的业务，有限的无线资源要求在尽可能低的带宽获得尽可能好的合成语音，所以语音编码一直受到人们的重视。波形编码主要描述的是语音的波形，而参数编码和混合编码多是描述产生语音的模型，即：声道模型和激励模型。

混合编码可以在编码速率和语音质量之间达到很好的折中。典型算法有多脉冲线性预测编码 (Multi-pulse linear prediction coding, MP-LPC)[1]、激励线性预测 (Code excited liner prediction, CELP)[2]，这两种算法对声道的描述都是一样的，只是激励模型不同。混合编码要解决的问题其实就是如何以最少的比特来描述声道和激励信号。后来的研究者对 MP-LPC 中激励脉冲幅度和位置的信息进行了一系列的简化。规则脉冲线性预测编码 (Regular-pulse excitation linear prediction coding, RPE-LPC) 中 [3]，采用等间隔的脉冲串作为激励。这样，只需要确定并传输第一个脉冲的位置和各个脉冲的幅度。在多脉冲最大似然量化 (Multipulse maximum likelihood quantization, MP-MLQ) 中 [4]，脉冲的位置为奇数或者偶数，幅度为 ± 1。在上述两种改进的 MP-LPC 方法中，通过使激励脉冲更加规则来减少需要传输的信息，但是并未考虑语音帧间复杂度的不同。文献 [5] 使用了可变阶数滤波器来建模声道，也就是根据产生不同语音的声道复杂度来确定滤波器的阶数，使得对声道的编码更加有效。相对于声道复杂度的差异，激励信号复杂度的差异则更加明显，有更多的压缩空间。本文则根据激励的复杂度来确定脉冲的个数，提出了一种个数可变脉冲线性预测编码 (Number-variable pulse linear prediction coding, NVP-LPC) 方法。这种方法的中心思想是能满足误差要求的条件下，脉冲的个数越少越好。我们把这个问题看作一个稀疏表示问题，并给出了具体的脉冲搜索步骤。最后，提出了 NVP-LPC 的编码方案，并与 G.723.1 和 G.729 进行了比较。

2 个数可变脉冲线性预测编码算法

2.1 问题描述

设帧长为 N 的语音帧，激励信号写作 $p(n)=\sum_{k=1}^{N} g_k * \delta(n-k)$，这里 g_k 表示位于 k 位置的脉冲幅度。语音编码的关键是确定 g_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

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

这里，$\hat{s}_0(n)$ 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

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

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

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

其中，$h_w(n)$ 是加权综合滤波器的单位冲激响应。

语音编码的思想就是通过确定每个 g_k，$k=1,2,\cdots,N$，力求 $\|e(n)\|=\left\|\bar{e}_w(n)-\sum_{k=1}^{N} g_k * h_w(n-k)\right\|$ 最小。当一组 $g_k, k=1,2,\cdots,N$ 满足

$$\sum_{k=1}^{N} g_k * h_w(n-n_k)=\bar{e}_w(n), \tag{4}$$

$\|e(n)\|$ 达到最小值，也就是 $\|e(n)\|=\left\|\bar{e}_w(n)-\sum_{k=1}^{N} g_k * h_w(n-k)\right\|=0$。

以上 N 个等式写作矩阵的形式为

$$\boldsymbol{H}_w \boldsymbol{g}=\boldsymbol{E}, \tag{5}$$

其中，

$$\begin{aligned} \boldsymbol{H}_w &= \begin{bmatrix} h_w(1-1) & \cdots & h_w(1-N) \\ \vdots & & \vdots \\ h_w(N-1) & \cdots & h_w(N-N) \end{bmatrix}_{N\times N} \\ &= \begin{bmatrix} h_w(0) & \cdots & h_w(1-N) \\ \vdots & h_w(0) & \vdots \\ h_w(N-1) & \cdots & h_w(0) \end{bmatrix}_{N\times N}, \end{aligned} \tag{6}$$

$$\boldsymbol{g}=[g_1,g_2,\cdots,g_N]^{\mathrm{T}}, \tag{7}$$

$$\boldsymbol{E}=[\bar{e}_w(1),\bar{e}_w(2),\cdots,\bar{e}_w(N)]^{\mathrm{T}}, \tag{8}$$

这里，令 $\boldsymbol{h}_w^m=[h(1-m),h(2-m),\cdots,h(N-m)]^{\mathrm{T}}$，所以 $\boldsymbol{H}_w=(\boldsymbol{h}_w^1,\boldsymbol{h}_w^2,\cdots,\boldsymbol{h}_w^N)$，则 $\boldsymbol{E}=\sum_{k=1}^{N}g_k*\boldsymbol{h}_w^k$。

所以，当LPC滤波器确定之后，$\boldsymbol{H}_w$ 也随之确定，公式(5)可以看作一个线性系统。当 $\boldsymbol{H}_w$ 非奇异时，线性系统(5)有唯一解；而在多数情况下 $\boldsymbol{H}_w$ 为奇异的，该线性系统可能存在无穷多个解。

我们可以令矩阵 $\boldsymbol{H}_w$ 的秩为 $R(\boldsymbol{H}_w)$，则 $\boldsymbol{h}_w^1,\boldsymbol{h}_w^2,\cdots,\boldsymbol{h}_w^N$ 中有 $R(\boldsymbol{H}_w)$ 个有效矢量，而剩余的 $N-R(\boldsymbol{H}_w)$ 为非有效矢量，非有效矢量可以由有效矢量的线性组合来表示。假设 $\mathbf{J}$ 为所有有效变量下标的集合，只要 $k\notin\mathbf{J}$，则 g_k 可以等于零，而 $\boldsymbol{E}$ 可以表示为

$$\boldsymbol{E}=\sum_{k\in\mathrm{J}}g_k*\boldsymbol{h}_w^k. \tag{9}$$

所以从理论上来说，$R(\boldsymbol{H}_w)$ 是 $\boldsymbol{g}$ 中非零元素个数的上限，也就是激励信号中脉冲个数的上限。从语音传输的角度来说，脉冲个数越少，则编码速率越低。所以如何在保证合成语音质量的前提下，尽量减少脉冲的数量是激励信号建模的核心问题。这个问题可以表示为

$$\min_{\boldsymbol{g}}\|\boldsymbol{g}\|_0,\ s.t.\ \boldsymbol{H}_w\boldsymbol{g}=\boldsymbol{E}. \tag{10}$$

2.2 脉冲提取步骤

公式(10)所表示的问题可以看作一个稀疏表示的问题，解这个问题的方法很多，包括匹配追踪(Matching pursuit, MP)算法[6]，基追踪(Basis pursuit, BP)算法[7]，内点法[8]等。在本文中，采用了正交匹配追踪算法(Orthogonal matching pursuit, OMP)来解 $\boldsymbol{g}$。每次解出一个非零脉冲最优的位置和幅度，直至合成语音与原始语音之间的误差足够小。具体算法步骤如下：

第1步：初始化：$\Omega_0=\varphi$，$\boldsymbol{\Omega}_j$ 为第 j 次迭代之后，脉冲位置的集合，初始化为空集；$\boldsymbol{e}_0=\boldsymbol{E}$，$\boldsymbol{e}_j$ 为第 j 次迭代之后的残差信号，初始化为 $\boldsymbol{E}$；$\boldsymbol{g}_0=\mathbf{0}$，$\boldsymbol{g}_0$ 为脉冲幅度矢量，初始化为 $\mathbf{0}$ 矢量；脉冲个数为 $M\leqslant R(\boldsymbol{H}_w)$，允许的误差为 ε。

第2步：while($\|\boldsymbol{e}_j\|_2>\varepsilon$ 且 $j<M$)，开始循环

(1) 确定一个与当前残差最相关的第 n_j 个列矢量 $\boldsymbol{h}_w^{n_j}$，即 $n_j=\arg\max_{n_j}\left|\left\langle\boldsymbol{h}_w^{n_j},\boldsymbol{e}_{j-1}\right\rangle\right|$，并更新 $\boldsymbol{\Omega}_j=\boldsymbol{\Omega}_{j-1}\cup\{n_j\}$。

(2) 更新 $\boldsymbol{g}_j=\arg\min_{\boldsymbol{g}_j}\|\boldsymbol{E}-\boldsymbol{H}_w(\boldsymbol{\Omega}_j)*\boldsymbol{g}_j\|_2^2$。

(3) 更新 $\boldsymbol{e}_j=\boldsymbol{E}-\boldsymbol{H}_w(\boldsymbol{\Omega}_j)*\boldsymbol{g}_j$，$j=j+1$。结束循环。

第3步：输出 $\boldsymbol{\Omega}_j$ 及 $\boldsymbol{g}_j$。

2.3 e-脉冲数

e-脉冲数定义为

$$N^e=\|\boldsymbol{g}^e\|_0, \tag{11}$$

这里 e 为可允许的 $\boldsymbol{E}$ 与 $\boldsymbol{H}_w\boldsymbol{g}$ 之间的均方误差，N^e 为误差为 e 时的脉冲数。在本文的问题中，e 体现了合成语音与原始语音之间的近似程度。较大的 e 会使算法忽略掉一些对产生合成语音信号作用不大的脉冲，而较小的 e 则会将这些作用不大的脉冲也纳入考虑。

3 实验结果

3.1 e-脉冲数的分布

为了讨论不同语音帧的 e-脉冲数的分布情况，采用的语音数据来自于两个不同的说话者，语音的采样率均为8000 Hz，每160个样点分为一帧，共697帧。对每帧语音用本文提出的算法计算其 $N^{0.5}$、$N^{0.1}$、$N^{0.01}$，其各自的分布如图1所示。从图1中可见，当 e 减小的时候，会引起脉冲数的增加，也就是需要增加脉冲来抵消原始语音和合成语音之间的误差。当 e 为0.5时，只有1帧需要19个脉冲，其余均需要不超过15个脉冲，最少的只需要2个脉冲就可以满足要求；而当 e 变为0.1时，所需的脉冲数有所增加，脉冲数在16~20之间的语音帧达到184帧，但是其中3帧依然只需要2个脉冲；当 e 减少为0.01，脉冲数在16~20之间的语音帧达到252帧，脉冲数最低的为3。实验结果显示，当 e 从0.5减少为0.1，有458帧需要增加脉冲数；当 e 从0.1 减少为0.01，则有360帧需要增加脉冲数。所有语音帧的脉冲数 $N^{0.5}$、$N^{0.1}$、$N^{0.01}$ 的均值分别为9.6212、12.3429和14.3214，也呈现出增加的趋势。

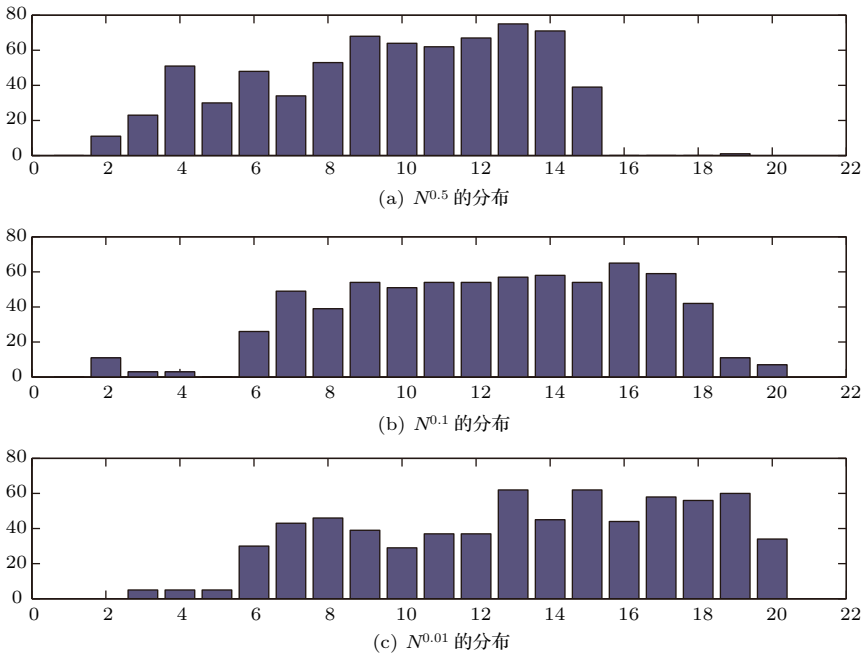


图 1　不同的 e-脉冲数分布

Fig. 1　Distribution of e-pulse number

其中一帧语音不同脉冲数的激励信号与合成语音结果如图 2 所示，其中图 2(a) 和图 2(b) 为原始语音帧的残差信号和语音波形。图 2(c)–图 2(l) 给出了分别具有 6、8、10、14、18 个脉冲的激励信号及对应的合成语音。从图 2 可以看出，随着脉冲数的增加，合成语音的波形会更加接近原始语音波形，信噪比会逐步增加。但是，当脉冲达到某个值，继续增加脉冲对信噪比的改善并不明显，这也说明了，开始搜索到的脉冲，能对匹配原始语音起到较大的作用，而在后面搜索到的脉冲，所起的作用较小。

3.2　个数可变脉冲线性预测编码方案与性能评价

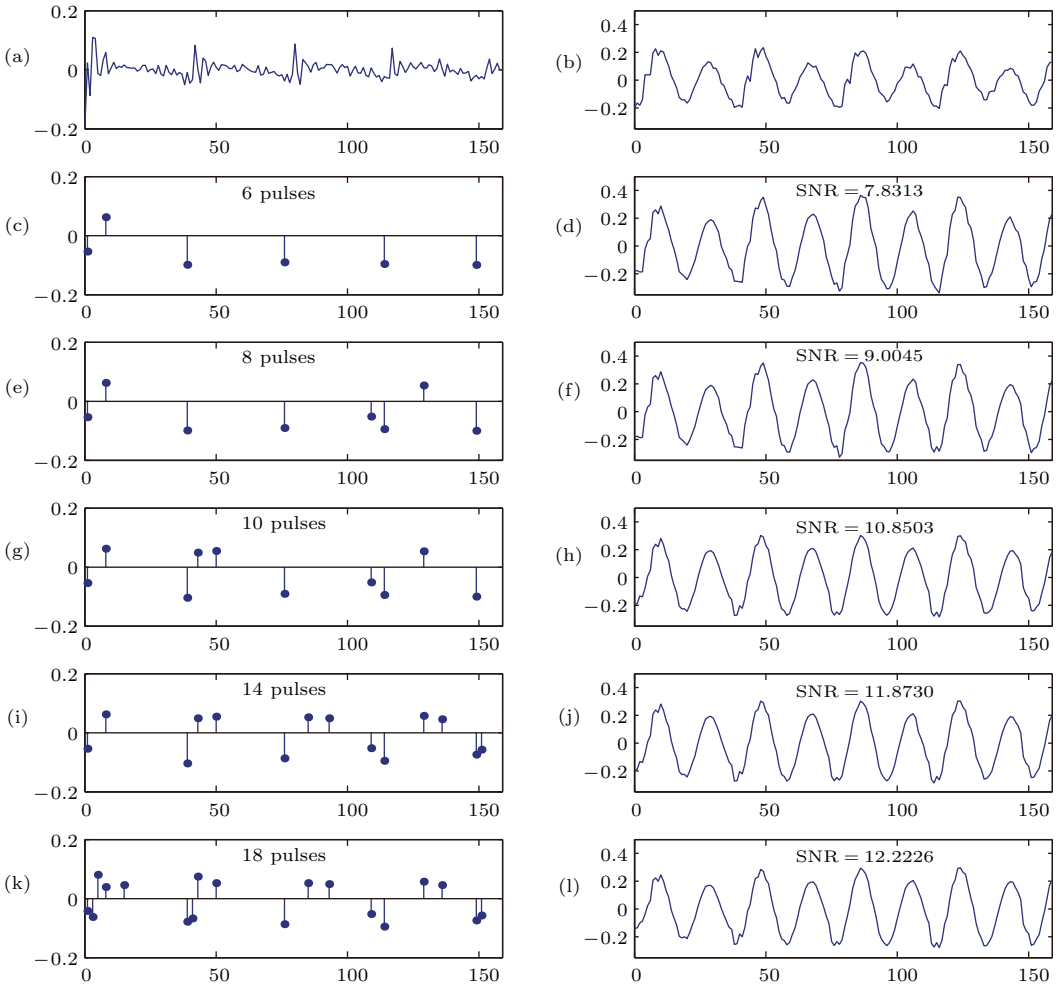
对于一般的语音帧，e 一般取值 0.1 即可取得较好的合成语音。在本文提出的个数可变脉冲线性预测编码的基础上，提出具体的编码方案如下。每 240 个样点 (30 ms) 分为一帧，e 取 0.1，应用本文提出的算法进行脉冲提取。脉冲个数的范围为 6∼20 个，用 4 比特来标识。每个脉冲位置用 8 比特来编码，而脉冲幅度根据脉冲个数分为基本脉冲和超出脉冲。脉冲个数在 12 个以下的称为基本脉冲，以 20 比特进行矢量量化；超出 12 个的脉冲称为超出脉冲，第 13∼16 个为一组，第 17∼20 个为一组，分别用 5 比特来矢量量化。这样设计是因为后面的脉冲对于合成语音质量的贡献相对较小，所以采用较少的比特进行编码。LSF 采用多级矢量量化的方法，以 18 比特编码。所以具体的比特分配如表 1 所示。

表 1　NVP-LPC 比特分配

Table 1　NVP-LPC bit allocation

参数	bits 数
LSF 系数	18
脉冲个数	4
每个脉冲位置	8
基本脉冲幅度矢量 (12 脉冲)	20
超出脉冲幅度矢量 (4 脉冲)	5

测试数据包括 83 个男声和 83 个女声，共 1560 句，内容是选自人民日报。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。从表中可以看出，对于测试语音，NVP-LPC 可以在约 4.4 kbps 的编码速率下获得优于 G.723.1 的合成语音，相比 G.729 略差。G.723.1 和 G.729 的平均编码时间较为接近，分别为 0.0576 s 和 0.0438 s，NVP-LPC 的平均编码时间为 0.2328 s。



(a) 残差信号 (b) 原始语音 (c−l) 包含6、8、10、14、18个脉冲的激励信号以及合成语音

图 2　一帧语音的脉冲提取与语音合成结果

Fig. 2　Excitation signals and the synthesized speech using them

表 2　NVP-LPC 与 G.723.1、G.729 的比较

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

编码方法/标准	编码时间 (s)	编码速率 (kbps)	PESQ_MOS
NVP-LPC	0.2328	约 4.4	3.762
G.723.1	0.0576	5.3	3.611
G.729	0.0438	8	3.815

4　结论

对于不同的语音帧，声门激励是不同的，这也决定了采用固定的激励模型来建模声门激励是不合理的。在本文中，对于复杂度不同的声门激励，采用个数不同的脉冲串来描述，可以更加有效地描述激励信号。本文提出了个数可变脉冲线性预测编码的方法，着重讨论了脉冲迭代搜索的算法。通过仿真实验给出了不同语音帧的脉冲数分布情况。实验发现脉冲数增加可以有效的降低合成语音和原始语音之间的误差。随着搜索过程的进行，后面搜索得出的脉冲对于改善语音质量的贡献要小于前面搜索得出的脉冲。

进而，给出了 NVP-LPC 的具体编码方案。对于测试语音，NVP-LPC 可以在约 4.2 kbps 的编码速率下获得优于 G.723.1 的合成语音，但是略差于 G.729。可见，NVP-LPC 可以根据不同语音帧激励信号的复杂度进行编码，比起传统的混合语音编码方法更加有效。本文所采用的测试语音数据库包括了日常生活中遇到的不同复杂度的语音，所以 NVP-LPC 所获得的 4.2 kbps 的编码速率具有一定代表性。这种根据语音信号的复杂度不同而进行的

变长编码可以更好的体现有效性，应该是未来语音编码领域着重考虑的课题。而本文算法的平均编码时间为0.2328 s，这距离实际应用还有较大差距，也是下一步需要解决的问题。

参 考 文 献

[1] ATAL B S, HANAUER S L. Speech analysis and synthesis by linear prediction of the speech wave[J]. The Journal of the Acousticol Society of America, 1971, 50(2): 637–655.

[2] 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.

[3] KROON P, DEPRETTERE E, SLUYTER R. Regular-pulse excitation–A novel approach to effective and efficient multipulse coding of speech[J]. IEEE Trans. Acoustics, Speech, and Signal Processing, 1986, 34(5): 1054–1063.

[4] YOON S, KANG H, PARK Y, et al. An efficient transcoding algorithm for G.723.1 and G.729A speech coders: Interoperability between mobile and IP network[J]. Speech Communication, 2004, 43(1–2): 17–31.

[5] 蒋保臣, 马震. 改进 MPLPC 语音编码方法 [J]. 计算机应用与软件, 2005, 22(8): 77–79.
JIANG Baochen, MA Zhen. Modified MPLPC speech coding method[J]. Computer Applications and Software, 2005, 22(8): 77–79.

[6] MALLAT S, ZHANG Z. Matching pursuit with time-frequency dictionaries[J]. IEEE Transactions on Signal Processing, 1993, 41(12): 3397–3415.

[7] CHEN S, DONOHO D L. Basis pursuit[C]. IEEE, 1995: 41–44.

[8] KIM S J, KOH K, LUSTIG M, et al. An interior-point method for large-scale $l1$-regularized least squares[J]. IEEE Journal on Selected Topics in Signal Processing, 2007, 1(4): 606–617.

◇ 声学新闻和动态 ◇

2016年度全国检测声学与物理声学会议在昆明成功举行

2016年11月10–13日，由中国声学学会检测声学分会和物理声学分会主办，中国科学院声学研究所和南京大学声学所承办，云南省地球物理学会协办的2016年度全国检测声学与物理声学会议在昆明召开。中国声学学会学术委员会主任程建春、中国声学学会检测声学分会主任王秀明、中国声学学会物理声学分会主任刘晓宙、云南省科协副主席、云南省地球物理学会理事长叶燎原、中国声学学会功率超声分会主任林书玉、中国声学学会检测声学分会副主任刘晓峻、中国声学学会物理声学分会副主任张碧星等人出席了开幕式。来自全国各大学、科研院所、公司企业等47家单位的160多名代表参加了会议，围绕检测声学及物理声学相关的基础、应用基础及其前沿技术等热点展开了热烈的学术交流。

大会开幕式由刘晓峻教授主持，王秀明研究员、刘晓宙教授分别致开幕词。叶燎原理事长代表协办单位致欢迎词。本次会议邀请了8位知名专家学者就声学超材料、超声检测、声波测井和声学传感器等热点问题进行主题报告，分别是：南京大学程建春教授介绍了基于声超材料的螺旋波生成；武汉大学刘正猷教授介绍了声子晶体中的谷涡旋态及输运；南京理工大学沈中华教授介绍了激光超声实现缺陷检测和薄膜力学性质表征的研究；北京航空航天大学周正干教授介绍了相控阵超声成像方法及其应用；中国科学院声学研究所陈浩研究员介绍了三维声波测井研究新进展；南京大学卢明辉教授介绍了声拓扑态的研究；厦门大学张宇教授介绍了海豚声波束形成物理机理及仿生应用研究；黑龙江大学刘盛春教授介绍了光纤声学传感器。同时会议安排了检测超声与光声检测、固体声学与深部钻测、声学超材料及应用、声学中的基本物理问题四个专题，专题报告以青年学者为主体，充满了热烈的讨论和激情的思维碰撞，主持人不得不经常打断讨论，以保证报告准时完成。报告结束后，会议学术委员会评选出15篇优秀论文，并为作者颁发了奖状以资鼓励。

大会闭幕式由刘晓峻教授主持，程建春教授在致辞中表示，本次会议报告人准备充分，听讲人聚精会神，热烈的讨论一直延续到会议结束，最后取得圆满成功。刘晓宙教授致闭幕词，宣布2016年度全国检测声学与物理声学会议胜利闭幕，感谢云南省地球物理学会的大力支持，特别感谢会务组为会议召开做出的巨大贡献。他指出，本次会议有两个显著的特点：一是创新性和实用性，如声拓扑、声螺旋、声谷涡旋等新概念的提出，代表了国际声学领域的最高水平；而相控阵技术、声波测井技术、激光超声技术、光纤传感技术、声场仿真技术等反应出我国声学技术的广泛应用。二是本次会议涌现出一大批具有博士学位的青年科技工作者，他们的报告都具有鲜明的特点。本次会议表明，我国的声学事业在原始创新和技术应用方面都取得了较大的进步，希望各位代表携手努力，为我国科学事业和经济发展做出更大贡献。

本次会议获得与会代表一致好评，普遍认为本次会议展示了检测声学和物理声学领域理论及应用研究的最新成果，极大的促进了学术交流，加深了各科研单位间的了解与合作，对我国检测声学和物理声学的理论和应用的发展产生了积极的促进作用。

（中国声学学会　　安志武）