语音信号产生模型的建立及应用

王莉华
(周口师范学院 物理与电子工程系, 河南 周口 466001)

摘 要:从人类语音产生的机理出发,介绍了语音信号的特征和语音信号的语谱图,引出了语音信号的产生模型。同时讨论了在语音信号产生的模型应用中,线性预测编码方法及语音产生模型在语音合成和语音识别中的应用原理,体现了语音产生模型在语音处理技术方面的重要地位。

关键词:模型;频率;线性预测编码
中图分类号:TP391.42 **文献标识码**:A

Design and application of the pronunciation signal producing model

WANG Li Hua
(Physics and Eelectronic Engineering Department, Zhoukou Normal College, Zhoukou 466001, China)

Abstrsct: This paper introduced the pronunciation signal characteristic and the pronunciation signal language spectrogram, which the mechanism that pronciation produces set off from human being simultaneously. It discussed the model application in pronunciation signal produces, line predictive coding method in the pronunciation signal and the pronunciation produces the model in the pronunciation synthesis and the speech recognition application principle, manifested the important status that the pronunciation produces the model in the pronunciation processing technology aspect.

Key words: model; frequency; LPC

语音由一连串的音所组成,这些音及其相互间的过渡就是代表信息的符号。这些符号的排列由语音的规则所控制。对这些规则及其在人类通信中的含义的研究属于语言学的范畴。但对语音信号加以处理以改善或提取信息时,有必要对语音产生的机理进行讨论。

图1为发音器官示意图。声道起始于声带的开口(即声门处)而终止于嘴唇,它包含了咽喉(连接食道和口)和口(或称为口腔)。声道的截面积取决于舌、唇、颌以及小舌的位置,它可以从0(完全闭合)变化到约20 cm^2,鼻道则从小舌开始到鼻孔为止。当小舌下垂时,鼻道与声道发生声耦合而产生语音中的鼻音。另外,图中还包含了由肺、支气管、气管组成的次声门系统,这个次声门系统是产生语音能量的源泉。当空气从肺里呼出时,呼出的气流由于声道某一地方的收缩而受到扰动,语音就是这一系统在此时辐射出来的声波。

语音的声音按其激励形式的不同可分为三类:浊音、摩擦音和爆破音。浊音:当气流通过声门时,如果声带的张力刚好使声带发生张弛振荡式的振动,就能产

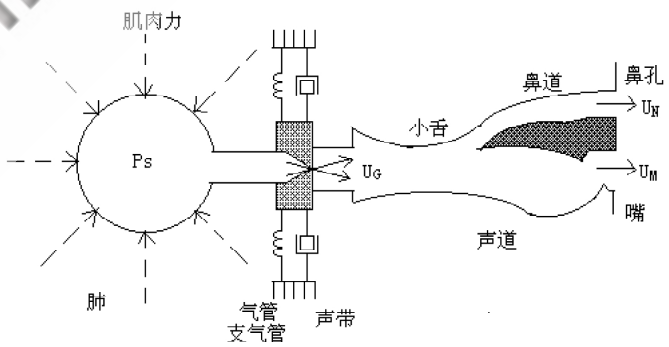

图1 发音器官示意图

生准周期的空气脉冲,这一空气脉冲激励声道得到浊音,如音标中的"U"、"d"、"w"、"i"、"e"等为浊音。摩擦音或称为清音:如果声道在某处(一般在接近嘴的那端)发生收缩,同时迫使空气以高速冲过这一收缩部分而产生湍流,从而得到摩擦音,此时建立的宽带噪声源激励了声道,如音标中的"ʃ"就是摩擦音;爆破音:如果使声道前部完全闭合,在闭合后建立起气压,然后突然释放,这样就得到了爆破音,如音标中的"tʃ"就是爆

破激励产生的。

1 语音信号的特征和语谱图

图1中声道和鼻道都表示为非均匀截面的声管，当声音产生以后就顺着声管传播，它的频谱形状会被声管的选择性所改变。这类似于人们在管风琴或管乐器中所看到的谐振现象。在此将声道管的谐振频率称为共振峰频率。共振峰频率和声道的形状与大小有关，每种形状都有一套共振峰频率作为其特征。改变声道的形状就产生不同的声音，因此，当声道形状改变时，语音信号的谱特性也随之改变。

语谱图是通过语谱仪画出的、以显示语音信号的通用图。它的垂直方向表示频率，水平方向表示时间。图2表示了一段英语语句的语音信号。

获得这些图的原理大致如下：

首先把语音信号拆成短的时段，一般为2 ms～40 ms，然后在合适的窗口长度上使用FFT找每一短时段的频谱。图中每一点表示在给定时间和给定频率范围内频谱的能量。段的长度是根据频率分辨率和时间分辨率要求折中选择的。目前数字信号处理技术水平已能够实时处理语音频谱随时间的变化，这就意味着，FFT和显示处理能够在下一段数据捕获前完成。例如，采样频率为8 kHz(由采样定理知，信号带宽的上限为4 kHz)，一段长度内有256个采样点，FFT和显示处理时间必须小于32 ms。

从英文字“rain”中字母a的实例表明：语音信号有周期的时域波形，如图2(a)所示；它的频谱类似于一串有间隔的谐波，如图2(b)所示。同样，字“storm”中的字母s的实例表明：摩擦音时域信号为噪声，如图2(c)所示，它的频谱如图2(d)所示。这个频谱证明对声音的2个主要源都存在共振峰频率的影响。

在图3中，图的下半部分是相应的语谱图，语音能量由颜色的深浅来表示，颜色越深，语音能量越强。

由图3可知，语音样例“他去无锡市，我到黑龙江”的每一个汉字的发音对应一组频谱，有其基音和谐波。基音和谐波的宽度不等说明有共振峰频率的影响。从短时稳定的频谱存在说明语音信号存在短期相关性，即尽管模拟声道的数字滤波器参数是随时间改变的，但是

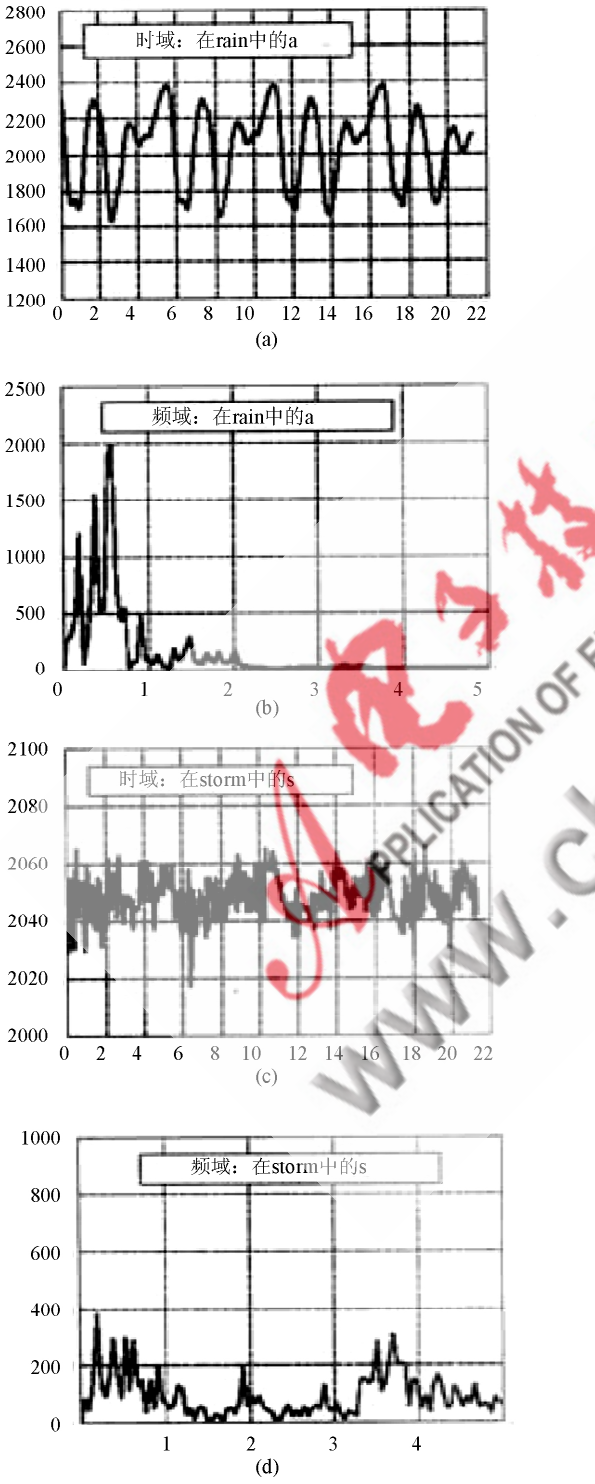

图2 语音信号的频域表示之一

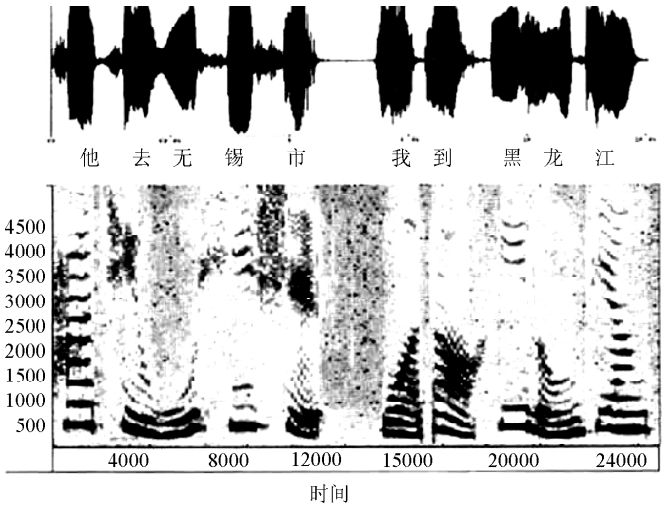

图3 语音信号的频域表示之二

在很短的时间(如几毫秒)内，由于存在确定的周期性频谱，因而可以认为，在该段时间内，数字滤波器参数不随时间而变化。可以使用线性预测方法，即一个语音采样值能够由前面若干个采样值的组合逼近，故称为线性预测。因此，每一个汉字语音对应一组线性预测系数，也就是对应一组确定的声道数字滤波器系数。

2 语音信号的产生模型

根据上面的分析，可以用近期所有语音合成和识别技术采用的人类语音模型来模拟语音信号的产生，如图4所示。

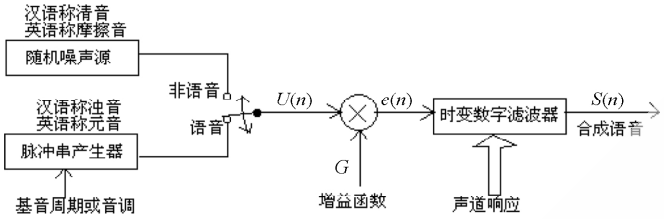

图4 语音信号产生模型

用随机噪声发生器产生噪声源模拟摩擦音(汉语称清音)，利用音调或称基音周期控制脉冲串产生器模拟元音(汉语称浊音)。用增益函数表示声音振幅。模拟声道的数字滤波器是一个线性时变滤波器。

3 线性预测编码(LPC)

线性预测编码LPC(Line Predictive Coding)方法在语音信号产生模型应用中是至关重要的，下面给出它的物理概念和方法。采样后的语音是离散信号，可以利用Z变换进行分析计算。设声道滤波器为一个全极点滤波器，其传递函数为$V(z)$，则输出信号为：

$$S(z)=E(z)\times V(z)=G\times E(z)/A(z) \tag{1}$$

式中，$E(z)$为声道滤波器的激励$e(n)$的Z变换；$A(z)$为声道滤波器的逆滤波器，是全零点滤波器；G为增益函数，表示声音振幅的一个参数；$S(z)$为合成的语音。在已知激励和滤波器参数后，可得到合成语音，故(1)式称为合成模型。由(1)式可得：

$$E(z)=S(z)\times A(z) \tag{2}$$

(2)式为(1)式的逆运算，故称为语音分析模型。

若逆滤波器为$A(z)$，输入语音信号为$S(z)$，则输出即为激励信号$E(z)$。然而，$A(z)$是未知的，需要使用线性预测的方法求得。

因为$A(z)$是全零点滤波器，其结构如图5所示。通过证明可得：

$$A(z)=1-\sum_{i=1}^{M}a_i z^i \tag{3}$$

即$A(z)$是由M节滤波器组成，式中i是滤波器的阶数，a_i是逆滤波器的系数，有待确定。把(3)式代入(2)式，并将Z变换的式子转换为离散值来写，则有：

$$e(n)=S(n)-\sum_{i=1}^{M}a_i S(n-i) \tag{4}$$

(4)式说明对样本序列值$S(n)$，n时刻序列值由它前面M个样本线性预测得到。即：

$$\hat{S}(n)-\sum_{i=1}^{M}a_i S(n-i) \tag{5}$$

同时表示，激励信号$e(n)$是语音信号$S(n)$与预测信号$\hat{S}(n)$之差，称为预测误差。(5)式可写为Z变换形式：

$$\hat{S}(n)=F(z)\times S(z) \quad 即 \quad F(z)=\sum_{i=1}^{M}a_i z^i \tag{6}$$

式中，$F(z)$为预测滤波器值，若输入$A(z)$，输出即为预测值$\hat{S}(n)$，见图5。

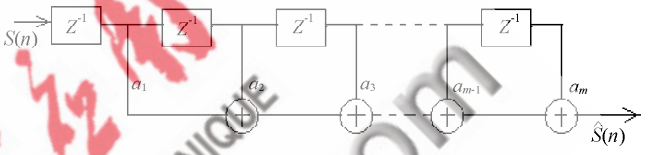

图5 线性预测编码系数计算

可见，这里存在2个滤波器，1个是预测滤波器$F(z)$，可用来求预测值；另一个为逆滤波器，它等于$1-F(z)$，可用来从激励信号求出重建的语音信号。使用这2个滤波器关键是求系数a_i。利用公式(4)，预测误差$e(n)$越小，预测值$\hat{S}(n)$越接近信号值$S(n)$。可采用$e(n)$的最小均方误差准则来确定a_i的系数。若$S(n)$已知，在短时间范围内(如20 ms)，在8 kHz采样频率下就有160个$S(n)$样本点，利用它来训练预测滤波器$A(z)$，系数a_i就可以确定。系数a_i是时变的，但在短的时限内是不变的。因此，在线性预测算法中，系数a_i的计算每帧都要进行1次，当前帧系数a_i计算值作为下一次计算时用。

4 语音产生模型的应用

语音产生模型说明一个短时的语音信号可以用3个参数来定义：(1)从周期性波和随机噪声中选择1个作为激发态；(2)如果使用周期性波，必须选择1个频率作为基音；(3)模拟声道响应所使用的数字滤波器系数。

4.1 语音产生模型在语音合成技术中的应用

早期产品中应用到的连续语音合成技术，是借助于大约以每秒40次速度修改上述的短时语音信号的3个参数来实现的。如适合儿童学习的“说和拼音机”。由于它仅仅采用26个英文字母作为音库，因而这种语音合成的声音质量不高，声音非常机械。

此后，用汉字语音作为库，用波形拼接方法进行语音合成，效果有所改进，但是库的存储量太大。解决的方案是，使用语音分析方法，即利用语音产生模型概念，把一个语音信号分解成下列特性参数：线性预测系

(下转第38页)

[2] CANNY J. A computational approach to edge detection [J]. IEEE Trans Pattern Analysis and Machine Intelligence,1986(12): 679-697.
[3] 牛连强，陈彦军，刘守仁，等.结焦图像的处理与识别方法研究[J].计算机工程与设计，2005，26(6): 1494-1496.
[4] GONZALEZ R C, WOODS R E. 数字图像处理[M]. 阮秋琦，译.(第2版). 北京：电子工业出版社，2007.
[5] 边肇祺.模式识别[M].(第2版).北京：清华大学出版社,2000.
[6] SHAPIRO J M. Embedded image coding using zerotrees of wavelet coefficients. IEEE Transactions on Signal Processing, 1993, 41 (12): 3445-3462
[7] MALLAT S, ZHONG S. Characterization of signals form multiscale edges. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1992, 14 (7): 710-732.
[8] SCHMEELK J. Wavelet transforms and edge detectors on digital images[J].Mathematical and Computer Modeling, 2005, 41(13): 1469-1478.
[9] HERIC D, ZAZULA D. Combined edge detection using wavelet transform and signal registration [J]. Image and Vision Computing, 2007, 25(5): 652-662.

（收稿日期：2009-01-05）

（上接第30页）

数(取 10 个)、基音周期范围、基音周期数目(基音持续时间)和清音存在时间等。根据 ITU-T G.729 语音编码方法，一帧语音信号特征参数仅需 80 bit，即 80 个 16 bit 样本压缩为 80 bit，缩小 16 倍。到合成需要该音时，再利用语音产生模型由所存的特征参数实时转换为语音。

4.2 语音产生模型在语音识别技术中的应用

与机器进行语音交流，让机器明白你说什么，这是人们长期以来梦寐以求的事情。语音识别技术就是让机器通过识别和理解过程把语音信号转变为相应的文本或命令的技术。其原理是：由于每一个短时语音信号包含一串语音特性参数，不同的汉字音有不同的特征参数，所以利用特征参数的差别来识别不同的汉字音。

近 20 年来，语音识别技术取得显著进步，开始从实验室走向市场。预计未来 10 年内，语音识别技术将进入工业、家电、通信、汽车电子、医疗、家庭服务、消费电子产品等各个领域。

参考文献

[1] 拉宾纳 L R，谢弗 R W. 语音信号数字处理[M]. 北京：科学出版社，1983.
[2] 戴逸民，梁晓雯，裴小平. 基于DSP的现代电子系统设计[M]. 北京：电子工业出版社，2002.
[3] 奥本海姆. 信号与系统[M]. 刘树棠，译 . 西安：西安交通大学出版社，1998.
[4] 何苏勤，王忠勇.TMS320C2000系列DSP原理及应用技术[M]. 北京：电子工业出版社，2003.

（收稿日期:2008-12-30）

（上接第33页）

Lena(128 × 128)图像 PSNR 分别高出 1.19dB 和 2.90dB，与传统小波变换方法相比，Lena(256 × 256)和 Lena(128 × 128)图像 PSNR 分别高出 0.07DB 和 0.02DB。

本文针对图像超分辨率过程中传统的插值方法误差较大，处理后的边缘细节及纹理不够理想，有时会出现方块效应或边缘退化的缺点，利用二元树复小波变换与边缘插值方法相结合放大图像，然后对放大图像的高频系数进行修改，最后通过小波逆变换得到重构后的图像。实验结果表明，与传统方法相比，本文算法可以明显提高图像的清晰度，既保留了丰富的细节,又抑制了边缘震铃效应，同时 PSNR 也有所提高。将二元树复小波变换与边缘插值方法相结合应用到图像超分辨率重建中来，具有一定的理论研究价值和实际应用价值。

参考文献

[1] ATES H F, ORCHARD M T. Image interpolation using wavelet-based contour estimation[J]. IEEE Acoustics, Speech, and Signal Processing, 2003, 3(4): 109-112.
[2] TAPIA D F, THOMAS T G, MURGULA M C. Wavelet-based interpolation algorithm for MRI images[J]. Journal of Alloys and Compounds, 2004, 369 (2004): 239-243.
[3] LI Xin, MICHAEL T. Newedge-directed interpolation[J]. IEEE Trans on Image Processing, 2001, 10(10): 1521-1527.
[4] NGUYEN N, MILANFAR P. A wavelet-based interpolation-restoration method for superresolution[J]. Circuits Systems Signal Process, 2000, 4(19): 321-338.
[5] YU Len Huang. Wavelet-based image interpolation using multilayer [J]. Neural Computing and Applications, 2005, 1(14): 1-10.
[6] KINGSBURY N. A dual-tree complex wavelet transform with improved orthogonality and symmetry properties[J]. IEEE Image processing, 2000, 1(2):375-378.

(收稿日期：2009-01-21)