

用逆滤波基音同步分析提取声门波的方法*

王成友 郑林华 黄小严

(国防科技大学电子技术系 长沙 410073)

摘要 根据语音研究的要求, 提出一种运用逆滤波基音同步分析提取声门波的方法, 此方法可以比较准确地提取声门波。文中讨论了这种方法的原理、实现及其结果。

关键词 基音同步, 声门波

分类号 TN912

大量的语音研究已表明, 声门波的波形对合成语音质量有着决定性的影响^[1]。为了合成高质量的语音, 对声门波作出一种比较精确的估计是非常必要的。本文介绍一种运用逆滤波基音同步分析提取声门波的方法, 并讨论这种方法的实现及其结果。

1 基音同步分析提取声门波的原理

根据线性预测理论^[2], 语音中浊音产生的线性预测模型如图1。

图中: $G(Z)$, $V(Z)$, $R(Z)$ 分别为声门波声道辐射的 Z 域表达式。

$e(n) \leftrightarrow E(Z)$ 为周期脉冲串, 周期的大小等于基音周期。

$U_c(n) \leftrightarrow U_c(Z)$,

$U_L(n) \leftrightarrow U_L(Z)$ 为图中相应位置的

时域和 Z 域表达式。假设声道为全极点模型结构, 则

$$V(Z) = 1 / (1 + \sum_{i=1}^K C_i Z^{-i}) \quad (1)$$

$$R(Z) = 1 - Z^{-1} \quad (2)$$

$$U_c(Z) = S(Z) / (V(Z)R(Z)) \quad (3)$$

$R(Z)$ 一般认为 $1 - Z^{-1}$ 足以表达, 这样, 只要求出 $V(Z)$, 声门波 $U_c(Z)$ 就能求出。

现设 $q(n) \leftrightarrow Q(Z)$ 为

$$Q(Z) = U_c(Z)R(Z) \quad (4)$$

$$q(n) = U_c(n) * r(n)$$

此时, 模型框图1可以转变图2。

声门关闭期稳定时, 理想的 $s(n)$, $U_c(n)$, $q(n)$ 波形如图3。

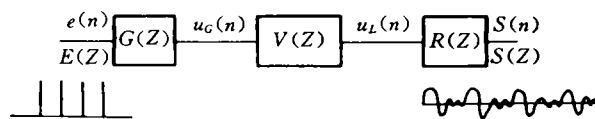


图1 浊音产生模型

* 1994年2月28日收稿

其中 $n=L_c, n=L_0$ 分别为声门关闭、开启点, 因为已设 $R(Z) = 1-Z^{-1}$, 所以, 对应于 $U_G(n) = 0, L_c \leq n \leq L_0$

有 $q(n) = 0, L_{c+1} \leq n < L_0$

根据假设

$$V(Z) = 1/(1 + \sum_{i=1}^K C_i Z^{-i})$$

$$s(n) = \sum_{i=1}^K C_i s(n-i) + q(n) \quad (5)$$

因为 $(L_c + 1) \leq n < L_0, q(n) = 0$

所以 $s(n) = - \sum_{i=1}^K C_i s(n-i), (L_c + 1) \leq n < L_0 \quad (6)$

若设 M 阶滤波器冲激响应为

$$A(Z) = \sum_{i=0}^M a_i Z^{-i}, \quad (a_0 = 1, K \leq M)$$

$s(n)$ 通过这一滤波器, 输出即为

$$e(n) = s(n) + \sum_{i=1}^K a_i s(n-i)$$

在 $(L_c + 1) \leq n < L_0$ 内, 将式 (6) 代入上式:

$$e(n) = \sum_{i=1}^K (a_i - c_i) s(n-i)$$

因此如果

$$\begin{aligned} a_i &= c_i, & i &= 1, \dots, K \\ a_i &= 0, & i &= K+1, \dots, M \end{aligned} \quad (7)$$

则 $e(n) = 0, L_c + 1 \leq n < L_0$

如果对 $s(n-M)$ 到 $s(n-N-M-1)$ 段进行 M 阶协方差分析, 在这段内整个误差平方和为:

$$\alpha_M(n) = \sum_{j=1}^{n+N-M-1} e^2(j) \quad (8)$$

因为协方差分析是按最小均方误差的准则 (即 $\alpha_M(n)$ 为最小), 提取出分析滤波系数 $\{a_i\}$ 的, 因此理论上如果 $n \geq L_c + 1, n + N - M < L_0$, 那么 $e(n)$ 在这段为 0, 这也就是提取出的 $\{a_i\}$ 满足式 (7), $\alpha_M(n) = 0$.

一旦 $\{a_i\}$ 求出, (下面字母上加 $\hat{\cdot}$ 为估计出的值)

$$\hat{Q}(Z) = S(Z)/\hat{V}(Z) \quad (9)$$

$$\hat{U}_G(Z) = \hat{Q}(Z)/(1 - Z^{-1}) \quad (10)$$

即: 声门波便求出。

在实际提取声门波中, 将涉及三个问题:

1. 分析窗的选择

在图 3 中有两种窗被显示, 一种窗第一点为 $L_c - M + 1$, 最后一点为 L_0 。设 $K \leq M$, 那么 (6) 式是成立的, $\alpha_M(n)$ 将会达到 0 值。另外一种, 窗第一点为 $L_c + 1$, 最后一点为 L_0 。此时误差序列 $e(j)$ 在最末 $(p+1)$ 点不准确, 因为它不在 $[L_c + 1, L_0 - 1]$ 内, 所以 $\alpha_M(n)$ 不能减少至 0。因此窗的选择一般应用第一种。

2. 误差平方和

因为模型与实际语音信号不是绝对一致时, 一般 $\alpha_M(n)$ 误差不大可能为 0, 因此, 一般用下面相

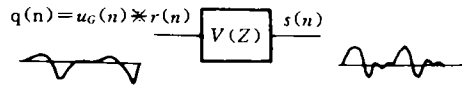


图 2 浊音产生模型变形

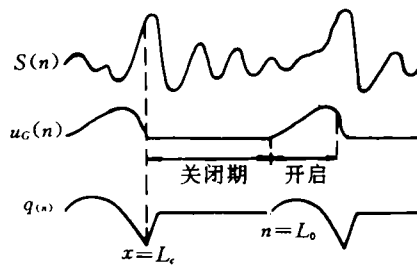


图 3 理想 $s(n)$ 、 $u_G(n)$ 、 $q(n)$ 波形

对误差度量：

$$\eta(n) = \alpha_M(n) / \alpha_0(n)$$

其中： $\alpha_0(n)$ 为输入相应段的语音能量。

3. 分析阶数

阶数的选择应该 $K \leq M_0$ 。

2 逆滤波基音同步分析提取声门波的实现

根据以上分析，作者制定并实现这一过程，其框图如图 4。

图中，线性相位高通滤波的目的在于去除一些由于记录设备引起的低频成分。语音以 16kb/s 采样率输入。

下面以提取元音 [a] 为例，来详细说明提取过程，元音 [a] 时域波形的一段示于图 5 (a)。

(1) 从图 5 (a) 中，取出一帧，帧长为 30 个点。一般在满足协方差分析条件下，应尽可能小于等于关闭期的长度，这样，当分析帧进入关闭期后协方差分析的结果可以确切只含声道特征。对取出的一帧数据进 16 阶行协方差分析得到 $A(Z)$ ，按前面讨论的方法计算 $\eta(n)$ 。

(2) 帧移动一点，再取一帧信号，做 (1) 中的计算，以此规律，完成一个基音周期。此时，计算出的 $\eta(n)$ 示于图 5 (b)。

(3) 定义 $\eta(n)$ 中变化最快点的值为 η_n ，在 $\eta(n) < \eta_n$ 内，找出 $\eta(n)$ 的最小值点，对应此点的线性预测系统便最接近于声道。

(4) 用 (3) 确定的预测滤波器对 $S(n)$ 作用，求出 $\hat{q}(n)$ ，如图 5 (c)。

(5) 对 $\hat{q}(n)$ 逆辐射求出 $u_G(n)$ 如图 5 (d)。

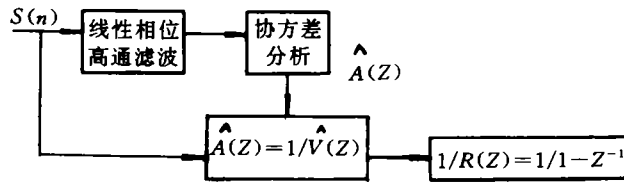


图 4 逆滤波基音同步分析提取声门波框图

3 实验讨论

(1) 线性预测阶数对提取声门波的影响

线性预测阶数按前面分析应该大于等于声道本身固有的阶数，但阶数取大，由协方差分析条件，帧长要求大，这样取出的帧可能超出声门关闭期，使得提取出的声门波不准。因而，实际中需要反复调整。在我们系统中取 16 比较合适。

(2) 预加重系数影响

图 4 中，线性相位高通滤波，实现时采用预加重，预加重系数直接控制着滤波器的下限频率，因此，它直接影响到声门波的波形。

(3) 帧长的影响

如果帧长取得较小，则 $\eta(n)$ 值离散性较大，以致于幅度变化到最大值，还不能正确指示声门关闭点。如果帧长取得较大，超过了声门关闭期，则提取的线性预测系数不准确，因此，帧长要取得合适(作者试验帧长，

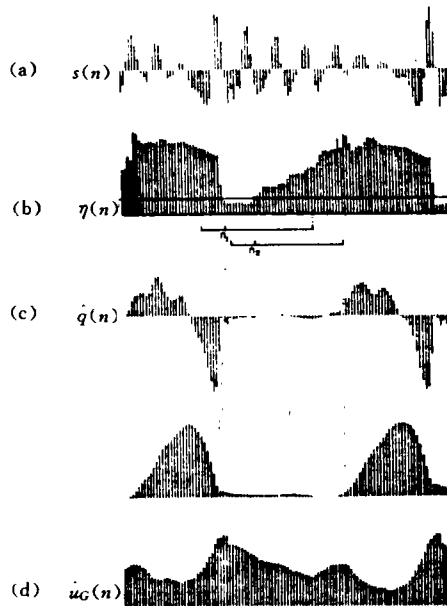


图 5 音 [a] 的声门波提取

一般稍大于 $2P$ 即可)。

一般试验中, (1) (2) (3) 参数需配合调整, 以提得实际语音声门波。

(4) 记录设备的低频影响

记录设备的低频影响使提出的声门波如图 6 (b), 因为此时对低频进行了积分, 出现整个信号上升或下降。

(5) 记录设备的相位影响:

记录设备的相位影响, 会导致声门波变形, 如图 6 (a)。

(6) 利用提取声门波进行 LPC 合成。作者利用前面提取的声门波进行 LPC 合成试验。其实验过程如下:

a. 先将音 [a] 的每个基音周期提取线性预测系数 $\{a_i\}$, 在稳定后的基音周期提取声门波, 并且利用中心削波法提取出基音周期, 同时计算出每个基音周期的幅度。

b. 利用 a 中提取出的参数, 按每个基音周期为一单元进行 LPC 合成。其框



图如右。

c. 将 b 中合成出的语音放音听辨。它与用单脉冲激励, 滤波器激励合成语音相比, 自然度高出许多。与 Rosenberg 提出的声门波激励相比要好。

此实验结果证实这种方法提取的声门波和线性预测系数比较准确。

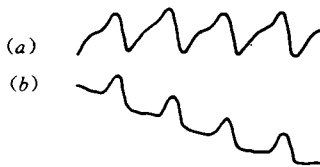


图 6 设备低频和相位的影响

4 结 论

这种方法不仅提取出的声门波非常准确, 相应提取出的线性预测参数也相当可靠, 它的结果可以运用到其它的语音处理中, 比如语音识别, 语音合成。

参 考 文 献

- 1 J Gauffin, J Sundberg. Spectral Correlates of Glottal Voice Source Wqveform Characteristics, J. Speech Hearing Res, 1989, 32
- 2 L R Rabiner 等著; 朱雪龙译. 语音信号数字处理, 北京: 科学技术出版社, 1983
- 3 王成友. 汉语语音分析与合成系统. 国防科技大学电子技术系硕士论文, 1993

An Inverse Filtering and Pitch Synchronous Analysing Method of Estimating Glottal Waveform of Voiced Speech

Wang Chengyou Zheng Linhua Huang Xiaoyan
(Dpartment of Electronic Technolog)

Abstract

In this paper, an inverse filtering and pitch synchronous analysing method of estimating glottal waveform of voiced speech is prsented. in accordance with the requiremerts of speech study. Reliable glottal waveform of voiced speech, can be provided by the method This paper describes and discusses the principles and realization of the method.

Key words pitch synchronous; glottal waveform