

文章编号 :1001 - 2486(2005)01 - 0089 - 04

基于 LPC 残差用频域法来实现基音周期变换*

李 波^{1,2}, 王成友¹, 杨 聪¹, 蔡宣平¹, 张尔扬¹

(1. 国防科技大学 电子科学与工程学院, 湖南 长沙 410073; 2. 空军工程大学 电讯工程学院, 陕西 西安 710077)

摘 要 基音周期变换是文—语转换和语音转换的重要内容。在语音转换中有很多作者主张基于 LPC 分析来实现语音转换, 通过直接或间接调整 LPC 系数来实现频谱包络的转换, 而基音周期的转换则通过对 LPC 激励谱或 LPC 残差的处理来实现。提出对 LPC 残差用频域法来进行处理, 得到改变了基音周期的 LPC 残差, 从而实现语音的基音周期的变换。这种方法也可以用于文—语转换的基音周期改变中。

关键词 文—语转换; 语音转换; 基音周期变换

中图分类号: TP391 文献标识码: A

Achieving Pitch Modification Using Frequency Method Based on LPC Residual

LI Bo^{1,2}, WANG Cheng-you¹, YANG Cong¹, CAI Xuan-ping¹, ZHANG Er-yang¹

(1. College of Electronic Science and Engineering, National Univ. of Defense Technology, Changsha 410073, China;

2. The Telecommunication Engineering Institute, Air Force Engineering University, Xi'an 710077, China)

Abstract Pitch modification is an important part in text-to-speech synthesis and voice conversion. In voice conversion, many authors in favor of realizing the voice conversion by the method based on LPC analysis, with which the spectrum envelope is converted by modifying the LPC coefficients directly or indirectly and the pitch modification is realized by processing the LPC excitation spectrum or LPC residual. Frequency method is proposed to process the LPC residual, then the pitch modified LPC residual is got. This method can also be used in the pitch modification of the text-to-speech synthesis.

Key words text-to-speech synthesis; voice conversion; pitch modification

基音周期变换(pitch modification)就是要保持语音频谱包络不变, 而只改变语音的基音周期^[2]。改变基音周期是文—语转换和语音转换(voice conversion)的重要内容。在文—语转换中, 要改变基本单元的基音周期以使其与包含多个基本单元的语音段的超音段特征相匹配。在语音转换中^[1,3], 基音周期是语音的一个重要的个性特征, 需要将源语音的基音周期转换为目标语音的基音周期。改变基音周期是实现高质量按规则语音合成难度最大的问题之一, 一般情况下当基音周期改变较大时, 就会引起合成语音质量的下降^[4]。源语音经过语音转换所得到的转换语音质量往往有所下降, 这其中的很大一部分是由基音周期变换引起的。

很多学者主张基于 LPC 分析来实现语音转换。即通过直接^[5]或者间接^[1,3]地调整 LPC 系数来实现语音的频谱包络的转换, 而基音周期的改变则通过对 LPC 的激励谱^[1]或者 LPC 残差^[2]的处理来实现。在文献 [2] 中提出首先确定重要激励(significant excitation), 再通过垫加加权的随机噪音的方法来实现基音周期的增大, 通过截取的方法来实现基音周期的减小, 从其实验的波形来看效果不是很理想。本文提出用频域法来处理 LPC 残差, 实现基音周期的改变。首先对 LPC 残差利用文献 [2] 中的方法来确定重要激励, 然后以重要激励为中心点以汉明窗取语音段进行频谱分析; 再基于类谐波语音模型, 通过零值插值或者零值删除法来实现基音周期的改变, 最后合成出改变了基音周期的 LPC 残差信号。实验结果表明, 本文所提出的方法效果良好。

* 收稿日期 2004 - 09 - 12

作者简介: 李波(1974—) 男, 博士生。

1 频域分析

在本文中所用的语音信号采样频率为 11 025Hz,量化精度为 16bit。假定已经通过 LPC 分析得到 LPC 残差信号 $R(n)$,本文主要研究用频域法处理 $R(n)$,改变基音周期。

对 LPC 残差采用群延迟的方法^[2]来确定重要激励,以各个重要激励为中心点用 401 点长的汉明窗截取 $R(n)$ 进行 1024 点的 DFT 变换,得实部 $DFT_{\cos}[k]$ 和虚部 $DFT_{\sin}[k]$,并求幅度谱 $S(k)$ ($k=0,1,\dots,511$)。采用 401 点的固定帧长,主要是因为长的帧长有利于保持好的幅度谱的谐波形状。

2 伪谐波语音模型表示

Moulines^[6]提出用抽取—插值法和压缩—扩展法两种方法来实现基音频率的改变。对于抽取—插值法,要求很高的基音频率的估计和频率精度的估计^[9],常常会引起很大的失真,笔者也曾基于本文下面提出的类谐波模型采用抽取—插值法来进行基音频率改变的实验,效果不好,因此对该方法不做过多探讨。压缩—扩展法相对来说是一种实现基音周期改变比较成功的方法,但是这种方法会使频谱的谐波结构与基音频率同变化率地扩展和压缩,这是不符合实际的。实际上当分析帧长足够长时,语音频谱的谐波结构稳定,浊音信号的频谱是一种梳状谱,每个谐波频率处有一个不太宽的梳齿,相邻两个梳齿之间常有较大间隙,几乎没有能量^[7],基音频率不同,主要表现在梳齿间间隙的大小不同,而梳齿相对来说比较稳定,变化不大。正弦模型则是一个频率谐波只取一个频率点,这样其对频谱的表示就不够准确。考虑两个方面的情况,即保持幅度谱谐波形状的稳定又尽量准确地表示原频谱,提出对每个谐波采用多个频点进行表示,称之为类谐波语音模型。基于这个模型,通过对谐波间的零值进行删除或者复制的方法来实现基音频率的改变。

本文对每个谐波取以谐波点为中心的 7 个频率点,这是因为本文实验所用语音的采样频率为 11 025Hz,利用 1024 点 DFT 变换进行频谱分析,而人的基音频率范围为 80~500Hz^[7],其所对应的 DFT 频率点数为 7.4~46.5,这样选择 7 频点的伪谐波语音模型在进行后面的基音频率变换时不会引起谐波的重叠。

由于基音频率求解的精度很难达到足够高,为提高谐波点求解的鲁棒性及其谐波点位置的准确性,对每个谐波区间采用 Stylianou^[8]提出的方法进行清浊判断,如果为清,则以该区间中间点作为该区间的谐波点;如果为浊,则以该区间的谱幅度的极大值点为该谐波区间的谐波点。然后以该区间的谐波点为基准点,根据所求解的基音频率^[10]确定下一个谐波区间,以相同的方法进行搜索。以上面所确定的各个谐波点位置为中心,保持 $DFT_{\cos}[k]$ 和 $DFT_{\sin}[k]$ 取 7 点不变,置其它频点为零值,这样就得到了 7 点的伪谐波语音模型表示。

3 进行基音周期变换

在上一节中,得到实部 $DFT_{\cos}[k]$ 和虚部 $DFT_{\sin}[k]$ 的 7 点伪谐波表示。对于基音频率的变换可以通过保持 7 点伪谐波值不变,对其中间的零值进行插值和删除来实现。这种方法将与 Moulines^[6]的压缩—扩展法存在一个共同的问题,那就是当基音频率增大时,频谱将超过 $F_s/2$ (F_s 为采样频率);而当基音频率变小时,频谱将会压缩,在高频处出现空白区域。对这个问题的处理方法是:当频谱扩展时,将超过 $F_s/2$ 的部分去掉,只截取 $0 \sim F_s/2$ 部分;当频谱压缩时,复制低频部分来填充高频谱部分,得到基音频率转换后的 $DFT'_{\cos}[k]$ 和 $DFT'_{\sin}[k]$,实现方法如图 1 所示。

4 合成出基音频率转换后的 LPC 残差

对 $DFT'_{\cos}[k]$ 和 $DFT'_{\sin}[k]$ 进行 IDFT 变换得到 $s'(n)$,将 $s'(n)$ 进行重要激励检测,以离中心点最近的重要激励为中心截取变换了基音周期长的激励段,进行串接,并进行适当的时长调整,即得变换了基音周期的激励,如图 2 所示。

5 实验

以汉语男声发音“a”为例用上述方法进行实验。采样频率为 11 025Hz,量化精度为 16bit,将基音周

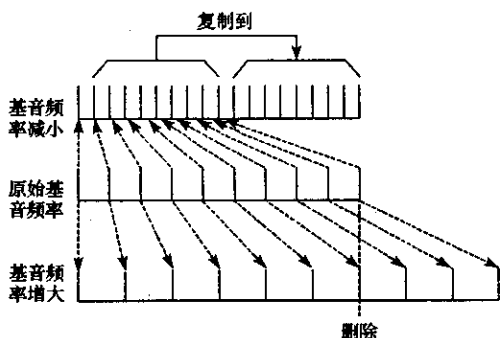


图1 改变基音频率

Fig. 1 Modify the fundamental frequency

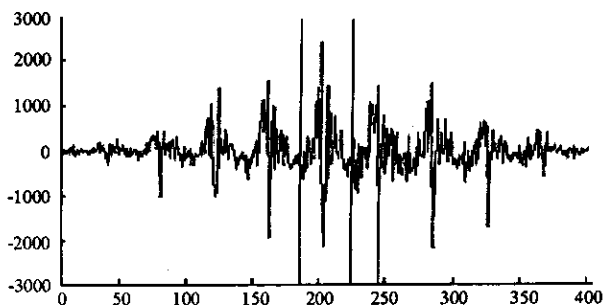


图2 截取一个周期的合成LPC残差

Fig. 2 Cut a pitch period of synthesized LPC residual

期缩短 1.8 倍。用 Durbin 法求解 32 阶 LPC 系数 a_k ($k = 1, 2, \dots, 32$), 根据公式 $R(n) = X(n) - \sum_{k=1}^{32} a_k X(n-k)$ 得到 LPC 残差 $R(n)$ 。利用本文提出的方法处理 LPC 残差 $R(n)$ 。图 3 和图 4 分别为基音周期变换前后的情况。

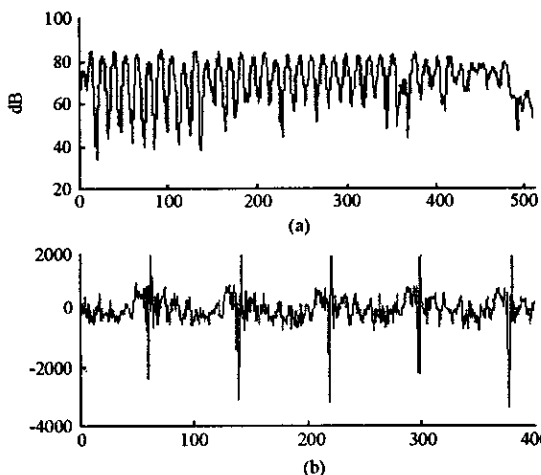
图3 (a) 基音周期变换前的 LPC 残差幅度谱;
(b) 基音周期变换前的 LPC 残差波形图

Fig. 3 (a) The spectrum of LPC residual before pitch modification;

(b) The waveform of LPC residual before pitch modification

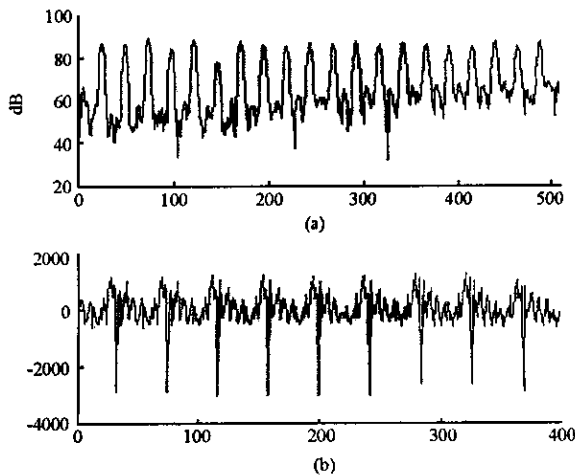
图4 (a) 基音周期变换后的 LPC 残差幅度谱;
(b) 基音周期变换后的 LPC 残差波形图

Fig. 4 (a) The spectrum of LPC residual after pitch modification;

(b) The waveform of LPC residual after pitch modification

为验证最终的改变基音周期的效果, 又将所得到的 LPC 残差用原来的 LPC 系数进行合成, 图 5 和图 6 分别为基音周期变换前后的情况。从图中可以看出基音周期变换是有效的。

6 结束语

本文提出用频域法处理 LPC 残差来进行基音周期变换。用伪谐波模型来表示语音的频谱, 且基于该模型用零插值法和零删除法在频域实现基音频率的改变, 不仅提高对频谱的表示精度, 而且可以方便有效地实现基音周期的变换。实验结果表明本文提出的方法是有效的。

参考文献:

- [1] Arslan L. M. Speaker Transformation Algorithm Using Segmental Codebooks (STASC) [J]. Speech Communication, 1999, 28(3): 211 - 226.
- [2] Rao K. S., Yegnanarayana B. Prosodic Manipulation Using Instants of Significant Excitation [A]. Int. Conf. Acoust., Speech Signal Processing [C], Apr. 2003.

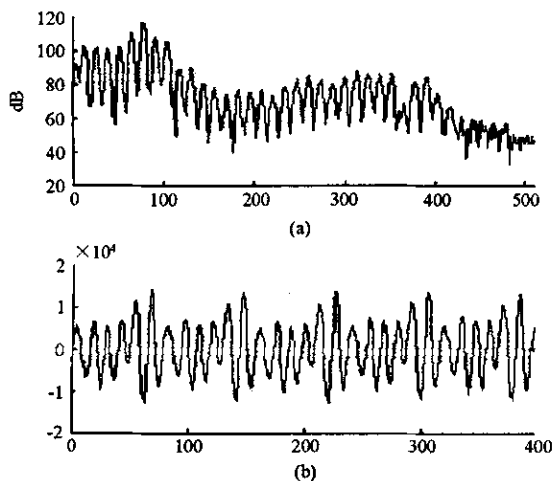


图5 (a) 基音周期变换前语音幅度谱;
(b) 基音周期变换前语音波形图

Fig. 5 (a) The speech spectrum before pitch modification;
(b) The speech waveform before pitch modification

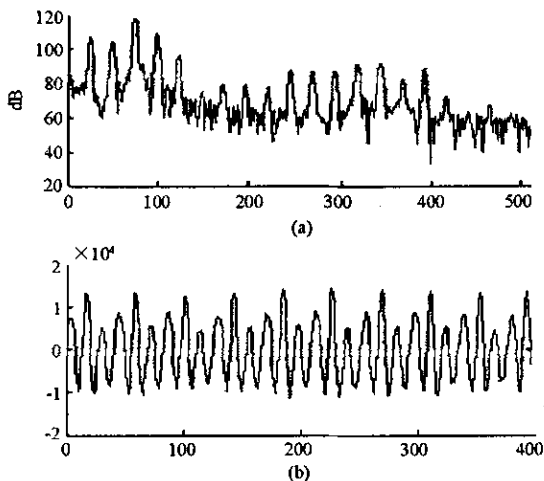


图6 (a) 基音周期变换后语音幅度谱;
(b) 基音周期变换后语音波形图

Fig. 6 (a) The speech spectrum after pitch modification;
(b) The speech waveform after pitch modification

- [3] Kain A B. High Resolution Voice Transformation[D]. Ph. D. Dissertation , OGI School of Science and Engineering at Oregon Health and Science University , October 2001.
- [4] Tanaka K , Abe M. A New Fundamental Frequency Modification Algorithm with Transformation of Spectrum Envelope According to F0[A]. IEEE ICASSP[C] , 1977 951 - 954.
- [5] Mizume H , Abe M. Voice Conversion Algorithm Based on Piecewise Linear Conversion Rules for Formant Frequencies and Spectrum Tilt [J]. Speech Communication , 1995 , 2(16) : 153 - 164.
- [6] Moulines E , Charpentier F. Pitch-synchronous Wave Form Processing Techniques for Text to Speech Synthesis Using Diphones[J]. Speech Communication , 1990 , 9 453 - 467.
- [7] 易克初 , 田斌 , 付强. 语音信号处理[M]. 北京 : 国防工业出版社 , 2000.
- [8] Stylianou Y. A Pitch and Maximum Voiced Frequency Estimation Technique Adapted to Harmonic Models of Speech[A]. IEEE Nordic Signal Processing Symposium[C] , Sept. 1996.
- [9] Oytun T K. Objective Tests on Spectral Envelope Estimation Methods for FD-PSOLA Pitch Scale Modification[J]. Sestek Inc. R&D. Dept. 2001.
- [10] Li B , Li Y Y , Wang C Y , et al. A New Efficient Pitch-tracking Algorithm[A]. Proceedings of the 2003 IEEE RISSP[C] , Changsha , China , 2003 : 1102 - 1107.

