

基于时频分析方法 LPFT 的语音处理 GUI 系统^①

李秀梅, 吕真, 姬宇佳

(杭州师范大学 信息科学与工程学院, 杭州 310036)

摘要: 由于语音的非平稳性及时变性, 时频分析方法是处理语音信号的重要工具. 然而线性短时傅里叶变换的时频聚集性较差, 而双线性维格纳变换在处理多分量信号时会受到交叉项的干扰. 为了克服以上两种时频分析方法的缺点, 利用短时傅里叶变换的扩展形式即局部多项式傅里叶变换 LPFT 来处理语音信号, 建立了基于 LPFT 的语音处理 GUI 系统, 实现了在时域、频域和时频域对语音的分析和对比. 并给出语音处理的例子, 验证了 LPFT 方法与其它方法相比所具有的优势. 该系统简明直观, 是语音处理的较好的平台.

关键词: 时频分析; 短时傅里叶变换; 维格纳变换; 局部多项式傅里叶变换; 语音处理; 图形用户界面系统

Speech Analysis GUI System Based on the Local Polynomial Fourier Transform

LI Xiu-Mei, LV Zhen, JI Yu-Jia

(School of Information Science and Engineering, Hangzhou Normal University, Hangzhou 310036, China)

Abstract: Since the speech is nonstationary and time-varying, the time-frequency analysis is a useful tool to process the speech. However, the linear short-time Fourier transform (STFT) suffers from low resolution in the time-frequency domain, and the bilinear Wigner-Ville distribution (WVD) suffers from the cross terms when processing multi-component signals. Therefore, in this paper, the generalization of the STFT, that is, the local polynomial Fourier transform (LPFT) is employed to process the speech, and a speech analysis GUI system based on the LPFT is proposed. In the proposed GUI system, the analysis and comparisons for speech are realized in the time domain, the frequency domain, and the time-frequency domain. It is shown by the example that the LPFT is more suitable to process the speech. The proposed system is user-friendly, and is a good platform to process the speech.

Key words: time-frequency analysis; short-time Fourier transform(STFT); Wigner-Ville distribution(WVD); local polynomial Fourier transform(LPFT); speech processing; GUI system

语音是一种非平稳的时变信号, 是人类相互传递信息的最重要、最有效和最常用的手段, 在语音编码、语音合成、语音识别、语音增强和人机交互等语音处理中都需要对语音信号进行处理, 提取语音中包含的各种信息^[1]. 因此, 语音信号处理在语音信号应用领域中具有举足轻重的地位.

语音的时域分析和频域分析是语音信号处理的两种重要方法. 语音信号的时域分析可以分析和提取语音信号的时域参数, 但是却不能直观了解语音信号的频率特性. 而对于非平稳的语音信号, 传统的傅里叶

变换只能描述信号频谱的分布情况, 因此无法反映其频率随时间变化的情况. 采用时频分析方法则可以把语音信号变换为时间和频率的二维函数^[2,3], 准确描述语音信号频率随时间变化的关系, 进一步阐释语音信号中的有用信息.

然而, 由于语音信号的非线性及非平稳性特点, 传统的时频分析方法, 例如短时傅里叶变换 STFT(short-time Fourier transform), 提取语音特征的准确度有限, 从而极大限制了后期语音处理效果的提升; 维格纳变换 WVD(Wigner-Ville distribution)虽然可以提供

^① 基金项目:国家自然科学基金(61102164)

收稿时间:2012-07-08;收到修改稿时间:2012-08-08

较高的时频聚集性,但是它在处理语音这类多分量时变信号时却会受到交叉项的干扰.因此,必须采用更为有效的时频分析处理方法,保证语音信号特征提取的准确度和对噪声的有效抑制,从而提高语音信号处理的性能.局部多项式傅里叶变换 LPFT(local polynomial Fourier transform)是一种优化的时频分析方法^[4],适合处理非平稳的语音信号,可以为语音提供更精确的分析.

本文在分析 LPFT 与 STFT 和 WVD 相比所具有的优势后,利用 LPFT 对语音信号进行处理,并基于 LPFT 设计出语音处理 GUI 系统,作为语音分析的可视化平台.仿真结果的图形化结果以人机交互的动态方式呈现,增加了语音处理的直观性,验证了 LPFT 在语音信号处理方面的优越性.

1 时频分析方法 STFT, WVD, LPFT

在本节我们将对语音处理系统中将要用到的短时傅里叶变换 STFT,维格纳变换 WVD,和局部多项式傅里叶变换 LPFT 做简要介绍.

1.1 短时傅里叶变换 STFT

在语音处理的应用如语音识别,语音增强和语音合成中,准确描述频率随时间的变化至关重要.在一般的语音频率估计中,大都采用短时傅里叶变换对其进行处理^[1].短时傅里叶变换是线性时频分析方法中应用最广泛的一种,对于信号 $s(t)$ 而言,其短时傅里叶变换定义如下:

$$STFT(s; t, \omega) = \int_{-\infty}^{\infty} s(\tau) h(t - \tau) e^{-j\omega\tau} d\tau \quad (1)$$

其中 $h(t)$ 为用以截断信号的窗函数.短时傅里叶变换的平方被称为谱图 spectrogram.短时傅里叶变换易于运算,而且在处理多分量信号时不会受到交叉项的干扰.但是,如式(1)所示,短时傅里叶变换采用固定的时间窗对语音信号进行截断,并假设在每个窗内语音信号频率都是恒定的,因此无法为语音信号提供足够高的时频聚集性.而时频聚集性是衡量时频分析性能的重要指标,我们总是希望时频分布的能量以较高的时频聚集性集中在信号能量聚集的时频区域内.时频聚集度越高,越有利于我们对信号特征进行提取,从而为非平稳信号提供更为准确的分析.另外,短时傅里叶变换假设语音的频率在短时间段内是恒定的,由于语音的时变性,这一假设可能与实际不符从而导致错误的频率估计.

1.2 维格纳变换 WVD

作为非线性时频分析方法代表的维格纳变换虽然可以提供较高的时频聚集性,但是如式(2)所示,它在处理语音这类多分量时变信号时却会受到交叉项的干扰.

$$WVD(s; t, \omega) = \int_{-\infty}^{\infty} s(t + \tau/2) s(t - \tau/2) e^{-j\omega\tau} d\tau \quad (2)$$

为了抑制维格纳变换中交叉项的干扰, Cohen 类双线性变换中的伪维格纳变换和平滑伪维格纳变换等变换受到了人们的关注^[2,3],并应用于语音中的基音检测.但是这些变换在抑制交叉项干扰的同时,却会削弱信号的时频聚集性.因此,既能提供较高的信号时频聚集性,同时对多分量信号处理时又能不受交叉项干扰的时频分析方法,成为人们关注的重点.

1.3 局部多项式傅里叶变换 LPFT

作为短时傅里叶变换的扩展,LPFT 这一时频分析方法由俄罗斯科学家 V. Katkovnic 提出^[4],其形式如下

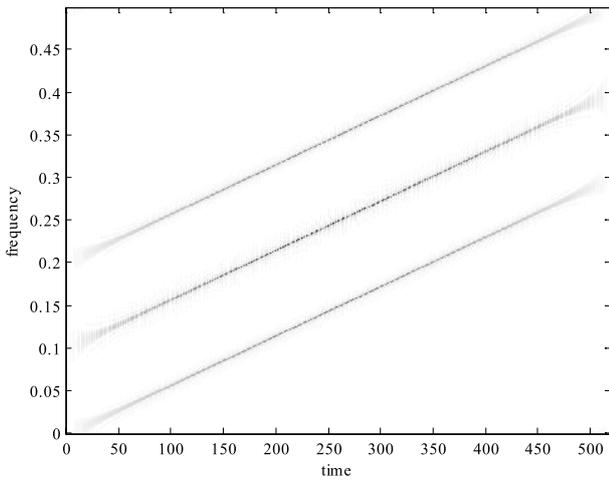
$$LPFT(x; t, \varpi) = LPFT(x; t, \omega, \omega_1, \dots, \omega_{M-1}) \\ = \int_{-\infty}^{\infty} x(t + \tau) h(\tau) e^{-j\theta(\tau, \varpi)} d\tau \quad (3)$$

式中 $h(\tau)$ 为窗函数,

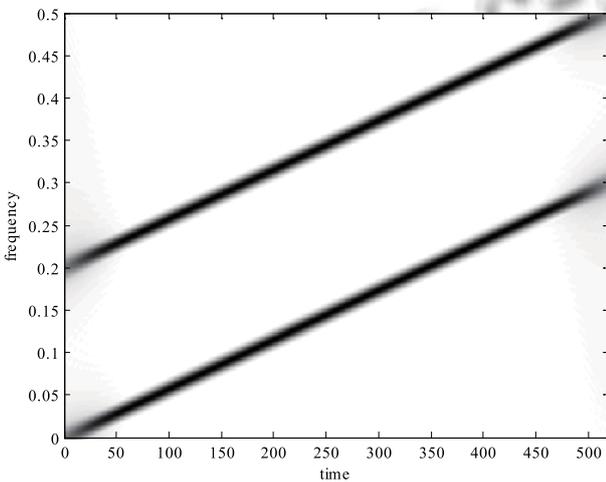
$\theta(\tau, \varpi) = \omega\tau + \omega_1\tau^2/2 + \dots + \omega_{M-1}\tau^M/M!$, $\varpi = (\omega, \omega_1, \dots, \omega_{M-1})$ 中的 $\omega_1, \dots, \omega_{M-1}$ 分别为所分析信号 $x(t)$ 的瞬时频率的一阶及高阶导数, M 为 LPFT 的阶数.当 $M=1$ 时, LPFT 对应为短时傅里叶变换.

根据 Weierstrass 逼近定理,可以用多项式逼近任意时变信号的相位.这种具有多项式相位的时变信号即为多项式相位信号(polynomial phase signals).与短时傅里叶变换不同, LPFT 使用多个参数逼近多项式相位信号的瞬时频率,因此可以提供比短时傅里叶变换更优越的时频聚集性.同时,因为 LPFT 是短时傅里叶变换的扩展形式,也满足线性特性,从而在处理多分量时变信号时不会受到交叉项的干扰.如图 1 所示,对两个平行的线性调频信号, LPFT 能给出比 STFT 更集中的时频聚集性,同时又能避免 WVD 中因多分量所产生的交叉项.因此, LPFT 是一种更为优化的处理时变信号的变换.

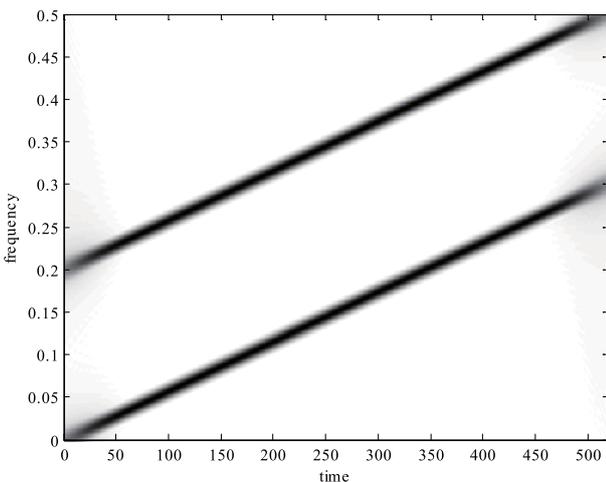
作为一种优化的时频分析方法, LPFT 已广泛应用到一系列不同的领域:在雷达成像中,当目标快速运动时,可借助 LPFT 提高成像的精度^[5,6]; LPFT 也用于通信系统中的干扰^[7];以 LPFT 为基础的波束形成技术能够对速度不同的源进行定位和跟踪^[8].



(a) STFT



(b) WVD



(c) LPFT

图 1 两个平行线性调频信号分量的 STFT, WVD, LPFT 表示

基于上述所分析的 LPTF 的优点,它能够对语音信号的时变性给出更精确的描述,更适合处理时变的语音信号,可以为语音提供更精确的分析和处理.在本文中,我们将利用 LPFT 对语音信号进行处理,以得到比 STFT 和 WVD 更精确的语音信号处理.

2 基于Matlab建立语音处理GUI系统

MATLAB 作为一种高级技术计算语言和交互式环境,将高性能的数值计算和可视化集成在一起,并提供了大量的内置函数,因此广泛应用于算法开发、数据可视化、数据分析以及数值计算等领域.在人机交互方面,matlab 中图形用户界面 GUI 具有强大的功能,由窗口、光标、按键、菜单和文本等对象(Object)构成.用户界面工具箱中具有多种界面控件,可以通过编程控制各个控件之间协调工作.GUI 同样提供了文件选择功能,即用户可根据自己的需求添加,删除文件,然后选择指定文件导入 GUI 中进行数据处理,从而实现用户的自主操作.图形用户界面开发环境 (GUIDE)提供了一系列创建 GUI 的工具.这些工具极大的简化了 GUI 设计和生成的过程.已有研究者就 Matlab 建立语音处理系统展开研究^[9,10].

本文将在 Matlab 开发环境下,完成语音信号处理的可视化 GUI 系统,包括前台界面设计和后台算法程序设计.利用 GUI 系统,用户只需选择所需处理的语音信号,并在编辑框中输入或设置相应的参数,即可实现对语音信号的分析以及其在时域、频域、STFT、WVD 和 LPFT 的对比.该系统的框架图如图 2 所示.

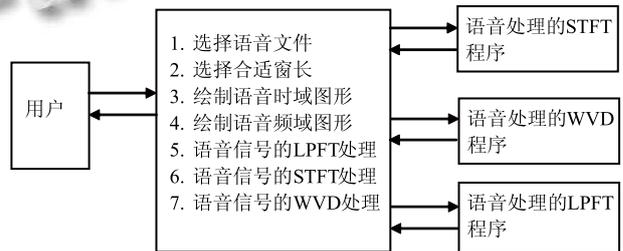


图 2 基于 Matlab 的语音信号处理 GUI 系统框架图

1) 控件的基本设计步骤: 在界面设计中, GUI 提供了按钮、滑动键、无线电选择按键、检录框、可编辑文本框、静态文本框、弹出框、列表框、双位按键、坐标系、面板、方框等控件的图标.在打开空白的 GUI 界面后,通过选择这些控件,及修改相关属性即可完

成 GUI 界面的设计. 之后进入 GUI 编程, GUIDE 会自动生成一个 M 文件, 在这个 M 文件中包含了每一个控件的回调函数的框架和初始化代码. 只需在回调函

数中添加要执行的代码即可完成 GUI 的全部设计. 所完成的基于 Matlab 的语音信号处理 GUI 界面如图 3 所示.

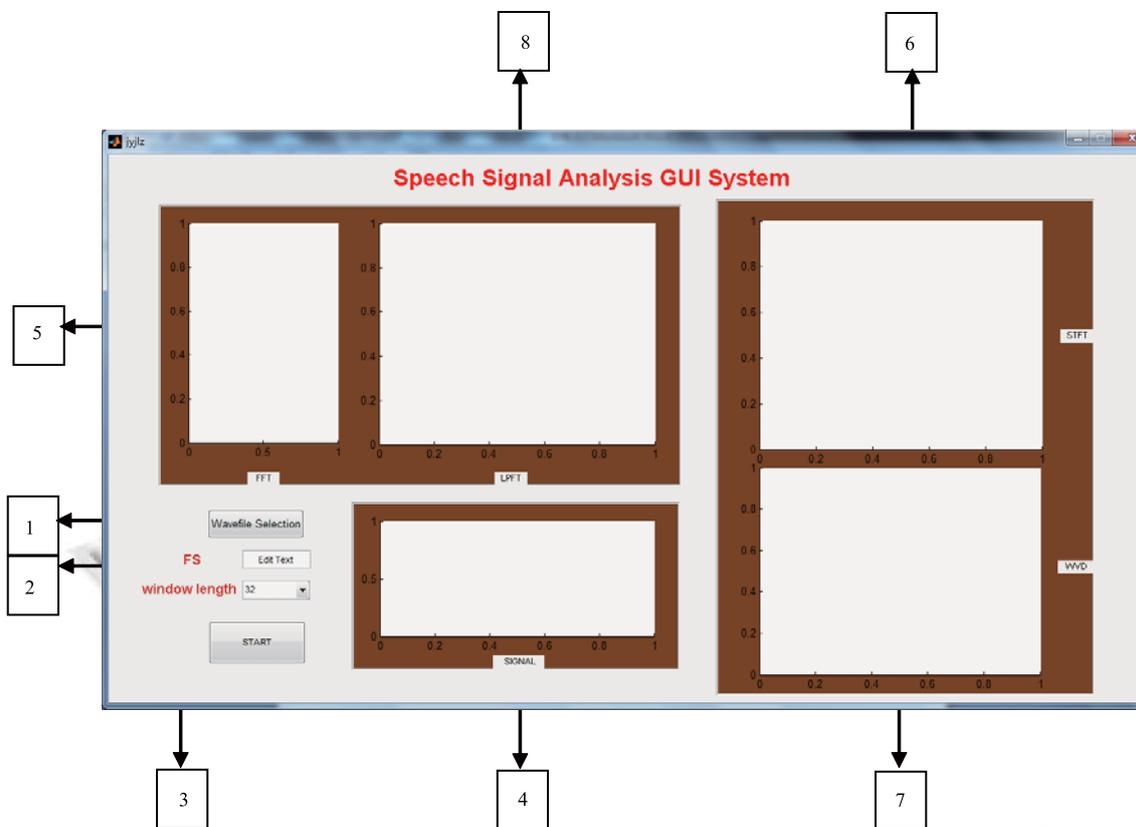


图 3 基于 Matlab 的语音信号处理 GUI 界面

图 3 中所标记的数字分别代表:

1. 输入语音信号选择;
2. 读取语音信号的采样频率;
3. 窗口长度, 现有 32, 64, 128, 256, 512, 1024 等六个选项;

4. 语音信号在时域上的波形表示
5. 语音信号在频域上的表示
6. 语音信号的 STFT 表示
7. 语音信号的 WVD 表示
8. 语音信号的 LPFT 表示

2) 该系统的可操作性以及应用性主要体现在文件选择功能. 我们可以通过路径选择要处理的语音文件, 利用 matlab 中的 `uigetfile` 函数显示一个对话框, 列举出当前目录下的文件名供用户检索或者修改将要被打开的文件名. 如果该文件名正确且存在, 用户点击 `open` 按钮时, `uigetfile` 返回文件名. 在打开文件之后,

wav 文件的数据被读入到 workspace 中, 在 `evalin` 的帮助下, 我们可以将 workspace 中的数据导入到 GUI 中, 从而完成文件的选择与读取.

利用语音信号处理 GUI 系统对语音信号 “your mail” 进行分析, 该语音信号采样频率为 5512.5Hz. 分别利用时域、频域、STFT、WVD、LPFT 对语音信号进行描述和处理, 所得到的处理结果如图 4 所示. 由图可见, 虽然语音信号的时域和频域也能提供一定的信息量, 但是它们都无法描述语音信号的频率随时间的变化情况, 而借助于时频分析方法, 如 STFT, WVD 和 LPFT 则可以将语音信号变换为时间和频率的二维函数, 描述语音信号频率随时间变化的关系. 而且, 由于 STFT 所提供的时频聚集性较低, 而 WVD 则受到多分量交叉项的干扰, 因此二者均不能对语音信号给出较好的表示. 而 LPFT 使用多个参数逼近多语音信号的瞬时频率, 因此可以提供比短时傅里叶变换更优

越的时频聚集性。同时,由于 LPFT 的线性特性,它在处理多分量语音信号时不会受到交叉项的干扰。LPFT 的优越性在语音的过渡带 0.2-0.3 范围内更为明显,因为这个部分语音的频率变化较为迅速。

3 结语

本文利用局部多项式傅里叶变换 LPFT 处理语音信号,并建立了基于 LPFT 的语音处理 GUI 系统。直观地将 LPFT 分析与时域、频域、STFT 和 WVD 分析进行对比,从而验证了 LPFT 是一种更为优化的处理语音信号的工具。该系统界面友好,易于操作,是有效处理语音信号的可视化平台。

致谢

该论文还受到以下项目资助:杭州师范大学科研启动基金(2011QDL021),杭州师范大学本科生创新能力提升工程项目,杭州师范大学实验室开放项目。

参考文献

- 1 胡航.语音信号处理.第2版.哈尔滨:哈尔滨工业大学出版社,2005.
- 2 Cohen L. Time-Frequency Analysis. New Jersey: Prentice-Hall, 1995.
- 3 张贤达.现代信号处理.第2版.北京:清华大学出版社,2002.
- 4 Katkovnik V. A new form of Fourier transform for time-varying frequency estimation. *Signal Processing*, 1995,47(2): 187-200.
- 5 Djurovic I, Thayaparan T, Stankovic L. SAR imaging of moving targets using polynomial Fourier transform. *IET Signal Processing*, 2008,2(3):237-246.
- 6 Li XM, Bi GA, Ju YT. Quantitative SNR analysis for ISAR imaging using LPFT. *IEEE Trans. on Aerospace and Electronic Systems*, 2009,45(3):1241-1248.
- 7 Djukanovic S, Dakovic M, Stankovic L. Local polynomial Fourier transform receiver for nonstationary interference excision in DSSS communications. *IEEE Trans. on Signal Processing*, 2008,56(4):1627-1636.
- 8 Katkovnik V, Gershman A. A local polynomial approximation based beamforming for source localization and tracking in nonstationary environments. *IEEE Signal Processing Letters*, 2000,7(1):3-5.
- 9 王光艳,赵晓群,王霞.语音信号时频特征显示系统的设计和仿真. *计算机工程与应用*,2010,46(29):73-75.
- 10 题原,张劲松.基于 MATLAB 的语音信号采集和分析系统的可视化设计. *齐齐哈尔大学学报*,2006,22(6): Aug, 2008,11(1):1-14.
- 4 陈成,肖俊,庄越挺.单目视频人体三维运动高效恢复. *计算机辅助设计与图形学学报*,2009,21(8):1118-1126.
- 5 雷涛,罗薇薇,樊养余,等.复杂环境下的运动人体骨架提取算法. *计算机应用研究*,2010,27(8):3194-3197.
- 6 于若飞.基于 MeanShift 的运动人体骨架重构方法. *科学与技术*,2011,11(21):5220-5222.
- 7 马颂德. *计算机视觉——计算理论与算法基础*.北京:科学出版社,1997.
- 8 Furukawa Y, Ponce J. Accurate, dense and robust multi-view stereopsis. *IEEE Pattern Analysis and Machine Intelligence*, 2002,7(8):752-758.
- 10 Taylor CJ. Reconstruction of Articulated Objects from point Correspondences in a Single Uncalibrated Image. *Computer Vision and Image Understanding*, 2000,349-363.
- 11 Zou BJ, Chen S, Shi C, et al. Automatic reconstruction of 3D human motion pose from uncalibrated monocular video sequences based on markerless human motion tracking. *Pattern Recognition*, 2009: 1559-1571.

(上接第4页)