

基于 MATLAB 的声波分析研究

许文仪 谷雨 指导老师：俞熹*
(复旦大学物理系，上海，邮编 200433)

作者简介：

许文仪 (1986~) 男，籍贯上海，现就读于复旦大学物理系本科，2004 级。

移动电话：13761735753；E-mail: niaotaiji@sohu.com。

谷雨 (1986~) 男，籍贯四川，现就读于复旦大学物理系本科，2004 级。

移动电话：13636397638；E-mail: robert_gu@163.com。

指导老师简介：

俞熹 (1978~)男，讲师，籍贯上海，现任复旦大学物理系实验中心副主任。

移动电话：13585710010；Email: whyx@fudan.edu.cn。

I 摘要

Matlab 是一款功能强大的数学软件，它附带大量的信号处理工具箱，如 sptool、fdatool、wavelet、simulink 工具箱等，为信号分析研究，特别是文中主要探讨的声波分析研究带来极大便利。声音的特征可用声压强，频律等各种参数描述，文中主要探究声音在频率上体现出的各种特性，并在此基础上在 Matlab 平台下对声纹识别技术做初步探索。

I 关键词

频谱分析，声纹识别，Matlab，Simulink，小波分析

I 中图分类号：TN912.3

Voice Analysis Based On MATLAB

Wenyi Xu Yu Gu (Instructor: Xi Yu)

(Physics Department of Fudan University, Shanghai China, Postcode200433)

Authors Introduction:

Xu Wenyi (1986~), Shanghai, China, E-Mail: niaotaiji@sohu.com

Gu Yu (1986~), Sichuan, China, E-Mail: robert_gu@163.com

Instructor Introduction:

Xi Yu (1978~), Lecturer, Shanghai, China, E-Mail: whyx@fudan.edu.cn

Vice Director of the physics lab center of Fudan University

I Abstract

Matlab is a powerful mathematic software with a mass of toolboxes dealing with signal processing, like 'sptool', 'fdatool', 'wavelet', 'simulink' etc. It gives a terrific shortcut to the research of signal processing, especially the wave analysis. We can characterize the sound with key parameters such as intensity, frequency etc. Our experiment is based on Matlab 7.0 to make a general research of the relations between voice and frequency. Besides, to have a simple

exploration of the voice recognition is another target.

I Keywords:

frequency analysis, speech recognition, Matlab, Simulink, wavelet analysis

I 正文

1. 引言

Matlab 原始程序于 20 世纪 70 年代中期由美国新墨西哥大学计算机系主任 Clever Moler 开发完成, 于 1984 年由 Mathworks 公司推向市场。在 20 多年的发展过程中软件版本不断更新, 功能逐步完善, 深受大学教师、学生及科研工作人员好评。Matlab 也成为数值计算领域耳熟能详的名词, 活跃在数值计算的各个舞台。

语音与自然界中存在的千变万化的声音的本质差别, 在于它本身的高度有序性和完美的作为信息交流的媒体即表意性。对语音特性的好奇激发了人们对语音的大量研究。在此基础上语音识别技术在最近几十年中逐步发展起来, 经历了从相对简单的有限集说话人身份识别(也称为声纹识别)到开放集识别, 进而到孤立词词义识别再到连续语音识别的发展过程, 是声学、语言学、信息处理技术等多门学科交叉综合应用的典范。作为人工智能中人机交互手段的首要实现步骤, 语音识别及合成技术成为了近年的研究热点, 也有不少实用化的技术成果被商家成功推向了市场, 让人们看到了它在 PDA、智能家电、工业控制等领域的广阔发展前景。本文就结合 Matlab 优秀的数值计算能力及强大的信号处理功能, 借助于对声音的部分基本特征的研究, 拓展 Matlab 在物理实验中的应用, 并初步揭开语音识别的奥秘。

2. 原理与背景知识

2.1 实验平台介绍:

在本实验中, 基本上最主要的仪器为话筒、耳机与 pc 机(另有定音器等发声工具), 整个实验平台建立在 Matlab 软件的基础上, 使得信号的采集、分析与数据转换完全一体化, 并且利用 Matlab 各工具包的强大功能, 还可以对数据进行即时的处理反送。由于 Matlab 各个工具包往往采用了业界中最尖端的算法, 因此相比与传统的实验方式, 本实验不但在设备搭建上方便易行, 而且在数据处理方面更加快捷、精确。

以下是在本实验中主要采用的几个 Matlab 工具包

(1) Simulink 工具箱

MATLAB 附带的 Simulink 工具包具有强大而快捷的实时分析功能, 为语音信号的实时处理带来便利。

图 1 为通过 simulink 实现的信号采集器。数据直接读入工作空间。可实现波形实时显示, 即时频谱显示, 即时回放等多项功能, 是本实验数据采集的主要平台。

(2) sptool 工具包

对于已经采集的数据, 可通过 Matlab 共有的工作空间导入至 sptool 工具包, 此间可以对信号时域、频域进行更详细的分析: 例如

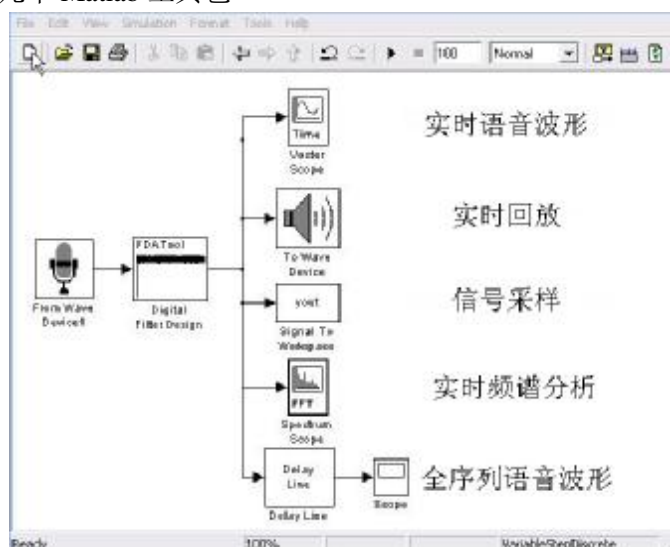


图 1: simulink 数据采集面板

选取某一段短时域部分发音，或者使用其内带的数字滤波器在频域中选取一段来发音，通过反复挑选发音段，可以发现人声与频率、音节长短、衰减快慢的定性关系。

(3)封装子模块

当需要对大量的声音信号作快速处理时（如不同人的同一发音或同一发音人的不同语言），此时可以先在 `simulink` 中设计子模块：再利用子模块构建一个复杂系统，通过延迟线，就可以使经过不同滤波处理的信号依次输出到耳机、音响等外设上。籍由这样连续的输出，人耳就能更加清晰地便别出不同频率成分的影响。

可见，凭借 `Matlab` 强大的实时信号处理能力，只要在其中将程式模块组装完毕，就可以将繁琐的仪器搭建、数模转换、信号分析等过程轻而易举地程序化，从而集中精力于研究发声机理。

2.2 发声的生理机理与发声模型

对声音特别是人声的研究，首先得从人发声的基本规律开始。人的发音机理如图所示：发音时由肺部收缩送出一股直流空气，经气管流至喉头声门处。在发声之初，声门处声带肌肉收缩，声带并拢间隙小于 1mm，这股直流空气冲过很小的缝隙，使声带得到横向和纵向速度，此时，声带向两边运动，缝隙增大，声门处压力下降，弹性复力将声带拉回平衡位置继续趋向闭合，即声带振动，且具有一定的振动周期。^[1]

为对发声模型作一个定性了解，考虑如图 4 的简化的一维模型。按以上发声机理，气流上冲，获得一个特定的频率，在口腔中形成发射，最后由口部出射由于口腔中气体的速度一定，故给定一个谐振动的话，可得相应波长为 $f \cdot \lambda = v$ ，于是初态入射波的空间部分可表达为： $\sin(k_0x)$ 。

口腔是产生不同声音的重要部位，现用最简单的驻波模型，易知其本征态为： $\sin(\omega_n x)$ ，

（ ω_n 与 L 有关，实际中 $L=L(x)$ ），故驻波模式数与频率特征位置对应，

$\omega_n = \frac{np}{L} = \frac{2p}{L} = \frac{2p}{v} \cdot f_n, \therefore f_n = \frac{v}{2L} \cdot n, n = 1, 2, 3L$ 。这就是声音频域图中产生基频与泛音的原理。

初 态 $t=0$ ，有 $\sum_{n=1}^{\infty} c_n \sin(\omega_n x) = \sin(k_0 x)$ ，求 得

$c_n \propto \int_0^L \cos(\omega_n - k_0)x dx + \int_0^L \cos(\omega_n + k_0)x dx$ ， $|c_n|^2$ 即为对应频率为 ω_n 的振动频率波的功率。由上式，仅当 $\omega_n - k_0 = 0$ 时，被积函数在 $0 \sim L$ 的周期内才会有很大的值，否则由于 \cos 函数的振荡性，积分值必不会很大。

2.3 采样与降噪

根据采样定理：一个连续变化的模拟信号，假设有最高频率或带宽 F_{max} ，若采样周期为 T ，则采样频率为 $F=1/T$ ，若能满足 $F=1/T > 2F_{max}$ ，即采样频率大于或等于模拟信号最高频率的两倍，那么采样后的离散序列就能无失真地恢复出原始连续模拟信号^[2]。为了同时兼顾运算数据量的要求，试验中采用 44.1kHz 采样率，16 位数据。

在录音过程中，环境噪声和系统噪声将会对试验结果带来干扰，故需经过滤波预处理，

除去噪声干扰^[3]。试验在无声试验室中实施，故环境噪声可略去。滤波可由滤波器实现，设置中心频率为 50Hz 的带阻滤波器即可除去电流噪声对信号产生的影响。滤波器的设计在 Matlab 中可由 fdatool 或 sptool 工具箱方便的实现。在命令行分别输入“fdatool”和“sptool”即可分别进入二者工作界面。fdatool 工具箱能方便而快捷的完成滤波器的设计，并可保存为 m 文件方便调用。sptool 是一个功能高度集成的工具箱，能实现信号的回放、滤波器设计、fft 等多种变换等多种功能^[4]。

2.4 小波分析方法对语音识别的改进

在 matlab 中重写 DTW 等经典识别算法即可实现比较高精度的识别率。但这些算法由于物理例外使用傅立叶变换直接处理信号，在其诞生之初本身即存缺陷。傅立叶变换在实现将信号转换至频域进行处理时，其缺点是只能对全时域信号作分析，而作为瞬变的语音信号，其瞬时特征得不到反映。针对这种不足，前人提出了可以提高辨识精度的改进方法，如通过加窗实现“短时变换”等，但仍然存在缺陷，主要表现在：

1. 时窗局域化是一次性的，即窗函数的形状和大小是固定的，不能敏感的反应信号的突变。而突变恰好是语音信号的特征之一。

2. 由于信号分析中的“不确定性”原理，高频部分时域相对窄，即需较高的时域分辨率而较低的频率分辨率；而低频部分，时域相对宽，即需较高的频率分辨率而较低的时域分辨率。^[5]

而小波变换恰能满足这些要求。它的窗宽随频率增高而减小，符合高频信号高时域分辨率的要求。故笔者考虑在算法中引入小波变换作为作为尝试作为改进。

3. 实验内容与数据分析

3.1 频谱分析及结果分析

语音感知过程与人的听觉系统具有频谱分析功能是紧密相关的。傅立叶变换为时域信号的频域化分析提供了有力的数学工具。而更为重要的是 FFT（快速傅立叶变换算法）的出现，大大降低了数据处理的计算量，为实际应用铺平了道路。在 MATLAB 中可以通过命令： $z=fft(s).*conj(fft(s))$ ，（fft 为快速傅立叶算法）实现该变换。其中 s 为信号序列，z 为频率功率谱。

人的声道和鼻道都是非均匀的声道管，声道管的谐振频率称为共振峰频率，或简称共振峰。它与发音器官的确切位置和形状有很大关系，即共振峰和声道的形状与大小有关。经研究，韵母音色差异可用前三个共振峰 f1、f2、f3 表示，f1 主要分布在 290 到 1000Hz 范围内，f2 主要分布在 500 到 2500Hz 范围内，f3 主要分布在 2500 到 4000Hz 范围内。^[6]

实际中，声带对给定的 f 是差不多的，改变口腔的形状、舌的位置等，相当于调整 $w_n - k_0 = 0$ ，因此引起主极大位置的变化。假定保持口腔位形几乎不变，多次发音，实得图形中主极大位置几乎不变，次极大位置也少有变化，两者高度对比的变化也不大。由于人声带起振频率偏低，因此频图中能量集中于低频段。见图 5，横轴为频率，单位 Hz，纵轴为功率。

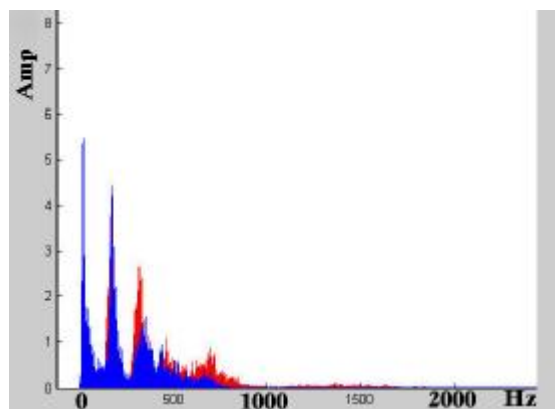


图 5：发声频谱

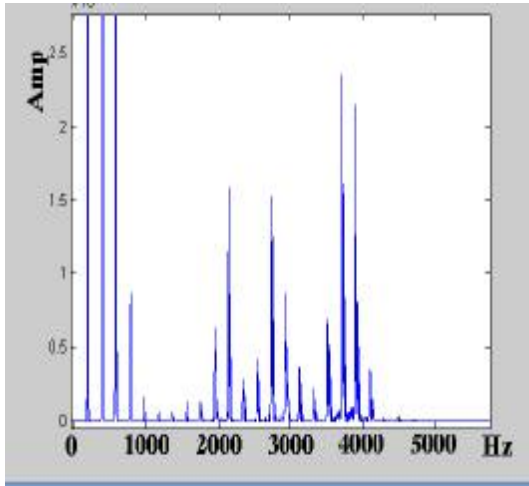


图 6

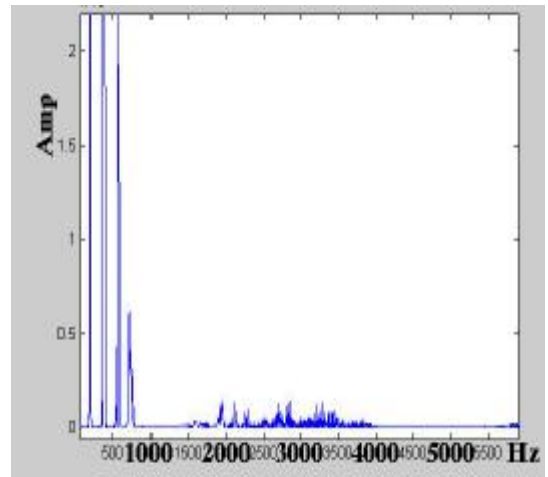


图 7

图 6 图 7 为两人同发“ A ”音时的频谱。可以看出，基频部分（小于 1000Hz）差异不大，但高频部分的衰减有显著的差距，这和不同人的声带结构不同，导致谐波结构不同是一致的。基于高频衰减行为随说话人改变而对应改变的特征，可作为声纹识别的初步方案。但易受发声人身体条件等的影响，这也是声纹识别技术最大的缺陷，即可通过刻意模仿，自然走音等状况产生漏判和误判^[2]。

再考虑模型中时间部分 $e^{-if_n t}$ ，假定时间部分为延时达 $2A$ 的谐波 $e^{-if_n t}$ ，则

$$F(f) \propto \int_{-A}^A e^{i(f-f_n)t} dt = \frac{\sin(f-f_n)A}{f-f_n}, \quad \text{功率} \propto \frac{\sin^2(f-f_n)A}{(f-f_n)^2}, \quad \text{半峰宽} \Delta f_n = \frac{x}{A}.$$

x 为方程 $\sin(x) = \frac{x}{\sqrt{2}}$ 的解，故如 A 为常数， Δf_n 不变。

实验得到的是高频率半峰宽显著增加，说明 A 随 n 的增加而减少，即高频段持续时间减少。见图 8 图 9。

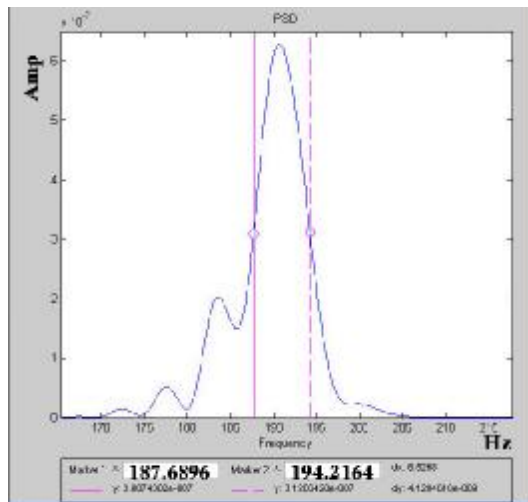


图 8: 低频半峰宽

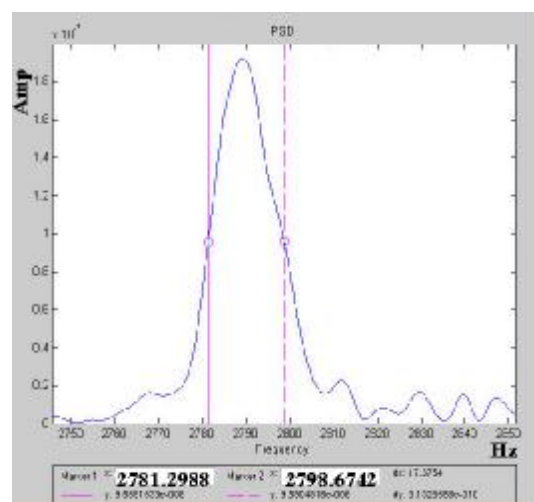


图 9: 高频半峰宽

图 8 为中心频率为 193Hz 附近的基频，半峰宽（图中圆形数据点之间）读数为 6.5Hz 左右；图 9 为中心频率 2790Hz 附近的共振峰中的一条主峰，半峰宽（圆形数据点之间）读数为 17.3Hz 左右。可见随频率的增高，半峰宽的增加是相当显著的。

实验中还发现衰减并不是单纯的 A 的减小，时间部分或许是类似 e 指数形的衰减因子，如图 10 所示，考虑到模型比较粗糙，故只能定性给出上述的一些基本说明。

通过傅立叶变换还可方便的读取发音源的频率，例如音叉、乐器等，可以简便可行的拓展 MATLAB 在物理实验中的应用。图 11 为定音器的三个基音（分别以不同的颜色标示）的频谱图，从主峰可方便的辨识频率。高频谐波的存在与定音器的加工工艺有关。

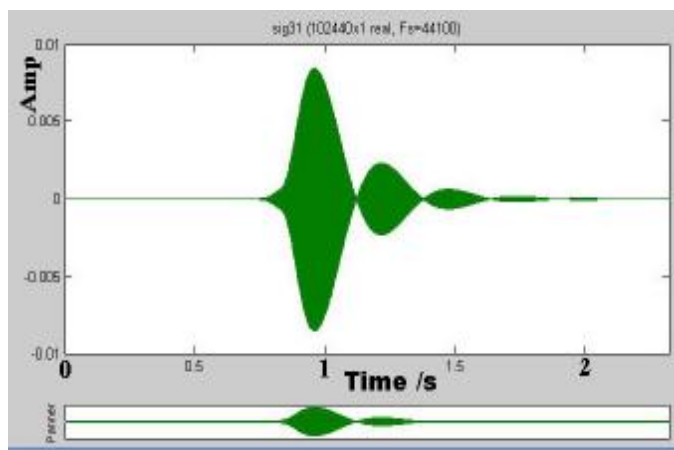


图 10: 峰的衰减现象

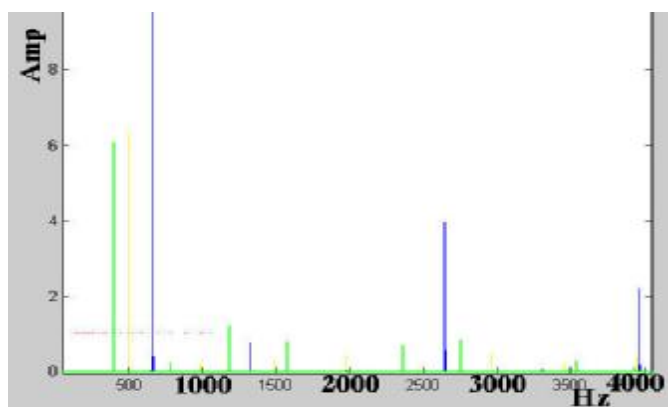


图 11: 定音器频谱

3.2 小波分析在语音识别中的应用研究

在 Matlab 中，小波分析可由“wavlet”工具箱实现，也可由命令行输入。分解过程如图 12，图中以 3 层分解为例。返回值以图中格式存于工作空间。

在命令行键入“wavemenu”即可进入小波工具箱界面。单击“wavelet 1-D”即可进入一维连续小波分析界面。小波函数种类繁多，每一种都有自己的特性，最适函数需要经验和反复试验寻求，不同的条件有不同的选择^[5]。为统一比较各种参数，笔者使用被广泛应用的“haar”小波作为分解小波函数，其函数形式为：

$$\Psi_H = \begin{cases} 1, & 0 \leq x \leq 1/2 \\ -1, & 1/2 \leq x \leq 1 \\ 0, & \text{其他} \end{cases}$$

为 1，滤波器长度为 2。

图 13 和图 14 中显示的分别是发“A”音以“haar”小波做 8 层分解的细节系数图和对应的经 FFT 变换至频域后的图形，随图的序号增加，频带范围递增。

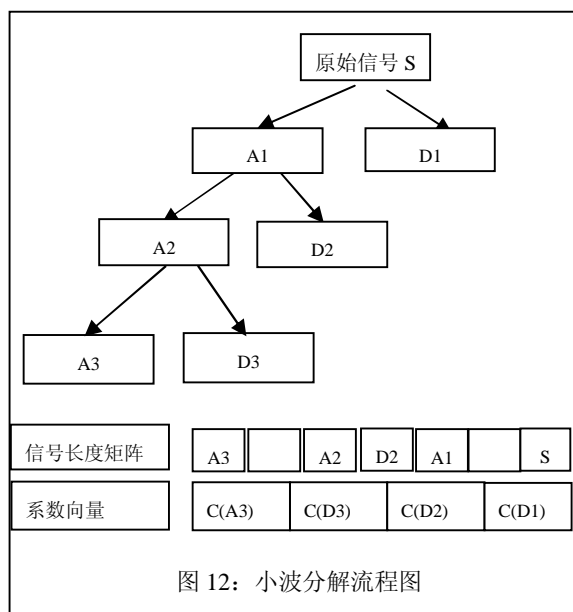


图 12: 小波分解流程图

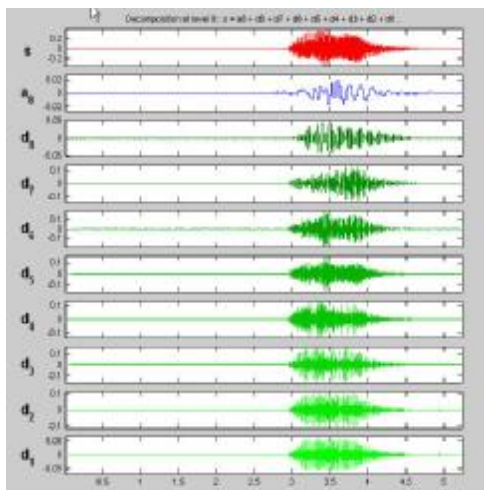


图 13 小波分解系数图

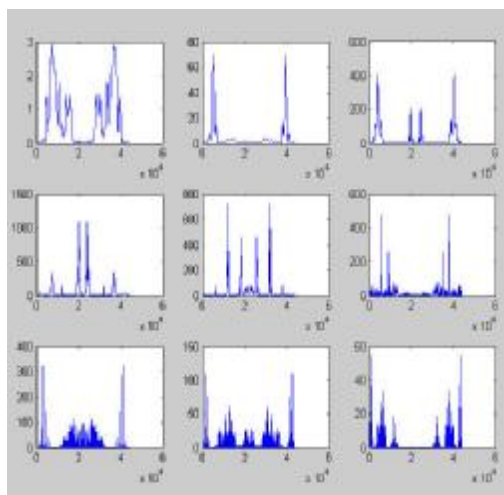


图 14 (横坐标为频率/Hz, 纵坐标为计算机的内部数据强度)

而我们更关心的是频域的结果，特别是高频部分细节系数变换的结果。根据发声原理，低频部分标志着基音的特征，所以当不同的人发同样的音时（本实验中受试者都以正常方式发“A”音），基频频谱峰值应大致相同，而不同的人由于声带构造不同，在基音大致相同的情况下（都发同一个音），高频谐波的衰减将会有很大不同，体现在频谱上就是高频区频谱差异显著，这是分辨不同人的特征之一。从此下面的图中我们可以看到以上预测是准确。

以上结果也可调用命令：`[c l]=wavedec (s, N, n)` 实现。其中，`c` 为分解得到的细节系数；`l` 为对应细节系数的长度；`s` 为信号名称；`N` 为分解层数；`n` 为小波函数名。

图 15、16 为同一人两次发声的结果。由于每次发声声带结构都会有细微不同，故两幅图仍有细微差别，但可以看出图形具有较高的一致性，具有同样的分布特征。图 17、18 为不同人发声的结果。同时结合对比图 15 或 16，可以看出，除坐标为 [1, 1]（图片标号自上而下）的图外，差别都是相当明显的，符合基音相同的要求。

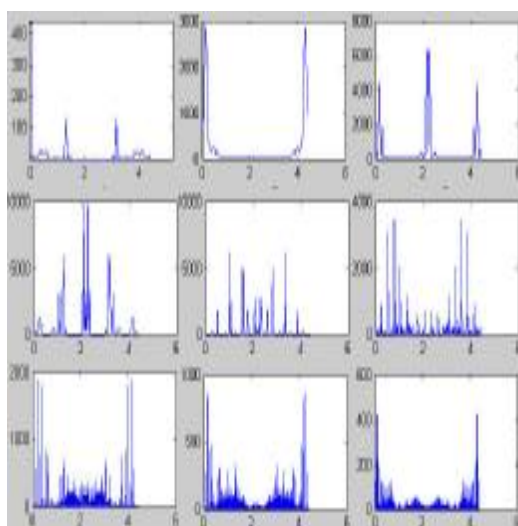


图 15

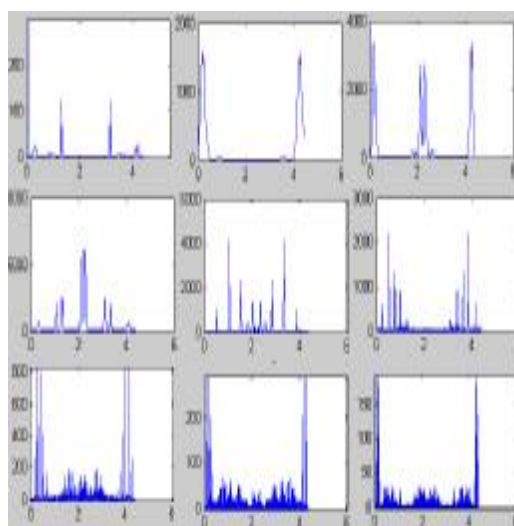


图 16

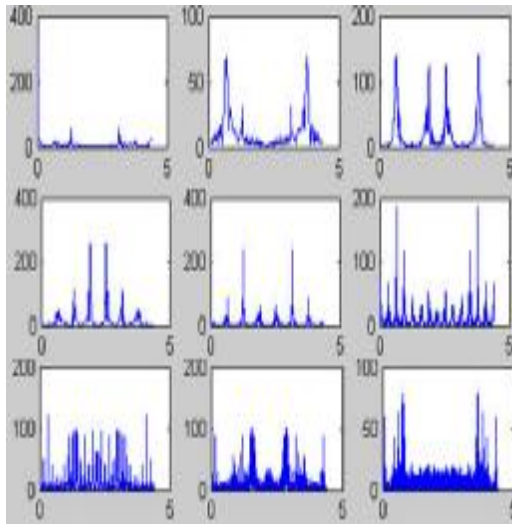


图 17

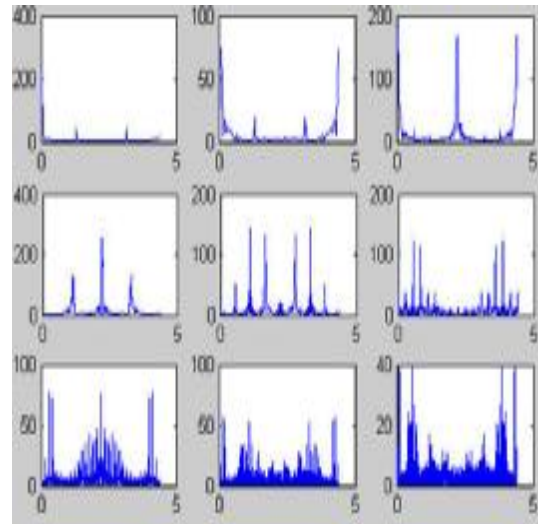


图 18

图 15-18 为采样数据运算结果，频率 (Hz) 的相对强度分布

经过采样对比，最后选用序号为[1, 3], [2, 1], [3, 1], [3, 2]的图形作为个人特征参数的提取源。在封闭集（实验人数为 8 人）中，可有效分辨说话人身份，正确率不小于 80%。

4. 结论和展望

通过试验的探索，充分展示了 MATLAB 强大而且专业的数值计算和信号处理能力。通过对其附带的各工具箱的使用，及 simulink 模块的快速实时开发，结合自由的程序编写，使得一些原本复杂的数据处理过程变得快捷简便，大大提高了实验的可操作性能。MATLAB 提供的强大的实时处理能力是信号处理特别是语音识别所梦寐以求的，结合软件提供的各类软硬件接口和高级语言编写环境，为实验的展开提供了可靠的软硬件支持，展示了 MATLAB 平台在物理实验领域广阔的应用前景。

在笔者基于 Matlab 环境下实现的初步说话人识别实验中，小波变换表现出的在信号处理中的强大优势，符合其在信号处理领域愈发受到人们重视的现状，为人们描绘了深入研究其在语音识别中的应用的的美好图景。在算法方面，试验可以考虑采用更为完善但同时更复杂的 HMM（隐马尔科夫）模型等提高辨识的准确率。这些都对未来的研究方向提出了可行的切入点。

致谢： 本文获得复旦大学青年科学基金（EYH1512025）以及国家基础科学人才培养基金（J0730310）资助，特此致谢

参考文献：

- [1]马大猷，理论声学,北京，科学出版社，2004
- [2]王炳锡等，实用语音识别基础，北京，国防工业出版社，2005 年
- [3]王炳锡，语音编码，西按，西安电子科技大学出版社，2002 年
- [4]飞思科技产品研发中心，MATLAB7 辅助信号处理技术与应用，北京，电子工业出版社，2003
- [5]董长虹，Matlab 小波分析工具箱原理与应用，北京，国防工业出版社，2004

- [6]杨家沅, 语音识别与合成, 成都, 四川科学技术出版社, 1994 年
- [7]王沫然, MATLAB 与科学计算, 北京, 电子工业出版社, 2003 年
- [8]董泽、谢华等, 小波变换模极大值消噪算法的研究, 电力科学与工程, 2005
, No. 3, 12-16
- [9]蒋文建、韦岗, 一种多层子带的噪声语音识别新方法, 数据采集与处理, 2002. 3,
Vol . 17. 1
- [10]Okawa S.Bocchieri E.Potamianos, A Multi-band speech recognition in noisy
environments[C],ICASSP,1998,641~644.