

## 基于元音分类度的帕金森病语音特征分析

张 涛<sup>1 2\*</sup> 洪文学<sup>2</sup> 常凤香<sup>2</sup> 刘旭龙<sup>2</sup>

<sup>1</sup> (燕山大学信息科学与工程学院, 秦皇岛 066004)

<sup>2</sup> (燕山大学生物医学工程研究所, 秦皇岛 066004)

**摘 要:** 帕金森病是一种常见的神经功能障碍疾病, 患有帕金森病给患者及其家人都带来巨大痛苦。帕金森病虽然无法治愈, 但为早期诊断对患者非常重要。语音障碍是帕金森病的早期症状之一, 本研究选择使用持续发音法, 针对不同的元音发音对语音样本进行采集, 并从周期变化特征、峰值变化特征、谐波信噪比以及非线性动力学等几方面进行语音特征分析。实验结果在证明语音特征用于帕金森病诊断有效性的同时, 也发现不同特征下不同元音的类间分离度, 为基于语音障碍的帕金森病自动诊断奠定基础。

**关键词:** 帕金森病; 语音障碍; 语音特征; 元音; 类间分离度

中图分类号 TP391.42 文献标识码 D 文章编号 0258-8021(2011)03-0476-05

## Speech Features Analysis of Parkinson's Disease by Vowel Class Separability

ZHANG Tao<sup>1 2\*</sup> HONG Wen-Xue<sup>2</sup> CHANG Feng-Xiang<sup>2</sup> LIU Xu-Long

<sup>1</sup> (College of Information Science and Engineering, Yanshan University, Qinhuangdao 066004, China)

<sup>2</sup> (Institute of Biomedical Engineering, Yanshan University, Qinhuangdao 066004, China)

**Key words:** Parkinson's disease; vocal disorder; speech feature; vowel; class separability

## 引言

帕金森病(Parkinson's Disease, PD)是人类常见的神经退行性疾病之一, 其发病率仅次于阿尔茨海默病(Alzheimer's Disease)。据流行病学调查, 在北美约有 100 万的帕金森病患者<sup>[1]</sup>。而在我国, 2001 年约有 200 万患者<sup>[2]</sup>, 且每年以 10 万人的速度递增。

同许多其他神经系统疾病一样, 对帕金森病的诊断主要依赖于专家的临床经验<sup>[3]</sup>。目前, 我国大约有 6 成的帕金森病患者被误诊或者漏诊。除了对帕金森病认识不够的因素外, 高昂的诊断费用与复杂的诊断过程往往令大部分早期患者望而却步。随着老龄化社会的到来, 急需一种可以应用于基层医疗机构甚至家庭的方便可行、费用低廉的帕金森病初步诊断方法, 这对于提高帕金森病的早期就诊率具有重大意义<sup>[4]</sup>。因此, 以应用简便为特色的帕金森病新型诊断方法也成为研究的热点。

在帕金森病的各种表现中, 语言障碍为早期帕金森病的典型症状之一<sup>[5]</sup>, 大约 90% 的帕金森患者会出现某种程度的语言障碍<sup>[6]</sup>。对语言障碍的测量具有测量方便、有利于远程诊断的特点, 使得该类方法得到了极大的关注<sup>[6]</sup>。Little 等对此进行了一系列的研究<sup>[7-8]</sup>, 并利用模式识别方法对基于语言障碍的帕金森病诊断进行了分析, 奠定了模式识别方法在基于语言障碍的帕金森病机器诊断方法中的理论基础。基于 Little 等的数据集, Guo、张涛等分别运用遗传算法和多维筛对其进行了分析<sup>[9-10]</sup>。从模式识别角度看, 现有的分类精度达到了较高水平。与此同时, 针对语言障碍与所选择语音特征及发音间关系的研究却未见报道。由于不同发音对应的神经控制不同, 而不同发音情况下其语音特征表现必然有所差异, 因此该研究的进展水平将影响基于语言障碍的帕金森病的检测与识别过程。

基于此, 本研究利用持续发音法, 对不同元音

doi:10.3969/j.issn.0258-8021.2011.03.026

收稿日期: 2010-11-10, 录用日期: 2010-12-26

基金项目: 国家自然科学基金(60904100)

\* 通讯作者。 E-mail: zhtao@ysu.edu.cn

进行语音特征分析与统计,寻找不同语音特征下不同元音的类间分离度并讨论其物理意义,为具有明确物理意义的帕金森病语音诊断奠定基础。

## 1 实验方法

### 1.1 采集方法与对象选择

目前,有多种方式对语言障碍进行测试,如持续发音法、连续演讲法等。其中,持续发音法只要要求受试者发出特定的音节,这不但可以有效地避免发音不清带来的干扰,而且不受语种与方言等因素的影响,因此得到了广泛的应用。考虑到帕金森病发病地域性广的特点,笔者采用持续发固定元音法进行研究。考虑到语音的普遍性,本研究选择的元音来源于英语国际音标中的5个,包括[ei][i:] [æ][ɔ:] [a:]。在语言学中,绝大多数的语言都有相同或类似的发音,因此该测试结果具有普遍性。

在受试对象的选择上,本实验选择50个健康受试对象,由其进行健康人健康组与模拟帕金森患者的发音。在实验前,每位受试者要进行帕金森患者发音状况的学习。由健康人模拟帕金森患者发音,除了考虑到测试所需时间较长、帕金森患者的身体状况难以承受之外,还可以对同一人、不同状况下的发音进行监测,从而发现递进变化规律。其中,正常发音录制的语音归类为健康组,模拟发音录制的语音归类为障碍组。

在样本采集过程中,测试者使用持续发固定元音法,针对[ei][i:] [æ][ɔ:] [a:]这五个特定元音分别录制样本,每段语音样本的长度在6~10 s不等,采样频率为8 kHz,由于人类日常语音的频率范围在300~3 400 Hz,符合奈奎斯特抽样定理,切实可行。录制完毕的样本随后被送入计算机,进行后续语音特征分析,以获得所需的有效信息。

### 1.2 语音特征选择

在对现有各种语音特征算法进行对比分析的基础上,结合Little等的研究成果<sup>[8]</sup>,选取了周期变化特征、峰值变化特征以及谐波信噪比、趋势波动分析法等作为研究重点,以期明确各特征算法的物理意义,为语音诊断提供更合理的理论依据。

## 2 实验结果与讨论

### 2.1 周期变化特征

由语音信号本身特点及持续发音法的测试特点可知,采集到的语音时域波形图具有明显的类周期性。但是,要想进一步了解各语音周期值的相对

变化情况,就需要对所获得的语音周期序列进行算法分析。对于周期变化特征,可采用跳动、相对幅度振动、5点周期振动商、周期间平均绝对差与平均周期比进行测量<sup>[11]</sup>,结果分别如表1~表4所示。

表1 相对跳动结果

Tab. 1 The results of relative jitter

元音样本	[ei]	[i:]	[æ]	[ɔ:]	[a:]
健康组/%	12.62	10.19	9.56	3.84	6.00
障碍组/%	10.64	8.85	15.67	5.36	5.17

表2 相对幅度振动结果

Tab. 2 The results of RAP

元音样本	[ei]	[i:]	[æ]	[ɔ:]	[a:]
健康组/%	7.92	6.57	6.11	2.51	3.89
障碍组/%	6.57	5.50	9.72	3.38	3.24

表3 5点周期振动商结果

Tab. 3 The results of PPQ5

元音样本	[ei]	[i:]	[æ]	[ɔ:]	[a:]
健康组/%	9.21	8.17	7.04	3.10	4.69
障碍组/%	6.79	6.47	10.70	4.03	3.68

表4 周期间平均绝对差与平均周期比结果

Tab. 4 The results of DDP

元音样本	[ei]	[i:]	[æ]	[ɔ:]	[a:]
健康组/%	23.76	19.73	18.33	7.55	11.69
障碍组/%	19.73	16.51	29.17	10.14	9.72

经过对各组数据的对比分析可以看出,人类的语音信号是一个类周期信号,基音周期值随时间会发生一定振动,而存在语言障碍的确可以使基音周期的振动情况发生变异,本研究中所使用的跳动(jitter)、相对幅度振动、*n*点周期振动商、周期间平均绝对差与平均周期比这些语音特征也确实可以把这种变异情况展现出来。

对各元音的运算结果进行横向比较后发现,在发[ei][i:] [a:]这3个元音时,健康测试者的特征值略大于模拟语言障碍测试者的特征值。而在发[æ][ɔ:]这两个元音时,模拟语言障碍测试者的数据反而会大于健康测试者的特征数据。这是由于人的发音过程类似于一个非线性动力系统。在发出不同的语音时,口腔及声道形状、肌肉群的紧张程度都会有所不同,造成了整个非线性动力系统特性的改变<sup>[13]</sup>。反映到语音特征值上,发不同的元音,其运算结果的特点也会有所不同。

### 2.2 峰值变化特征

与周期变化特征的研究类似,为更清晰地体现

语音峰值的相对变化情况,同样需要对所获得的语音峰值序列进行特征算法分析。基于此,引入闪烁、 $n$ 点幅度振动商( $n=3,5,11$ )以及相邻周期幅度差的平均绝对差的概念<sup>[14]</sup>。其中,闪烁分为绝对闪烁(shimmer dB)和相对闪烁(relative shimmer)。表5~表10分别为各特征对语音数据样本的运算结果。

表5 绝对闪烁结果

Tab. 5 The results of shimmer

元音样本	[ei]	[i:]	[æ]	[ɔ:]	[a:]
健康组/dB	0.24	0.24	0.31	0.27	0.33
障碍组/dB	0.29	0.24	0.26	0.31	0.35

表6 相对闪烁结果

Tab. 6 The results of relative shimmer

元音样本	[ei]	[i:]	[æ]	[ɔ:]	[a:]
健康组/%	2.80	2.84	3.45	3.12	3.71
障碍组/%	3.41	2.85	3.09	3.66	3.79

表7 3点幅度振动商

Tab. 7 The results of APQ3

元音样本	[ei]	[i:]	[æ]	[ɔ:]	[a:]
健康组/%	1.58	1.63	1.94	1.88	2.04
障碍组/%	2.03	1.64	1.76	2.13	2.03

表8 5点幅度振动商

Tab. 8 The results of APQ5

元音样本	[ei]	[i:]	[æ]	[ɔ:]	[a:]
健康组/%	1.75	1.95	1.99	2.04	2.15
障碍组/%	2.12	1.85	1.93	2.40	2.18

表9 11点幅度振动商

Tab. 9 The results of APQ11

元音样本	[ei]	[i:]	[æ]	[ɔ:]	[a:]
健康组/%	1.80	2.09	2.04	2.16	2.35
障碍组/%	2.30	1.95	1.99	2.60	2.29

表10 相邻周期幅度差的平均绝对差

Tab. 10 The results of DDA

元音样本	[ei]	[i:]	[æ]	[ɔ:]	[a:]
健康组/%	4.75	4.90	5.84	5.65	6.13
障碍组/%	6.09	4.92	5.28	6.40	6.10

对比实验数据结果可以看出,类周期性的语音信号,其幅度峰值随着时间推移也会发生一定振动,并且不同语音样本的峰值振动情况均可以由闪烁(shimmer)、 $n$ 点幅度振动商以及相邻周期幅度差

的平均绝对差这些语音特征展示出来。

经过这六个特征值的提取运算,所使用的元音[ei][i:][æ][ɔ:][a:]分别体现了不同的变化特点。其中,针对[ei]与[ɔ:]的运算,模拟语言障碍测试者的数据明显大于健康测试者的特征数据;而在元音[æ]的特征值提取结果中,健康测试者的数据均大于模拟语言障碍测试者的数据。与这两种情况不同,[i:]和[a:]的特征值提取结果虽表现出了峰值的振动特点,但两类测试者的数据基本相同,暂时无法把它们作为诊断依据使用。

以上实验结果再次证明,存在某种语言障碍会影响人的发音动力系统。但由于发不同的语音其动力系统的特点不同,就造成了针对不同发音,同一种语音特征也会以不同形式将语言障碍体现出来。该结论符合语音模型和帕金森病的发病机理。

### 2.3 谐波信噪比

自相关性是衡量数据之间关联程度的一种常用方法,而谐波信噪比就是一种基于自相关理论的特征值算法。

以元音[ei]与[i:]的相应语音样本的自相关为例,结合其自相关运算结果图,具体分析其在谐波信噪比这一语音特征中的应用。

图1和图2分别为为健康组和障碍组针对元音[ei]的自相关运算结果。对于一个稳定时间信号 $x(t)$ ,其自相关函数为

$$r_x(\tau) = \int x(t)x(t+\tau)dt \quad (1)$$

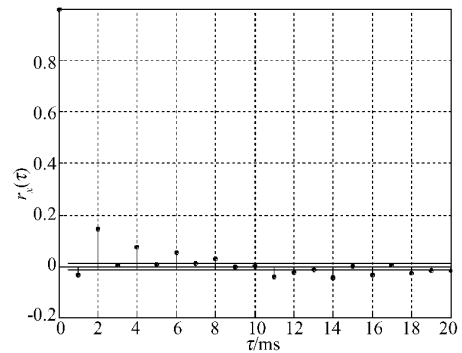


图1 元音[ei]健康组数据自相关

Fig. 1 The self-correlation of [ei] from healthy group

由于语音信号是一种类周期信号,所以自相关运算结果并没有周期性出现最大值,均出现了一些局部最值,当 $\tau=2$ 时,自相关函数取到局部最值。中的最大值,用此值比上 $\tau=0$ 时的函数值,即为语音信号中的谐波比重。

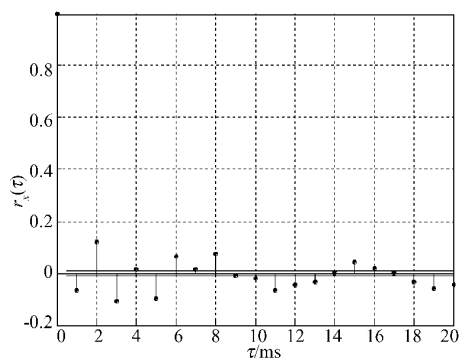


图 2 元音 [ei]障碍组数据自相关

Fig. 2 The self-correlation of [ei] from disorder group

与元音 [ei]类似 ,针对 [i:] 的语音样本数据进行自相关运算后 ,其结果也没有周期性出现最大值 ,而只是出现有一些局部最值 ,如图 3 和图 4 所示。当  $\tau = 2$  时 ,自相关函数值取到局部最值中的最大值 ,用此值比上 0 点的函数值 ,同样代表了语音样本信号中的谐波成分。

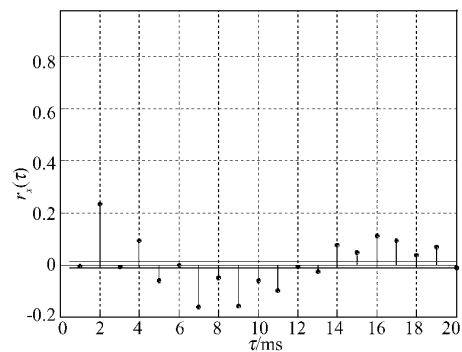


图 3 元音 [i:]健康组数据自相关

Fig. 3 The self-correlation of [i:] from healthy group

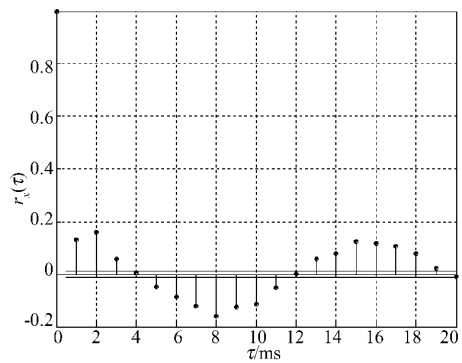


图 4 元音 [i:]障碍组数据自相关

Fig. 4 The self-correlation of [i:] from disorder group

人的语音信号或多或少都会夹杂一定的噪声

成分 ,若不考虑外界环境的干扰 ,这种噪声主要就与发音动力系统的特点有关 ,所使用的谐波信噪比就是一个用来度量信号中谐波成分与噪声成分比例的特征值。表 11 列出函数谐波信噪比对语音数据样本的运算结果。

从数据分析不难看出 ,由于加入了颤抖、发音间隔及较多的气息声 ,使障碍组的谐波信噪比明显小于健康组 ,而谐波信噪比取值较小说明其噪声成分较大。所以 ,借助本研究所使用的谐波信噪比这一特征的取值 ,可以展示出发音者是否存在语言障碍。

表 11 谐波信噪比结果					dB
Tab. 11 The results of HNR in dB					
元音样本	[ei]	[i:]	[æ]	[ɔ:]	[a:]
健康组	-7.63	-5.20	-5.92	-8.16	-5.82
障碍组	-8.56	-7.19	-7.49	-9.26	-7.38

## 2.4 趋势波动分析

趋势波动分析 (detrended fluctuation analysis , DFA) 是一类应用于研究语音信号中随机噪声自相似程度的语音特征。这种随机噪声大多是在发声时气流经过声带而产生的。对于那些存在语言障碍的帕金森病患者 ,由于其声带存在病变 ,就会使夹杂进语音信号中的随机噪声发生改变。

表 12 和表 13 为针对健康组与障碍组测试语音样本在不同窗口长度下的计算结果。如前所述 ,趋势波动分析是一类基于非线性动力系统理论的新型语音特征 ,其具体实施过程主要分为两大部分 ,即语音数据变化趋势求取和语音数据围绕其变化趋势上下波动情况分析。通过对随机噪声特性分析 ,可以观测到帕金森患者的病变情况。

表 12 健康组样本的趋势波动分析					
Tab. 12 The results of DFA for healthy group					
窗口长度	[ei]	[i:]	[æ]	[ɔ:]	[a:]
100	193.57	172.07	171.63	138.49	143.66
200	259.03	222.47	240.61	186.99	204.62
300	297.03	257.30	288.56	220.11	259.17
400	351.84	282.14	315.80	251.28	289.63
500	416.49	321.25	339.23	276.22	311.33
600	490.18	330.89	372.65	292.06	342.11
700	430.91	351.81	388.02	336.80	378.17
800	482.58	389.75	416.18	359.38	434.89
900	508.72	421.46	452.30	366.68	409.52
1000	505.50	419.10	440.84	390.98	445.71

表 13 障碍组样本的趋势波动分析

Tab. 13 The results of DFA for disorder group

窗口长度	[ei]	[i:]	[æ]	[ɔ:]	[a:]
100	169.43	175.04	147.93	158.55	150.81
200	221.12	238.28	201.81	205.02	205.35
300	273.78	278.61	246.69	260.16	252.26
400	293.99	308.92	308.24	278.89	313.12
500	331.48	346.42	327.41	333.32	361.92
600	365.19	406.95	343.21	351.28	383.76
700	434.36	425.03	360.81	412.51	429.51
800	437.31	468.80	406.59	439.53	456.89
900	433.52	555.66	427.52	484.62	465.22
1000	499.69	438.46	471.80	482.60	503.16

针对趋势波动分析,元音 [ei] [æ] [ɔ:] 在两组数据中表现出了明显的特异性,可以作为诊断依据做进一步分析,而 [i:] [a:] 差异不明显。该结论符合语音信号分析中的共振腔模型。

### 3 结论

经过对所得数据的分析处理,进一步阐明了各语音特征的物理意义及其应用特点,为基于语言障碍检测的帕金森病早期诊断提供了可靠依据。在证明语音特征用于帕金森病诊断有效性的同时,分析了不同特征下不同元音的类间分离度,为基于语音障碍的帕金森病自动诊断奠定了基础。

在后续的工作中,除了加强语音特征的数量以达到更好的检测效果外,在简化实验过程的前提下,引入真实的各阶段帕金森患者进行语音样本采集是重要的工作方向。

### 参考文献

[1] Lang AE,Lozano AM. Parkinson's disease-first of two parts [J].

New England Journal of Medicine ,1998 ,339:1044 – 1053.

[2] Zhang Zhenxin ,Gustavo CR , Zhen Hong ,et al. Parkinson ' s disease in China: prevalence in Beijing ,Xian ,and Shanghai [J]. The Lancet 2005 ,365(9459) :595 – 597.

[3] 张振馨. 神经系统疾病流行病学调查方法和问题 [J]. 中华神经科杂志 2005 ,38(2) :65 – 66.

[4] 韩艳 ,张晓红 ,陈彤 ,等. 帕金森病诊治现状调查 [J]. 中华保健医学杂志 2008 ,10(1) :18 – 20.

[5] Duffy JR. Motor speech disorders: substrates , differential diagnosis ,and management [M]. Saint Louis: Elsevier Mosby , 2005:80 – 142.

[6] Sapir S ,Spielman JL ,Ramig LO ,et al. Effects of intensive voice treatment on vowel articulation in dysarthric individuals with idiopathic parkinson disease: acoustic and perceptual findings [J]. J Speech Lang Hear Res 2007 ,50:899 – 912.

[7] Little MA ,Patrick EM ,Eric JH ,et al. Suitability of dysphonia measurements for telemonitoring of Parkinson ' s disease [J]. IEEE Transactions on Biomedical Engineering ,2009 ,56(4) : 1015 – 1022.

[8] Little MA ,Patrick EM ,Stephen JR ,et al. Exploiting nonlinear recurrence and fractal scaling properties for voice disorder detection [J]. BioMedical Engineering OnLine ,2007 ,6(1) :1: 19.

[9] Guo Peifang ,Prabir B ,Nawwaf K. Advances in detecting parkinson ' s disease [J]. Lecture Notes in Computer Science , 2010 ,6165:306 – 314.

[10] 张涛 ,洪文学 ,李铭婷. 基于多维筛分类器的可视化帕金森病诊断 [J]. 燕山大学学报 2010 ,34(2) :180 – 184.

[11] Li Xi ,Tao Jidong ,Michael TJ. Stress and emotion classification using jitter and shimmer features // ICASSP 2007 , Hawaii: IEEE 2007:1081 – 1084.

[12] Paul B. Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound [J]. J Acoustical Society of America ,1993 ,17:97 – 110.

[13] Kantz H ,Schreiber T. Nonlinear time series analysis [M]. Cambridge:Cambridge University Press ,1999:39 – 154.

[14] Andre G A. Modeling prosodic different for speaker recognition [J]. Speech Communication 2007 ,49:277 – 291.

(上接第 475 页)

[4] 王永红. 经络的实质探析 [J]. 世界中医药 ,2010 ,5(4) : 236 – 237.

[5] 王研. EIT 固定参考点图像重建和电极结构优化方法研究 [D]. 北京:中国协和医科大学 ,2002: 1 – 5.

[6] Polydorides N , Lionheart WRB. A Matlab toolkit for three-dimensional electrical impedance tomography:a contribution to the electrical impedance and diffuse optical reconstruction software project [J]. Measurement Science and Technology , 2002 ,13:1871 – 1883.

[7] 聂守平 ,魏晓燕. 数字图像的奇异值分解 [J]. 南京师大学报 ,2001 ,24(1) : 59 – 61.

[8] 黄小为 ,吴传生 ,朱华平. 基于奇异值分解建立的一种新的正则化方法 [J]. 数学物理学报 ,2005 ,25A(3) : 331 – 336.

[9] 黄小为 ,吴传生 ,李卓球. TSVD 正则化方法的参数选取及数值计算 [J]. 华中师范大学学报 ,2006 ,40(2) : 154 – 157.

[10] 董超峰 ,孙方裕. 基本解方法求解各向异性材料中热传导方程的时间反向问题 [J]. 浙江大学学报 ,2007 ,34(1) : 33 – 39.