构音障碍话者与正常话者发音的比较分析

原 梦　王洪翠　王龙标　党建武

摘要 言语障碍话者是指由于脑神经中枢、周围神经或末梢器官受损造成无法正确发音的特殊人群，在临床表现中脑瘫是其中之一。TORGO 数据库采集了正常话者和脑瘫引起的构音障碍话者的发音数据。本文对 TORGO 数据库进行了详细的标注、整理，对构音障碍话者和正常话者语音数据及电磁式发音动作数据进行了整合。从声学分析和发音运动分析两方面对比了构音障碍话者和正常话者的发音特点。在声学分析方面，利用 Mel 频率倒谱系数对音频数据进行特征提取，在此基础上，从连续语音中提取特定音素，利用隐马尔可夫模型从声学模型参数对说话人发特定音素进行分析。在发音运动方面，从连续语音中提取特定音素，对说话人发特定音素的发音空间位置变化情况开展研究。实验表明使用隐马尔可夫模型的方法可以有效区分正常话者和构音障碍话者；两者在发音运动方面也有显著差异，从侧面印证了人的发音机理。

关键词 言语障碍，发音空间运动，语音分析

Comparative Analysis of Articulation between Disorder and Normal Speakers

YUAN Meng　WANG Hongcui　WANG Longbiao　DANG Jianwu

Abstract Speech disorder speakers refer to people whose pronunciation is incorrect due to the brain central or peripheral nervous damage disease, in which cerebral palsy is one cause for speech disorder. The damage influences the control of the speech organs without impacting the comprehension ability. We refined the TORGO database by relabeling data, and conducted acoustic and articulatory analyses. On the acoustic analysis, we use the Mel Frequency Cepstrum Coefficient to extract feature parameters from the audio data. On this basis, we further extract specific phonemes from continuous speech and adopt the Hidden Markov Model method to analyze the parameters of acoustic model. The experiment results indicate that Hidden Markov Model method can effectively investigate the pronunciation mechanism of disorder speakers and normal speakers. On the analysis of moving trajectory, we analyze speaker's articular movements for several specific phonemes out of continuous speech. The results showed that the disorder speakers' tongue swings back and forth when they pronounce since they cannot control the tongue stably.

Key words Dysarthria, Articulatory movement, Speech analysis

1. 引言

言语障碍是由于先天性和外伤性神经运动障碍使人无法正常发音。这些损伤会影响说话人对发音器官的正常控制，但对于其他人话语可以正常理解，也可以产生有意义的且语法正确的语言。脑瘫作为言语障碍的一种，在北美儿童中约占 0.5%，其中 88%在成年后依旧具有言语障碍[1]。在我国，脑瘫的发病率为 0.186%-0.4%，并且发生率仍

有上升趋势[2]，目前，医学角度对言语障碍的评估主要是 Frenchay 构音障碍评定法[3]和中国康复研究中心制定的构音障碍评定法，这些方法依赖于医生的主观判断，缺少量化度量[4]。在对障碍话者的语音研究中多以声学特征为主要分析对象，缺乏对于发音器官运动情况的分析[5]。因此，对于言语障碍的康复缺少客观、可视化的指导。

2. 数据处理

本文中我们采用的是 TORGO 数据库。该数据库采集利用完全自动化校准的 3DAG500 电磁式发音运动记录系统（EMA）采集运动空间数据和同步的声学数据。音频的采样率为 16 000Hz，位置信息的采样率为 200Hz。TORGO 数据库中包含了大约 23 个小时的英语语音数据，这些数据是从 8 个患有脑瘫或者肌萎缩侧索硬化症的言语障碍话者和 7 个正常话者获取的[6]。语料由非词短（要求说话者重复/ iy-p-ah /，/ ah-p-iy /，/ p-ah-t-ah-k-ah /）、短语和限制句、非限制句组成。

2.1 数据标注

本文中的语音研究基于音素级，利用 praat 软件对音频信号进行手工标注。我们分别标注了两个构音障碍话者（两名女性）和两个正常话者（一名女性、一名男性）的数据。标注的音素标号是基于语音识别中常用的 Arpabet 标号，共有 39 个。标注好后对文件进行转码，提取标注信息，包括音素名、起止时间。

表 1: 标注数据的详细信息。

标号	DF01	DF02	NF01	NM02
类型	构音障碍		正常	
性别	女性	女性	女性	男性
语句数	146	236	383	615
音素数	1 425	2 296	3 925	5 263

2.2 特征提取

Mel 频率倒谱系数(MFCC)是基于人耳听觉频域特性，将线性幅度谱映射到基于听觉感知的 Mel 非线性幅度谱中，再转换到倒谱上。有以下步骤：

步骤 1——预加重：将一组语音信号 s(n)通过高通滤波器。

高通滤波器关系可以表示为：

$$H(z) = 1 - az^{-1}([a \in [0.9,1]) \quad (1)$$

经过预加重后的信号 $s'(n)$ 表示为：

$$s'(n) = s(n) - as(n-1) \quad (2)$$

本文中 a 值取 0.95。

步骤 2——加窗：本文中取 20ms 为一帧，由于帧边界处频谱能量的可能存在泄漏情况，对每一帧都进行加窗处理，本文中我们选用汉宁窗。

步骤 3——快速傅里叶变换（FFT）：对每一帧进行 FFT 变换，从时域数据转变为频域数据，并计算其谱线能量。

步骤 4——Mel 滤波：把求出的每帧谱线能量通过 Mel 滤波器，并计算在 Mel 滤波器中的能量。

步骤 5——计算 DCT 倒谱：把 Mel 滤波器的能量取对数后计算 DCT，就可以得到 Mel 频率倒谱系数 MFCC。

本文中提取 MFCC 为 14 阶，未提取其一阶差分系数和二阶差分系数。特征值维度为 14 阶。

3. 声学分析

我们选择单元音 ɔ、ɑ、i、u、ɛ、ɪ、ʊ、ʌ/ə、æ 和双元音 eɪ、aɪ、oʊ、aʊ、ɔɪ 作为实验研究的音素。训练高斯混合隐马尔可夫模型（GMM-HMM）来区别正常话者和构音障碍话者，分别对将对应音素的 MFCC 特征值作为模型的观察序列，采用 10 倍交叉验证的方式，对于每个音素都训练出正常话者的 GMM-HMM 模型和构音障碍话者的 GMM-HMM 模型。在本文中，我们的高斯混合模型中使用 8 个模型，训练迭代次数为 10 次。隐马尔科夫模型是无跨越的从左向右模型，它的状态数为 3 个，训练的迭代次数为 10 次。

准确率、敏感性和特异性作为评测分类

表 2: 单元音识别结果。

标号	AO	AA	IY	UW	EH	IH	UH	AH	AE
IPA	ɔ	ɑ	i	u	ɛ	ɪ	ʊ	ʌ/ə	æ
准确率	84.10%	92.53%	95.25%	88.26%	88.73%	93.23%	66.67%	91.82%	90.53%
特异性	100%	100%	99.24%	98.65%	100%	96.97%	85.19%	95.73%	99.47%
敏感性	41.54%	69.88%	98.17%	64.62%	61.67%	85.25%	25%	82.61%	68.83%

结果的指标。构音障碍话者代表正样本，正常话者代表负样本。准确率是真正数与真负数的和除以所有测试样品的数量。敏感性指标可以反映出构音障碍话者被诊断为正常话者的概率。特异性指标反映了正常话者被诊断为构音障碍话者的概率。

$$\text{准确率} = \frac{\text{真正数} + \text{真负数}}{\text{样本总数}} \quad (3)$$

$$\text{特异性} = \frac{\text{真负数}}{\text{真负数} + \text{假正数}} \quad (4)$$

$$\text{敏感性} = \frac{\text{真正数}}{\text{真正数} + \text{假负数}} \quad (5)$$

表 3: 双元音识别结果。

标号	AY	OW	AW
IPA	aɪ	oʊ	aʊ
准确率	92.54%	87.66%	72.22%
特异性	99.46%	99.40%	100%
敏感性	76.83%	59.42%	0%

标号	OY	EY
IPA	ɔɪ	eɪ
准确率	76.19%	90.94%
特异性	100%	100%
敏感性	0%	64.06%

表 2 给出了单元音音素识别结果，表 3 给出了双元音音素识别结果。这两个表中，由于“UH”、“AW”、“OY”这三个音素障碍话者的训练样本数低于 10 个，数量太少，模型训练不充分，造成敏感性结果较差。

4. 发音运动分析

在这个数据库中，为采集空间数据，传感器线圈的三个点被贴到舌表面，即舌尖（TT - 舌尖后 1 厘米处）、舌中（TM - 舌尖传感器后 2 厘米）和舌后（TB - 舌中传感器后 3 厘米），所有传感器测量声带运动通常都在矢状切面上，本文中我们选用 TT、TM、TB 这三点关键点矢状切面上的位置信息来进行分析。

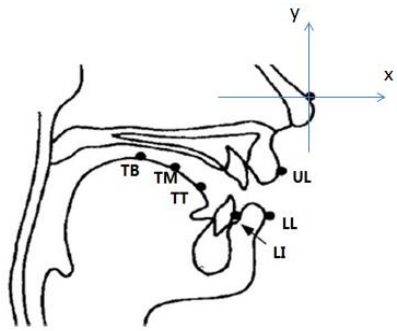

图 1: 关键点传感器位置分布。

空间分析是从正常话者和构音障碍话者的发音中提取特定音素，研究出其在发音过程中舌头位置的变化情况，为了能够清楚地看到发音过程中舌头的变化过程，我们首先用舌头上三点传感器的采样值进行研究，发音时候的空间分布由空间舌头的关键点和时间信息的动态特性表示。

从表 2 中，我们发现标号为“AO”与“EH”单元音音素的声学分析对于区别正常话者和构音障碍话者的效果不显著，我们要探究在发音运动方面是否能区别正常话者

和构音障碍话者。单元音音素“AO”、“EH”，对应的国际音标表示为“ɔ”、“ɛ”。

观察发音位置时，一个元音通常能够用声道中的收缩位置表示，我们通过传感器的值观察正常话者和构音障碍话者在发这些音时候的舌面高度情况。为了使结果具有可比较性，我们分析女性构音障碍话者 DF02，并以女性正常话者 NF01 发音进行对比。同时，我们为了消除连续语音中选定研究的音素受到前一音素的影响，我们选择出前一音素是静音时刻，也就是说选出以此元音音素开始的语料进行分析，同时，也保证正常话者和构音障碍话者后一音素相同，在分析“AO”时，后一音素为“L”，对应的国际音标表示为“l”；在分析“EH”时，后一音素为“V”，对应的国际音标表示为“v”。

图中的横轴代表的是口腔从前到后的水平方向，纵轴代表的是舌位的高度。黑色粗线是舌头上三个观测点的位置变化，由左至右分别为舌根、舌中、舌尖，同时以鼻尖传感器的值基准点，所有舌头传感器的值需要减去鼻尖传感器的值，差值之后的第一个点为原点，其他差值点与原点的相对位置进行画图，黑色细线为每一帧的三个观测点的连线情况，代表此时舌头的轮廓。

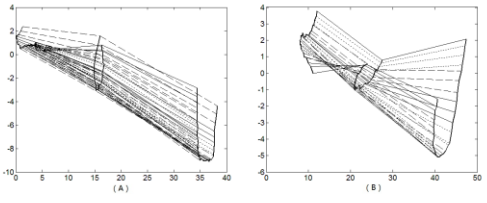

图 2：正常话者发“AO”舌面变化图。

再来看构音障碍话者 DF02 发“AO”音的情况。

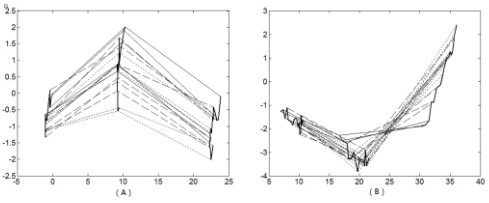

图 3：构音障碍话者发“AO”舌面变化图。

构音障碍话者和正常话者发“AO”相比，对于舌头控制得不好，对于 DF02 来说，这名患者的舌头运动从图 3（A）和图 3（B）中间黑色粗线条可以明显看出运动波动非常大，前一时刻的位置和下一时刻的位置相差非常远，图 3（A）出现很明显的来回抖动现象。正常话者的抖动情况氛围十分小，如图 2（A）和（B）黑色粗线条的变化基本呈现平稳状态。同时，在图 2 中正常话者在发此音素时，由横轴可看出舌头的运动范围广泛灵活，舌头的位置趋势为由舌根到舌尖降低，但是在图 3 中构音障碍话者的发音则呈现出较为混乱的状态，图 3（A）舌头呈现出倒 U 状，图 3（B）舌头呈现出 U 状。

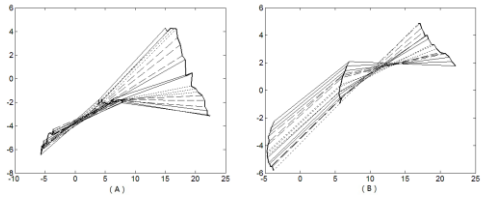

图 4：正常话者发“EH”舌面变化图。

再来看构音障碍话者 DF02 发“EH”音的情况。

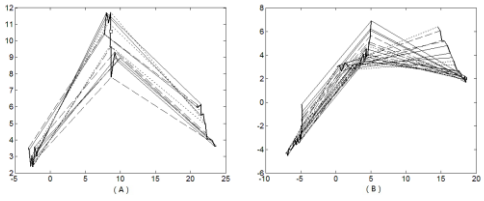

图 5：构音障碍话者发“EH”舌面变化图。

从这个音素的分析中我们可以得出与上一图相同的结论。即，从图 5（A）和图 5（B）中间黑色线条可以看出构音障碍话者在舌头前一位置和下一位置的即黑色粗线条波动非常大，出现很明显的来回抖动现象，但是图 4（A）和图 4（B）中正常话者的黑色粗线条抖动情况范围十分小，变化基本呈现平稳状态。同时我们可以看到，图 4 中正常话者在发此音素时舌头的位置趋势为由舌根到舌尖增高，但图 5 中构音障碍话者在发此音素时，舌头呈现的状态是倒 U 状。

与此同时，对比图 2 与图 4，我们可以发现，对于正常话者来说，在发不同音素时，舌头的形状和轮廓是不同的。但对于构音障

碍话者来说，图 3 与图 5 在发不同音素时，舌头的位置和轮廓无规律。

5. 结语

本文通过对 TROGO 数据库的标注整理，对构音障碍话者的数据和正常话者的声学数据和 EMA 数据进行了整合，可以作为今后同类研究的基础。在此基础上，从声学分析和空间运动分析两方面对比构音障碍话者和正常话者的数据。

在声学分析上，我们在数据预处理阶运用了 Mel 频率倒谱系数对音频数据提取特征，将这些特征作为观察数据训练出正常话者和构音障碍话者的元音的 HMM 模型，并对模型进行测试。

在发音运动分析上，其结果显示构音障碍话者和正常话者在发特定音素时，正常话者的舌头在发音时会有轻微抖动现象，但是构音障碍话者的舌头发音时有剧烈的上下抖动现象，且变化范围很大；正常话者的舌头发相同音素时，舌头的位置状态基本相同，但是构音障碍话者的舌头位置状态不稳定，呈现多种类型。

正常话者在发不同音素时，发音运动形状和轮廓区别很大，构音障碍话者在发不同音素时，发音运动形状和轮廓很难区别。

由于标注数据较少，HMM 模型在某些音素上效果不好；现阶段只是将声学分析和空间分析隔离开进行，接下来，我们希望能建立合适的模型将声学数据和空间分析结合起来；此外，我们目前的分析是元音中的部分音素，对于元音中的其他音素和辅音都还要做进一步的分析。

6. 致谢

本研究得到国家重点基础研究发展计划（973 计划）项目（编号：2013CB329301）和国家自然科学重点基金项目（编号：61233009 和编号 61303109）的经费支持。

7. 参考文献

[1] Hasegawa-Johnson, M., Gunderson, J., Huang, T. 2006. *Audiovisual Phonologic-Feature-Based Recognition of Dysarthric Speech*. Johnson.

[2] 林庆、李松（2004）《小儿脑性瘫痪》第 2 版。北京：北京医科大学出版社。

[3] Pamla, M., Endreby. 1983. Frenchay Dysarthria Assessment. California: College-Hill Press, 34~ 53

[4] Selouani, S. A., Dahmani H., Amami R., et al. 2012, Using speech rhythm knowledge to improve dysarthric speech recognition. *International Journal of Speech Technology*, 15(1): 57-64.

[5] Rudzicz, F. 2011. Articulatory knowledge in the recognition of dysarthricspeech. *IEEE Tran- sactions on, Audio, Speech, and Language Processing*, 19(4): 947-960.

[6] Kent, R.D., Rosen, K. 2004. Motor control perspectives on motor speech disorders. In: Maassen B., Kent R.D., Peters H., van Lieshout P., Hulstijn W. (eds.) *Speech Motor Control in Normal and Disordered Speech*, Oxford University Press, Oxford, chap.12, 285-311.

原 梦 天津大学天津市认知计算与应用重点实验室，在读硕士，主要研究领域为病理语音处理。
E-mail: yuanmeng@tju.edu.cn

王洪翠 天津大学天津市认知计算与应用重点实验室讲师，博士，主要研究领域为语音信息处理。
E-mail: hcwang@tju.edu.cn

王龙标 天津大学天津市认知计算与应用重点实验室教授，博士，主要研究领域为语音识别，声学信号处理。
E-mail: longbiao_wang@tju.edu.cn

党建武 天津大学天津市认知计算与应用重点实验室教授，博士，主要研究领域为人的语音生成和感知机理、语音认识的研究、语音个人特性的识别与合成、言语康复的研究。
E-mail:dangjianwu@tju.edu.cn