

DOI:10.11798/j.issn.1007-1520.202423050

· 论著 ·

梅尔频率倒谱系数在声带息肉手术 前后嗓音分析中的价值研究

刘茉,葛鑫颖,赵晓畅,郝青青,李祖飞

(首都医科大学附属北京朝阳医院耳鼻咽喉头颈外科,北京 100020)

摘要: **目的** 本研究拟通过提取患者嗓音中的梅尔频率倒谱系数(MFCC)指标,探讨其在声带息肉手术前后嗓音分析中的临床价值。**方法** 回顾性分析于2018年1月—2019年8月行声带息肉手术且术前及术后1个月均行嗓音评估的患者41例,男31例,女10例;平均年龄(42.9 ± 11.4)岁。另选取无声嘶且无声带病变的正常受试者21例作为基线对照。使用基于Python编程语言的librosa语音处理包进行MFCC特征提取,分别提取每位患者的MFCC均值、MFCC方差与MFCC标准差,使用配对样本 t 检验比较声带息肉手术前后上述各MFCC特征的差异。**结果** 声带息肉患者术后MFCC均值 1.25 ± 1.01 、MFCC方差 561.34 ± 154.98 及MFCC标准差 21.74 ± 4.03 比术前MFCC均值 6.81 ± 2.05 、MFCC方差 1019.66 ± 295.87 及MFCC标准差 34.37 ± 6.63 显著下降,差异具有统计学意义($t=18.596, P=0.000; t=10.338, P=0.000; t=11.852, P=0.000$)。声带息肉组患者术后1个月其MFCC均值、MFCC方差及MFCC标准差与正常受试者相比差异均无统计学意义,表明绝大部分声带息肉患者术后嗓音得到良好的恢复。**结论** 本研究首次探索了MFCC在声带息肉手术前后嗓音分析中的价值, MFCC各特征可作为评估声带息肉术后嗓音恢复的指标。

关键词: 声带息肉;声嘶;梅尔频率倒谱系数;嗓音分析;手术

中图分类号:R767.92

Value of Mel frequency cepstrum coefficient in voice analysis before and after vocal polyp surgery

LIU Mo, GE Xinying, ZHAO Xiaochang, HAO Qingqing, LI Zufe

(Department of Otorhinolaryngology Head and Neck Surgery, Beijing Chaoyang Hospital, Capital Medical University, Beijing 100020, China)

Abstract: **Objective** Mel frequency cepstrum coefficient (MFCC) has a wide range of applications in the field of speech recognition, but its application has not been reported in the field of voice analysis at home and abroad. This study intends to analyze its research value in voice analysis before and after vocal polyp surgery by extracting the MFCC index. **Methods** A total of 41 patients who underwent vocal polyp surgery in our hospital from January 2018 to August 2019 and received voice evaluation before and 1 month after the surgery were retrospectively analyzed. In addition, 21 normal subjects who had neither hoarseness nor vocal cord lesions were selected as the baseline control. The librosa speech processing package based on Python programming language was used for MFCC feature extraction. The mean value, variance and standard deviation of MFCC in each patient were extracted respectively. The paired sample t -test was used to compare the differences of the above MFCC features before and after surgery. **Results** The enrolled patients included 31 males and 10 females with an average age of (42.9 ± 11.4) years. Their postoperative mean value, variance and standard deviation of MFCC got decreased significantly compared to the preoperative ones (6.81 ± 2.05 vs 1.25 ± 1.01 , $t=18.596$, $P=0.000$; 1019.66 ± 295.87 vs 561.34 ± 154.98 , $t=10.338$, $P=0.000$; 34.37 ± 6.63 vs 21.74 ± 4.03 , $t=11.852$, $P=0.000$). The differences of mean value, variance and standard deviation of MFCC between the patients one month after surgery and the normal subjects were statistically insignificant, indicating that the voice of most patients recovered well after surgery. **Conclusions** This study is the first to explore the value of MFCC in voice analysis before and

after vocal polyp surgery. The characteristics of MFCC can be used as indexes to evaluate the voice recovery after vocal polyp surgery.

Keywords: Vocal polyp; Hoarseness; Mel frequency cepstrum coefficient; Voice analysis; Surgery

声带息肉是成人最常见的引发声音嘶哑的疾病,是声带固有层浅层的良性局限性病变,疾病早期可以通过休声、嗓音矫治等方法进行治疗,声带病变较严重的患者,一般需要通过手术治疗恢复正常嗓音^[1]。嗓音分析是评估声嘶及其治疗后嗓音功能恢复的重要手段之一,既往对嗓音质量的客观声学分析主要采用基频、基频微扰、振幅微扰、标准化噪声能量和谐噪比等指标进行研究和评价^[1-2],此外,一些基于傅里叶变换等理论基础而形成的嗓音软件,如美国 Tiger 公司研发的 Dr. Speech 也得到一定程度的应用^[3]。

随着计算机技术的逐渐发展,研究人员提出了更多的语音评估指标,如梅尔频率倒谱系数(Mel-frequency cepstral coefficients, MFCC)。MFCC 目前被广泛应用基于人工智能的语音识别、音频特征分析等方面^[4-5]。MFCC 在嗓音医学领域,尤其是声带息肉手术前后声嘶症状变化的评估方面,国内外尚属于空白,本研究通过对声带息肉患者术前术后和正常人群的嗓音 MFCC 各指标进行差异分析,挖掘 MFCC 在声带息肉预后中的价值研究,初步探讨 MFCC 在嗓音医学领域的研究价值,为嗓音医学分析提供新的评价指标。

1 资料与方法

1.1 临床资料

回顾性分析 2018 年 1 月—2019 年 8 月在首都医科大学附属北京朝阳医院耳鼻咽喉头颈外科收治的声带息肉患者共 41 例,男 31 例,女 10 例;平均年龄为(42.9 ± 11.4)岁。所有患者均在术前行声音采样备用,手术均采用经口插管全身麻醉,使用德国 STORZ 公司的支撑喉镜暴露声门,在显微镜下用 CO₂ 激光(1~2 W)行病变切除,术后病变送病理回报均为符合声带息肉,术后患者予以布地奈德混悬液(1 mg)2 支,每天 2 次雾化吸入及口服广谱抗生素 2~3 d 后出院,术后常规禁声 3 d。术后 1 个月行嗓音声学评估和电子喉镜检查。

同期选取了无吸烟酗酒、嗓音疾病、听力及构音正常和近期无上呼吸道感染的正常成年人 21 例作为对照组,男 9 例,女 12 例;平均年龄为(36.3 ±

14.5)岁。声带息肉组患者与对照组在年龄方面差异无统计学意义($t = -1.826$, $P = 0.077$)。声带息肉组与对照组相比男性比例略高($\chi^2 = 6.507$, $t = 0.011$)。

1.2 声学录音采样

采用德国 LingWAVES 嗓音分析系统的专业级声级计对受试人员声音进行采样,检查室环境噪声要求 <45 dB,受试者采用自然舒适坐位,双脚平放,声级计用三角支架固定,放置于受试者正前方,略倾斜 45°角,与受试者口唇距离保持 10~15 cm。录音前先指导受试者如何正确发元音/a:/音并持续 3 s 以上,录音成功后取稳定发音的 2 s 音频样本进行分析。所有音频资料以 wav 格式进行保存。

1.3 MFCC 特征提取

采用基于 Python 编程软件(3.9 版本)的 librosa 语音处理包对所有 wav 格式语音文件进行分析并提取其 MFCC 特征,采用 matplotlib 包进行 MFCC 频谱图绘制,由于每一个患者其 MFCC 提取到的是一系列系数矩阵,无法直接用于量化分析,因此,采用 pandas 包计算每一例患者的 MFCC 均值、MFCC 标准差及 MFCC 方差,这 3 个参数将作为每个患者独立的评价嗓音的 MFCC 量化指标,并进行进一步的组间比较与分析。

1.4 统计学分析

采用 SPSS 20.0 软件进行统计学分析,声带息肉术前及术后 MFCC 各指标差异采用配对样本 t 检验分析,声带息肉患者术后与正常受试者其 MFCC 各指标差异采用独立样本 t 检验分析,计数资料组间比较采用四格表 χ^2 检验,以 $P < 0.05$ 为差异具有统计学意义。

2 结果

2.1 声带息肉患者术前与对照组 MFCC 系数比较

对照组其 MFCC 均值、MFCC 方差及 MFCC 标准差等 MFCC 量化指标,采用配对样本 t 检验分析,均显著低于声带息肉患者,差异具有统计学意义。具体数据见表 1。

表1 声带息肉组患者术前与对照组 MFCC

参数差异分析 ($\bar{x} \pm s$)

MFCC 量化指标	声带息肉组术前 (n=41)	对照组 (n=21)	t	P
MFCC 均值	6.81 ± 2.05	1.36 ± 1.42	-10.888	0.000
MFCC 方差	1 019.66 ± 295.87	666.68 ± 227.94	-5.199	0.000
MFCC 标准差	34.37 ± 6.63	22.89 ± 3.92	-7.293	0.000

注:MFCC(梅尔频率倒谱系数)。下同。

2.2 声带息肉患者手术前后 MFCC 差异分析

所有声带息肉患者均进行嗓音声学分析,并成功提取 MFCC 均值、MFCC 方差及 MFCC 标准差,与术前相比,采用配对样本 t 检验分析声带息肉术后患者其 MFCC 均值、MFCC 方差及 MFCC 标准差均显著下降,差异具有统计学意义。具体数据见表2。

图1A及图1B分别展示了术前及术后1个月声带息肉患者嗓音 MFCC 频谱图,可见声带息肉患者术后嗓音 MFCC 频谱图各频带清晰,而术前嗓音 MFCC 频谱图各频带较为粗糙,不同频带间界限较为模糊。

2.3 声带息肉术后患者与对照组 MFCC 差异分析

41例声带息肉患者术后1个月的 MFCC 均值、MFCC 方差及 MFCC 标准差与对照组相比,差异均无统计学意义,具体数据见表3。表明绝大部分声带息肉患者术后嗓音得到良好的恢复,接近正常人嗓音。

图1B与图1C分别展示了声带息肉组患者术后与对照组其嗓音 MFCC 频谱图。

表2 41例声带息肉组患者术前及术后各 MFCC

参数差异分析 ($\bar{x} \pm s$)

MFCC 量化指标	术前	术后	t	P
MFCC 均值	6.81 ± 2.05	1.25 ± 1.01	18.596	0.000
MFCC 方差	1 019.66 ± 295.87	561.34 ± 154.98	10.338	0.000
MFCC 标准差	34.37 ± 6.63	21.74 ± 4.03	11.852	0.000

3 讨论

既往研究表明人耳感知的声音频率和声音的实际频率并不是线性的,梅尔刻度就是依据人的听觉实验结果来分析音频的频谱,MFCC就是在梅尔刻度频率域提取出来的倒谱系数^[6]。其原理为:使用计算机技术,将语音信号从低频到高频的不同频率条带,按照临界条带宽度的大小,使用带通滤波器由密到疏对语音信号进行滤波处理,可对不同频带信号进行能量输出,得到信号能量基本特征,通过对上述特征的进一步处理,即可得到 MFCC 频谱图,并统计 MFCC 均值、方差及标准差等参数用于统计分析或人工智能模型构建^[7]。

目前 MFCC 在与音频相关的医学领域亦有不同程度的应用,如呼吸音、心音等。Palaniappan 等^[8]通过搜集患者的呼吸音,对呼吸音进行 MFCC 特征提取,采用机器学习算法建立识别肺部疾病的模型;Zhou 等^[9]通过 MFCC 特征提取,发现足月新生儿与早产儿间存在呼吸音的差异;Deng 等^[4]通过对心音的 MFCC 特征提取,联合深度学习算法,建立预测心脏疾病的人工智能模型。此外,有学者通过对患者的语音分析,提取 MFCC 等音频特征,建立了阻塞性睡眠呼吸暂停低通气综合症的预测模型,其模型受试者操作特征曲线下面积达 83% ~ 87%^[10-11]。有关 MFCC 在嗓音医学领域的研究,国外 Anumanchi-

表3 声带息肉组患者术后与对照组各 MFCC

参数差异分析 ($\bar{x} \pm s$)

MFCC 量化指标	声带息肉组 术后(n=41)	对照组 (n=21)	t	P
MFCC 均值	1.25 ± 1.01	1.36 ± 1.42	-0.369	0.714
MFCC 方差	561.34 ± 154.98	666.68 ± 227.94	-1.904	0.067
MFCC 标准差	21.74 ± 4.03	22.89 ± 3.92	-1.071	0.289

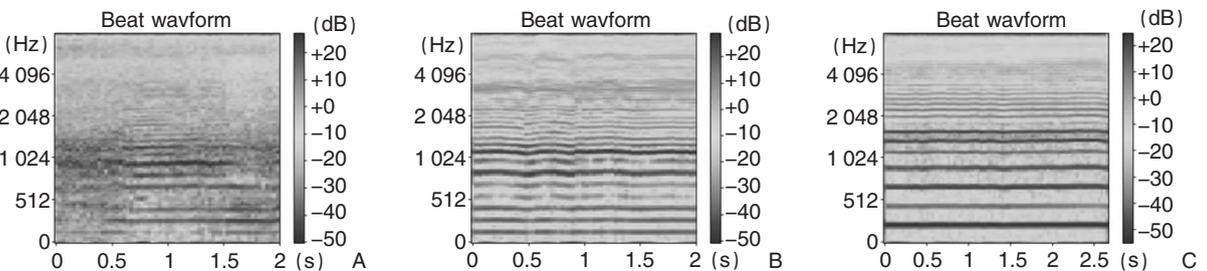


图1 嗓音 MFCC 频谱图 A:声带息肉组患者术前嗓音 MFCC 图; B:同一患者声带息肉术后1个月嗓音 MFCC 图; C:对照组嗓音 MFCC 图

palli 等^[12]将发声的运动学特征使用 MFCC 将其解码为声学特征,最终将颅内的脑电活动翻译生成语音。然而关于 MFCC 在临床嗓音评估方面的研究,尚未见报道。本研究首次探讨了 MFCC 在声带息肉手术前后嗓音功能变化中的研究价值,其主要参数为 MFCC 均值, MFCC 方差与 MFCC 标准差,这 3 种 MFCC 特征在声带息肉手术后均显著下降,且术后 1 个月患者均接近正常嗓音,与对照组相比,上述 MFCC 特征均无显著差异,提示 MFCC 或可作为一种新型的嗓音医学评估手段,用于声带病变的嗓音声学评价。

本文所用的 MFCC 量化指标主要为 MFCC 均值, MFCC 方差与 MFCC 标准差,声嘶患者上述各指标显著上升,由于 MFCC 在嗓音医学领域研究几乎为空白,因此,其机制尚不清楚。MFCC 系数在提取后会生成一个系数矩阵,而 MFCC 均值是单个语音信号的 MFCC 系数的平均值,其标准差及方差即为 MFCC 标准差与 MFCC 方差可表示 MFCC 系数的变化程度,可以用于反映语音信号的变化情况^[6-7]。从声带息肉术前与术后以及正常受试者语音的 MFCC 频谱图对比发现,术前患者各频率条带模糊,边界不清,这可能与声带息肉病变导致的黏膜波幅度改变、声带振动振幅不对称、声门闭合不全^[13]有关,其频率条带清晰度或可反应患者的声嘶程度。MFCC 提取后可根据不同频段的清晰度得到一组数据矩阵,其频率条带越粗糙,数据矩阵上的各数值越大,所提取到的 MFCC 均值、MFCC 方差与 MFCC 标准差也越大。因此,上述参数可能可以反应 MFCC 频谱图上各频率条带而间接反应嗓音的粗糙程度。其具体机制有待进一步研究。

本研究也存在以下几点局限:①本研究纳入样本量相对较少,且为单中心数据,未来还需扩大样本量进行深入研究;②本研究仅选择声带息肉患者作为研究对象,纳入的声带息肉患者术后大多嗓音恢复接近正常人嗓音, MFCC 在声带白斑与早期声门型喉癌等其他声带病变中的嗓音声学分析价值也有待进一步实验深入探索。

综上所述,本研究首次探索了 MFCC 在声带息肉患者手术前后嗓音声学评估中的研究价值,本文提出的 MFCC 均值、MFCC 方差与 MFCC 标准差等量化参数,或可作为声带病变的新型嗓音声学评估指标,这些指标值得在未来的嗓音医学领域深入探索,为嗓音医学研究提供全新的视角。

参考文献:

- [1] Karlsen T, Sandvik L, Heimdal JH, et al. Acoustic voice analysis and maximum phonation time in relation to voice handicap index score and larynx disease[J]. J Voice, 2020, 34(1):161. e27 - 161. e35.
- [2] 徐萌,金晓彤,孙毓哈,等. 早期声门型喉癌低温等离子射频消融术后联合嗓音训练的嗓音学分析[J]. 中国耳鼻咽喉颅底外科杂志, 2022, 28(2):95 - 98.
- [3] Kim HK, Gao SH, Yi B, et al. Validation of the dysphonia severity index in the Dr. Speech Program[J]. J Voice, 2019, 33(6):948. e23 - 948. e29.
- [4] Deng M, Meng T, Cao J, et al. Heart sound classification based on improved MFCC features and convolutional recurrent neural networks[J]. Neural Netw, 2020, 130:22 - 32.
- [5] Iqtidar K, Qamar U, Aziz S, et al. Phonocardiogram signal analysis for classification of Coronary Artery Diseases using MFCC and 1D adaptive local ternary patterns[J]. Comput Biol Med, 2021, 138:104926.
- [6] Hao X, Bao Y, Guo Y, et al. Multi-modal neuroimaging feature selection with consistent metric constraint for diagnosis of Alzheimer's disease[J]. Med Image Anal, 2020, 60:101625.
- [7] Fernando T, Ghaemmaghami H, Denman S, et al. Heart sound segmentation using bidirectional LSTMs with attention[J]. IEEE J Biomed Health Inform, 2019, 24(6):1601 - 1609.
- [8] Palaniappan R, Sundaraj K, Sundaraj S. A comparative study of the svm and k-nn machine learning algorithms for the diagnosis of respiratory pathologies using pulmonary acoustic signals[J]. BMC Bioinformatics, 2014, 15:223.
- [9] Zhou L, Marzbanrad F, Ramanathan A, et al. Acoustic analysis of neonatal breath sounds using digital stethoscope technology[J]. Pediatr Pulmonol, 2020, 55(3):624 - 630.
- [10] Ding Y, Sun Y, Li Y, et al. Selection of OSA-specific pronunciations and assessment of disease severity assisted by machine learning[J]. J Clin Sleep Med, 2022, 18(11):2663 - 2672.
- [11] Ding Y, Wang J, Gao J, et al. Severity evaluation of obstructive sleep apnea based on speech features[J]. Sleep Breath, 2021, 25(2):787 - 795.
- [12] Anumanchipalli GK, Chartier J, Chang EF. Speech synthesis from neural decoding of spoken sentences [J]. Nature, 2019, 568(7753):493 - 498.
- [13] Krasnodębska P, Szkielkowska A, Miaskiewicz B, et al. Objective measurement of mucosal wave parameters in diagnosing benign lesions of the vocal folds [J]. Logoped Phoniatr Vocol, 2019, 44(2):73 - 78.

(收稿日期:2023 - 02 - 26)

本文引用格式:刘茉,葛鑫颖,赵晓畅,等. 梅尔频率倒谱系数在声带息肉手术前后嗓音分析中的价值研究[J]. 中国耳鼻咽喉颅底外科杂志, 2024,30(2):102 - 105. DOI:10. 11798/j. issn. 1007 - 1520. 202423050

Cite this article as: LIU Mo, GE Xinying, ZHAO Xiaochang, et al. Value of Mel frequency cepstrum coefficient in voice analysis before and after vocal polyp surgery [J]. Chin J Otorhinolaryngol Skull Base Surg, 2024,30(2):102 - 105. DOI:10. 11798/j. issn. 1007 - 1520. 202423050