

[DOI] 10.12016/j.issn.2096-1456.2022.06.012

· 综述 ·

计算机科学应用于唇腭裂语音诊疗的研究进展

郭毅波^{1,3}, 蔡鸣^{2,3}

1. 上海交通大学口腔医学院, 上海(200125); 2. 上海交通大学医学院附属第九人民医院口腔颌面科, 上海(200011); 3. 国家口腔医学中心, 国家口腔疾病临床医学研究中心, 上海市口腔医学重点实验室, 上海(200011)

【摘要】 唇腭裂术后针对性的语音治疗是唇腭裂综合序列治疗的重要组成部分, 主观语音评价是目前应用最广泛的评价体系, 由于无法避免主观倾向的局限性, 精准的腭裂语音诊疗仍有进一步完善的空间。随着医工交叉的广泛推进, 计算机科学已成为前沿医疗的基石, 人工智能在言语障碍领域已开展初步的临床应用。对于腭裂语音的计算机辅助识别和评价技术已有较多研究成果, 辅音省略、咽擦音等代偿性发音错误是关键的分析特征, 结合软硬件技术的系统能达到较高的识别准确度; 但目前研究大多仅局限于独立的声学特征识别, 有待进一步开发多参数的分析函数来避免普遍评价的系统偏差; 且计算机科学在腭裂语音评价及治疗方面仍缺乏治疗方案的尝试, 在线诊疗平台未得到充分使用, 因此运用人工智能媒介与主观评价体系的结合应是唇腭裂综合序列治疗的前景模式之一。

【关键词】 唇腭裂; 语音评价; 腭裂语音; 声学特征; 人工智能; 在线平台

【中图分类号】 R78 **【文献标志码】** A **【文章编号】** 2096-1456(2022)06-0453-04

【引用著录格式】 郭毅波, 蔡鸣. 计算机科学应用于唇腭裂语音诊疗的研究进展[J]. 口腔疾病防治, 2022, 30(6): 453-456. doi: 10.12016/j.issn.2096-1456.2022.06.012.



微信公众号

Research progress of computer science in cleft lip and palate speech therapy GUO Yibo^{1,3}, CAI Ming^{2,3}. 1. College of Stomatology, Shanghai Jiao Tong University, Shanghai 200125, China; 2. Department of Oral & Cranio-Maxillo-facial Surgery, Shanghai Ninth People's Hospital, Shanghai Jiao Tong University School of Medicine; Shanghai 200011, China; 3. National Center for Stomatology; National Clinical Research Center for Oral Diseases; Shanghai Key Laboratory of Stomatology, Shanghai 200011, China

Corresponding author: CAI Ming, Email: zidanecm500@126.com, Tel: 86-13918490900

【Abstract】 In addition to surgical treatment, postoperative targeted speech therapy is an important part of the comprehensive team approach of cleft lip and palate (CLP). As the most widely used assessment for speech recognition, the main limitation of subjective assessments is its personal tendencies on results. Therefore, the precise system remains to be further improved. With the promotion of multidisciplinary crossing and integration, computer science has become the cornerstone of a high-quality health care system, and artificial intelligence (AI) has been applied to the field of speech disorders preclinically. Various studies focused on computer-aided recognition and evaluation technology of cleft palate speech, in which compensatory pronunciation errors, such as consonant ellipsis and pharyngeal fricative, were the key features for recognition and analysis. Although the combination with software and hardware improved the accuracy of recognition, current studies were primarily limited to independent acoustic feature identification, and a multiparameter analysis is needed to avoid systematic deviation. It lacks the attempt of treatment with computer science in speech evaluation of cleft palate, and the online platform has not been fully used. The use of AI and subjective evaluation systems is one of the prospect models of the comprehensive team approach of CLP.

【Key words】 cleft lip and palate; voice evaluation; cleft palate speech; acoustic feature; artificial intelli-

【收稿日期】 2021-08-12; **【修回日期】** 2021-09-20

【基金项目】 教育部高等教育司 2021 第二批产学研协同育人项目(202102457003); 国家重点研发计划(2016YFC1000503); 上海交通大学医学院附属第九人民医院临床研究型 MDT 项目(201701012)

【作者简介】 郭毅波, 住院医师, 博士生, Email: guoyibo98@163.com

【通信作者】 蔡鸣, 主任医师, 博士, Email: zidanecm500@126.com, Tel: 86-13918490900

gence; online platform

J Prev Treat Stomatol Dis, 2022, 30(6): 453-456.

【Competing interests】 The author declares no competing interests.

This study was supported by the grants from The Second Batch of Industry-University-Institute Cooperative Education Projects in 2021 (No.202102457003); National Key R&D Program of China (No.2016YFC1000503); Clinical Research MDT Project of Shanghai Ninth People's Hospital, Shanghai Jiao Tong University School of Medicine (No.201701012).

语音治疗是唇腭裂综合序列治疗的重要组成部分,基于音标的主观语音评价被认为是腭裂语音治疗的“金标准”^[1]。主观语音评价是通过专业人员主观判听来判断语音清晰度和发音错误类型的测试,需要专业团队实施,存在易受环境干扰和个体偏差影响等局限性^[2]。利用仪器进行检查分析的客观语音评价能对主观评价进一步佐证与补充,常用方法包括X线片、鼻咽内窥镜、语音频谱分析仪等,但无法避免空间限制、侵入性等问题^[3]。近年来,计算机语音处理系统和频谱分析逐渐被应用到唇腭裂语音评价中,其具有简单、无创、不受年龄限制等优势,可弥补传统主观和客观评价的不足。笔者就计算机科学在腭裂语音识别、分析和治疗方面的研究现状进行综述,为临床腭裂语音治疗提供研究思路。

1 在线语音治疗平台

唇腭裂语音治疗作为一种长期随访的持续性治疗,无需过多的辅助设备支持,因此可以适应在线智能平台诊疗的模式。

国外已有基于物联网(The internet of things, IOT)架构的远程语音治疗平台用于评估病理语音,APLo、SpeakGeek等系统在初步使用中均卓有成效^[4]。在线平台能够使医患双方进行便捷的实时交互,保证医生密切掌握患者治疗进度,以便制定个体精准、适应患者身心发育的治疗计划。

我国的唇腭裂语音治疗尚未有大范围的远程在线治疗实施,医疗资源紧张、专业化团队缺乏、高精尖技术欠缺都是亟待解决的因素,并且当下在线医疗尚处于发展阶段,其知名度与认可度仍有待提高。

在线唇腭裂语音治疗要求高质量的语音传输来保证主观评价的准确性,需克服手持式电子设备计算能力有限的问题,目前已有较为成熟的自适应模型来更好地使语音特征序列与单词网络对齐^[5]。包括科大讯飞等在内的科技公司对移动终

端处理能力的提升已有相当的进展,硬件条件基本满足语音治疗的使用需求。

2 智能腭裂语音识别

以人工智能媒介进行唇腭裂语音治疗的基础是通过语音信号处理和模式建立让计算机自动识别和分析患者语音特征。这避免了个体听力敏感度和倾向性对评价结果造成的系统误差,也减少了主观语音评价专业化学习及培训的难度。在医疗领域,已有研究发现人工智能对于自闭症谱系障碍患者的语音分析优于语音治疗师的主观判断,其能够降低结果的假阳性率与假阴性率^[6];而结合多种语音信号特征或建立语音障碍数据库的人工智能分析已能在多种不同的环境下较为精确地分辨帕金森病症语音^[7]。

近年来,针对高鼻音、咽擦音、声门塞音等腭裂语音中特征性代偿发音错误的针对性研究也逐渐增多,其识别分析技术已有较为成熟的结果。辅音省略是最典型、最常见的代偿性发音,使用机械设备克服病理性语音数据采集困难的客观条件,并通过波形差分分析方法可以达成对辅音省略88.9%的准确检测^[8];通过听觉滤波器和多通道模型相结合的方法能够对咽擦音的频谱特征进行提取并增强,使自动检测能够实现高达94.95%的正确率^[9];除此之外,利用1/3倍频程算法提取特征谱线的技术也可以实现咽擦音与正常擦音的自动分类检测^[10];而将声门塞音的多种声学特征参数结合识别的检测系统可以达到90%的信息熵正确率^[11]。结合以上研究,不难发现硬件和软件的进步和优化都是腭裂语音识别技术逐步成熟的重要基石。由于单一声学参数的使用在普遍评价上易出现系统偏差,对于各种代偿性发音不同侧重的谱系剖析也能使自动识别的准确性和特征性不断提高。同时,相较主观评价仅能通过增加检验人员数量或是重复测评来避免差错而言,借助智能语音识别能够从切分算法、能量分布、物理收音等

多角度来保证和提高判断评价的准确性^[12],这是精准医疗所必须的。

目前对于腭裂语音识别方面的研究大多仅局限于独立的声学特征辨别,且没有大规模应用于临床的报道出现,综合多参数建立判别函数并结合前沿硬件设施的识别系统可能是打破这一瓶颈的研究方向之一。

3 人工智能与言语障碍诊疗

物联网和涉及智能设备的人工智能平台在言语障碍等疾病方面已有了初步尝试,已有研究证明此类治疗方式有助于患者临床症状的改善^[13]。从广义上划分,唇腭裂也属于器质性的言语障碍,可以尝试借鉴此类治疗模式。

例如,在言语失用症领域,人工智能可辅助检测患者嘴部姿势来协助治疗人员进行语言康复训练,使医生能够同期对多位患者进行治疗,从而提高临床效率^[14];基于个人移动终端的治疗应用程序在初步临床试验中也已证明对慢性表达性失语症有益^[15],这些运用人工智能辅助治疗的案例对唇腭裂语音治疗的程序开发有极大的参考意义。

不难发现,尽管未有国际通用的治疗软件推行,但人工智能在言语障碍领域已有初步的临床使用,其在易用性、依从性、普适性、准确性及治疗效率上均有优势。结合目前已有的唇腭裂语音识别进展,人工智能技术结合主观评价的腭裂语音治疗模式应当是切实可行的,该治疗模式可以提供更频繁和更有针对性的方案,辅之以语音生成的视觉交互,视听结合的治疗模式对于患儿而言更符合其知识摄取的客观需求,大大提高交互效率^[16-17]。

唇腭裂对患者的情绪表达、幸福指数、环境适应能力上均有不同程度的影响,患者在接受语音治疗时正处于心智成熟和建立价值观的关键时期,接受器质性治疗的同时心理治疗也不容忽视。目前以人工智能为基础的社交机器人对于自闭症儿童言语表达上的治疗有较多临床研究支持,其外观及治疗形式相比传统治疗而言能显著减低患儿的社交焦虑并提高自我表达的积极性,从而提升治疗效果^[18]。完善的智能治疗平台能涵盖教学功能、激励特征和定量特征功能,可在没有医务人员的情况下用于全局评估,从而促进治疗效率的提高和合理的医疗资源分配^[19],有助于唇腭裂综合序列治疗的多地域普遍开展。

4 展望

目前,计算机科学在唇腭裂语音评价及治疗方面尚未出台具体治疗方案和治疗规范,与主观评价结合的治疗模式也有待进一步临床试验后进行补充完善。但从言语障碍领域的应用来看,其对于患儿年龄和术后时间点的要求相比常规治疗无显著差异,因此可沿用目前广泛认可的序列治疗时间线。同时,在主观评价中,噪声干扰的影响是致命性的,人工智能可以通过优化算法或是增设滤波设备来消除混响,减少残留噪声,增强识别精度,降低环境要求^[20];柔性压电声学传感器等最新设备的推出在语音接收和分析方面已有了接近甚至超越人耳的潜质^[21],加之当下普及的各类移动终端设备,其治疗模式已没有明显的硬件设备障碍;深度神经网络的出现确保人工智能能够进行自我补足与完善。因此就发展潜力而言,基于计算机科学的唇腭裂语音评价与治疗体系是超越传统主观评价模式的。由于尚无具体临床案例治疗效果的参考,计算机科学辅助的语音治疗模式在实施过程中也可能遭遇尚未发现的障碍。

远程在线平台在目前唇腭裂综合序列治疗中亦未获得较高水平的开发和利用,除了软硬件条件限制之外,临床医生、研究人员以及计算机工程师三者之间合作有限,无法进行细致的交流反馈从而进行共同开发也是阻碍因素之一^[22]。应当大力推动医工交叉项目的落实与研究,促进科研成果的临床转化。

5 总结

计算机科学在唇腭裂语音治疗领域的使用是客观需要和未来发展的双重要求,基于智能平台的语音治疗可以成为唇腭裂患者语音治疗的有效辅助手段,灵活运用人工智能媒介与主观评价体系的结合是唇腭裂综合序列治疗的前景模式之一。智能唇腭裂语音识别已有一定的研究基础和进展,但目前尚未应用于临床。未来研究方向可着眼于高效精准的语音识别和能够自我学习完善的人工智能治疗程序,依托于医工交叉项目实现人工智能在唇腭裂语音治疗领域的成果转化。

【Author contributions】 Guo YB wrote the article. Cai M revised the paper. All authors read and approved the final manuscript as submitted.

参考文献

- [1] Fitzpatrick B, Coad J, Sell D, et al. Assessing speech at three

- years of age in the cleft palate population: a scoping review of assessment practices[J]. *Int J Lang Commun Disord*, 2020, 55(2): 165-187. doi: 10.1111/1460-6984.12517.
- [2] Chapman KL, Baylis A, Trost-Cardamone J, et al. The americleft speech project: a training and reliability study[J]. *Cleft Palate Craniofac J*, 2016, 53(1): 93-108. doi: 10.1597/14-027.
- [3] Kobayashi Y, Kanamori D, Fujii N, et al. Velopharyngeal closure analysis using four - dimensional computed tomography: a pilot study of healthy volunteers and adult patients with cleft palate[J]. *BMC Med Imaging*, 2019, 19(1): 54. doi: 10.1186/s12880-019-0350-4.
- [4] Toki EI, Pange J, Mikropoulos TA. An online expert system for diagnostic assessment procedures on young children's oral speech and language[J]. *Procedia Computer Science*, 2012, 14: 428-437. doi: 10.1016/j.procs.2012.10.049.
- [5] Lee LM, Le HH, Jean FR. Improved model adaptation approach for recognition of reduced-frame-rate continuous speech[J]. *PLoS One*, 2018, 13(11): e0206916. doi: 10.1371/journal.pone.0206916.
- [6] Nakai Y, Takiguchi T, Matsui G, et al. Detecting abnormal word utterances in children with autism spectrum disorders: machine-learning-based voice analysis *versus* speech therapists[J]. *Percept Mot Skills*, 2017, 124(5): 961-973. doi: 10.1177/0031512517716855.
- [7] Kuresan H, Samiappan D, Masunda S. Fusion of WPT and MFCC feature extraction in Parkinson's disease diagnosis[J]. *Technol Health Care*, 2019, 27(4): 363-372. doi: 10.3233/THC-181306.
- [8] He L, Wang XY, Zhang J, et al. Automatic detection of consonant omission in cleft palate speech[J]. *Int J Speech Technol*, 2019, 22(1): 59-65. doi: 10.1007/s10772-018-09570-w.
- [9] 付佳, 田婷, 唐铭, 等. 结合 PECGTFs 和 SSMC 的腭裂语音咽擦音自动检测算法[J]. *计算机工程与应用*, 2019, 55(24): 102-109. doi: 10.3778/j.issn.1002-8331.1808-0265.
- Fu J, Tian T, Tang M, et al. Automatic detection algorithm for cleft palate speech pharyngeal fricatives combined with PECGTFs and SSMC[J]. *Comput Eng Appl*, 2019, 55(24):102-109. doi:10.3778/j.issn.1002-8331.1808-0265.
- [10] 何飞, 孟雨璇, 田维维, 等. 基于多延迟四阶累积量倍频程谱线的腭裂语音咽擦音自动检测算法[J]. *计算机科学*, 2020, 47(1): 144-152. doi: 10.11896/j.sjcx.180701349.
- He F, Meng YX, Tian WW, et al. Automatic detection algorithm of pharyngeal fricative in cleft palate speech based on multi-delay fourth-order cumulant octave spectral line[J]. *Comput Sci*, 2020, 47(1): 144-152. doi: 10.11896/j.sjcx.180701349.
- [11] He L, Zhang J, Liu Q, et al. Automatic detection of glottal stop in cleft palate speech[J]. *Biomed Signal Process Control*, 2018, 39: 230-236. doi: 10.1016/j.bspc.2017.07.027.
- [12] 王熙月, 黄毅鹏, 钱佳慧, 等. 基于声学特征的腭裂语音声韵母切分[J]. *计算机工程与应用*, 2018, 54(8): 123-130, 136. doi: 10.3778/j.issn.1002-8331.1611-0388.
- Wang XY, Huang YP, Qian JH, et al. Initial and final segmentation in cleft palate speech based on acoustic characteristics[J]. *Comput Eng Appl*, 2018, 54(8): 123 - 130, 136. doi: 10.3778/j.issn.1002-8331.1611-0388.
- [13] Furlong L, Erickson S, Morris ME. Computer-based speech therapy for childhood speech sound disorders[J]. *J Commun Disord*, 2017, 68: 50-69. doi: 10.1016/j.jcomdis.2017.06.007.
- [14] Castillo JC, Álvarez-Fernández D, Alonso-Martín F, et al. Social robotics in therapy of apraxia of speech[J]. *J Healthc Eng*, 2018: 7075290. doi: 10.1155/2018/7075290.
- [15] Stark BC, Warburton EA. Improved language in chronic aphasia after self-delivered iPad speech therapy[J]. *Neuropsychol Rehabil*, 2018, 28(5): 818-831. doi: 10.1080/09602011.2016.1146150.
- [16] Kersey AJ, Emberson LL. Tracing trajectories of audio - visual learning in the infant brain[J]. *Dev Sci*, 2017, 20(6): 12480. doi: 10.1111/desc.12480.
- [17] Franceschini S, Trevisan P, Ronconi L, et al. Action video games improve reading abilities and visual-to-auditory attentional shifting in English-speaking children with dyslexia[J]. *Sci Rep*, 2017, 7(1): 5863. doi: 10.1038/s41598-017-05826-8.
- [18] Sartorato F, Przybylowski L, Sarko DK. Improving therapeutic outcomes in autism spectrum disorders: enhancing social communication and sensory processing through the use of interactive robots [J]. *J Psychiatr Res*, 2017, 90: 1 - 11. doi: 10.1016/j.jpsychires.2017.02.004.
- [19] Vucovich M, Hallac RR, Kane A, et al. Automated cleft speech evaluation using speech recognition[J]. *J Craniomaxillofac Surg*, 2017, 45(8): 1268-1271. doi: 10.1016/j.jcms.2017.05.002.
- [20] Lee SC, Wang JF, Chen MH. Threshold-based noise detection and reduction for automatic speech recognition system in human-robot interactions[J]. *Sensors (Basel)*, 2018, 18(7): 2068. doi: 10.3390/s18072068.
- [21] Jung YH, Hong SK, Wang HS, et al. Flexible piezoelectric acoustic sensors and machine learning for speech processing[J]. *Adv Mater*, 2020, 32(35): e1904020. doi: 10.1002/adma.201904020.
- [22] Jesus LM, Santos J, Martinez J. The table to tablet (T2T) speech and language therapy software development roadmap[J]. *JMIR Res Protoc*, 2019, 8(1): e11596. doi: 10.2196/11596.

(编辑 张琳)



官网