

一种应用虚拟发音头像的普通话聋儿言语康复系统

刘晓千 燕楠 王岚

(中国科学院深圳先进技术研究院 深圳 518055)

摘要 本文介绍了一种新型的包含虚拟发音人头像的语言训练系统并对该系统在听障儿童语言康复训练中的实用性进行了研究。该系统以虚拟三维发音人头像的形式再现了汉语中常见的 20 个单音节字的发音动作,并将之用于语言康复训练。系统运用 EMA(三维电磁发音仪)设备采集内部及外部发音器官运动轨迹作为驱动数据,有效模拟了汉语发音。此外,系统还包括一个听力训练模块和一个朗读训练模块,用于辅助聋儿的发音康复训练。文章对系统发音模拟功能的准确性进行了人工评估,并应用系统对两名 4 至 5 岁之间的聋儿进行了初步的康复训练。初步评估及训练的结果表明:系统的发音模拟功能可有效模拟发音器官的动作,有助于提高听障儿童的语言能力;同时,系统的设计风格较易获得幼儿认同。研究证明,本系统在听障儿童发音训练方面是有效的,具有较高的应用价值。

关键词 听障儿童;交互式言语康复系统;三维模拟发音指导;虚拟现实

An Interactive Speech Training System with Virtual Reality Articulation for Hearing Impaired Mandarin-speaking Children

LIU Xiao-qian YAN Nan WANG Lan

(Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China)

Abstract In this study, a novel interactive speech training system based on virtual reality articulation was introduced and its efficiency in the rehabilitation of hearing impaired (HI) children was investigated as well. 20 meaningful words in Mandarin were presented by using a 3D talking head in this system for articulatory training. The Electro-Magnetic Articulography (EMA) and graphic transform technology were used to develop the reality inner articulatory movements of 3D talking head. Furthermore, relevant corpus was organized in a listening module and a speaking module of the system to help to recover the HI children's language skills. Then, the accuracy of virtual reality (VR) articulatory movement model was evaluated through a series of experiments. Finally, a pilot test was performed on the training of two HI children with the system. The preliminary results show that there is an improvement in the speaking abilities of the tested HI children. Meanwhile, the system was recognized as acceptable and interesting for younger children. The study indicates that the presented training system is effective to articulation training for the HI children.

Keywords hearing impaired children; interactive speech training system; 3D articulatory tutor; virtual reality

1 引 言

听障儿童的康复是长期以来得到社会各界广泛关

注的公益领域。其中,听障儿童的语言能力通常较为欠缺。Resmi 等人的研究成果表明,幼儿时期的听力损失对其语言能力、交流能力的成熟和发展有很大影响^[1]。Engwall 等人在研究中也指出,幼儿在耳聋后

基金项目: 国家自然科学基金重点项目(61135003); 国家自然科学基金(90920002); 广东省创新型研究团队(201001D0104648280)。

作者简介: 刘晓千,研究方向为计算机辅助言语康复, E-mail: liu.xq@siat.ac.cn; 燕楠,副研究员,研究方向为构音障碍、失语症等言语障碍的康复与评估; 王岚,研究员,研究方向为言语康复与大规模语音识别。

其听觉十分有限,无法靠自身获取声音信息,进而缺少发声训练中所必须的模仿对象,因此聋儿的说话能力发育趋于停滞^[2]。同时,最新研究表明,可视化的发音教学有助于听障儿童的语言学习,视听结合的语音教学方式比单独的视觉或听觉教学效果更好,这是由于发音的动态模拟过程可刺激聋儿对内部发音器官运动轨迹的理解,并使聋儿形成发音时对发音器官位置的控制^[3]。通过建立和应用三维动态发音模型,可以有效刺激聋儿视觉反馈的形成,从而使聋儿的发音动作逐渐规范化。

同时,最新的神经生理研究还表明,大脑具有与手、足、口等部位运动有关联的镜像神经元系统^[4,5]。而语音的产生功能也很可能与镜像神经元系统有关。

Franceschini 等人的研究成果表明,镜像神经元在发音中的作用可分为两个方面^[6]:一方面,当人们对某些词汇有模仿经验后,再次发音会强化练习效果;另一方面,中枢神经系统中负责某些特定动作执行的区域也和表示这些动作的词语有关。因此,在有构音障碍的患者(例如听障儿童)的康复训练中,可以通过加强模仿来形成发音动作镜像,从而加强发音练习的效果。

传统的语言康复训练一般通过教师人为产生发音动作的示例让受训者模仿。这样的训练方法难以使听障儿童观察到内部发音器官(如舌头等)的准确运动。针对传统训练方法的缺陷,近年来涌现了一批视听结合的聋儿言语康复系统。Rathinavelu 等人设计了一款计算机辅助聋儿康复系统,该系统使用磁共振成像技术(MRI)探知发音器官在发音过程中的运动轨迹并生成对应的模拟发音图片序列用于聋儿的发音教学^[7,8]。Olle Bälter 等人开发出了一款 ARTUR(Articulatory Tutor, 发音向导)系统,该系统包含一个用于发音模拟和教学的三维虚拟发音人头像。同时,该系统还可记录说话人的发音过程并将

其与标准发音进行对比教学^[9]。法国和中国台湾地区的研究人员也开发了类似的语言康复训练系统,这些系统使用计算机发音模拟过程来进行聋儿发音教学^[10,11]。受训儿童首先模仿一个音素或字的发音,然后系统使用语音识别技术来检验发音的正确与否并给出相应反馈。但现有的研究缺少一款专门针对汉语普通话的聋儿言语康复系统,也缺少对发音器官运动的准确模拟。

因此,本研究针对现有系统缺陷,开发出一款专门针对汉语普通话训练的交互式聋儿言语康复系统,采用三维发音模拟功能给予聋儿发音器官正确发音位置的真实模拟。三维发音模拟头像可以使受训儿童自由观察下巴、舌头、嘴唇等发音器官的运动。通过使用系统进行训练,聋儿能够模仿并掌握三维发音头像所模拟的正确发音动作。系统同时还应该包含针对汉语中常用字和易混淆字的听力和朗读训练,用来辅助发音动作训练,提高儿童的语言运用能力。

2 系统设计

完善的聋儿言语康复系统的需求应包括:

- (1) 能够生成难度适宜的聋儿训练课程;
- (2) 具有容易为聋儿接受的界面风格(例如卡通风格等);
- (3) 朗读模块应具有语音识别和对比功能,可以自动接受并处理训练者的发音;
- (4) 系统能够给出恰当的反馈,鼓励听障儿童发音。

根据实际需求,我们设计了新型聋儿言语康复系统。图 1 是交互式言语康复系统的体系结构图。系统共包括三个主要部分:听力训练模块、三维模拟发音指导以及朗读训练模块。

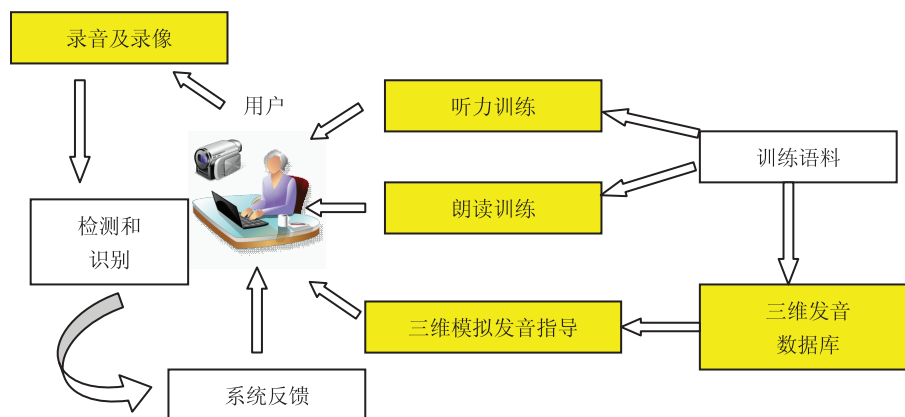


图 1 聋儿言语康复系统结构图

2.1 听力训练模块

听力训练模块共包含 18 组, 每组 3 个汉语中常见近音字的发音及其对应图片。在训练时, 系统随机播放每组字中一个的读音, 受训儿童根据听到的内容进行辨别及选择。系统根据选择情况做出不同反馈: 若选择正确则给出正确提示, 并进入下一组近音字训练; 若选择错误则重新播放当前一组近音字的提示音, 并重新选择。通过不断的训练可以提高聋儿对不同字的发音的掌握。图 2 是听力训练界面。



图 2 听力训练界面

2.2 朗读训练模块

朗读训练模块主要是为了巩固发音练习内容、训练聋儿说话能力而设计。此模块共包含“果园”、“动物园”、“家庭”、“课堂”四个儿童较为熟悉的场景, 每一场景均包含若干常见名词的图片。当录音时, 每一场景中的名词图片会依序高亮显示, 受训儿童需大声说出高亮显示的图片对应的词汇。系统后台会自动对其发音进行语音识别, 并根据识别结果给出正确或者错误反馈。当一个场景中的发音任务全部结束之后, 系统会播放提示音, 对受训儿童进行鼓励。图 3 是系统朗读训练模块界面。

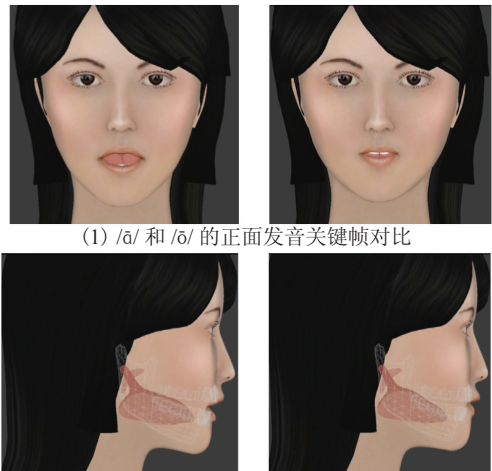


图 3 朗读训练界面

2.3 三维模拟发音指导

听障儿童的主要发音障碍是发音器官运动的障

碍, 因此本系统的主要任务是建立发音器官的三维运动模型。驱动这一模型的发音点位移数据可以使用电磁发音仪(EMA)设备采集。EMA 的位置传感器可以粘贴于口腔内的舌头、牙齿、下颚等用常规手段不易观察到的部位, 通常用于采集同步的音频和位置数据^[12,13]。本次实验包括 4 个粘贴在唇部的传感器和 3 个粘贴在舌头上的传感器, 因此所建立模型可以精确反映柔软发音器官的动作。运用 EMA 采集的位置数据建立三维发音人模型的步骤可参见^[12]。3D 发音人头像模型包括小舌、舌头、上下唇、上下齿、下颚的三维模型。而关于 3D 说话人模型的变形算法很多, 这里采用的是 DFFD(Dirichlet Free-form Deformation)算法, 通过引用新的坐标系即自然邻居坐标 Sibson 坐标可以把控制点设置在网格的任意位置, 同时控制网格的形状可以是任意的。DFFD 算法更适合于复杂多变的几何变形^[14]。因此, 本文采用 DFFD 算法为变形算法, 使虚拟 3D 说话人头产生了符合人体生理特征的发音动作。图 4 是 /ā/ 和 /ò/ 的三维发音模拟对比。



(1) /ā/ 和 /ò/ 的正面发音关键帧对比

(2) /ā/ 和 /ò/ 的侧面发音关键帧对比

图 4 /ā/ 和 /ò/ 的三维发音模拟示意图

图 5 是三维发音指导界面的示意图。儿童在训练

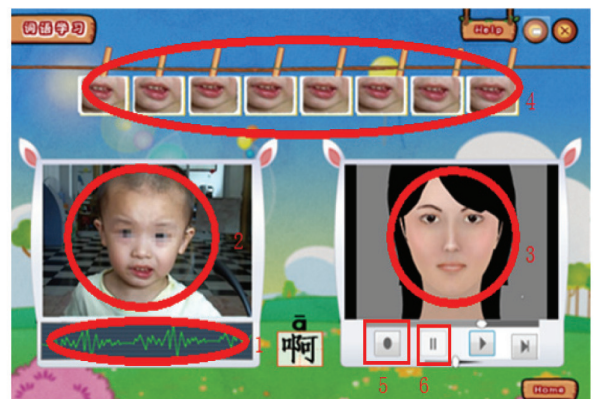


图 5 三维发音模拟指导界面

时首先观看发音演示视频, 然后面对摄像头重复发音, 发音视频经系统后台处理, 提取出发音关键帧并显示在界面最上方。通过自身发音动作与标准发音动作的对比, 儿童可以更好地掌握发音器官的关键动作。

3 人工评测

为了评估三维发音人头像模拟真实发音过程的准确度, 共设定两组人工评测实验。

第一组人工评测实验的目的: 测试三维发音人头像模拟发音过程在区分易混淆音节上的性能。实验选取 10 对易混淆音节的模拟发音动画以及 10 名普通话水平优秀的测试者, 每位测试者都需要对 10 对易混淆音节的模拟发音动画进行区分。测试方法是: 假设需要区分的发音动画分别为 /cī/ 和 /chī/, 则需要告知测试者该对发音分别为 /cī/ 和 /chī/, 然后先后随机播放两个发音动画, 由测试者来对其进行辨认, 并记录辨认结果。每对结果的辨认真确率统计如表 1 所示: 平均辨认正确率达到 90% 以上, 证明系统可以

正确区分易混淆发音的发音动作。

第二组人工评测实验的目的: 测试单个字的发音人头像是否能正确模拟真实的发音。实验邀请了另外 8 位普通话水平优秀的测试者对每个发音的模拟动画进行打分, 其中, 1 分表示“差”, 2 分表示“普通”, 3 分表示“好”, 4 分表示“较好”, 5 分表示“极好”。每个模拟发音动画的平均得分如表 2 所示。所有模拟动画的平均得分为 4.2 分, 这说明测试者们普遍认为: 三维模拟发音人头像对真实发音的模拟效果较好且基本正确。

4 初步训练及训练结果

为了验证系统是否有助于聋儿言语康复, 本实验还包括对聋儿进行初步训练。从深圳市晴晴言语康复中心选择了 1 位男童和 1 位女童(均在 4 到 5 岁之间), 二者均为耳聋患者且有构音障碍, 分别使用助听器或人工耳蜗来恢复听力。二者均接受过 1 年的普通言语训练, 未接受过计算机辅助训练, 但对计算机辅助训练的方式表现出较浓厚的兴趣。

表 1 易混淆音节的区分正确率

易混淆音节对	辨别率	易混淆音节对	辨别率	易混淆音节对	辨别率
/ā/ vs. /ō/	90%	/lē/ vs. /nē/	85%	/rī/ vs. /chī/	85%
/ē/ vs. /ī/	100%	/gē/ vs. /dē/	80%	/zī/ vs. /zhī/	100%
/ū/ vs. /yū/	90%	/xī/ vs. /qī/	80%	/cī/ vs. /chī/	100%
/sī/ vs. /shī/	90%				

表 2 单个音节的人工评分平均分

音节	平均分	音节	平均分	音节	平均分	音节	平均分
/ā/	4.1	/yū/	4.4	/xī/	4.4	/cī/	4.0
/ō/	4.4	/gē/	4.1	/qī/	4.1	/sī/	3.1
/ē/	3.8	/dē/	4.0	/jī/	4.1	/zhī/	4.1
/ī/	3.8	/lē/	4.0	/rī/	3.8	/chī/	4.4
/ū/	4.0	/nē/	4.8	/zī/	4.0	/shī/	3.8

表 3 儿童发音清晰度评分以及用户体验评分

	发音清晰度评分		错误最多音节		用户体验评分		
	训练前	训练后	训练前	训练后	学习兴趣	学习动力	可接受程度
S1	64	77	/nē/, /cī/, /sī/, /shī/ /jī/, /xī/, /zhī/	/jī/, /xī/, /zhī/	4	5	4
S2	61	72	/chī/, /shī/, /rī/, /jī/, /sī/	/rī/, /jī/, /sī/	4	4	3

初步训练过程持续两周,两位受训者每天需接受至少 40 分钟的计算机辅助训练,每周 5 次,共 10 次训练。每天训练的内容包括 5 组听力易混字训练、10 个发音训练和 2 组朗读训练,由两位语言治疗师辅助儿童进行训练。系统会记录下每位受训儿童每天训练的时间以及每个发音正确与否。在全部训练开始之前和之后,分别使用文献[15]中的方法对两位儿童进行语言清晰度测试。通过比较初步训练之前和之后的语言清晰度差异,即可评价系统训练的有效程度。此外,在 5 次训练结束后还需要对系统的用户体验做一次评价,共有 3 个评价标准:“乐趣”、“学习动力”和“可接受程度”。儿童在治疗师的帮助下,按以上 3 个标准对系统进行 5 分制打分。

表 3 给出了初步训练之前、初步训练之后的语言清晰度测试结果以及用户体验评分结果。语言清晰度测试的得分是指在 100 个常见音节中儿童发音正确的个数。语言清晰度测试的结果表明,两个聋儿的语言清晰度在训练之后有明显提高,特别是在双唇音、唇齿音、舌尖音、舌根音的发音上提高较大;而在舌尖后音、舌面音的发音上提高较小。用户体验评价的结果表明,听障儿童在训练中体会到较强的乐趣(平均得分 4.0)和较强的学习动力(平均得分 4.5),对于初步训练的学习内容,接受程度尚可(平均得分 3.5)。

5 结 论

本研究的主要目的是设计一款有效的、针对汉语普通话的聋儿言语康复系统。在系统中嵌入了一个新式的、基于 EMA 数据的三维发音模拟模块。为了验证三维发音模拟的准确性,我们评价了每个单独发音模拟动画的准确度以及易混淆音节的发音动画的可区分程度;结果表明,设计的三维发音模拟头像可以较好地表现所对应的真实发音过程,并且跟易混淆音节有一定的区分度。初步的临床训练表明,虽然参与训练的儿童数量较少,但他们的发音清晰度提高较明显。此外,系统的界面、提示以及反馈均采用卡通风格,易于为儿童所接受。

然而,该系统只是初步的系统模型,在临床试用过程中仍存在一些缺陷,主要表现在:①现有的三维发音模拟主要模拟了嘴唇、下巴、牙齿和舌的运动,而对于发音过程中的气流控制的模拟不足,因此在接下来的研究中需要引入气流运动的表现形式,例如声

带的振动;②系统中的训练语料较少,需要进一步扩充三维模拟发音人头像的内容。因此,我们下一步将根据临床反映对系统进行进一步的优化和升级,使之能够更好的为聋儿的言语康复提供帮助。

参 考 文 献

- [1] Resmi K, Kumar S, Sardana H K, et al. Graphical speech training system for hearing impaired [C] // Proceedings of International Conference on Image Information Processing (ICIIP 2011), 2011: 1-6.
- [2] Engwall O, Bälter O, Öster A M, et al. Designing the user interface of the computer-based speech training system ARTUR based on early user tests [J]. Behaviour & Information Technology, 2006, 25: 353-365.
- [3] Badin P, Youssef A B, Bailly G, et al. Visual articulatory feedback for phonetic correction in second language learning [J]. Actes de SLATE, 2010: 1-10.
- [4] Buccino G, Solodkin A, Small S L. Functions of the mirror neuron system: implications for neurorehabilitation [J]. Cognitive and Behavioral Neurology, 2006, 19: 55-63.
- [5] Small S L, Buccino G, Solodkin A. The mirror neuron system and treatment of stroke [J]. Developmental Psychobiology, 2012, 54: 293-310.
- [6] Franceschini M, Agosti M, Cantagallo A, et al. Mirror neurons: action observation treatment as a tool in stroke rehabilitation [J]. European Journal of Physical and Rehabilitation Medicine, 2010, 46: 517-523.
- [7] Rathinavelu A, Thiagarajan H, Rajkumar A. Three dimensional articulator model for speech acquisition by children with hearing loss [C] // Proceedings of 4th International Conference on Universal Access in Human Computer Interaction, 2007, 4554: 786-794.
- [8] A. Rathinavelu, Yuvaraj G. Data visualization model for speech articulators [C] // Proceedings of AICERA, 2011: 155-159.
- [9] Bälter O, Engwall O, Öster A M, et al. Wizard-of-Oz test of ARTUR: a computer-based speech training system with articulation correction [C] // Proceedings of 7th International ACM SIGACCESS Conference on Computers and Accessibility, 2005: 36-43.
- [10] Rooney E, Carraro F, Dempsey W, et al. HARP: an autonomous speech rehabilitation system for hearing-impaired people [C] // Proceedings of Third International Conference on Spoken Language Processing, 1994: 2019-2022.
- [11] Hsiao M L, Li P T, Lin P Y, et al. A computer based software for hearing impaired children's speech training and learning between teacher and parents in Taiwan [C] // Proceedings of 23rd Annual International Conference of the IEEE, 2001, 2: 1457-1459.
- [12] Wang L, Chen H, Li S, et al. Phoneme-level articulatory

-
- animation in pronunciation training [J]. *Speech Communication*, 2012, 54: 845-856.
- [13] Chen H, Wang L, Liu W, et al. Combined X-ray and facial videos for phoneme-level articulator dynamics [J]. *The Visual Computer*, 2010, 26: 477-486.
- [14] Moccozet L, Thalmann N M. Dirichlet free-form deformations and their application to hand simulation [C] // *Proceedings of Computer Animation*, 1997: 93-102.
- [15] Sun X B. Evaluation criteria and methods of hearing impaired children [M]. Beijing: Sanchen Video Press, 2009.