

12-13-2019

Rehabilitation Training System of VR-based Speech Apraxia

Huanhuan Jiao

1. School of Information Science and Engineering, Hangzhou Normal University, Hangzhou 311121, China; ;3. Virtual Reality and Intelligent Systems Research, Hangzhou Normal University, Hangzhou 311121, China; ;

Zhigeng Pan

2. Alibaba Business School, Hangzhou Normal University, Hangzhou 311121, China; ;3. Virtual Reality and Intelligent Systems Research, Hangzhou Normal University, Hangzhou 311121, China; ;4. Guangzhou nine Digital Technology Co., Ltd., Guangzhou 510000, China;

Shuchang Xu

1. School of Information Science and Engineering, Hangzhou Normal University, Hangzhou 311121, China; ;

Qingshu Yuan

1. School of Information Science and Engineering, Hangzhou Normal University, Hangzhou 311121, China; ;

Follow this and additional works at: <https://dc-china-simulation.researchcommons.org/journal>



Part of the Artificial Intelligence and Robotics Commons, Computer Engineering Commons, Numerical Analysis and Scientific Computing Commons, Operations Research, Systems Engineering and Industrial Engineering Commons, and the Systems Science Commons

This Paper is brought to you for free and open access by Journal of System Simulation. It has been accepted for inclusion in Journal of System Simulation by an authorized editor of Journal of System Simulation.

Rehabilitation Training System of VR-based Speech Apraxia

Abstract

Abstract: Aiming at the problem that the enthusiasm of active training in patients with traditional speech apraxia rehabilitation is gradually weakened, a virtual reality-based rehabilitation training system for speech apraxia is designed. The system simulates the 26 monosyllabic pronunciation mouths shapes made by virtual human in Chinese and uses it for the learning module. *It is proposed to identify 26 monosyllabic words by using the isolated word speech recognition algorithm, which improves the accuracy of speech recognition.* The system includes a training module and a scene module for assisting the patient's pronunciation rehabilitation training. The evaluation results show that the virtual human articulation process in the system can realistically simulate the movement of real articulation organs, which plays a positive role in improving the patients' language ability.

Keywords

Virtual reality, Speech Aphasia, Virtual mouth animation, Speech recognition of isolated word

Recommended Citation

Jiao Huanhuan, Pan Zhigeng, Xu Shuchang, Yuan Qingshu. Rehabilitation Training System of VR-based Speech Apraxia[J]. Journal of System Simulation, 2019, 31(11): 2344-2349.

基于 VR 的言语失用症康复训练系统

焦欢欢^{1,3}, 潘志庚^{2,3,4}, 徐舒畅¹, 袁庆曙¹

(1. 杭州师范大学信息科学与工程学院, 浙江 杭州 311121; 2. 杭州师范大学阿里巴巴商学院, 浙江 杭州 311121;
3. 杭州师范大学虚拟现实与智能系统研究院, 浙江 杭州 311121; 4. 广州玖的数码科技有限公司, 广东 广州 510000)

摘要: 针对传统言语失用症康复训练中患者主动训练积极性逐渐减弱的问题, 设计了一种基于虚拟现实的言语失用症康复训练系统。采用虚拟人嘴型动画模拟汉语中 26 个音节发音嘴型, 并用于学习模块。利用孤立字语音识别算法识别 26 个音节, 提高识别的准确率。系统还用训练模块、场景模块辅助患者康复训练。结果表明: 系统中的虚拟人发音过程能够逼真地模拟真实发音器官的运动, 对提高患者的语言能力具有积极作用。

关键词: 虚拟现实; 言语失用症; 虚拟人嘴型动画; 孤立字语音识别

中图分类号: TP391.9 文献标识码: A 文章编号: 1004-731X (2019) 11-2344-06

DOI: 10.16182/j.issn1004731x.joss.19-FZ0334

Rehabilitation Training System of VR-based Speech Apraxia

Jiao Huanhuan^{1,3}, Pan Zhigeng^{2,3,4}, Xu Shuchang¹, Yuan Qingshu¹

(1. School of Information Science and Engineering, Hangzhou Normal University, Hangzhou 311121, China; 2. Alibaba Business School, Hangzhou Normal University, Hangzhou 311121, China; 3. Virtual Reality and Intelligent Systems Research, Hangzhou Normal University, Hangzhou 311121, China; 4. Guangzhou nine Digital Technology Co., Ltd., Guangzhou 510000, China)

Abstract: Aiming at the problem that the enthusiasm of active training in patients with traditional speech apraxia rehabilitation is gradually weakened, a virtual reality-based rehabilitation training system for speech apraxia is designed. The system simulates the 26 monosyllabic pronunciation mouths shapes made by virtual human in Chinese and uses it for the learning module. *It is proposed to identify 26 monosyllabic words by using the isolated word speech recognition algorithm, which improves the accuracy of speech recognition.* The system includes a training module and a scene module for assisting the patient's pronunciation rehabilitation training. The evaluation results show that the virtual human articulation process in the system can realistically simulate the movement of real articulation organs, which plays a positive role in improving the patients' language ability.

Keywords: Virtual reality; Speech Apraxia; Virtual mouth animation; Speech recognition of isolated word

引言

言语失用是因脑损伤造成的言语产生时的语音-运动障碍, 常常伴随失语症的出现^[1]。言语失

用症的存在严重影响患者日常生活活动能力及回归社会能力^[2]。目前, 国内外对言语失用症的评估治疗方法仍较少。传统的康复治疗中, 医护人员需要对患者提供大量的感觉提示, 如视觉、触觉、肌肉运动知觉提示、触摸提示和言语提示等, 工作量巨大且枯燥乏味^[3], 已不能满足治疗的需求。

虚拟现实技术因其独特优势, 开始被越来越多地应用到康复治疗领域。例如, Ogourtsova 等^[4]提



收稿日期: 2019-05-30 修回日期: 2019-07-16;
基金项目: 国家重点研发计划(2017YFB1002803), 广州创新创业领军团队项目(CXLJTD-201609), 浙江省健康智慧厨房系统集成重点实验室开放课题(2017F04);
作者简介: 焦欢欢(1993-), 女, 安徽宿州, 硕士生, 研究方向为虚拟现实与人机交互。

<http://www.china-simulation.com>

• 2344 •

出了一种沉浸式虚拟现实系统,可较好地用于训练患者的空间知觉能力;Maier 等^[5]提出了一种康复治疗游戏系统(虚拟现实系统),并将其用于脑卒中患者的康复训练。该系统在提高患者的记忆力、空间想象能力具有独特优势;王辉等^[6]通过电脑反馈界面搭建了一个用于脑卒中患者康复训练的虚拟现实系统,并用于患者的日常生活能力训练。但以上 3 个训练系统都缺乏对患者的言语失用康复训练。刘婷婷等^[7]针对在特殊人群康复中虚拟现实技术的应用研究进行了总结,具体梳理了该技术在心理问题、脑卒中和自闭症儿童 3 类人群中的应用。提出了将虚拟智能体引入到康复治疗当中,通过结合体感和语音技术建立情景化虚拟康复环境的构想。可看出语音技术或言语残疾康复训练本身在康复治疗中也能发挥重要作用。

大量的言语失用症康复训练研究表明,重复发音、重复声音顺序和重复动作模式是康复训练的必要内容^[8]。根据运动康复理论和大脑神经可塑性理论,丰富的训练环境可促进康复的进程^[9-10],也可有效地提高患者的训练兴趣和积极性,从而提高康复训练效率。

传统的康复治疗方法通常是在医院由康复师示范发音器官动作并让患者模仿,使用这样的训练方法,患者可在一定程度上模仿康复师的行为,但是难以使患者观察到发音器官的准确运动,并且其康复治疗过程单调、昂贵,患者难以产生兴趣。针对传统训练方法的缺陷,近年来涌现了一系列基于视听相结合的言语康复训练系统。Miller^[11]提出了一种基于虚拟现实的训练系统,主要应用于聋哑儿童的语言能力训练。Eden^[12]提出了一种用于聋哑儿童叙事能力训练的虚拟现实系统。杨雪和雷江华^[13]对虚拟现实技术在聋哑儿童生活能力(包含语言能力)训练中的应用研究进行了总结。Rathinavelu 等^[14-15]设计了一款计算机辅助言语康复训练系统,该系统的原理是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,并根据该轨迹数据生成对应的模拟发音图片序列用于聋儿的发音教学。

Olle 等^[16]开发出了一款 ARTUR (Articulatory Tutor, 发音向导)系统,该系统由一个用于模拟发音和教学的三维虚拟发音人头像组成,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。

但是,目前专门针对脑卒中引起的言语失用症康复训练系统的研究比较少,也缺少对发音器官运动的准确模拟。为此,本文设计了一种基于虚拟现实技术的康复训练系统,将治疗手段和康复练习融于游戏活动中,有利于提高患者对于康复治疗的兴趣和主动性,从而减少康复师的工作量,提高康复训练效率。

1 系统设计

相较于传统的言语失用症康复训练中重复构音器官发音训练、重复发声训练等问题,虚拟康复训练系统的主要优点是通过提供沉浸式的虚拟场景,提高了患者康复训练的兴趣和主动性,减少了康复师的工作量,降低了患者的康复训练成本,有效地提高患者的训练效率。整个系统使用 Unity 3D 作为开发软件,交互设备采用 HTC Vive 头戴式显示设备,编程语言采用 C#,能够实现数据采集、实时语音识别和测评的功能。

虚拟康复训练系统的总体设计如图 1 所示。

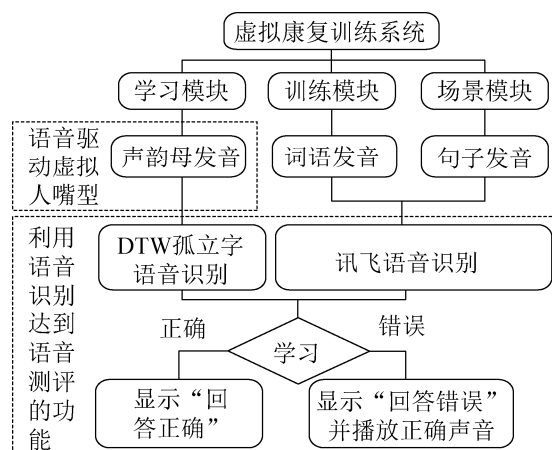


图 1 基于虚拟现实的康复训练系统组成

Fig. 1 Composition of rehabilitation training system based on virtual reality

系统包括学习模块、训练模块和场景模块,这

3 个模块分别从声韵母发音、词语发音、句子发音进行康复训练。

学习模块采用虚拟人发音嘴型的形式模拟汉语中 26 个单音节发音嘴型动画，利用 DTW 孤立字语音识别算法识别 26 个单音节；训练模块和场景模块通过调用讯飞开放平台的语音识别 SDK 实现词语和句子的识别，从而实现语音测评的功能。

2 模块结构

2.1 学习模块

学习模块主要包括 26 个声韵母单音节的嘴型学习和发音学习。在学习时，系统随机分配一个单音节，患者根据每个单音节的嘴型动画和发音进行学习。系统根据检测到的声音作出不同的提示：若学习正确则显示“回答正确”界面，并进入下一个声韵母单音节的学习；若学习错误则显示“回答错误”界面，并播放正确发音。通过不断地进行构音器官训练和发音训练，学习模块可提高患者对声韵母单音节的学习能力。学习模块场景如图 2 所示。

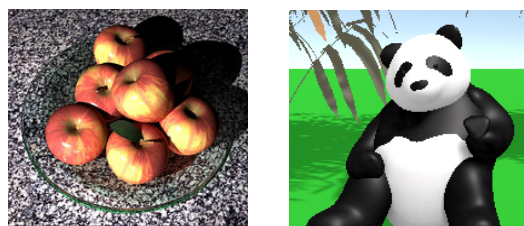


图 2 学习模块场景
Fig. 2 Scenario of module

2.2 训练模块

训练模块设计的目的在于巩固发音练习内容、训练患者说话能力。该模块主要包括“动物”、“水果”两类共 16 个模型。在训练过程中，系统会随机分配一个模型(如苹果)，患者需要按住手柄进行录音并回答相应模型对应语音。当系统检测到语音时，会自动识别检测的信息，并对其进行判断。根据判断结果向患者作出“正确”或者“错误”的提示。当一个场景中的发音训练任务全部结束后，系统会播放相应的提示界面，该提示界面会对患者进行言语鼓励，来提高患者训练的积极性。训练模块中的

部分模型如图 3 所示。



(a) 苹果 (b) 熊猫

图 3 训练模块的部分模型

Fig. 3 Partial model of training modules

2.3 场景模块

场景模块是基于学习模块和训练模块构建的，可向患者提供一个能够进行社交能力训练的三维虚拟应用场景。应用场景包含了水果店、动物园等日常典型的场景。在进行语音训练的时候，患者可在水果店里与虚拟店员进行互动并购买水果，或在动物园里进行参观互动。设计场景应用模块设计是为了向患者提供一个虚拟的三维互动环境，使得患者在康复训练的过程中更能体验到身临其境的感觉，从而提高患者参与互动的积极性和兴趣，进而巩固学习模块和训练模块的成果。系统中部分应用场景如图 4(a)、(b)所示。



(a) 在水果店中购买水果



(b) 在动物园中参观

图 4 应用场景

Fig. 4 Application scenarios

3 虚拟人嘴型驱动与语音识别

3.1 虚拟人嘴型驱动

创建三维人脸动画的第一步就是创建一个三维的人脸模型。系统使用 3D MAX 建模, 并在此基础上创建虚拟人嘴型动画。虚拟人嘴型驱动流程图如图 5 图所示。

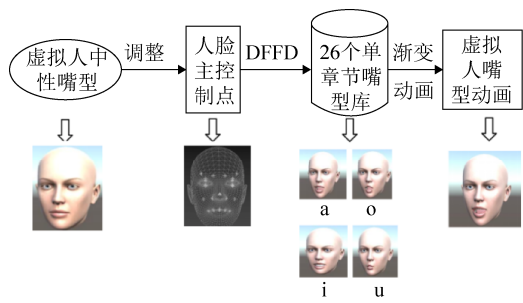


图 5 虚拟人嘴型驱动流程图
Fig. 5 Flow chart for mouth driver of virtual human

图 6 为控制人脸嘴型的主要特征点, 通过控制主要特征点的位移得到变形后的虚拟人单音节嘴型, 从而创建 26 个单音节嘴型库。有关 3D 说话虚拟人模型的变形算法很多, 本文采用 Dirichlet 自由变形算法(Dirichlet free-form deformation)^[17], 该方法对控制点进行调整, 打破了传统变形算法(FFD)对控制拓扑结构的限制, 主要针对选出的控制点进行变形, 模拟虚拟人的肌肤运动, 从而使 3D 说话人虚拟人产生了符合真实人体生理特征的发音动作。

虚拟人嘴型驱动的最后一步就是获得平滑逼真的人脸动画, 本文主要采用基于 DirectX 的渐变动画(Morphing)技术^[18], 对 2 个动画关键帧间进行插值。

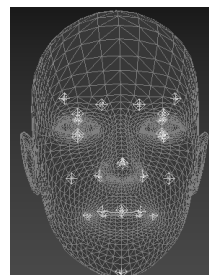


图 6 控制人脸嘴型的主要特征点
Fig. 6 Main feature points controlling the facial mouth shape

3.2 语音识别

系统通过调用讯飞开放平台的语音识别 SDK, 对检测到的声音进行识别。由于讯飞不能直接对汉语中单音节进行识别(如输入 a, o, e 读音不能识别为 a, o, e), 所以把单音节的语音识别转换为相同读音的汉字识别(如单音节 a, 对应“啊”、“阿”等所有发/a/读音的汉字), 然后与正确的嘴型汉字库相匹配, 从而实现语音测评的功能。经过实验测试, 讯飞对孤立字的识别率比较低, 平均识别正确率为 46.1%, 对词语和句子的识别率比较高, 平均正确识别率为 92%。所以, 需要找到一个对孤立字识别率比较高的算法。

目前, 针对孤立字的语音识别算法有很多, 本文采用 DTW (Dynamic Time Warping, 动态时间归整)算法^[19]。因为 DTW 算法比其他算法运算量小, 在孤立字小词汇量的语音识别中有着广泛应用。

基于以上算法, 对 26 个单音节读音进行语音识别, 每个音节识别 10 次, 分别采用讯飞开放平台的语音识别与 DTW 孤立字语音识别算法后, 其识别的对比结果如表 1 所示。

表 1 讯飞和 DTW 算法对比结果

Tab. 1 Comparison of results between XunFei and DTW algorithm																												/%
音节	a	o	e	i	u	ü	b	p	m	f	d	t	n	l	g	k	h	j	q	x	z	c	s	r	y	w		
讯飞	90	90	20	60	60	10	10	10	40	20	70	20	70	20	90	80	90	20	80	20	80	10	10	10	60	60		
DWT 算法	90	80	80	50	70	60	80	60	70	80	70	50	80	60	80	90	80	70	90	70	80	50	70	80	80	70		

由表 1 知，讯飞和 DTW 算法对 26 个声韵母单音节的识别率分别为 46.1%，72.6%。因此，系统对声韵母单音节进行识别时，调用 DTW 孤立字语音识别算法；对词语和句子进行识别时，调用讯飞开放平台的语音识别算法。

4 系统评估

本系统共设定 2 组人工测评实验，通过对比实验数据来评估虚拟人模拟真实发音过程的准确度。

第 1 组人工测评的目的是在易混淆音节的区分上，测试虚拟人模拟发音的性能。实验对象由两部分组成：8 对易混淆音节的模拟发音动画和 10 名普通话水平优秀的测试者，每位测试者都需要区分 8 对易混淆音节的模拟发音动画。

测试方法如下：假设/a/与/o/为需要区分的发音动画，首先要把/a/与/o/的发音告诉测试者，然后将两个发音动画随机播放，由测试者来对其进行分辨，并记录辨认结果。每对结果的辨认正确率统计见表 2。可看出，平均辨认正确率达到 78.75%以上，这说明该系统可正确区分混淆发音的发音动作。

表 2 易混淆音节的区分正确率

Tab. 2 Classification accuracy of confusable syllables

易混淆音节对	辨别率/%	易混淆音节对	辨别率/%
/a/与/o/	100	/n/与/l/	90
/u/与/ü/	80	/g/, /k/与/h/	60
/b/与/p/	90	/j/, /q/与/x/	70
/d/与/t/	80	/z/, /c/与/s/	60

第 2 组人工评测实验的目的是对于单个音节的发音嘴型动画，测试虚拟人对其是否能正确模拟。这组实验由 5 位优秀的普通话水平测试者组成，在实验中，分数分为 5 个等级，1 分代表“差”，2 分代表“普通”，3 分代表“好”，4 分代表“较好”，5 分代表“极好”。通过 5 位测试者对每个发音的模拟动画进行打分，并记录其结果和分析，不难得到每个模拟发音动画的平均分如表 2 所示。由表可知测试所得到的所有模拟动画的平均分得分为 3.68 分，这说明测试者们普遍认为：虚拟人嘴型发音动画对真实发音的模拟效果比较好。

表 3 单个音节的人工评分平均分
Tab. 3 Average score of a single syllable

音节	平均分	音节	平均分
/a/	3.78	/l/	3.64
/o/	3.84	/g/	3.52
/e/	3.64	/k/	3.64
/i/	4.02	/h/	3.44
/u/	3.98	/j/	3.72
/ü/	3.80	/q/	3.86
/b/	3.74	/x/	3.68
/p/	3.48	/z/	3.92
/m/	3.64	/c/	3.70
/f/	3.56	/s/	3.66
/d/	3.30	/r/	3.56
/t/	3.44	/y/	3.88
/n/	3.56	/w/	3.82

5 结论

本文针对言语失用症患者设计了一套基于虚拟现实的康复训练系统。该系统采用虚拟人发音嘴型模拟汉语中 26 个音节嘴型动画，为验证虚拟人模拟的准确性，本文对易混淆音节的发音动画辨别率以及每个音节发音模拟动画的准确度进行评价；结果表明，设计的虚拟人发音嘴型动画能够很好地模拟真实发音动作。目前，该系统还处于初步模型，在临床试用时仍有不足之处，表现在：

(1) 现有的三维发音模拟器官包括嘴唇、下巴、牙齿和舌，该系统可模拟这些器官的基本运动，而对于发音过程中的具体运动，如声带震动、气流控制等模拟不足，因此在接下来的研究中需要关注这些模拟表现形式；

(2) 系统中的训练语料不够丰富，仍需要进一步扩充虚拟人发音嘴型的内容，完善语音数据库。因此，我们需要根据临床反映对系统进行进一步的优化和升级，使得该系统能够为言语失用症的患者康复提供更加完善的帮助。

参考文献：

[1] 汪洁. 运动性失语症口语表达障碍的语言治疗[J]. 现代康复, 2001(19): 20-22.
Wang Jie. Speech therapy for motor aphasia with speech disorders[J]. Modern Rehabilitation, 2001(19): 20-22.

- [2] 王芹, 陶静, 刘斐雯, 等. 失用症的分类及治疗研究现状[J]. 实用医学杂志, 2016, 32(5): 689-692.
Wang Qin, Tao Jing, Liu Feiwei, et al. Classification and treatment of apraxia[J]. Journal of Practical Medicine, 2016, 32(5): 689-692.
- [3] 潘翠兰, 毛韶丽. 言语失用症康复训练报告[J]. 现代康复, 2000(6): 851.
Pan Cuilan, Mao Shaoli. A report on rehabilitation training for speech apraxia [J]. Modern Rehabilitation, 2000(6): 851.
- [4] Ogourtsova T, Archambault P, Lamontagne A. Post-stroke unilateral spatial neglect: virtual reality-based navigation and detection tasks reveal lateralized and non-lateralized deficits in tasks of varying perceptual and cognitive demands [J]. J Neuroeng Rehabil (S1743-0003), 2018, 15(1): 34.
- [5] Maier M, Banuelos N, Ballester B, et al. Conjunctive rehabilitation of multiple cognitive domains for chronic stroke patients in virtual reality [C]. International Conference on Rehabilitation Robotics. New Jersey: IEEE, 2017: 947-952.
- [6] 王辉, 吴吉生. 虚拟现实训练对认知障碍的脑卒中偏瘫患者的影响[J]. 中国康复, 2017, 32(4): 299-301.
Wang hui, Wu Jisheng. Effect of virtual reality training on lower extremity function and ADL in hemiplegics with cognitive impairment after stroke[J]. China Rehabilitation, 2017, 32(4): 299-301.
- [7] 刘婷婷, 刘箴, 钱平安, 等. 虚拟现实在特殊人群康复中的应用研究[J]. 系统仿真学报, 2018, 30(9): 3229-3237.
Liu Tingting, Liu Zhen, Qian Pingan, et al. Application of virtual reality in rehabilitation of special populations[J]. Journal of System Simulation, 2018, 30(9): 3229-3237.
- [8] 祝路, 吴春凤, 叶卉. 综合康复护理在脑卒中后言语障碍患者中的应用效果观察[J]. 中国医学创新, 2019, 16(12): 80-83.
Zhu Lu, Wu Chunfeng, Ye Hui. Application effect of comprehensive rehabilitation nursing in patients with post-stroke speech impairment [J]. Chinese Medical Innovation, 2019, 16(12): 80-83.
- [9] 孙咏虹, 吴冰洁. 丰富康复训练与神经可塑性[J]. 中国康复理论与实践, 2010, 16(7): 635-637.
Sun Yonghong, Wu Bingjie. To enrich rehabilitation training and neuroplasticity[J]. Rehabilitation Theory and Practice in China, 2010, 16(7): 635-637.
- [10] 王进嘉, 张喆. 虚拟现实技术在医疗领域应用的发展建议[J]. 电信网技术, 2017(2): 14-18.
Wang Jinjia, Zhang Zhe. Development suggestions on application of virtual reality technology in the medical field [J]. Telecommunication Network Technology, 2017(2): 14-18.
- [11] Miller P. Learning with a missing sense: what can we learn from the interaction of a deaf child with a turtle[J]. American Annals of the Deaf(S0002-726X), 2009, 154(1): 71-82.
- [12] Eden S. Virtual intervention to improve storytelling ability among deaf and hard of hearing children[J]. European Journal of Special Needs Education (S0885-6257), 2014, 29(3): 370-386.
- [13] 杨雪, 雷江华. 听障儿童基于虚拟现实技术的干预研究[J]. 昆明学院学报, 2019, 41(2): 39-44.
Yang Xue, Lei Jianghua. A review of virtual reality research into interventions for children with hearing impairment[J]. Journal of Kunming University, 2019, 41(2): 39-44.
- [14] 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. Berlin: Springer, 2007, 4554: 786-794.
- [15] Rathinavelu A, Yuvaraj G. Data visualization model for speech articulators [C]. Proceedings of AICERA. New Jersey: IEEE, 2011: 155-159.
- [16] 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. New York: ACM, 2005: 36-43.
- [17] Moccozet L, Laurent N. Dirichlet free-form deformations and their application to hand simulation [C]. Proceedings of the Computer Animation. New Jersey: IEEE, 1997: 93-102.
- [18] Adams J. Direct X. 高级动画制作[M]. 刘刚译. 重庆: 重庆大学出版社, 2005.
Adams, J. Direct X. Advanced animation[M]. Liu Gang, Trans.. Chongqing: Press of Chongqing University, 2005.
- [19] 周炳良, 邓立新, 洪民江. 一种新的基于 DTW 的孤立词语音识别算法[J]. 计算机技术与发展, 2018, 28(4): 127-131.
Zhou Bingliang, Deng Lixin, Hong Minjiang. A novel isolated word algorithm of speech recognition based on DTW[J]. Computer Technology and Development, 2018, 28(4): 127-131.