

文章编号:1001-9081(2015)01-0257-05

doi:10.11772/j.issn.1001-9081.2015.01.0257

基于发音特征的声效相关鲁棒语音识别算法

晁 浩^{*}, 宋 成, 彭维平

(河南理工大学 计算机科学与技术学院, 河南 焦作 454000)

(* 通信作者电子邮箱 chaohao@hpu.edu.cn)

摘要:针对声效(VE)相关的语音识别鲁棒性问题,提出了基于多模型框架的语音识别算法。首先,分析了不同声效模式下语音信号的声学特性以及声效变化对语音识别精度的影响;然后,提出了基于高斯混合模型(GMM)的声效模式检测方法;最后,根据声效检测的结果,训练专门的声学模型用于耳语音识别,而将发音特征与传统的谱特征一起用于其余4种声效模式的语音识别。基于孤立词识别的实验结果显示,采用所提方法后语音识别准确率有了明显的提高:与基线系统相比,所提方法5种声效的平均字错误率降低了26.69%;与声学模型混合语料训练方法相比,平均字错误率降低了14.51%;与最大似然线性回归(MLLR)自适应方法相比,平均字错误率降低了15.30%。实验结果表明:与传统谱特征相比发音特征对于声效变化更具鲁棒性,而多模型框架是解决声效相关的语音识别鲁棒性问题的有效方法。

关键词:语音识别; 声效; 发音特征; 多模型框架; 孤立词

中图分类号: TP391.42; TN912.34 **文献标志码:**A

Robust speech recognition algorithm based on articulatory features for vocal effort variability

CHAO Hao^{*}, SONG Cheng, PENG Weiping

(College of Computer Science and Technology, Henan Polytechnic University, Jiaozuo Henan 454000, China)

Abstract: Aiming at the problem of robust speech recognition for Vocal Effort (VE) variability, a speech recognition algorithm based on multi-model framework was presented. Firstly, changes of acoustic characteristics under different VE modes, as well as influence of these changes on speech recognition, were analyzed. Secondly, a VE detection method based on Gaussian Mixture Model (GMM) was proposed. Finally, the special acoustic models were trained to recognize whisper speech if the result of VE detection was whisper mode; otherwise articulato features, in company with spectrum features, were introduced to recognize speech of the remaining four VE modes. The experiments conducted on isolated-word recognition show that significant improvement of recognition accuracy can be achieved when using proposed method: compared with the baseline system, the mixed corpus training method and the Maximum Likelihood Linear Regression (MLLR) adaptation method, the average character error rate of five VE modes is reduced by 26.69%, 14.51% and 15.30% respectively. These results prove that articulato feature is more robust than the traditional spectrum feature when confronting VE variability, and the multi-model framework is an efficient method for robust speech recognition related to VE variability.

Key words: speech recognition; Vocal Effort (VE); articulatory feature; multi-model framework; isolated-word

0 引言

现有的语音识别研究通常针对正常情况下的语音,忽视了语音信号的声效(Vocal Effort, VE)变化。而在现实的环境中,人们不可能一直都在同一种声效水平下交流:在图书馆或者自习室里需要通过耳语的方式交流;在嘈杂的场合需要大声说话对方才能听见;而在嘈杂的工厂车间可能就需要通过高喊的方式才能够交流。声效的改变不仅仅意味着语音的强度变化,还会使得语音信号的其他声学特性如基频、共振峰发生偏移^[1],进而影响语音识别系统的精度。因此,研究声效相关的鲁棒语音识别算法对于提高语音识别的准确率是十分必要的,也成为近年来语音识别领域里的研究热点^[2-3]。

声音效果(声效)被研究人员定义为正常人的一种发音变化的衡量,而这种发音变化是人出于正常交流的需要,根据

交流时双方距离的远近或背景噪声的高低自动调整发音方式所产生的^[1]。通常将声音效果由低到高分为5个量级/模式:耳语(whisper)、轻声(soft)、正常(normal)、大声(loud)、高喊(shouted)。

近年来研究人员探讨了语音信号在不同声音效果量级下的声学特性和语音特性,尝试发现蕴含声音效果量级的显著性信息的声学参数。如文献[4]发现声强级(Sound Intensity Level, SIL)会随着声效量级的提高而上升,而且能量相关的参数蕴含了重要的声音效果量级信息;文献[5]发现除耳语模式外声音效果的量级升高会伴随着频谱倾斜的下降;文献[6]阐述了随着声音效果变化对基频(Fundamental frequency, F0)的影响;文献[7]对耳语模式下塞尔维亚语中辅音的声学特性进行了分析。上述研究表明声效变化会造成语音信号特性的改变。

收稿日期:2014-08-19;修回日期:2014-09-30。 基金项目:国家自然科学基金资助项目(61300124);河南省基础与前沿技术研究计划资助项目(132300410332);河南省科技厅科技攻关计划项目(132102210123);河南省教育厅科技攻关计划项目(13A520321)。

作者简介:晁浩(1981-),男,河南许昌人,讲师,博士,主要研究方向:语音识别、模式识别; 宋成(1980-),男,河南信阳人,讲师,博士,主要研究方向:信息安全; 彭维平(1979-),男,湖北天门人,副教授,博士,主要研究方向:智能计算、信息安全。

声效相关的鲁棒语音识别算法是近几年才兴起的研究领域,5种声效模式中耳语音发音机制与其他4种声效模式有着明显的区别,所以国内外研究人员主要对耳语音的语音处理进行了广泛而深入的研究^[8-11],取得了较好的效果。对于其他几种声效模式下的语音识别研究则鲜有报道。相关的研究主要包含以下几方面:1)特征层面。文献[12]通过提取鲁棒性谱特征来降低轻声和大声这两种声音效果模式对语音识别性能的影响,但识别结果只取得了轻微的提高。2)模型参数优化。采用自适应方法对声学模型参数进行优化,然后识别对应声效模式下的语音^[13]。这种方法对于发音方式差异较大的声效模式(如耳语和正常语音)效果不甚理想。文献[13]还利用多模式下的混合语料对声学模型进行训练,其对语音识别性能的提高有限,理论上说其根本达不到使用对应模型进行识别时的精度。

本文主要在特征层面研究声效相关的鲁棒语音识别算法。针对耳语外其他4种声效发音方式未有明显改变的特点,将蕴含发音方式信息的发音特征引入语音识别系统,以提高声效鲁棒性。同时考虑到耳语音自身独特的发音机制,提出基于耳语音检测的多模型语音识别框架。

1 不同声效模式下声学特性分析及其影响

声效模式的改变会造成语音信号特性的变化,下面针对语料库训练集,从声强级、时长、帧能量分布以及频谱倾斜等4个方面分析声效模式对语音信号的影响。

1.1 声强级

声强是指声音在传播途径上单位面积的声能流密度。人们对声音强弱的感觉并不是与声强成正比,而是与其对数成正比的。这正是人们使用声强级来表示声强的原因,而声强级是以分贝衡量音质的重要参数。图1为5种声效模式下语音信号的声强级对比,分别描述了各个声效模式下语音的声强级均值和标准差。其中:Wh表示耳语,So表示轻声,No表示正常,Lo表示大声,Sh表示高喊。

从图1中可以看出,随着声效模式从耳语到高喊的转变,声强级逐渐提升,其均值变化比较明显。此外,观察图1中声效模式对应声强级的标准差,5种声效中正常语音的标准差最低(No:5.11),而当声效从正常模式向上转向高喊模式或者向下转向耳语模式时,声强级的标准差均会有显

著的提高(Sh:6.89;Wh:7.62)。

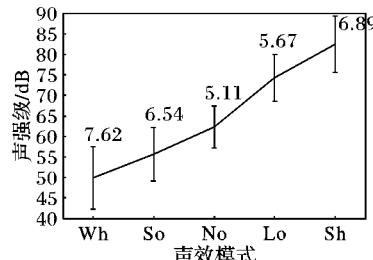


图1 5种声效模式的声强级均值及标准差

1.2 音节时长

本文对5种声效模式下音节的时长变化同样作了分析。音节的边界通过人工切分去除静音部分。为了消除音节不同带来的时长变化,耳语、轻声、大声和高喊模式下每个音节的时长都利用该音节在正常模式下的时长进行归一化处理,而后每种声效模式下音节的时长求均值,具体如图2所示。

正常模式下(No)的音节时长经过归一化处理后均为1,所以该模式下所有音节时长的均值也为1.00。剩下4种声效模式,耳语(Wh)中音节的平均时长最短,为0.78;高喊模式下音节的平均时长最高,为1.16;轻声模式的平均时长位于耳语和正常模式之间,大声模式的平均时长则位于正常模式和高喊模式之间。

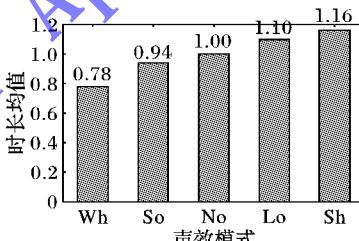


图2 5种声效模式下归一化以后的音节时长均值

1.3 帧能量分布

图3分别为5种声效模式下语音的帧能量分布柱状图,从中可以看出不同声效模式下语音的能量分布发生了明显的变化。图3中的低能量帧可以视为静音和摩擦音,由于发音时声带基本不震动,耳语音中的静音和摩擦音明显高于其他声效模式。另外根据柱状图总体形状的改变和峰值位置的转移可以看出,随着语音的声强级提高,语音帧的平均能量也会提高,同时代表响音的高能量帧的数目也会提升。

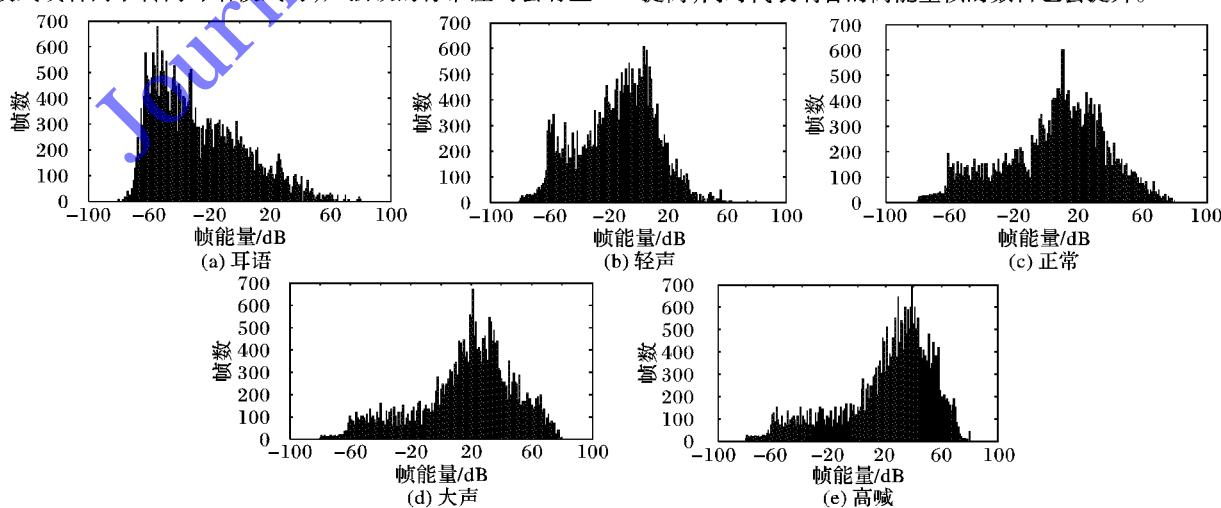


图3 5种声效模式下帧能量分布

1.4 频谱倾斜

最后是频谱倾斜的变化。频谱倾斜的计算方法参考了文献[14]中的方法。根据语音帧的能量分布情况确定一个阈值。能量高于阈值的语音帧作为潜在的发音部分被选出;而低于阈值的语音帧,作为潜在的静音和摩擦音则被移除。对于选出的语音帧,计算其谱周期图斜度;然后对得到的周期图求平均;最后利用一元线性回归来计算频谱的斜度(单位为分贝/倍频程,dB/Octave)。表1给出了5种声效模式下语音信号的频谱倾斜均值。

表1 5种声效下的平均频谱倾斜 dB/Octave

声效模式	频谱倾斜	声效模式	频谱倾斜
耳语	-2.65	大声	-7.61
轻声	-5.92	高喊	-6.42
正常	-7.87		

从表1中可以看出,与正常语音相比,无论是耳语、轻声、大声还是高喊,其频谱倾斜均有所下降。耳语和轻声模式下,语音强度降低的同时高频带能量却有所提高,这就造成了频谱的倾斜度下降;大声和高喊模式下语音频谱倾斜下降则可能是由形状更规范的声门脉冲造成的。

1.5 声效变化对语音识别的影响

上述分析表明,声效模式的不同会造成语谱的改变,进而影响语音识别的精度。为了评估这种影响,本文搭建了孤立词识别系统来进行实验。对于每种声效模式,分别利用训练集中对应声效模式的语料训练其专用声学模型,总有5个声学模型集合;然后利用每种声效模式的专用声学模型集,对测试集中各种声效模式下的语料进行识别。识别结果如表2所示。

表2 声效模式专用声学模型的测试集字错误率结果 %

声学模型	测试集字错误率				
	耳语	轻声	正常	大声	呼喊
耳语	0.32	41.21	32.15	48.62	52.48
轻声	59.34	2.86	1.22	63.75	72.42
正常	55.87	12.95	0.42	58.65	63.51
大声	52.15	38.93	18.89	0.97	19.62
呼喊	51.44	43.76	38.69	20.12	2.18

表2中第3行显示,正常情况下的语音识别系统(用正常模式下语料训练)在识别其他声效下的语音时,其字错误率均有大幅度的提升。实际上,不管是耳语、轻声、正常,还是大声和呼喊模式,该模式专用的声学模型用于识别对应模式的测试语料时,识别精度均很高;而识别其他声效模式下的测试语料时,由于声学模型的训练环境和测试环境不匹配,其精度均有大幅度的下降。特别是对于声效量级差别比较大的声效模式(如耳语和呼喊,轻声和呼喊等),识别对方模式的语料时其精度更低。这就表明声效模式改变对造成语音识别精度的急剧下降,对语音识别系统的性能影响较大。

2 基于发音特征的声效相关语音识别算法

2.1 基于多模型框架的解码算法

耳语音发音时激励源是噪声,声带不振动,韵母部分和浊辅音声母部分没有基频,发音机制的改变使得耳语音与其他声效语音的声学特性有着明显的不同。剩余4种声效模式其

发音的最大区别是声道激励的强度,尽管也会造成语音的声学特性和语音特性的改变,但其发音方式和发音器官动作并未有明显的区别。这也就意味着除耳语外,声效模式的变化不会对发音动作和发音方式造成较大的影响。一般来说,蕴含发音方式和发音器官动作信息的发音特征相对独立于声学环境的变化,这其中就包括说话人自身语音的多变性,即发音特征不易受到说话人生理、心理、情绪、说话速率等因素的影响,而声音效果本质上也是属于说话人自身语音的特性。因此,本文考虑将蕴含发音信息的发音特征引入到声效相关的鲁棒语音识别系统。而对于耳语音的识别,则为其训练了专门的声学模型,整个识别过程采用多模型框架:即为耳语音训练专门的声学模型集,而剩余4种声效模式则对应一个声学模型集。具体的识别算法如图4所示。

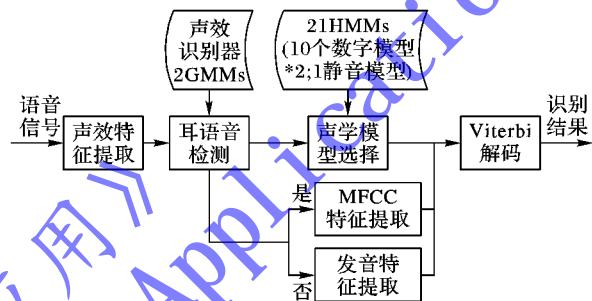


图4 多模型框架的语音识别算法示意图

首先提取用于检测声效模式的区分性特征,这里所用的特征即为第1章分析的4种语音/声学特性:声强级、时长、帧能量均值以及频谱倾斜。声效的检测利用高斯混合模型(Gaussian Mixture Model, GMM)来实现,对于耳语模式训练一个对应的GMM,剩下的4种声效模式训练一个GMM。对于待识别的语音信号,根据2个GMM概率得分的大小判断该信号是否是耳语音;然后根据检测结果选择声学模型集合。共有两个声学模型集合:一个对应耳语模式,另一个对应非耳语模式。每个集合中包含10个声学模式,分别对应10个阿拉伯数字(0~9),每个数字训练一个隐马尔可夫模型(Hidden Markov Model, HMM),除此之外还有1个共用的静音模型,所以声学模型集共有21个HMM。同时根据声效的检测结果提取特征,如果判断是耳语音则提取谱特征如梅尔频率倒谱系数(Mel-Frequency Cepstral Coefficient, MFCC);如果检测结果非耳语音则提取发音特征,将发音特征与MFCC一起用于后续的解码。确定了声学模型和前端特征后,利用Viterbi算法进行解码。

2.2 发音特征

除耳语音外,声效模式的变化不会对发音动作和发音方式造成较大的影响,那么发音特征对于其余4种声效模式的语音识别会具有较好的鲁棒性。因此,在图4描述的算法中,如果判断当前语音不是耳语音,那么就提取发音特征,并利用发音特征完成后续的解码。

本文所用的发音特征集主要包括:“发音位置”“发音方式”“是否送气”等8种发音特征,每种发音特征包含各自的所属类别,总共41类,如表3所示。

发音特征的提取主要采用统计方法:首先训练统计模型;然后利用模型计算语音信号属于发音特征的各类别的后验概率。在本文采用神经网络计算语音帧属于发音特征类别的后

验概率,如图 5 所示。对于表 3 中的每一种发音特征,都需要训练一个多层感知器与之对应,因此一共要训练 8 个多层感知器(Multi-Layer Perception, MLP)。MLP 的输入数据采用频谱特征(如 MFCC),输出数据的维数(或者说输出层节点的个数)与每个发音特征的所属类别数相同。例如,发音位置对应的 MLP,其输出层节点的个数为 8;然后将每个 MLP 的输出连接起来形成一个新的矢量;最后通过主成分分析(Principal Component Analysis, PCA)降维来消除冗余信息,从而得到发音特征。

表 3 发音特征及其包含的类别

发音特征名称	类别	类别数目
发音位置	双唇音 唇齿音 舌尖前音 舌尖中音 舌尖后音 舌面前音 舌面后音 NULL	8
发音方式	塞音 鼻音 擦音 边音 塞擦音 NULL	6
是否送气	送气音 非送气音 其他辅音 NULL	4
声带振动	清音 浊音	2
舌位高低	高 央高 央 央低 低 NULL	6
舌位前后	前 中前 中 中后 后 NULL	6
嘴唇形状	圆唇 非圆唇 NULL	3
声调类型	一声 二声 三声 四声 轻声 NULL	6

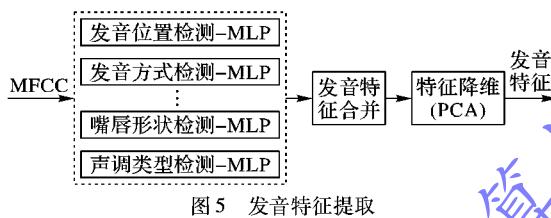


图 5 发音特征提取

3 实验及分析

3.1 实验语料

实验所用的训练集包含 2500 个数字(0~9),其中每种声效模式(耳语、轻声、正常、大声、呼喊)包含 500 个数字,分别由相同的 10 个男说话人录制,即每个说话人录制 5 次全部 0~9 的数字。测试集中每种声效模式包含 300 个数字,仍然由上述 10 个男说话人录制,每个说话人录制 3 次全部 0~9 的数字。录音环境为安静的实验室环境,采样率 16 kHz,采样位数 16 bits。

前端声学特征包括 12 维梅尔频率倒谱系数(MFCC)及 1 维标准化能量,以及它们的一阶及二阶差分,窗长为 25.6 ms,帧移为 10 ms。

3.2 声效检测模型及声学模型

声效检测模型采用 GMM,混合数为 8。声学模型均采用 HMM,其中 10 个数字的声学模型由 5 状态组成,3 个为发射状态,每个发射模型由混合数为 16 的 GMM 进行模拟,静音段模型的 HMM 共有 3 个状态。声学模型的训练以及最终的解码均采用语音识别软件工具 HTK V3.2.1 构建^[15]。发音特征提取使用 8 个 MLP,每个 MLP 的输入层有 39 个节点,中间隐含层的节点数为输入层的 2 倍,输出层节点数与每个发音特征的所属类别数相同。

3.3 系统搭建

实验中共搭建了 4 种孤立词识别系统,除了用正常模式下的语料训练声学模型建立起的基线系统和本文所提算法外;还有一个系统利用 5 种声效模式下的混合语料对声学模

型进行训练;最后一个系统采用声学模型优化方法,同样采用正常模式下的语料训练声学模型,识别非正常模式下的语音时利用最大似然线性回归 (Maximum Likelihood Linear Regression, MLLR) 自适应方法对声学模型参数进行优化,然后再进行识别。

3.4 实验结果分析

首先对测试集中语料的声效模式进行了识别,检测结果如表 4 所示。可以看出,耳语和其他声效的检测精度都比较高。特别是对其他 4 种声效的检测,其准确率达到 99.8%。这是由于耳语音的发音机制与其他声效模式有着明显的区别,反映在声学特性和语音特性上也有明显不同,所以耳语和其他声效之间有着较高的区分性。而考虑到声效检测错误会影响后续的语音识别精度,可靠的声效检测结果也是必需的。

表 4 声效模式的准确率结果 %

实际声效	声效模式准确率	
	耳语	非耳语
耳语	97.6	2.4
非耳语	0.2	99.8

最后根据本文提出的基于发音特征的语音识别算法,本文对测试集中的语料进行了孤立词识别实验,结果如表 5 所示。

表 5 孤立词的测试集字错误率结果 %

系统类型	测试集字错误率					
	耳语	轻声	正常	大声	呼喊	平均
基线系统	55.87	12.95	0.42	58.65	63.51	38.28
混合训练模型	34.36	27.43	17.33	23.28	28.12	26.10
MLLR 自适应	49.86	9.68	0.42	32.19	42.31	26.89
本文算法	1.64	8.76	7.53	18.92	21.12	11.59

从表 5 中可以看出,与基线相比后面的 3 种系统的平均字错误率均有不同程度的提高。其中:混合训练模型对于耳语、轻声、大声以及呼喊模式下的测试语料,其识别精度有着均衡的提高;而对于正常模式下的语料,其字错误率则从 0.42% 提高到 17.33%。该方法对语音识别性能的提高有限,理论上说其根本达不到使用对应模型进行识别时的精度。而 MLLR 自适应方法对于与正常语音的发音方式差异较大的声效模式(如耳语)效果不甚理想;而对于与正常语音的发音方式接近的声效模式(轻声和大声)则效果显著。与混合训练模型方法和 MLLR 自适应方法相比,本文所提算法除正常模式下语音识别的精度略有下降外,其他 4 种声效模式下语音识别精度均有大幅度的提高。特别是对应耳语,由于训练了专门的声学模型,在检测出耳语后由专门的声学模型进行识别,其准确率有着大幅度的提升。

在计算复杂度方面,发音特征的提取主要是多层感知器的计算。由于实验采用的 8 个多层感知器的各层节点总数 T 大约为 1000 左右。对于每一帧语音,特征提取大约需要 T 个 sigmoid 函数的运算,时间复杂度为 $O(T)$ (T 的量级为 10^3)。计算某时刻的某个状态的前向变量需要比较前一时刻的 N 个状态,此时时间复杂度为 $O(N)$,每个时刻有 N 个状态,此时时间复杂度为 $N * O(N) = O(N^2)$ (N 为状态数,量级为 10^3)。因此,特征提取的计算复杂度要远远小于解码的时间复杂度,特征的提取并不会显著增加总的解码时间。

4 结语

针对语音信号的声效变化,本文提出了一种基于发音特征的语音识别算法。根据耳语音与其他声效的发音机制差异明显的特点,训练了耳语对应的声学模型集,并建立了基于声效检测的多模型语音识别框架。同时利用人工神经网络获取蕴含发音特性信息的发音特征,然后将发音特征和MFCC特征一起用于解码。实验中根据所提方法的字错误率从基线系统的38.28%下降到11.59%,特别是对耳语音的识别,其字错误率从55.87%降到1.64%。实验结果表明:1)将发音特征应用与声效相关的鲁棒性语音识别能够取得较好的效果;2)当声效模式检测准确时,基于多模型框架的语音识别算法对于降低声效对语音识别精度的影响具有很大的潜力。

参考文献:

- [1] TRAUNMÜLLER H, ERIKSSON A. Acoustic effects of variation in vocal effort by men, women, and children [J]. Journal of the Acoustical Society of America, 2000, 107(6): 3438 – 3451.
- [2] ZELINKA P, SIGMUND M, SCHIMMEL J. Impact of vocal effort variability on automatic speech recognition [J]. Speech Communication, 2012, 54(6): 732 – 742.
- [3] RAITIO T, SUNI A, POHJALAINEN J, et al. Analysis and synthesis of shouted speech [C]// INTERSPEECH 2013: Proceedings of the 14th Annual Conference of the International Speech Communication Association. [S. l.]: ISCA, 2013: 1544 – 1548.
- [4] ZHANG C, HANSEN J H L. Analysis and classification of speech mode: whispered through shouted [C]// INTERSPEECH 2007: Proceedings of the 8th Annual Conference of the International Speech Communication Association. [S. l.]: ISCA, 2007: 2289 – 2292.
- [5] BOU-GHAZALE S, HANSEN J H L. HMM-based stressed speech modeling with application to improved synthesis and recognition of isolated speech under stress [J]. IEEE Transactions on Speech Audio Processing, 1998, 6(3): 201 – 216.
- [6] LU Y, COOKE M. The contribution of changes in F0 and spectral tilt to increased intelligibility of speech produced in noise [J]. Speech Communication, 2009, 51(12): 1253 – 1262.
- [7] JOVICIC S T, SARIC Z. Acoustic analysis of consonants in whispered speech [J]. Journal of Voice, 2008, 22(3): 263 – 274.
- [8] ZHANG C, HANSEN J H L. An entropy based feature for whisper-island detection within audio streams [C]// INTERSPEECH 2008: Proceedings of the 9th Annual Conference of the International Speech Communication Association. [S. l.]: ISCA, 2008: 2510 – 2513.
- [9] ZHANG C, HANSEN J H L. Advancements in whisper-island detection within normally phonated audio streams [C]// INTERSPEECH 2009: Proceedings of the 10th Annual Conference of the International Speech Communication Association. [S. l.]: ISCA, 2009: 860 – 863.
- [10] LI X, DING H, XU B. Entropy-based initial / final segmentation for Chinese whiskered speech [J]. Acta Acustica, 2005, 30(1): 69 – 75. (栗学丽, 丁慧, 徐柏龄. 基于熵函数的耳语音声韵分割法[J]. 声学学报, 2005, 30(1): 69 – 75.)
- [11] YANG L, LIN W, XU B. Isolated word recognition in Chinese whispered speech [J]. Applied Acoustics, 2006, 25(3): 187 – 192. (杨莉莉, 林伟, 徐柏龄. 汉语耳语音孤立字识别研究[J]. 应用声学, 2006, 25(3): 187 – 192.)
- [12] MEYER B T, KOLLMEIER B. Robustness of spectro-temporal features against intrinsic and extrinsic variations in automatic speech recognition [J]. Speech Communication, 2011, 53(5): 753 – 767.
- [13] ITOH T, TAKEDA K, ITAKURA F. Analysis and recognition of whispered speech [J]. Speech Communication, 2005, 45(2): 139 – 152.
- [14] HANSEN J H L. Analysis and compensation of stressed and noisy speech with application to robust automatics recognition [D]. Atlanta: Georgia Institute of Technology, 1988.
- [15] YOUNG S, EVERMANN G, GALES M, et al. Hidden Markov model toolkit [EB/OL]. [2013-12-22]. <http://htk.eng.cam.ac.uk/docs/docs.shtml>.

(上接第251页)

- [5] MEISEL F, BIERWIRTH C. A unified approach for the evaluation of quay crane scheduling models and algorithms [J]. Computers & Operations Research, 2011, 38(3): 683 – 693.
- [6] LEGATO P, TRUNFIO R, MEISEL F. Modeling and solving rich quay crane scheduling problems [J]. Computers & Operations Research, 2012, 39(9): 2063 – 2078.
- [7] LE M, ZHAO Y, LIU X. Shore-mounted gantry crane scheduling for single vessel under time window-using mathematical programming and rule heuristic algorithm [J]. Computer Engineering and Applications, 2014, 50(9): 242 – 248. (乐美龙, 赵彦营, 刘秀玲. 时间窗下单船岸桥调度——基于数学规划和规则的启发式算法[J]. 计算机工程与应用, 2014, 50(9): 242 – 248.)
- [8] MEISEL F. The Quay crane scheduling problem with time windows [J]. Naval Research Logistics, 2011, 58(7): 619 – 636.
- [9] LEGATO P, MAZZA R M, TRUNFIO R. Simulation-based optimization for discharge/loading operations at a maritime container terminal [J]. OR Spectrum, 2010, 32(3): 543 – 67.
- [10] CHUNG S H, CHOY K L. A modified genetic algorithm for quay crane scheduling operations [J]. Expert Systems with Applications, 2012, 39(4): 4213 – 4221.
- [11] SU N, ZHANG M, JOHNSTON M, et al. Hybrid evolutionary computation methods for quay crane scheduling problems [J].

- Computers & Operations Research, 2013, 40(8): 2083 – 2093.
- [12] LIU Z, HAN X, XI L, et al. A heuristic for the quay crane scheduling problem based on contiguous bay crane operations [J]. Computers & Operations Research, 2012, 39(12): 2915 – 2928.
- [13] CHEN J, LEE D-H, CAO J. Heuristics for quay crane scheduling at indented berth [J]. Transportation Research Part E: Logistics and Transportation Review, 2011, 47(6): 1005 – 1020.
- [14] FAN Z, LE M. A genetic algorithm to minimize the makespan and waiting time for the bi-objective quay crane scheduling problem [J]. Journal of Systems and Management, 2013, 22(1): 120 – 127. (范志强, 乐美龙. 最小化最大完工时间与等待时间的岸桥作业调度双目标优化及其遗传算法[J]. 系统管理学报, 2013, 22(1): 120 – 127.)
- [15] DIABAT A, THEODOROU E. An integrated quay crane assignment and scheduling problem [J]. Computers & Industrial Engineering, 2014, 73: 115 – 123.
- [16] CAI G, DONG E. Comparison and analysis of generation algorithm and ant colony optimization on TSP [J]. Computer Engineering and Applications, 2007, 43(10): 96 – 98. (蔡光跃, 董恩清. 遗传算法和蚁群算法在求解TSP问题上的对比分析[J]. 计算机工程与应用, 2007, 43(10): 96 – 98.)