

纪念应崇福院士诞辰 95 周年

语音中元音和辅音的听觉感知研究*

颜永红[†] 李军锋 应冬文

(中国科学院声学研究所 中科院语言声学与内容理解重点实验室 北京 100190)

摘要 本文对语音中元音和辅音的听觉感知研究进行综述。80 多年前基于无意义音节的权威实验结果表明辅音对人的听感知更为重要, 由于实验者在学术上的成就和权威性, 这一结论成为了常识, 直到近 20 年前基于自然语句的实验挑战了这个结论并引发了新一轮的研究。本文主要围绕元音和辅音对语音感知的相对重要性、元音和辅音的稳态信息和边界动态信息对语音感知的影响以及相关研究的潜在应用等进行较为系统的介绍, 最后给出了总结与展望。

关键词 语言声学, 语音产生, 语音感知, 可懂度, 元音, 辅音

中图分类号: O429

文献标识码: A

文章编号: 1000-310X(2013)03-0231-06

Perception of vowels and consonants in speech

YAN Yonghong LI Junfeng YING Dongwen

(Key Laboratory of Speech Acoustics and Content Understanding Institute of Acoustics,
Chinese Academy of Sciences, Beijing 100190, China)

Abstract This paper reviews the recent advances of speech perception research on relative importance of vowels and consonants in speech utterances. Almost 80 years ago, experiments conducted on meaningless syllables by a highly respected speech authority indicated that consonant was more important. This conclusion is widely accepted as common sense during the years after. In mid 1990s, a new experiment on continuous speech challenged this conclusion and thus incurred a new round of research on this subject. The main attention is paid to the relative importance of vowels and consonants in speech perception, the perception of steady-state acoustic cues and the transition cues at the vowel-consonant boundary, and the potential applications of their related research. The short conclusion and prospect are described.

Key words Speech acoustics, Speech production, Speech perception, Speech intelligibility, Vowel, Consonants

2013-04-18 收稿; 2013-04-21 定稿

*国家自然科学基金(10925419, 90920302, 61072124, 11074275, 11161140319, 91120001, 61271426), 中国科学院战略性先导科技专项(面向感知中国的新一代信息技术研究, XDA06030100, XDA06030500), 国家 863 计划(2012AA012503), 和中科院重点部署项目(KGZD-EW-103-2)资助项目

作者简介: 颜永红 (1967-), 男, 江苏无锡市人, 研究员, 研究方向: 语音信号与信息处理。

李军锋 (1979-), 男, 河南安阳人, 研究员。

应冬文 (1975-), 男, 湖北黄陂人, 副研究员。

[†]通讯作者: 颜永红, E-mail: yanyonghong@hcl.ioa.ac.cn

1 引言

语言声学是用声学方法来研究和人类语言相关的声音的一门科学,是声学的一个重要分支。其研究内容包括:语音生成、传播、感知、理解及其在计算机上的实现等。由于语音本身所具有的复杂性,相关研究不可避免地呈现出跨学科的特点,涉及的学科包括声学、语言学、生理学、心理学、认知科学、计算科学和信息学等。这部分地反映在“言语链”(图1)所示意的人类言语交际过程中。

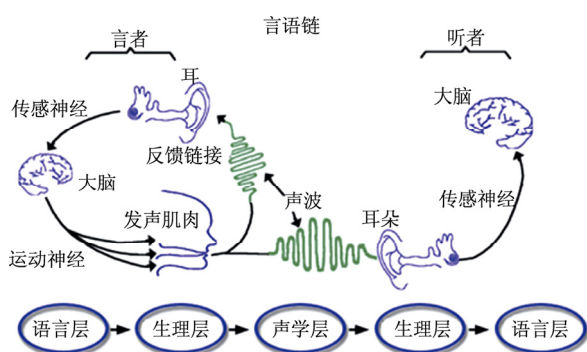


图1 人类言语交际的“言语链”^[1]

作为语言声学的一个重要内容,语音听觉感知主要研究人耳是如何对语音信号进行听取、感知和理解的,是语音识别、语音人机交互和语音通信等众多应用的基础。本文将主要对语音中元音(Vowel或V)和辅音(Consonant或C)的听觉感知研究(尤其是它们对语音理解的贡献)进行较为系统的综述。

2 元音和辅音对语音听觉感知的相对重要性研究

能听会说作为人们与生俱来的一种能力,其获得似乎是很容易的。这掩盖了语音本身潜在的复杂性。其实直至今日,对于人为什么能够处理、听懂语音的机理尚不完全清楚,还有很多的方面需要长期深入的探索。语音可以抽象地描述为由一连串的字或音素组成,这是非常简单的;但是对语音进行声学分析是相当困难的。从语音理解的角度来讲,语音信号不容置疑的携带了充足的信息。但是什么是人类对语音信号进行感知的基本单元?哪一个单元携带了更多的感知信息呢?针对这两个相关问题,人们已经进行了大量的研究并取得了重大进

展。一般来讲,音素常常被认为是语音信号的最小单元,如同物理学中的原子。传统地来讲,音素可以分为元音和辅音两大类。语音可懂度研究的一个方向就是在语音信号的这两类基本特征上,试图理清清楚两者的相对重要性。

2.1 辅音对语音感知的重要性

在第一次世界大战之后对通信传输系统的强烈需求的驱动下,Fletcher从清晰度、元音和辅音引起的错误模式出发,研究了元音和辅音在CV、VC和CVC的单音节字和无意义音节中的相对重要性^[2]。研究结果是:辅音为孤立词识别提供了更多的有用信息,即语音的大部分信息是由辅音携带的。近年来,Owren和Cardillo利用分段替换方法产生了只有辅音或者只有元音的多音节字,进行语音可懂度和说话人辨别主观测听实验^[3]。结果表明:辅音对语音可懂度更为重要,元音携带了更多的说话人信息。总之,对语音理解或可懂度来讲,辅音比元音更为重要的结论已经被语音和临床听力科研人员所广泛接受,并成为了一个常识。

语音感知中辅音比元音更为重要的这一结论常用和文本处理相似的方法来解释。直观的解释为:因为辅音的数量比元音更多,因此辅音更具信息量;语音信息是冗余的,辅音携带了更多信息从而能够理解字的意思。从信息论的角度出发,Stilp和Kluender以文本单词为例,说明了字母不确定性决定了文本单词的可读程度^[4]。例如,在书写英文中辅音字母有20-21个,元音字母有5-6个。由于辅音字母的不确定性程度高于元音字母的不确定性,因此在保留辅音字母而去除元音字母情况下,一个单词的可读性高于保留元音字母而去除辅音字母的情况。美式英语口语有15个元音或双元音,25-26个辅音。因此,只保留辅音的字的可懂度要同样高于只保留元音的字的可懂度。

2.2 元音对语音感知的重要性

尽管辅音比元音对语音听觉感知更为重要已经广为接受并成为了“常识”,但是这一结论首次受到了Cole和Yan等人的质疑^[5]并得到了文献^[6]的呼应。他们采用噪声替换实验方法(noise-replacement paradigm),即用噪声替换选自TIMIT数据库的语音中的辅音或者元音,并将所得语音信号提供给年轻听力正常人进行主观测听实验。他们的研究表明:替换掉辅音的语音比替换掉元音

的语音具有更高的识别率(比率约为2:1)。也就是说,对语音感知来讲,元音比辅音携带了更多的信息,因此更为重要。

最近, Kewley-Port 等人采用文献[5]的噪声替换实验方法,研究了听力正常年轻人和听力损伤老年人而言,元音和辅音对语音感知的相对重要性^[7]。实验结果表明:听力损伤老年人的识别率远远低于听力正常年轻人的识别率;对听力正常年轻人和听力损伤老年人(当辅音的可听范围达到4000 Hz后)而言,仅保留元音的语句识别率和仅保留辅音的语句识别率的比值大约为2:1;相对于孤立词语来讲,元音比辅音携带了更多对语句可懂度有用的信息。Fogerty 等人将 Cole 等的研究延伸到老年人群,考察了年龄和听力损失是否影响元音和辅音在单词及句子听辨中的相对作用^[8-9]。他们的实验采用了 TIMIT 语料库中的语句和一些选自其他数据库中的单音节词,同样使用噪声分别替换元音部分或者辅音部分后作为听辨材料。此外,为了让听力正常老年人和听力受损老年人能在相似的听力水平上对单词和句子进行判听,他们针对听力受损老年人把语音频谱的低频和低频进行了抬升。实验结果表明:只有在保留元音并且语料为语句的情况下,听力受损老年人的单词听辨准确率才能和正常听力的年轻人和老年人在一个水平上,其他情况下听力受损老年人的准确率比另外两组被试的准确率低。他们的结论是:在句子语境下,元音对可懂度的贡献要高于在孤立词语境下的贡献;而且只要语音能达到听力受损老年人听阈的水平,元音携带的关键性听觉感知特征也能够被他们利用。对听力损失老年人来讲,可听区域的减小严重影响了他们融合利用多种信息对被噪声污染的语音信号进行理解的能力,相对于只有辅音或其他语音片段的语音来讲,只保留元音信息的信号对语句感知与理解具有更重要的益处。

由于自然语音具有多个相互作用、且不易单独控制的声学属性, DiGiovanni 等人利用合成语音的方法来实现对元音和辅音中不同属性(如个体特性、时长和强度)的控制,按照文献[5]的噪声替换实验方法,研究了在自然语音和合成语音中元音和辅音对可懂度的相对贡献^[10]。自然语音选自 TIMIT 语料库,使用噪声替换实验方法产生了未处理自然语音、元音保留自然语音和辅音保留自然语音。通

过主观测听实验,再次证明了自然语音中元音对可懂度的贡献高于辅音,这和文献[5]的实验结果是一致的。合成语音是由 ATT 的波形拼接技术产生的,形成未处理合成语音和元音保留合成语音(稳态元音)两种听辨条件。主观测听实验结果表明:稳态元音条件下的可懂度明显低于自然语音和合成语音的可懂度。可能的原因包括:稳态元音段缺少自然元音内在频谱变化和时域变化,同时也缺少与前后音素的协同发音特征。总的来讲,元音比辅音对自然语音可懂度贡献更大;未处理合成语音的可懂度与未处理自然语音的可懂度相当,元音保留合成语音与元音保留自然语音的听辨准确率相似。

尽管元音对语音感知理解更为重要的原因尚未完全明白,一些研究指出可能的原因包括^[5]:元音包含了更多的可以恢复的丢失辅音的协同发音信息;噪声替换实验方法中的替换噪声幅度包络信息可以给辅音提供更多的信息;元音比辅音具有更多的语句信息。

2.3 元音和辅音相对重要性的解释

强调辅音对语音感知理解(即可懂度)具有重要作用的研究主要是集中于单音节词或者多音节词;而强调元音对语音感知理解具有重要作用的研究都是采用自然语句的语音进行实验得到的。单音节和多音节词的言语处理主要依赖于分段的、从下到上的信息;而语句可懂度任务包含了大量的从上到下的预测信息。文献[2]中音素误差分析是基于离散符号来表示单音节词的。然而,错误识别的音素并不能正确反映一个特定的音素对词识别的影响,这是因为声学信息往往是分散在音素中的。噪声替换方法去除了特定的声学信息,从剩余的分散信息中来直接测量字的识别率。这样,“元音比辅音携带了更多的信息”的结论可以直接从没有任何特定干涉符号分析的听觉系统处理过后的声学信息来估计。人类通过自然流利语音中的基频包络、幅度包络和句子时长等信息来帮助理解语音,单个的字或词仅仅能够对自然口语语音的理解提供部分信息。因此,在自然流利语音的理解中,元音比辅音具有更重要的作用。

3 元辅音过渡段对语音可懂度的作用

将高度动态变化的声音信号分割为对应于元

音和辅音的离散单元本身就是模糊不清的,特别是将共振峰的变化分为辅音或元音,因为这些变化既包括辅音信息又包括元音信息。除了声音结构中的变化外,元音和辅音的声学信息很大程度上是重叠的、协同发音的,所以他们不能够很清晰地区分开来。进一步来讲,一个音素的声学结构总是强烈受到它周围音素的影响,以至于一个音素常常包含有其他音素的大量信息。元音、辅音及其协同发音的感知效果能够在多个语音片段观察到^[11]。

声学信号中与元音和辅音相关的感知信息相互作用。Cooper 等人首次研究了辅音中的声学信息和感知的关系,指出这些特定信息的感知取决于他们和邻近元音的位置关系^[12]。进一步来讲,语音的感知高度依赖于声学-音素上下文关系。因此,元音和辅音的感知以及他们对语句可懂度的贡献都是由分散和重叠的声学信息引起的。

Fogerty 和 Kewley-Port 通过将元音和辅音的边界移动到极限情况,发现共振峰过渡段变化对语句识别中的元辅音的贡献是不一样的^[8]。因此,相对于元音和辅音在语音感知理解中的作用,过渡段的影响已经达到了被感知理解的程度。Fogerty 等人进一步调查研究了元辅音过渡段在语音听觉感知中的影响^[9]。他们通过将过渡段增加到辅音段或者从元音段减掉,从而生成了 C+VP 和 V-VP 两种条件。C+VP 的激励信号中包含了辅音信息、部分元音过渡段、元音由噪声替换。V-VP 的激励信号包含了元音信息、在元辅音边界用噪声替换元音和剩余的元音过渡段。研究表明:对 C+VP 来讲,识别率随着加进来的元音过渡段的增加而增加;在 V-VP 条件下,在用噪声替换的辅音过渡段不超过 30%的情况下,只有元音的语音可懂度不受影响。

Lee 和 Kewley-Port 调查研究了在四种子段情况下(稳态中心、动态边缘、起始过渡段和终止过渡段),动态过渡信息对 CVC 音节和语句识别的作用^[13]。每一种条件下的子段信息包括相同比例的辅音和元音。实验结果表明:当元音和辅音的比例相同时,任何类型的子段信息都没有表现出相对优势;分段的过渡部分和中心部分对语句可懂度具有相同的贡献,然而尚未全完明白这两种类型的声学信息是如何相互影响的。

4 元音和辅音对语音可懂度影响的客观描述

上述的元音、辅音及其过渡信息对语音可懂度贡献的研究,主要是基于主观测听实验来进行的。由于主观测听实验的复杂性、耗时性和结果的相对不确定性,如何对元音和辅音在语音听觉感知中的作用进行客观描述(即语音可懂度的客观预测)已经成为一个研究热点。

4.1 传统语音可懂度客观预测方法

1929 年, Fletcher 与 Steinberg 对语音清晰度进行了比较系统和广泛的研究,并首次正式提出了“清晰度”的概念,并建立了清晰度指数的基础理论:语音信号各个频带对可懂度的贡献是相互独立的,并且一个语音信号的可懂度是各个频带信号的贡献之和。在此基础上,1950 年, Fletcher 和 Galt 提出了一种在特定环境下(如:语音音量、信噪比受到某种程度的干扰、通过某种频率响应的系统等),用以预测一个人在该环境下的语音可懂度的方法,称之为语音清晰度指数,简称 AI(Articulation Index)^[14]。在此基础上,美国国家标准局在 1969 年发布了计算 AI 的 ANSI S3.5-1969 标准^[15]。随着实验语料从无意义的音节变为单词、单句,便出现了语音可懂度指数(Speech Intelligibility Index, SII),美国国家标准局在 1997 年对清晰度指数(AI)的标准进行了修正,发布了计算 SII 的 ANSI S3.5-1997 标准^[16]。SII 可以用来预测在噪声干扰和滤波传输条件下的语音可懂度,还被应用到助听器和人工耳蜗的验配上^[17]。

另一个常用的语音可懂度预测方法是以调制传输函数(Modulation Transfer Function, MTF)为基础发展起来的语音传输函数(Speech Transmission Index, STI),它是由 Houtgast 和 Steeneken 于 1971 年首次提出来的^[18]。本质上,STI 反映的是声音传输通路上的物理特性,而语音可懂度依赖于很多实验因素,因此很难在 STI 和语音可懂度之间建立唯一的对应关系。但只要保持这些因素不变,就可以建立唯一的对应关系。通过大量实验,将其引入各种声音的传输通路,证明了 STI 方法的有效性^[19]。Elhilali 等人采用听觉模型,对噪声、混响等情况对语音信号带来的频域和时域调制指数的影响进行联合分析,提出了一种新的语音可懂度客观评价方法——频谱-时间调制指数(Spectro-Temporal

Modulation Index, STMI)^[20]。研究表明: STMI能够在一定程度上预测出噪声干扰、房间混响、非线性失真、相位抖动以及频率搬移等失真条件下的语音可懂度, 但它的研究仍然不够成熟。

4.2 语音可懂度客观预测的新进展

基于听觉感知系统主要对外界物理刺激的变化进行反映这个事实, Stilp 和 Kluender 认为使用香农信息论中的熵 (entropy) 来描述物理刺激的不确定性更加合理^[4]。在此基础上, 提出了一个心理声学度量法, 即基于耳蜗刻度的频谱熵 (cochlear-scaled spectral entropy, CSE), 来估计小的语音段的潜在信息量, 并指出它应该能够更好地预测语音可懂度。CSE 表征经过耳蜗听觉滤波器处理之后的语音谱片段与其之前谱片段的相似程度, 用欧氏距离来刻画。CSE 度量提供了一个对语音片段中的潜在信息的逐步估计, 也可以在 CSE 级别上进行替换和删除操作。因此不需要关于语音分类的假设, 也不需要手工选择元音和辅音的边界。实验结果表明: 相对于传统的元/辅音、辅/元音过渡来讲, 基于 CSE 的大小来预测语句的可懂度取得了更好的预测效果。在平均意义上, 元音比辅音具有更多的信息量, 因此在自然口语中元音具有更高的可变性。这个结果主要取决于用来估计熵的耳蜗刻度, 因为它是一个准对数, 赋予共振峰附近的频率更大的权重, 也就是说, 这些频率是大多数元音结构的“重心”, 随着时间变化较快, 因此具有更多潜在信息量。尽管辅音也包含了快速变化, 但是它们往往都在高频处, 因此被赋予较低的权重^[6]。

最近, Chen 和 Loizou 假设在可懂度预测过程中, 仅仅包括有用信息的片段能够提高与主观可懂度测试结果的相关性, 进一步调查研究了句子中携带更多信息的语音片段, 并将其应用到语音可懂度预测中^[21]。他们的实验结果表明: 以元音为主的高能量片段比以辅音为主的低能量片段具有更好的语音可懂度预测能力; 中级 RMS (root mean square) 语音片段比低级 RMS 语音片段和高级 RMS 语音片段具有更好的语音可懂度预测能力, 这是由于中级 RMS 语音片段包含了大量的元音和辅音的边界过渡段信息。这个结果也和文献^[4]的结果是一致的, 即具有较高熵 (对应于包含大量频谱变化) 的语音片段能够更好地预测语音可懂度。总的来讲, 可懂度预测模型应该更加关注语音谱变化, 这个可以用

中级 RMS 和基于熵刻度的分割方法来表征。相对于熵刻度方法, 中级 RMS 方法更多强调了具有元辅音过渡段的谱变化, 因此被认为在噪声中更加鲁棒。

5 元辅音听觉感知应用的新进展

语音中元辅音听觉感知研究的结果也已经逐步在实际系统中得到了应用。本节将以语音分离和语音保密为例, 对其最新应用进展进行介绍。

语音分离的研究在语音通信、声学目标检测和声音信号增强等方面有着重要的理论意义和实用价值。计算听觉场景分析研究的进步使得语音分离 (尤其是浊音分离) 取得了很大进步。然而由于清音缺少谐波结构、能量微弱、更容易受到干扰影响等特点, 清音分离显得更加困难。听觉感知研究结果为清音分离提供了指导。正如文献^[5]所述, 浊音常常与清音具有强烈的协同发音的特点, 对语音可懂度来讲携带了更多的信息。以此研究结果为基础, Hu 和 Wang 等人提出了一个从非语音干扰中分离清音的方法^[22]。这个分离过程包括: 分离和重组。在分离阶段, 通过对声音事件的起始和终止进行分析, 将输入混合信号分解为连续的时频段。清音的重组是基于声学-音素特征对分离得到的时频段进行贝叶斯分类来实现的。结合已有的浊音分离模型, 将形成一套能够处理清音和浊音的语音分离系统。

语音隐私保护一直是音频通信中的一个重要问题, 在远程协同应用和安全监控等系统中具有广泛的应用前景。音频信号不仅提供了语音内容本身信息, 而且具有其它的环境声音等信息。如果通过改变语音信号使其不可懂并保留其它有用的环境声信息, 那么隐私保护问题就可以得到一定程度的解决。以语音听觉感知研究成果^[5,7-8]为基础 (如利用噪声替换元音大大降低了语音可懂度), Chen 等人提出了一种通过自动替换元音的音频隐私保护方法^[23], 它能够混淆语音信号, 而保留语调以及处理后环境声音的感知。具体来讲, 首先利用 LPC 从语音信号中把激励信息和声道信息分离开来; 然后用事先保存的具有不同声道形状的不相关元音的 LPC 系数来替换输入语音中元音部分的 LPC 系数, 从而保留了音频信号的语音性和大部分环境声音, 大大降低了语句的可懂度, 提高了语音隐私的保密性。

6 总结与展望

总之,作为语言声学的重要研究内容之一,语音听觉感知研究是一个充满活力的多学科交叉的研究方向。对影响语音感知和理解的基本单元、他们的相互作用以及在人脑中是如何反应的研究是理解人对语音信号进行感知和认知的一个关键问题。这个研究领域的进步能够对听觉障碍患者有一定的帮助,能够为语音识别和语音通信等应用性能的提高提供理论指导,为建立自然和谐的人机交互奠定基础。

参 考 文 献

- [1] DENES P, PINSON E. The Speech Chain[M]. 2nd ed. Worth Publisher, New York, 1993.
- [2] FLETCHER H. Speech and Hearing[M]. Van Nostrand, New York, 1929.
- [3] OWREN M J, CARDILLO G C. The relative roles of vowels and consonants in discriminating talker identity versus word meaning[J]. J. Acoust. Soc. Am., 2006, 119(3): 1727-1739.
- [4] STILP C E, KLUENDER K R. Cochlea-scaled entropy, not consonants, vowels, or time, best predicts speech intelligibility[J]. Proc. Natl. Acad. Sci. U.S.A., 2010, 12387-12392.
- [5] COLE R A, YAN Y, MAK B, et al. The contribution of consonants versus vowels to word recognition in fluent speech[J]. Proc. ICASSP, 1996, 853-856.
- [6] LEWICKI M S. A signal take on speech[J]. Nature, 2010, 821-822.
- [7] KEWLEY-PORT D, ZACHARY BURKLE T, LEE Jae Hee. Contribution of consonant versus vowel information to sentence intelligibility for young normal-hearing and elderly hearing-impaired listeners, [J] J. Acoust. Soc. Am., 2007, 122(4): 2365-2375.
- [8] FOGERTY D, KEWLEY-PORT D. Perceptual contributions of the consonant-vowel boundary to sentence intelligibility[J]. J. Acoust. Soc. Am., 2009, 126(2): 847-857.
- [9] FOGERTY D, HUMES L E. The role of vowel and consonant fundamental frequency, envelope, and temporal fine structure cues to the intelligibility of words and sentences[J]. J. Acoust. Soc. Am., 2012, 131(2): 1490-1501.
- [10] DIGIOVANNI J, PREWITT J M, NAGARAJ N K et al. The relative contribution to speech intelligibility from consonants and vowels using synthesized and naturally-spoken sentences[J]. Proc. the 160th Meeting of Acoustical Society of America, 2010.
- [11] KENT R D, MINIFE F. Coarticulation in recent speech production models[J]. J. Phonetics, 1977, 5: 115-133.
- [12] COOPER F, DELATRE P, LIBERMAN A, et al. Some experiments on the perception of synthetic speech sounds[J]. J. Acoust. Soc. Am., 1952, 24: 597-606.
- [13] LEE J H, KEWLEY-PORT D. Intelligibility of interrupted sentences at subsegmental levels in young normal-hearing and elderly hearing-impaired listeners[J]. J. Acoust. Soc. Am., 2009, 125(2): 1153-1163.
- [14] FLETCHER H, ROGERS H G. The perception of speech and its relation to telephony[J]. J. Acoust. Soc. Am., 1950(22): 89-151.
- [15] KRYTER K D. Methods for the calculation and use of the articulation index[J] J. Acoust. Soc. Am., 1962, 34(11): 1689-1697.
- [16] ANSI S3.5-1997, Methods for calculation of the speech intelligibility index[J]. American National Standards Institute, New York, 1997.
- [17] HALPIN C, THORNTON A, HOU Z. The articulation index in clinical diagnosis and hearing aid fitting[J]. Current Opinion in Otolaryngology Head and Neck Surgery, 1996(4): 325-334.
- [18] HOUTGAST T, STEENEKEN H J M. Evaluation of Speech Transmission Channels by Using Artificial Signals[J]. J. Acustica, 1971(25): 355-367.
- [19] STEENEKEN H J M. HOUTGAST T. A physical method for measuring speech-transmission quality[J]. J. Acoust. Soc. Am., 1980(67): 318-326.
- [20] ELHILALI M, CHI T, SHAMMA S A. A spectro-temporal modulation index (STMI) for assessment of speech intelligibility [J] Speech Communication. 2003, (41): 331-348.
- [21] CHEN F, LOIZOU P C. Contribution of cochlea-scaled entropy and consonant-vowel boundaries to prediction of speech intelligibility in noise[J]. J. Acoust. Soc. Am., 2012, 131(5): 4104-4113.
- [22] HU G, WANG D L. Segregation of unvoiced speech from nonspeech interference[J]. J. Acoust. Soc. Am., 2008, 124(2): 1306-1319.
- [23] CHEN F, ADCOCK J, KRISHNAGIRI S. Audio privacy: reducing speech intelligibility while preserving environmental sounds[J]. Proc. 16th ACM International Conf. on Multimedia, 2008: 733-736.

勘误: 2013年第2期封三右上角误加页码159, 特此更正。