

我国已于1992年开始执行“汉语综合资料库和语言信息处理系统评价方法”研究,为期五年。现在资料库的录音(48kHz采样,16bit量化DAT录音)工作已基本完成。汉语孤立词识别系统评价方法已有草案。准备采取在研究工作的同时,开展试用推广工作,以便使基础与应用更好的结合。

863项目支持的汉语语音识别数据库,由中国科学技术大学完成,并于1993年12月通过鉴定。

1993年12月还在合肥召开了言语资料库和言语输入/输出评价方法研讨会。会上成立了汉语言语资料库和言语输入/输出系统评价方法协调委员会,简称Chinese COCOSDA。该委员会由张家骏任召集人;分成三个工作组:

言语资料库,言语识别和言语合成,分别由王仁华、黄泰翼和张家骏任召集人。863项目“智能计算机”专家组成员参加了研讨会,并给予了大力支持。无疑这将对我国语言信息处理事业的发展与与国际相关组织的合作有巨大推动作用。

参 考 文 献

- [1] Castagneri G. (Editor), Proc. Workshop on Internat. Cooperation and Standardization of Speech Databases and Speech I/O Assessment Methods, Chiavari 26—28 Sept. 1991 (Italy).
- [2] Jones K. and Mariani J. (Editors), Proc. the 1992 Workshop of the Internat. Coordinating Committee on Speech Databases and Speech I/O Systems Assessment, 12 Oct. 1992 Banff, Canada.

汉语综合资料库的设计*

张家骏 齐士铃 吕士楠

(中国科学院声学研究所 北京 100080)

1993年5月18日收到

语言是人类最重要的交际工具,随着现代信息技术的发展,语言也是人与机器之间交际的有效工具。近年来世界各国纷纷建立本国的言语资料库作为言语科学研究和言语技术开发的基础。汉语综合资料库的语音材料有:汉语全部有调音节、数字串、单词、韵律特征材料,以及语言清晰度试验用音节表、词表、句表和有代表性的短文等。汉语综合资料库在语言学和语音学特征以及声学特征方面充分体现汉语的基本特点。首先要解决语料选取问题,考虑各种语言单位的使用频率,不仅要包括全部高频词,也要反映较全面的语音现象。数据库在结构上是开放的模块式的,同时配有灵活的数据库管理系统。

ABSTRACT

Speech is the most important means of communication among people, and speech I/O an effective manner of communication between human and machine. Speech databases lay a foundation of speech science and speech technology. A comprehensive speech database of standard Chinese which might reflect the linguistic and phonetic characteristics of standard Chinese is composed of all tonal syllables used in real speech, all commonly used words, digit string, some sentences, and speech articulation test lists including phonetically balanced syllable lists, word lists and sentence lists. The speech database is an open set with module structures.

表1 国外几个有代表性言语资料库的用途、材料内容和发音人条件

国 家	数据库名称	用 途	材 料	发 音 人
美 国		DARPA 研究集体内: 1.长期言语研究 2. 言语识别和理解技术 开发3. 言语识别和理解技术评价 语音识别的声学语言学知识训练识别器 识别器评价的共同测试基础 海军资源管理 1000 词语言模型连续语言识别(发音人有关, 发音人适应, 发音人无关) 扩充训练材料 航空旅游服务(面向自然言语) 1.电话言语的声学分析 2. 开发电话言语识别算法 3. 宽带识别算法对电话言语训练	每人 2 句“方言校准”每人 5 句“语言密集” 每人 3 句“自然语言”共 10 句 21000 语句 2652 句/人 初样 1000 语句	630 人代表美国英语的主要方言 (八个区)70%男,30%女 160 人(仍认为不够) 4 人
	TIMIT			
	RM1			
	RM2			
	ATIS NTIMIT			
欧 共 体 八 国 28 实 验 室 (SAM)英,德,法, 丹,意,荷,挪,瑞典	EUROMO EUROM1 SUNSTAR	属 ESPRIT Project 2589 1.言语识别评价 2.言语合成评价 3.技术与研究 有代表性的噪声数据库	EUROM-0 五种语言五小时 EUROM-1 数字 0-9999 五句有段落韵律 扩充 CVC 表,每条重读多次(CVC 嵌入 两词之中)合成评价: 无意义词表,语义不 可预测句(SUS),1990 已向 CCITT 建议 4 套: 1.国内电话噪声 2.国际电话噪声 3.环境噪声(办公室)4.环境噪声(室外)	20 人(5 语种, 每种 4 人)多发音人 (8 语种每种60人)丹麦语荷兰语已 完成少发音人(8 语种每种 10 人)
	British Telecom Research Lab.	开发电话言语识别算法		1100 发音人,每人发 12 分钟
	GRECO数据库 BDSNS (语音) BDLEX (词 汇) 1990 .3 公布了试验标准 French Standard- S31-115-1, “自动言语处理系 统评价”;第一部分, 自动言语识 别系统的定义与评价方法正在录 制的数据库; GRECO, BREF 以及 ESPRIT SAM database			10 人(5 男,5 女)
瑞 典	CTH	言语研究 言语编码	语音库: 23 元音变体(9 个基本元音),100 常用双音 (CV 或 VC)单句库; 230 语音 平衡句70选自文本库,文本库: 5种权威报 纸共 4935 词 352 句 102 段	1.4人标准发音(新闻播音员 2 人 一男一女有经验的发音人 2 人) 2.一般人 共 2.5 小时采样

荷兰	Noise data-base Post-Production Factors (ENST 也有一类系统叫 PSOLA)	测试言语产生、言语知觉模型 主要用于声学语音学研究 Speech I/O 评价 言语通讯信道数字模拟系统	1.元音、辅音、音节、单词(孤立读) 2.孤立单句3.一页小说4.一段对话 24种噪声,每种至少4分钟 80—10kHz 模拟频谱形状、附加噪声非线性失真、混响、自动增益控制、数字调制方法共68种信道	目标: 10男 10女 10童 现在: 4男 1女
西班牙	ALBAYZIN政府资助 ALBAYZIN 5 个研究组 Phonetic database	言语识别、自然言语处理	4500句 1.语音平衡,任务无关 2.任务有关(西班牙地理数据库查询) 3.有噪声言语 200句分成20组每组语音平衡	100人包括各地方言男女各半(六个区) 100人男女各半
意大利	Telephone grade quality speech corpora	评价电话言语识别	利用现有数据库和 Telephone Line Emulator	模拟通路失真、近端远端回声、噪声、频移、卫星延迟相位颤动
澳大利亚	NSLD 标准(National Spoken Language Database)			
日本	JEIDA Noise Database DSR	通用言语资料库 模拟标准环境 (Pseudo-Environment)	323条×4次×150人193800取样其中: 110单音节178孤立词35四数字串17种环境噪声 1.元音和数字2.216语音平衡词(ATR表) 3.110单音节(JEIDA表) 4.连续言语 a.东北大学报(70句5分钟包括可能重音) b.问句(11个)(45s) c.言语质量测试句,由日常对话出现率最高的20音节组成 d.故事:伊索寓言“北风”(50s) e.天气预报(50s) f.陈述句:用于语调研究 5229孤立词,选自词典216语音平衡词 数字与100音节115短句,用于自动翻译电话连续言语:孤立句503选自报刊 语音平衡1542语音平衡词212词3000词连续言语	75男75女(20—60岁)每条每人读四次 10男10女每条读四次每人两小时
	ATR ETL Tohoku University ATR 对话库 Dialog database Continuous Speech Database Committee,	多用途	270000词	10男10女专业发音人将来100多人

表 2 国外几个有代表性的言语资料库的录音条件存储介质和实用情况

录音条件	存储介质	实用情况
美国 双通道: 1.Seunheiser HMD414头戴式抗噪声传声器 2. B&K 4165 1/2" 压力传声器, 距唇 30cm, 20 度偏角 20kHz 采样, 16bit 量化降至 16kHz 采样 Double-High 字符屏幕显示给发音人, 录音后发音人和主持人共同听录音, 重录检出的错误发音 TIMIT 峰值 S/N:29dB ATIS 1.一般办公室具有声级44dB A 2. 指向性传声器距发音人 2 英尺	CD-ROM 680Mbyte 符合 ISO-9660 Data-Format	87 年开始应用开始发行 TIMIT Disc1-1.1 美国 \$100 国外 \$300 RM1, 2Discs 美国 \$400 国外 \$950 连续言语, 开发, 测试和评价数据, 评分和 Speech Header 软件, 1 Disc 美国 \$750 国外 \$950 RM2, 2Disc (连续数字), 美国 \$750 国外 \$950 TIDIGITS 3Disc (连续数字) 美国 \$750 国外 \$950
欧共体 双通道: 传声器和声门阻抗仪为了 Post-Production Effects 可加到录音材料上 16kHz (EUROMO) 20kHz 采样新库或 AU22 3. 1Mbyte 扩充存储 CD-ROM Reader 工作语言 C SUNSTAR 电话噪声 8kHz 采样环境噪声 20kHz 采样	EUROMO CD ROM—张已发行	最低硬件要求: 1.IBM PC AT 或兼容 2.模拟接口OROS-AU21
英国 消声室, 1/2" 电容传声器, 44. 1kHz 或 20kHz 采样		
法国 为数据库录音在 CNET-LANNION 专门建造了录音室		
瑞典 CTH: B&K 4165 传声器 B&K 2804 电源 SONY PCM-FI Audio Processor SONYSL-FIE Vidio Tape Recorder 16 bit 量化 KTH: SONY F-1, B&K 电容传声器, 16 kHz 采样 16 bit 量化只标音位不标变体		
日本 JEIDA Noise Database 用三种传声器 1. 立体声式 2. 驻极体式 3. 动圈式普通传声器, 隔声间	DAT 76 盒 共 120 小时 DAT 18 盒 共 120 分钟	分发到 50 个日本言语研究组织 1986—1987

一、建立汉语综合资料库的意义

建立言语综合资料库和研究言语信息系统评价方法,是近年来人机言语通讯和计算机技术发展的客观需要。语言是人类最重要的交际工具。因而言语信息处理技术已成为当前信息技术发展的前沿。言语所能传递的信息是无限的,使用言语进行交际的人数也是非常大的,言语交际的方式又是多种多样的。因此,如何选择具有代表性的言语样本,作为言语处理系统(如:言语合成、言语识别、言语理解)的设计和评价的依据,便成为一个很重要的研究课题。必须设计完善的言语资料库。早期采用元音和辅音作为样本加以研究,但很快发现协同发音的问题:以后又有了孤立单词的样本,但局限性也很大。一方面单词数量很大,另一方面在自然言语中单词之间也还有相互影响。所以必需根据声学、语言学、心理学和逻辑学的规律,在无限的言语表达中选出有限的、有代表性的模式作为样本。这种方法和材料不但应在本国尽量统一,而且在国际上要能与其他语种可比。当前国际上新兴的研究课题——电话自动翻译系统,更要求言语资料库要考虑到语言翻译的职能。语言是具有极强的民族特点和地区特点的。因此许多国家都分别建立本国官方语言的资料库。

如何客观地评定言语通讯系统的性能和质量,并指出其改进方向的科学方法,也是十分需要的。对一个智能系统的评价犹如对学生进行考试;什么样的考题,怎样的考法,直接关系到得分。为此必需选取有代表性的试验材料,采用科学设计的方法来进行评价。特别是现在言语信号处理方式多种多样,数字化系统越来越多,信号失真形式也与模拟系统很不相同,如何检验这些失真对言语质量的影响便成了新课题。

语言是有很强的民族特点的,其他国家的经验只能借鉴而不能照搬。汉语是世界上使用人数最多的语种,除大陆、台湾、香港、澳门以

外,还有东南亚地区与众多华侨。研究建立汉语综合资料库是一件工作量大,涉及面广的基础性工作。现在应开始分阶段有步骤地逐步建立和不断完善汉语大型综合资料库。

二、国际发展状况

自八十年代以来,国外纷纷建立本国语言的资料库,不少国家已经初步建成并公开发布有关资料供言语工程等方面研究应用。从总的发展趋势看有以下几个特点:1. 规模大——表现在言语材料内容数量大,发音人多,2. 用途广——既考虑应用于言语处理系统,如训练和评价识别系统等,也可用于如言语产生、言语知觉模型等声学语音学的基本研究,3. 层次高——从元音、辅音、孤立音节开始,逐渐扩大到词、句、乃至文章段落,向自然言语发展,4. 标准化——用统一的选材原则和方法建立资料库,使得各种言语的数据库之间数据可以交流,不同语言之间可以比较。

美国 DARPA^[1]连续言语识别用 1000 单词资源管理数据库是从 630 名成年发音人中选出 130 人的录音,对不同地区有广泛的代表性(1988)^[1]。近年来又将已建的数据库经不同类型电话传输,发展为连续言语、多说话人、电话带宽的言语数据库。现公开以激光盘片发行的就有 15 种之多。英国也由电信局建立大型电话数据库(1988)^[2]。日本国际电气通信基础技术研究所(ATR)自从 1986 年成立以来便从事大型日语资料库的研究,内容有 8500 多条目^[3]。法国在国家研究中心的言语通讯协调小组支持下已建立法语语音资料库 BDSNS 和音系与词汇资料库 BDLEX(已有两个版本)可供使用^[4]。资料库中对大量说话人搜集的口呼数字(孤立的和连续的),便是很有应用价值的资料。它可直接用于口呼拨号电话、计算机数据输入、电话办理银行业务等。荷兰等其它国家也都在建立自己的数据库。详细情况参见表

* 本课题为国家自然科学基金重点项目。

三、汉语综合资料库的选材与内容

从美国、法国和日本已初步建成的言语资料库来看,在言语材料的选取上有很大差别,这一方面取决于资料库所针对的应用目标,也和言语本身的语音学特征有关。汉语语音资料库拟包括汉语普通话全部音节和95%以上的常用词和适量的单句和短文。对语音材料综合设计的原则是:以尽量少的材料能反映出汉语普通话的各种语音现象。

目标是建立一个既适应于言语工程应用又满足言语科学研究用的综合资料库。这是一个要求较高的长远目标,很难一次完成。言语资料库的内容包括三个不同层次——语言学的、语音学的和声学的。根据应用目标又可分成三种类型:(1).言语识别和言语合成系统训练、测试评价用的;(2).言语识别和言语合成系统诊断改进用的,(3).声学语音学研究用的。现在主要侧重资料库的语音学的和声学的层面,但是要考虑到向语言学层面的发展。目标也主要侧重在言语分析、合成和识别的应用方面,结合声学语音学研究需要。在资料库建立的过程中同时也要开展必要的研究工作,如自动分段标音问题。在初步建成后将扩展,加重语言学层面的内容并可供机器翻译应用。

所建资料库研究的对象是汉语普通话,但要考虑到今后向不同方言和民族语言扩展提供基础。主要包括:

①. 汉语语音数据库:(1).十个单元音,22个辅音(取自双声词),(2).1200个有调音节,(3).标准试验句(“他去无锡市,我到黑龙江”)和短文“北风与太阳”,(4).十个数字,孤立的(包括不同读法,如1(么)7(拐),和两个、三个、四个、六个、七个组合,其中六联组合是考虑到我国邮政编码,七联组合考虑到电话号码),(5).清晰度试验音节表、词表、句表(语音平衡),(6).基本句型表,共300句。

语音资料除语音信号波形和声门波外,应

包括必要的语言特征标记,如音位界、音节界、重音等供语音规则控制用。

② 词汇资料库: 它是对语音数据库的补充,为人机对话双向文语转换工作提供资料,为小规模机器翻译提供词典。它包括,大词汇资料库,约45000个词,覆盖率为95%。考虑到轻声和儿化都是汉语构词手段,还包括有轻声词库和儿化词库。

③ 资料库管理系统: 资料库是综合声学数据、语音特征、语音规则和语法语义于一体的庞大系统。现在还没有成熟的管理系统可用,需独立设计合于汉语资料库的专门管理系统,便于检索与调用。

四、方法

发音人: 由于中国地域广大,社会成分复杂,为使资料库有足够代表性,发音人需包括各个方面,如不同的性别、年龄、文化修养、母语背景以及地域上差别等。语音资料库发音人为40人,按代表性要加以选择。

技术条件: 为便于国际交换,需数字化录音,48kHz采样,16bit量化,使用测量用电容传声器(B&K 1/2"),声门阻抗仪(Laryngograph)在消声室中同时录音于左右两声道,同步记录语音波形和声门波。发音时进行声级监视,以保持发音平稳。存贮媒介用DAT盒式磁带,由于数据量庞大,有条件时向光盘存储过渡。

• 文件内容及注释: 在汉字文本中包括同音字、同音词(常用范围)。

音标注释: 语音库做音位标记,以汉语拼音稍加补充为标音符号,采用声、韵、调音位系统,标定音位界。

声学参数: 预留空间,准备在下一阶段完成之。

五、结束语

为使汉语综合资料库的进一步完善,在此

基本框架基础上,除需加重语言学层面的内容和声学参数外,还考虑加入:1.环境数据库包括:各种噪声环境、声场环境等,2.电话言语资料库,3.典型(特定)工作条件语音数据库(如:不同紧张程度下的言语通讯),4.汉语语料库部分内容的言语资料库,5.自然口语资料库。

汉语综合资料库是一项系统工程,有很大的难度和极大的工作量,要从一点一滴做起,逐步建立、逐步应用、逐步完善。

参 考 文 献

- [1] Patti Price, Fisher, W. M., Bernstein J. and Pallet, D.S. *Proc. ICASSP-88*, 651—654.
- [2] Millar, P.C. Cameron, I., R. Greaves, A. J. and Mopeake, C.M. *Proc. ICASSP-88*, 647—650.
- [3] Kuwabaral, H. Takeda, K. Sagisaka Y. et al., *Proc. ICASSP-89*, 560—563.
- [4] Carre, R. Descout, R. Eskenazi M. et al., *Proc. ICASSP-84*, 42, 10, 1—4.

脉冲 Doppler 系统中混叠现象的克服和测速最大值的提高

汪 源 源 王 威 琪

(复旦大学电子工程系 上海 200433)

1993 年 2 月 22 日收到

脉冲 Doppler 系统混叠现象的存在使其最高速度的测量范围受到限制。本文先介绍一种克服混叠现象的方法:平均频率跟踪法,然后报导了我们研究的两种克服混叠现象的方法:最大频率跟踪法和混合跟踪法。通过计算机的模拟实验,指出:利用混合跟踪法不仅可较好地克服混叠现象,而且便于仪器实现。这对于脉冲 Doppler 仪器性能的提高,具有重要意义。

一、引 言

超声 Doppler 技术在医学临床诊断中占有十分重要的地位, Doppler 技术的一个重要方面:脉冲(PW) Doppler 出现于 70 年代初,经过几十年的应用,目前已成为临床的一种重要诊断手段。PW Doppler 可以测量人体中不同深度血管中的血流速度,它需要重复发射声脉冲,以及相对发射脉冲延迟一段时间设置一个选通门(称为距离选通)。重复发射脉冲和距离门选通构成了时间上、空间上的采样,其结果使原先的 Doppler 信号频谱展宽。上述过程可以用图 1 说明^[1]:图 1(a)为用连续波 Doppler 得到的信号频谱,其中正向最大血流速度 V'_{max} 对应的 Doppler 频移为 f_H ,反向最大血流速度 V'_{max} 对应的 Doppler 频移为 f_H 。图

1(b)为 PW Doppler 的信号频谱,PRF 表示脉冲的重复频率。根据采样定理, f_H 和 f_H 必须小于 $\frac{1}{2}$ PRF,否则如图 1(c)所示,正向血流 f' 大于 $\frac{1}{2}$ PRF,其 oa 部分进入前一个频谱的反向区,发生频谱混叠(Aliasing)现象。正向 oa 部分被误视为反向成分,在声谱图上表现为声谱顶部削平,被削去的部分出现在反向声谱中,如图 1(d)所示。因此 $\frac{1}{2}$ PRF 被称为 PW Doppler 频谱分析的奈奎斯特极限。这种极限的出现,使得 PW Doppler 最大测速范围受到限制。

当然提高脉冲重复频率 PRF 可以扩大测高速的范围,即减轻频谱的混叠现象。但 PRF 的提高会使最大检测距离 d_{max} 缩短,因为