面向言语工程应用的语音学研究

——中国社会科学院语言所语音研究室近期研究工作和成果简介

李爱军
中国社会科学院语言所
liaj@cass.org.cn

提要：近几年，语音研究手段有了很大提高，言语工程应用领域不断扩大，使得语音学研究在理论和应用层面都有了新的发展。这里介绍中国社会科学院语言所“语音与自然话语重点学科”和语音研究室在近期开展的语音学研究工作和所取得的一些成果，以面向言语工程应用方面为主，也有一些是涉及语音学基础研究的。

关键词：语音学 中国社会科学院语言所语音研究室

1. 引言

中国社会科学院语言所“语音与自然话语重点学科”成立于 2002 年，主要依托于语音研究室和当代语言学研究室，该学科以自然话语处理中的语音问题为主要研究对象，还包括与语音问题相关的句法、语义和语用等方面的问题。

近年来，围绕自然话语处理和言语工程应用中的语音问题，我们承担了国家863项目、国家社科基金项目、国家社科基金项目以及一些横向合作项目，有针对性地设计并建立起了一些大规模的语音语料库，如863项目支持的10大方言点普通话大型地方普通话语音语料库，中国社会科学院重大项目支持的普通话口语对话库，社科基金项目支持的语篇语音语料库。并在此基础上开展了以下几个方面的研究工作：（1）自然话语的语调韵律特征研究；（2）情感语音研究；（3）方言口音问题研究；（4）音段与音段音变的研究；（5）关于音段的语音实现与超音段的韵律结构的关系的探索；（6）语音的生理研究。下面先对这些已经或正在开展的研究工作进行简要介绍，然后谈谈我们今后的研究思路。

2. 研究工作介绍

2.1. 语音语料库的收集和标注

2.1.1. 自然情感语音库CASSESC

CASSESC(CASS Expressive Speech Corpus)的基础库部分包括4名专业演员的6种基本情绪录音。同时，收录了电视和电台的情感访谈节目，对这些自然口语进行了转写和部分语音标注。

2.1.2. 厦门普通话的对话和独白、以及厦门方言库

收集了 100 人的厦门普通话对白，20 个对话，厦门方言的单音节、两音节等基础数据。是一个针对方言语音识别和方言普通话研究的较大规模语音库。

2.1.3. 普通话基础语音语料库

内容主要分为单音节（覆盖全部有调音节）、词语（覆盖轻声、儿化，以及连读变调等现象）、句子（包含不同句类和句式以及各类复句关系，并争取做到三音子的最大覆盖）、篇章（包含散文、故事、新闻、情书等不同体裁）以及话题表述（包括看图说话、内容复述、即兴解说等不同方式）等五大部分。录制完这套基础语音语料库之后，我们将对之进行细致地语音标注。

2.1.4. 现场即席话语多模态语料库

现场即席话语多模态语料库指按一定的取样标准采录的事先无文字准备的话语活动数字化视频库。目前英国爱丁堡大学构建的多模态语料库和中国社会科学院语言所构建的多模态语料库可以说是世界上此类研究的首批成果。我们侧重于研究话语活动中的“说－做”这两种行为的动态集成模式和结构，同时对动态集成模式和结构在视频流上进行切分和标注。这项研究在国内外尚属首次。

2.1.5. 汉语语音库收集与标注通用规范

在中文语音交互技术标准工作组领导下，我们承担了汉语语音库收集与标注通用规范的标准制定工作，旨在为言语工程与语音研究，提供语音库制作和标注规范，便于语音库的资源共享与分发。其内容包括语音库的分类标准、术语定义、语音库技术规范、音字转写规范、韵律标注规范和音段标注规范等。

2. 2. 自然话语语调以及相关韵律特征

语音科学目前正处于从处理孤立话语向处理连续话语过渡的重要时期，许多新的问题会不断涌现出来，当前最重要的就是自然口语的韵律问题。过去的韵律研究大多是基于朗读语料和小语料库基础上的研究，无法真实全面地反映人们实际语言的韵律特点。有研究表明，自然话语与朗读语料之间存在着很大的差异。要想真正弄清自然话语的韵律特征和规

律，就应该以自然话语材料作为研究对象。

2.2.1. 汉语句调及疑问句研究

这一研究利用 973 电话对话库，研究疑问语气和陈述语气的表达问题，通过知觉实验的辨认测验方法证实，在脱离语境的条件下，人们对疑问强度的感知有可能被弱化；并通过合成实验证明，疑问和陈述的区分只能来自于边界调。

2.2.2. 关于汉语口语语调结构的探索

以声学-语音学的实验分析为基础，探索汉语声调和语调生成的过程，客观地揭示汉语语调的深层本质及内部结构。集中讨论（1）什么是话语语调的基本模式？它是怎样形成的？（2）具有区别词义功能的声调是怎样跟话语的总体语调有机结合的？考察结果表明，话语语调的基本模式主要跟不同层次话语单元的音阶运动、即自然音高水平的起落有关，它的起伏波动取决于言语产生的生理制约和语言学（如语气、情感及语义等）表达的需要。而声调模式既跟各个局部音节或词的音高变化的曲拱调形相关，又跟它们音高变化的音阶特征有关。在实际话语中，各个局部的音节或词，一方面基本保持它们的音高曲拱模式，以满足词义区别的需要，另一方面通过它们的音阶的相对起落变化，满足话语总体语调调阶起伏的需要。实际话语里局部的声调跟总体语调是通过它们的音阶的相互叠加而有机结合起来的。

2.2.3. 普通话声调与边界调音高特征及其实现规则

这一研究基于 10,240 个句子的音高数据分析了普通话声调与边界调的音高特征及其音高实现规则，确定普通话各类声调、轻声与边界调的音高特征点，然后以此为基础，概括出非轻声与非轻声、非轻声与轻声、轻声与轻声以及非轻声与边界调、轻声与边界调的音高特征及其结合实现方式,得到的音高特征点、六条规则以及边界调的功能，在描写连接音高模式时，成为有效的根据。

2.2.4. 清辅音对后接元音F0的扰动及该扰动对感知的影响

对清辅音后接元音 F0 的扰动及该扰动对感知的影响进行了分析，发现不同的清辅音对相同元音产生的扰动不同，相同的清辅音对不同的后接元音的扰动也有区别，具有一定的规律性。此外，还考察了相同发音部位的清辅音和浊辅音对后接元音的扰动情况，并在句子层面进行了感知试验，结果发现，扰动对语句的自然度产生影响很小，通过听辨实验验证了 F0 的扰动对感知没有影响。

2.2.5. 两字组音高下倾现象研究

这一研究采用了国外语调分析理论（AM）中处理音高曲拱的方法，把普通话语流中的音高曲拱分解为音高目标段和音高过渡段，并采用国外音系学理论中的一条“音高降阶”原则简明而精确地解释了普通话两字组连调模式中存在的 4 类有规则的音高异常现象：（1）当两个去声相连时，后一去声的音高起点相对较低；（2）当阴平调处于去声之后时，其音高通常会整体下降；（3）处于上声之后的阳平调，通常具有相对较高的调尾；（4）处于去声之后的四声音域会相对较窄。这一研究表明，在复杂的语音现象背后起作用的往往是一些简明的音系规则，这为我们认识和解释语音现象提供了一条比较好的思路。

2.2.6. ‘吗’字是非问句的语调韵律特征研究

这一研究采取声学语音学的实验手段细致地考察了普通话中“吗”字是非问句的句末音高特征。其研究结果表明，在朗读条件下，不管是否带有句末语气词“吗”，是非问句都具有同一类型的句末边界调，音高表现为{H}或{Raised}；其研究结果还表明，句中焦点位置的改变不会对句末边界调的音高表现产生本质的影响，句中声调的改变也不会对句末边界调的音高表现产生本质的影响。基于此，该研究认为，普通话里的句末边界调是一个相对独立于焦点重音和声调的语调音高特征，有着相对稳定的音高表现。

2.2.7. 结构轻音音节的声学分析和考察

轻音音节在语音上的表现具有时长缩短、能量减弱、清声母浊化、元音央化、音高随前音节变化等声学表现。本研究选取振幅和时长考察，目的是考察轻音的稳定性问题，并且不同人的比较也可以看其变化的典型性。个体分析与比较研究结果表明，振幅曲线形状比较一致，时长比较一致。不同人之间差异不明显，不同音节间差异明显。从研究结果中可以看出，每个轻音音节具有自身的一致性，这从一人多样本与多人的数据可以看出。这种一致性表现在振幅曲线的形状和时长平均值上。结果发现在连续话语中，结构轻音音节表现出各音自身的特点，这表现在每个音节振幅曲线有一致的模式，时长比较一致。不同人之间差异不明显，不同音节间的差异更大。

2.2.8. 语调研究中心理和声学等价单位

由于人耳听觉系统非常复杂，迄今为止人类对它的生理结构和听觉特性还不能从生理解剖角度完全

解释清楚。所以，对人耳听觉特性的研究目前仅限于在心理声学和语言声学。人们对语调的特性研究目前也只能从心理感知和声学分析方面来揭示。语调研究的物理实体一般是基频曲线 F0，不同发音人的基频变化范围有很大的差异。一般语调音高曲线 F0 以赫兹（Hz）为单位，研究中也采用与心理-声学（psycho-acoustic）相关的等价单位，如半音 semitones, 镁 mel, 巴克 Bark 以及 ERB 等。到底在语调研究中使用哪个单位更好？本研究总结了国内外关于这个问题的研究，说明半音是最好的反应心理-声学的语调研究单位。还用半音分析了汉语情感语音库中几种情感语音音高曲线的特征。

2.2.9. 普通话疑问词韵律的语音学分析

本研究从语音学的角度考察普通话疑问词在特殊问句、是非问句、回声问中的韵律表现。来自四个发音人的语音材料表明：疑问词与相应的动词结构在不同类型的句子中有韵律上的区别。在特殊问句中，疑问词是句子的焦点；而在是非问句中，动词结构是句子的焦点。句子的焦点成分在语音上实现为语调重音，因此，其词调拱度得以保留，有时得以加强，然而，相应的非焦点成分的词调拱度会被压缩，有时就被简化成平调。说话人通常会拉高整个回声问的调域，以及扩张句中疑问词的频率范围，来表达惊讶的意义。然而，材料并不支持音长和音强与焦点、非焦点之区分相关联。

2.2.10. 汉语普通话重音结构研究

在同言语工程学界合作的过程当中，逐渐对汉语的重音及其分布规则积累了一些新的认识。结合语音合成试验，考察了汉语重音的层次结构及其与其它韵律特性的关系。实验结果表明，语句中的重音分布受节奏层次和语调结构的制约，也具有一定的层次和等级结构。而这种结构又跟语法层次存在一定的相关关系：首先，常规的重音等级取决于它所处的句法层次的高低；其次，在同一韵律单元内部，重音的分布又主要取决于该单元的语法结构特点（详见《基于文本信息的韵律结构预测及其在合成系统中的应用》）。同时，还跟它在韵律单元中的位置前后具有一定关系。事实上，句法上的层次是底层语义结构的外部表现，所以，从这个意义上来说，语句中的重音分布本质上是受语义结构的制约。

2.2.11. 普通话两音节词焦点重音的语音学和音系学研究

针对普通话不同结构不同声调组合的两音节焦点词所传达的焦点重音的语音实现和音系表征进行了系统的分析和研究。在对焦点重音的语音表现和音系表征进行描述时，把音节底层的音调目标值作为描写的基本单元。主要发现有：⑴在焦点域处，焦点重音将焦点词两个音节的音域拉大，拉大的实质在于两个音节的 H 调都被显著抬高；焦点重音对焦点处 L 调的作用不明显，作为焦点词的两个音节，其基频均表现出被抬高的情况，这是汉语作为以双音节词为主的声调语言所特有的现象。⑵焦点词后音节的音域被焦点重音压缩，压缩的实质在于焦点词后阴平和阳平的 H 调被压低；但对于所有 L 调和去声的 H 调来说，该作用不显著。⑶焦点前音节音域受焦点重音影响非常微弱。⑷焦点重音的作用范围可以延伸至句末，无论负载焦点词的句子的音节数量是 7 个还是 11 个，焦点重音都可将句末音节的音阶显著压低。⑸ 不同结构的焦点词所传达的焦点重音的实现方式不同。⑹在优选论的框架下，我们通过分析发现，焦点重音的分布是由一组等级排列的制约条件决定的。具体地说，当焦点词两个音节底层具有的 H 音调数量不同时，H 调多的音节负载焦点重音；当焦点词的两个音节底层具有的 H 调数量相同时，后重制约条件决定焦点重音的分布；对于重叠结构，轻声永远为轻，焦点重音必须落在其前面的非轻声音节上；当焦点词的结构为“前缀＋词根”且后重时，与同样声调环境的“修饰语＋中心词”结构相比，前缀要轻于修饰语，这是形态句法制约条件决定的。

2.2.12. 韵律词的组构规则研究

根据对相关语料库的文本信息和语音信息的实验分析，我们发现，在实际的话语中，为了满足韵律上的需要，单音节的语法词都会或者通过延长该音节，或者前附或者后附到另一个标准音步上，或者跟前后其它的单音节词组合，设法凑足一个音步，从而构成一个韵律词。不过，这种韵律词的组词并不是任意的，而是遵循一定的原则，具有一定的规律。

实际言语中的韵律词组构的主要原则是：第一，单音节词处于重读地位，或者因为左邻右舍已经是完整的音步而无所依附时，就通过延长而构成独立的音步；第二，单音节词彼此粘附或与其它双音节词组合成音步时，虽然具有很大的自由度，但仍然要受各自词性的制约。

最近，我们在对ASCCD朗读语料库进行手工分词和词性标注的基础上，对比分析了韵律词和文本词之间的区别，研究数据表明，绝大部分单字词不能独立构成韵律词，需要跟与其相邻的词语进行合并，这是

导致韵律词和文本词不一致和很难对齐的重要原因。单字词跟相邻词语合并时，是选择向左合并还是向右合并，在很大程度上受控于其自身的词类：单音节的量词、助词、方位词、名词和介词等倾向于向左合并；单音节的形容词、数词、副词、代词和动词倾向于向右合并；此外，单字词的合并方向还可能会受到相邻词语的词类和词长等因素的影响。通过规则把某些单字词跟其相邻词语进行合并之后，能够在一定程度上提高韵律词的识别正确率。

2.2.13. 语速问题

根据对朗读话语的实验分析，对语速问题得出如下几点看法：

（1）作为语速快慢的度量，每秒钟所说的音节数，不能笼统地用包含停顿在内的话语总时长来计算。尤其是比较不同话者的语速时，必须计算出他们的实际音速，才能真正反映听觉上感知到的语速差异。

（2）话语中通过音节的时长伸缩而实现的语速变化及其调节方式，跟音节在语句中的位置密切相关，并且显示出一定的层次和规律。总的说来，语句内部中间位置上的语速相对较快，而首尾位置上的语速相对较慢。而且，起首位置上的调节主要是通过声母的加长实现的，而末尾位置上的调节主要是通过韵母的延长或缩短实现的。

（3）作为总体语速变化的客观变量，不仅在于时长的伸缩，还涉及音高的调节。总的趋势是，语速加快，语句的总体音高就相对抬高，总体音域相对扩大；语速减慢，语句的总体音高就相对降低，总体音域相对缩小。

（4）根据本实验的观察结果，总体语速的改变并不是均匀地分布在整个语句上，而是一种非线性分布。

同时，我们还对英语语句在快中慢三种语速中的韵律特征进行了对比分析，考察了语速对音高、时长、重音以及韵律结构等各个方面的影响。具体结论见本文集文章《不同语速的英语语句韵律特征对比分析》。对汉语的相关问题的研究也正在进行，目的是探索篇章级局部语速变化策略，使合成系统在篇章范围的语速/节奏变化更趋于自然。

2.2.14. 朗读语篇中重音与词类的关系

考察在朗读语篇中，重音在不同词类中的分布情况以及与词类的关系，另外考察在两音节词结构中，构词结构对重音的影响情况，焦点与重音的关系，不同发音人重音处理的异同问题。

通过统计分析，得到初步结论：

（1）重音音节占到总音节数的 38%和 40%，不同重音等级的音节在整体时长上有明显的不同。

（2）在不同的词类中，音节的单、双对重音的分布有一定的影响。在单音节词中，动词、名词、副词出现重音最多。在双音节词中，名词、动词、形容词、副词出现重音最多。

（3）双音节词中，两位发音人的词首位置出现重音分别占 70%和 77%。词首重音占多数。双音节词中，述宾和状中结构的词出现词首重的百分比略低于其他结构。但都大于 50%。结构对重音位置的影响不明显。双音节词中，不同词类中，前重的占多数，只有数词的影响不明显。

2. 3. 情感语音研究

国内外对“Emotional speech & Expressive speech”的研究转向自然口语，因为口语对话系统中，情绪和语气表达自然丰富。如法国的 CNRS 的情感对话库收集的是急诊救助热线的真实内容，通过对这个情感库在语音层和情感分类等多个层次的标注，应用到正面和负面情绪的识别系统中。

我们收集了电台和电视台情感类节目语音作为情感语音库CASS_ESC的口语部分，开展情感语音表达和情感词关系的研究。对基本情绪库，开展了情感重音的研究、音质（Voice Quality）声学分析等.

2.3.1. 普通话高兴语音的基频曲拱分析

利用基本情绪库，对普通话“高兴”语音与相应的“中性”语音进行“基频曲拱”对比分析。通过统计分析发现，基频曲拱对高兴语音的表达具有显著作用。中性语音与高兴语音的句子基频曲拱都是下倾的；但是高兴下倾趋势明显小于中性下倾趋势。而且相比中性，高兴韵律词的调阶整体抬高，高兴不同位置韵律词的词尾更加上翘，这种“上翘”在高兴的句尾表现尤其明显。

2.3.2. 情感句重音模式研究

情感句中重音模式对合成情感语句至关重要。我们从感知的角度，分析了两位发音人的高兴、害怕、难过、生气这四种情感句的重音模式与对应中性句重音模式之间的关系，同时，考察发音人之间的情感句重音模式。对于相同的语句，两位发音人的重音分布模式在各种情感状态下相似。不管中性句的重音模式如何，对应的“高兴”，“害怕”，“难过”和“愤怒”情感句重音有向句末转移的趋势，转移程度有所区别，从高到低依次为“愤怒”，“害怕”，“难过”，和“高兴”。详见本文集《情感句重音模式》。

2.3.3. 情感语音合成基频抖动的分析与建模

基频抖动特性是情感语音音质的特性之一，我们研究了高兴、愤怒、悲伤、难过以及对应的中性语音的极品抖动特性。Jitter(TJ) 分为两类：Random Jitter(RJ)和 Deterministic Jitter(DJ)。DJ 的产生主要是情感变化，RJ 是许多小随机事件的累加，比如说左右声带的不对称，声带上粘液的影响，气流造成的湍流噪音，等等。用两种方法对 jitter 进行了建模。

2.3.4. 口语中的副语言和非语言现象与情感和态度表达的关系

口语中的非语言现象和非语言现象非常重要，表达说话人的风格以及传达说话人的情感、情绪和态度等，这些都被称为非书面语信息（non-verbal information），如喘息、拖音、句间沉默、笑声、哭声等等。

我们以常见的喘息段（breath segment）为例，研究韵律结构和情绪唤醒度（valence）、情绪活跃度（activation）等跟喘息段声学特性的关系，包括喘息段的强度和时长。通过对不同参数的分析，我们得到了朗读语料时喘息段在不同情绪唤醒度和活跃度状态下出现的一些普遍性规律，利用合成的部分语料，进行了听辨实验，得到比较理想结果。详见本文集《朗读语流中的喘息音段分析研究》。其他的口语特性如沉默的表达作用，也在探究之中。

2. 4. 口音问题研究

关于口音问题的研究，国内过去主要是以方言自身的研究及普通话推广教学为目的，在此基础上编写大量的汉语方言大词典及各地学习普通话指南。语音上的对比研究作的相对较少，即便有也是某方言与普通话或方言与方言之间的对比，且大多只是作定性研究。近期虽然有人试图应用实验语音学的方法，从声学表现出发作定量研究，但也只是局限于方言和普通话内部，作标准普通话与地方普通话对比研究的基本没有。但是不管怎样，方言及普通话自身的研究仍有很大的价值，因为要作地方普通话与标准普通话的对比研究必须以这两方面的研究为基础。使得研究结果为普通话学习和普通话客观评测提供语音学方面的指导。

言语工程在语音评测和教学上的应用，渐渐明朗。除了地方普通话，涉及一些中介语语音问题的研究。我们尝试使用语料库的方式开展了英语为第二语言的中国人的英语语音分析。

2.4.1. 武汉、厦门、上海普通话标准普通话的语音对比研究

本研究结合实践研究和应用的需要，基于语音语料库，运用对比语言学的观点，用声学语音学的途径，详细分析上海普通话、厦门普通话、武汉普通话和标准普通话的音段和韵律特征的声学相关物，系统对比考察上海普通话与标准普通话的声学语音学差异。

按照语音的标注和标注人主观评测结果，对地方普通话的口音进行归类：重、中、轻三个级别；韵律方面，重点考察中度口音的发音人与标准普通话的声调调形、轻重音分布模式及声学表现，以及语调变化上的异同；音段方面对比方言普通话与普通话声韵系统的差异，并考察由方音引起的音变现象，初步进行数据统计及分析，找出音系上的对应关系；重点研究中度口音发音人与标准普通话整个元音系统差异，包括单元音，二合元音，三合元音及带鼻尾的复合元音，分别对比各自的声学元音图及共振峰模式图，归纳出相关规律。

从统计分析结果可知，仅仅以音段的错误结果参与评分与主观评分的较大差异，说明仅仅以音段的评分结果作为最终的评分标准是不行的，韵律的因素在评测中占有重要的因素。也说明，在机器自动客观评测中，不但要识别出音段的错误，而且也要发现韵律的问题，两者结合才能给出一个客观的评分。在普通话的学习中也是一样，音段和韵律的习得都很重要。

2.4.2. 台湾国语与普通话的单字调对比研究

这一研究采取实验语音学的方法和手段对比分析台湾国语和普通话中的四声（阴平、阳平、上声和去声）调值，并考察台湾地区发音人的方言声调系统是否会对其国语的四声调值产生显著影响。研究结果表明，台湾国语和普通话中的四声调值存在以下差异：（一）台湾国语的声调音域范围相对较窄，平均音域为 7.6 个半音；普通话的声调音域范围相对较宽，平均音域为 12.8 个半音。（二）台湾国语中阴平字的调值大多实现为 44，略低于普通话中的阴平调。（三）台湾国语中的上声字多为低降或中降调形，很少出现类似于普通话中的升调尾。（四）台湾地区的发音人在读阳平字的时候，通常会先有一个较大幅度的音高下降，或者出现一段相对较长的音高持平，然后才上升，以至于实际的音高上升点比较靠近调型段的末尾部分。为了考察这些调值差异的产生根源，这一研究还对比分析了台湾地区发音人朗读的闽南方言材料，发现两地语音差异在很大程度上可以从发音人的方言背景里找到的根源，发音人的方言声调系统在一定程度上对其国语的四声调值产生了显著影响，这一研究方法为今后同类研究提供了可借鉴的参考价值。

2.4.3. 四个地区的普通话声调对比研究

对上海、厦门、武汉普通话以及台湾国语的单字调和两字连读变调模式进行分析，与标准普通话进行对比研究，我们发现，对于中度口音的发音人，他们的上声在单说和位于两音节词末的时候与普通话的上声有差异；即使发音人单说的上声调“很到位”，在两字组连读时，出现在后字也可能出现问题，说明单字调的“习得”不能解决连读中声调问题；此外通过比较标准普通话、方言和带口音的普通话之间的调阶和调域，发现发音人在讲有口音的普通话或者国语的时候，会尽量保持母语方言中的调阶和调域，在语言学习中，母语方言的调域被顽固地保留，其次是是调阶，最容易改变的是调形，这也是因为调形有区分词义的功能，而其他俩者却没有。 通过对有口音的普通话以及国语上声的感知实验我们可以看出，发音人的声调偏误，都是因为母语缺少上声这样的调形引起的，必然用一个母语方言中一个较为接近的声调来代替。

2.4.4. 中国英语学习者英语语句的韵律特征差异分析

从实验语音学的角度，对比研究了中国英语学习者 CL2 (Chinese English Learners)朗读英语语句的节奏模式与标准美国英语发音人和标准英国英语发音人 EL1(English as first language)之间的节奏模式。包括音高重音（pitch accent）在语句中分布模式、语句的韵律结构、时长结构等。初步结果如下：(1) 通过分析韵律边界发现 CL2 与 EL1 的各级韵律边界出现的个数没有明显分类差异，语调短语含有中间短语的个数也没有明显的分类差别。尽管 EL1 之间的边界一致率普遍高于 CL1,但是 CL2 与 EL1 的韵律边界一致率高低与英语水平没有绝对直接的关系；（2）CL2 的英语水平跟音高重音位置相关，音高重音位置跟母语习得者相同个数越多，习得者的英语水平越高;(2)CL2 的英语水平跟重音实现的方式相关，EL1 的音高重音所在重读元音时长比 CL2 长，英语水平高的 CL2，其音高重音所在元音时长显著长于水平低的发音人，接近 EL1 发音人的长度；(3) 通过分析韵律单元的时长以及包含的词汇词个数，发现 EL1 与 CL2 的韵律结构差异主要表现在语调短语层，从节奏模式看， CL2 与 EL1 都表现出稳定的音高重音实现域 ID（或稳定的中间短语）和语调短语。(4) 通过分析韵律单元后的停顿，发现 EL1 与 CL2 的停顿策略在韵律短语单元之后有差异。(5) 音高重音的个数与句子长度相关，但是两者有差异，随着句子长度增加，EL1 的音高重音的增加个数明显小于 CL2。也说明韵律结构的差异在较长的句子表现更明显。

2.5. 音段与音段音变的研究

2.5.1. 音段音变研究

对减音问题进行深入考察，并对语音变化的稳定性问题进行分析。研究语料包括朗读语料和口语对话语料，研究范围包括对个体发音与不同个体发音之间的比较。声学分析包括对出现减音的词语进行基频、能量、时长、共振峰的测量。精细实验包括考察语速的影响问题，个体发音的稳定性问题等。通过实验发现，连续话语中出现的减音有一定的规律性，轻读和在韵律词内位置、语速及音节内部结构都会引起减音的出现。减音所表现的稳定性是个体发音比较一致，不同人所表现的趋势比较稳定。

对语音不变量问题的考察与研究。语音不变量理论是 Stevens 和 Blumstein 在 1978 年提出的，其中心内容是在语音音段自身存在反映语音特征的声学不变量，因此可以从塞音爆破段本身寻找线索来判定其发音部位的特性。

对声母浊化的听辨实验，实验结果表明，在词语内声母的浊化对听辨没有明显的影响，孤立音节时声母的浊化特别是冲直条的有无直接影响听辨者对音节的判断。

2.5.2. 鼻尾丢失

在研究过程中我们还考察了音段因素、韵律因素以及这两种因素之间相互作用对鼻尾丢失的影响。我们发现带鼻尾音节的后继音节的声母的发音方式，带鼻尾音节自身的声调以及带鼻尾音节之后的韵律边界对鼻尾的丢失有显著影响。带鼻尾音节自身的声调与带携带重音的类型之间的相互作用、带鼻尾音节自身的声调与带鼻尾音节之后的韵律边界的类型、带鼻尾音节自身的声调与带鼻尾音节的后继音节的声母的发音方式对鼻尾的丢失也有显著影响。

2.6. 关于音段的语音实现与超音段的韵律结构的关系的探索

这部分涉及两方面主要内容：

第一，通过分析音节的时长结构，研究各种韵律延长的类型及其不同特点。研究和讨论的基础主要是对连续话语语料库语音的测量和分析。结果表明，存在着三种不同类型的音段延长，每一种类型都与一定的韵律事件相对应，都有各自独特的时域变化方式。具体说来，边界前的音段延长以音节韵母的延长为主，边界后音段延长以声母辅音为主，而与重音凸显

相关的音段延长则是涉及整个音节的、相对平衡的延长。因此，音段的不同延长效应，也许可以作为了解言语韵律结构的有价值的信息。

第二，采用声学分析与动态的电子腭位测量相结合的方法，通过基于普通话语料库语音的分析考察，研究自然话语中音段的语音实现跟韵律结构的关系，并在此基础上探索音段发音的增强与减缩的语言学动因及语音学机理。

研究结果表明，跟其他许多语言里的情况相仿，汉语自然话语里同样存在着韵律导向的、有规律的发音增强（strengthening）和减缩(reduction)。这是一种在言语交际过程中普遍存在的适应性发音调节现象。它出自两种对立面的统一：交际双方要求之间的矛盾统一和言语交流的需求跟发音局限之间的矛盾统一。其本质就是，通过规律性地调节音段发音实现的程度，突出话语的结构关系，最大程度地实施有效的言语交际。因此，它不但为韵律结构提供了又一个可靠的客观标志，而且进一步提高了我们对言产生机制的认识。

2.7. 语音的生理研究

2.7.1. 元音、双元音产生的机制研究

以汉语宁波方言为例，检视了元音、双元音产生的发音和声学特性。研究使用电磁发音仪采集了七位发音人的发音器官运动数据，结合声学数据，对元音、双元音产生的静态发音位置、发音器官运动学（主要是舌运动学）等内容进行了考察。

对单元音的研究表明：元音的高低与舌位的高低没有明确的对应关系，这证明了传统的元音描写术语，如元音高低、前后，并没有在生理上和发音上的实证基础，而只是有语言学上的分类意义，在语音学上，与之相关联的，正如前人所指出的，应该是声学上或者是心理声学上的证据。在发音上，对材料所进行的平行因子分析结果显示，表面上纷繁复杂的舌位信息可以成功地分解出两个底层的舌移动机制，“前举”和“回缩及后举”，所有元音的发音都可以用这两个机制来模拟，而传统的元音描写维度如高低、前后，也可以用这两个机制来作重新的阐释。与前人对英语及德语等西方语言研究结果的一致性也表明，这两个舌移动机制在元音发音上是具有普遍性的。

国内外文献中，双元音的研究主要集中在对其静态特性的分析上面，已有的动态特性方面的研究，也只集中在从声学的角度出发，比如研究双元音共振峰变化的速度。我们的研究不仅进行了双元音产生的声学分析，同时运用电磁发音仪所采集的六个发音人的数据，从舌运动学的角度探讨汉语宁波方言双元音产生的发音机制。结果表明：在区分前响双元音和后响双元音的条件下，舌平均运动速度和舌峰值速度基本能将不同的双元音区分开来,即不同的前响或后响双元音分别具有不同的舌平均速度和舌峰值速度。但是，这两个参数都不能将前响双元音与其相应的后响双元音区分开来。研究发现，另一个舌运动学参数，峰值速度出现时间，是区分前、后响双元音的较好的参数。一般说来，在前响双元音中，舌运动的峰值速度出现得较晚；而在后响双元音中，舌运动的峰值速度出现得较早。这是因为速度峰值出现在发音动作变化过程中（对应于声学上的谱变化），并由宁波双元音自身的时间结构特点决定。至于短双元音，在发音时舌运动明显加速，但发音幅度变化不明显；有意思的是，当音节变短时，峰值速度出现时间相当稳定。

2.7.2. 不同发音器官之间的协同关系研究

在语音产生的过程中，不同的发音器官之间存在着协同关系，发音、发声器官之间也会有一定程度的相互影响，有时，甚至非发音器官（如头部运动）与语音也有某种关联。

运用电磁发音仪所采集的宁波方言材料，我们检视了元音产生过程中舌和下颚之间的协同关系。结果发现，二位男性发音人在元音发音中倾向于使用不同的发音策略。大致上说，发音人一更多地利用颚位的区分来对不同高度元音之间相似的舌位进行策略性的“补救”，反之亦然；而发音人二侧相对齐整地运用舌、颚协同，高颚位往往与高舌位相伴随。

其它条件相同，高元音伴随高声调，低元音低声调，这被称为元音的固有声调（instrinsic pitch）。固有声调在非声调语言中是广泛存在的。运用电磁发音仪所采集的材料，我们检视了宁波方言中高调与低调对元音发音时的舌位是否存在影响。结果发现：同一个元音，低调（低升调、低短调）与相应的高调（高升调、高短调）在发音上是不同的，在发低调时舌位相对靠后。然而对于同一个高降调内部的高调成分和低调成分之间，则没有发音上舌位的不同。

在对所采集的一位普通话女性发音人的发音材料的分析中，我们发现：在元音发音时，第三声与其它声调存在较明显的不同，虽然不是在所有的元音或声调对比中存在统计上的显著性。这种不同不仅体现在舌、颚、唇，而且还表现在发音时的头部移动（head movement）上：在发第三声时，发音人的头部较低、较后。

2.7.3. 汉语方言部分鼻化辅音研究

晋语、广东中山附近的粤语、闽南话、广东清新县的客家话等汉语方言都有历史上来源于纯鼻音声母的部分鼻化辅音。本研究录制了来自这四个方言区的共64位发音人的口、鼻气流/口腔间气压和/或声学语料。到目前为止，研究发现在这些区域的方言中，部分鼻化辅音是较广泛地存在的。晋语（尤其是晋中和晋西地区）的部分鼻化辅音最为丰富。在五个发音部位都发现有部分鼻化塞音：双唇（labial）、齿龈（alveolar）、后齿龈（postalveolar）、腭龈（alveolopalatal）、硬腭（palatal）、软腭（velar）；另外还有一个部分鼻化齿龈擦音。其它区域的受调查方言都有部分鼻化双唇、齿龈、软腭塞音；其中潮汕闽南话还有一个部分鼻化齿龈擦音。

有趣的是，部分鼻化后齿龈塞音和腭龈塞音可以各自和它们相对应的部分鼻化塞擦音互换，即形成自由变体；类似地，部分鼻化齿龈擦音也可以和它相对应的部分鼻化齿龈塞擦音互换，形成自由变体。

定性、定量研究均发现闽南话的部分鼻化辅音与其它方言中的不同——前者的鼻音段较弱。研究表明：闽南话处于去鼻音化（denasalization）这一历史音变的后期，而其它方言则处于早期。

3. 进一步的研究与发展思路

我们的研究与发展思路要围绕当前国际上语音领域的热点问题开展，为言语工程和言语教学服务，并且深入研究语音学相关理论。当前国际上语音在三个方面的特点概括如下：

生理方面，研究范围、力量手段不断扩大和更新，研究内容更加深入。生理语音学集中在嗓音声源研究和口鼻腔调音机理研究两大块。能够利用更多的生理分析设备进行研究，如动态腭位仪器(EPG)、喉头仪（EGG）、高速摄影和高速成像仪、肌电仪(Electromyography, EMG)、电磁发音动作描迹仪(Electro-Magnetic Articulograph, EMA)、口鼻气流气压计等。

心理方面，用 FMRI 等手段，研究语音和语言在大脑中的分区和功能。国内开始研究情感语音在大脑的成像与认知过程。儿童语言习得主要集中在语法、词汇方面，儿童语音研究仍然没有大规模的开展。

声学方面，对自然口语的声学研究仍然是热点。语音识别中目前面临的巨大困难就是自然口语中的噪声如口音、发音人讲话风格和环境噪声等；自然口语的韵律特征、音段和音质特征与情感和语气的表达的关系；自然口语的韵律建模；从社会语言学的角度、认知的角度对自然口语进行收集、分析和标注；口语中的语音变化模式研究。

自然口语是即兴的、事先没有准备的口头交际话语，与书面语和朗读话语相比，自然口语是最自然的、使用频率最高的、使用范围最广的言语交际形式。因此研究自然口语的语音特征，探索并掌握自然口语的语音规律，能够在一定程度上提高机器对自然口语的识别和理解能力，并有助于合成出更自然、更符合人类语感的话语来，这无疑有助于促进人机对话系统的发展。另外，口语语料所提供的是自然的而不是人工诱导出的言语，这意味着口语语料研究更有可能反映实际生活中真正使用的语言的特点，而过去的汉语语音学研究大多以实验语句或朗读语料为研究对象，这显然不利于考察自然口语语音的真实表现和功能。因此从应用和理论两个方面看，当前我们的研究应该加强自然口语语音的研究工作。

3. 1. 加强语调韵律方面的研究

按研究对象的不同可以将语音研究分为两大基本领域：音段研究和韵律研究。音段特征与语音的可懂度密切相关，韵律特征与语音的自然度密切相关。当前通用的语音合成系统大多采取音段拼接合成的方式，这种系统合成出来的声音相对具有较高的可懂度，但在自然度上还需大力改进。此外，语句的语调韵律特征还可以用来表达语气类型、篇章结构和语义焦点等言语信息，以及交际者的情绪状态等情感信息。因此要识别和合成出抑扬顿挫的、充满感情色彩的、能够满足各种交际需要的言语来，需要加强语调韵律方面的研究工作。

3.1.1. 焦点重音和语速的研究

现在正在进行的项目主要涉及焦点重音对非轻声两字组和五字组时长变化的影响。对于两字组时长变化模式的研究，主要关注不同声调组合的两字组，其在负载句的不同位置时，受焦点重音的影响其时长变化规律。对于五字组而言，主要关注全阴平、全阳平、全上声和全去声组合的五字组，在负载句不同位置的时候，受焦点重音的影响其时长变化模式。

考察普通话朗读话语的语速在句子层面和篇章侧面的变化策略也是正在研究的课题。

3.1.2. 集中精力解决好功能语调问题，并为情感语调的研究打好基础

语调是作用于语句层面上的具有区别语句意义（理性意义和/或情感意义）的韵律变化模式，从功能上讲可以分为功能语调和情感语调两个层

面。功能语调在结构上由两个主要成分组成：边界调和句重音，通过这两个成分来表达语句的语气类型、篇章结构以及语义焦点等理性意义，具体说来大致包括以下几个方面：1）.分界功能，把句子从语流(话语)中切分出来；2）.篇章功能，表示句子在语篇中的组织结构；3）.言语行为功能，表示语句的表述目的，从而提醒听者对其做出合适的反应；4）.突显功能，通过逻辑重音突显句子的语义焦点。在这几种功能中，最后一种功能的实现主要和句重音密切相关，前三种功能的实现主要和边界调密切相关。情感语调主要用来表达说话人的情绪、态度和感情等情感意义。就当前的语音学研究和言语工程应用来看，最迫切需要解决的是功能语调，以实现这四类基本功能的表达和理解，满足基本的言语交际需要，同时还要为情感语调的研究打好基础，如情感的分类，情感语料的搜集、分类和标注等等，以逐步推进自然语句的语调韵律特征研究。

3.1.3. 大力研究句间和超句的语调韵律特征

过去汉语语音学研究主要考察语句内部的韵律特征，而对超句和句间的语调韵律特征考察较少，这不利于全面地了解韵律特征及其功能。一方面，只有掌握了语句之间的语调韵律特征，考虑句子之间在韵律上的衔接与过渡，才能生成出连贯自然的段落篇章来。另一方面，句子是最基本的言语表达和理解单元，要理解一段话语，首先必须将句子从语流中切分出来。要将句子组织成段落篇章或者从段落篇章中将句子切分出来，都必须充分掌握自然话语中语句边界的韵律特征，因此开展语句边界的语调韵律特征研究无疑是必要的。

3.1.4. 考察语调韵律特征和话语交际功能（语用）之间的关联

过去汉语语音学研究主要考察韵律和句法之间的关系，而很少探讨韵律和语用交际等因素之间的关系，这不利于深入地认识韵律特征及其功能。从言语交际的角度研究语调韵律特征，考察其分界功能、言语行为功能、篇章结构功能和突显焦点功能等,一方面可以揭示不同韵律表现的背后成因，从而可以透过现象看本质，通过话语的韵律表现了解说话人的交际意图；另一方面可以通过有效地运用韵律手段来实现不同的交际功能。

3.1.5. 定量研究话语的语调韵律特征，建立精确的语调韵律模型。

过去对语调韵律特征的研究大多是定性的，如轻重、高低、升降、长短等，而这些定性描述比较适合于语言教学，而对于言语识别和合成等工程应用领域来说是远远不够的。研究语调韵律特征的定量模型，采用一些学习模型，如 NN 模型生成基频曲线，这无疑有助于推动言语识别的自然度和言语合成的精确性。

逐步开展韵律组织和重音分布方面的研究工作，对“汉语的韵律结构及重音分布”进行研究，考察话语韵律结构与其句法结构之间的关系和异同，分析韵律组织和重音分布的控制因素，以期在语音合成中能够生成出合适的韵律模式来，以改善语音合成的自然度。

语气词在自然口语中的作用非常突出，它的语言学和语用功能是我们关注的热点之一。

3.2. 汉语普通话学习和评测中的语音学问题研究

这个课题涉及了两个方面的研究内容：一个是中国方言区的人讲的地方普通话的语音特性研究，另外一个是外国人讲的普通话的语音特性研究。前者是针对地方普通话的语音研究，后者是对外汉教学中国外汉语学习者的语音研究。

语音学教学国际会议（phonetic teaching and learning conference –PTLC）反映了语音教学的一些新的动态，比如：Grit Mehlhorn提出了个人语音教练IPC (Individual Pronunciation Coaching)概念，帮助学习者对语言的认识和学习的自主性。他发现学习者个人学习能力不同，在音段、韵律、发音技巧、学习认知方式、学习策略、对语言的意识、自我控制能力、学习目的以及期望等方面都与学习者的语音学习水平有关。IPC的功能包括了诊断学习者的个人发音，制定语音学习方案，帮助学自者将主要目标化解成一个个子目标，制定适当的学习策略，对学习者的进步进行反馈等。我们将以IPC功能为核心，在中介语语音库基础上系统研究音段和韵律发音偏误；并且结合生理发音仪提供的数据，给学习者提供可视化的学习助理。

3.3. 生理分析方面

近期的计划和目标主要有下面几方面：一、大规模采集汉语普通话的发音数据，建立数据库，并在此基础上面探索语音发音机制；二、在此基础上面，加强基于生理的语言和音系模型研究，推动本身学科内的理论建设；三、加强与言语工程学的合作，为建立语音产生的精确生理模型作出自身的贡献。同时，如果条件许可，我们寻求与拥有 fMRI 一起的科研单位或者医院的合作，深入进行语音的

生理方面的研究，在此基础上面，探索生理语音学的前沿性理论课题。而且，在进行生理语音学的深入研究基础上，结合婴幼儿语音的研究、心理方面的研究，是可以为学界另一个前沿性热点，认知科学，做出我们自身的贡献的，这是我们的中远期目标。

3.4. 加强自然口语中语音特征分析和会话言语行为分析

利用现有的多媒体技术（包括视频的和音频的）记录真实场景下的自然口语会话活动，然后对这些连续的视频和音频信号进行切分和标注，并在此基础上研究人类的会话策略，这是近年来会话分析和人机对话系统的研究热点。

我们将继续研究自然口语中表达个人风格、情绪和情感的语音特征，同时在面向任务的对话中，如会议和网络心理咨询这两种应用，针对通用言语行为标注进行一些探索性研究，确定言语行为基本分类，研究言语行为与语音和其他模态如姿势之间的关系。

致谢：感谢参加本综述撰写的各位老师和同学，他们是林茂灿、曹剑芬、顾曰国、 蔡德和、熊子瑜、胡方、陈肖霞、华武和贾媛。

参考文献：

《语音研究报告》，http://ling.cass.cn/yuyin/index.htm