面向情感的语音合成系统

陶建华[1]　许晓颖[1][2]

[1] 中国科学院自动化研究所模式识别国家重点实验室　北京　100080

[2] 北京师范大学文学院 北京 100068

摘要： 情感语音合成是近几年语音合成的研究热点，现有的研究多以语音的韵律和声学特征为指导因素。在情感语音合成的研究中，必须解决如下三个核心问题:(1)如何确定情感状态和情感语音的声学特征参数?(2)如何建立情感状态与语音的声学特征的关联关系，建立情感声学参数的综合控制模型？(3)如何结合文本分析和场景因素建立情感的预测机制？本文将围绕着这三个问题进行深入的探讨，在实验分析的基础上，对情感语音合成中需要处理的情感特征，以及情感特征与语音特征的相互关系问题进行分析，并针对这些关联关系提出了情感声学模型和情感韵律建模的思路，初步实现了一个情感语音合成系统的原型。

关键字： 情感语音合成、情感分析、影响情感因素、情感焦点、情感关键词。

1 引 言

语音是人类交际的最重要的工具之一。人类的说话中不仅包含了文字符号信息，而且还包含了人们的感情和情绪的变化。例如，同样一句话，往往由于说话人的情感不同，其意思和给听者的印象就会不同，所谓“听话听音”就是这个道理。而传统的语音处理系统多着眼于语音词汇传达的准确性，而忽略了包含在语音信号中的情感因素。情感特征的人工处理，在信号处理和人工智能等领域具有重要意义。近几年来，在自然语言处理、信号处理、随机过程处理等方法的推动下，语音合成技术获得了很大的发展，突破了传统的单纯语音计算算法的研究。情感语音合成的研究，适应了语音技术的未来发展趋势，由于它能够很好的将语音的口语分析、情感分析与计算机技术有机的融合，为实现以人为本，具有个性化特征的语音合成系统，奠定基础。

有关情感的论述可以从 19 世纪末的 William James[1]追述到二十世纪末的 James Russell [2]。从语音信号中提取情感特征，分析人的感性与语音信号的关联，将情感特征应用于语音合成方面的研究，只是国外近几年刚刚兴起的研究课题，大量的模型还没有得到很好的解决。人的情感被分为基本类和扩展类两种（Rene Descarte [4]），基本类对情感的描述起到重要的作用，通常情况下，情感语音的研究主要集中在情感基本类与语音声学参数的关联分析上，目前，针对情感基本类的常见的定义有：害怕、发怒、高兴、悲伤、惊奇和厌恶等六种，尽管如此，针对不同的场合，其分类标准依然会有所区别。通常的扩展，包括区别发怒的特征，增加蔑视、懊恼、厌倦、担心、傲慢和爱慕等，这些可以由性别特征以及其它特征区别开。每一种语言均包含着一些特殊的情感用语，Whissell [9]收集了 107 种反映情感状态的词，Plutchik[10] 则列出了 142 种，这些词覆盖了很大范围的情感状态，只有很少一部分可以被归纳到基本类。一些科学家通过分析，将人的表达方式从“憎恶”一直细分类到“狂怒”[11]。而这一分类则与具体的语言和文化密切相关。通常意义下，人们对情感的理解，主要集中在情绪的变化上，然而将情感进行细致扩展，则衍生到自然口语的表现方式，它相对于普通朗读风格，更贴近人的生活和接近人自然的情感流露和表达方式。

情感发音的实现，需要通过语音的声学参数体现人的情感特性，Sylvie J.L. Mozziconacci 在 IPO ('t Hart et al., 1990)语调方法的基础上初步加入了情感控制参数，增加了语音合成的表现力。Cohn[1]针对情感的声学特性编写了简单的情感编辑器，使研究人员可以细致的观测情感控制参数对语音输出的影响，对情感语音合成的研究起到了较好的推动作用。已有的研究多局限在零散和片面的领域，为建立较为完整的情感语音合成系统，涉及到情感语料库设计，

情感韵律特征分析及情感建模，语法、语义对情感发音的影响，面向口语的韵律分析及建模，情感语音声学模型的建立，场景分布对情感发音的影响，以及韵律个性化等一系列的研究。本文将针对其中的几项作一些较为细致的分析和论述。

2 影响情感的因素

研究情感语音合成，首先我们必需进行影响情感因素的分析。A. Paeschke & W. F. Sendlmeier[1]在他们的工作中，论述了英语中情感语音的韵律特性，他将影响情感发音的因子归结到激励、态度和反复三个基本因素，并在此基础上初步探讨了它们之间的一些联系。情感虽然与有机体的生理唤醒状态有着密切的关系，但它不是单纯地由生理唤醒状态决定的。情感产生的源泉是客观现实，但是，情感又不是客观现实直接、机械地决定的。作用于人的外部世界的各种事件与人的各种需要的联系是发生在认知活动之中的。客观事物对人的作用必须通过人的认知过程，而且由于人的认识的每一次活动又不是单独地被孤立的一件件事物决定的，人在生活实践中积累的知识和经验制约着当前的认识，并与人的态度或愿望结合起来。因此，人们对作用于他们的事物的判断与评估，才是情感的直接原因；同一事件对不同的人或在不同的时间、条件下出现，可能被做出不同的评估或料想，从而产生不同的情绪。

正是由于过去经验制约着人对当前事件的认识和评价，当事件是符合或加强人的认识和愿望时，就产生肯定的情绪。偶然的好友重逢，能引起旧日友谊的重现，因而符合主体的道德需要；意料之外的成功，生活或工作中困难的突然拓通，主体愿望的实现，这些都会引起不同程度的喜悦和快乐。但是，当出现的事件被判断为并非是所愿望的，被料想为难以控制这些不利事件的影响的存在，这时就容易产生否定的情绪。因此，情绪和情感是通过认知活动的"折射"而产生的。所谓认知的折射就是指人在过去经验中所形成的愿望与渴求的系统对当前认识活动的影响。

因此，现代研究一般地支持这样一种观点，即情感为三种因素所制约：环境影响、生理状态和认知过程。其中认知因素在情感的产生中起关键性的作用。如上面分析，情感语音的研究，需要与人们对语言文字的认知和理解、对环境等其它诸多因素的理解，紧密结合起来。

3 情感语音的声学分析与建模

情感语音的声学分析是情感语音处理最易入手的步骤，通过声学特性分析过程，为得到情感状态下的声学参数综合控制模型带来帮助。[1][2][3][4][8][11][12]均对此进行了较为详细的分析，然而针对汉语的情感声学特征的研究，却少有人进行，为得到情感状态下的声学关联关系，本文在分析情感语料的基础上，进行了一定的总结。

3.1 情感语料

情感语料是进行情感语音合成研究的重要基础，目前，国内外现在还没有提出用于情感分析的语料设计标准。大部分已经存在的西方语言情感语料库多采用演员录制的方法[6][7]，由于区别特征明显，这为分析带来了很大的便利，但经过艺术加工的声音，在很大程度上并不能反映真实生活中的语音情感特征。真实生活中的语料与不同的文化、发音人和背景有较大的关联，语料收集存在着很大的难度。为达到情感语音建模的目的，本文则采用了演员录制和真实场景相结合的方法，在演员录制中，选用了 28 个演员充当说话者，其中 14 个男声 14 个女声，每个人录制了 1580 句具有 5 种不同情感的语音，包括陈述句、疑问句和感叹句。自然场景中的语料则选用了由社科院语言所提供的 CADC 语料库，共 1613 个即兴对话语句。该语料使用了 praat 工具进行标注，包括基频、音节边界、副语言学信息等。考虑到情感因素，语料处理中进一步加入了情感状态和情感关键词属性的标注内容。

3.2 情感语音的声学分析

由于人对语音的感知是非常多样化，全面考虑情感的声学特征是一个非常困难的工作，考虑到计算机的处理能力，只能通过部分参数从一定程度上对情感语音的声学特性进行了概括。一般情况下，语音的情感相关性的表示形式可以通过说话人模型或者声学模型来实现。Cahn[1]将其归结为四类。由于汉语的韵律多以音节为处理单位，在这种有调音节的韵律分析中，音节的韵律特征起着非常重要的作用，因而，为便于在汉语中处理，本文将情感语音的声学特征直接分为三类：韵律类、音质类和清晰度类。概述如下：

3.2.1 韵律类

韵律类主要用来表征不同情感状态下语气的变化，它包括如下韵律参数描述：

平均基频：整个语句的基频平均值。

基频范围：整个语句的基频范围，基频范围在很大程度上能够反应人的情绪状态（积极情绪或消极情绪）。

重音的突变特性：在情感语句中，重音多体现情感焦点特性，经常由情感关键词承载，在积极的情绪中，它多能体现情感状态的激烈程度。如：发怒时，情感关键词往往出现突然的重音加强特性。

停顿的连贯性：用以表示语句的停顿是否连贯。人在情绪受到压抑或快速膨胀时，有时会出现由于概念表述不清而导致的语气断续特征。

语速：用以表征语气的缓急程度，人在焦急、恐惧时多出现语速加快的现象，有时欢快的语气也能带来类似效果。

重音频度：重音的频度在一定程度上能够体现情感状态的持续性。

音强：音强也是用于情感确定的重要参数，经过实验分析，在情感语音中，音强的变化往往表现与基频范围的变化的一致性。即、基频范围增大时，音强也多表现为增强。但是，相对基频变化来说，大部分音强变化并不明显。

音节基频高线倾斜程度：语句中音节基频高点连线的变化情况（上升、水平和下降）。

音节基频低线倾斜程度：语句中音节基频低点连线的变化情况（上升、水平和下降）。

基频抖动：对于焦虑语音特征的则会出现“f0 抖动”现象，这一现象描述了基频从一个区域到另一个区域之间的快速和反复的变化。在此情况下，有时音节会失去其固有调型。

3.2.2 音质类

音质类用来表征在情感状态的语音的音质发生的变化，它通过如下参数描述：

呼吸声：在语音流中，出现呼吸气等声音。当一个人处于紧张或欢快状态时，出现的快速呼吸停顿，或当一个人由于恐惧而牙齿紧压产生的回旋气流噪声。

明亮度：低频能量和高频能量的比值，用以反映语音的清亮特性。

喉化度：发音时，声门出现不连续的脉冲震动特性，经常出现在极度恐惧的情感状态中。

3.2.3 清晰度类

情感信息与人的声道同样具有一定的关联。清晰度可分为正常、焦急、模糊和准确。清晰度描述了元音质量的变化和清辅音是否变化为相应的浊辅音。比如：人在厌恶时，有时说话“嘟嘟朗朗”，表达不清。

由于情感表现的多样性和复杂性，导致情感声学参数的数值分布多呈现较大的离散特性，表 1 则针对五种基本情感状态列出了几种基本声学参数的较为平均的体现。

	喜悦	发怒	悲伤	恐惧	厌恶
语速	较快，但有时较慢	稍快	稍慢	很快	非常慢
平均音高	很高	非常高	稍低	非常高	非常低
音高范围	很宽	很宽	稍窄	很宽	稍宽

音节基频高线变化	平滑，上升变化	陡峭，在重读音节处	下降变化	正常	宽，下降终端变化
音节基频低线变化	平滑，上升变化	没有太多的变化	下降变化	正常	下降终端变化
音强	较高	较高	较低	正常	较低
音质	有呼吸声，响亮	有呼吸声，胸腔声调	共鸣生	不规则发声	嘟囔的胸鸣声
清晰度	正常	焦急	模糊	准确	正常

表1，五种主要情感的声学特征

本文得到结果与Pereira [8]、Banse 和 Scherer [13]在平均基频、基频范围和音强上是较为接近的。然而，受限于语料的规模、语种以及情感表现的复杂性，部分参数的声学表现并不具有太多的比较性。

3.3 情感语音的声学模型

在情感声学特性分析的基础上，本文进而初步总结了语音合成中几种常见的韵律参数的调解方法，并成功的构筑了五种基本情感状态下的情感语音声学模型。

模型中的一些声学特征控制规则列表如下：

情感	语速	音强	音节基频高线变化	音节基频低线变化	平均音高	音高范围	其它
喜悦	+40%	+80%	平滑： +2	平滑：-2	+100%	+100%	轻微的呼吸
发怒	+20%	+50 %	较陡峭的重音形式+3	较陡峭的重音形式+3	+80%	+80%.	
悲伤	-20%	-50%	平滑： +4	平滑： -2	无变化	-60%	
恐惧	+20%	无变化	平滑： -2, flat	平滑： -2, 有抖动	+20%	+20%	有呼吸声
厌恶	-20%	-30%	平滑：+2, 向下	平滑：+2	-20%	-30%	

表2，汉语合成语音中的情感声学表达的控制规则

表2中，所有的“＋”号代表增加，“－”号代表减少。音节基频高低线的平滑程度表示是否会产生f0抖动。它为－5到＋5的一组数字，负值表示会产生基频颤抖特性。系统采用Klatt合成算法，通过多元激励模型实现了高性能的韵律调解、频谱变化和嵌入呼吸气声等主要功能。

4 语音合成中的情感预测

由于语言中的情感是直接与说话人的思想状态相对应的，在特定情感状态下，说话人的行为会直接影响发音的结果。人的情感并不单纯由文本信息决定，它与人所处的场景等信息，以及与人的感知密切相关。人在发音时，大脑会通过信息综合，酝酿情绪。目前虽没有有效的方法能够将针对所有的语法信息做出准确的概念分析和判断，但通过情感关键词、标注、标点等信息预测情感状态，依然能够进行情感的判断。本文阐述了一种基于情感状态预测网络(Emotional Status Prediction Network, ESiN)的方法对文本负载的情感状态进行预测。

4.1 情感焦点

情感状态预测网络的初始步骤，首先是确定情感的焦点。情感焦点在通常情况下，由情感关键词驱动，多出现在情景对话和具有剧烈变化的情感状态中。这一结论在表1中得到了充分的验证。在发怒语气中，承载发怒的情感关键词得到了突然加强。在文本的阅读过程中情感的表达是通常通过不同的情感焦点实现的，情感焦点受句法结构、声调结构和功能词的影响。在语句结构中，功能词体现了句子的主要意思，而情态词用于加强情绪。“激动”的情绪能够通过加强功能词或情态词而得到明显的表现。例如：我非常生气。短语“*非常生气*”表示了句子的关键的情感状态，并且在愤怒的情感时会得到有力地加强。其它的一些词语，如：

“不好”，“很”，“非常”等等也会达到同样的效果。与此同时，由于汉语是一个有调语言，这为声调模式下的情感焦点的确定带来了难度。Reyelt, Grice（1996）等人通过声调序列的模型，在对句法结构和信息结构的分析中融入了声调序列的作用，得到的更好的情感焦点预测效果。在已知情景的情况下，有时汉语的情感焦点也可以通过声调组合的分析来得到。为实现情感焦点的预测，作者构筑了大规模的情感标注词典，通过在语音合成的词典中加入情感状态描述属性来辅助情感焦点和状态的判断。

4.2 情感状态预测网络（ESiN）

ESiN 网络的基本组成部分是结点和链路。结点（情感载体）可承载情感信息而链路（情感传播者）可传播情感信息。ESiN 网络中的每个结点中有三个属性：词语（包括属性，如：词类、情感属性等）、情感矢量（$\vec{E}_t = \{e_{t0}, e_{t1}, \cdots, e_{tN}\}$，t 表示不同的节点。情感矢量表征一个节点的基本情感状态。情感矢量的分量分别代表不同的情感状态，N 表示情感状态的个数每一个分量的值从 0 至 1 分布，表示该情感状态的程度。其中 1 为最高，0 表示该情感状态不存在）、情感触发器（用以综合情感矢量，用以计算到当前节点为止的情感状态）。

ESiN 的链路表示情感的传播路径。根据不同语法规则，网络中有不同类型的链路。情感可以无延时传播，每个链路包含三个信息：方向（链路的起源和终结）、情感延时函数（延时函数控制情感传播）、情感生成概率函数（用于决定情感激励状态的概率）。

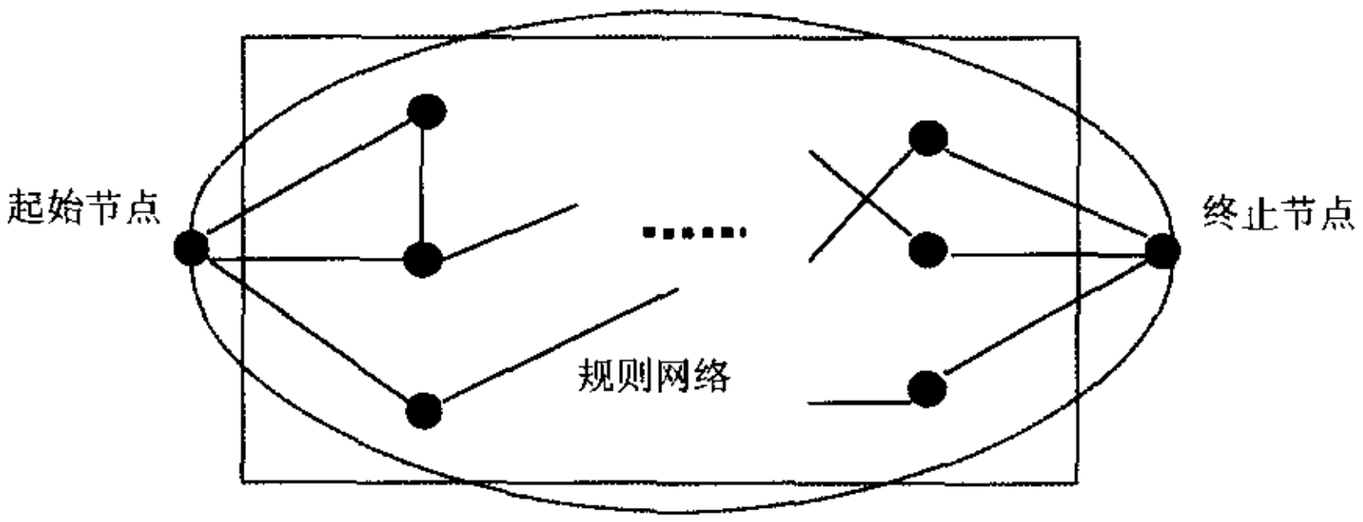

图 1：ESiN 情感状态预测网络

ESiN 网络如图 1 所示，整个工作过程包括：情感初始化、情感触发和情感传播等过程。

情感初始化包括：分析语句中是否存在情感标注或标点符号等信息、分析语句中是否存在情感关键词并确定情感焦点等步骤，同时确定网络初始节点的情感矢量值 $\vec{E}_0$。

在情感传播中，一个输入句子包含情感关键词（情况（a））这个词就用作传播源。若情感传播关键词没出现，我们就寻找功能词、情感标注或符号等信息（情况（b））。若情况（a）和情况(b)都没发生，则有情感触法值的词被用作传播源。这构成网络的初始化过程，情感传播值 Ep 从传播源计算并用转换延时函数(1)与情感矢量结合起来。延时函数(1)定义来确保情感适于传播。为了确保情感的汇集，情感传播在经一些阶段后延时为零。据以上标准，我们定义以下函数：

$$\vec{E}_t = D(\vec{E}_{t-1}) = \vec{\delta}(t) \times \alpha + \vec{E}_{t-1} \exp(-.005 \times t^2) + \vec{C}_t \tag{1}$$

$$\text{其中：}\ \vec{\delta}(t) = \begin{cases} 1 & \text{，当前节点为情感关键词} \\ \{P_0, P_1, \cdots, P_N\} & \text{，当前节点为非情感关键词} \end{cases} \tag{2}$$

$$P_n = P(O_t \mid (O_{t-1}, O_{t-2}, \cdots, O_{t-M}), n) \tag{3}$$

E_{t-1} 表示在与当前节点相关的前一个节点情感矢量。$\delta(t)$ 表示当前的情感激励，若当前节

点为情感关键词是，则输出一个新的情感激励源 1；若当前节点为非情感关键词时，则通过情感生成概率函数，确定当前的情感激励状态，其中 O_t 为节点 t 的词类标注。α 表示情感抑制系数，用以表示情感激烈程度。C_t 表示情感矢量的修正值，用以人为修正当前节点中的情感矢量值，主要是针对一些特殊的情感用语需要做一些调整。在正常情况下，情感矢量 E_t 在失去持续激励的条件下，经过一定的节点，将逐渐衰减为零。

通常情况下，在 ESiN 网络的终止节点中，情感矢量只有一个分量不为零。这表示，在一般情况下的情感状态通常由主要的情感关键词和标注信息等决定。但在，有些场合，会出现情感状态歧异的情况，这主要由情感关键词、标注本身出现多意性导致。有时句型结构和词性组合也能导致这种歧异。情感触发器的目的，是将多个不为零的情感矢量分量根据其节点的历史记录进行综合，得到确定的到当前节点为止的情感状态。

$$M_t = \arg\max_n (\sum_{i=0}^{t} e_{n,i}) \tag{4}$$

情感状态确定过程如图 2 所示：

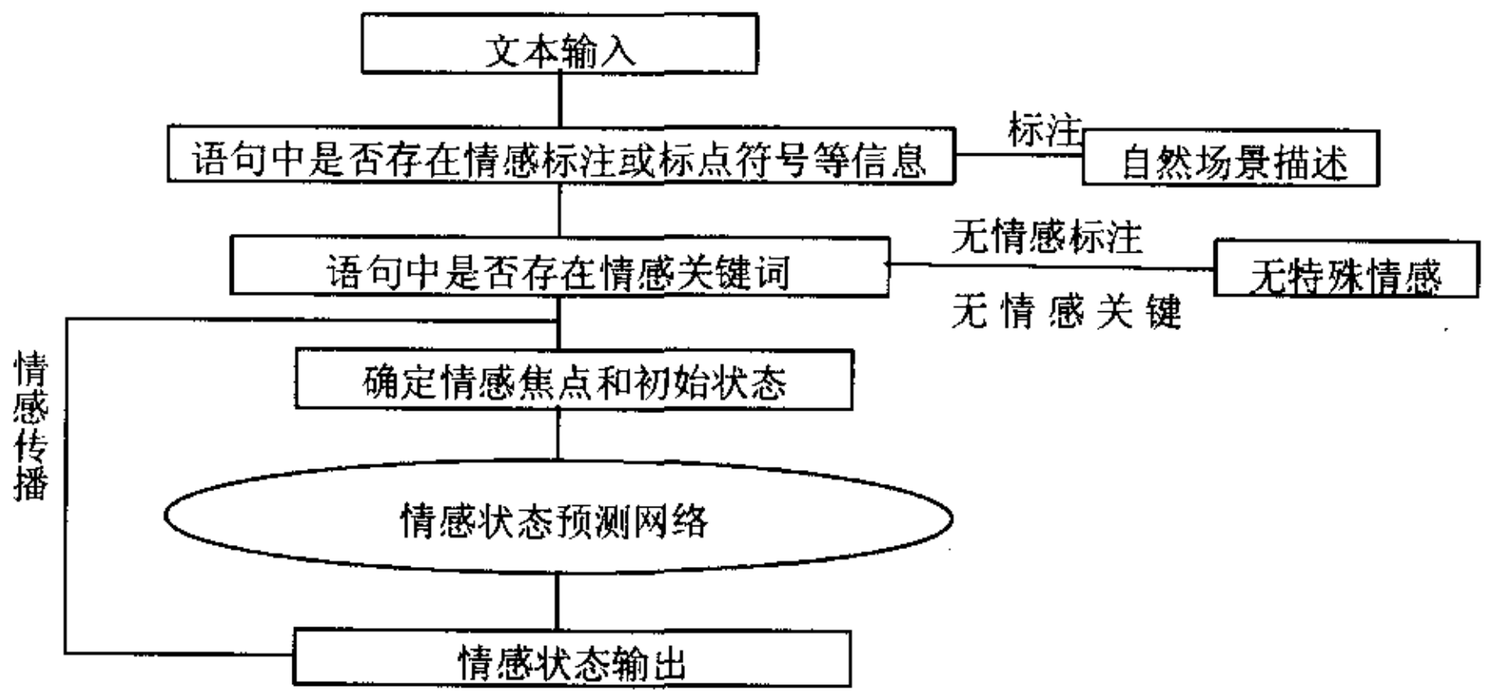

图 2：情感状态确定流程

情感生成概率，是一种情感转移过程中的规则，在现有的情况下，它反映了语法信息到情感状态转移特性的一种折射。通过以下几个准则确定：

1）可区分性。必须使语义上不具有包含关系的语义范畴用不同的类别进行标注。而对语义上具有包含关系的语义范畴，则可用几种类别同时进行标注。

2）完整性。表示一个完整的语义类的规则要尽量分为一类。主要是为了保证所有该项语义类能够完全地被表示，而不会出现某些语句组成模式由于不在类别标注之内而不能被正确确定。

3）简单性。考虑到要进行动态分类，活动规则集如果过于复杂，由于语法信息在标注时的歧义性，反而导致效果下降。应尽量避免将无序型、长程型及交叉型规则列入标注。

4）粒度选择。活动规则集不能起到描述整句的功能，粒度不能太小，从而起不到对情感分类路径的有效约束。

4.3 自然场景的驱动作用

研究情感发音，不能离开外界因素对人发音的影响。情境,指的是与某人的主观世界直接相关的那部分特定环境。古人谈及创作诗文的妙诀之一就是做到“情景合一”,即主观感情与某情境二者天衣无缝的结合。场景对情感特征的影响，则由教育工作者和一些社会学研究的

工作者进行了较早和较详细的分析，他们从感性的世界观理解不同场景，分析人的意境与场景的融合，并阐述了表达方式的异同，以社会学视角分析问题，虽然不能直接用于计算机的情感韵律建模，但为这一研究提供了较好的参考。一般意义上，影响我们情绪的环境可以分为三类：强敏感场景因子，弱敏感场景因子和非敏感场景因子。在对话语音中，面部表情、姿态、举止等通常会引起说话人迅速的反应。至于背景音乐、周围环境的色调，它们通常不是很强烈的影响说话人的情绪能力，从而构成弱敏感环境因子。这一划分并不是绝对的，它们随着具体的情况，影响情绪的能力会出现变化。

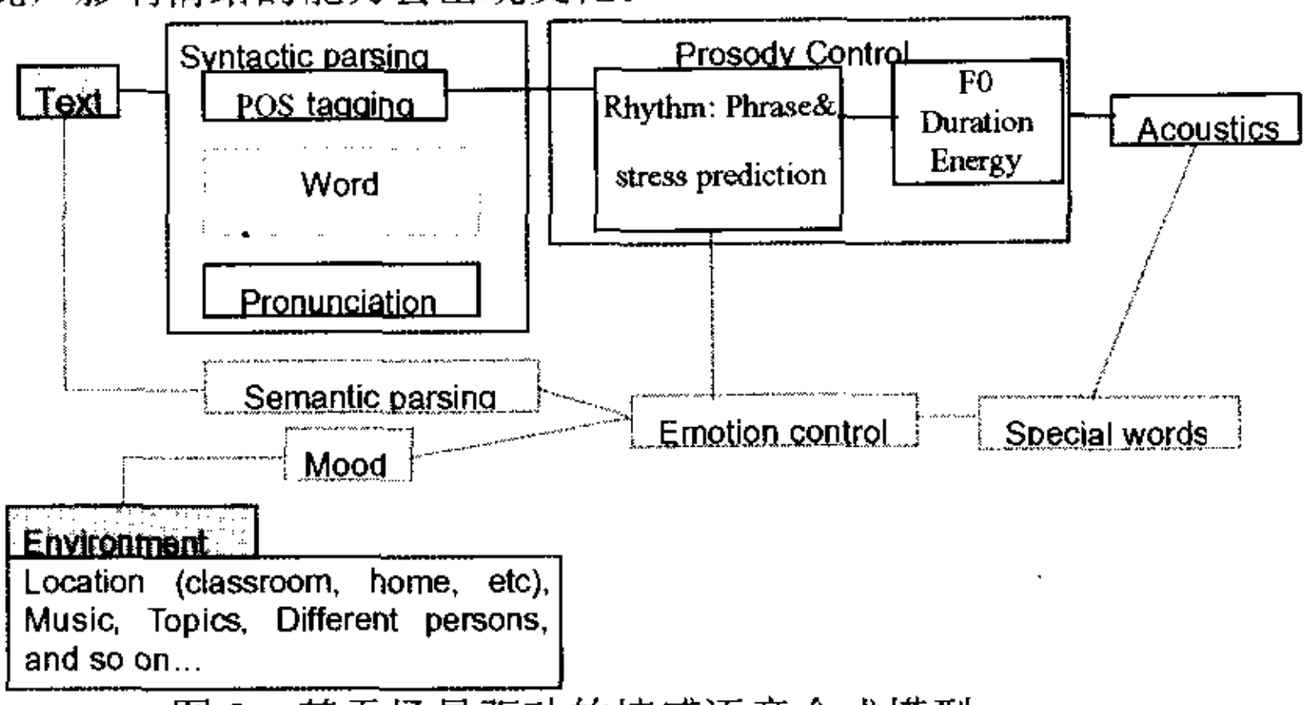

图 3：基于场景驱动的情感语音合成模型

与传统的由三分模型（文本分析、韵律模型和声学模型）组成的语音合成系统相比，完整的情感语音合成系统需要引入了环境感知和情感控制模型，如图 2 所示。为了使系统更适应环境，我们定义了一种情感描述语言 ECML 1.0，它融合了 SABLE 和 EML 标注语言，是一种能灵活使用情感命令的工具。用户可以定义情感状态并使之按照用户的定义工作。目前，系统原型已综合了表情分析和表现以产生多模态的接口，并融合了部分环境参数的背景音乐和场景信息。

5 情感语音合成结果的辨识实验

与通常的语音合成的结果相比，情感使得系统更具有表现力。但是，由于情感与人的感知密切相连，是不是所有的情感都能够在语音合成系统有效地表达？为了验证语音合成结果中负载情感的质量，本文通过 250 个学生来进行判断。这些学生来自小学、中学和大学等不同年龄阶段，年龄分布为 6 到 25 岁。在没有任何的系统设计经验的情况下，他们被要求写出所有听到的语音合成结果的情感状态。最后，其结果采用如下公式统计。

$$e = {M'_n}/{M_n} \tag{5}$$

这里，e 是正确率，M'_n是情感状态 n 中被正确判断的数目，M_n是情感状态 n 的总数目。

表 3 为得到的结果：

年龄(岁)	发怒	喜悦	悲伤	恐惧	厌恶
6-12 （小学）	89%	86%	56%	72%	50%
12-18（中学）	85%	78%	85%	56%	68%
18-25（大学）	55%	73%	72%	60%	75%

表 3，具有不同年龄和教育背景的听者对合成结果的评估

从表 3 中，我们发现通过系统实现的大多数情感语音合成结果在大部分情况下，能够被认同。表中的数据似乎反映年龄和教育背景是影响情感感知的一个因素，尽管这还需进一步

的试验证实。由于情感状态的定义在不同的人群和文化中是不明确的，进一步有关情感感知的实验还有待继续。

6 结论及展望

本文分析了汉语的情感语音合成中涉及的诸多关键技术，针对情感状态下的声学特征分析、情感焦点、情感预测网络以及情感与环境的关系等问题进行较为详细的阐述。基于以上分析，建立了一个新的环境感知的情感语音合成系统原型。大量的研究表明人们的情感受到人们的认知以及所表达的概念的很大影响。而通过机器学习的方法，实现对人的认知的完全理解，还需要一个较长的过程。进一步的工作，将通过深入分析情感语音的声源和声道特性、分析多种场景下情感语音的几种主要表现形式，并融合多模式的人机交互环境，以得到更为丰富的交互信息，从而实现更为人性化和个性化的语音合成系统。

参 考 文 献

[1] J. Cahn, " Generating Expression in Synthesized Speech," Master's thesis, MIT, 1989.

[2] Keikichi Hirose, Nobuaki Minematsu, etc, "Analytical and perceptual study on the role of acoustic features in *realizing emotional speech*", *ICSLP2000*

[3] Donna Erickson,1 Arthur Abramson, "Articulatory characteristics of emotional utterances in spoken English", ICSLP2000

[4] Schrö der M1, Cowie R2, Douglas-Cowie E2, "Acoustic Correlates of Emotion Dimensions in View of Speech Synthesis", Eurospeech2001

[5] Schrö der M, "Emotional Speech Synthesis: A Review", Eurospeech2001

[6] Kazuhito Koike, Hirotaka Suzuki, Hiroaki SAITO, " Prosodic Parameters in Emotional Speech", ICSLP98

[7] J.M. Montero, J. Gutiérrez-Arriola, etc, "Emotional speech synthesis: from speech database to tts", ICSLP98

[8] Cécile Pereira, Catherine Watson, "Some acoustic characteristics of emotion", ICSLP98

[9] Oatley, K. (1987). Cognitive science and the understanding of emotions. Cognition and Emotion. 3(1), 209-216.

[10]Murray, I. R.; Arnott, J.L.: 1992, 'Towards the simulation of emotion in synthetic speech: A review of the literature on human vocal emotion'. Journal of the acoustic society of America, *93*, 1097-1108.

[11]*Williams, C.E.; Stevens, K.N.: 1981, 'Vocal correlates of emotional states'.* In: J.K.Darby (eds.): Speech evaluation in psychiatry , New York, Grune & Stratton, pp. 221-240.

[12]Banse, R. and Scherer, K. R., Acoustic Profiles in Vocal Emotion Expression, Journal of Personality and Social Psychology, 70(3):614-636, 1996

[13]Jianhua Tao, Emotion Control of Chinese Speech Synthesis in Natural Environment, Eurospeech2003, Genever

[14]Marc Schrö der*, Roddy Cowie+, Ellen Douglas-Cowie+, Machiel Westerdijk° & Stan Gielen° , Acoustic Correlates of Emotion Dimensions in View of Speech Synthesis, Eurospeech2001, Scandinavia, Sep, 2001

[15]Noam Amir, Classifying emotions in speech: a comparison of methods, Eurospeech2001, Sep. 2001

[16]Jiahong Yuan, Liqin Shen, Fangxin Chen, The acoustic realization of anger, fear, joy and sadness in Chinese, ICSLP2002, Denver, Sep.2001