情感语音的韵律特征分析与转换

熊文婷，崔丹丹，孟凡博，蔡莲红

摘要 本文分别以整句和句内音段（句首、句中、句末）为单位对情感语音的六种典型韵律特征进行了分析，研究了当语音信号由中性变换成愤怒、高兴、惊奇和悲伤情感时，其声学特征的变化特点。根据分析结果，进一步地选择时长、能量、最大基频、基频范围这四个韵律特征，进行了由中性语音到情感语音的转换实验，并对全局和局部分音段两种转换模型的转换效果进行实验比较，也对局部分音段情感转换模型的有效性进行了主观评测。实验结果表明：情感语音的韵律特征对于不同情感，语句内不同音段，其变化不尽相同。这种音段间韵律特征的局部差异对情感的表达感知有贡献，并可用于指导情感转换。

关键词：情感语音 韵律特征 局部 情感转换

1 引言

语音是人们在日常生活中重要交流手段，它不仅能够表达文字所含的语义信息，还可以通过说话者的说话方式，如语气、音调变化等表达出说话者的态度和情绪，即我们常说的“言外之意”[4]。情感语音的研究已成为语音研究的热点。

情感语音的研究表明，语音情感信息主要体现在韵律特征的变化上[1] [2] [4]。有的进一步指出，基本情感的声学特征差异，主要反映在基频的高低，能量的增减和语速的快慢[4]。运用少量的韵律特征，可以较为有效地识别悲伤，高兴，愤怒，惊奇等情感[5][8]，并且通过修改韵律参数，可以在一定程度上表现出情感[5]。但遗憾的是，对于高兴，愤怒，惊奇等高激活的情感区分，一直较为困难。

已有的研究，大都是以整个情感句为单位分析语音信号的韵律特征[6] [8]。但语音特征具有时变性：时长的伸缩，音调升降在句中不同位置具有不同的变化规律[3]。情感的表现和感知也会随音段的位置改变。因此，有必要研究句内局部音段的韵律特征和音段间的韵律特征变化特点。这对发掘情感语音声学表达的规律，实现更高质量的情感语音转换，是非常有意义的。

为了研究韵律特征在语句内部的局部差异对情感表达的影响，本文设计了文本对齐的无意义短句，并用五种情感录制了情感句，分别从整句话和局部音段两个角度对六个情感韵律特征进行统计分析与对比，研究了不同情感下韵律特征的变化特点，总结出基于全局（整句话）韵律特征和局部（句内音段）韵律特征的两种情感转换模型。实验评测结果表明，局部分音段情感转换模型在惊奇和愤怒情感方面效果明显。

2 语料与特征

本文研究韵律特征参数的时域变化对情感表达的贡献，因此分析中应尽可能的消除文本语义对情感韵律特征的影

* 本文研究工作受国家自然科学基金资助（60433030）

响，并尽量确保用于分析的语音同种情感表达方式前后一致。先以短句为单位对录入的情感语音的平均基频、最大基频、最小基频、基频变化范围、平均能量和平均时长进行统计分析；再将短句分为句首、句中、句末三个音段，并以音段为单位对以上 6 个声学特征进行统计，得到情感语音句内不同位置音段的局部韵律特征。

2.1 分析用语料

本文设计的语料是文本对齐的短句（五字--六字），共 48 个。为了避免情感表达方式受文本内容、音调的影响，文本涵盖所有单、双音节声调组合的 24 种情况，每种组合在句首、句中、句末各出现一次。每个短句均为 5 个或 6 个音节组成。为了研究句语音韵律结构的局部特点，根据词的位置将每个短句分成句首、句中、句末三个音段，每段为 1-2 个音节。

录音者用中性、愤怒、高兴、惊奇、悲伤五种情感共录制了 240 句情感句。虽然一种情感可以有多种表达方式，但在录音时要求录音者对同一种情感保持相同的表达方式。

然后，对采集的数据进行评估，选取情感表达明显的样本，进行语句切分；再利用分析工具进行音节切分、基频提取；最后手工进行修正，剔除野点，进行平滑，并用四次曲线对基频曲线进行拟合。

2.2 韵律特征提取

本文分析研究的特征是短时能量、时长、平均基频、最大基频、最小基频、基频范围。短时能量计算的帧长 20ms，帧移 10ms，以 db 为单位。为排除噪音干扰，只计算句中能量过某特定阈值的那些音节。时长计算的是从语句第一个音节的开始到最后一个音节结束的时间间隔，包括了音节之间的停顿部分。其它四个参数可从标注数据中得到。

为便于对比情感语音与中性语音韵律特征差异，将同文本语句的非中性情感的参数与中性情感的参数相比，得到归一化的表示。

例如愤怒的全局基频均值 $\overline{F0}_{愤怒}$ 的计算公式如下，其中 $\overline{F0}_{愤怒}(x)$ 表示的是愤怒情感中第 x 句的基频均值：

$$\overline{F0}_{愤怒} = \frac{\sum_{1}^{48}\overline{F0}_{愤怒}(x)}{\sum_{1}^{48}\overline{F0}_{中性}(x)}$$

局部分析局内音段的韵律特征时，则是将将句子分为句首、句中、句末三个部分，依次统计这三个音段的愤怒、高兴、惊奇、悲伤四种情感下的局部韵律特征（相对值）。局部统计时，每个位置（句首、句中、句末）各有 48 个音段。

3 韵律特征分析对比

3.1 全局韵律特征分析

首先，考察统计的 6 种全局韵律特征在语音由中性转为愤怒、高兴、惊奇、愤怒时的变化区别（图 1）

从图 1 看出，悲伤较中性时有所延长，而其它三种的语速均有提高。在能量方面，悲伤的能量明显小于其他三种情感。通过全局特征可以容易区别悲伤情感。愤怒、高兴和惊奇明显提高了基频均值。同时该图也反映出，单从全局的角度，以语音的韵律特征来区分激活度较高的

愤怒、高兴和惊奇比较困难。特别是当这三种情感的时长、最大基频、最小基频、基频范围的统计方差相对较大时，更是对这三种情感的区分造成困难（表1）。

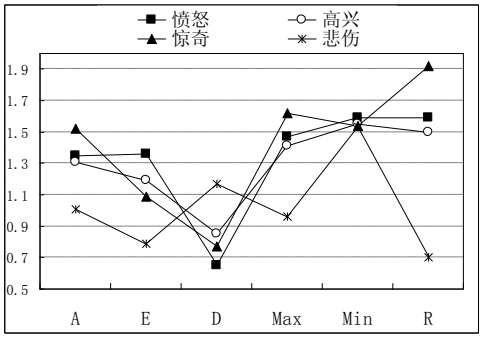

图 1 不同情感的全局韵律特征变化对比

A: 平均基频；E 平均短时能量；D: 时长；Max: 最大基频；Min: 最小基频；R: 基频范围

表 1 全局韵律特征变化的平均值与方差

μ：归一化均值；σ：标准差

		愤怒	高兴	惊奇	**悲伤**
基频均值	μ	1.36	1.41	1.50	**1.03**
	σ	0.15	0.11	0.11	**0.06**
平均能量	μ	1.29	1.21	1.08	**0.77**
	σ	0.02	0.01	0.01	**0.01**
时长	μ	0.59	0.82	0.82	**1.11**
	σ	0.07	0.07	0.07	**0.11**
最大基频	μ	1.36	1.53	1.63	**0.94**
	σ	0.24	0.18	0.19	**0.10**
最小基频	μ	1.46	1.47	1.39	1.44
	σ	0.72	0.64	0.76	0.68
基频范围	μ	1.29	1.57	1.80	**0.57**
	σ	0.59	0.49	0.61	**0.25**

音节的韵律特征与其所在的韵律结构位置相关，即其特征会随音节所在的韵律词内、韵律短语、语调短语的位置变化。为了进一步的研究不同情感，特别是愤怒、高兴与惊奇之间的韵律特征的差异，本文将一句话分为句首、句中、句末三个部分，分别考察语音在情感从中性情感变至愤怒、高兴、惊奇、悲伤时的局部变化情况。

3.2 句首/句中韵律特征分析

统计句首音段的韵律特征（图 2）。结果表明，愤怒的能量明显高于高兴和惊奇；高兴的时长大于愤怒和惊奇，在频率范围上则比他们要窄得多。

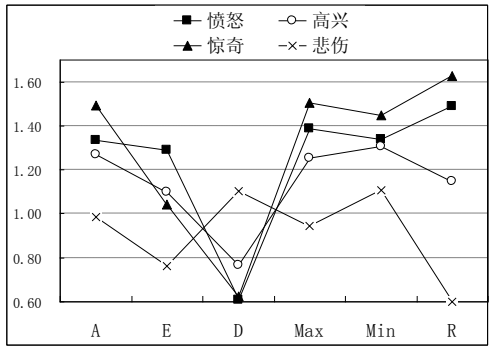

图 2 不同情感的句首韵律特征对比

统计句中音段的韵律特征。我们发现：惊奇在基频最大、最小和基频均值上是可以与愤怒和高兴区分开的，而愤怒和高兴的这三个特征参数值则十分接近。高兴在句中部分的时长要比愤怒和惊奇稍长。

3.3 句末韵律特征分析

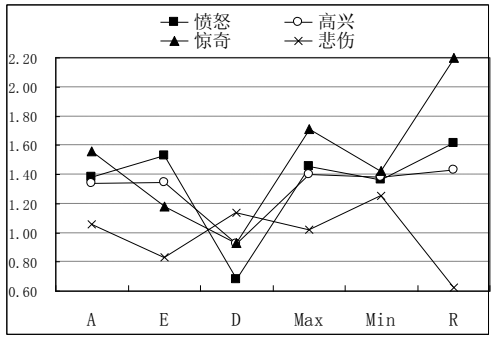

图 3 不同情感的句末韵律特征变化对比

统计句末音段的韵律特征（图 3）。结果表明，在句末部分，惊奇的基频范围的变化幅度明显增大，增幅是愤怒的两倍，是中性情感下的 2.2 倍。同时惊奇在最大基频方面进一步的增大，已能较为明显地和愤怒、高兴区别开来。此外，愤怒在句末音段的时长明显短于愤怒和高兴，是愤怒区别于其他情感的又一特点。

3.4 综合分析对比

综合 3.2-3.3 的局部分析结果，激活度较高的愤怒，惊奇的区分能力已有较大改善。愤怒情感的平均短时能量在句首、句中、句末均为最大，并且其句末的时长最为短促；惊奇的句末的基频范围明显大于其他情感，句末音调提高，基频范围的显著增宽也是惊奇的特点之一。相较于愤怒和惊奇，高兴的局部特点并不十分显著，只是比愤怒、惊奇在句首、句中的语速更加缓慢，句首部分的基频范围更窄。

另一方面，横向比较这三种情感可以发现，句末音段中不同情感的韵律特征差异最为突出。这表明句末部分的说话方式是情感表达的关键之一，从而也印证了句内分音段研究角度的有效性。

除此之外，还可以看出局部分析对区别愤怒、高兴、惊奇有较大贡献的主要是平均短时能量、最大基频、基频范围这几个特征。

4 转换实验及结果分析

根据以上分析，选取与情感表达感知最为密切相关的四种特征——平均短时能量、时长、最大基频、最小基频——分别从全局和局部两个角度对语音信号由自然中性语音到愤怒、高兴、惊奇悲伤的情感语音的转换过程进行建模，得到全局和局部两种基于韵律特征的情感转换模型。例如，局部转换模型的句末段部分如下：

表 2 局部转换模型——句末段

句末	愤怒	高兴	惊奇	悲伤
平均能量	1.22	1.13	1.02	0.76
时长	0.56	0.74	0.65	1.03
最大基频	1.28	1.37	1.52	0.95
基频范围	1.19	1.37	1.63	0.55

为了验证上文的分析结果，证明局部音段韵律特征对情感表达和感知的影响，进行了偏向性测试，比较了上述全局情感转换模型和基于句内分音段（句首、句中、句末）的局部模型的转换效果，然后对局部模式的情感转换结果进行了主观测试并给予评价。

4.1 实验语料

一共录制了 10 组集外测试用句，每组有四个不同的情感句。各组中每句话都可分成两个部分：含有情感倾向的不同文本的半句和不含情感倾向的相同文本半句。录音者先用不带情感的中性语音朗读四种情感句，然后再用相应的情感语音朗读一遍。随机选择其中两组进行本次测试。

4.2 转换方法

为了给测试者提供一个具体情感语境，令其更好的评价情感转换结果的有效性，分别截取两个半句，对其用全局模型和局部模型进行中性到四种情感的转换。再将转换结果和该句中未转换部分的录入情感语音进行拼接。为了避免录入的情感语音段给测试者造成听音干扰，对每组句测试语句都给出了被转换

内容的中性录入语音作为参照。

因此，四种情感各有测试句 4 组，每组包括 1 个参考句，2 个转换结果句。共 16 组 48 句。参与测试的 10 位同学分别就每组中的两种转换结果进行比较，选出更加符合相应情感语境的转换结果。另外，也请这些同学单独对基于分音段的局部情感转换模型的转换效果进行具体评价，评测语料为 16 句，测试时提供情感转换句的同文本自然中性语音作为参考。

4.3 偏向性测试结果

统计每种情感中，测试者认为较好的转换模型的分布情况，结果如表 3 所示。

表 3 偏向性测试结果

	1	2	3	4	5
愤怒	0%	5%	32.5%	57.5%	5%
高兴	0%	35%	50%	15%	0%
惊奇	2.5%	25%	27.5%	35%	10%
悲伤	0%	12.5%	52.5%	35%	0%

可以看出，大多数人（77.5%）认为惊奇情感的转换使用局部模型明显优于全局模型。在高兴及愤怒上，局部略有优势。而对于悲伤情感，更多人认为全局模式的转换效果要好于局部模式。测试结果基本符合分析结果。

4.4 局部模型主观评价结果

对分音段的局部情感转换模型的转换结果的进行主观评价。评价标准共分 5 级：1—没有情感信息， 2—有正向的情感信息，但目标情感不清楚，3—稍有目标情感，4—能较好的表达目标情感，但可以听出与自然情感语音的差别，5—非常好的表达了目标情感，非常自然。

统计出每种情感大家评价等级的分布，用百分数形式表示，如表 4 所示。

表 4 分音段局部情感转换模型的主观评测结果

	全局更优	局部更优	相当
愤怒	12.5%	17.5%	70%
高兴	22.5%	30%	47.5%
惊奇	15%	77.5%	7.5%
悲伤	35%	10%	55%

对局部情感转换模型的转换结果的主观评测表明：局部情感转换模型能较好的实现由中性语音到目标情感的转换。其中，愤怒转换效果最好，其次是惊奇和悲伤，最后是高兴。

4.4 实验结果分析

综合两项主观测试的结果，可以得到以下启示。

对于悲伤情感，采用全局韵律特征已能很好地与其它情感区分，但全局模型的基频低于局部模型。

对于高兴情感，两种转换模型的区分度不大，并且转换的结果语音在听感上虽然有较强的激活性，但与理想的愉悦性还存在一定距离。

对于惊奇，测试结果显示，韵律特征参数的局部变化，特别是句末音段的变化对该情感的表达有明显贡献。

5 结语

本文验证了语音的韵律特征承载着主要情感信息这一结论，并进一步地通过将语句分为句首、句中、句末三个音段，以音段为单位对情感语音的局部韵律特征进行了研究。研究发现：时长、平均短时能量、最大基频和基频范围这四个韵律特征的局部（尤其是句末）的变化对语音情感的表达与感知有着显著贡献，并能够在一定程度上区分激活程

度相近的愤怒、高兴和惊奇 3 种情感。

通过对基于全局情感转换模型和局部转换模型的中性到情感语音的转换实验，证明了上述结论。

不可否认，情感的语音表达具有多样性。但是，由于研究是基于排除语义、声调影响的语料进行的，所以分析得到的情感表现方式具有一定的普遍意义。这为进一步描述语音情感，及建立相应的语音情感转换模型提供了有价值的参考。

当语义本身具有情感倾向时，或者情感焦点改变时，都将会对本文研究的韵律特征产生影响，其规律也将更为复杂。而激活度较高的愤怒、高兴、惊奇情感的特征差异也还需要进一步的研究。我们愿和语音学界的各位同行共同努力，不断探索。

参考文献

[1] Sobin C.; Alpert M., Emotion Speech: The Acoustic Attributes of Fear, Anger, Sadness and Joy, Journal of Psycholinguistic Research, Volume28, Number 4, July 1999 , 347-365(19)

[2] Dimitra Vergyri, Andreas Stolcke, Venkata R. R. Gadde, Luciana Ferrer, Elizabeth Shriberg, Prosodic Knowledge sources for automatic speech recognition, ICASSP 2003

[3] 曹剑芬，“普通话节奏的声学语音学特性”，第四届全国现代语音学学术会议论文集, 1999

[4] 蒋丹宁，“情感语音的声学特征分析及建模”，清华大学，2005

[5] 崔丹丹，“情感语音分析与变换的研究” 清华大学”，2007，

[6] 赵力， 将春辉， 邹采容，吴镇洋，“语音信号中的情感特征分析和识别的研究”，电子学报, 2004, 32(4) , 606-609

[7] 韩纪庆，邵艳秋，“基于语音信号的情感处理研究发展”，电声技术，2006, 05, 58-62

[8] 姜晓庆，田岚，崔国辉，“多语种情感语音的韵律特征分析和情感识别研究”，声学学报, 2006, 31(3), 217-221

(熊文婷 北京邮电大学 100876
wentingxiong@gmail.com 本工作在清华大学人机语音交互实验室完成,
崔丹丹、孟凡博、蔡莲红 清华大学
100084 clh-dcs@tsinghua.edu.cn)

Analysis of Prosodic Features of Emotional Speech and the Experiment of Conversion

This paper analyzed the global and local prosodic features of emotional speech. In addition to the traditional unit of whole sentence, the segments at different place (head, body, tail) within it are considered for the first time. The acoustic differences of speech signal from neutral to anger, surprise, joy and sadness were investigated by statistical analysis.

With this result, duration, energy, maximum pitch and pitch range were used in following conversion experiments, based on global and local models respectively. The results of two scopes were compared and a subjective test was also scheduled to evaluate the local conversion model. The result shows that the local prosodic variations are different for different emotion and such variations contribute to the expression and perception of emotional speech