英语焦点重音声学参数分析与建模

孟凡博、蒙美玲、蔡莲红

摘要： 本文针对英语焦点重音表达的特点，提出了一种从中性语音到含有焦点语音的转换方法。数据分析表明，中性语音到焦点语音各音节声学特征的变化与该音节与焦点的相对位置有关。因此以音节为单位，本文根据音节与焦点的相对位置，将焦点语音的音节分成 7 类，分析了当一句中性语音变换成带有焦点重音的语音时，基频、时长和能量等声学特征的变化。在此基础上，建立了由中性语音到含有焦点重音语音的转换模型。本文将转换模型应用于中性语音，并对其转换结果进行了主观评测实验。实验结果表明：该转换模型较好的表达了焦点重音，达到了提高了英文语音表现力的目的。

关键词: 焦点、基频、时长、能量

1. 引言

语音是人们在日常生活中重要的交流手段，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、音调变化等表达出焦点等其他含义。焦点是句法语义学的一个概念，根据广泛认可的焦点-重音理论（Focus-to-Accent），在音高重音语音中，成为焦点的词或成分会以音高重音的形式在口语中表现出来，即形成焦点重音[1][2][3]。而焦点重音的研究已经成为了语音研究的热点之一。

现有研究表明，基频、时长的变化，对于焦点重音的表达感知具有重要贡献[2][4][5]，焦点字（词）的基频、能量会升高，部分语种的焦点词之后词的基频、能量被降低[5]。Costa 在他的论文中比较了焦点重音与中性语音中，元音及辅音基频和时长的差异，并且得出，“high vowel”的时长比“low vowel”短，而基频比“low vowel”高。另一位学者根据距离重音距离的不同，对音节单元的时长进行分析，分析结果表明，越靠近重音，时长越长[6]。

已有的研究，大都是对中性语音与焦点语音焦点词的声学特征分析，并且是以词[5]或音素为单位[4]，缺乏对英语多音节词以及重读音节等先验知识的使用，并且缺乏对转换建模的研究。而研究表明，距离重音距离的位置对时长的变化具有重要影响[6]，相应的，焦点的表现和感知也会随声学特征的变化而改变。因此，有必要根据音节与焦点单词及重读音节相对位置的不同，分析研究声学特征变化的特点。这对发掘焦点重音声学表达规律，实现高质量焦点语音转换，提高英文语音表现力，是非常有意义的。

为了研究声学特征在相对焦点不同位置的变化对焦点感知的影响，本文设计了 23 句分析用语料文本，并分别录制了中性及含有焦点的语音。并借鉴[6]的分类思想，根据相对焦点位置的关系，将录音的音节分成 7 类，然后对每类音节的 7 个声学特征进行了统计分析与对比，研究了相对于中性语音，焦点语音各类音节的变化特点。然后根据分析结果，从声学特征中选取 5 个建立了由中性到焦点语音的转换模型，并进行了主观评测实验。实验结果表明，该转换模型较好的表达了焦点重音，达到了提高英文语音表现力的目的。

2. 分析建模语料库

英文每字可以由多个音节组成，并且重读音节的位置也有变化。因此，为了分析研

究由中性语音到焦点语音声学特征的变化规律，需要针对英语设计恰当的语料文本，再进行录音，提取特征和数据分析。

2.1 语料文本

本文的研究对象是英语中性语音到焦点语音声学特征的变化，英语每字可能由多个音节组成，而重读音节也可以在不同的位置。针对这个特点，本文设计了 23 句语料文本。其中句式包括了陈述句、疑问句以及选择疑问句，焦点单词由 3 位标注者标注，结果具有很好的一致性。焦点单词在每句文本中的句首、句中和句末等不同位置，焦点单词包括了单音节词、双音节词以及多音节词，并且重读音节也在单词中的不同位置。

2.2 语料录制

为了研究中性语音到焦点语音的声学特征的变化，本文为邀请了一位资深英语女性发音人为 2.1 节中的每句文本分别录制了中性语音与焦点语音，要求在录制中性语音时发音平白不带语调，在录制焦点语音时发音突出焦点，如果某句话录音不符合要求则重新录制。共录制 2 遍，一共 92 句语音。每句录音以 Microsoft Windows Wav 格式保存（单声道，16 位，采样率 16kHz）。每句录音进行了手工音节切分和基频标注，并且在数据分析之前，对基频曲线进行了中值平滑。

3. 焦点重音声学特征分析

为了分析从中性语音到焦点语音声学特征的差异，根据相对于焦点的位置，本章以音节为单位，将语音的音节分成 7 类，然后分析了每一类的 7 个声学特征的变化（基频最大值，基频最小值，基频范围，基频平均值，基频斜率，能量，时长）。

3.1 音节分类规则

已有研究表明声学特征的变化规律与重音的相对位置有关[6]，因此，根据相对于焦点单词以及其重读音节的位置，本文以音节为单位，将语音的音节分成 7 类：对于焦点单词，1-重读音节，2-重读音节的前一音节，3-重读音节的后一音节，4-其他音节；对于非焦点单词的音节，5-焦点单词的前一音节，6-焦点单词的后一音节，7-其他音节。图 1 是本文音节分类的举例，其中单词 California 是焦点音节。

Teresa is flying to California tomorrow.
7 5 4 2 1 3 6 7

图 1：音节分类示意图

3.2 声学特征的选取

本文的研究对象为中性语音到焦点语音的声学特征的变化，现有研究表明，与焦点有关的声学特征主要有基频，时长，能量三个特征，因此，本文选择以下 7 个参数进行分析：

基频最大值（Max，Hz），基频范围范围（R，Hz），基频最小值（Min，Hz），基频平均值平均值（Mean，Hz），基频斜率绝对值（S，Hz/ms），短时能量（E，dB），平均时长（D，ms）。

以音节为单位，本文分别从中性语音以及焦点语音中提取这 7 个声学特征，然后计算同文本的各特征焦点语音与中性语音的变化比例，最后进行统计分析。

3.3 焦点重音声学特征分析

表 1 显示了第 1 类（焦点单词重读音节，例如图 1 中的 for）音节由中性语音到焦点语音声学特征的变化。与中性语音相比，焦点语音的基频最大值有显著的提高，但是基频最小值无明显变化。焦点语音的基频斜率有很大的变化，是中性语音的 3.8 倍，并且能量增加，时长加长。并且在分析数据时发现，焦点语音的所有 1 类音节的基频斜率都是正的。这说明，对于 1 类音节，发音人提高了基频，但是维持最小值不变，因此斜率为正变大以及基频最大值升高，同时音量增大，语速变慢，以突出焦点。

表 2 显示了第 2 类（焦点单词重读音节前一音节，例如图 1 中的 li）音节由中性语音到焦点语音声学特征的变化。这一部分基频和能量变化不大，但是时长缩短，基频斜率绝对值变小，这是因为，这一类音节往往

是轻读音节，例如单词“apartment”中的“a”。

表 1: 焦点单词重读音节声学特征变化

	Max	R	Min	Mean	S	E	D
中性	241	53	187	212	134	52	146
焦点	326	137	188	268	519	57	187
Ratio (%)	135	257	100	126	385	110	128

表 2: 焦点单词重读音节前一音节声学特征变化

特征	Max	R	Min	Mean	S	E	D
中性	234	54	179	207	261	49	123
焦点	246	58	187	214	233	50	119
Ratio (%)	105	106	104	103	89	103	97

表 3: 焦点单词重读音节后一音节声学特征变化

特征	Max	R	Min	Mean	S	E	D
中性	222	51	171	194	298	46	142
焦点	298	131	167	230	913	51	134
Ratio (%)	134	257	97	118	304	112	94

表 3 显示了第 3 类（焦点单词重读音节后一音节，例如图 1 中的 nia）音节由中性语音到焦点语音声学特征的变化。这部分音节的基频最大值变大（达到中性语音的 1.3 倍），这是由于基频曲线的连续性，焦点单词重读音节的基频斜率为正，在音节末达到最大值，因此，之后的音节的基频最大值也比较大。这一类音节的基频斜率为负，基频逐渐降低，并且基频最小值变小。

表 4 显示了第 4 类（焦点单词的其他音节，一般是 4 音节单词，例如图 1 中的 Ca）音节由中性语音到焦点语音声学特征的变化。焦点语音的这部分的基频最小值和时长与中性语音的相等，但是基频最大值略有升高，导致基频范围和基频斜率略有增加。

表 5 和表 6 显示了第 5、6 类（焦点单词前和焦点单词后一音节，例如图 1 中的 to）音节由中性语音到焦点语音声学特征的变化。这两部分的声学特征变化不大，区别在于，第 5 类音节的基频略微高于第 6 类音节的基频，这与人们说话音调逐渐降低是一致的。

表 7 显示了第 7 类（其他音节，例如图 1 中的 Teresa is flying 和 morrow）由中性语音到焦点语音声学特征的变化。这部分音节基频略有升高，时长缩短，但是基频斜率明显降低。

表 4: 焦点单词其它音节声学特征变化

特征	Max	R	Min	Mean	S	E	D
中性	209	38	171	189	139	47	158
焦点	247	73	174	206	254	48	147
Ratio (%)	118	192	102	108	183	101	93

表 5: 焦点单词前一音节声学特征变化

特征	Max	R	Min	Mean	S	E	D
中性	234	54	179	207	261	49	123
焦点	246	58	187	214	233	50	119
Ratio (%)	105	106	104	103	89	103	97

表 6: 焦点单词后一音节声学特征变化

特征	Max	R	Min	Mean	S	E	D
中性	222	55	167	198	403	49	138
焦点	241	81	159	193	534	49	128
Ratio (%)	108	148	95	97	132	99	92

表 7: 其它音节声学特征变化

特征	Max	R	Min	Mean	S	E	D
中性	297	159	138	220	207	50	119
焦点	328	181	146	228	145	51	116
Ratio (%)	110	114	105	103	70	103	97

为了验证根据与焦点单词及其重读音节的相对位置对音节进行分类的标准是否合理，本文计算了各维声学特征变化与音节类的相关性（表 8）。从中可以看到，基频最大值、基频均值、能量和时长均与音节类有较大的相关性，基频范围有较弱的相关性，基频最小值与音节类相关性不大，这与之前的数据分析中，各音节类的基频最小值变化不大相一致。此外，基频斜率与音节类无相关性，这是由于数据中基频斜率绝对值变化较大，规律性较弱。表 9 显示了焦点语音各音节类中基频斜率符号分布的百分比，从中

可以看到，对于含有焦点的语音，尤其在焦点单词重读音节附近，基频斜率符号具有很强的一致性。

以上分析表明，焦点单词音节的声学特征变化较大，焦点单词附近音节的声学特征有一定的变化，距离焦点单词越远，声学特征变化越小。基频在焦点单词重读音节逐渐升高，在下一音节逐渐降低。焦点单词重读音节的声学特征变化最为剧烈。

表 8: 声学特征变化与音节类的相关性

特征	Max	R	Min	Mean	S	E	D
相关性	-0.5	-0.3	0.1	-0.5	0.0	-0.5	-0.4

表 9: 焦点语音基频斜率符号分布

音节类	1	2	3	4	5	6	7
正(%)	87	91	17	65	34	30	17
负(%)	13	9	83	35	66	70	83

4. 焦点重音转换算法

第 3 章分析结果表明，中性语音到焦点语音声学特征的主要变化在焦点单词及其附件，并且焦点单词重读音节的声学特征具有最显著的变化。焦点单词的基频具有重读音节部分斜率为正，其后一音节斜率为负的特性。在第 3 章分析基础上，本章基于 TD-PSOLA 语音修改算法，建立了一个中性语音到焦点语音的转换模型。

4.1 焦点重音转换模型

数据分析表明，中性语音到焦点语音的基频范围方差较大，不适合用来建模，而对于刻画基频曲线的变化，基频最大值和基频最小值要优于基频平均值，虽然基频斜率绝对值的比值与音节类相关性较低，但是由于基频斜率是重要的刻画基频曲线调型的特征，以及对于同一音节类，基频斜率的符号具有很强的一致性，因此最终本文选择基频最大值、基频最小值、基频斜率、能量和时长进行建模，其中基频最大值、基频最小值、能量和时长为焦点语音相对于中性语音的比例，而基频斜率为绝对值。

本文从分析语料中随机选取 64 句作为训练语料，统计特征参数。表 10 为本文最后建立中性语音到情感语音各维声学特征的修改模型。它表示对于各类音节，相应的 5 维声学特征的修改幅度应该多少。

4.2 基于 TD-PSOLA 的焦点重音修改算法

TD-PSOLA[7]是一个修改基频和时长的语音修改算法，它通过调整峰值点位置修改基频、通过增加或删除基音周期修改时长。

对于本文的修改模型，基于 TD-PSOLA 实现从中性语音到焦点语音的修改算法分为 4 步：1）修改基频最大值、基频最小值和时长；2）修改基频斜率；3）根据目标基频曲线采用 TD-PSOLA 修改语音；4）修改能量。

表 10: 中性语音到焦点语音的预测模型

音节类	Max (%)	Min (%)	S (Hz)	E (%)	D (%)
1	135	100	519	110	128
2	112	111	236	103	113
3	134	97	913	112	94
4	115	99	265	102	87
5	105	104	233	103	97
6	108	95	534	99	92
7	110	105	145	103	97

1、设 $\mathbf{P}_i(n)$为中性语音第 i 音节（从 b_i 开始，到 e_i 结束）的基频序列，$\mathbf{D}_i(n)$为相应的基频点的时间序列，设 $P_{\mathrm{Min},i}$ 和 $P_{\mathrm{Max},i}$ 为该音节的基频最小值和最大值。设 R_{Max} 和 R_{Min} 为模型中基频最大值和最小值的比例，并且 R_{Duration} 为时长的变化比例。那么目标基频序列 $\mathrm{P}_i^{'}(\mathrm{n})$ 以及相应的时间 $\mathbf{D}_i^{'}(n)$ 序列应为：

$$P_{\mathrm{Min},i}^{'} = P_{\mathrm{Min},i} \times R_{\mathrm{Min}} \quad (1)$$

$$P_{\mathrm{Max},i}^{'} = P_{\mathrm{Max},i} \times R_{\mathrm{Max}} \quad (2)$$

$$k = \frac{P_{\mathrm{Max},i}^{'} - P_{\mathrm{Min},i}^{'}}{P_{\mathrm{Max},i} - P_{\mathrm{Min},i}} \quad (3)$$

$$\mathbf{P}_i^{'}(n) = P_{\mathrm{Min},i}^{'} + k \times \left(\mathbf{P}_i(n) - P_{\mathrm{Min},i}\right), n \grave{o} [b_i, e_i]$$
$$\mathbf{D}_i^{'}(n) = \mathbf{D}_i(n) \times R_{\mathrm{Duration}}, n \grave{o} [b_i, e_i] \quad (4)$$

2、首先采用最小二乘法对 $\mathbf{P}_i^{'}(n)$，$\mathbf{D}_i^{'}(n)$ 拟合得到直线 $\mathbf{f}_1(\bullet)$，设 $P_{\mathrm{slope}}^{'}$ 为模型预测的基频斜率，令 $\mathbf{f}_1(\bullet)$ 的中点为不动点，那

么可以得到目标基频曲线的拟合直线 $\mathbf{f}_2(\bullet)$，并且目标基频曲线 $\mathbf{P}_i^{*}(n)$ 计算方法如下：

$$\mathbf{f}_2(n) = P'_{\text{slope}} \times (n - \frac{e_i - b_i}{2}) + \mathbf{f}_1(\frac{e_i - b_i}{2}), n \in [b_i, e_i] \quad (5)$$

$$\mathbf{P}_i''(n) = \mathbf{P}_i'(n) \times \frac{\mathbf{f}_2(n)}{\mathbf{f}_1(n)}, n \in [b_i, e_i] \quad (6)$$

3、在有目标基频及时间序列之后，设原始中性语音的波形序列为 $\mathbf{S}_i(n)$，则修改后的语音波形序列 $\mathbf{S}_i'(n')$ 为：

$$\mathbf{S}_i'(n') = \mathrm{f}(\mathbf{S}_i(n), \mathbf{T}_i(n), \mathbf{T}_i'(n)) \quad (7)$$

$$n \in [b_i, e_i], n' \in [b_i', e_i']$$

其中 $\mathrm{f}(\bullet)$ 表示 TD-PSOLA 算法[7]。

4、设 $\mathrm{R}_{\text{Energy}}$ 为模型预测的能量比例，则在汉明窗 $\mathrm{W_i(n)}$ 平滑下对波形 $\mathrm{S_i'(n')}$ 调整，得到最终音节 i 的转换结果 $\mathrm{S_i^*(n)}$：

$$\mathbf{S}_i''(n) = \mathbf{S}_i'(n) R_{\text{Energy}} \mathbf{W}_i(n), n \in [b_i', e_i'] \quad (8)$$

其中汉明窗 $\mathrm{W_i(n)}$ 定义为：

$$\mathbf{W}_i(n) = 0.53836 - 0.46164 \cos\left(\frac{2\pi(n - b_i')}{e_i' - b_i'}\right) \quad (9)$$

$$n \in [b_i', e_i']$$

最后，N 个音节的转换结果拼接在一起构成目标语音：

$$\mathbf{S}''(n) = \{\mathbf{S}_1''(n), \ldots, \mathbf{S}_i''(n), \ldots, \mathbf{S}_N''(n)\} \quad (10)$$

5. 实验与讨论

本文在第四章建立了中性语音到焦点语音的转换模型以及基于 TD-PSOLA 算法的转换算法，本章将通过主观评测实验，验证转换模型和转换算法的质量。

5.1 实验语料与实验方法

本文从分析语料中随机选取 64 句作为训练语料建立焦点转换模型，剩余 28 句作为测试语料。通过焦点转换模型修改测试语料得到焦点语音转换结果。实验中，为每位听音人提供标注有焦点单词的文本以及相应的转换语音。

本文采取准确度和确信度作为主观实验的评测指标。对于准确度，听音人选择转换结果“是”或“否”评价转换结果是否正确表达了语音中的焦点信息。对于确信度，采用 MOS 评分的方法，听音人分为 5 个等级对准确度进行打分（5 非常确定，4 确定，3 可能，2 不确定，1 不知道）。

本文邀请 6 位听音人参加主观评测实验，分别统计选择是或否占总样本的百分比，以及确信度的平均值和置信度为 0.95 的置信区间。

5.2 实验结果与分析

基于主观评测结果，本文分别统计了准确度和确信度。表 11 显示了实验结果中，选择是否焦点准确表达出来的样本百分比。表 12 显示了确信度的平均值和置信区间。

表 11: 转换语音焦点是否表达出来的百分比

选项	是	否
百分比	97	3

表 12: 确信度的平均值及置信区间

确信度	平均值	置信区间
实验结果	4.5	0.0044

实验结果显示，本文提出的转换模型对于焦点的转换具有较高的准确度，并且听音人对于转换结果具有较高的确信度。其主要原因在于，本文根据音节与焦点相对位置将音节分成 7 类进行数据统计分析和建模，数据分析表明，由中性语音到焦点语音声学特征的变化与音节类具有较高的相关性，因此，本文提出的模型较准确较完备的刻画了由中性语音到焦点语音声学特征的变化。

6. 结语

本文以音节为单位，根据与焦点单词以及其重读音节的相对位置关系，将音节分成 7 类，分别统计分析了由中性语音到焦点语音各类音节基频最大值、基频范围、基频最小值、基频平均值、基频斜率、能量和时长的变化规律。分析结果表明，中性语音到焦点语音声学特征的变化与和焦点单词及其重读音节的相对位置具有较强的相关性。声学特征在焦点单词有较大的变化。

本文以数据统计分析为基础，建立了中性语音到焦点语音的转换模型，并通过TD-PSOLA 语音修改算法实现了转换算法。为了验证转换模型的有效性，本文设计了主观评测实验，实验结果表明模型的转换结果较好的表达了焦点，达到了提高英文语音表现力的目的。

本文的中性语音到焦点语音的转换模型还有许多可以改进的地方，例如在建模时引入句式、引入韵律层级等特征的影响，采用更合适的音节基频曲线构造参数来描述基频变化（例如，Pitch-Target），会达到更好的转换效果。

6. 致谢

该研究工作受国家自然科学基金（编号：60805008，60928005，60910130）的经费支持。

7. 参考文献

[1] 王韫佳、初敏、贺琳（2006）汉语焦点重音和语义重音分布的初步实验研究。《世界汉语教学》，第 2 期。

[2] Botinis, A., Fourakis, M., Gawronska, B. 1999. Focus identification in English, Greek and Swedish. *Proc. of The 14th ICPhS* San Francisco 1557-1560

[3] Rump, HH., Collier, R. 1996. Focus conditions and the prominence of pitch-accented syllables. *Proc. of Language and Speech*, Vol.39. 1-17

[4] Costa. 2004. Intrinsic Prosodic Properties of Stressed Vowels in European Portuguese. *Proc. of Speech Prosody*, 53-56

[5] Chen, S.-w., Wang, B., Xu Yi. 2009. Closely related languages, different ways of realizing focus. *Proc. of Interspeech*, 1007

[6] Barbosa, A., Arantes, P., Silveira, L.S. 2004. Unifying Stress Shift and Secondary Stress Phenomena with a Dynamical Systems Rhythm Rule. *Proc. of Speech Prosody*, 49-52

[7] Moulines, E., Charpentier, F. 1990. Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones. *Proc. of Speech Communication*, v.9 n.5-6 (1990) 453-467

孟凡博 清华大学计算机科学与技术系 100084

Professor Helen Meng CUHK Human-Computer Communications Laboratory

蔡莲红 清华大学计算机科学与技术系 100084