

## 两岸口语语流韵律比较初探

——以音强及音节时程分布为例

台湾中央研究院语言学研究所 郑秋豫 李岳凌 郑云卿

### 一 引言

中文口语语流最显著的特征为短语成组(phrase grouping),在口语篇章里最为明显(吴为章、田小琳,2002)。郑秋豫透过量化处理大批口语语料得到的证据,对这个现象提出“阶层式多短语语流韵律架构 HPG(Hierarchical Prosodic Phrase Grouping)”一说,每一层韵律单位对于最终的语流韵律表征都有贡献,而语流韵律则是各层级韵律单元贡献度的总和(Tseng et al, 2004, 2005)。这个架构最重要的特点,是强调并界定各短语间彼此的韵律关联性,而非各短语的孤立句调。她采用语料库语音学研究方法,从感知角度出发分析口语语流韵律,在声学语音信号(基频走势、音节时长、边界信息、边界停顿和停延段等)上,找到各韵律阶层贡献的证据及声学模板(template),这些模板的存在,是言语链里语言产生和语音处理的关键;也意味着人在连续讲话时,说话人在实时产生与听话人在实时处理语音信号时,都涉及语流的规划单元及声学模板。

阶层式多短语语流韵律架构 HPG 理论也明确说明跨短语之间的韵律关系。不同于传统语言学对语流或句调的看法,这个韵律架构把句群(即意义连贯的语段)中的短语或语句视为彼此之间具有亲属关系的韵律成分,同一层级的韵律单元是姊妹成分,韵律短语的句调一旦进入语段中的特定位置形成语段(言语段落)时,势必将有所调整以达到语流的最终韵律。一个语段可通过其如何开始、如何维持以及如何结束来加以阐明。对韵律组织而言,阶层性韵律单元和对应的韵律边界及边界信息十分重要,并有其相对应的声学韵律特征。而阶层式架构表示上层管辖下层,架构中的韵律单元包括了音段(声母、韵母)、音节 SYL(syllable)、韵律词 PW(Prosodic Word)、韵律短语 PPh(Prosodic Phrase)、韵律句群 PG(Prosodic Phrase Group),以及相对应的五级边界停顿 B1 到 B5,这些韵律单元和边界之间,呈现层级性关系,换言之,在此韵律架构下的韵律单元、韵律边界和边界停顿停延,都有系统性的韵律关系。

我们先前已发现,语流中的时程分布及音节长短上可以由音段成分与韵律组织信息来预测(Tseng et al, 2004)。在本研究中,我们想要了解:(1)语流音强组型是否如时程分布一般,可由语料中通过量化程序导出?(2)韵律组型韵律是否可由组织的管辖效应(governing effect of prosody organization)解释?(3)台湾国语与北京普通话在口音上的区分是否表现在韵律组型上?

### 二 实验设计

#### (一) 实验语料介绍

本实验分析了三批国语及普通话语料库,其中包括两位台湾

国语播音员的朗读语料(一男一女,以下简称 TMS 和 TFS)以及一位北京女性播音员朗读语料(以下简称 BFS)。① 朗读文本包含 15 段长段落,段落字数从 85 到 981 个音节不等,每一个段落平均字数 377 音节。这三批语料在语速表现上,有很大的不同。其中说得最快的是台湾男性播音员 TMS,平均音节长度为 182ms,平均语速(包含停顿)每秒 4.3 个音节;最慢则为北京女性播音员 BFS,平均音节长度为 267ms,每秒 2.8 音节,显示出慢语速的语流韵律特征。三批语料的语速特征摘要如表 1 所示:

表 1 语料语速特征分析

语料 特征	TFS	TMS	BFS
音节总数	5655	5655	5483
音节平均时长 (ms) (仅考虑音段时长)	202.45	181.84	266.91
语速 (Syllable #/sec) (包含停顿)	4.23	4.33	2.816

## (二) 语料处理

首先以 HTK Toolkit 程序判定音段成分(声母、韵母)再以 SAMPA-T(Tseng et al, 1999)自动标记音段边界,然后人工标记感知到的韵律边界。音段边界标记的人工检查和韵律边界标记都是由训练有素的标记员所完成的。

在实验数值截取方面,首先使用 ESPS Toolkit 计算音段 RMS 值,每一音段以十帧(frame)为一单位,计算 10 帧的 RMS 平

均值,少于 10 帧的音段时长则直接取平均数。除此以外,为了排除因为每一段落在录音过程中的些微改变所造成的韵律层级差异,每一段落中的 RMS 值皆经过标准化过程,得到 NRMS 值。如此一来,我们即可将研究焦点放在韵律单元内的音强组型的比率差异,而不是每一音节所取得的绝对音强数值。

## (三) 分析过程

我们使用在(Tseng et al, 2004)的研究中提出的研究步骤及方法,也就是线性回归统计中的逐步回归技术(step-wise regression technique)来分析估算语料。在本研究中,我们发展出的一个四阶层的线性模型来预测发音人的音强及时长组型。此一阶层式的回归模型由下到上、一层一层地往上叠加,依序为:音节层、韵律词层(PW)、韵律短层(PPh)以及韵律句群层(BG),最后把每一层的评估结果加总,得到最后的评估结果,再与实际语料比对,以检视评估的比例是否有意义。全部过程都是要研究自变量(IVs):音段成分(segment identities),韵律架构(prosodic organization)与因变量(DVs):音节时长(syllable duration),NRMS 之间的关系,其分析步骤摘要如下:(1) DVs 直线转换;(2) 决定音段成分的分群;(3) 建立线性回归模型;(4) 删减无效之 IVs;(5) 无法预测之残差值(residual)由紧接着的更上层解释之,重复步骤(3)与(5)。

为了能达到更好的预测,DVs 必须尽可能地呈现正态分布。DVs 与其转换值(log, square root, etc)必须使用 N-plot 加以检验、比较何者接近正态分布,最接近正态分布者因而入选。于是,在接下来的回归分析过程中,我们决定使用三批语料原始时长数值以及 NRMS 平方根。

① 本研究承清华大学计算机系蔡莲红教授提供北京女性播音员普通话语料,特此致谢。

间断数据的线性模型使用 DataDesk 的部分平方和 (partial sum of squares) (type3) 建立之。在音节层中, 将语料中所有的音段分成若干的群组, 以减少回归分析的复杂度, 是必要的做法。以音强回归预测来说, 我们决定分成 10 组辅音群与元音群; 而时长回归部分, 则分为 8 组。分组的标准是依音节层中每个 IVs 项目的 DV 平均值而决定。换言之, 具有相近的时长/ NRMS 平均数之音段分在同一组, 音节声母类型、韵母类型以及字调类型则被当作自变项。同时, 音节前后位置因素亦是考虑因素之一, 另外, 我们也考虑了其二因子交互作用, 所以音节层之回归分析可用下列公式表示:

$$\begin{aligned} \text{DVs} & & \text{IVs} \\ \text{Dur (ms)/ Square root NRMS} & = & \text{constant} + \text{CT}_y + \text{VT}_y \\ & & + \text{Ton} \\ & + & \text{PCT} + \text{PVT} + \text{PrT} + \text{FCt} + \text{FVt} + \text{FlT} \\ & + & \text{2-way factors of each factor above} \\ & + & \text{Delta 1} \end{aligned}$$

回归分析之后, 将较小影响力的因素 ( $p\text{-value} > 0.1$ ) 予以排除, 无法由音段成分所预测之残差 Delta 1, 则移到在上一层分析。从 PW 层到 BG 层, 我们以韵律结构当作自变项, 所提取出的系数表示韵律单元在特定音节位置的效应值 (时长或音强的增加或减少)。

在 PW 层中, 我们的目的在于了解 DV 是否受其在 PW 的位置所影响, 因此 PW 层模型之公式表示如下:

$$\text{Delta 1} = f(\text{PW length}, \text{PW sequence}) + \text{Delta 2}$$

无法由 PW 结构所预测之 PW 层的残差 Delta 2, 在紧接的上

一层 PPh 层继续分析。从我们最新的研究中发现, PPh 组织造成的效应, 在 PPh 中的第一与最后 4 音节的位置最为显著, 因此我们标记少于 8 音节的 PPh 的音节韵律序列信息为 [PPh 的长度, 该音节在 PPh 中的序列位置], 对大于或等于 9 音节者, 我们逐一标记前 4 与后 4 音节, 而其余中间音节为 [M], 也就是 {[I1], [I2], [I3], [I4], [M]…… [M], [F1], [F2], [F3], [F4]} 的标记方式。如此一来, 我们略过语料中长 PPh 不足的问题。PPh 层之公式表示如下:

$$\text{Delta 2} = f(\text{PP length}, \text{PP sequence}) + \text{Delta 3}$$

无法由 PW 与 PPh 所预测之 PPh 层的残差 Delta 3, 在紧接的上一层 BG 层继续分析。我们标记 BG 中的第一与结尾 PPh, 而其余的 PPh 则一视同仁, 当作中段的 PPh。在每一个 PPh 中, 少于 7 音节的 PPh 的音节韵律序列信息为 [PPh 长度, 该音节在 PPh 中的序列位置], 对大于或等于 7 音节者, 我们逐一标记前 3 与后 3 音节, 而其余中间音节则一视同仁, 于是, BG 内的起首 PPh 的音节位置可标记为 {[II1], [II2], [II3], [IM]…… [IM], [IF1], [IF2], [IF3]}。而 BG 层之公式表示如下:

$$\text{Delta 3} = f(\text{PPh I | M | F}, \text{PPh length}, \text{PPh sequence}) + \text{Delta 4}$$

采用两项数值评估预测结果: (1) 相关系数 (Correlation Coefficient,  $r$ ), 表示预测结果与原始资料的关联程度。(2) 总残差误差 (T. R. E.) 为残差平方和与原始资料平方和百分比。T. R. E. 意味着由底层之音节层到目前这一层为止无法说明的残差误差比率。

### 三 实验结果

本节包括二项实验结果,一为音强比较;一为音节时长预测。我们首先说明音强预测的分析结果,然后是音节时长预测结果。在此二预测过程中,资料皆是由底层的音节层分析到最高的BG层。在每一层中,一同比较三批语料的分析结果,以显出其中的差异性。

#### (一) 音强比较

##### 1. 音节层

在这一层中,每一批语料的回归分析使用不同的音段分群,表2到4描述了每一批语料中在音强回归分析中的音段分群。其分群标准系根据每一音段成分的NRMS的平方根数值的平均数,将相近平均数的音段成分分在同一群。语料间的辅音群分群比起元音群分群更加地近似。同时使用单因子和二因子之回归分析,并删减(忽略)p值大于0.1之因素。

表2 TFS—音强音段分群

Group	Consonant	Group	Vowel
CRMS1	h, s, k, p, s, ts', t	VRMS1	iau, @, y, ua, ia, U
CRMS2	tj, sj, dj, dz	VRMS2	i, U, ai, u, a, ou, iE
CRMS3	f, s', dz', g	VRMS3	au, yEn, an, iou, iaN
CRMS4	Z', b, d, l	VRMS4	yE, uai, @', yn, uan, iEn, aN, uei
CRMS5	m, n	VRMS5	uaN, @n, iN, @N, o
CRMS6	Zero	VRMS6	u@n, uo, oN, in, yoN, ei

表3 TMS—音强音段分群

Group	Consonant	Group	Vowel
CRMS1	ts, s, p, h, dz, tj, ts', k	VRMS1	y, U, i, u, U'
CRMS2	dj, t, sj, s'	VRMS2	iN, in, @N, iE, yE, @, yn, iEn
CRMS3	f, dz', g	VRMS3	lou, u@n, oN, @n
CRMS4	Z', l, b	VRMS4	iaN, ei, iau, yEn, o, yoN, ou, uei
CRMS5	d, m, n	VRMS5	uan, aN, an, ai, @', uo, uaN, uai
CRMS6	Zero	VRMS6	ia, au, a, ua

表4 BFS—音强音段分群

Group	Consonant	Group	Vowel
CRMS1	ts, s, h, f, p, dz	VRMS1	@, yEn, U, u, y, iaN, iN
CRMS2	dj, t, ts', k	VRMS2	yoN, aN, u@n, iEn, i
CRMS3	sj, tj	VRMS3	@N, oN, an, @n, au
CRMS4	g, l, dz', Z'	VRMS4	U', iau, ai, uai, ia, in, uan, uaN, ua
CRMS5	m, s'n, d, b	VRMS5	yE, uei, iou, ou, iE, ei, yn
CRMS6	Zero	VRMS6	uo, o, a

表5 三批语料音强预测之全面性评估

Corpus Layer	TFS		TMS		BFS	
	T. R. E.	r	T. R. E.	r	T. R. E.	r
Syllable	63.80%	0.616	47.65%	0.732	48.00%	0.724
PW	62.10%	0.621	47.02%	0.736	40.53%	0.737
PP	48.19%	0.666	37.43%	0.766	23.42%	0.775
BG	49.04%	0.693	35.37%	0.787	27.88%	0.814

表6 TFS-时长音段分群

Group	Consonant	Group	Vowel
CDUR1	d, b, g	VDUR1	@, o, U, U
CDUR2	dz', l, f	VDUR2	i, u, a, ei
CDUR3	n, Z'	VDUR3	yE, y, @n, in
CDUR4	m, dz, dj	VDUR4	uo, iE, ai, ou, uei
CDUR5	t, p, k, h	VDUR5	@N, oN, iN, an, au
CDUR6	s', ts', sj, s	VDUR6	yn, iau, aN
CDUR7	ts, tj	VDUR7	ia, iou, u@n, @', iEn, ua
CDUR8	Zero	VDUR8	uan, yEn, iaN, uaN, uai, yoN

表7 TMS-时长音段分群

Group	Consonant	Group	Vowel
CDUR1	d, b, g	VDUR1	@, o
CDUR2	dz', l	VDUR2	o, U, ei
CDUR3	n, f, Z'	VDUR3	i, u
CDUR4	m, t, p, dz, dj, k	VDUR4	a, in, uo, y, @n, iE, yE, ou
CDUR5	h, ts'	VDUR5	uei, iN, ai, @n, yn
CDUR6	ts, sj, tj	VDUR6	oN, iou, au, aN, iau, an
CDUR7	s, s'	VDUR7	ia, iEn, u@n, uaN, uai, ua, uan
CDUR8	Zero	VDUR8	yEn, iaN, yoN

## 2. PW 层

图 1-3 显示出在第三批语料中 PW 层的音强组型。每一条线代表在一个韵律词中特定位置音节所对应之回归系数。Y 轴代表 NRMS 平方根的预测值, 正值表示在特定位置音节的音强具大于平均值, 负值则表示音强小于平均值。PW 层的一般组型清楚地

显示出语料间的差异。愈长的韵律词, 需要的能量越大, 起始的音强越大。除此之外, 以 2 音节与 3 音节的韵律词来说, TFS 的回归系数范围为 0.016 和 0.019, TMS 的回归系数范围为 0.003 和 0.012, BFS 则是 0.028 和 0.054, 显示出两批台湾国语语料在音强方面, 较北京普通话低。

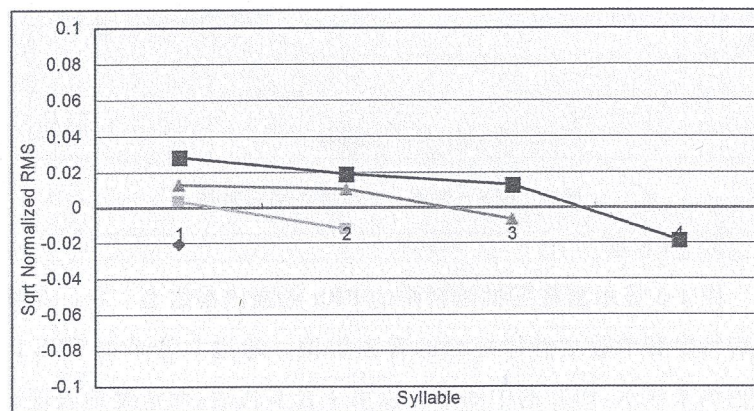


图1 TFS-音强-PW层之回归系数

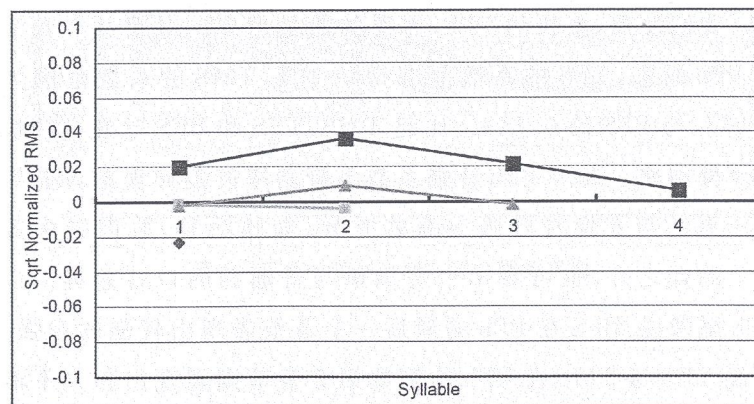


图2 TMS-音强-PW层之回归系数

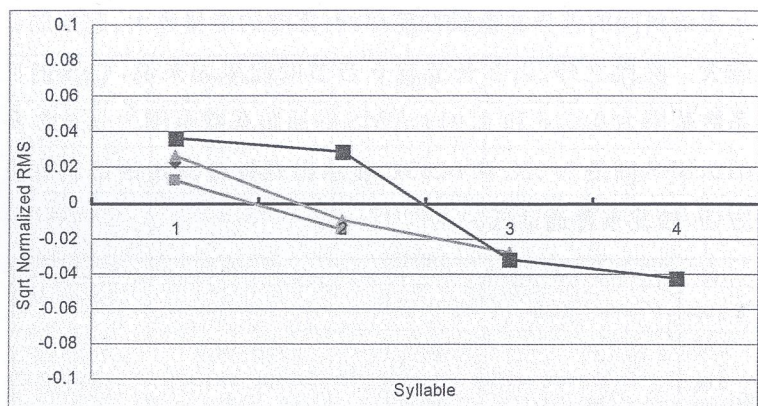


图3 BFS—音强—PW层之回归系数

### 3. PPh 层

图 4-6 显示出在第三批语料中的 PPh 音强比率组型。每一条线代表特定音节数之韵律短语的音强组型。超过 9 音节的 PPh 以紫色线条显示, PPh 的中间部分以第 4 音节代表, 而清楚地表现出前 4 与后 4 音节的音强组型。

我们发现, 愈长的 PPh, 需要的能量越多。对超过 9 音节的 PPh 而言, TFS 的系数范围为 0.101, TMS 的系数范围为 0.088, 与 BFS 的 0.118。比起 TMS、TFS 和 BFS 皆具有较快的下降斜率。此一 PPh 音强系数下降斜率的差异现象可能导因于男性与女性在呼吸上有些差异, 如肺活量、肌肉强度等等。除此之外, 也可看出台湾国语与普通话的口音差异。以 PPh 尾来说, BFS 在 PPh 的最后一个音节表现出较强的音强, 反观 TFS 与 TMS 在 PPh 的倒数第 2 音节则表现出相当不清楚的突显 (prominence)。超过 7 音节的 PPhs 之回归系数范围, TFS 是 0.1, TMS 是 0.073, 而 BFS 则为 0.096。超过 7 音节

中段位置 PPhs 的回归系数范围, TFS 是 0.081, TMS 是 0.056, 而 BFS 则为 0.094。

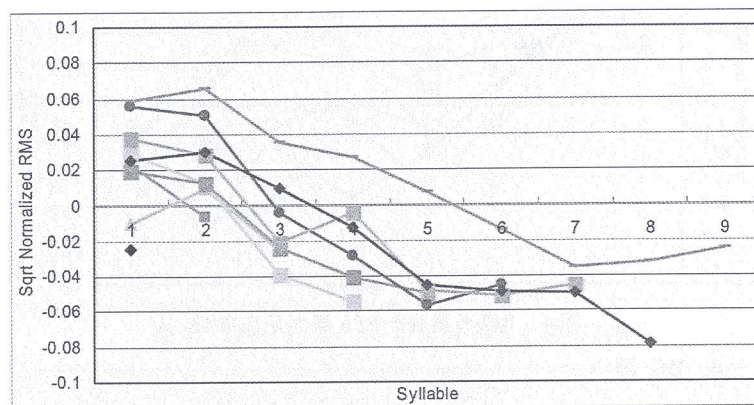


图4 TFS—音强—PPh层之回归系数

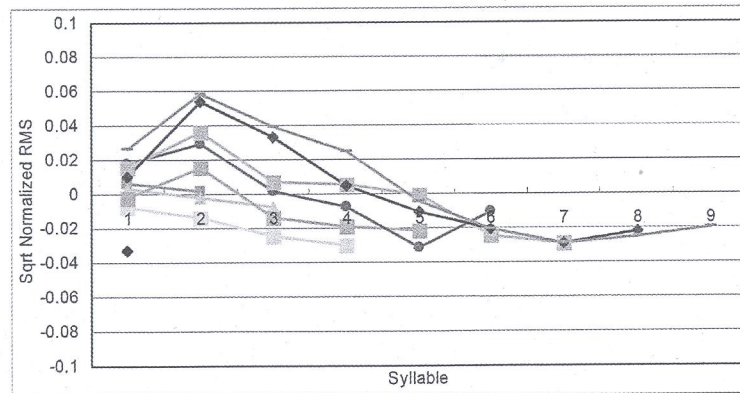


图5 TMS—音强—PPh层之回归系数

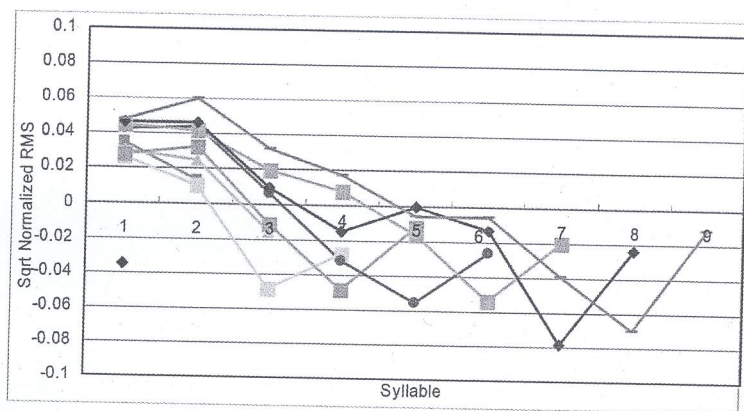


图6 BFS-音强-PPh层之回归系数

#### 4. BG 层

在 BG 中,我们假设在 BG 中不同位置的 PPh 的贡献度并不相同,因而进行个别分析。BG 中第一与最后的 PPh 视为起首与结尾 PPh,而其他的则为中段 PPh。

##### (1) BG 中的起首 PPh(BG-initial PPh)

图 7-9 显示在 BG 中的起首 PPh 音强组型。每一条线代表不同长度 PPh 的组型。超过 7 音节的起首 PPh 以紫色线条显示,我们发现:相对于前文中已出现过的 PPh 组型,当起首 PPh 比较长的时候,它反而没有较强的音强。一般来说,除了第一与最后音节以外,BG 中的起首 PPh 音节和所导出的预测平均值比较起来,具有较强的音强。

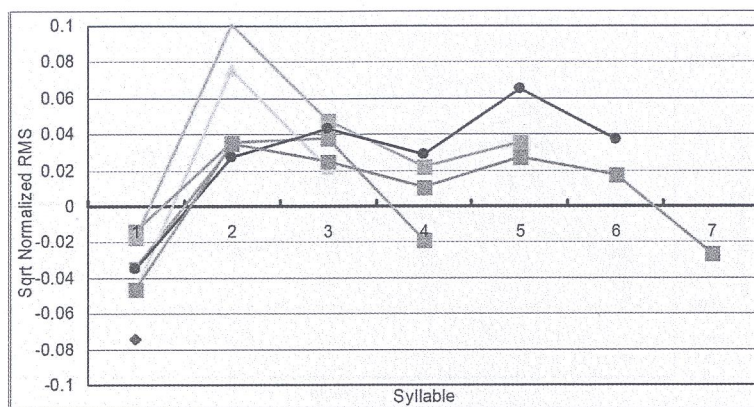


图7 TFS-音强-BG中起首PPh之回归系数

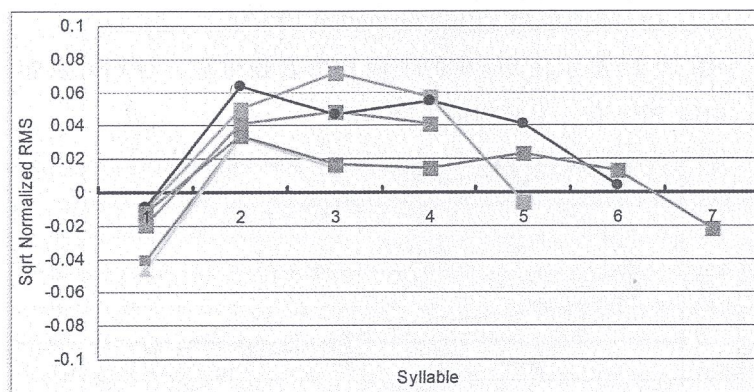


图8 TMS-音强-BG中起首PPh之回归系数

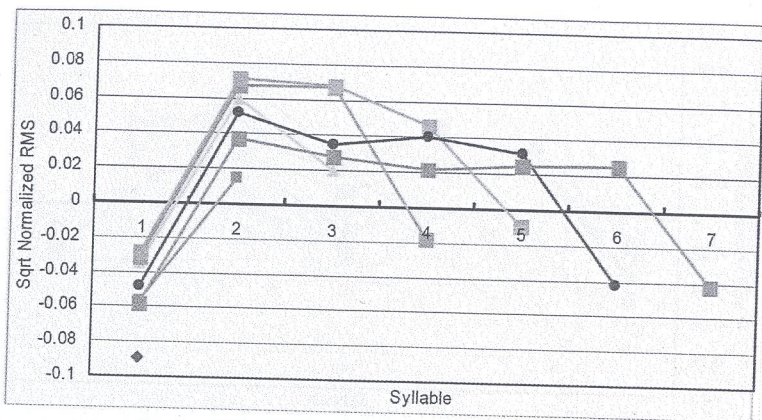


图9 BFS-音强-BG 中起首 PPh 之回归系数

## (2) BG 中的中段 PPh(BG-medial PPh)

图 10-12 显示在 BG 中的中段 PPh 音强组型。与 PPh 层相类似,中段 PPh 有一个相似的下组型。

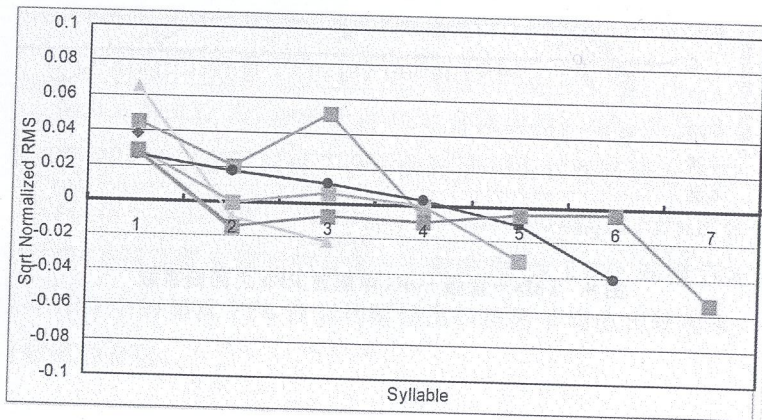


图10 TFS-音强-BG 中段 PPh 之回归系数

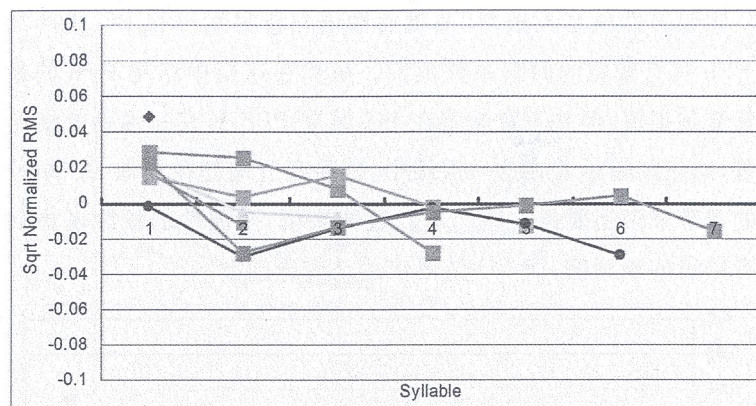


图11 TMS-音强-BG 中段 PPh 之回归系数

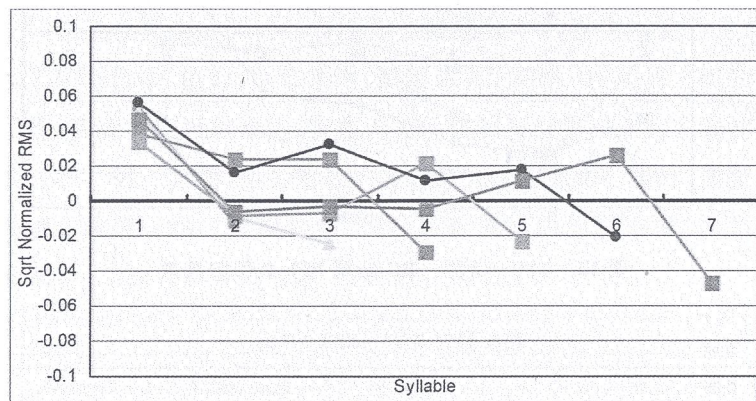


图12 BFS-音强-BG 中段 PPh 之回归系数

## (3) BG 中的结尾 PPh(BG-final PPh)

图 13-15 显示在 BG 中的结尾 PPh 音强组型。所有的语料均显示出,结尾 PPh 并没有表现出如前所见的 PPh 音强减弱,反而呈现一相反趋势:在 BG 结尾伴随着一个较强的音节。较短的结尾 PPh 具有一较宽的系数范围。对超过 7 音节的结尾 PPh 而言,系数范围 TFS 是 0.077, TMS 是 0.059, 而 BFS 则为 0.094。



比起男性发音人语料, 两批女性语料皆显示出在 BG 中的结尾 PPh 具有较宽的回归系数范围。虽然在这层中结尾 PPh 的最后音节是正值, 但并不意味着在 BG 结尾 PPh 具有一强劲的结尾音节, 这仅能暗示着在这一层级中, 这些音节比起在其他 BG 单元里的音节来说还要强一些。若以整体来说, 全面性的预测考虑到所有层级的预测值, 而 BG 结尾依然是变弱的。

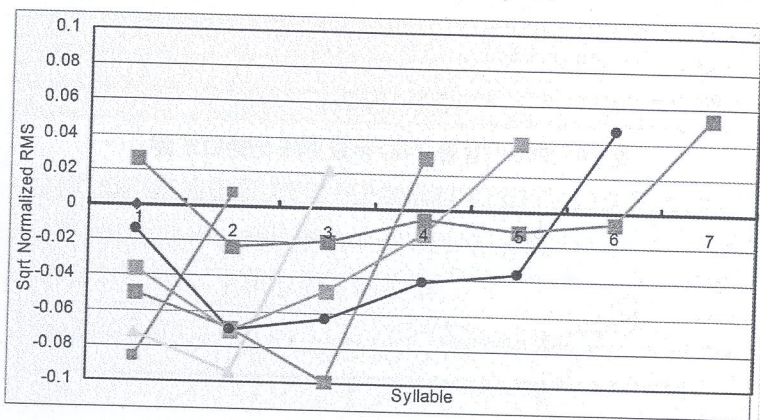


图 13 TFS-音强-BG 结尾 PPh 之回归系数

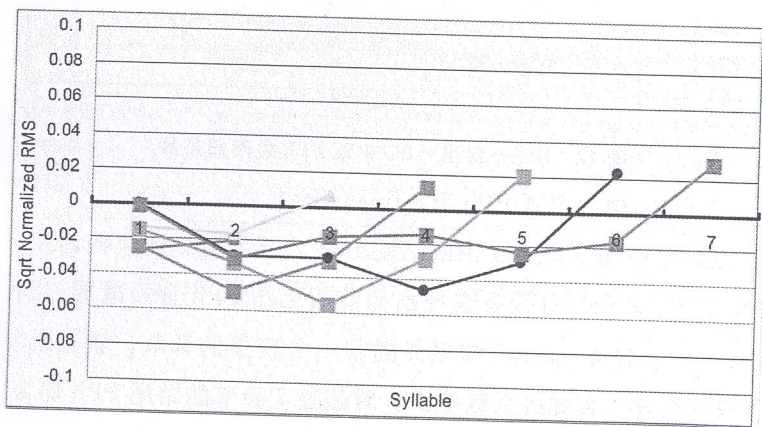


图 14 TMS-音强-BG 结尾 PPh 之回归系数

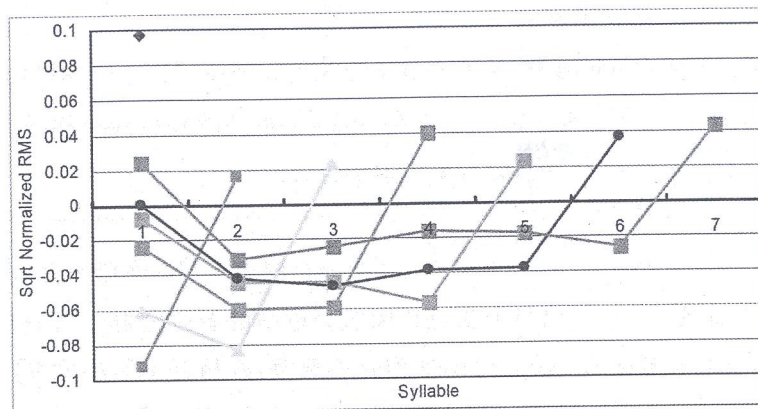


图 15 BFS-音强-BG 结尾 PPh 之回归系数

全面性预测之评估如表 8 摘要描述如下。值得注意的是: 原始 DV(NRMS 值)与总和预测(预测总和之平方)转换值之间的评估值, 因此我们可以推论: 残差误差的增加可能导因于此非线性转换过程。另一方面, 相关系数  $r$  意味着当增加更高一层时, 预测力获得了改善。在第三批语料中, BFS 具有最好的预测结果(相关系数  $r = 0.814$ ), 而最差的是 TFS(相关系数  $r = 0.693$ )。

表 8 BFS-时长音段分群

Group	Consonant	Group	Vowel
CDUR1	d, b	VDUR1	@, U, U'
CDUR2	g, l	VDUR2	u, o
CDUR3	n, dz, Z', m	VDUR3	a, l, yE, ou, ua
CDUR4	dz, dj	VDUR4	ei, ai, uo, au
CDUR5	t, p, f, k, h	VDUR5	oN, in, @n, u@n, uaN, @N, an, aN, iE, iN
CDUR6	ts'	VDUR6	uei, y, ia, yn, uan, iau
CDUR7	tj, s', sj, s, ts	VDUR7	iEn, @', iaN, iou
CDUR8	Zero	VDUR8	yEn, uai, yoN

## (二) 时长预测

在时长预测部分,除了音节层的音段分群外,所有过程都与音强的回归分析一样。以三批语料来说,原始音段时长分布称得上接近正态分布,在回归分析中作为 DV。

## 1. 音节层

在这一层中,每一批语料的回归分析使用不同的音段分群,表 6-8 描述了每一批语料中在时长回归分析中的音段分群。其分群标准系根据每一音段成分的平均时长数值,将相近平均数的音段成分分在同一群。同时使用单因子和二因子之回归分析,并删减(忽略)p 值大于 0.1 之因素。

表 9 TFS—时长音段分群

Group	Consonant	Group	Vowel
CDUR1	d, b, g	VDUR1	@, o, U, U
CDUR2	dz, l, f	VDUR2	i, u, a, ei
CDUR3	n, Z'	VDUR3	yE, y, @n, in
CDUR4	m, dz, dj	VDUR4	uo, iE, ai, ou, uei
CDUR5	t, p, k, h	VDUR5	@N, oN, iN, an, au
CDUR6	s', ts', sj, s	VDUR6	yn, iau, aN
CDUR7	ts, tj	VDUR7	ia, iou, u@n, @, iEn, ua
CDUR8	Zero	VDUR8	uan, yEn, iaN, uaN, uai, yoN

表 10 TMS—时长音段分群

Group	Consonant	Group	Vowel
CDUR1	d, b, g	VDUR1	@, o
CDUR2	dz, l	VDUR2	o, U, ei
CDUR3	n, f, Z'	VDUR3	i, u
CDUR4	m, t, p, dz, dj, k	VDUR4	a, in, uo, y, @n, iE, yE, ou
CDUR5	h, ts'	VDUR5	uei, iN, ai, @n, yn

CDUR6	ts, sj, tj	VDUR6	oN, iou, au, aN, iau, an
CDUR7	s, s'	VDUR7	ia, iEn, u@n, uaN, uai, ua, uan
CDUR8	Zero	VDUR8	yEn, iaN, yoN

表 11 BFS—时长音段分群

Group	Consonant	Group	Vowel
CDUR1	d, b	VDUR1	@, U, U
CDUR2	g, l	VDUR2	u, o
CDUR3	n, dz, Z', m	VDUR3	a, l, yE, ou, ua
CDUR4	dz, dj	VDUR4	ei, ai, uo, au
CDUR5	t, p, f, k, h	VDUR5	oN, in, @n, u@n, uaN, @N, an, aN, iE, iN
CDUR6	ts'	VDUR6	uei, y, ia, yn, uan, iau
CDUR7	tj, s', sj, s, ts	VDUR7	iEn, @, iaN, iou
CDUR8	Zero	VDUR8	yEn, uai, yoN

表 12 TFS—时长—音节层因素之 ANOVA 摘要表

Source	df	Sums of Squares	Mean Square	F-ratio	Prob	Source	df	Sums of Squares	Mean Square	F-ratio	Prob
CTy	7	615273	87896.1	42.767	<0.0001	CTy * VTy	49	199615	4073.77	1.9821	<0.0001
VTy	7	367437	52491.1	25.54	<0.0001	CTy * FTh	35	140762	4021.78	1.9568	0.0006
Ton	4	109460	27365	13.315	<0.0001	FCT	8	31730.5	3966.32	1.9299	0.0514
Ton * FCT	32	356105	11128.3	5.4146	<0.0001	PCT * PTh	25	95864.1	3834.56	1.8658	0.0056
CTy * Ton	25	183320	7332.81	3.5679	<0.0001	CTy * PTh	35	126427	3612.19	1.7576	0.0039
VTy * Ton	23	152705	6639.33	3.2305	<0.0001	PCT * PVT	49	173919	3549.36	1.727	0.0013
PVT	7	42277.2	6039.59	2.9386	0.0045	PVT * PTh	23	76195.2	3312.83	1.6119	0.0324
PTh	4	23319.3	5829.82	2.8366	0.023	FVT * FTh	23	74873	3255.35	1.5839	0.0377
PCT	8	37596.2	4699.52	2.2866	0.0193	Ton * PCT	31	96176.4	3102.46	1.5095	0.0348
FCT * FVT	56	262766	4692.25	2.2831	<0.0001	VTy * FCT	56	156598	2796.39	1.3606	0.0387
FCT * FTh	29	134237	4628.86	2.2522	0.0001	Const	1	2.31E+08	2.31E+08	1.12E+05	<0.0001
Error	5118	1.05E+07	2055.23			Total	5654	2.40E+07			

表 13 TMS—时长—音节层因素之 ANOVA 摘要表

Source	df	Sums of Squares	Mean Square	F-ratio	Prob	Source	df	Sums of Squares	Mean Square	F-ratio	Prob
CTy	7	175279	25039.9	14.276	<0.0001	CTy * FTh	35	122844	3509.82	2.001	0.0004
PVT	9	74244.5	8249.38	4.7031	<0.0001	PCT * PTh	31	103806	3348.59	1.9091	0.0018
Ton * FTh	20	163881	8194.05	4.6716	<0.0001	PVT * PTh	26	80244.6	3086.33	1.7596	0.01
FCT	8	56309.5	7038.69	4.0129	<0.0001	VTy * FVT	49	148264	3025.8	1.7251	0.0013
FVT * FTh	26	178220	6854.63	3.9079	<0.0001	FCT * FVT	50	146530	2930.6	1.6708	0.0022
VTy	7	45882.4	6554.62	3.7369	0.0005	VTy * PCT	56	148779	2656.77	1.5147	0.0081
VTy * Ton	26	146987	5653.35	3.2231	<0.0001	VTy * FTh	28	72556.2	2591.29	1.4773	0.0504
CTy * VTy	43	226817	5274.82	3.0073	<0.0001	VTy * PVT	49	119575	2440.3	1.3913	0.0373
PTh	4	18084.9	4521.23	2.5776	0.0356	VTy * FCT	56	136265	2433.31	1.3873	0.0301
CTy * Ton	26	116595	4484.44	2.5567	<0.0001	PTh * FCT	39	92503.4	2371.88	1.3522	0.0709
FCT * FTh	29	128627	4435.42	2.5287	<0.0001	Const	1	2.08E+08	2.08E+08	1.19E+05	<0.0001
Error	5030	8.82E+06	1754.03			Total	5654	2.05E+07			

表 14 BFS—时长—音节层因素之 ANOVA 摘要表

Source	df	Sums of Squares	Mean Square	F-ratio	Prob	Source	df	Sums of Squares	Mean Square	F-ratio	Prob
CTy	7	417148	59592.5	14.966	<0.0001	FTh	4	30036.5	7509.13	1.8858	0.11
Ton	4	96536.5	24134.1	6.0609	<0.0001	VTy * FTh	28	197720	7061.43	1.7734	0.0073
CTy * VTy	46	655573	14251.6	3.5791	<0.0001	VTy * PTh	33	223588	6775.41	1.7015	0.0075
FVT	7	85042.7	12149	3.051	0.0033	VTy * FCT	56	371720	6637.86	1.667	0.0014
Ton * FTh	16	184510	11531.9	2.896	<0.0001	FCT	8	53079.5	6634.94	1.6663	0.1013
VTy * Ton	26	294887	11341.8	2.8483	<0.0001	Ton * PCT	32	208780	6524.38	1.6385	0.0132
FCT * FTh	27	298044	11038.7	2.7722	<0.0001	VTy * PCT	56	351417	6275.3	1.5759	0.0041
Ton * FCT	32	312549	9767.16	2.4529	<0.0001	PVT * PTh	30	185741	6191.37	1.5549	0.0275
FVT * FTh	26	235986	9076.38	2.2794	0.0002	PCT * PVT	59	343430	5820.84	1.4618	0.0125
FCT * FVT	44	342737	7789.47	1.9562	0.0002	VTy * FVT	49	262980	5366.94	1.3478	0.0537
CTy * Ton	27	209032	7741.94	1.9443	0.0024	Const	1	3.91E+08	3.91E+08	98096	<0.0001
Error	4865	1.94E+07	3981.94			Total	5482	3.89E+07			

## 2. PW 层

图 16-18 显示出在第三批语料中之 PW 层的回归系数。这些图当中,每一条线代表不同长度的韵律词单元,而 Y 轴中每一点代表在一个韵律词中特定位置音节所得来的回归系数,这些系数表示相较于低一层所获得的预测,该音节时长延长/缩短的程度。PW 层的一般组型像一勺子状曲线——PW 的中间音节缩短而最后音节拉长。在第三批语料中,TFS 的延长/缩短效应最小,在四音节的 PW 中最短和最长音节相差 19.68ms 是最为显著的差异;BFS 在 3 音节 PW 则有 39.52ms 的差异。

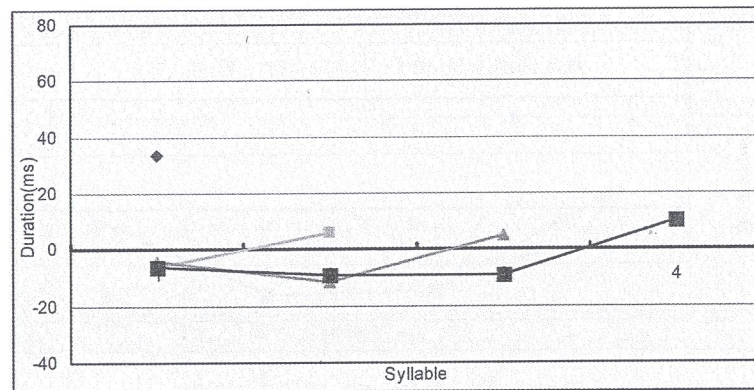


图 16 TFS—时长—PW 层之回归系数

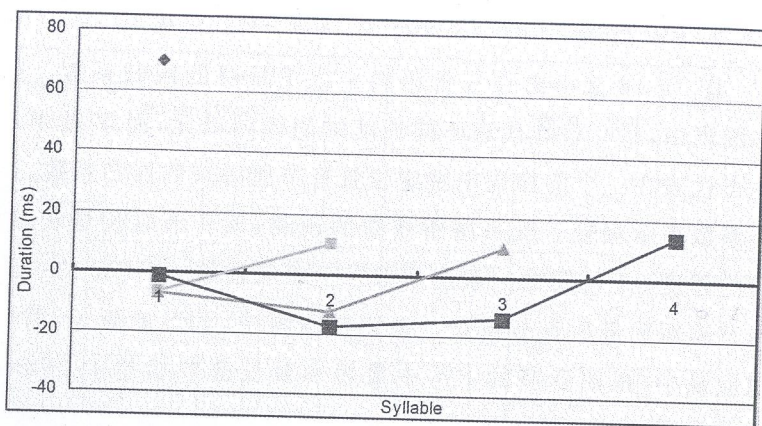


图 17 TMS—时长—PW 层之回归系数

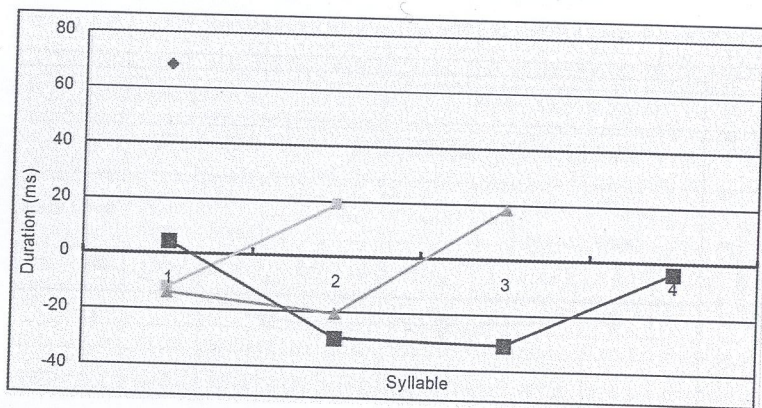


图 18 BFS—时长—PW 层之回归系数

### 3. PPh 层

图 19-21 显示了三批语料中 PPh 层的回归系数。PPh 层的一般组型可从 PPh 中的最后 4 音节中所看出。不管什么音节长度的韵律短语,从最后倒数第 4 个音节开始皆维持相同的组型,(倒数)第三音节明显最短,而最后音节则拉长,这便是众所熟知的“短语结尾效应”(phrase final effect),其实它就是

不同韵律层级的边界效应。TMS 和 TFS 在短语结尾表现出相同的组型,BFS 则有倒数第 3 音节最短,最后音节却没有拉长的倾向。

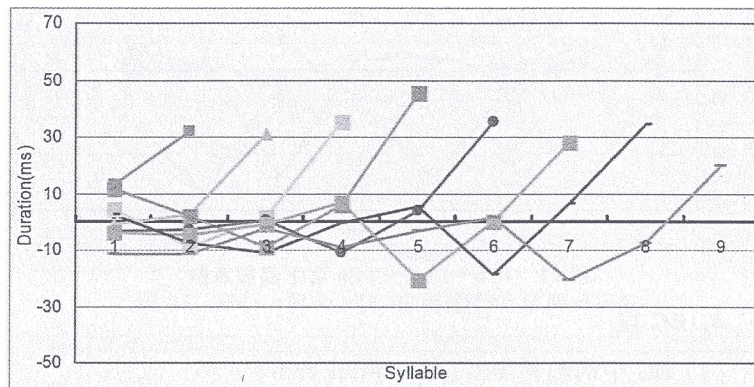


图 19 TFS—时长—PPh 层之回归系数

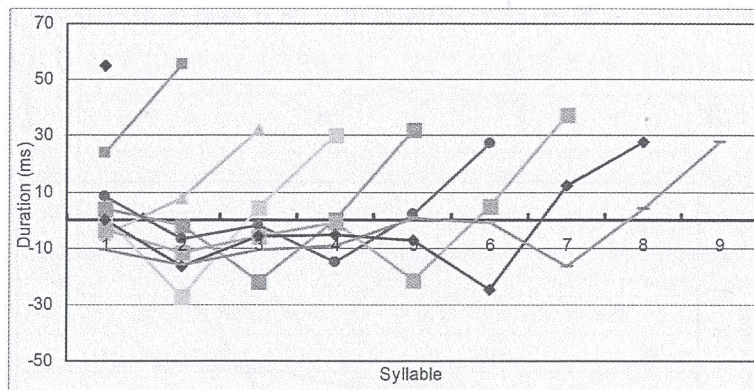


图 20 TMS—时长—PPh 层之回归系数

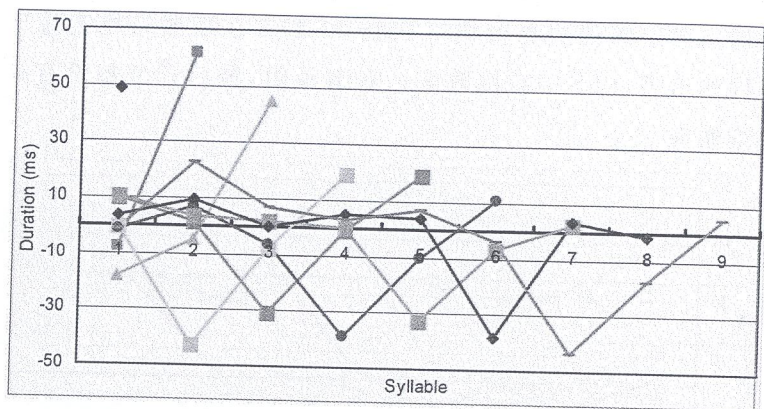


图 21 BFS—时长—PPh 层之回归系数

#### 4. BG 层

##### (1) BG 中的起首 PPh(BG-initial PPh)

三批语料中,由起首 PPh 所贡献之一般时长组型十分相似。相较于 PPh 的预测,起首 PPh 短语末音节是更加地延长,起首 PPh 的最后音节均被延长了 23.85(TFS)到 29.55ms (BFS)不等的长度。

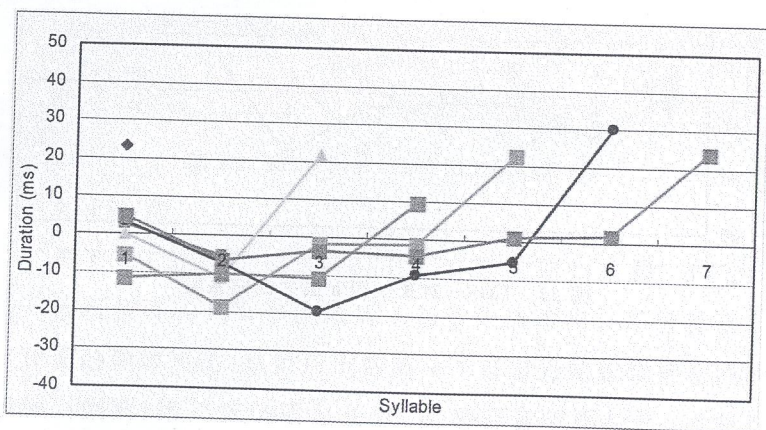


图 22 TFS—时长—BG 中起首 PPh 之回归系数

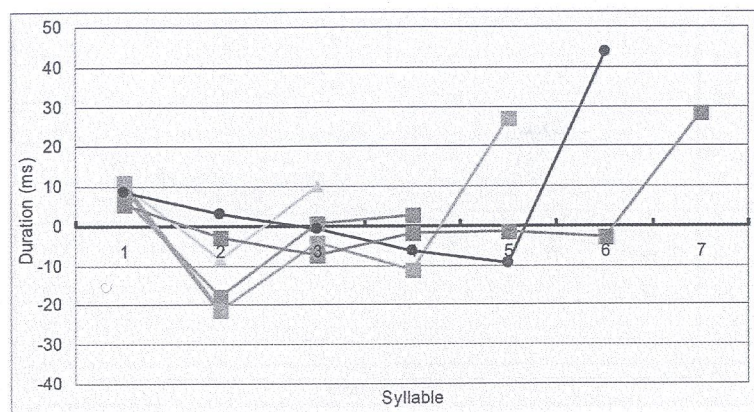


图 23 TMS—时长—BG 中起首 PPh 之回归系数

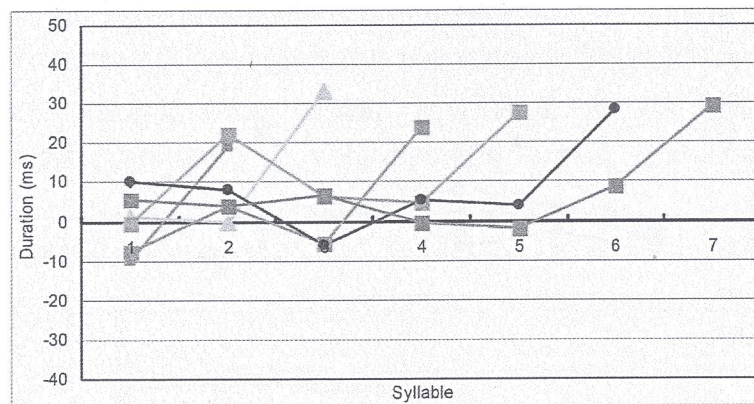


图 24 BFS—时长—BG 中起首 PPh 之回归系数

##### (2) BG 中的中段 PPh (BG-medial PPh)

除了已知的 PPh 结尾延长外,在这一层中,BG 的中段 PPh 更加地延长。在这里,必须注意起首 PPh 与中段 PPh 之间的差异:(1)起首 PPh 比中段 PPh 多了 10ms 左右,具有较大的延长效应。(2)中段 PPh 的第一个音节时长缩短了 10ms,这样的现象在起首 PPh 并未发现。

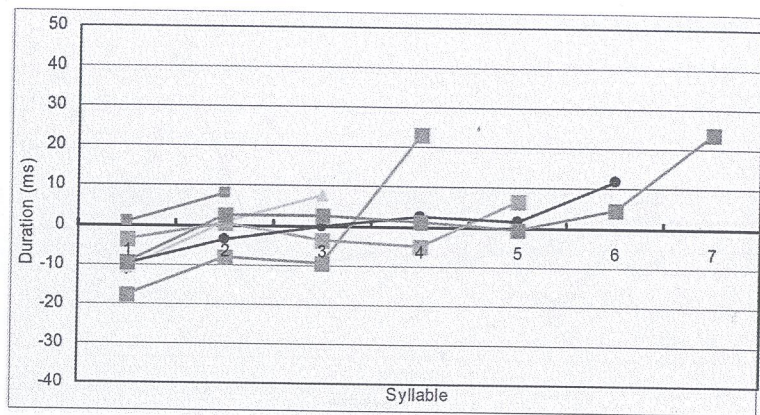


图 25 TFS-时长-BG 中段 PPh 之回归系数

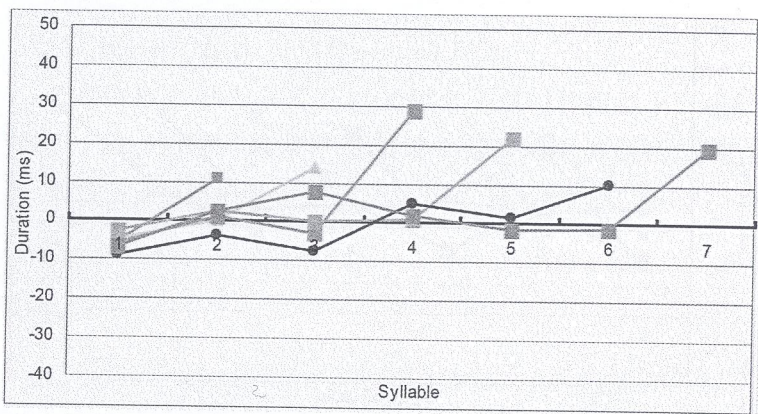


图 26 TMS-时长-BG 中段 PPh 之回归系数

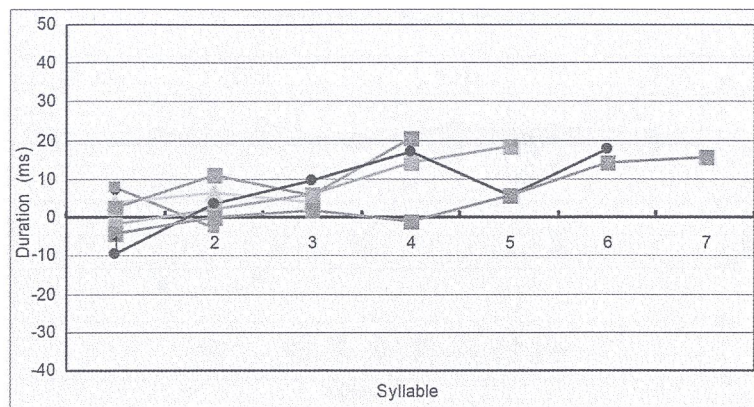


图 27 BFS-时长-BG 中段 PPh 之回归系数

(3) BG 中的结尾 PPh (BG-final PPh)

BG 的结尾 PPh 的最后音节并没有如起首与中段 PPh 一样拉长,而是明显地缩短。对超过 7 音节的结尾 PPh, TFS 缩短了 31.3ms, TMS 缩短了 45.55ms, BFS 则为 18.8ms。然而, BG 最后音节的全面性效果应是预测结果的累积总和,其实际韵律表现依然呈现延长现象。

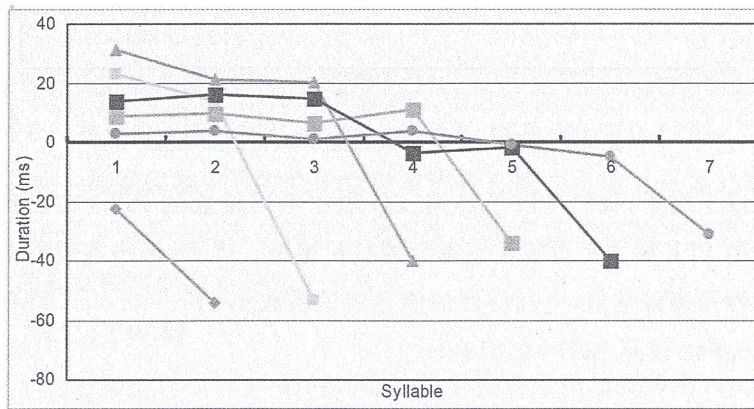


图 28 TFS-时长-BG 结尾 PPh 之回归系数

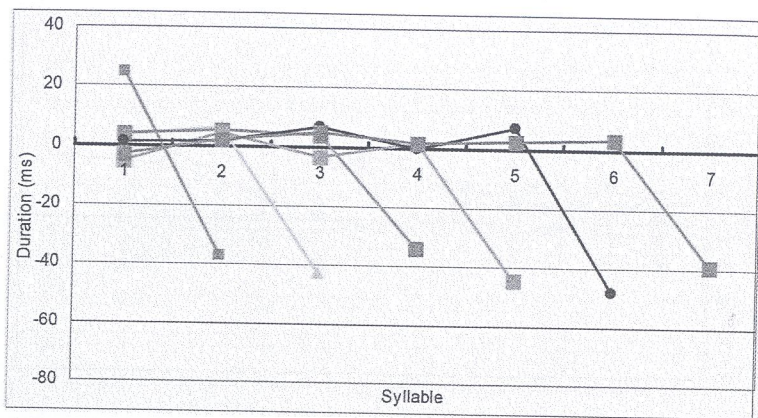


图 29 TMS—时长—BG 结尾 PPh 之回归系数

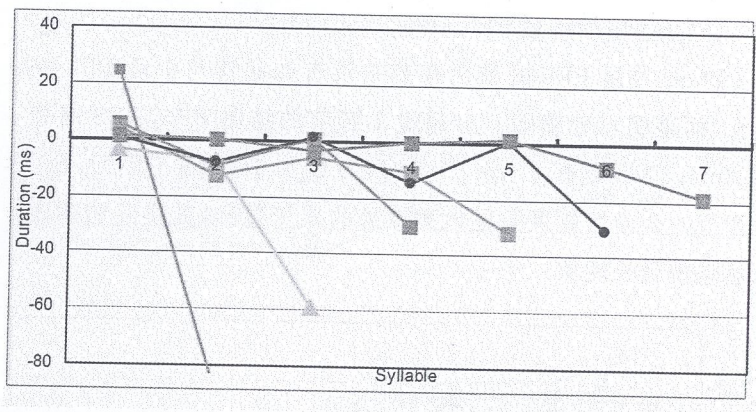


图 30 BFS—时长—BG 结尾 PPh 之回归系数

从以上结果,我们可以了解到:BG 层中不同位置的 PPhs 在已预测的 PPh 层之上贡献了它本身的时长组型。表 15 列出各层的全面性预测评估值,全面性评估值之相关系数  $r$ , TFS=0.806, TMS=0.842 以及 BFS=0.819。

表 15 三批语料时长预测之全面性评估

Corpus Layer	TFS		TMS		BFS	
	T. R. E.	r	T. R. E.	r	T. R. E.	r
Syllable	43.88%	0.749	43.11%	0.754	49.97%	0.709
PW	42.43%	0.759	39.87%	0.776	43.22%	0.755
PP	38.07%	0.789	33.41%	0.822	36.13%	0.803
BG	35.13%	0.806	29.37%	0.842	33.56%	0.819

#### 四 结果分析与讨论

我们发现了在阶层式结构中的典型的韵律组型,如同时长组型一般,音强组型也可从音段成分与韵律结构中得到。除此以外,北京普通话与台湾国语间的差异也可由每一韵律单元层级中看出。在本节中,我们将分别讨论每一层的实验结果。

##### (一) 音节层

从音节层的实验分析结果中,我们可以看到,将近 50% 的预测值可从音段成分因素所获得。以音强预测来说,音节层预测了接近 52% 的原始资料是(T. R. E. = 47.65% in TMS and 48% in BFS),最差的是 TFS,只解释了 36.2%。至于时长预测部分, TFS 已被解释了 56.12%, TMS 56.89% 以及 BFS 50.03%。因此,此一结果证明了为何韵律结果对语音合成在本质上是如此重要,语音合成若是只考虑到音节的串接与过渡,那将因缺少了更上层的韵律阶层信息而听起来不自然。

##### (二) PW 层

在三批语料中, PW 的时长与音强组型是可预估获得的,音强的逐渐下降与时长像带把勺子状的组型,均显示了韵律词组型之

原型(prototype)。比起其他两者,较慢速的 BFS 语料的韵律组型是更加地明显,而且 PW 层韵律组型被解释的比率在 BFS 也是比较大的。以音强预测来说,BFS 的 T. R. E. 减少了将近 8%,反而 TFS 只有 1.7%与 TMS 的 0.63%。以时长预测来说,BFS 的 T. R. E. 减少了将近 6.75%,反而 TFS 只有 1.45%与 TMS 的 2.24%。除此之外,我们也认为北京普通话与台湾国语之间口音上的差异,除了音段成分的差异,也有韵律组型的不同,北京普通话比起台湾国语具有较明显的轻重音与缩短/延长组型。

### (三) PPh 层

从三批语料中所取得之 PPh 层的一般韵律组型,清楚地显现出音强的下降与如 PPh 结尾一般的时长终止式(cadence)。当时长组型中可以发现类似的型态时,北京普通话与台湾国语的清楚差异也被发掘。相对于结尾的弱化,BFS 却加强了短语的最后音节;在其时长预测部分,和 PW 层的结果比较起来,PPh 结尾也没有更进一步拉长,反而结果是最后音节的前面一个音节的缩短。在三层韵律层级中(PW, PPh, BG layer),PPh 层在音强与时长的预测上说明了绝大部分在 BFS 的音强预测上,PPh 层将近解释了 17%的 T. R. E. 值。

### (四) BG 层

在 BG 的起首 PPh、中段 PPh 与结尾 PPh 都发现了清楚的韵律组型。以起首 PPh 来说,最后音节会拉长,而中间的音节比 BG 其他位置的音节还要念得较为强烈。以中段 PPh 来说,第一音节被加重与缩短,而最后音节则被弱化与延长。至于结尾 PPh,在时长与音强组型上都发现了与 PPh 相反的组型,最后两音节呈现出重音上的轻重组型(weakened-stressed pattern)。除此以外,和相

同 BG 的其他 PPh 比较起来,最后音节同时也缩短了。

## 五 结语

在本文中,我们通过配合“阶层式多短语语流韵律架构 HPG”的回归统计模型,检验大量语料的声学语音特性,在音强分布和语流中音节在时程配置二方面作预测,发现当上层的韵律单元增加时,预测力也增加。进一步的证明语流的韵律最终输出现象,无论是音强或时程配置,都是各韵律层级分层贡献的总和。通过实验比较两岸口语在语流韵律,我们得到以下的结论:

(1) 由各层分析显示,语流由多层次的架构所组合而成,每一层皆有典型的特征,证明口语语流韵律存在着一定的规划性及规则性。

(2) 韵律结构中,音强和时长都具有规则性,而以音节长短在时程配置上规则性最是明显,在各阶层间显现类似的边界效应及特征。层层叠加的结果,是口语语流韵律与孤立句的句调在听感上显现极大差异的最主要原因。

(3) 两岸虽有口音的差异,在语流韵律的特性上大致还是相当一致的,亦即基本的韵律差异并不大。我们认为,必须进行更进一步的实验,以厘清两岸口音差异还有哪些语音层面,如音段及声调系统,甚或是否还包括了语速差别。

(4) 我们通过量化所获得的韵律模型,可当做语音合成的基础,作为语音科技开发方面的应用。

本研究除以上发现外,在研究方法方面,采用语料库语音学,因而大幅增加了语料的量及变异性;并因此配合了量化的处理。



在研究课题方面,走出汉语语音学研究一向重字调或句调、轻韵律及实际说话现象的框架,将口语篇章的特点有系统地带进语音学的研究。

### 【参考文献】

1. 吴为章、田小琳《汉语句群》,商务印书馆 2002 年。
2. Tseng, Chiu-yu, Pin, Shao-huang and Lee, Yeh-lin. 2004. Speech prosody: Issues, approaches and implications. In Fant, G., Fujisaki, H., Cao, J. and Xu, Y. (ed.), *From Traditional Phonology to Modern Speech Processing* (语音学与言语处理前沿). pp. 417-437. Beijing, China: Foreign Language Teaching and Research Press, 外语教学与研究出版社。
3. Tseng Chiu-yu, Pin, ShaoHuang, Lee, Yeh-lin, Wang, Hsin-min and Chen, Yong-cheng. 2005. Fluent Speech Prosody: Framework and Modeling. *Speech Communication (Special Issue on Speech Prosody)*, Vol. 46(3-4): 284-309.
4. Tseng, Chiu-yu and Chou, Fu-chiang. 1999. Machine readable phonetic transcription system for Chinese dialects spoken in Taiwan. *The Journal of the Acoustical Society of Japan (E)* 20(3): 215-223.