

兩岸口語語流韻律比較初探——以音強及音節時程分佈為例

鄭秋豫¹，李岳凌¹，蔡蓮紅²，鄭雲卿¹

¹中央研究院語言學研究所

cytling@sinica.edu.tw

²北京清華大學計算機系

1. 引言

中文口語語流最顯著的特徵為短語成組 (phrase grouping)，在口語篇章裏最為明顯 (吳為章&田小琳，2002)。鄭秋豫透過量化處理大批口語語料得到的證據，對這個現象提出階層式多短語語流韻律架構一說，每一層韻律單位對於最終的語流韻律表徵都有貢獻，而語流韻律則是各層級韻律單元貢獻度的總和 (Tseng et al 2004, 2005)。這個架構最重要的特點，是強調並界定各短語間彼此的韻律關連性，而非各短語的孤立句調。她採用語料庫語音學之研究方法，從感知角度出發分析口語語流韻律，在聲學語音訊號 (基頻走勢、音節時長、邊界資訊、邊界停頓和停延段等) 上，找到各韻律階層貢獻的證據及聲學範本 (template)，這些模版的存在，是言語鍊裏語言產製和語音處理的關鍵；也意謂著人在連續講話時，說話人在即時產製與聽話人在即時處理語音訊號時，都涉及語流的規劃單元及聲學範本。

「階層式多短語語流韻律架構」理論也明確說明跨短語之間的韻律關係。不同于傳統語言學對語流或句調的看法，這個韻律架構把句群 (即意義連貫的語段) 中的短語或語句視為彼此之間具有親屬關係的韻律成份，而不是孤立、不相關的韻律單元；而階層式架構表示上層管轄下層，同一層級的韻律單元則是姊妹成分。韻律短語的句調一旦進入語段中的特定位置形成語段 (言語段落) 時，勢必將有所調整以達到語流的最終韻律，一個語段可藉其如何開始、如何維持以及如何結束加以闡明。對韻律組織而言，階層性韻律單元和對應的韻律邊界及邊界訊息十分重要，並有其相對應的聲學韻律特徵。架構中的韻律單元包括了音段 (聲母、韻母)、音節 SYL (syllable)、韻律詞 PW (Prosodic Word)、韻律短語 PPh (Prosodic Phrase)、韻律句群 PG (Prosodic Phrase Group)，以及相對應的五級邊界停頓 B1 到 B5，這些韻律單元和邊界之間，亦呈現層級性關係，換言之，在此一韻律架構下的韻律單元、韻律邊界和邊界停頓停延，都已發現有系統性的韻律關係。

我們先前已發現，語流中的時程分佈及音節長短上可以由音段成分與韻律組織訊息所預測 (Tseng et al, 2004)。在本研究中，我們想要瞭解：(1) 語流音強組型是否如時程分佈一般，可由語料中通過量化程式導出？(2) 韻律組型韻律是否可由組織的管轄效應 (governing effect of prosody organization) 解釋？(3) 臺灣國語與北京普通話在口音上的區分是否表現在韻律組型上？

2. 實驗設計

分析方法

2.1 實驗語料介紹

本實驗分析了三批國語暨普通話語料庫，其中包括兩位臺灣國語播音員之朗讀語料（一男一女，以下稱之為 TMS 和 TFS）以及一位北京女性播音員朗讀語料（以下稱之為 BFS）。其朗讀文本包含 15 段長段落，段落字數從 85 到 981 個音節不等，每一個段落平均字數 377 音節。這三批語料在語速表現上，即有很大的不同。其中說的最快的是臺灣男性播音員 TMS，平均音節長度為 182 ms，平均語速（包含停頓）每秒 4.3 個音節；最慢則為北京女性播音員 BFS，平均音節長度為 267ms，每秒 2.8 音節。而且 BFS 語料也顯示出：較長的音節時長與語流中較多停頓等慢語速的語流韻律特徵。三批語料的語速特徵摘要如表一所示：

特徵 \ 語料	TFS	TMS	BFS
音節總數	5655	5655	5483
音節平均時長 (ms) (僅考慮音段時長)	202.45	181.84	266.91
語速 (Syllable #/sec) (包含停頓)	4.23	4.33	2.816

表 1. 語料語速特徵分析

2.2 語料處理

首先以 HTK Toolkit 程式判定音段成分（聲母、韻母）再以 SAMPA-T（Tseng et al, 1999）自動標記音段邊界，而後以人工標記感知到的韻律邊界。不管是音段邊界標記人工檢查、抑或停延標記皆是由訓練有素的標記員所完成。

在實驗數值截取方面，首先使用 ESPTS Toolkit 取音段 RMS 值，每一音段以十個音框 (frame) 為一單位，計算 10 個音框的 RMS 平均值，少於 10 個音框的音段時長則直接取平均數。除此以外，為了排除因為每一段落在錄音過程中的些微改變所造成的韻律層級差異，每一段落中的 RMS 值皆經過標準化過程，得到 NRMS 值。如此一來，我們即可將研究焦點放在韻律單元內的音強組型的比率差異，而不是每一音節所取得的絕對音強數值。

2.3 分析過程

我們使用在（Tseng et al, 2004）的研究中提出的研究步驟及方法，也就是線性回歸統計中的逐步回歸技術（step-wise regression technique）來分析估算語料。在本研究中，我們發展出一個四階層的線性模型來預測發音人的音強及時長組型。此一階層式的回歸模型由下到上、一層一層地由下到上往上疊加，依序為：音節層、韻律詞層(PW)、韻律短層(PPh)以及韻

律句群層(BG)，最後把每一層的評估結果加總，得到最後的評估結果，再與實際語料比對，以檢視評估的比例是否有意義。全部過程都是要研究自變項(IVs)：音段成分(segment identities), 韻律架構(prosodic organization)與依變項(DVs)：音節時長(syllable duration), NRMS 之間的關係，其分析步驟摘要如下：(1.) DVs 直線轉換, (2.) 決定音段成分的分群 (3.) 建立線性回歸模型, (4.) 刪減無效之 IVs, (5.) 無法預測之殘差值(residual)由緊接著的更上層解釋之，重複步驟(3.)與(5.)。

爲了要有更好的預測，DVs 必須盡可能地呈現常態化分佈。DVs 與其轉換值(log, square root, etc)必須使用 N-plot 加以檢驗、比較何者接近常態化分佈，最接近常態化分佈者因而入選。於是，在接下來的回歸分析過程中，我們決定使用三批語料原始時長數值以及 NRMS 平方根。

間斷資料之線性模型使用 DataDesk 的部份平方和(partial sum of squares)(type3)建立之。在音節層中，將語料中所有的音段分成若干的群組，以減少回歸分析的複雜度，是必要的作法。以音強回歸預測來說，我們決定分成 10 組子音群與母音群；而時長回歸部份，則分爲 8 組。分組的標準系依音節層中每個 IVs 項目的 DV 平均值而決定。換言之，具有相近的時長/ NRMS 平均數之音段分在同一組，音節聲母類型、韻母類型以及字調類型則被當作自變項。同時，音節前後位置因素亦是考慮因素之一，另外，我們也考慮了其二因子交互作用，所以音節層之回歸分析可用下列公式表示：

$$\begin{array}{l}
 \text{DVs} \qquad \qquad \qquad \text{IVs} \\
 \text{Dur (ms)/ Square root NRMS} = \text{constant} + CTy + VTy + Ton \\
 \qquad \qquad \qquad \qquad \qquad \qquad + PCT + PVT + PrT + FCt + FVt + FIT \\
 \qquad \qquad \qquad \qquad \qquad \qquad + 2\text{-way factors of each factor above} \\
 \qquad \qquad \qquad \qquad \qquad \qquad + \text{Delta 1}
 \end{array}$$

回歸分析之後，將較小影響力的因素(p-value >0.1)予以排除，無法由音段成分所預測之殘差(Delta 1)，則移到在上一層分析。從 PW 層到 BG 層，我們以韻律結構當作自變項，所擷取出的係數表示韻律單元在特定音節位置的效應值（時長或音強的增加或減少）。

在 PW 層中，我們的目的在於瞭解 DV 是否受其在 PW 的位置所影響，因此 PW 層模型之公式表示如下：

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

無法由 PW 結構所預測之 PW 層的殘差：Delta 2，在緊接的上一層 PPh 層繼續分析。從我們最新的研究中發現，PPh 組織造成的效應，在 PPh 中的第一與最後 4 音節的位置最爲顯著，因此我們標記少於 8 音節的 PPh 的音節韻律序列資訊爲[PPh 的長度,該音節在 PPh 中的序列位置]，對大於或等於 9 音節者，我們逐一標記前 4 與後 4 音節，而其餘中間音節爲[M]，也

就是{[I1], [I2], [I3], [I4], [M]... [M], [F1], [F2], [F3], [F4]}的標記方式。如此一來，我們略過語料中長 PPh 不足的問題。PW 層之公式表示如下：

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

無法由 PW 與 PPh 所預測之 PPh 層的殘差：Delta 3，在緊接的上一層 BG 層繼續分析。我們標記 BG 中的第一與結尾 PPh，而其餘的 PPh 則一視同仁，當作中段的 PPh。在每一個 PPh 中，少於 7 音節的 PPh 的音節韻律序列資訊為[PPh 長度, 該音節在 PPh 中的序列位置]，對大於或等於 7 音節者，我們逐一標記前 3 與後 3 音節，而其餘中間音節則一視同仁，於是，BG 內的起首 PPh 的音節位置可標記為{[II1], [II2], [II3], [IM]...[IM], [IF1], [IF2], [IF3]}。而 BG 層之公式表示如下：

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

採用兩項數值評估預測結果：(1.) 相關係數 (Correlation Coefficient, r)，表示預測結果與原始資料的關聯程度。(2.) 總殘差誤差(T.R.E.)為殘差平方和與原始資料平方和百分比。T.R.E. 意謂著無法由底層之音節層到目前這一層為止無法說明的殘差誤差比率。

3. 實驗結果

本節包括二項實驗結果，一為音強比較；一為音節時長預測。我們首先說明音強預測的分析結果，然後是音節時長預測結果。在此二預測過程中，資料皆是由底層之音節層分析到最高的 BG 層。在每一層中，一比較三批語料的分析結果，以顯出其中的差異性。

3.1 音強比較

3.1.1 音節層

在這一層中，每一批語料的回歸分析使用不同的音段分群，表 2 到 4 描述了每一批語料中在音強回歸分析中的音段分群。其分群標準系根據每一音段成分的 NRMS 的平方根數值的平均數，將相近平均數的音段成分分在同一群。語料間的子音群分群比起母音群分群更加地近似。同時使用單因數和二因數之回歸分析，並刪減(忽略)p 值大於 0.1 之因素。表 5 到 7 顯示出剩餘因素之 ANOVA 摘要表。在所列因素中，CRM/CR 代表子音因素，VRM/VR 代表母音因素，而 Ton 代表字調因素。以 P 和 F 為首碼者代表前/後音節的貢獻因素。

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 2. TFS – 音強音段分群

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	la, au, a, ua

表 3. TMS – 音強音段分群

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 4. BFS – 音強音段分群

Source	df	Sums of Squares	Mean Square	F-ratio	Prob	Source	df	Sums of Squares	Mean Square	F-ratio	Prob
CRM	5	7.19217	1.43843	107.64	<0.0001	VRM*FTn	21	0.577025	0.027477	2.0561	0.0031
PVR	7	2.42141	0.345916	25.885	<0.0001	PTn*Ton	19	0.51406	0.027056	2.0246	0.0053
FCR*FVR	37	4.13819	0.111843	8.3691	<0.0001	CRM*VRM	25	0.674099	0.026964	2.0177	0.002
VRM*Ton	20	1.04553	0.052276	3.9118	<0.0001	VRM*FCR	34	0.838438	0.02466	1.8453	0.002
FTn	4	0.196686	0.049171	3.6794	0.0054	VRM*PCR	35	0.796716	0.022763	1.7034	0.0061
FCR*Ton	27	1.28257	0.047503	3.5546	<0.0001	CRM*PTn	25	0.527438	0.021098	1.5787	0.0335
FTn*Ton	16	0.735772	0.045986	3.4411	<0.0001	FVR*Ton	20	0.402831	0.020142	1.5072	0.0681
PCR*PVR	33	1.12543	0.034104	2.5519	<0.0001						
FVR*FTn	19	0.631157	0.033219	2.4857	0.0003	Const	1	1662.33	1662.33	1.24E+05	<0.0001
Error	5308	70.668	0.013364			Total	5655	113.783			

表 5. TFS – 音強 – 音節層因素之 ANOVA 摘要表

Source	df	Sums of Squares	Mean Square	F-ratio	Prob	Source	df	Sums of Squares	Mean Square	F-ratio	Prob
Ton	4	0.318537	0.079634	11.675	<0.0001	Ton*PTn	16	0.279218	0.017451	2.5585	0.0006
CRM*VRM	25	1.11392	0.044557	6.5324	<0.0001	CRM*PCR	30	0.466823	0.015561	2.2813	<0.0001
CRM	5	0.21493	0.042986	6.3021	<0.0001	CRM*FVR	25	0.350033	0.014001	2.0527	0.0015
VRM	5	0.189595	0.037919	5.5593	<0.0001	Ton*FTn	20	0.2714	0.01357	1.9895	0.0054
PTn	4	0.142174	0.035544	5.211	0.0003	PCR*PVR	30	0.37841	0.012614	1.8493	0.0033
PCR	7	0.234709	0.03353	4.9158	<0.0001	VRM*PVR	25	0.301156	0.012046	1.7661	0.0107
FCR*FVR	30	0.80378	0.026793	3.928	<0.0001	CRM*FCR	30	0.328262	0.010942	1.6042	0.0197

VRM*Ton	20	0.518252	0.025913	3.799	<0.0001	PVR*PTn	20	0.214018	0.010701	1.5688	0.0509
PCR*PTn	18	0.396491	0.022027	3.2294	<0.0001	VRM*PCR	30	0.307593	0.010253	1.5032	0.0384
CRM*Ton	19	0.400155	0.021061	3.0877	<0.0001	Ton*PCR	24	0.242049	0.010085	1.4786	0.0621
FCR*FTn	23	0.437687	0.01903	2.7899	<0.0001	FVR*FTn	20	0.197949	0.009897	1.4511	0.088
Ton*PVR	20	0.357034	0.017852	2.6172	0.0001	VRM*PTn	20	0.19212	0.009606	1.4083	0.1062
Ton*PVR	20	0.357034	0.017852	2.6172	0.0001						
FVR	6	0.105766	0.017628	2.5844	0.0168	Const	1	2286.83	2286.83	3.35E+05	<0.0001
Error	5178	35.3185	0.006821			Total	5654	77.7589			

表 6. TMS – 音強 – 音節層因素之 ANOVA 摘要表

Source	df	Sums of Squares	Mean Square	F-ratio	Prob	Source	df	Sums of Squares	Mean Square	F-ratio	Prob
Ton	4	4.71969	1.17992	115.67	<0.0001	CRM*Ton	19	0.53875	0.028355	2.7796	<0.0001
CRM	5	0.808926	0.161785	15.86	<0.0001	Ton*PCR	23	0.576665	0.025072	2.4578	0.0001
FCR	6	0.520646	0.086774	8.5063	<0.0001	PCR*PTn	19	0.467188	0.024589	2.4104	0.0006
PTn	4	0.340328	0.085082	8.3404	<0.0001	FVR*FTn	30	0.69715	0.023238	2.278	<0.0001
Ton*PTn	16	1.25228	0.078268	7.6725	<0.0001	VRM*FCR	30	0.626958	0.020899	2.0487	0.0006
Ton*FTn	16	0.978621	0.061164	5.9958	<0.0001	CRM*FTn	20	0.374104	0.018705	1.8336	0.013
PCR	6	0.317725	0.052954	5.191	<0.0001	CRM*FCR	30	0.558874	0.018629	1.8262	0.0039
VRM	5	0.254008	0.050802	4.98	0.0001	CRM*PCR	30	0.558667	0.018622	1.8255	0.0039
FTn	4	0.157095	0.039274	3.8499	0.004	VRM*FVR	30	0.546392	0.018213	1.7854	0.0053
CRM*VRM	23	0.883748	0.038424	3.7666	<0.0001	Ton*FVR	28	0.482779	0.017242	1.6902	0.013
PVR	6	0.179831	0.029972	2.9381	0.0073	VRM*Ton	20	0.310572	0.015529	1.5222	0.0635
PVR*PTn	20	0.571958	0.028598	2.8034	<0.0001	Const	1	1716.77	1716.77	1.68E+05	<0.0001
Error	5088	51.9034	0.010201			Total	5482	112.8			

表 7. BFS – 音強 – 音節層因素之 ANOVA 摘要表

3.1.2 PW 層

圖 1 到 3 顯示出在第三批語料中之 PW 層的音強組型。每一條線代表在一個韻律詞中特定位置音節所對應之回歸係數。Y 軸代表 NRMS 平方根的預測值，正值表示在特定位置音節的音強具大於平均值，負值則表示音強小於平均值。PW 層的一般組型清楚地顯示出語料間的差異。愈長的韻律詞，需要的能量越大，起始的音強越大。除此之外，以 2 音節與 3 音節的韻律詞來說，TFS 的回歸係數範圍為 0.016 和 0.019，TMS 的回歸係數範圍為 0.003 和 0.012，BFS 則是 0.028 和 0.054，顯示出兩批臺灣國語語料在音強方面，較北京普通話低。

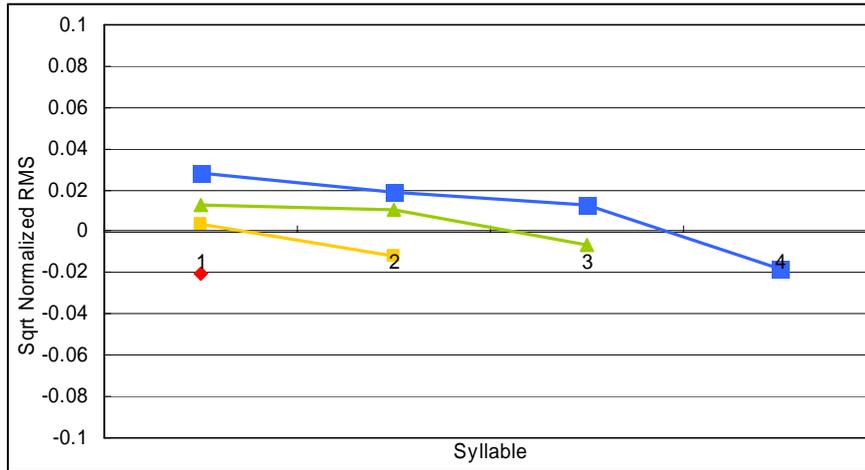


圖 1. TFS – 音強 – PW 層之回歸係數

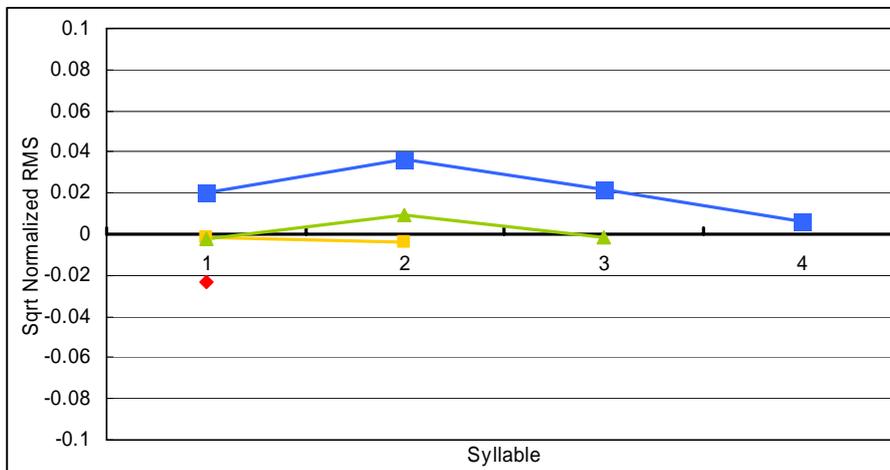


圖 2. TMS – 音強 – PW 層之回歸係數

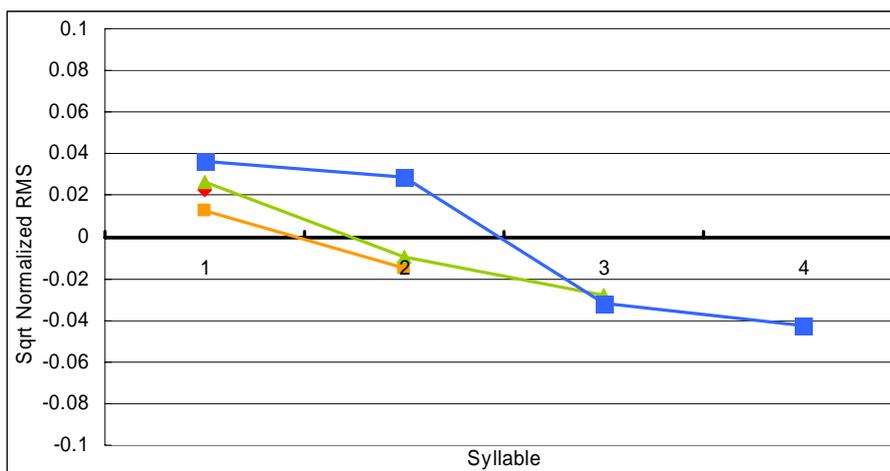


圖 3. BFS – 音強 – PW 層之回歸係數

3.1.3 PPh 層

圖 4 到 6 顯示三批語料中的 PPh 音強比率組型。每一條線代表特定音節數之韻律短語的音強組型。超過 9 音節的 PPh 以紫色線條顯示，PPh 的中間部份以第 4 音節代表，而清楚地表現出前 4 與後 4 音節的音強組型。

我們發現，愈長的 PPh，需要的能量越多。對超過 9 音節的 PPh 而言，TFS 的係數範圍為 0.101，TMS 的係數範圍為 0.088，與 BFS 的 0.118。比起 TMS，TFS 和 BFS 皆具有較快的下降斜率。此一 PPh 音強係數下降斜率的差異現象可能導因於男性與女性在呼吸上有些差異，如肺活量、肌肉強度等等。除此之外，也可看出臺灣國語與普通話的口音差異。以 PPh 尾來說，BFS 在 PPh 的最後一個音節表現出較強的音強，反觀 TFS 與 TMS 在 PPh 的倒數第 2 音節則表現出相當不清楚的突出 (prominence)。超過 7 音節的 PPhs 之回歸係數範圍，TFS 是 0.1，TMS 是 0.073，而 BFS 則為 0.096。超過 7 音節中段位置 PPhs 的回歸係數範圍，TFS 是 0.081，TMS 是 0.056，而 BFS 則為 0.094。

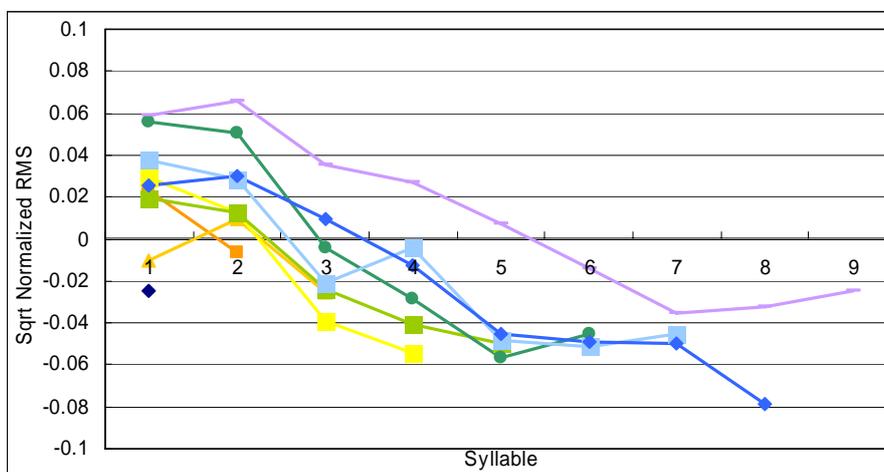


圖 4. TFS – 音強 – PPh 層之回歸係數

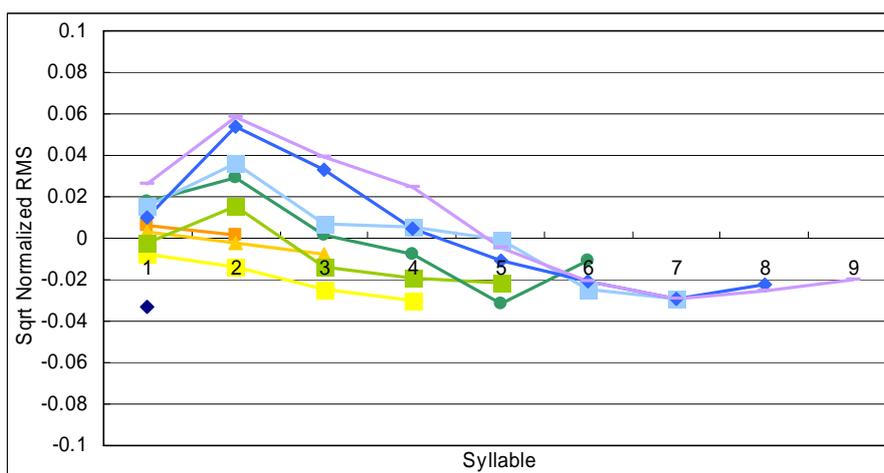


圖 5. TMS – 音強 – PPh 層之回歸係數

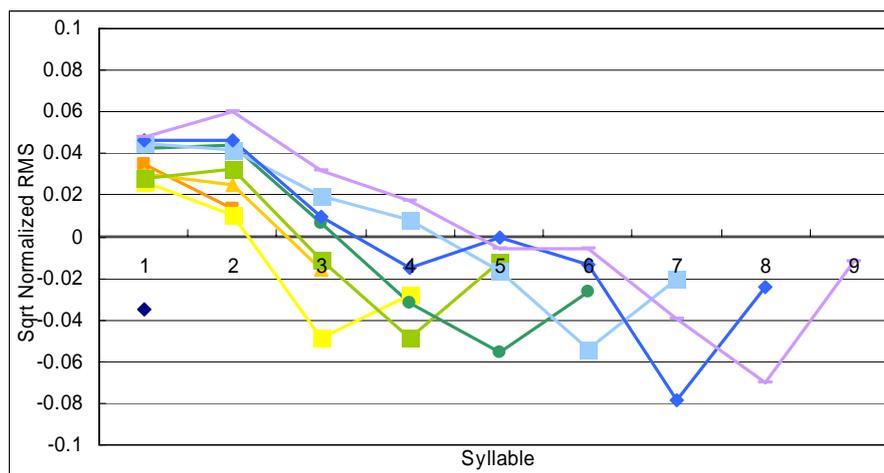


圖 6. BFS – 音強 – PPh 層之回歸係數

3.1.4 BG 層

在 BG 中，我們假設在 BG 中不同位置的 PPh 的貢獻度並不相同，而進行個別分析。BG 中第一與最後的 PPh 視為起首與結尾 PPh，而其他的則為中段 PPh。

BG 中的起首 PPh (BG-initial PPh)

圖 7 到 9 顯示在 BG 中的起首 PPh 音強組型。每一條線代表不同長度 PPh 的組型。超過 7 音節的起首 PPh 以紫色線條顯示，我們發現：相對於前文中已出現過的 PPh 組型，當起首 PPh 比較長的時候，它反而沒有較強的音強。一般來說，除了第一與最後音節以外，BG 中的起首 PPh 音節和所導出的預測平均值比較起來，具有較強的音強。

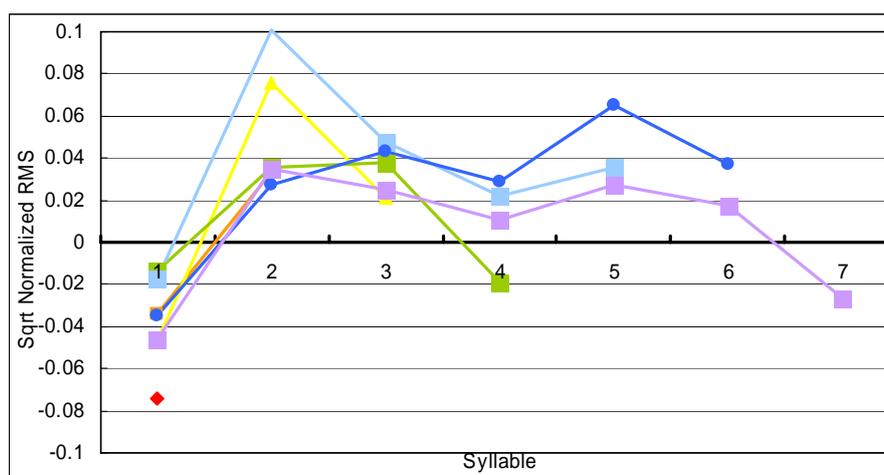


圖 7. TFS – 音強 – BG 中起首 PPh 層之回歸係數

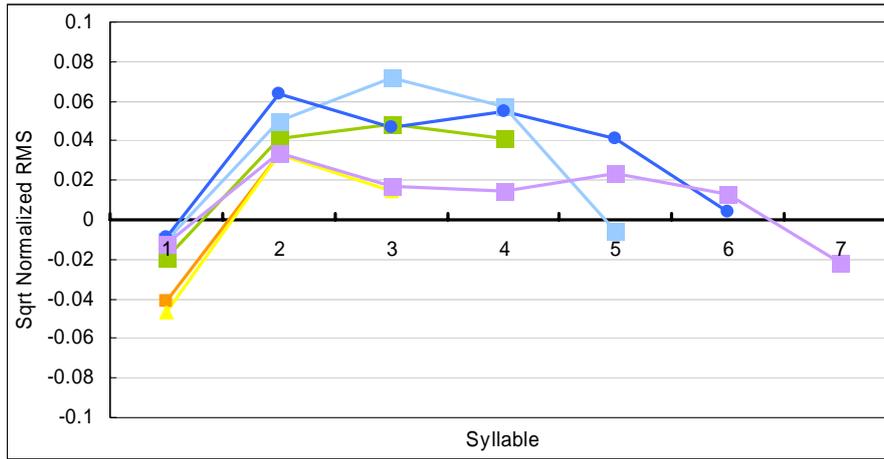


圖 8. TMS - 音強 - BG 中起首 PPh 之回歸係數

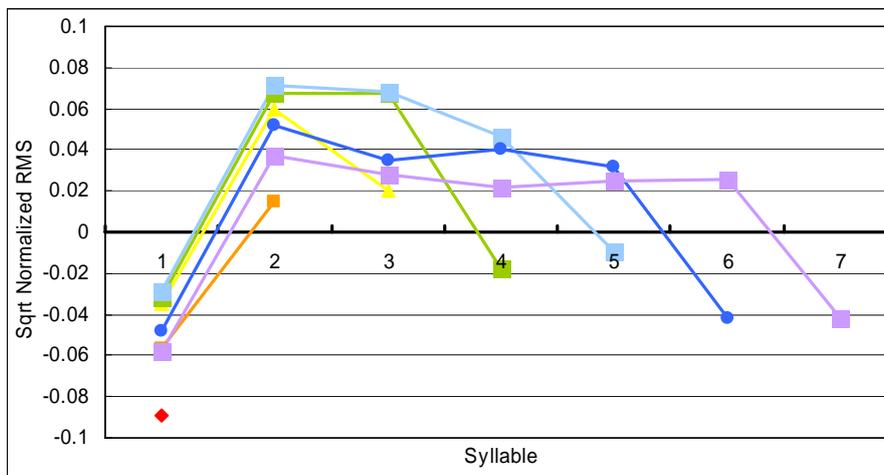


圖 9. BFS - 音強 - BG 中起首 PPh 之回歸係數

BG 中的中段 PPh (BG-medial PPh)

圖 10-12 顯示在 BG 中的中段 PPh 音強組型。與 PPh 層相類似，中段 PPh 有一個相似的下陷組型。

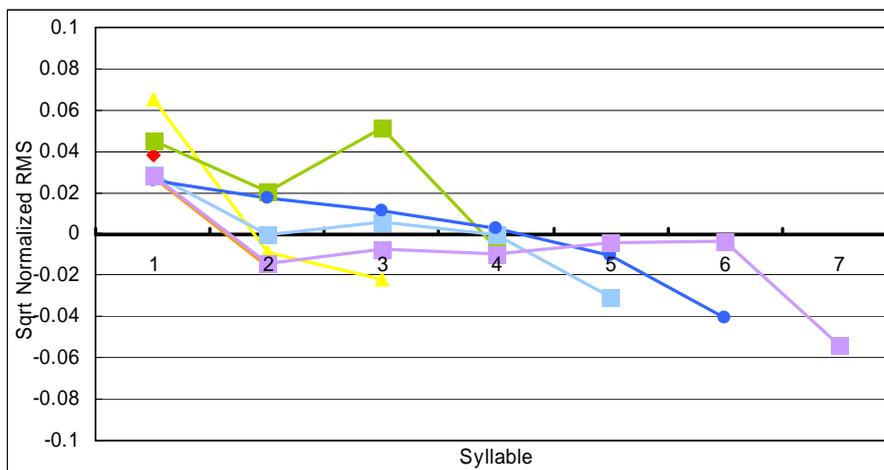


圖 10. TFS—音強—BG 中段 PPh 之回歸係數

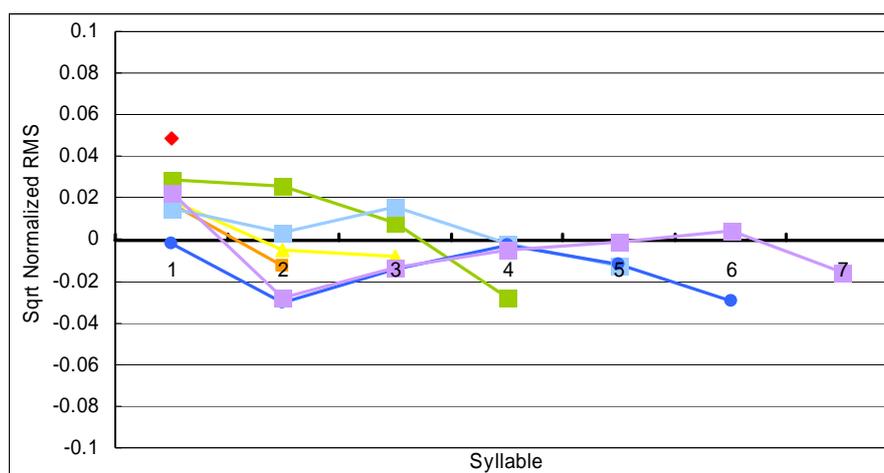


圖 11. TMS—音強—BG 中段 PPh 之回歸係數

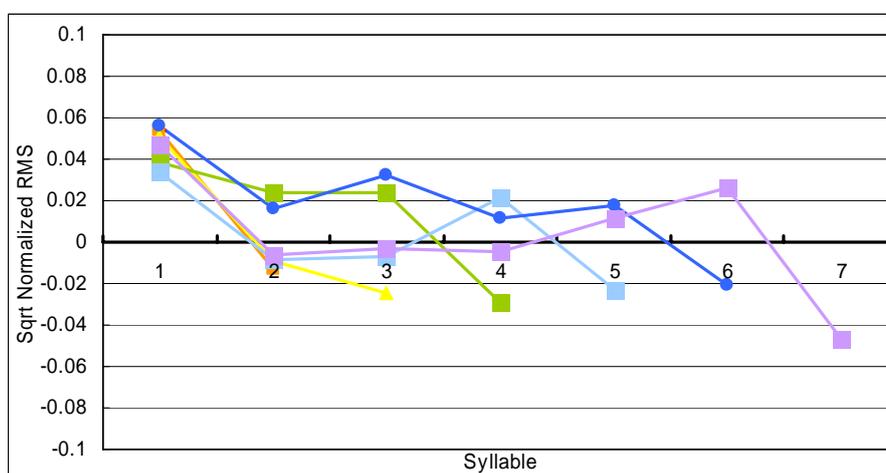


圖 12. BFS—音強—BG 中段 PPh 之回歸係數

BG 中的結尾 PPh (BG-final PPh)

圖 13-15 顯示在 BG 中的結尾 PPh 音強組型。所有的語料均顯示出，結尾 PPh 並沒有表現出如前所見的 PPh 音強減弱，反而呈現一相反趨勢：在 BG 結尾伴隨著一個較強的音節。較短的結尾 PPh 具有一較寬的係數範圍。對超過 7 音節的結尾 PPh 而言，係數範圍 TFS 是 0.077，TMS 是 0.059，而 BFS 則為 0.094。

比起男性發音人語料，兩批女性語料皆顯示出在 BG 中的結尾 PPh 具有較寬的回歸係數範圍。雖然在這層中結尾 PPh 的最後音節是正值，但並不意謂者在 BG 結尾 PPh 具有一強勁的結尾音節，這僅能暗示著在這一層級中，這些音節比起在其他 BG 單元裏的音節來說還要來得強一些。若以整體來說，全面性的預測考慮到所有層級的預測值，而 BG 結尾依然是變弱的。

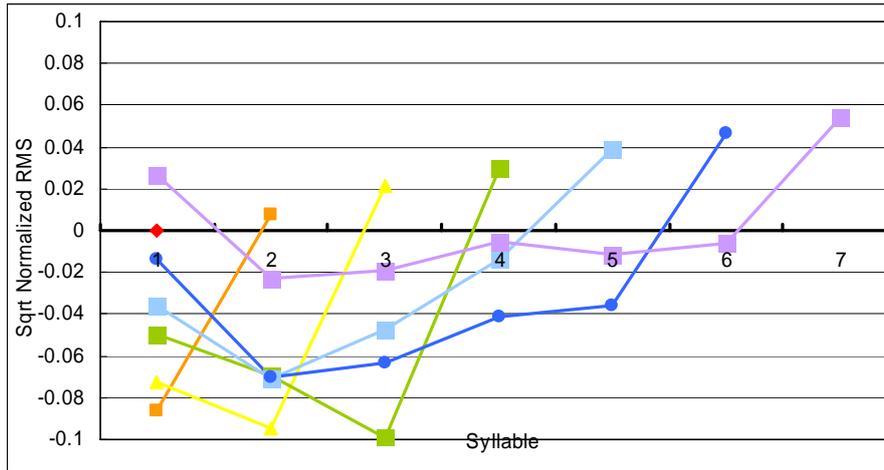


圖 13. TFS—音強—BG 結尾 PPh 之回歸係數

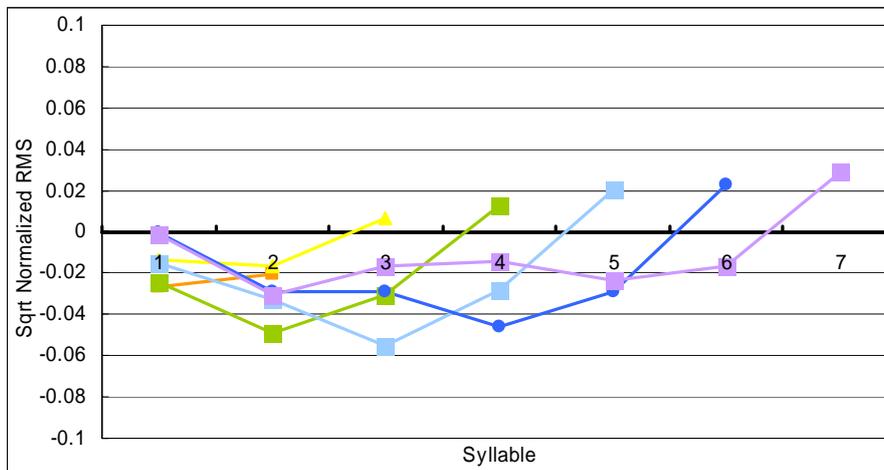


圖 14. TMS—音強—BG 結尾 PPh 之回歸係數

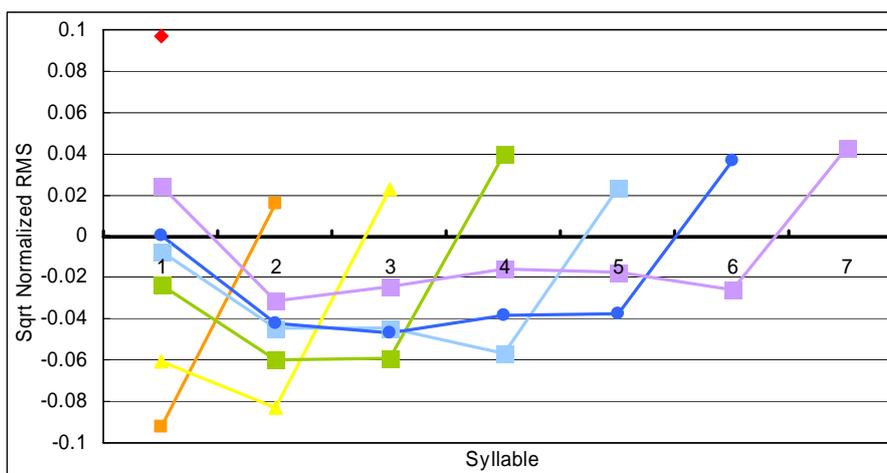


圖 15. BFS—音強—BG 結尾 PPh 之回歸係數

全面性預測之評估如表 8 摘要描述如下。值得注意的是：原始 DV (NRMS 值) 與總和預測 (預測總和之平方) 轉換值之間的評估值，因此我們可以推論：殘差誤差的增加可能導因於此非線性轉換過程。另一方面，相關係數 r 意謂著當增加更高一層時，預測力獲得了改善。在第三批語料中，BFS 具有最好的預測結果 (相關係數 $r = 0.814$)，而最差的是 TFS (相關係數 $r = 0.693$)。

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 8. 第三批語料音強預測之全面性評估

3.2 時長預測

在時長預測部份，除了音節層的音段分群外，所有過程都與音強的回歸分析一樣。以第三批語料來說，原始音段時長分佈稱得上接近常態化，在回歸分析中作為 DV。

3.2.1 音節層

在這一層中，每一批語料的回歸分析使用不同的音段分群，表 9 到 11 描述了每一批語料中在時長回歸分析中的音段分群。其分群標準系根據每一音段成分的平均時長數值，將相近平均數的音段成分分在同一群。同時使用單因子和二因子之回歸分析，並刪減(忽略)p 值大於 0.1 之因素。表 12 到 14 顯示出剩餘因素之 ANOVA 摘要表。在所列因素中，CTy/CT 代表子音因素，VTy/VT 代表母音因素，而 Ton/Tn 代表字調因素。以 P 和 F 為首碼者代表前/後音節的貢獻因素。

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 9. TFS—時長音段分群

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 10. TMS – 時長音段分群

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 11. BFS – 時長音段分群

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*FTn	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*PTn	25	95864.1	3834.56	1.8658	0.0056
CTy*Ton	25	183320	7332.81	3.5679	<0.0001	CTy*PTn	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*PTn	23	76195.2	3312.83	1.6119	0.0324
PTn	4	23319.3	5829.82	2.8366	0.023	FVT*FTn	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*FTn	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 12. TFS – 時長 – 音節層因素之 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*FTn	35	122844	3509.82	2.001	0.0004
PVT	9	74244.5	8249.38	4.7031	<0.0001	PCT*PTn	31	103806	3348.59	1.9091	0.0018
Ton*FTn	20	163881	8194.05	4.6716	<0.0001	PVT*PTn	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*FTn	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*FTn	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
PTn	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	PTn*FCT	39	92503.4	2371.88	1.3522	0.0709
FCT*FTn	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 13. TMS – 時長 – 音節層因素之 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	FTn	4	30036.5	7509.13	1.8858	0.11
Ton	4	96536.5	24134.1	6.0609	<0.0001	VTy*FTn	28	197720	7061.43	1.7734	0.0073
CTy*VTy	46	655573	14251.6	3.5791	<0.0001	VTy*PTn	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*FTn	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*FTn	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*PTn	30	185741	6191.37	1.5549	0.0275
FVT*FTn	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			

表 14. BFS—時長—音節層因素之 ANOVA 摘要表

3.2.2 PW 層

圖 16 到 18 顯示出在第三批語料中之 PW 層的回歸係數。這些圖當中，每一條線代表不同長度的韻律詞單元，而 Y 軸中每一點代表在一個韻律詞中特定位置音節所得來的回歸係數，這些係數表示相較於低一層所獲得的預測，該音節時長延長／縮短的程度。PW 層的一般組型像一勺子狀曲線——PW 的中間音節縮短而最後音節拉長。在第三批語料中，TFS 的延長／縮短效應最小，在四音節的 PW 中最短和最長音節相差 19.68ms 是最為顯著的差異；BFS 在 3 音節 PW 則有 39.52ms 的差異。

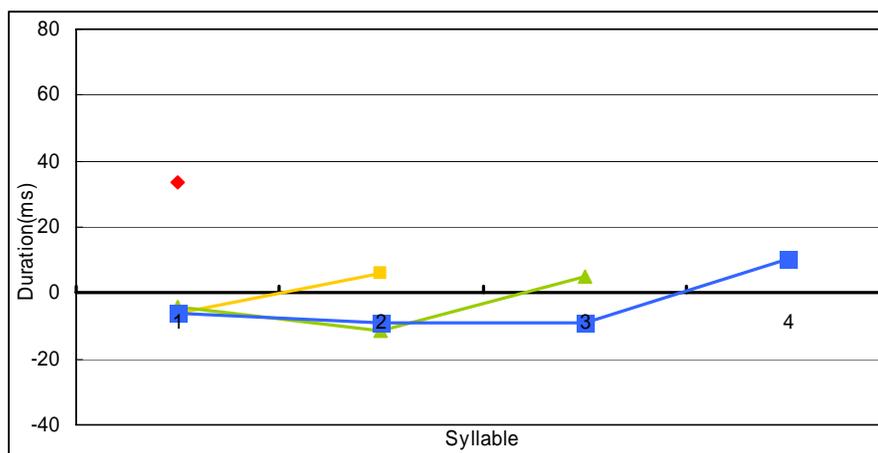


圖 16. TFS—時長—PW 層之回歸係數

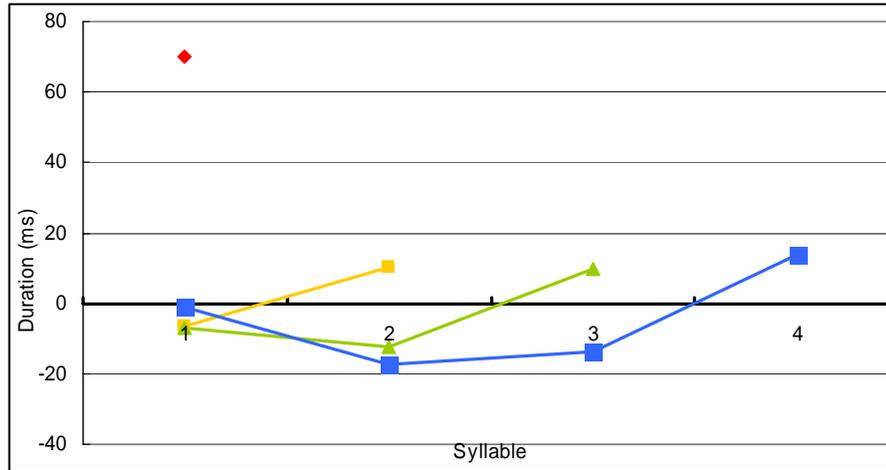


圖 17. TMS – 時長 – PW 層之回歸係數

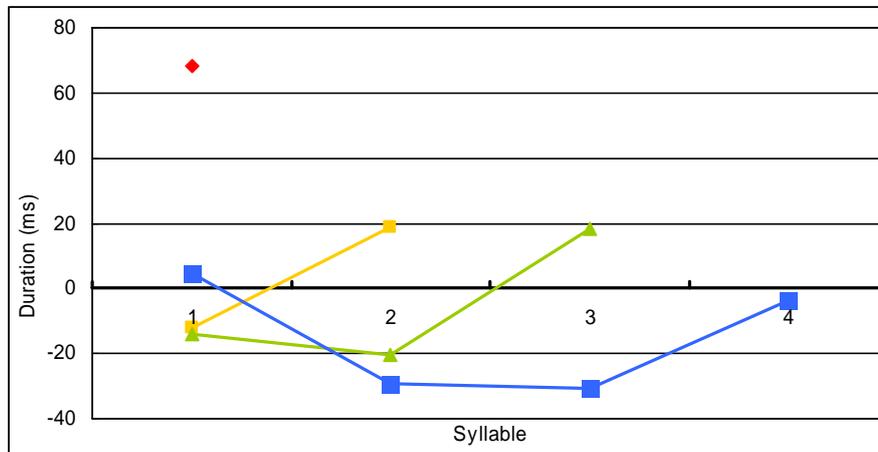


圖 18. BFS – 時長 – PW 層之回歸係數

3.2.3 PPh 層

圖 19 到 21 顯示了三批語料中 PPh 層的回歸係數。PPh 層的一般組型可從 PPh 中的最後 4 音節中所看出。不管什麼音節長度的韻律短語，從最後倒數第 4 個音節開始皆維持相同的組型，(倒數) 第三音節明顯最短，而最後音節則拉長，這便是眾所熟知的「短語結尾效應」(phrase final effect)，其實它就是不同韻律層級的邊界效應。TMS 和 TFS 在短語結尾表現出相同的組型，BFS 則有倒數第 3 音節最短，最後音節卻沒有拉長的傾向。

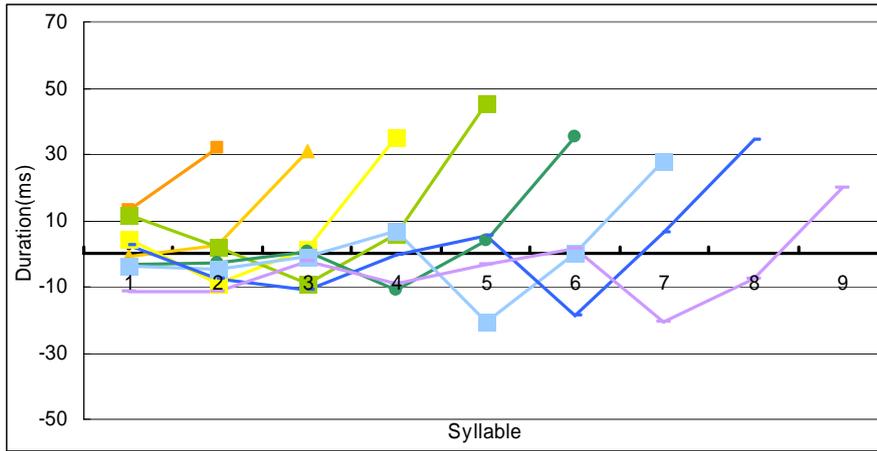


圖 19. TFS - 時長 - PPh 層之回歸係數

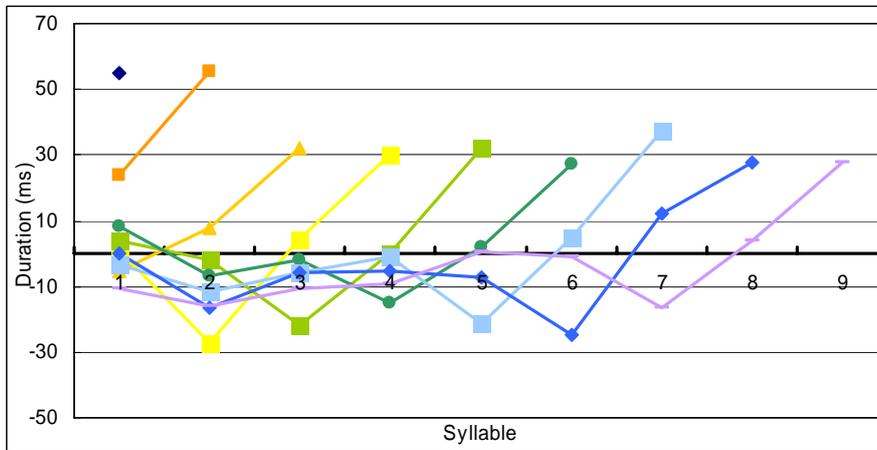


圖 20. TMS - 時長 - PPh 層之回歸係數

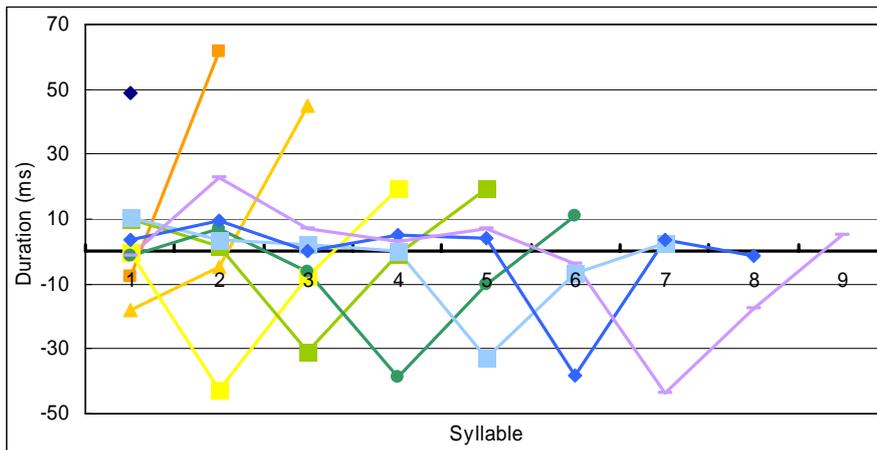


圖 21. BFS - 時長 - PPh 層之回歸係數

3.2.4 BG 層

BG 中的起首 PPh (BG-initial PPh)

三批語料中，由起首 PPh 所貢獻之一般時長組型十分相似。相較於 PPh 的預測，起首 PPh 短語末音節是更加地延長，起首 PPh 的最後音節均被延長了 23.85(TFS)到 29.55 ms (BFS)不等的長度。

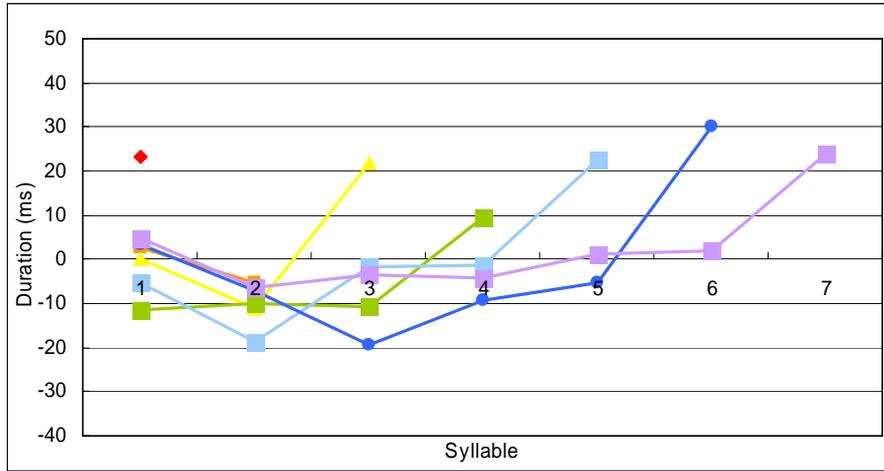


圖 22. TFS—時長—BG 中起首 PPh 之回歸係數

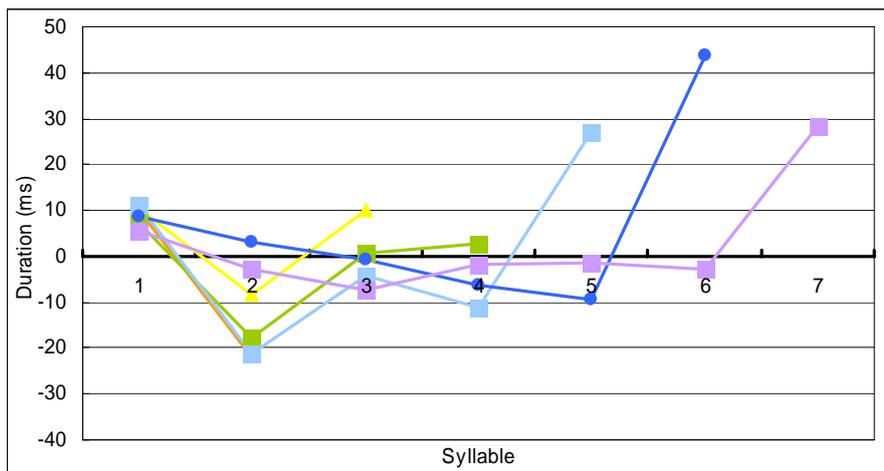


圖 23. TMS—時長—BG 中起首 PPh 之回歸係數

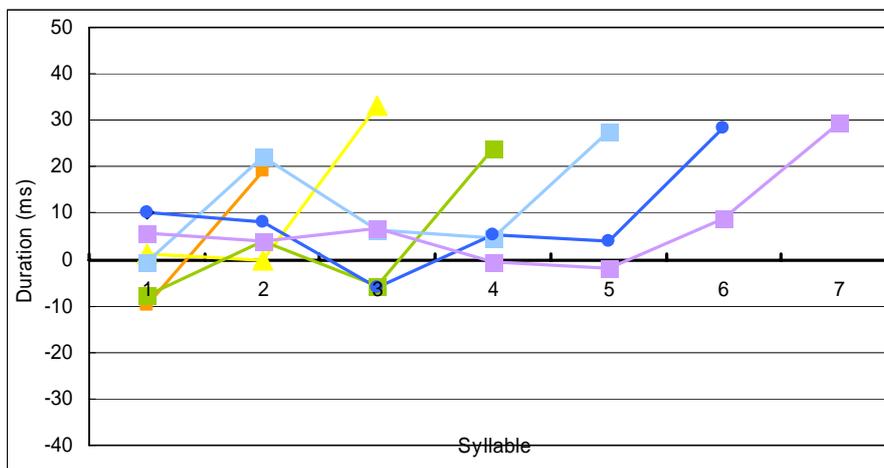


圖 24. BFS—時長—BG 中起首 PPh 之回歸係數

BG 中的中段 PPh (BG-medial PPh)

除了已知的 PPh 結尾延長外，在這一層中，BG 的中段 PPh 更加地延長。在這裏，必須注意起首 PPh 與中段 PPh 之間的差異：(1.) 起首 PPh 比中段 PPh 多了 10ms 左右，具有較大的延長效應。(2) 中段 PPh 的第一個音節時長縮短了 10 ms，這樣的現象在起首 PPh 並未發現。

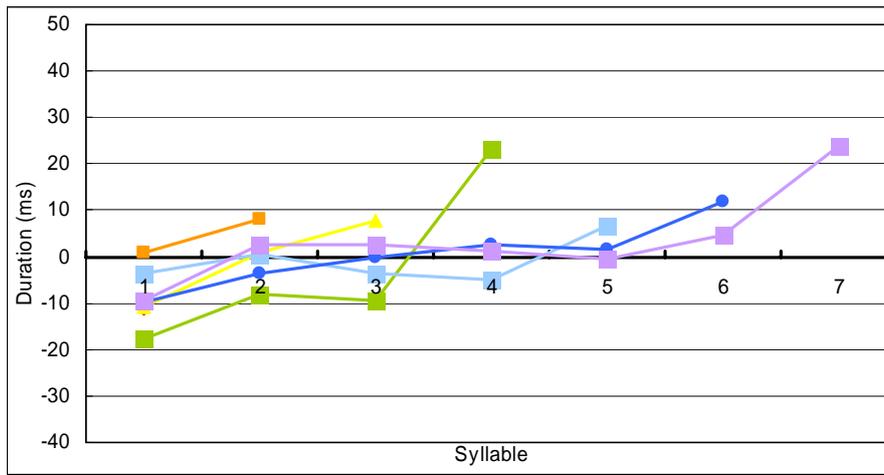


圖 25. TFS—時長—BG 中段 PPh 之回歸係數

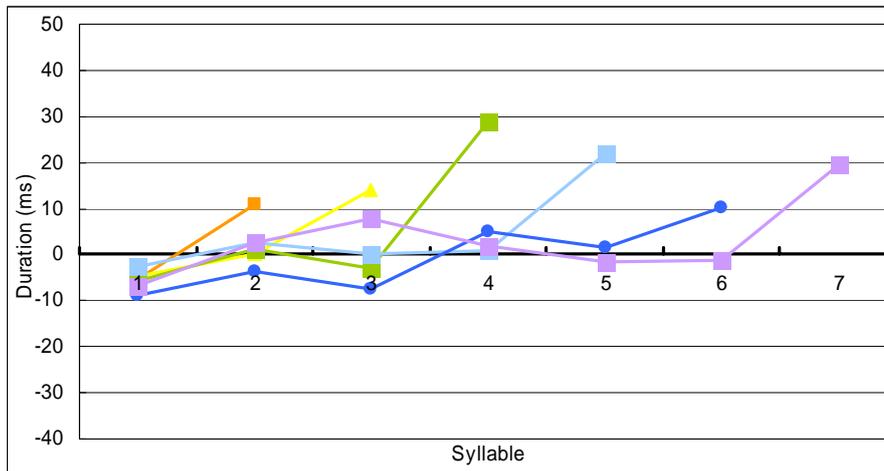


圖 26. TMS—時長—BG 中段 PPh 之回歸係數

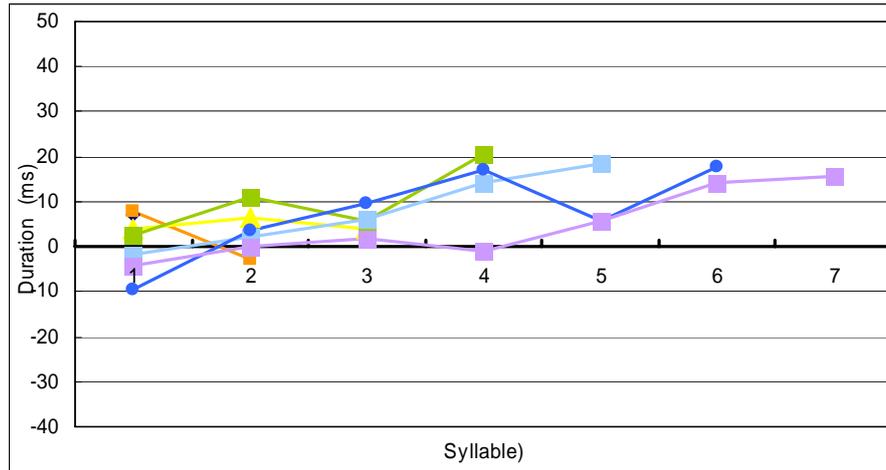


圖 27. BFS—時長—BG 中段 PPh 之回歸係數

BG 中的結尾 PPh (BG-final PPh)

BG 的結尾 PPh 的最後音節並沒有如起首與中段 PPh 一樣拉長，而是明顯地縮短。對超過 7 音節的結尾 PPh，TFS 縮短了 31.3ms，TMS 縮短 45.55ms，BFS 則為 18.8ms。然而，BG 最後音節的全面性效果應是預測結果的累積總和，其實際韻律表現依然呈現延長現象。

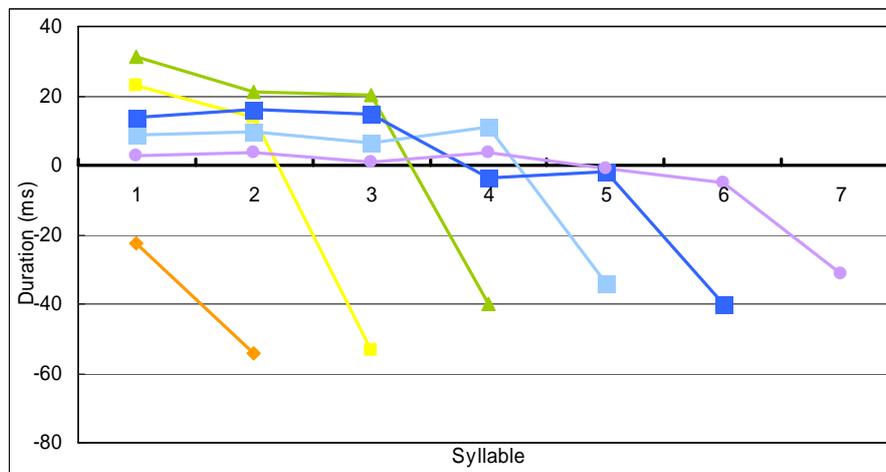


圖 28. TFS—時長—BG 結尾 PPh 之回歸係數

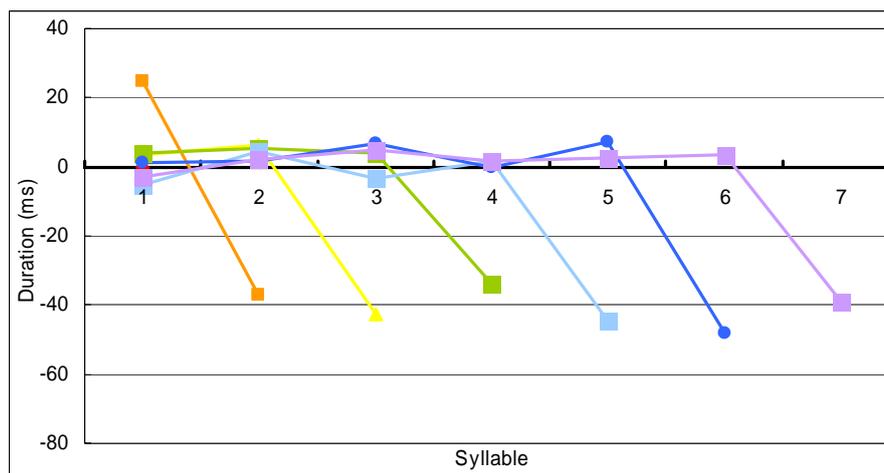


圖 29. TMS—時長—BG 結尾 PPh 之回歸係數

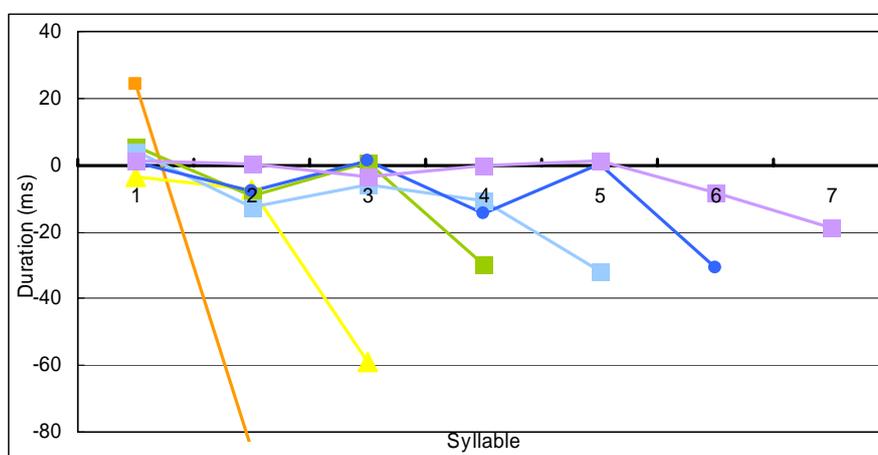


圖 30. BFS—時長—BG 結尾 PPh 之回歸係數

從以上結果，我們可以瞭解到：BG 層中不同位置的 PPhs 在已預測的 PPh 層之上貢獻了它本身的時長組型。表 15 列出各層的全面性預測評估值，全面性評估值之相關係數 r ，TFS=0.806，TMS=0.842 以及 BFS=0.819。

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

表 15. 三批語料時長預測之全面性評估

4. 結果分析與討論

我們發現了在階層式結構中的典型的韻律組型，如同時長組型一般，音強組型也可從音段成分與韻律結構中得到。除此以外，北京普通話與臺灣國語間的差異也可由每一韻律單元層

級中所看出。在本章中，我們將分別討論每一層的實驗結果。

音節層

從音節層的實驗分析結果，我們可以看到，將近 50% 的預測值可從音段成分因素所獲得。以音強預測來說，音節層預測了接近 52% 的原始資料是 (T.R.E = 47.65% in TMS and 48% in BFS)，最差的是 TFS，只解釋了 36.2%。至於時長預測部份，TFS 已被解釋了 56.12%，TMS 56.89% 以及 BFS 50.03%。因此，此一結果證明了為何韻律結果對語音合成在本質上是如此重要，語音合成若是只考慮到音節的串接與過渡，那將因缺少了更上層的韻律階層訊息而聽起來不自然。

PW 層

在第三批語料中，PW 的時長與音強組型是可預估獲得的，音強的逐漸下降與時長像帶把勺子狀的組型，均顯示了韻律詞組型之原型 (prototype)。比起其他兩者，較慢速的 BFS 語料的韻律組型是更加地明顯，而且 PW 層韻律組型被解釋的比率在 BFS 也是比較多的。以音強預測來說，BFS 的 T.R.E 減少了將近 8%，反而 TFS 只有 1.7% 與 TMS 的 0.63%。以時長預測來說，BFS 的 T.R.E 減少了將近 6.75%，反而 TFS 只有 1.45% 與 TMS 的 2.24%。除此之外，我們也認為北京普通話與臺灣國語之間口音上的差異，除了音段成分的差異，也有韻律組型的不同，北京普通話比起臺灣國語具有較明顯的輕重音與縮短/延長組型。

PPh 層

從第三批語料中所取得之 PPh 層的一般韻律組型，清楚地顯現出音強的下降與如 PPh 結尾一般的時長終止式 (cadence)。當時長組型中可以發現類似的型態時，北京普通話與臺灣國語的清楚差異也被發掘。相對於結尾的弱化，BFS 卻加強了短語的最後音節；在其時長預測部份，和 PW 層的結果比較起來，PPh 結尾也沒有更進一步拉長，反而結果是最後音節的前面一個音節的縮短。在三層韻律層級中 (PW, PPh, BG layer)，PPh 層在音強與時長的預測上說明了絕大部份，在 BFS 的音強預測上，PPh 層將近解釋了 17% 的 T.R.E 值。

BG 層

在 BG 的起首 PPh、中段 PPh 與結尾 PPh 都發現了清楚的韻律組型。以起首 PPh 來說，最後音節會拉長，而中間的音節比 BG 其他位置的音節還要念得較為強烈。以中段 PPh 來說，第一音節被加重與縮短，而最後音節則被弱化與延長。至於結尾 PPh，在時長與音強組型上都發現了與 PPh 相反的組型，最後兩音節呈現出重音上的輕重組型 (weakened-stressed pattern)。除此以外，和相同 BG 的其他 PPh 比較起來，最後音節同時也縮短了

5. 結語

在本文中，我們通過配合「階層式多短語語流韻律架構」的回歸統計模型，檢驗大量語料的聲學語音特性，在音強分佈和語流中音節在時程配置二方面做預測，發現當上層的韻律單元增加時，預測力也增加。進一步的證明語流的韻律最終輸出現象，無論是音強或時程配置，都是各韻律層級分層貢獻的總和。經由實驗比較兩岸口語在語流韻律，我們得到以下的結論：

- (1) 由各層分析顯示，語流由多層次的架構所組合而成，每一層皆有典型的特徵，證明口語語流韻律存在著一定的規劃性及規則性。
- (2) 韻律結構中，音強和時長都具有規則性，而以音節長短在時程配置上規則性最是明顯，在各階層間顯現類似的邊界效應及特徵。層層迭加的結果，是口語語流韻律與孤立句的句調在聽感上顯現極大差異的最主要原因。
- (3) 兩岸雖有口音的差異，在語流韻律的特性上大致還是相當一致的，亦即基本的韻律差異並不大。我們認為，必須進行更進一步的實驗，以厘清兩岸口音差異還有哪些語音層面，如音段及聲調系統、甚或是否還包括了語速差別。
- (4) 我們經由量化所獲得的韻律模型，可當做語音合成的基礎，作為語音科技開發方面的應用。

本研究除以上發現外，在研究方法方面，採用語料庫語音學，因而大幅的增加了語料的量及變異性；並因此配合了量化的處理。在研究課題方面，走出漢語語音學研究一向重字調或句調輕韻律及實際說話現象的框架，將口語篇章的特點，有系統的帶進語音學的研究。

6. 參考書目

- [1] 吳為章，田小琳。《漢語句群》。北京：商務印書館，2002年。
- [2] Tseng, Chiu-yu, Pin, Shao-huang and Lee, Yeh-lin (2004). "Speech prosody: Issues, approaches and implications" in *From Traditional Phonology to Modern Speech Processing* (語音學與言語處理前沿), edited by Fant, G., Fujisaki, H., Cao, J. and Xu, Y., Foreign Language Teaching and Research Press (外語教學與研究出版社), 417-437, Beijing, China.
- [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.