Coarticulation and prosodic hierarchy[i]

Yuling Zheng , Jianfen Cao**& Huaiqiao Bao**

*Institute of Ethnology and Anthropology & **Institute of Linguistics
Chinese Academy of Social Sciences
*zhyl-cass@263.net **jianfencao@hotmail.com *hqbao@ihw.com.cn

Abstract

This paper attempts to discuss the relationship between segmental coarticulation and prosodic structure in Mandarin Chinese. The discussion is mainly based on the EPG measurements to a set of phrases and short sentences produced in moderate speed. Main parameters measured in this study include (1) inter-gesture timing: temporal intervals overlapped between different articulatory gestures of V1#C2 at prosodic boundaries with different strength. (2) articulatory integrality: linguo-palatal contact state of different articulatory gestures for the related segments.

According to the data obtained so far, the strength of coarticulation is decreasing gradually with the increasing of prosodic level, and this tendency is consistent and identified in different speakers' speech.

These evidences indicate that segmental coarticulation and suprsegmental (prosodic) organization is originally indivisible in speech production. Therefore, in modeling of articulatory planning of speech production, one must take account of the close link between prosodic structure and segmental phonetic manifestation.

1. Introduction

Traditionally, coarticulation and prosody were treated as two independent areas either in theoretical research and spoken language processing. While various evidences come from natural speech strongly remind that segmental coarticulation and suprsegmental (prosodic) organization is originally indivisible in speech production. Therefore, in recent years, more and more interesting has been paid to the relationship between segmental processes and prosody controls of connected speech.

Experimental results from many languages have shown that the phonetic realization of consonant or vowel varies not only with phonemic identity, but also with suprasegmental factors such as stress / accent and prosodic position in the context. It indicates that segmental and prosodic planning in speech production is virtually not independent, since planning segmental articulation depends crucially on prosody [2]. Thus, the relative magnitude and timing of vocal articulatory gestures are the important correlates, which can cue the prosodic organization including prosodic boundary strength and accent status [3][4]. In Chinese, we have also found that articulation of the segments at prosodically stronger positions, i. e., at the edges of prosodic domains and the accented positions, are usually strengthened[1], and as an effect of counteraction, such strengthening likely result in relative weaker coarticulation between adjacent segments at prosodically stronger positions.

To make a deeper understanding, we have conducted a further investigation through a case study upon the variation of segment articulation at different prosodic positions. The preliminary results from EPG measurements have shown that either articulatory strengthening and coarticulation are highly sensitive to prosodic structure. However, the goal of this paper is just concentrate on how the strength of cross-boundary coarticulation varies depending on prosodic hierarchy. As for articulatody strengthening, it will be specified in Cao and Zheng in SP2006[1].

2. Test materials and methods

2.1. Test materials

In order to examine how the inter-segmental coarticulation sensitive to prosody structure in real speech, here the observing is concentrate on the cross-boundary segments in (c1)V1#C2(v2) sequences in different prosodic levels.

Test materials employed here are the following sentences:

说着/说着///来了/一个//走道儿的
(**Shuo1zhe0/shuo1zhe0///lai2le0/yi1ge4//zou3dao4er0de0**).

它们俩//就/商量/好了///说
(Ta1men0lia3//jiu4/shang1liang0/hao3le0///shuo1).

These sentences were selected from a speech corpus of EPG in Mandarin Chinese[5], which was produced by one male and one female native speaker. Here the prosodic hierarchy was annotated in terms of perceived boundary strength, the symbols of "/" , "//" and "///" represent the boundaries of prosodic word (PW) , prosodic phrase (PP) and intonation phrase (IP) respectively.

The examined tokens were chosen from above sentences, they are **"zhe0/shuo1", "ge4//zou3", "zhe0///lai2"** and **"jiu4/shang1", "lia3//jiu4", "le0///shuo1"**, where the boundaries between PWs, PPs and IPs are involved respectively.

2.2. Methods

The means used in this investigation is EPG, the main parameters involve: (1) inter-gestural timing, including durational ratio of intervals overlapped (RDIO) between different articulatory gestures in (c1)V1#C2(v2), and duration of interval (DI) between target V1 and v2, at PW, PP and IP boundaries respectively; (2) articulatory integrality represented by the RCA (ratio of linguo-palatal contact area) of the target V1, C2 and the transition of V1#C2 at PW, PP and IP boundaries.

3. Results and analysis

Generally, in the process of speech production, coarticulation is realized as articulatory gesture overlapping between adjacent segments and target reduction of the segments. The corresponding results, including the data on inter-gesture timing and articulatory integrality, are summarized in Table1 and Fig. 1, through which we can observe how the inter-segmental coarticulation is sensitive to prosodic structure in spoken Chinese.

3.1. Inter-gestural timing

Table 1 shows a comparison on inter-gesture timing of cross-boundary segments in different prosodic levels, in which the RDIO is durational ratio of interval overlapped between V1 and C2, and the DI is the duration of interval between the onsets of V1 and v2. The onset time of V1 and v2 were determined in terms of their linguo-palatal contact state.

Table 1. Articulatory data obtained from EPG measurements

Level / Item	PW		PP		IP	
Test token	(zh)e#sh(uo)	(j)iu # sh(ang)	(g)e # z(ou)	(l)ia # j(iu)	(zh)e # l(ai)	(l)e # sh(uo)*
RDIO: Male	0.139	0.148	0.108	0.135	-0.235	-0.206
Female	0.127	0.109	0.093	0.192	-0.270	0.113
DI: Male	221	207	238	270	460	521
Female	223	243	258	283	509	229
RCAv: Male	37	13	5	1	7	3
Female	37	25	13	0	3	21
RCAt: Male	29	18	6	8	6	3
Female	38	26	16	22	2	23
RCVc: Male	41	46	57	70	46	35
Female	48	49	61	71	43	48

* Here it is a PW boundary in female's utterance

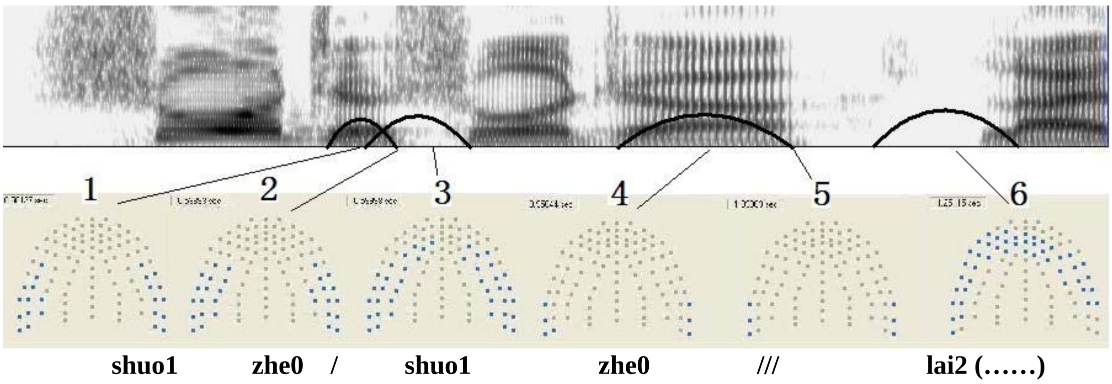

Fig. 1 Illustration on coarticulation between cross-boundary segments at different levels: **1** /e/'s target, **2** transition of /e#sh/ and **3** /sh/'s target at PW boundary; **4** /e/'s target, **5** transition of /e#l/ and **6** /l/'s target at IP boundary

Fig.1 gives the details both of the variation of overlapping interval and the variation of linguo-palatal contact state depending on the prosodic structure. In this picture, the upper shows the spectrogram, where the dark lines superposed on them illustrate the inter-gesture timing between different articulatory phases for tested segments. The bottom line presents the linguo-palatal contact state for V1, C2 and the transition of V1#C2.

From Fig. 1, We can see that, if there is gesture overlapping between V1 and C2, such as the case of **zhe/shuo,** then the corresponding dark lines are crossed, and the greater the overlapped interval, the larger the RDIO value is; if there is no obvious gesture overlapping between V1 and C2, for example in the case of **zhe///lai**, then the corresponding dark lines are separate, and the greater the distance between V1 and C2, the smaller the RDIO value is.

According to the RDIO data listed in Table 1, the value of RDIO is decreasing with the increasing of prosodic level, i. e., the overlapped interval is largest in PW boundary, then relatively smaller in PP boundary, and it is a minus value in IP boundary, apparently, it means that there is no obvious coarticulation between pre-boundary vowel and post-boundary consonant in this case.

At the same time, the data of DI indicate that the interval duration between target V1 and v2 also varies depending on prosodic hierarchy. Specifically, the higher the prosodic level, the greater the interval duration is, and the less the coarticulation exists.

3.2. Articulatory integrality

Articulatory integrality is another indicator which reflected the extent of coarticulation, here it is examined through linguo-palatal contact state and described in terms of RCA, the ratio of linguo-palatal contact area. The specific contact state can be observed from the relevant figures listed in Table 1, where the RCAv, RCAt, RCAc represent the linguo-palatal contact state of V1, C2 and transition of V1#C2 respectively.

From these data, a systematic difference on the extent of coarticulation can be observed in different prosodic levels. First, in the sentence **Shuo1zhe0/shuo1zhe0///lai2le0/yi1ge4// zou3dao4er0de0**, there are 3 /e/ vowels to be tested. Generally, for typical /e/, there should be few linguo-palatal contact during its articulation; however, the linguo-palatal contact state of the 3 /e/ is quite different. Specifically, in the case of **zhe/shuo,** i. e., at the PW boundary, the articulatory gesture of /e/ is obviously undergoing a interference from the planned gesture of the next segment /sh/. Consequently, the tongue position of this /e/ must be higher and much more front than that in **ge//zou** at PP boundary and in **zhe///lai** at IP boundary. It indicates that the target of /e/ in **zhe/shuo** is quite undershoot. On contrary, at the IP boundary**,** the articulation of **/e/** in **zhe///lai** is very integrity, its tongue position is highly closed to that of /e/ in isolation. According to the relevant data in Table 1, the RCA value of the 3 /e/ can be ranked as: /e/ in **zhe/shuo** > in **ge//zou** > in **zhe///lai**. It reveals that the articulation of vowel /e/ is most undershoot at PW boundary and more integrallty at PP and IP boundaries. Apparently, such articulatory difference is mainly caused by the strength different of boundaries that they located.

In addition, the same tendency is also found in **Ta1men0lia3//jiu4/shang1liang0/hao3le0/// shuo1**, too. In which, the tongue position of target vowel /u/ in **jiu/shang,** i. e., in the case of PW boundary, is clearly interference from the gesture of the adjacent /sh/; while the gesture of vowels /a/ in “**lia//jiu**” at PP boundary and /e/ in “**le///shuo**” at IP boundary are almost not interfered from the gesture of next consonant.

Another interesting phenomenon may be worth to mention that, according to the female’s speech, the prosodic hierarchy of sentence **Ta1men0lia3//jiu4/shang1liang0/hao3le0/shuo1** is some what different from that of the male speaker’s speech. Specifically, the boundary between **le/shuo** is a PW boundary, instead of an IP boundary like that in the male’s speech. Thus, its RCA is quite different from that of corresponding male’s one. Obviously, it is result in the strength change of prosodic boundary, a stronger interference from the gesture of followed /sh/ is introduced. This phenomenon further reveals that the extent of coarticulation is crucially sensitive to the variation of prosodic structure.

Table 2. Result of correlation analysis

Proximity Matrix

	Correlation between Vectors of Values					
	PL	DI	RDIO	RCAv	RCAt	RCAc
PL	.000	.900	-.824	-.753	-.850	-.053
DI	.900	.000	-.954	-.531	-.708	-.405
RDIO	-.824	-.954	.000	.382	.673	.550
RCAv	-.753	-.531	.382	.000	.847	-.380
RCAt	-.850	-.708	.673	.847	.000	.021
RCAc	-.053	-.405	.550	-.380	.021	.000

This is a similarity matrix

Table 3. Result of probability analysis

Test Statistics[a,b]

	DI	RDIO	RCAv	RCAt	RCAc
Chi-Square	9.035	6.369	7.731	8.788	8.469
df	2	2	2	2	2
Asymp. Sig.	.011	.041	.021	.012	.014

a. Kruskal Wallis Test

b. Grouping Variable: PL

3.3. Data analysis

As a case study, the data obtained so far are relatively limited. To make sure whether the findings observed here is credible or not, a further statistical analysis were conducted as well. Considering of the small tokens and less speakers, we examined all the data by using the NPar Tests. The result summarized in Table 2 shows that all the variables tested in this study are highly correlated with the distinction of prosodic level (PL) though the linguo-palatal contact area of C2 (RCAc) show a lower correlation.

However, according to the result summarized in Table 3, all these prosodic strength-dependent articulatory differences are quite significant. Consequently, we would claim that, all these coarticulatory correlates may be served as one of the markers for the prediction of prosodic hierarchy. Of course, as what pointed by Yohann Meynadier[4], they still cannot be considered as very good "predictors" of the hierarchical nature of prosodic constituent structure, because they appear to be largely both speaker and phonotactic context-specific. The similar situation also found in this study. However, we believe that there must be certain rule hidden in such complex phenomenon. Exploring the rule should be the next research object.

4. Conclusions

The preliminary EPG evidences obtained from this study show that, all the pre-boundary vowels (V1) at PW level strongly coarticulate with post-boundary consonants (C2); while there are less cross-boundary coarticulation between V1 and C2 at PP and IP levels; especially in IP level, the pre-boundary vowels are almost not interfered from the post-boundary consonants. These evidences suggest a close link between segmental articulation and prosodic control in speech production.

The results of data analysis show that there exist a negative correlativity between the extent of coarticulation and the height of prosodic hierarchy, it offers an account of that the high the prosodic level, the less the cross-boundary segmental overlapping, and the more sufficient segmental articulation.

Accordingly, we would claim that prosodic structure also can be cued by inter-segmental coarticulation in terms of its extent distinction.

5. Main References

[1]Cao, Jianfen. & Yuling. Zheng, 2006. Articulatory strengthening and prosodic hierarchy, Proc. of SP2006, Dresden, Germany, May 2-5, 2006.

[2]Keating, P. & Shattuck-Hufnagel, S., 2002. A prosodic view of word form encoding for speech production. UCLA working Papers in Phonetics, 101.

[3]Keating, P. A., 2004. Phonetic encoding of prosodic structure. *UCLA Working Papers I Phonetics, No. 103*.

[4]Meynadier, Y., 2004. Gradient linguopalatal variations due to a 4-level prosodic hierarchy in French. *9th Conference on Laboratory Phonology*, Urbana-Champain, IL, USA, June 2004.

[5]Zheng, Y. and Zu, S., 2001. The dynamic EPG corpus of Mandarin speech and the flat for corresponding research. *Acoustics and Electronic Engineering*, pp. 3-13.

[i] Supported by the National Fund of Natural Sciences, NO. 10374117