MODELING THE SPEECH RHYTHM OF BEIJING CHINESE IN THE CCI FRAMEWORK

Na Zhi, Pier Marco Bertinetto & Chiara Bertini

Laboratorio di Linguistica, Scuola Normale Superiore, Pisa, Italy
na.zhi@sns.it; p.bertinetto@sns.it; c.bertini@sns.it

ABSTRACT

This study describes the application of CCI (Control/Compensation Index) [3, 4] to a corpus of spontaneous Beijing Chinese. CCI is a modification of the PVI algorithm [8], devised to provide an improved representation of the rhythmic tendencies of natural languages. The CCI algorithm was previously applied to the modeling of Italian [3, 5]. The present findings refer to Beijing Chinese.

Keywords: Control/Compensation Index (CCI), speech rhythm, spontaneous Beijing Chinese

1. INTRODUCTION

This study is part of an ongoing project concerning the rhythmic behavior of different languages by means of the CCI algorithm, also in comparison with results obtained by other methods. In previous work [3, 4, 5], these measures were applied to corpora of spontaneous and read Pisa Italian, and is currently being applied to spontaneous German and Brasilian Portuguese. This paper deals with the speech rhythm of Beijing Chinese as contrasted with (Pisa) Italian.

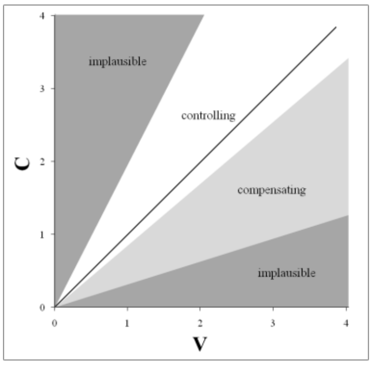
The basic idea of CCI (Control/Compensation Index) consists of relativizing the PVI model [13] to the number of segments composing each V and C interval. The duration of each interval is divided by the number of segments in it, according to the following formula, where *m* stands for ‘number of intervals’ (vocalic or consonantal, as separately considered), *d* for ‘duration’ (in ms), *n* for ‘number of segments within the relevant interval’:

$$CCI = \frac{1}{10(m-1)}\sum_{k=1}^{m-1}\left|\frac{d_k}{n_k} - \frac{d_{k+1}}{n_{k+1}}\right| \qquad (1)$$

Due to its very conception, the CCI model takes into account not only the speech durational behavior, but also the degree of phonotactic complexity as reflected in the number of segments composing each V and C interval. The model aims at providing a more realistic representation of the rhythmic tendencies of natural languages. Indeed, it makes a big difference, in terms of phonotactics, whether a C interval contains a single C, or a geminate, or a C cluster, and the same holds for the V intervals (with a single V, a long V, or a V sequence). In ideal situations, a perfectly “controlling” language should present tendentially identical C and V local durational fluctuations, thus falling on the bisecting line, or at least it should exhibit stronger stability in the V intervals. By contrast, “compensating” languages should fluctuate more in the V than in the C component, for they presuppose substantial V-reduction. Fig. 1 (which modifies the initial proposal in [3]), depicts these ideal situations, obviously to be interpreted *cum grano salis*. Since the use of CCI is still in the initial phase, one should allow for some approximation in the formulation of the relevant predictions.

One should keep in mind that the Control/Compensation model, in its full realization, is based on a two-level conception [4]. This paper only refers to level-I (phonotactics). Future work will address level-II (phrasal), where phrase prominences play a rhythmically crucial role.

Figure 1: Schematic representation of the major rhythmic types according to the CCI model.



2. APPLYING CCI TO SPONTANEOUS BEIJING CHINESE

2.1. The corpus

2.1.1. Data selection

The materials used in this study are utterances stemming from the *Chinese Spontaneous*

Conversation Corpus [9, 10], produced by the Chinese Academy of Social Sciences, Beijing. The corpus consists of 12 units of daily conversations between native Beijing speakers. Each unit is a 1-hour dialogue between two speakers of the same gender.

For this study, 607 utterances produced by 7 speakers (4 females and 3 males) were selected from the corpus and manually labeled at the segmental level by one of the authors (a native speaker).

The utterances were selected according to the same criteria used for the Italian corpus. They were tendentially neutral from the intonational point of view, and consisted of at least 8-syllables. In addition, they should not present disturbing phenomena such as speakers' overlap, laughters, background noise, etc. Moreover, each utterance-final syllable was trimmed, in order to minimize the durational distortion due to final lengthening. Finally, all sentence-initial syllable-onsets consisting of a plosive [p p^{h} t t^{h} k k^{h}] or an affricate [ts ts^{h} tʂ tʂ^{h} tɕ tɕ^{h}] were trimmed, due to the impossibility to exactly measure the initial C interval's duration. Occasionally, even an initial nasal- [m n], liquid- [l], or fricative-onset [s f ʂ ʐ ɕ x] had to be trimmed due to very small signal energy.

2.1.2. Regular vs. irregular sound elisions

- Deletions, insertions and shifts

In spontaneous speech, one can detect several unpredicted sound changes, consisting of deletions, insertions and shifts. In our corpus, deletions by far outnumbered insertions and shifts.

As for shifts, the following phenomenon deserves particular attention. Occasionally, the phonological constitution of a segment was modified, such that a phonological V was realized as a C. Namely, a high back vowel /u/, as part of the nucleus of an onsetless syllable, was sometimes produced as a labial-dental [v], especially in sequences such as /uən/, /uan/, /uaŋ/, /uai/ and /uei/. One and the same speaker could interchangeably articulate the two sounds. There are 32 such cases, while in 80 cases /u/ was not changed.

Only one case of sound-insertion was found in our data. By contrast, we found many cases of deletion. We divide them into two categories: "regular" and "irregular" sound deletions.

- Regular sound deletions

Regular deletions involve frequent words presenting allomorphy, involving routine casual speech pronunciations. Due to frequency of usage, these may become an intentional target.

We found that Beijing speakers often produce the bound morpheme 们 (a grammatical particle often used in plural personal pronouns) as *m* instead of the citational *mən*. E.g., *uo mən* 'we' could be realized as *uo m*, *ta mən* 'they' as *ta m*, *ni mən* 'you' (plural) as *ni m*. There were 63 such cases, and only 11 cases where *mən* was not shortened. We thus consider *m* an intentional deletion, due to the weak status of this plural particle in casual speech, except for instances of contrastive or emphatic focus.

A similar case involved the bound-morpheme 么 *mə*, to be found in words such as 什么 *ʂən mə* (question word 'what'), 怎么 *tsən mə* (question word 'how'), or 那么 *n*a *mə* (conjunction 'then'). The speakers often omitted the vowel. There were 36 such cases, and 52 cases where the citational syllable was pronounced. In sum, we considered *m* an acceptable pronunciation for both *mən* and *mə*.

- Irregular sound deletion

Irregular deletion are cases where, due to hypoarticulation, the phonetic output did not correspond to the speaker's phonological intention. This is strictly related to speech rate and casualness of speaking style. We found 61 cases of irregular V deletion, and 396 cases of irregular C deletion (out of which, 323 were onset deletions and 73 were coda deletions). This is not surprising, considering the relatively high speech rate in the corpus.

2.1.3. Phonetic and phonological analysis

The following table details the data selected from the corpus. By "phonological segments" we mean the intended phonemes, by "phonetic segments" the phonemes actually produced, due to irregular deletions:

Table 1: Number of phonological and phonetic segments analyzed in this study.

	phonological segments	phonetic segments
vowels	10264	10203
consonants	6976	6580
total	17240	16783

In the CCI model, the phonetically inaudible, but phonologically intended segments, might be assigned zero duration, for one may assume that they were part of the speaker's articulatory plan, irrespective of the actual phonetic output. However, we performed a double computation:

"phonological" and "phonetic". In the latter case, only the actually produced segments were taken into account.

2.1.4. Vowel clusters

In previous studies on Chinese syllable structure [1, 2, 6, 7, 11, 12, 13], no final agreement was reached on the phonological status of the initial segment /i/, /y/ or /u/ in multi-vowel sequences such as /ia/, /iau/, /iou/, /ya/, /ye/, /yu/, /ua/, /uo/, /uə/, /uai/, /uei/. When the initial segment is /i y/, there appears to be no convincing evidence to consider it a glide. We thus regard the relevant diphthongs and triphthongs as V sequences, and treat all segments involved as belonging to a tautosyllabic V interval. In other words, since all Vs in such sequences belong to one and the same syllable, they do not give rise to a hiatus (which in the CCI model would involve two adjacent V intervals).

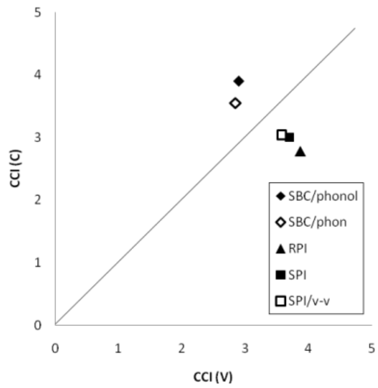
As for the V sequences beginning with /u/, the latter segment may correspond (as noted in §2.1.2) to two alternative phonological categories: either the vowel /u/, or the consonant /v/. In the CCI computations, we considered such [u] and [v] as belonging to a V or C interval, respectively. In the latter case, a citational two/three-vowel sequence was treated as a CV(V) sequence.

The number of two- and three-vowel sequences in our corpus was 2360 and 626, respectively. Their sum amounts to 45.0% of the V intervals. They thus represent a very substantial phonotactic component. By contrast, the number of Italian diphthongs was 145, i.e. 5.2% of the V intervals (in all, there were 2870 V and 3621 C segments). Since the glide status of the relevant segments is universally admitted for Italian, they were assigned to the preceding or following C interval. However, in order to better compare our present and previous results, we ran a further analysis (cf. SPI/v-v in fig. 2), where such segments were treated as part of the V interval, in analogy with the Chinese data.

3. RESULTS AND COMPARISON

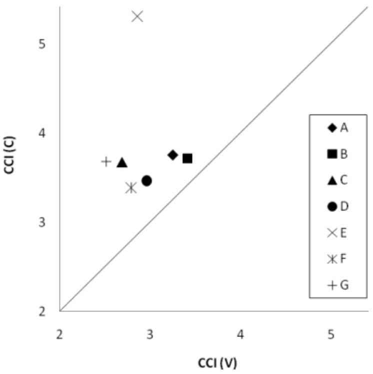
In contrast to Italian, Beijing Chinese falls, according to the CCI computations, to the left of the bisecting line (fig. 2). This indicates a strongly controlling behavior. Interestingly, this arises despite the presence of so many two- and three-V sequences, which might in principle introduce a great deal of compression among the relevant V segments. Despite this, and despite the virtual absence of C clusters, there is more "local fluctuation" (as measured by CCI, in analogy with PVI) in the C than in the V intervals. This is evidently due to the longer duration of aspirated consonants as opposed to non-aspirated ones. The higher stability of the V as opposed to the C component is a strong indication of a controlling tendency, particularly evident in the phonological analysis (cf. § 2.1.3), due to the non-negligible number of irregular C deletions. No substantial difference emerged, by contrast, between the SPI and SPI/v-v analyses.

Figure 2: Rhythmic tendencies of Spontaneous Beijing Chinese (SBC) in two analyses (*phonol*ogical and *phon*etic), Read Pisa Italian (RPI) and Spontaneous Pisa Italian (with glides assigned to C intervals [SPI] or V intervals [SPI/v-v]).



The disaggregated behavior of the individual speakers is shown in figure 3. As it happens, apart from one speaker (but limited to the C component), there was considerable inter-individual consistency:

Figure 3: Individual differences of speech rhythm. Capitals from A to G indicate the 7 Beijing speakers.



Speech rate is known to exert a crucial role in the rhythmical behavior of natural languages, as often noted in the specialized literature. To check for this factor, the Beijing speakers' productions were divided into three tempo-groups, as measured in segments per second (segm/s):

(I) low: <16.1 (average: 14.3);
(II) medium: >16.1, < 18.8 (average: 17.4);
(III) high: ≧18.8 (average: 20.5).

There were 205 utterances in group I, 203 in group II, and 199 in group III. The respective projections (according to the phonological analysis) are shown in figure 4 alongside the corresponding Italian data. To strengthen the comparison, the Italian data are shown with glides analyzed as part of either C (filled squares) or V intervals (white squares). As it happens, speed accelerations exerted a tendentially linear effect on both V and C intervals, as predicted for controlling languages.

Figure 4: Slow (1), medium (2) and fast (3) speech-rate for SBC, as compared with SPI and SPI/v-v.

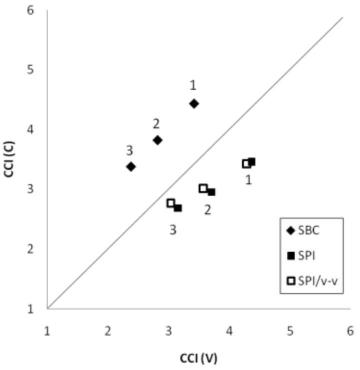
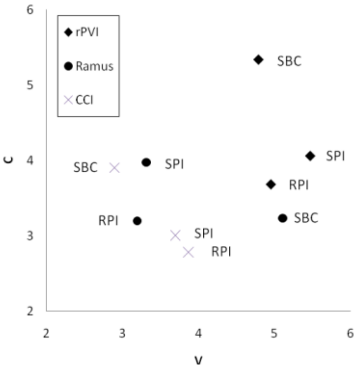


Figure 5 shows the results obtained by using the rPVI (raw PVI) and Ramus algorithms. The results of nPVI (normalized PVI), on the other hand, are not reported for they would hardly fit into the same graphic. Needless to say, the various metrics should only be compared in topological terms, rather than with respect to the actual values shown in the figure, for these largely depend on the multiplying factor applied for convenience. As it happens, rPVI separates SBC from SPI/RPI on the C axis, while the Ramus measure emphasizes the V component. CCI, by contrast, separates SBC from SPI/RPI on both axes, thus proving to be the most sensitive measure. As compared with rPVI, CCI reveals that the V component of SBC exhibits a striking stability even when relativized to the number of segments in each interval. This indicates that the V intervals' duration tends to be linearly correlated to the number of Vs included in each interval, thus suggesting a definitely controlling behavior, largely immune from V-reduction. As for the Ramus approach, it does capture the salience of the V intervals; however, it only yields an overall picture rather than the actual speech dynamics, due to its static nature which prempts any detailed phonotactic interpretation. Besides it does not capture the different behavior of the two languages with respect to the C component (see the comment to fig. 2).

Figure 5: Speech rhythm tendencies of SBC, SPI, and RPI as analyzed with rPVI**,** Ramus and CCI algorithms.



4. REFERENCES

[1] Bao, Z. 1990. Fanqie languages and reduplication. *Linguistic Inquiry* 21, 317-350.

[2] Bao, Z. 2001. The Asymmetry of the medial glides in middle Chinese. *Proc. 7th International and 19th National Conferences on Chinese Phonology* 11, 7-27.

[3] Bertinetto, P.M., Bertini, C. 2008. On modelling the rhythm of natural languages. *Speech Prosody 4th International Conference* Campinas, 427-430.

[4] Bertinetto, P.M., Bertini, C. 2010. Towards a unified predictive model of natural language rhythm. In Russo, M. (ed.), *Prosodic Universals. Comparative Studies in Rhythmic Modeling and Rhythm Typology*. Naples: Aracne, 43-77.

[5] Bertini, C., Bertinetto, P.M. 2007. Propezioni sulla struttura ritmica dell' italiano basate sul corpus semispotaneo AVIP/API. *Atti del 4º Convegno AISV* Cosenza.

[6] Duanmu, S. 1990. *A Formal Study of Syllable, Tone, Stress and Domain in Chinese Languages.* Ph.D. Dissertation, MIT.

[7] Duanmu, S. 2007. *The Phonology of Standard Chinese*. (2nd ed.). New York: Oxford University Press Inc.

[8] Grabe, E., Low, E.L. 2002. Durational variability in speech and the rhythm class hypothesis. *Laboratory Phonology*. Berlin: Mouton de Gruyter, 7, 515-546.

[9] Li, A.J. 2002. Chinese prosody and prosodic labeling of spontaneous speech. *Proc. of 1st International Conference on Speech Prosody,* Aix-en-Provence.

[10] Li, A.J., Yin, Z.G., Wang, M.L., Xu, B., Zong, C.Q. 2001. A spontaneous conversation corpus CADCC. *Oriental COCOCSDA Workshop,* South Korea.

[11] Wan, L.P. 2002. *Alignments of Prenuclear Glides in Mandarin*. Taipei: Crane Publishing.

[12] van de Weijer, J., Zhang, J. 2008. An X-bar approach to the syllable structure of Mandarin. *Lingua*. 118, 1416-1428.

[13] Yip, M. 2003. Casting doubt on the onset-rime distinction. *Lingua*. 113, 779-816.