# Rhythm in Mandarin Chinese and Italian: the Role of Sentence Accents

*Pier Marco Bertinetto, Chiara Bertini, Na Zhi*

Scuola Normale Superiore, Pisa, Italy

p.bertinetto@sns.it, c.bertini@sns.it, na.zhi@sns.it

## Abstract

Two perception experiments were run with Mandarin Chinese and Italian speakers. The aim was to comparatively examine the rhythmical role played by sentence accents. This study pursued previous work on the rhythmic features of Chinese and Italian as analyzed by means of the Control/Compensation model, according to which rhythm produces effects at two levels: level-I (phonotactical) and level-II (sentential). While previous work addressed the situation at level-I, the present study dealt with level-II, to check the respective strength of the accentual and syllabic oscillators.

**Index Terms**: rhythm, Control/Compensation model, sentence accent.

## 1. Introduction

### 1.1. The Control/Compensation model

In previous work [1, 2, 3, 13], the rhythmical tendencies of Mandarin Chinese and Italian were studied and compared by means of the Control/Compensation model (henceforth **CC**). CC is a model developed at Laboratorio di Linguistica of Scuola Normale Superiore, Pisa, aiming at improving the understanding of the natural languages' rhythmical tendencies. The terms "control" and "compensation" refer to the different degree of flexibility that the speakers of various languages exhibit, as stemming from specific and acquired articulatory routines. This articulatory attitude normally transfers to the pronunciation of foreign languages; this paper, however, only concerns L1.

CC differs from most rhythmical models in its very architecture, for it is a two-level model [2]. Level-I concerns the rhythmical consequences of the phonotactic organization of speech, while level-II concerns the sentential rhythmical structure, where sentence prominences play a crucial role. In order to understand the specific claims put forth by CC, one should best consider level-I first.

### 1.2. The architecture of rhythm: Level-I

CC incorporates claims developed in the articulatory phonology framework. It assumes that the sequencing of consonants (**C**) and vowels (**V**) may be viewed as the coupling of two oscillators. The C and V oscillators are very much in-phase when the phonotactics is very simple, but their mutual relationship becomes increasingly complex the more complex is the language phonotactics. CC claims that the mean duration of the segments composing any given C or V interval is a better rhythm predictor than the mean duration of the intervals themselves. This radically departs from many models proposed in the last two decades.

CC aims at providing a more realistic representation of the rhythmic tendencies of natural languages. It makes indeed a big difference, in terms of phonotactics, whether a C interval contains a single C, or a geminate, or a C cluster. The same holds for the V intervals, which may contain, e.g., a single V, a long V, or a V sequence in hiatus.

To flesh out the above assumptions, CC exploits at level-I a modified version of the PVI algorithm [5], whereby the interval duration is relativized to the number of segments composing it, according to the CCI formula (= CC Index); where *m* stands for 'number of intervals' (vocalic or consonantal, as separately considered), *d* for 'duration' in sec., *n* for 'number of segments within the relevant interval':

$$CCI = \frac{100}{(m-1)}\sum_{k=1}^{m-1}\left|\frac{d_k}{n_k}-\frac{d_{k+1}}{n_{k+1}}\right| \qquad (1)$$

The formula incorporates the essential merit of PVI, i.e. its being a dynamic model sensitive to the actual sequencing of segments and syllables, as opposed to static models such as that of Ramus [10]. But in addition, CCI takes into account an essential component of the actual constitution of the various C and V intervals, namely the number of segments they contain.

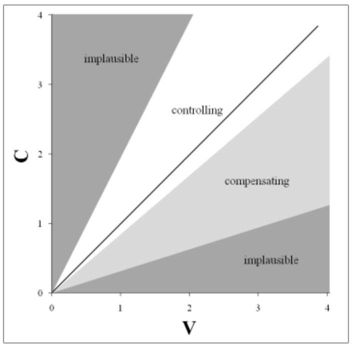

Figure 1: *Schematic representation of the major rhythmic types according to the CC model.*

Fig. 1 – which modifies the initial proposal in [1] – is an abstract representation of the CC predictions at level-I. In a purely ideal situation, a perfectly controlling language should present tendentially identical C and V local durational fluctuations, thus falling on the bisecting line, or it should at least exhibit stronger stability in the V than in the C intervals. By contrast, strongly compensating languages should fluctuate more in the V than in the C component, due to substantial V reduction. Needless to say, this should be interpreted *cum grano salis*: since CC is still in the testing phase, one should allow for some approximation in the initial predictions. One feature that needs to be taken into account in future work is distance from the origin of the Cartesian plane. Conceivably, typical controlling languages project at a shorter distance than heavily compensating ones. This points to an important feature of the CC charts, as opposed to those generated by most alternative models: projections should be read in mathematical, rather than merely topological terms.

Formula (1) was applied to corpora of spontaneous and read Pisa Italian (henceforth **PI**), and to a corpus of spontaneous Beijing Chinese (**BC**). Fig. 2 (adapted from [13]) shows that both languages exhibit a controlling behavior, with BC more strongly characterized than PI in this respect.

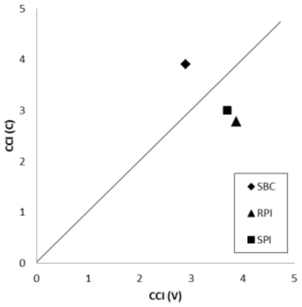

Figure 2: *Rhythmic tendencies of Spontaneous Beijing Chinese (SBC), Read Pisa Italian (RPI) and Spontaneous Pisa Italian (SPI).*

### 1.3. The architecture of rhythm: Level-II

Level-II is also based on the coupling of two oscillators: the accentual and the syllable-peak oscillators. Adopting suggestions by O'Dell & Nieminen [9], the relative coupling of the oscillators is expressed by formula (2), inspired by the so-called "Averaged Phase Distance" theory: *I* stands for 'duration of the inter-accentual intervals', *n* for 'number of syllable-peaks', $\omega_1$ and $\omega_2$ for the angular frequency – or velocity – of the two oscillators (assumed to be constant and expressing the 'natural' rhythm of each oscillator), and *r* indicates their relative-strength parameter, i.e. the degree of dominance of the accentual phrase over the syllable:

$$I(n) = \frac{r}{r\omega_1 + \omega_2} + \frac{1}{r\omega_1 + \omega_2} n \qquad (2)$$

This can be rewritten as $I = a + bn$, where the coefficients *a* and *b* stem from the linear regression computed on the data, with *r* in (2) expressing the ratio between *a* and *b*. The formula relates the duration of the inter-accentual interval to the number of syllable-peaks comprised in it. If *r* is greater than 1, then the overarching accentual oscillator predominates; if *r* is less than 1, the subordinated syllable-peak oscillator prevails. This is a revisitation of an important aspect of the Pikean model of rhythm, except that in CC this is just one component of rhythmic structure, rather than the whole of it. In the CC framework, dominance of the accentual oscillator translates as propensity towards compensation, for it indicates a relatively high degree of flexibility in the temporal organization of the syllables. By contrast, dominance of the syllabic oscillator points towards the controlling behavior.

The rhythmic behavior at level-II is based on sentence intonation. Sentence accents (henceforth **SA**) are prominent syllables playing a salient syntactic-pragmatic role. In practice, they are the subset of word-stresses that are perceived as prominent at sentence level. Obviously, not all SAs are equally salient (contrastive accents are the exception that confirms the rule). Nevertheless, one can assume that, normally, every sentence presents some SA. Their number depends on various factors, like: pragmatic underlining, speech rate, syntactic structure (e.g., interrogative sentences emphasize the SA component more than declarative sentences with neuter intonation).

On top of this, one should consider that SAs are not equally salient in every language. It is for instance common opinion, among Chinese phonologists, that Chinese does not present particularly salient SAs as compared, for instance, with Italian. But the problem is not that simple in Italian either, for many SAs are fairly weak, so as to leave in doubt as for their actual presence. Ultimately, identifying SAs is a tricky perceptual issue. Unless for the most prominent ones, their identification is a matter of degree, i.e. a probabilistic matter.

## 2. Perception experiments

In order to check for the presence of SAs in BC and PI, two perception experiments were run with native speakers. The task was to identify the SAs in a number of sentences. The BC materials were extracted from the *Chinese Spontaneous Conversation Corpus* [7, 8] and from the *Chinese Multext* corpus [6]; the Italian materials stemmed from the *API/AVIP corpus* [11], supplemented with a selection of the original sentences read by a subset of the same speakers [12]. The aim was twofold: (a) checking the intersubjective convergence in the task of SA localization; (b) measuring the rhythmical inclination at level-II of the languages under investigation.

### 2.1. Experiment design

The participants had to listen to a series of sentences, pointing out all SA-bearing syllables according to their own perception. They could listen to each utterance at will, and were free to modify their decisions before moving on to the next sentence.

#### *2.1.1. Stimuli*

For both languages, the experimental set consisted of 60 utterances: 30 spontaneous and 30 read ones. The length of each utterance was 7～16 syllables; very long utterances were avoided to reduce the task difficulty. The utterances presented a fairly neutral intonation contour, with no emphatic stress. To minimize the participants effort, the stimuli were divided into 3 lists: A, B and C, each consisting of 20 utterances (10 spontaneous + 10 read).

#### *2.1.2. Subjects*

15 Chinese and 15 Italian speakers, all native, took part in the experiment. They were randomly arranged in three groups. The list x group schema was as follows: Group I = Lists A+B; Group II = Lists A+C; Group III = Lists B+C.

### 2.2. Method

Here follows the instruction sheet:

*This study aims at examining the sentence accents in fluent speech. You will have to point out the syllables bearing a sentence accent based on your own perception. You are going to listen to one utterance at a time. The sentences are 40. You can either do the whole task at one time or divide it into two sections. After the first 20 sentences you will be offered the chance to rest.*

*A short training for locating the accents is necessary. Two sample utterances will be used to this purpose.*

*UTTERANCE 1* [with strong emphatic stresses]

*Click on the arrow to listen to the recorded version. In this sentence, the highlighted syllable is "mu", which stands out most from the others. If you agree on this, then click on the corresponding box in the top tier for confirmation. You can now move on to the next sentence by clicking on the arrow at the bottom right corner of the screen.*

*UTTERANCE 2* [with neutral intonation]

*Most of the sentences in this experiment look like this, with no strong* *emotions or emphases. You will however be provided with some hints, pointing out the potentially accented syllables, as shown on the bottom tier. You are expected to identify the most salient syllables according to your perception, by selecting the most prominent syllables among the ones highlighted. You are however free to select any syllable, even among the ones that are not highlighted. Try and click on any box of the top tier to verify that you can activate your own selection.*

*For any given sentence of the experiment, you will first have to click on the arrow to listen to it. On the screen you will see a sequence of boxes on two tiers. The bottom tier contains a few highlighted syllables, while the top tier is empty. By clicking on any empty box of the top tier you will make your own choice. You can listen to each sentence as many times as you like, and can modify your decisions until you are happy with them. Click on the bottom right arrow to move on to the next sentence.*

## 2.3. Results and discussions

Table 1 presents the percentages of perceived SAs in read and spontaneous speech for both languages. The data of both languages are presented with respect to four criteria (60%, 70%, 80% and 90%), indicating the degree of intersubjective convergence on SA identification (6 to 9 speakers out of 10, respectively). Each output probabilistically defines the inter-accentual intervals' boundaries. Obviously, the number of identified SAs decreases from left to right – as indicated by N in table 1-2 – alongside with the tightening of the constraint: i.e., more syllables are highlighted with the 60% criterion than with the much tighter 90% criterion.

A notable feature is the unequal number of SAs detected in the two languages. Table 1 shows that PI turned out to be much more stable than BC, both w.r.t. the various criteria and by comparing the two speaking styles. With the most tight criterion (90%), the intersubjective convergence remained quite substantial among the Italian participants, whereas it became marginal among the Chinese participants. One can thus safely conclude that SAs are part of the prosodic competence of the Italian speakers, whereas they should be regarded as a fairly elusive feature in Chinese prosody.

| | spontaneous | | | | read | | | |
|---|---|---|---|---|---|---|---|---|
| | 60% | 70% | 80% | 90% | 60% | 70% | 80% | 90% |
| BC | 29,2 | 21,6 | 13,0 | 5,3 | 37,6 | 25,5 | 13,8 | 3,8 |
| PI | 26,1 | 23,1 | 19,5 | 14,0 | 26,2 | 25,0 | 21,9 | 17,1 |

Table 1: *Percentage of highlighted syllables.*

In terms of the CC model, the *r* values of formula (2) shown in table 2 altogether point to a definitely controlling behavior. In this connection, one should note the steady decrease of *r* in both languages from the 60% to the 90% criterion. The dominance of the syllable-peak oscillator increases alongside the inter-accentual interval's duration. This phenomenon is particularly striking in the BC data. Evidently, with shorter inter-accentual intervals the varying segmental composition imposes its own rights, yielding some amount of syllabic compensation, whereas with longer intervals such local variations are smoothed out. This further emerges by inspecting the correlations between the inter-accentual intervals' duration and the number of segments they include, as shown in tables 2-3. As it happens, the correlation is almost perfect in BC, although it also yields a fairly robust value in PI (Kendall's *tau* was used with non-normal distributions). One is invited to conclude that the extraordinary stability of the syllabic oscillator in BC is a compensation for the relative weakness of the SA component. Presumably, in languages like Chinese level-II plays a reduced role, with the phonotactic component doing most of the rhythmic job.

| Beijing Chinese | | | | |
|---|---|---|---|---|
| | 60% | 70% | 80% | 90% |
| a | 0,04 | 0,03 | 0,03 | 0,01 |
| b | 0,14 | 0,15 | 0,15 | 0,16 |
| r | 0,31 | 0,23 | 0,22 | 0,06 |
| N | 197 | 139 | 79 | 27 |
| Pearson | 0,921** | 0,923** | 0,940** | 0,928** |
| Pisa Italian | | | | |
| | 60% | 70% | 80% | 90% |
| a | 0,08 | 0,09 | 0,08 | 0,07 |
| b | 0,12 | 0,11 | 0,12 | 0,13 |
| r | 0,71 | 0,83 | 0,64 | 0,58 |
| N | 309 | 285 | 246 | 186 |
| Kendall's tau | 0,695** | 0,705** | 0,717** | 0,751** |

Table 2: *Application of the level-II formula to the BC and PI corpora. (** = p<0,01).*

| | 60% | 70% | 80% | 90% |
|---|---|---|---|---|
| Spontaneous Beijing Chinese | | | | |
| r | 0,36 | 0,22 | 0,26 | 0,32 |
| Pearson | 0,905** | 0,912** | 0,912** | 0,897** |
| N | 88 | 65 | 39 | 16 |
| Read Beijing Chinese | | | | |
| r | 0,33 | 0,29 | 0,22 | 0,09 |
| Pearson | 0,934** | 0,935** | 0,970** | 0,977** |
| N | 109 | 74 | 40 | 11 |
| Spontaneous Pisa Italian | | | | |
| r | 1,21 | 1,30 | 0,82 | 0,62 |
| Kendall's tau | 0,683** | 0,703** | 0,726** | 0,728** |
| N | 138 | 122 | 103 | 74 |
| Read Pisa Italian | | | | |
| r | 0,37 | 0,08 | 0,12 | 0,38 |
| Kendall's tau | 0,714** | 0,716** | 0,713** | 0,767** |
| N | 171 | 163 | 143 | 112 |

Table 3: *Level-II features of spontaneous vs. read speech in BC and PI (** = p<0,01).*

Separate inspection of spontaneous and read materials adds further details to the picture. As table 3 shows, the situation is fairly stable in BC. The only minor divergence is to be noted at the 90% criterion, which is however hardly relevant, due to the small number of SAs. PI offers a more dynamic picture. While read speech conforms to the above-described controlling behavior, spontaneous speech switches to compensating behavior at the less demanding criteria (60-70%). This suggests two observations: (a) there seems to be a rhythmical divergence between spontaneous and read speech not to be observed in BC; (b) spontaneous PI highlights a hidden rhythmical ambiguity, which emerges in particular with the most liberal criteria (namely, with shorter inter-accentual intervals), where intimations of compensating behavior emerge. While the reason may ultimately be the same as that described above (i.e., an effect of the syllable's segmental composition), the unstable behavior of Italian should be underlined as an important datum, already partially pointed out in the investigation reported in [2].

The somehow unstable picture of PI invites further inspection into the speech rate factor, a well-known predictor of rhythmic behavior. Table 4 presents the results, with the corpus sentences equally divided in two subsets (T1 = slow, T2 = fast). For simplicity's sake, only the results relative to the intermediate criteria (70-80%) are reported.

| | T1 | T2 | T1 | T2 |
|---|---|---|---|---|
| | SPONTANEOUS | | | |
| | 70% | | 80% | |
| r | 1,86 | 0,66 | 0,98 | 0,28 |
| N | 66 | 56 | 58 | 45 |
| | Read | | | |
| r | 0,17 | 0,68 | 0,03 | 0,84 |
| N | 86 | 77 | 78 | 65 |

Table 4: *Level-II features of spontaneous vs. read PI relative to speech rate (T1 = slow, T2 = fast).*

Although the limited corpus size dictates caution, an interesting trend seems to emerge, such that spontaneous and read speech once again diverge. In spontaneous PI, the controlling behavior constantly emerges at T2, where the *r* value is systematically lower with each criterion used. This was expected: at higher speech rates there is less freedom for the syllable-peak oscillator to adjust w.r.t. a superordinate (and obviously unconscious) rhythmical target. The surprise comes from read speech, where the contrary tendency was found: T2 yielded higher *r* values. Although the corpus size does not allow strong inferences, one can propose the following interpretation, supported by data reported in [2], where (with 5 speech rates) alternating results were observed: i.e., *r* increased from T1 to T2, then decreased with T3 to rise again with T4 and finally decreased with T5. As it happens, the effect of speech rate increase is not uniform, for it goes together with the reduction of the SAs number. Depending on speed, the mean dimension of the inter-accentual intervals allows varying degrees of internal syllabic flexibility. In the present experiment something similar must have occurred, for the average speed of spontaneous and read speech differed (the respective boundaries between T1 and T2 were: spontaneous = 15,8 segments/s, read = 13,8 segments/s). One could thus arrange the data into the following scale of increasing speed: read T1 < spont. T1 < read T2 < spont. T2. Although the figures in table 4 do not not exactly alternate in this way, there is a trend in this direction. One can surmise that this tendency would be further endorsed by a larger corpus. Further data will be collected to this aim.

## 3. Conclusions

The above reported experiments were conceived to check the rhythmical tendencies of Chinese and Italian at level-II. One major conclusion is the relative SA-deafness of the Chinese speakers. The comparatively low inter-subjective convergence in SA identification confirms the traditional view and rejects the tentative hypothesis put forth in [2], where it was dubitatively proposed that Chinese might be a level-I controlling / level-II compensating language. The present results show that Chinese is indeed controlling, but possibly characterized by reduced salience of the level-II component.

As for Italian, the situation is more dynamic, for the level-II rhythmical organization, although oriented towards the controlling pole, appears to be more variegated. The actual polarization heavily depends on the speech rate factor, intersecting the style factor (read vs. spontaneous). Although the present results essentially confirm the ones in [2], more research is needed to settle the matter. A larger corpus is being tested to this effect.

## 4. References

[1] Bertinetto, P. M. and Bertini, C., "On modelling the rhythm of natural languages", Speech Prosody International Conference Proc., 4: 427-430, 2008.

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

[3] Bertini C., Bertinetto P. M. and Zhi N., "Chinese and Italian speech rhythm, normalization and the CCI algorithm", Interspeech Conference Proc., 12: 1853-1856, 2011.

[4] Delmonte, R., Bristot A., Chiran, L., Bacalu C. and Tonelli S., "Parsing the Oral Corpus AVIP/API", Atti del Convegno Internazionale'Il Parlato Italiano, 2004.

[5] Grabe, E. and Low, E. L. "Durational variability in speech and the rhythm class hypothesis", In C. Gussenhoven and N. Warner [Eds], Papers in Laboratory Phonology 7, 515-546, Berlin: Mouton de Gruyter, 2002.

[6] Komatsu, M., "Chinese MULTEXT: recordings for a prosodic corpus", Sophia Linguistica 57: 359-369, 2009.

[7] Li, A. J., "Chinese Prosody and Prosodic Labeling of Spontaneous Speech", International Conference on Speech Prosody Proc., 1, 2002.

[8] Li, A. J., Yin, Z. G., Wang, M. L., Xu, B. and Zong, C. Q., "A spontaneous Conversation Corpus CADCC", Oriental COCOCSDA Workshop, 2001.

[9] O'Dell, M. L. and Nieminen, T., "Coupled oscillator model of speech rhythm", International Congress of Phonetic Sciences Proc., XIVth: 1075-1078. 1999.

[10] Ramus, F., Nespor, M. and Mehler, J., "Correlates of linguistic rhythm in the speech signal", Cognition, 73(3), 265-292, 1999.

[11] Savy, R., Crocco, C. and Cutugno, F., "API – Archivio del Parlato Italiano", CIRASS-Università degli Studi di Napoli Federico "II". online, accessed on 7 July, 2011: http://www.parlaritaliano.it/index.php/it/corpora/673-corpusavip-api.

[12] Taranto, M. A., Bertini, C. and Bertinetto, P. M., "Rhythmic Index Elaborator (RIE) come strumento di indagine della struttura ritmica. Un'applicazione al Pisano semi-spontaneo vs. letto", Atti del 7$^{o}$ convegno AISV Lecce, 2011.

[13] Zhi N., Bertinetto P. M. and Bertini C., "The speech rhythm of Beijing Chinese, in the framework of CCI", International Congress of Phonetic Sciences Proc., XVIIth: 2316-2319, 2011.