Pier Marco Bertinetto

Chiara Bertini

Towards a unified predictive model of Natural Language Rhythm

(submitted paper)

This paper presents a natural language rhythm model, conceived so as to comply with the basic epistemological requirements of: explicitness, predictivity, unification. The last requirement refers to the fact that, over and above the terminological convenience, the competing rhythmical types (such as the traditional constrast 'syllable- vs. stress-timing') should be regarded as the two extremes of a continuum, rather than radically alternative types.
The proposed model is based on two independent levels: level-I (phonotactic) and level-II (sentential). At each level, languages may be characterized as more or less "controlling" or "compensating", along a continuum that ideally goes from a maximum of rigidity to a maximum of flexibility. Most importantly, languages may present competing tendencies at the two levels. This possibly accounts for the often elusive character of the rhythmic tendencies of the individual languages.
As far as level-I is concerned, a new algorithm (Control/Compensation Index) is proposed in order to check the rhythmical inclination of the languages. As for level-II, the O'Dell & Nieminen algorithm is exploited. Although provisional, the results obtained demonstrate that it is possible to base research on speech rhythm on entirely predictive models, allowing for direct falsifiability.

To Olle Engstrand, rhythmically

1. Epistemological requirements

Research on Natural Language Rhythm (NLR) entered a new phase around the turn of the new Millennium, when an entirely new algorithm to compare the rhythmical inclination of individual languages was proposed (see Ramus et al. 1999). The suggestion was soon followed by other scholars, suggesting revised or modified versions. The present authors will not even attempt at quoting them all. Among the revised versions, one should especially consider the Varco (Dellwo 2004) and the “semi-syllable” models (Rouas & Farinas 2004). These, like the Ramusian proposal, may be called “static” models, for the actual sequence of the relevant intervals (consonantal and vocalic) does not play a role. The results are not affected by any possible permutation of intervals; the algorithms provide an overall measure

characterizing any speech stretch in its entirety, be it the standard variation of the relevant interval's duration, their mean error, the global percent value etc. Among the modified versions, it is worth mentioning the method proposed by Wagner (2007) and most notably the PVI model (Grabe & Low 2002); the latter should be characterized as "dynamic", insofar as it takes into account the local durational fluctuations between any two adjacent intervals.

Despite the merit of revitalizing the topic of NLR, all these recent proposals (inspiring and even exciting as they undoubtedly have been) seem to be somewhat defective on epistemological grounds. In order to grasp this, let us list the three requirements that any NLR theory should fulfill, namely: (a) EXPLICITNESS, (b) PREDICTIVITY, (c) UNIFICATION. The first two are self-explaining; the third is strictly related to this particular research domain. The succinct survey that follows has no historiographic ambition; it is only meant to show that none of the models so far proposed fulfill all three requirements. The proposal put forth in this paper aims at remedying this fault.

Pike (1947) was a good start. The theory was perfectly explicit and predictive. It stated that languages belong to two types, each characterized by isochronicity within a specific domain: the syllable or the stress phrase (the latter to be intended as the inter-stress interval, i.e. the stretch comprised between the onset of a stressed syllable – or, alternatively, vowel – and the next one): hence, the contrast SYLLABLE- vs. STRESS-TIMING. This theory should be praised for its explicitness. The crude linguistic facts soon falsified it (for a more recent disconfirmation, see Van Santen & Shih 2000), but one should take this as a welcome result: falsified theories pave the way for better ones. There is another reason to be grateful to Pike: he pointed out the way towards the experimental testing of a prominent linguistic feature, something that still keeps people busy. As for the third requirement (unification), the Pikean theory was obviously orthogonal to it, for it postulated that languages belong to two radically alternative types (one syllable-sensitive, the other stress-sensitive). We take this to be a flaw, for assuming the existence of mutually unrelated rhythmical types looks unattractive. Indeed, since all languages have syllables, one wonders why only a subset of them should select the syllable as rhythm regulator. One should rather start from the assumption that all natural languages share the same structural features (e.g., syllables): the differences should best be conceived of in terms of degrees along a continuum, rather than as irreconcilable.

The Pikean model's failure gave rise to a number of attempts to save its basic intuition (for a detailed survey, mirroring the situation at the end of the Eighties, cf. Bertinetto 1989). Once it was ascertained that the original formulation did not correspond to the facts, the solution was sought in other directions, among which, most notably: (i) perceptual constructs feeding impressionistic judgments (see references in Bertinetto 1989); (ii) syllabic duration

compensation in the word or accentual domain (Lindblom & Rapp 1973, soon followed by others). The phonologically-oriented proposals by Bertinetto (1981) and Dauer (1983, 1987) also deserve mention: they pointed out a number of prosodic features variously feeding the rhythmical classification of languages, including – among others – the following: (a) V-reduction *vs.* full articulation in unstressed syllables; (b) complex *vs.* simple syllable structure; (c) relative flexibility *vs.* rigidity in word-stress placement; (d) tempo acceleration mainly due to compression of unstressed syllables *vs.* proportional compression. It immediately appears that the latter proposals presupposed a unified theory. Unfortunately, however, they were both wanting in explicitness and predictivity: although the features indicated (or a subset of them) are likely to have a bearing on NLR, their exact contribution was not spelled out. Altogether, the intermediate post-Pikean period might be characterized as a time for rethinking: lacking a predictive theory, the main effort was put into trying to collect arguments conducive to a unified theory, based on a broad typological view of the prosodic systems of natural languages.

The most recent models, although differing in the details, share one fundamental feature with the Pikean model: they are all explicit, for they offer algorithms capable of generating the desired segregation of the alleged syllable- vs. stress-timed languages. Whether they also exhibit predictivity, is another matter. In a sense, they should be regarded as at least weakly predictive, due to their explicitness. However, they cannot be regarded as fully (or strongly) predictive, for they are reticent on the unification issue. To avoid misunderstanding, one should add that the latter remark should not be read as referring to the position actually maintained by the individual models' proponents: what is meant here is that the models as such do not allow any specific inference as to whether the theory presupposes a unified design, or a two-modal one based on radically alternative rhythmical types. Since the authors do not state what the alleged rhythmical contrast should be based on, it is impossible to shed light on the issue. Actually, considering that most scholars agree that languages cluster around two rhythmical types, one might even suppose that this should be accepted as a basic postulate. But scientific enterprises cannot merely stem from intuition. The weakness of this state of affairs is obvious. In the absence of explicit predictions at the outset, the recent NLR models run a severe risk of circularity: any such algorithm is claimed to be working fine whenever it produces the correct grouping (where "correct" can only mean "consistent with the experimenter's expectations"; see Arvaniti, in press, for a similar criticism). Thus, the interpretation can only arise *post factum*, in terms of relative positioning. The models yield a topological arrangement, whereby languages of group A vs. B (whose existence is assumed, rather than independently explained) are shown to occupy different areas on the Cartesian

plane. This, however, does not tell us anything about the actual property that a language should exhibit in principle, in order to belong to the one or the other type. Consequently, none of these models can specify which language type should occupy which portion of the graphic, depending on which finely attuned structural properties. As a further consequence, the models lack an explicit metrics to effectively measure the distance between languages; hence, the "intermediate" types are merely accepted as a classification residue, rather than predicted. The recent literature on NLR abounds in sentences such as: "contrary to expectations, language X clusters with stress- rather than syllable-timed languages" or "language X is intermediate between the two types". There is nothing intrinsically wrong in this, except that belonging to the one or the other type is inferred *a posteriori* from the clustering results, rather than defined on independent grounds.

Needless to say, the above criticism is not meant to deny the validity of the general consensus on the existence of contrasting rhythmical tendencies. This does not merely stem from intuition, but is based on objective data, two of which are worth mentioning here. One source of data is the different ease with which the various languages may be fitted into musical-rhythmic frames. Although any language ultimately admits this possibility, the specific ways in which this may be obtained vary a lot. A dramatic contrast of this sort is hinted at by Cummins (2002), comparing the behavior of English speakers with that of Italian and Spanish speakers. Although a cross-linguistic detailed and large-scale study of the relation of words to music has not been undertaken to date (but see e.g., Dell & Halle, to appear), one may surmise that it would produce exciting results. Another important source of data is the different organizational basis of traditional versification systems. Each system captures the most relevant prosodic features of the given language, turning them into a (set of) organizing principle(s), such as: inter-stress distance, syllable counting, mora or syllable quantity, tone dynamics etc., often combining more than one principle. For instance, stress-syllabic systems regulate the inter-stress distances in terms of syllable counting, using foot-measures reminiscent of the Greek and Latin tradition, although the latter implemented a quantity-syllabic system. Since metrically regulated speech is intentionally aiming at rhythmicity, one is immediately drawn to the conclusion that the rhythm organizational basis differs from language to language, for otherwise every linguistic community would have adopted the same system.

Having said this, however, one should also acknowledge that no scientific enterprise can ignore its epistemological obligations. To put it succinctly: the basic intuition should first be connected to explicit structural properties about which detailed predictions can be formulated; these predictions should then be tested by appropriate tools, until they are falsified. In recent

NLR studies, however, the reverse happened: various tools have been devised to ascertain the initial intuition concerning rhythmic typology, without previously defining the exact structural properties on which NLR rests.

The general lesson to be learned from the brief survey in this section is that, although there seems to have been a constant – albeit discontinuous – progress in NLR theorizing, none of the models so far developed exhibited all three epistemological properties required by this particular research domain, as summarized in the following table:

NLR models	*Unification*	*Explicitness*	*Predictivity*
Pike 1947	-	+	+
Bertinetto 1981; Dauer 1983, 1987	+	-	-
Lindblom & Rapp 1973	+	+	-
Ramus, PVI, Varco	?	+	-

Table 1. Fulfillment of the epistemological requirements by selected NLR models.

It follows that the most urgent task consists in devising a unified, fully explicit and predictive theory, capable of generating the appropriate expectations as for what a language should be like (and do) in order to be assigned to a given rhythmical type.

2. Towards a new model

In a recent work (Bertinetto & Bertini 2008), the present authors offered the first outline of such a NLR model. The model will be further developed here. The reader should however be warned that this section does not exhaust the matter: § 3 will present the complete design, while § 4 will suggest possible expansions. Following the example of other scholars, the traditional terminology (syllable vs. stress-timing) will be abandoned, to avoid any misunderstanding tied to its original meaning. For simplicity's sake, this model also comprises two ideal types: CONTROLLING vs. COMPENSATING (henceforth: CONTR vs. COMPS), except that these should be conceived of as the extremes of a continuum and thus referred to for purely descriptive reasons. The terms are borrowed from Hoeqvist (1983), although the interpretation is quite different (namely, along the lines proposed in Vékás & Bertinetto 1991, who presented an embryonic sketch of the theory developed here).

The basic idea, inspired by work in articulatory phonology and earlier on by the seminal work of Fowler (1977), is as follows: languages may differ in terms of how V and C gestures are coupled in the speech flow. An ideally CONTR language should be conceived of as a language in which all segments receive the same amount of expenditure – or articulatory effort – and tend to have the same duration. This is obviously impossible, due to the varying

points and manners of articulation; yet, this view acquires plausibility as soon as one considers how languages diverge in terms of the coupling of V and C gestures. Some languages admit – or rather require – a much larger segmental overlap (i.e. co-articulation, co-production) than others, as evidenced by research in articulatory phonology. In the present authors's view, such languages correspond to the COMPS type. Here again, the ideal maximum – whereby all strictly adjacent C and V gestures overlap entirely – is physically impossible. It should thus be immediately clear that both extremes (CONTR / COMPS) are artificial constructs, only used to designate two ideal cases. What one actually finds in the real world are higher or lower degrees of control / compensation. This inspires the CONTROL / COMPENSATION (**C/C**) hypothesis.

The actual position along the continuum depends, to a very large extent, on the phonotactic structure of the individual language (see below, in this section, for further qualification; at any rate, the relevance of phonotactics for rhythm research was also pointed out by Eriksson 1992). A simple phonotactics naturally inclines towards the CONTR setting. Note that a language consisting of just one C and one V would be perfectly rhythmical (e.g., *ba-ba-ba*). This does not follow from any musical or dancing predisposition of human beings; it is a mere "emergent" property of gestural coordination, as task-dynamics has shown for quite some time. If rhythmicity is indeed the simplest way to cope with complex coordination problems, then language is an obvious candidate for it, for speech production involves the fine intertwining of several articulators. Languages, however, are complex systems, based on a number of (possibly competing) structural components. Not only do their phonologies involve an often fairly rich segment inventory, but word and sentence prosody interact with the segmental level in a number of ways, producing in the long run all sorts of phonological restructurings. As a result, languages often present a rich phonotactics, which forces the speaker to adopt a flexible (COMPS) articulatory setting. The natural result of this is the overlapping of C and V gestures, as Goldstein et al. (2007) have empirically shown with respect to syllabic structure: languages with a simple phonotactics have a greater chance of presenting a fairly in-phase coupling of the consonantal and vocalic oscillators. Thus, departing (more and more) from the CONTR ideal is the almost automatic consequence of a (more and more) complex phonotactics. The most typical sites for gestural overlap are the unstressed syllables, where the V nucleus offers itself as the privileged target for co-articulation. Needless to say, unstressed syllable reduction also occurs in CONTR languages, but to a lesser extent; conversely, and crucially, intra-syllabic durational compensation is larger in COMPS than CONTR languages, especially (but not only) in unstressed syllables. It will not go unnoticed that this view departs radically – and somehow paradoxically – from the

traditional one, despite the factual coincidence (at least in the prototypical cases) of COMPS and so-called stress-timed languages, as well as CONTR and syllable-timed ones. This should, however, cause no surprise, considering the empirical inadequacy of the Pikean view.

An important qualification is in order. Since complex phonotactics is naturally associated with complex syllable structure, the implication creeps in that syllable structure is the ultimate trigger of rhytmical differences. This, however, does not correspond to the authors' view. Following suggestions by Dziubalska-Kolaciuk (2002), Ohala & Kawasaki-Fukumori (1997), Vennemann (1994), Steriade (1998), Blevins (2003) and Dressler & Dziubalska-Kolaciuk (2006), syllable structure should be regarded as an epiphenomenal consequence of phonotactics, supplemented by domain constraints at the relevant prosodic level.[1] Thus, on the one hand, phonotactics is structurally superordinate with respect to the syllable; on the other hand, syllabification involves domain properties that are largely irrelevant to the rhythm issue.

The C/C model here described directly fulfills, due to its very conception, one of the three fundamental epistemological requirements, namely unification. The remaining two (explicitness and predictivity) need to be satisfied by appropriate computational tools. In Bertinetto & Bertini (2008) the following modified version of the PVI algorithm – called CONTROL/COMPENSATION INDEX (**CCI**) – was proposed, where d_k expresses the duration in milliseconds of the k^{th} V-(or C-)interval, n_k the number of segments in the relevant interval and m the number of V-(or C-)intervals composing the stretch of speech considered:

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

In practice, CCI relativizes the PVI measure to the number (n) of segments composing each V- or C-interval. The model thus inherits the PVI's dynamic virtue, adding to it the complexity of the phonotactic structure, for it obviously makes a difference whether a given C interval comprises one or several segments.

It is important to realize that CCI is a phonologically-driven model. Geminates and long Vs count as two segments (cf. Finnish), just like two Vs in synaloepha (whereby one of two abutting Vs is deleted); by consequence, hyper-long segments count as three (cf. Estonian). Conversely, Vs in hiatus count as separate (monosegmental) V intervals for – as detailed in § 3 – each syllable nucleus implements a vocalic oscillator's period. Similarly, instances of occasional C elision – especially frequent in *allegro*-style – should be taken care of by counting all the target Cs irrespective of the elision (provided, of course, the elision is not

[1] Depending on the language, the relevant domain should be identified with morpheme, word or phrase.

typical of the specific language variety adopted by the speaker; in that case, one should acknowledge the existence of a different rhythmical inclination). While applying the CCI algorithm one should thus carefully consider the phonological structure of the languages under study, possibly adopting a double counting in critical instances. Glides are a case in point: their treatment as either C or V segments varies from language to language. It is thus advisable to apply the algorithm in both ways, in order to ensure cross-linguistic comparison.[2]

In Bertini & Bertinetto (in press) the criteria adopted for the coding of a semi-spontaneous Italian corpus were spelled out. The materials consisted of excerpts of map-task dialogues, carefully segmented and labeled (the source corpus is available at: http://www.cirass.unina.it/ricerca/studi%20parlato/raccolta%20corpora/api/api.htm/). Each excerpt was at least eight (phonetically realized) syllables in length. Ten speakers were involved and the intervals numbered nearly 3000 for both Cs and Vs; almost 8500 segments were included in the measurements. One detail worth mentioning is that the final portion of any sentence, from the last stressed V onward, was neglected.[3] The use of "trimmed sentences" is justified by the fact that the final portion has its own (language-specific) prosodic properties as a boundary signal, that should be studied on its own independently of rhythm proper.

CCI makes explicit and directly verifiable predictions, as shown in fig. 1. Languages oriented towards the CONTR type should fall in the proximity of the bisecting line, showing that the local fluctuation of Cs and Vs tends to be of the same magnitude, whereas COMPS languages should exhibit more V than C fluctuation.

The analyses carried out by Mairano & Romano (2008) on a corpus of read speech passages, produced in 8 different languages, yielded results in line with the CCI model's predictions, as shown in fig. 2: German, American and RP English (traditionally considered stress-timed) tend to be COMPS, since they present comparatively more V than C local variation, as a consequence of the large amount of V-reduction in unstressed syllables. Conversely, Finnish, French, Canadian French and Italian (traditionally considered syllable-timed) lie in the vicinity of the bisecting line. Note that the data in fig. 2 stem from read speech, with the exception of those indicated as SpoIT, corresponding to the spontaneous Italian data analysed in Bertinetto & Bertini (2008); the latter presumably underwent some shifting towards the COMPS pole, due to hypo-articulation.

[2] It can be anticipated here that the application of this double measurement strategy to the Italian data described below produced a statistically irrelevant difference. Needless to say, languages with a much larger presence of diphthongs are expected to yield a significant contrast; in the present case, the segments involved in diphthongs (V plus adjacent glide) were 5.1% of the total.

[3] In addition, any C preceding the last stressed V was trimmed, on the assumption that the final lengthening phenomenon might involve at least part of that interval.

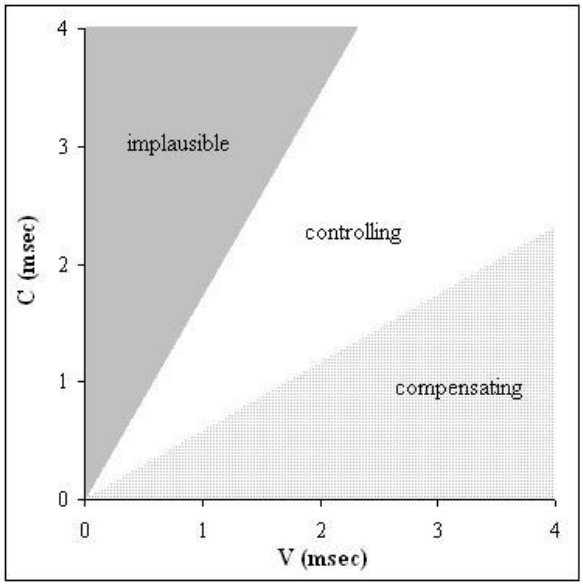

Figure 1: Schematic representation of the two ideal rhythmic types according to the C/C model.

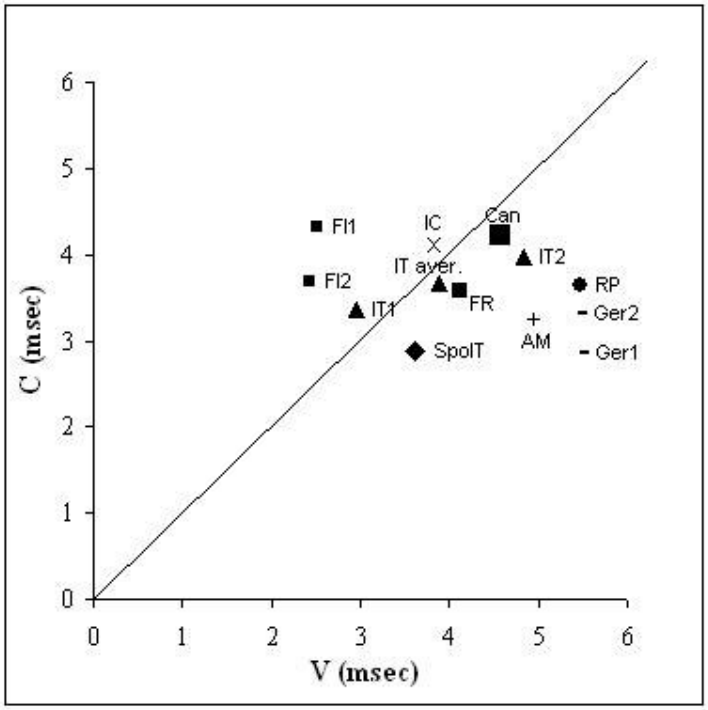

Figure 2: Application of CCI by Mairano & Romano (in prep.): AM = Amer. Eng., Can = Can. Fr., FI(1,2) = Finnish, FR = French, Ger(1,2) = German, IC = Icelandic (average of 10 speakers), IT(1,2)

= Ital. (+ IT aver.), RP = Eng. RP. SpoIT = spontaneous Ital. corpus as analyzed in Bertinetto & Bertini (2008).

Two caveats should be pointed out. First, with the exception of SpoIT and IC (Icelandic), each point on fig. 2 refers to a single speaker. However, as may be seen for FI (Finnish), Ger and IT (and as is known from previous studies, e.g.: Dellwo et al. 2005, Barbosa 2006), different speakers may do quite different things, suggesting that no generalization should be drawn from small observational bases. Second, the position of IC may appear to be somehow surprising, considering that it is a Germanic language like English and German, with a comparable phonotactic richness. Icelandic, however, seems to exhibit a low degree of V-reduction (Mairano, pers. communic.), which is compatible with its position in the figure. This datum suggests an important theoretical consequence that should be duly stressed to integrate the observations put forth at the beginning of this section. A rich phonotactics is not by itself conducive to COMPS behavior, although this is the default situation: the ultimate factor seems to be the amount of V-reduction, which may in some cases dissociate from phonotactic richness. This suggests that the degree of V-elasticity is in general higher as compared to C-elasticity, as also shown by the smaller range of variation of C as compared to V in fig. 2. Further support for the above-mentioned dissociation is provided by Singapore English, as opposed to British English (Low 1998); Western, as opposed to Eastern, Catalan (Gavaldá & Dellwo in preparation); Cantonese, as opposed to Mandarin, Chinese (Mok & Dellwo 2008). This dissociation is also to be found in L2 pronunciation of COMPS languages (e.g., English as spoken by Chinese speakers, Mok in preparation; see also White & Mattys 2007). The exact articulatory setting of such language varieties should be thoroughly investigated, also regardless of the NLR issue.[4]

Speech tempo variations provide a valuable test to assess the C/C hypothesis. The predictions are as follows: (i) Ideal CONTR languages should tend to reduce the segments' duration in a proportional way, whereas in COMPS languages Vs should be more affected than Cs; (ii) In CONTR languages, reduction should be much more drastic between slow than between fast rates, whereas in COMPS languages reduction should be relatively robust even at fast rates. The latter prediction stems from the larger articulatory flexibility of COMPS languages, allowing further freedom in terms of co-production of V and C gestures, while CONTR languages meet their compressibility threshold earlier (Bertinetto & Fowler 1989; cf. also Price 1980 and Davidson 2006).

[4] The notion "articulatory setting", already mentioned in § 2, should be understood as the language-specific implementation of the articulatory commands as based on specifically set coefficients, whereby CONTR languages exhibit higher stiffness coefficients – in the sense of Vatikiotis-Bateson (1988) – than COMPS ones.

These predictions were tested against the afore-mentioned spontaneous Italian corpus (Bertini & Bertinetto, in press). The speech materials were divided into 3 naturalistically obtained tempo groups (T1, T2, T3). The assignment of each utterance to a given group was done *a posteriori*, rather than directly elicited from the speakers: this avoids any possible distortion induced by the conscious effort to comply with the experimenter's demand. The rate measure used was segments/sec., which proved to be definitely more reliable than syllables/sec.[5] The rate classes were obtained so that each group consisted of an equal number of V- and C-intervals; as a consequence, the number of utterances in each class was not perfectly identical. This yielded the following classes:

Segments/sec.: T1 ≤ 15,6 (av. 14.2); T2 > 15.6 & ≤ 17.65 (av. 16.6); T3 > 17.65 (av. 19.2)
Syllables/sec.: T1 ≤ 6.75 (av. 6.1); T2 > 6.75 & ≤ 7.75 (av. 7.3); T3 > 7.75 (av. 8.9)

Interestingly, the general trend of the models is linear, with the exception of %V, nPVI(C) and Varco(C):[6]

T1 // T2 // T3	CCI(V), Ramus(V+C), rPVI(V+C), Varco(V), RF(V+C)
T1 <> T2 <> T3	nPVI(C), Varco(C)
T1 <> T2 // T3	nPVI(V)
T1 // T2 <> T3	CCI(C), %V

Table 2. Statistical analysis, based on three rate classes, according to alternative rhythm models: CCI, Ramus, PVI (raw and normalized), Varco, Rouas & Farinas (RF). Rate measure = segments/sec. The diacritics <> and // stand, respectively, for statistically 'not-separable' vs. 'separable' according to pairwise t-tests carried out on T1 vs. T2, and T2 vs. T3.

The first row in table 2 indicates that the relevant models are very sensitive to the rate differences as considered here; the second row, on the contrary, indicates that no difference is detected. The third row presents a rather implausible situation, whereby the contrast appears to be sharp only at fast rates; the last row, instead, suggests that the incompressibility threshold is reached between T2 and T3, which is definitely more reasonable. It appears that CCI is among the most sensitive models and it is the only one to capture the plausible propensity of Cs to attain incompressibility before Vs. With respect to the predictions spelled out above, the statistical analysis based on CCI suggests that Italian does not conform entirely

[5] This should not come as a surprise, considering that in the syllable/sec. measure a structural factor – syllable complexity – ends up compounded with a duration factor.

[6] The data reported below refine those of Bertinetto & Bertini 2008, where rate was only measured in syllables/sec. and the groups were equalized with respect to the number of utterances involved, so that the number of V- and C-intervals in each class was not perfectly identical.

to the idealized CONTR type. Indeed, the acceleration's effects are not strictly proportional for Vs and Cs, for only the latter reach threshold in the T2 vs. T3 comparison.

Although this cannot be regarded as the last word on the matter, the results, together with the ones reported in fig. 2, look promising. One aspect of the model is especially worth highlighting here: namely, its predictive character. This enables the researcher to put forth meaningful predictions with respect to tempo variations within a single language. Inter-language comparison is a useful – and typology-wise unavoidable – perspective, but is not necessary to validate the model. This solves the circularity problem referred to in § 2.

3. A two-level model of NLR

The above sketch of a NLR model was devised to capture (to a large extent) the rhythmical consequences of phonotactic structure. In the default case, a complex phonotactics (yielding a complex syllable structure) is conducive to a COMPS behavior, although one and the same language may be pronounced with varying rhythmical inclinations (see the examples provided above). This, however, does not exhaust the picture, for languages are based on a complex phonological architecture. Over and above the segments' concatenation, they present overarching levels, among which ACCENTUAL DOMAINS (as defined below) are especially relevant to the present concern. The model should thus be extended in the direction of a two-level architecture, conceived of as two pairs of coupled oscillators, comprising:

- Level-I (PHONOTACTIC), based on the coupling of the vocalic and consonantal oscillators, along the lines suggested by Goldstein et al. (2007).[7]
- Level-II (SENTENTIAL), based on the coupling of the accentual and syllable-peak oscillators, adopting suggestions by O'Dell & Nieminen (1999).

In O'Dell & Nieminen's model, the subordinated oscillator is called "syllabic". It is important to realize, however, that the property of the syllable which is relevant here consists in its being a peak- (or nucleus-) carrier, rather than an organizational principle emerging from the phonotactic flow. In this paper the term "syllable-peak" was thus preferred to avoid confusion, namely to avoid any possible conflict with the substantive content of level-I, which was defined above as independent of syllable structure. Note however that – as proposed in § 4 – there is factual convergence of the level-I vocalic oscillator and the level-II syllable-peak oscillator; the two levels should thus be seen as intersecting in an interesting way.

The above two levels should be understood first and foremost as levels of structural description, although they are ultimately grounded in articulatory control. In this respect, the

[7] Actually, the latter model assumes as many C-oscillators as there are Cs in a cluster, but this is irrelevant for the present concern, although definitely relevant for the purpose of modeling the fine syllabic structure.

notion "oscillator" deserves a comment. Any rhythmical behavior presupposes the periodic activity of some sort of oscillator. This should be interpreted in a physical, rather than metaphorical sense. In the case at issue, the level-I oscillators receive a straight-forward physiological interpretation: the V- and C-oscillators are explicitly mentioned in articulatory phonology, based on a solid tradition of empirical studies. It follows, then, that the nature of the level-I oscillators is universally specified, for all languages consist of recurring Vs and Cs. This, however, does not hold for level-II. Here again we have to do with physical objects – namely the more or less regularly recurring sentence accents – but the actual implementation of the level-II oscillating system varies from language to language, depending on the specific set of acoustic cues making up the accents' physical substance. Thus, although both pairs of coupled oscillators should be interpreted in physical terms, only the first pair receives a direct physiological interpretation, while the implementation of the level-II oscillators not only varies from language to language, but heavily depends on the language-specific perceptual tranfer from the acoustic signal. Indeed, it is not even granted that one and the same cocktail of acoustic cues gives rise to identical judgments in different languages.

Multi-level conceptions of NLR have already been advanced in the literature (e.g., Barbosa 2007; O'Dell et al. 2007). The major claim to originality of the model proposed here, apart from its specific shape, lies in the possible divergence of the two levels, as detailed below.

As an aside, one could observe that the content of level-I and –II is vaguely reminiscent, respectively, of the syllable- and stress-timing division of labor. The major depart from the Pikean tradition lies in the fact that both levels are considered relevant for any given language. Since, in the present model, syllable and accent are no longer regarded as the source of two alternative rhythmical tendencies, they should be regarded as basic prosodic features necessarily shared by all languages. Specifically, although the phonological role of word stress differs from language to language, one may assume that sentence accents are universally present as rhythm regulators, whatever their language-specific phonetic implementation. The latter is no doubt the product of several intermixed acoustic components, as emphasized by Kohler (in press). For example, dynamic tones – especially descending ones – yield an impression of longer duration, as opposed to static tones, adding further complications at the perceptual level. Besides, in stress languages, like English or Italian, there is an interplay between word-stresses (including secondary ones) and sentence accents: at very slow rates, the latter tend to coincide with the former, whereas at faster rates only the most salient stresses are preserved. As a result, the average number of syllables per accentual

domain increases along with the tempo, preserving some sort of durational regularity. This fact will be directly observed in the data reported below.

Of paramount importance is the contrast 'rigid' vs. 'mobile' word stress. In Italian, word stress may be downplayed or even (at faster rates or in stress clashes) deleted, but – with rather few exceptions – it cannot be shifted. In English, on the contrary, a large part of the lexicon may undergo optional stress shift, as in words like *coronal*, *exponent*, *contribute*, *subsidence*, *exquisite*, *satiety* etc. (also depending on specific sociolects). This may be regarded as the level-II equivalent of the level-I C/C-continuum: the more mobile the stress is, the more flexible (COMPS) the accentual structure, for the speaker may then have a larger degree of freedom in regulating the inter-accentual distances. Once again, one finds a gradient between two extremes, in accordance with the unification requirement. The underlying assumption is that speakers follow their spontaneous inclination towards rhythmicity as long as the language intricacies do not constrain their behavior. Word stress rigidity/flexibility is such a constraint at level-II, just as simple/complex phonotactics is at level-I.

It is important to underline, however, that the C/C parameter does not necessarily converge at both levels. The interaction may be complex, due to the vagaries of linguistic typology. The examples in the table below may not be the most prototypical ones, but will suffice for the present purpose:

TYPE	*LEVEL-I*	*LEVEL-II*	*EXAMPLE*
1	CONTR	CONTR	*Italian*: relatively simple phonotactics, fairly rigid word stress pattern
2	COMPS	COMPS	*English*: fairly complex phonotactics, fairly mobile word stress pattern, density of secondary stresses yielding further prominence sites
3	COMPS	CONTR	*Polish*: very complex phonotactics (Bertinetto et al. 2007), rigid word stress pattern
4	CONTR	COMPS	*Japanese*? *Chinese*? (see the text for comments)

Table 3. Interplay of level-I and level-II with respect to the C/C contrast.

To avoid confusion, one should speak of CONTR-I, COMPS-II etc., with integers referring to the appropriate level. Needless to say, several other components may cooperate to yield the final result, most notably word structure. For instance, a language with many polysyllables and rigid stress pattern (cf. again Polish, with fixed penultimate stress) offers little ease to the speaker to produce regularly recurring prominences. Conversely, a language whose lexicon mainly consists of mono- or disyllables has a much greater chance of presenting regular inter-accentual distances.

What is especially relevant is that a two-level model of NLR seems to justify the often vague intuitions that people have, with respect to the rhythmical inclinations of the languages.

The suggestion underlying the model proposed here is that no single measure can capture the actual behavior of any language: both level-I and level-II should be taken into account. Their possible divergence justifies the sometimes elusive character of rhythm judgments, including scholarly judgments. This may, for instance, explain why Polish has been alternatively assigned, impressionistically, to syllable- or stress-timing, depending on the perceiver.

Three important caveats should be put forward here. The first is to the effect that the accentual phrase as intended here (i.e., the stretch comprised between two sentence accents), should not be confused with the foot, occasionally invoked in relation to rhythmical matters. The foot acts at a lower level than the accentual phrase (indeed, in metrical phonology it is considered to be an infra-lexical unit). Besides, unlike the accentual phrase, the foot is less pervasive than often assumed. In the view of the present authors, not all languages exhibit this unit of phonological analysis (Marotta 2003).

The second caveat concerns the lack of objective criteria for locating the sentential prominences. Individual subjects may or may not detect as prominent a given syllable in a speech chain, and even one and the same speaker may provide different judgments at different times. Apart from the very salient sentential prominences, there is a grey zone of ambiguity often to be noted in spontaneous speech. Indeed, the "news reading" style, some version of which seems to be practiced in most language communities, sounds so peculiar precisely because of the constantly emphatically realized prominences. One should thus be aware that the individuation of sentential prominences is not a straightforward process. It is advisable to adopt multiple measures – as in the results to be presented – e.g., limited to the most prominent peaks (MEASURE α) or including the intermediate ones (MEASURE β) .

The third caveat is even trickier. As it happens, dynamic stress – as conceived of for English, Italian, Polish etc. – is not a feature of every language. For instance, it does not play a role in Chinese, Japanese, Korean, Tamil and Mongol (Akamatsu 1997, Nolan 2008), although even there polysyllables normally have a prominent syllable – or rather, as in Japanese, a mora – presenting distinctive tonal features.[8] Table 3, in any case, is based on the assumption that sentential accent, however realized, is a universal trait as rhythm regulator. Every language is assumed to present sentential prominences whose more or less regular distribution accounts for an important share of rhythm perception. Their presence is normally tied to word-prominence locations, although the relation is not one-to-one, for sentence accents only exist beyond the word, at the intonational level. Their function is purely

[8] Incidentally, the literature often hints at the notion "mora-timing", as applying to languages such as Japanese, Korean, Sinhalese, Tamil, Hindi. In the present authors' view, however, mora-timing is not regarded as an autonomous type, but rather as the most extreme form of level-I CONTR behavior.

communicative-pragmatic: they partition the speech chain into conveniently sized chunks, providing anchoring points that help the hearer to process the intended meaning. Interestingly, this sort of chunking – besides possibly subserving the respiratory activity at the lowest production level – also seems to matter with respect to memory processes (Boucher 2006). One might thus want to consider this a kind of expansion into the cognitive domain of the task-dynamics proposal, concerning the emergent nature of rhythmical behaviors: supposedly, the rhythmical organization of speech at the sentential level is exploited by both speaker and hearer for the sake of thought coordination. If this is true (at least in part), then Chinese and Japanese – plus any phonotactically simple language where dynamic word-stress does not play a role – are good candidates for type (4) in table 3. Should this not be the case, then one should limit the role of level-II to a subset of the languages, reducing somewhat the scope of the two-level model presented here. There is, in any case, little doubt as for the extremely simple phonotactics of languages such as Japanese and Chinese (Bauer 1995, Akamatsu 1997). This suggests them as very likely candidates as CONTR-I languages.

What one needs in order to validate the level-II hypothesis is, once again, a convenient algorithm. The one proposed by O'Dell & Nieminen (1999), exploiting the "Averaged Phase Difference" theory (APD), is a viable option. It has the following shape, where *I* stands for 'duration of inter-stress intervals', *n* for 'number of syllables', ω_1 e ω_2 for the angular frequency of the two oscillators, and *r* indicates their relative strength parameter:

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

In practice, the formula relates the duration of the inter-accentual interval to the number of syllables composing it. If the resulting *r* is greater than 1, then the overarching (accentual) first oscillator predominates; if *r* is less than 1, the subordinate (syllable-peak) second oscillator prevails.

This allows to put forth exact predictions as for level-II, again with respect to speech rate variations: (i) At slow rates, the accentual oscillator should predominate in all languages, following the universal tendency towards rhythmicity alluded to above; (ii) At faster rates, the syllable-peak oscillator should prevail; however, its dominance is expected to emerge earlier, and more emphatically, with CONTR-II languages. The rationale is as follows: COMPS-II languages present a relatively flexible structure, allowing the speaker more freedom to adjust the inter-accentual distances. In a language like English, this may be obtained by downgrading some of the word prominences and possibly promoting some of the secondary ones, and above all by shifting the word prominences as the case requires. In a language, like

Chinese or Japanese, this may supposedly be achieved by appropriately redistributing the sentential prominences, assuming that they are exceedingly flexible due to the non-dynamic character of word stress: the speaker can preserve the correct tonal assignments largely irrespective of sentence accent location. By contrast, since none of these possibilities is available to CONTR-II languages, the dominance of the syllable-peak oscillator should tend to emerge as soon as the speech rate begins to increase, although some restructuring is available to the speaker, mostly by way of accent deletions.

These predictions were tested against the same Italian corpus as exploited in § 2.[9] Since the number of observations was higher than in the CCI calculus, it was possible to partition the materials not only into 3, but into 5 rate classes, with segments/sec. as criterion. Table 4a-b presents the results according to MEASURE α and β, respectively:

Tempo	*segments/sec.*	*N*	*r*	*Tempo*	*segments/sec.*	*N*	*r*
T1	< 15,7 / av. 14.2	264	1.15	t1	< 14.9 / av. 13.6	166	1.05
T2	< 17.8 / av. 16.6	277	1.03	t2	< 16.1 / av. 15.4	157	1.30
T3	> 17.7 / av. 19.2	275	0.71	t3	< 17.2 / av. 16.6	161	0.84
				t4	< 18.9 / av. 17.9	167	0.91
				t5	> 18.8 / av. 20,0	165	0.57

Tempo	*segments/sec.*	*N*	*r*	*Tempo*	*segments/sec.*	*N*	*r*
T1	< 15,7 / av. 14.2	264	1.29	t1	< 14.9 / av. 13.6	166	1.20
T2	< 17.8 / av. 16.6	277	1.06	t2	< 16.1 / av. 15.4	157	1.36
T3	> 17.7 / av. 19.2	275	0.54	t3	< 17.2 / av. 16.6	161	1.02
				t4	< 18.9 / av. 17.9	167	0.76
				t5	> 18.8 / av. 19.9	165	0.46

Table 4a/4b: Output of the APD algorithm as applied to rate classes naturalistically extracted from a spontaneous Italian corpus. Coupled oscillators: accentual vs. syllable-peak. 2a: MEASURE α = limited to the most prominent peaks. 2b: MEASURE β = including the intermediate peaks.

The results show that, at slow rates, the accentual oscillator does indeed predominate; however, as rate increases, the syllable-peak oscillator definitely prevails. This tendency is emphasized by MEASURE β, whereby the intermediate-level accentual peaks are included: at the slow tempos, the intermediate peaks contribute to regularize the inter-accentual distances, whereas at faster tempos they obtain the contrary effect. In the present case this occurred despite the relative rarity (4.3%) of intermediate peaks vis-à-vis the most salient ones. Interestingly, when 5 rate classes are considered, the above tendency turns out to be non-monotonic, showing that tempo variation is accompanied by some restructuring in the

[9] In this case, the C-interval preceding the last stressed V was not trimmed, in order to preserve the integrity of the accentual interval (see fn. 3 for the different strategy used in connection with level-I measurements). Note however that the intervals considered at level-II are much larger; hence, the effect of any possible slowing down during the last C-interval is negligible.

implementation of accentual prominences (i.e., deletions or insertions). For instance, t2 allows a more regular accent distribution, yielding a sharper dominance of the accentual oscillator. Apart from this detail, Italian appears to behave by and large as a CONTR-II language.

This result was expected, but what really matters is that it was autonomously derived: it stems from behavioral measures mirroring relevant structural properties. As noted above, this avoids the risk of circularity implicit in basing one's interpretation on the mere contrastive distribution on the Cartesian plane of allegedly prototypical languages. At the present stage of our knowledge, no language can really be considered prototypical.

4. Expanding the model

The algorithm described in § 2 aims at capturing the rhythmical behavior carried by the segments flow, which in turn affects (and is possibly affected by) the overarching accentual oscillations, as described in § 3. Which of these components is the dominant factor remains – for the time being – unclear, although one may want to assign this role to level-I due to the pervasive nature of phonotactics. What one can emphatically assert, in any case, is that the inter-level relation is not deterministic, for the two levels may diverge with respect to the C/C continuum. This, however, does not imply that no attempt should be undertaken to combine the two levels into a single design. It is indeed very tempting to reduce the two pairs of coupled oscillators described above to a three-fold cascade of hierarchically organized oscillators: accentual > syllable-peak/vocalic > consonantal. Although this goal will not be pursued here, we would like to briefly sketch the view behind it. As a matter of fact, the level-I vocalic oscillator, implementing the syllabic nucleus, may be conflated with the level-II syllable-peak oscillator. As for the consonantal oscillator, it clearly acts upon the vocalic one very much in the same way as the syllable-peak oscillator acts upon the accentual one at level-II. In terms of coupling, these tow sets of oscillators are strictly equivalent.

To check the latter claim, the formula in (2) was applied to the level-I oscillators, relating the duration of inter-V-onset intervals – from one V-onset to the next – to the number of intervening Cs.[10] Once again, *r* greater than or less than 1 indicates whether the overarching (vocalic) or the subordinated (consonantal) oscillator prevails.

The predictions are as follows: (i) In general, the consonantal oscillator should emerge as the dominant factor along with tempo increases, for the Cs comprised between two V gestures cannot be compressed beyond a certain threshold, whereas Vs allow for more compression;

[10] The relevance of the inter-V-onset interval as a rhythmic unity is underlined, e.g., by Keller & Port 2007.

(ii) In CONTR languages, however, due to the relative incompressibility of unstressed Vs, the higher stiffness of the vocalic oscillator should partly compensate the previous effect.

The computation's results (as applied to the same spontaneous Italian materials as above), whether calculated for 3 or 5 rate classes, appeared to be compatible with both expectations. As table 5 shows, the dominance of the consonantal oscillator increases from speed-1 to speed-2, but then begins to decrease towards the fastest rates. Needless to say, these predictions should be checked against other languages, particularly those expected to follow the COMPS pattern. The present authors are currently engaged in such a task.

Tempo	*segm.s/sec.*	*N*	*r*	*Tempo*	*segm.s/sec.*	*N*	*r*
T1	< 15,7 / av. 14.2	913	0.97	t1	< 14.9 / av. 13.6	561	1.01
T2	< 17.8 / av. 16.6	947	0.72	t2	< 16.1 / av. 15.4	573	0.74
T3	> 17.7 / av. 19.2	951	0.84	t3	< 17.2 / av. 16.6	549	0.78
				t4	< 18.9 / av. 17.9	556	0.81
				t5	> 18.8 / av. 20.0	572	0.84

Table 5. Output of the APD algorithm as applied to the rate classes naturalistically extracted from a spontaneous Italian corpus. Coupled oscillators: vocalic vs. consonantal.

The next step might possibly consist in extending the algorithm in (2) to a system of three cascaded oscillators – accentual, syllable-peak/vocalic, consonantal – thus attempting to model the combined effect of level-I and -II. This however should best be left for future research.

5. Conclusion

The primary purpose of this paper was to propose a unified and predictive NLR theory. Inevitably, the hypothesis presented here will in due time – perhaps very soon – be disconfirmed, but the present authors will not be upset about that: any theory's crisis, or even death, should be viewed as a step forward, paving the way to improved conceptions. It remains to be seen whether this sketch of a theory will be globally disconfirmed or only with respect to some of its predictions. Should the latter be the case, there would be room for reformulation of the details; alternatively, an entirely new hypothesis should be devised. Whatever the case, future theories will necessarily presuppose the spelling-out of explicit language features, from which specific rhythmical consequences can be derived. Returning somehow to the original spirit of the Pikean proposal, one should realize that NLR is the observable consequence of precise – albeit so-far poorly understood – structural properties, rather than a sort of phonetic primitive. The ultimate goal is to isolate and define those basic structural properties.

As noted above, the first results obtained should be checked against other linguistic materials. These should be selected out of conveniently sized corpora, for it is now clear that no meaningful conclusion can be drawn from scanty data. One should thus compare the present Italian data, stemming from spontaneous speech, both with read speech from the same language, and with speech from other languages, both read and spontaneous. It is however important to note that, once the topic is addressed within a sound epistemological perspective, cross-linguistic comparison becomes a useful – indeed necessary – tool for theory testing, rather than being the precondition for the results' assessment. The latter should rather follow from the constant interplay between predictions and results, progressively extended to a larger array of data.

It is equally important to observe that whoever engages in this research domain should be aware that this is a cumulative scientific enterprise. Whatever new insight one develops will rest on previous successes and failures, just as the model presented in this paper exploits a number of ideas developed by other scholars, whose inspiration is gratefully aknowledged by the authors. Hopefully, by joining efforts, a better understanding of this fascinating language aspect will be achieved.

References

Akamatsu, T. (1997). *Japanese Phonetics. Theory and Practice*. München / Newcastle: LINCOM.

Arvaniti, A. (in press). Rhythm, timing and the timing of rhythm. *Phonetica*.

Barbosa, P. (2006). *Incursões em torno do ritmo da fala*. Campinas: Pontes.

Barbosa, P. (2007). From syntax to acoustic duration: A dynamical model of speech rhythm production. *Speech Communication, 49*, 725-742.

Bauer, R.S. (1995). Syllable and word in Cantonese. *Journal of Asian Pacific Communication*, *6*, 245-306.

Bertinetto, P.M. (1981). *Strutture prosodiche dell'italiano. Accento, quantità, sillaba, giuntura, fondamenti metrici*. Firenze : Accademia della Crusca.

Bertinetto, P.M. (1989). Reflections on the dichotomy 'stress- vs. syllable-timing'. *Revue de Phonétique Appliquée*, *91/93*, 99-130.

Bertinetto, P.M. & Bertini, C. (2008). On modeling the rhythm of natural languages. *Proceedings of the 4th Speech Prosody Conference*. University of Campinas.

Bertinetto, P.M., Scheuer, S., Dziubalska-Kolaczyk, K. & Agonigi, M. (2007). Intersegmental cohesion and syllable division in Polish. *Proceedings of the 16th Int. Congress of Phonetic Sciences* (1953-1956). Universität Saarbrücken.

Bertinetto, P.M. & Fowler, C.A. (1989). On sensitivity to durational modifications by Italian and English speakers. *Rivista di Linguistica*, *1*, 69-94.

Bertini, C. & Bertinetto, P.M. (in press). Prospezioni sulla struttura ritmica dell'italiano basate sul corpus semispontaneo AVIP/API". *Atti del Convegno AISV 2007*. Università della Calabria.

Blevins, J. (2003). The independent nature of phonotactic constraints. In Féry, C. & Van De Vijver, R. (eds.), *The Syllable in Optimality Theory*. Cambridge: Cambridge University Press.

Boucher, V.J. (2006). On the function of stress rhythms in speech: Evidence of a link with grouping effects on serial memory. *Language and Speech*, *49*, 495-519.

Cummins, F. (2002). Speech rhythm and rhythmic taxonomy. *Proceedings of the 1st Speech Prosody Conference* (121-126) Aix en Provence.

Dauer, R.M. (1983). Stress-timing and syllable-timing reanalyzed. *Journal of Phonetics 11*, 51-62.

Dauer, R.M. (1987). Phonetic and phonological components of language rhythm. *Proceedings of the 11th Int. Congress of Phonetic Sciences* (vol.5, 447-50). Tallinn (former URSS).

Davidson, L. (2006). Schwa elision in fast speech: segmental deletion or gestural overlap? *Phonetica, 63*, 79-112.

Dell, F. & Halle, J. (to appear). Comparing musical textsetting in French and in English songs. In Aroui, J.-L. & Arleo, A.(eds.) *Towards a Typology of Poetic Forms.* Amsterdam: Elsevier.

Dellwo, V. (2004). Rhythm and speech rate: A variation coefficient for ΔC. In Karnowski, P., Szigeti, I. (eds.), *Language and language-processing* (231-241). Frankfurt am Main: Peter Lang.

Dellwo, V., Steiner, I., Aschenberner, B. & Dankovicova, J. & Wagner, P. (2005). Bonn-Tempo Corpus and Bonn-Tempo Tools: A database for the study of speech rhythm and rate. *Proceedings of Interspeech 2004 – 8th International Conference of Spoken Language Processing*. Jeju Island, Korea.

Dressler W. U. & Dziubalska-Kołaczyk, K. (2006). Proposing morphonotactics. *Italian Journal of Linguistics, 18*, 249-266.

Dziubalska-Kolaczyk, K. (2002). *Beats and Binding Phonology*. Frankfurt a.M.: Lang.

Eriksson, A. (1992). *Aspects of Swedish Speech Rhythm*. PhD dissertation, University of Göteborg.

Fowler, C.A. (1977). *Timing Control in Speech Production*. Indiana University Linguistics Club.

Gavaldá, N. & Dellwo, V. (in preparation). Vowel reduction and Catalan speech rhythm.

Goldstein, L., Chitoran, I. & Selkirk, E. (2007). Syllable structure as coupled oscillator modes: Evidence from Georgian vs. Tashlhiyt Berber. *Proceedings of the 16th Int. Congress of Phonetic Sciences* (241-244). Universität Saarbrücken.

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

Hoeqvist, Ch. Jr. (1983). Syllable duration in stress-, syllable- and mora-timed languages. *Phonetica*, *40*, 203-237..

Keller, E. & Port, R. (2007). Speech timing: Approaches to speech rhythm. *Proceedings of the 16th Int. Congress of Phonetic Sciences* (327-329). Saarbrücken.

Kohler, K. (in press). Rhythm in speech and language. A new research paradigm. *Phonetica.*

Lindblom, B. & Rapp, K. (1973). Some Temporal Regularities of Spoken Swedish. *Papers of the Institute of Linguistics*, *21*. University of Stockholm.

Low, E.L. (1998). Prosodic Prominence in Singapore English. PhD dissertation, Univ. of Cambridge.

Marotta, G. (2003). What does Phonology tell us about Stress and Rhythm? Some Reflections on the Phonology of Stress. In Solé, M.J., Recasens, D., Romero, J. (eds.), *Proceedings of the 15th International Congress of Phonetic Sciences* (vol. 1, 333-336). Barcelona.

Mairano, P. & Romano, A. (2008). A comparison of four rhythm metrics for six languages. Poster presented at the conference on Empirical Approaches to Speech Rhythm. University College London.

Mok, P.P.K. (in preparation). Using durational measures with non-native speech rhythm.

Mok, P.P.K. & Dellwo, V. (2008). Comparing native and non-native speech rhythm using acoustic rhythmic measures: Cantonese, Beijing Mandarin and English. *Proceedings of the 4th Speech Prosody Conference*. University of Campinas.

Nolan, F. & L. Asu (in press). The Pairwise Variability Index (PVI) and co-existing rhythms in language, again. *Phonetica.*

O'Dell, M. & Nieminen, T. (1999). Coupled oscillator model of speech rhythm. *Proceedings 14° Int. Congress Phonetic Sciences* (vol. 2, 1075-1078). Berkeley University.

O'Dell, M., Lennes, M., Werner, S. & Nieminen, T. (2007). Looking for rhythms in conversational speech. *Proceedings of the 16th Int. Congress of Phonetic Sciences* (1201-1204). Saarbrücken.

Ohala, J. J. & Kawasaki-Fukumori, H. (1997). Alternatives to the sonority hierarchy for explaining segmental sequential constraints. In Eliasson, S., Jahr, E. H. (eds.), *Language and Its Ecology: Essays in memory of Einar Haugen* (343-365). Berlin: Mouton de Gruyter.

Pike, K.L. (1945). *The Intonation of American English*. Ann Arbor, Mich..

Price, P.J. (1980). Sonority and syllabicity: acoustic correlates of perception. *Phonetica 37*, 327-43.

Ramus, F., Nespor, M. & Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition 73*, 265-292.

Rouas, J.L. & Farinas, J. (2004). Comparaison de méthodes de caractérisation du rythme des langues. *Workshop MIDL*. Paris.

Steriade, D. (1998). Alternatives to syllable-based accounts of consonantal phonotactics. In O.Fujimura, B.D.Joseph & B.Palek (eds.), *Proceedings of LP 1998* (205-245).

Van Santen, J.P.H. & Shih, C. (2000). Suprasegmental and segmental timing models in Mandarin and American English. *Journal of the Acoustical Society of America*, *107*, 1012-1026.

Vatikiotis-Bateson, E. (1988). *Linguistic Structure and Articulatory Dynamics*. Indiana University Linguistics Club.

Vékás, D. & Bertinetto, P.M. (1991). Controllo vs. compensazione: sui due tipi di isocronia. In Magno Caldognetto, E,, Benincà, P. (eds.), *L'interfaccia tra fonologia e fonetica* (155-162). Padova: Unipress.

Vennemann, Th. (1994). Universelle Nuklearphonologie mit epiphänomenale Silbenstruktur. In Ramers, K. H.,Vater,, H,, Wode H. (eds.), *Universale phonologische Strukturen und Prozesse*. Tübingen: Niemeyer.

Wagner, P. (2007). Visualizing levels of rhythmic organization. *Proceedings of the 16th Int. Congress of Phonetic Sciences* (1113-1116). Universität Saarbrücken.

White, L. & Mattys, S. (2007). Calibrating rhythm: First language and second language studies. *Journal of Phonetics*, *35*, 501-522.