COUPLED OSCILLATOR MODEL OF SPEECH RHYTHM

Michael L. O'Dell and Tommi Nieminen
University of Tampere, Finland

ABSTRACT

Attempts to verify instrumentally tendencies toward stress timing or syllable timing have been less than successful, leading many researchers to abandon the rhythm dichotomy as too simplistic. One empirical finding which suggests an underlying unity in the rhythms of different languages is a statistical tendency for duration of stress group to be a linear function of the number of syllables contained in it, with languages differing mainly in the constant term of the linear function. We use a robust mathematical model of simple coupled oscillators to argue that the linear relation observed may reflect a very general tendency of rhythmic systems organized in hierarchical interaction, regardless of concrete details, which may differ in complex ways. In addition, the oscillator model provides a rich conceptual tool for elucidating rhythmical differences in speech behavior.

1 INTRODUCTION

Speech rhythm commonly refers to a temporal patterning of elements, typically syllables, in the flow of speech. Such patterning requires that syllables be classified into at least two categories, such as stressed vs. unstressed, or strong vs. weak. Since Pike [18], in what may be called the traditional account of speech rhythm, languages have been categorized either as **stress-timed** or **syllable-timed**. Stress-timed languages are supposed to have a simple rhythmical pattern with stresses (or, to be precise, the rhythmical beats of the stressed syllables) at equal distances, whereas syllable-timed languages are supposed to have a simple rhythmical pattern with equal-length syllables. In short, the dichotomy presents speech rhythm ultimately as a simple phenomenon of equal beats connected with either of two linguistic units.

This notion has been widely rejected on the basis of empirical data [4, 20], since attempts to verify instrumentally tendencies toward regular intervals either of stresses or of syllables have been less than successful. Many researchers have abandoned the speech dichotomy as too simplistic, claiming that speech rhythm can only be regarded as a complex and multi-variabled phenomenon. On this account, no simple rhythm-generator is available to study, and the perceived rhythm, whatever it is like, does not manifest simple patterning of elements. For example, Dauer [3] sees rhythm as 'the result of the interaction of a number of components', and transforms the dichotomy into a difference between languages with 'stronger' or 'weaker' rhythm.

However, there is also an empirical finding that seems to suggest an underlying unity in the rhythms of different languages. Eriksson [4] has pointed out that there appears to be a strong statistical tendency for duration of the stress group (measured as **interstress interval** or **ISI**, i.e. the duration between two successive stresses) to be a simple linear function $I = a + bn$ of the number of syllables n contained in it, with languages differing mainly in the constant term a. Such a linear relation for total duration is also compatible with the 'minimum duration' equations used by Klatt [9] and others for segment durations. Eriksson used linear regression to reanalyze Dauer's [2] data of the mean durations of n-syllable stress groups, with n ranging from 1 to 4, in five languages ('stress-timed' English and Thai and 'syllable-timed' Spanish, Greek, and Italian). The results of the regression analysis are given in Table 1. As can be seen, the values cluster around 100 ms for the slope coefficient b and 100 ms or 200 ms for the constant a. The only thing markedly differing between languages is the constant, which falls roughly into two groups following the traditional timing dichotomy: about 100 ms for syllable-timed languages, about 200 ms for stress-timed languages.

English	$I = 201 + 102n$	$r = 0.996$
Thai	$I = 220 + 97n$	$r = 0.973$
Spanish	$I = 76 + 119n$	$r = 0.997$
Greek	$I = 107 + 104n$	$r = 1.000$
Italian	$I = 110 + 105n$	$r = 1.000$

Table 1. Linear regression equations and correlation coefficients (r) for five languages using Dauer's [2] data.

The benefits of this analysis can be summarized thus: (a) the traditional timing dichotomy without a common term is transformed into a one-variable scale; (b) this scale groups the languages in exactly the same way as the timing dichotomy so that (c) an empirical validation is given to the traditional dichotomy *without* the suggestion of simple rhythmical organization of the stress groups. The correlation coefficients are remarkable, ranging from 0.973 in Thai to 1.000 in Greek and Italian, but one must remember that the formulas were calculated from averaged data. Admittedly, Eriksson's is not the first proposal for changing the timing dichotomy to a one-variable scale, but it differs from most others by deriving its impetus from empirical data. Of course, definite conclusions as to the generalizability of the account can only be reached when languages from outside the stated timing distinctions are studied—languages such as 'mora-timed' Japanese [8, 19] or 'foot-timed' Finnish and Estonian [13, 21]—but even with this reservation, Eriksson can be claimed to have found a feature worth further research. Nieminen [15] calculated a linear regression for his Finnish material, resulting in $I = 132 + 143n$, which is roughly midway between the stress- and syllable-timed extremes, as far as the constant is concerned. This perhaps is not surprising considering that Finnish has been notoriously difficult to place in the timing dichotomy—cf. [14].

Mathematical formulas estimated from empirical data do not explain anything by themselves, they are just a means of categorizing languages. Explanation demands at least a suggestion of the mechanism *underlying* the rhythmical units; Eriksson acknowledges this too. The main difference between the language groups seems to be in the constant term. The 'natural' interpretation (as Eriksson says) is that the constant term represents an extra duration included in the stressed syllable. This would mean that the difference between stress-timed and sylla-

ble-timed languages is only that stressed syllables are longer in stressed-timed languages. Actually it is easy to demonstrate that the figures given in Table 1, especially as they are mean values of a speech sample, do not tell us anything about the internal organization of the stress group [4]. We cannot say if there is syllabic compression at all, and if there is, which syllables it applies to. The 'extra' duration could be a part of the stressed syllable, part of the stress group as a unit (for instance, as final lengthening), distributed evenly to all syllables, or even to various syllables at random. In fact, the 'natural' interpretation is contradicted by empirical data. For instance in French (allegedly a syllable-timed language) the stressed syllables are markedly longer than the unstressed ones, even more so than in the stress-timed languages [5]. Also, it has long been established that in numerous languages compression does occur (in all syllables) as the number of syllables increases (cf. eg. [12, 16] and references therein).

2 COUPLED OSCILLATORS AND APD THEORY

In recent years there has been much success in the modeling of biological rhythmic behavior using coupled oscillators. The basic idea is to assume the existence of subrhythms which would exhibit simple oscillatory behavior if observed in isolation. When oscillators are combined into larger systems so that they influence each other, the resulting patterns of rhythm may be much more complex than those of the component oscillators. In some cases, enough is known about the mechanisms underlying a particular behavior, that detailed models of component oscillators and the ways they influence each other (coupling) may be attempted. In many other cases the mechanisms leading to rhythmic behavior are not understood in detail, or can only be guessed at. Fortunately, however, much of the macroscopic behavior of systems of oscillators is relatively insensitive to the exact details of the oscillators or the couplings involved. A mathematical technique called APD theory (for *Averaged Phase Difference*) has been developed which is abstract enough to derive qualitative conclusions about collections of oscillators in spite of minimal knowledge of the details of the components (cf. [10]). The essence of this technique is twofold. First, any descriptions of oscillating subsystems are reparameterized in coordinates of phase relative to the system's own limit cycle attractor, or 'natural oscillation', reducing the variables involved to phase. If no previous physical description is available we may assume this transformation has been applied and start with a simple phase description. Operating on its own, such a subsystem will be characterized by

$$\dot{\theta} = \omega \tag{1}$$

that is, the derivative (or rate of change) of the oscillator's phase (θ) is a constant (ω) expressing the oscillator's 'natural' rhythm or eigenfrequency. The next step is to consider the interaction of two (or more) such oscillators, each with its own eigenfrequency. Even with the above simplification, this interaction could in general be a complicated function of the phases of each of the subsystems, but a further simplification is utilized in APD theory. For each subsystem the effects at each phase *difference* are averaged over an entire cycle, giving a simple characterization of the total system in terms of constant eigenfrequencies (ω) along with couplings dependent only on phase differences (ϕ). For instance, for a coupled system of two oscillators, we have

$$\begin{aligned} \dot{\theta}_1 &= \omega_1 + H_1(\phi) \\ \dot{\theta}_2 &= \omega_2 - H_2(\phi); \quad \phi = \theta_2 - \theta_1 \end{aligned} \tag{2}$$

It then becomes possible to study the behavior of a relatively simple model which nonetheless qualitatively reflects the behavior of the more complex underlying system in a wide range of situations and with very mild assumptions. Using this technique it should be possible to model speech rhythms as collections of coupled oscillators, and possibly draw some general conclusions. In the case of syllables and stress groups, we need a coupling function which changes according to the number of (intended) syllables per stress group. In other words, the idea will be to assume a 'stress group oscillator' and a 'syllable oscillator' coupled together by a function that depends on *n*, the number of syllables per stress group. Each oscillator will have its own eigenfrequency which we designate ω_1 for the stress group oscillator and ω_2 for the syllable oscillator. We assume the coupling influences may be expressed as a function of a quantity

$$\phi_n = \theta_2 - n\theta_1 \tag{3}$$

with *n* the number of syllables per stress group. If we further assume that the two coupling functions are identical in form but opposite in sign, varying only in relative strength, we arrive at the following system:

$$\begin{aligned} \dot{\theta}_1 &= \omega_1 + H(\phi_n) \\ \dot{\theta}_2 &= \omega_2 - rH(\phi_n) \end{aligned} \tag{4}$$

where *r* indicates the relative strength (or dominance) of the stress group over the syllable. To find an equilibrium solution, we set the derivative of ϕ_n to zero:

$$\dot{\phi}_n = (\omega_2 - n\omega_1) - (r+n)H(\phi_n) = 0 \tag{5}$$

which gives the coupling function an equilibrium value of

$$H(\phi_n) = \frac{\omega_2 - n\omega_1}{r+n} \tag{6}$$

The period of the stress group oscillator (eg. the time from stress to stress, or interstress interval) at such an equilibrium (if it exists) can then be calculated as a function of *n*:

$$T_1(n) = \frac{1}{\omega_1 + H(\phi_n)} = \frac{r}{r\omega_1 + \omega_2} + \frac{1}{r\omega_1 + \omega_2} n \tag{7}$$

The period is thus a linear function in *n* of the form $I = a + bn$ used by Eriksson [4]. Therefore it is to be expected on the basis of this general model of two rhythms hierarchically coupled (ie. 1 : *n*) that the period of the slower rhythm will tend to a linear function of *n*, at least when there exists a stable solution to equations (3) and (4).

Interpretation of regression coefficients. If *a* and *b* in Eriksson's formula are estimated empirically, as in Table 1, then the 'relative strength' parameter *r* of equation (4) can be estimated as a/b. If Eriksson's formula is rewritten as $I = b(r + n)$, it can be seen that the regression line intersects the *n*-axis at $-r$, and has *b* as its slope. Differences between syllable- and stress-timed languages could be mainly a matter of this relative influence between syllables and stress groups, with

stress group dominating in stress-timed languages ($r > 1$) but not in syllable-timed languages ($r \leq 1$). The slope parameter b depends on the oscillator eigenfrequencies and will therefore reflect differences in tempo. Figure 1 shows a plot of r vs. b for the languages in Table 1. A point for Finnish is also shown based on [15]. If languages are categorized according to r instead of a, Finnish would appear to fit in the syllable-timed group.

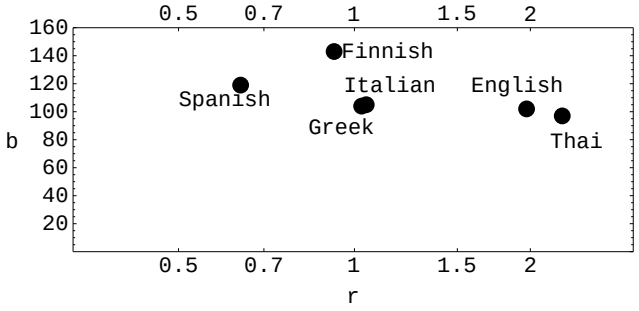

Figure 1. Relative coupling strength (r) vs. regression slope (b) for various languages.

3 LIMITATIONS

Admittedly the model expressed by equations (3) and (4) is very abstract. Indeed, this was the main motivation for using it. What would be the consequences if some of the assumptions of the model are violated when a more realistic description is available? Under what conditions will the conclusion be upheld that there is a tendency for ISI to be a linear function of the number of syllables? In this section we consider the severity of some limitations of the basic model.

Phase walk-through. The relation expressed in equation (7) depends on the existence of an equilibrium solution. However, if the coupling function $H(\phi_n)$ is finite, there must be an n large enough that no equilibrium exists. When this happens, a phenomenon known as 'phase walk-through' occurs [7] in which the oscillators don't keep step, but continuously 'slip' relative to each other. At this point the linear relation of equation (7) will break down and no general formula for the ISI can be found. It is not clear whether this phenomenon has any relevance for speech, since it could well be that this phase walk-through would occur only at values of n not observed in speech.

Duration of stressed syllable. It is well known that in many languages stressed syllables are longer than unstressed syllables, *ceteris paribus*. In our model this is equivalent to saying that the equation for syllable frequency includes a 'stress function' depending on stress group phase, which will slow the syllable down in the vicinity of some particular phase representing stress. This change obviously means that the equilibrium point will vary throughout each stress cycle. However, integrating the stress function over one stress cycle will still contribute only a constant value, so that the linear form of the interstress interval remains unchanged. The same comment applies to any other change in duration at a particular position within the stress group—such as final lengthening.

Relaxing the limit cycle requirement. One aspect of the general model which may well be questioned is whether each subsystem would exhibit oscillation on its own. Actually, APD theory can be extended to at least some systems with components which are only nearly oscillatory (excitable systems or 'one shot oscillators', cf. [10]). We have found through computer simulation that a model with a 'one shot syllable' can nevertheless exhibit a roughly linear relation between number of syllables and ISI.

Discrete (multiple pulse) interactions. Also, it may well be that interactions between stress group and syllable production are not continuous but discrete, confined to a few points around the cycle. Of course such influences can be formally averaged as APD theory requires, but the question arises to what extent the (qualitative) behavior of the system is preserved under this averaging. Ermentrout and Kopell considered this question [6] and found that the distortion introduced by averaging is small providing there are enough interaction pulses around the cycle. Again using computer simulation, we found that a system with interactions limited to very few pulses can indeed exhibit a roughly linear relation between number of syllables and interstress interval. However, the qualitative behavior predicted by APD theory is not valid when interactions between oscillators are confined to one pulse per superordinate period. For this reason the present model is not directly applicable, for example, to the interesting results of so called 'speech cycling' experiments [1], in which repetitive speech is synchronized with an external pulse-like signal.

4 ELABORATIONS AND APPLICATIONS

We next consider briefly some ways in which the basic model can be modified to accommodate additional aspects of rhythmic variation.

Effect of differing syllable types. We may wish to classify different syllables according to some *a priori* scheme in an attempt to explain some of the variation evident in syllable rate. It is certainly to be expected that different syllable types will correspond to different values of eigenfrequency (ω_2). How can these differences be incorporated into the model? From equations (4) and (6) we can easily solve for the period of the syllable cycle at equilibrium, giving $T_2 = a/n + b$, with a and b the same as in Eriksson's equation, both depending on ω_2 (as well as ω_1). If ω_2 is allowed to assume different values for different syllable types, then each syllable type i will have its own a_i and b_i, and interstress interval can be computed as an appropriate sum:

$$T_1(n) = \sum n_i \left(\frac{a_i}{n} + b_i\right); \quad n = \sum n_i \tag{8}$$

with n_i syllables of type i. If the data contains enough cases, the a_i and b_i parameters can be estimated by multiple regression, as illustrated in [15, 17] for short syllables (open syllables with short vowels) vs. long syllables (all other types) in the Finnish material. If the coupling is assumed to be independent of syllable type, each a_i and b_i pair provides an estimate of $r = a_i/b_i$. For Nieminen's Finnish data regression yields $a = 95$, $b = 117$ ($r = 0.81$) for short syllables, and $a = 148$, $b = 163$ ($r = 0.91$) for long syllables. An estimate of r can also be obtained directly using non-linear regression. This may be more desirable especially if the range of n is small. This technique gives an estimate of $r = 0.88$ for Nieminen's data. Also the differences (but not the absolute values) of the eigenfrequencies for different syllable types can then be estimated as $1/b_i - 1/b_j$. For Nieminen's data this gives an estimated difference of 2.57 Hz between short and long syllables.

Adding hierarchical levels. What happens to the linear

relation between number of cycles (say syllables) at one level and the period of a superordinate cycle (say stress group) when the model is expanded to include several hierachical levels? This is an interesting question, since speech rhythm has often been described, at least for some languages, in just these terms, for instance with an additional mora level below the syllable, or a foot level between the syllable and stress group. If we expand the above model to include $k + 1$ oscillators instead of two, keeping the assumption of strict hierarchy so that the oscillators form a chain with coupling between neighbors only, we get the following set of equations:

$$\begin{aligned}
\dot{\theta}_1 &= \omega_1 + H_1(\phi_1) \\
\dot{\theta}_2 &= \omega_2 - r_1 H_1(\phi_1) + H_2(\phi_2) \\
&\cdots \\
\dot{\theta}_k &= \omega_k - r_{k-1} H_{k-1}(\phi_{k-1}) + H_k(\phi_k) \\
\dot{\theta}_{k+1} &= \omega_{k+1} - r_k H_k(\phi_k)
\end{aligned} \tag{9}$$

Here $\phi_i = \theta_{i+1} - n_i\theta_i$, and n_i is the number of θ_{i+1} cycles per θ_i cycle, for all i from 1 to k. Setting the derivatives of all the $\phi_i = 0$ for the equilibrium and solving for $H_1(\phi_1)$ leads eventually to an expression for the period of the slowest (θ_1) oscillator which is a linear function of all the numbers of different subunits contained in it: $T_1 = c_0 + c_1(n_1) + c_2(n_1n_2) + \ldots + c_k(n_1n_2\ldots n_k)$. The 'relative strength' parameters in equation (9) are equal to the ratios of adjacent coefficients: $r_i = c_{i-1}/c_i$. This suggests using multiple regression on empirical data to estimate the relative strengths of the couplings between the various levels of such a hierarchical model.

Variation within a single language. It has been suggested that a model of interacting oscillators may help to organize observed rhythmical differences between languages. It remains, however, to investigate the extent to which the parameters estimated in the model vary within a single language, from speaker to speaker, or in different situations. To this end, we offer an analysis of some additional data for Finnish. The first is based on an extensive list of measured word durations included by Laurosela in his 1922 description of the Ostrobothnian dialect of Finnish [11]. Estimating r with non-linear regression as above for Nieminen's data gives a rather higher value of $r = 1.84$, which would seem to place Finnish (at least this dialect) in the stress-timed category. However, we suggest that the difference may be in large part due the very different speech situation: pronouncing a list of words. Interestingly Laurosela's data give an estimate for the difference between short and long syllables very close to that for Nieminen's data: 2.36 Hz. We have also made a preliminary analysis of data from an informal interview with a Finnish woman. Besides being a different speaker, this data differs from Nieminen's data in representing a very spontaneous speaking style. Preliminary results indicate a much smaller value for relative coupling strength: $r = 0.44$. It would appear that caution is warranted in interpreting relative coupling strength as an indicator of language type. Also Eriksson's regression data for individual speakers of Swedish in an identical reading task [4] show a wide range of estimated relative coupling strength, from $r = 1.09$ to $r = 5.76$.

5 CONCLUSION

Rhythm is indeed a complex phenomenon influenced by many factors, and the mechanisms responsible for producing rhythmic patterns may differ in complex ways from language to language. However, our model of interacting oscillators leads us to the conclusion that certain traits of speech rhythms, such as the linear relationship noted by Eriksson [4], may be very general in spite of differences in details, perhaps reflecting tendencies for any hierarchically organized rhythmic behavior. The general oscillator model utilized here allows an interpretation of some such aspects without recourse to unknown details of the underlying mechanisms involved.

REFERENCES

[1] Cummins, F., R. F. Port 1998. Rhythmic constraints on stress timing in English. *Journal of Phonetics* 26, 145–171.

[2] Dauer, R. M. 1983. Stress timing and syllable-timing reanalyzed. *Journal of Phonetics* 11, 51–62.

[3] Dauer, R. M. 1987. Phonetic and phonological components of language rhythm. *Proceedings XIth ICPhS*. U.S.S.R. Academy of Sciences of the Estonian S.S.R., Tallinn. Vol. 5, 447– 450.

[4] Eriksson, A. 1991. *Aspects of Swedish Speech Rhythm*. University of Göteborg, Göteborg.

[5] Delattre, P. 1966. A comparison of syllable length conditioning among languages. *International Review of Applied Linguistics* 4, 183–198.

[6] Ermentrout, G. B., N. Kopell 1991. Multiple pulse interactions and averaging in systems of coupled neural oscillators. *Journal of Mathematical Biology* 29, 195-217.

[7] Ermentrout, G. B., J. Rinzel 1984. Beyond a pacemaker's entrainment limit: phase walk-through. *American Journal of Physiology* 246, R102–R106.

[8] Han, M. S. 1994. Acoustic manifestations of mora timing in Japanese. *Journal of the Acoustical Society of America* 96, 73–82.

[9] Klatt, D. H. 1973. Interaction between two factors that influence vowel duration. *Journal of the Acoustical Society of America* 54, 1102–1104.

[10] Kopell, N. 1988. Toward a theory of modelling central pattern generators. A. H. Cohen, S. Rossignol and S. Grillner (eds.): *Neural Control of Rhythmic Movements in Vertebrates*. John Wiley and Sons, New York. 369-413.

[11] Laurosela, J. 1922. *Foneettinen tutkimus Etelä-Pohjanmaan murteesta*. SKS, Helsinki.

[12] Lehiste, I. 1970. *Suprasegmentals*. The M. I. T. Press, Cambridge (Mass.).

[13] Lehiste, I. 1990. Phonetic investigation of metrical structure in orally produced poetry. *Journal of Phonetics* 18, 123–133.

[14] Miller, M. 1984. On the perception of rhythm. *Journal of Phonetics* 12, 75–83.

[15] Nieminen, T. 1996. *Suomen kielen puherytmi*. Master's Thesis. Department of Finnish and General Linguistics, University of Tampere.

[16] Nooteboom, S. G. 1972. *Production and Perception of Vowel Duration*. A study of durational properties of vowels in Dutch. Philips Research Laboratories, Eindhoven.

[17] O'Dell, M. L., T. Nieminen 1998. Reasons for an underlying unity in rhythm dichotomy. *Linguistica Uralica* XXXIV 3, 178–185.

[18] Pike, K. L. 1945. *The Intonation of American English*. University of Michigan Press, Ann Arbor.

[19] Port, R. F., J. Dalby, M. O'Dell 1987. Evidence for mora timing in Japanese. *Journal of the Acoustical Society of America* 81, 1574–1585.

[20] Roach, P. 1982. On the distinction between 'stress-timed' and 'syllable-timed' languages. David Crystal (ed.): *Linguistic Controversies*. Essays in linguistic theory and practice in honour of F. R. Palmer. Edward Arnold, London. 73–79.

[21] Wiik, K. 1991. On a third type of speech rhythm: foot timing. *ICPhS / XII Aix-en-Provence 1991*. Aix-en-Provence: Université de Provence 1991. Vol. 3, 298–301.