MICHAEL L. O'DELL & TOMMI NIEMINEN (Tampere)

REASONS FOR AN UNDERLYING UNITY IN RHYTHM DICHOTOMY

PREFACE: RHYTHM DICHOTOMY OR NOT?

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 (1945), 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 (see Eriksson 1991: 20–36 for summary; Roach 1982), 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. Therefore, no simple rhythm-generator should be available to study, and the perceived rhythm, whatever it is like, should not manifest simple patterning of elements. For example, Dauer (1987) 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 (ibid., 447).

However, there is also an empirical finding that seems to suggest an underlying unity in the rhythms of different languages. 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 of the function (Eriksson 1991: 40–44[1]). Eriksson used linear regression to reanalyze Dauer's (1983) 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: for syllable-timed languages it is about 100 ms, for stress-timed languages about 200 ms.

Table 1. Linear regression equations and correlation coefficients for five languages using Dauer's (1983) data.

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	$I = 110 + 105n$	$r = 1.000$

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 (r) 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

[1] A linear relation for total duration is also compatible with the 'minimum duration' equations used by Klatt (eg. 1973) and others.

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 (Port et al. 1987, Han 1994) or ‘foot-timed’ Finnish and Estonian (Lehiste 1990, Wiik 1991)—but even with this reservation, Eriksson can be claimed to have found a feature worth further research. (Nieminen (1996: 92) calculated a linear regression for 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 could be expected considering that Finnish has been notoriously difficult to place in the timing dichotomy—cf. Miller 1984.)

The formulas obtained mathematically 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 syllable-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 (Eriksson 1991: 46–47). 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 (Delattre 1966: 190, 193). Also, it has long been established that in numerous languages compression does occur (in all syllables) as the number of syllables increases (cf. eg. Lehiste 1970: 40–41, Nooteboom 1972: 62–71 and references therein).

COUPLED OSCILLATORS AND APD THEORY

In recent years there has been much success in the modelling 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. Kopell 1988). 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

$$\dot{\theta} = \omega \qquad (1)$$

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

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

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 eigen-frequency 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:

$$\dot{\theta}_1 = \omega_1 + H(\phi_n)$$
$$\dot{\theta}_2 = \omega_2 - r\,H(\phi_n) \tag{4}$$

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

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

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 (1991). 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 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*. On a plot of *a* vs. *b*, this can be seen as the slope of a line through the origin and the point (*b*,*a*) as illustrated in Figure 1 for the coefficients in Table 1. A point for Finnish is also shown in Figure 1 (filled circle) based on Nieminen 1996. If languages are categorized according to *r* instead of *a*, Finnish would appear to fit in the 'syllable-timed' group.

If ω_1 and ω_2 are held constant in the model while *r* is allowed to vary, *a* and *b* will be inversely proportional with intercepts $1/\omega_1$ and $1/\omega_2$. Actually, it would appear that the languages of Table 1 are *roughly* related in this fashion (cf. the dashed line in Figure 1), which invites the interpretation that these languages *could* have syllables and stress groups with roughly the same eigenfrequencies, the differences being mainly a matter of the relative influence between syllables and stress groups.

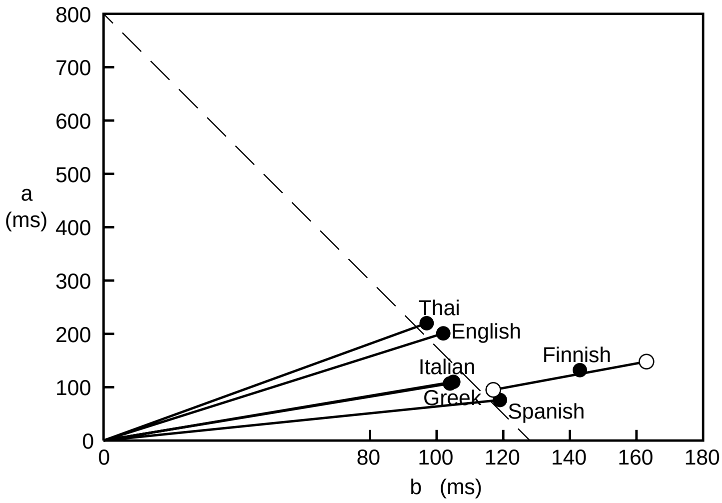

Figure 1. Relationships between coefficients *a* and *b*.

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 (Ermentrout & Rinzel 1984) 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 occurs at values of *n* not observed in speech.

MODIFICATIONS OF THE MODEL

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 of adding details in an attempt to make the model more realistic? 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? We now consider briefly some modifications of the basic model and their qualitative consequences.

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' $K(\theta_1)$ depending on stress group phase, which will slow the syllable down in the vicinity of some particular phase, eg. $\theta_1 = 0$, representing stress: $\theta_2 = \omega_2 - K(\theta_1) - rH(\phi_n)$. 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.

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 equation (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:

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

with n_i syllables of type *i*. If there are enough cases, the a_i and b_i parameters can be estimated by multiple regression, as illustrated in Nieminen (1996) for short syllables (open syllables with short vowels) vs. long syllables (all other types) in his Finnish material. In Figure 1 there are two (connected) open circles shown for Finnish based on Nieminen's data: closer to the origin the point $a = 95$, $b = 117$ for short syllables, farther from the origin the point $a = 148$, $b = 163$ for long syllables. If the coupling is assumed to be constant, these two points each provide an estimate of $r = a_i/b_i$ and should lie on a line through the origin. Interestingly the point for short syllables is roughly in line with the other languages plotted, while the point for long syllables is markedly different.

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. Kopell 1988). 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 (Ermentrout & Kopell 1991) 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.

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 interstress interval) 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: $T_1 = c_0 + c_1(n_1) + c_2(n_1n_2) + \ldots + c_k(n_1n_2\ldots n_k)$, of all the numbers of different (sub)units it contains. 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.

CONCLUSION

The model developed here can be considered a synthesis of the simple and complex conceptions of speech rhythm. We believe that 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 roughly linear relationship noted by Eriksson, may be very general in spite of differences in details. We suggest this result may in fact be even more general, reflecting tendencies for any hierarchically organized rhythmic behavior.

REFERENCES

D a u e r , R. M. 1983, Stress-timing and syllable-timing reanalyzed. — Journal of Phonetics 11, 51–62.

---- 1987, Phonetic and phonological components of language rhythm. — Proceedings Xith ICPhS. Tallinn: U.S.S.R. Academy of Sciences of the Estonian S.S.R. 1987. Vol. 5, 447– 450.

E r i k s s o n , A. 1991, Aspects of Swedish Speech Rhythm. Göteborg: University of Göteborg.

D e l a t t r e , P. 1966, A comparison of syllable length conditioning among languages. — International Review of Applied Linguistics 4, 183–198.

E r m e n t r o u t , G. B., K o p e l l , N. 1991, Multiple pulse interactions and averaging in systems of coupled neural oscillators. — Journal of Mathematical Biology 29, 195-217.

E r m e n t r o u t , G. B., R i n z e l , J. 1984, Beyond a pacemaker's entrainment limit: phase walk-through. — American Journal of Physiology 246, R102–R106.

H a n , M. S. 1994, Acoustic manifestations of mora timing in Japanese. — Journal of the Acoustical Society of America 96, 73–82.

K l a t t , D. H. 1973, Interaction between two factors that influence vowel duration. — Journal of the Acoustical Society of America 54, 1102–1104.

K o p e l l , 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. New York: John Wiley and Sons. Pp. 369-413.

L e h i s t e , I. 1970, Suprasegmentals. Cambridge (Mass.): The M.I.T. Press.

---- 1990, Phonetic investigation of metrical structure in orally produced poetry. — Journal of Phonetics 18, 123–133.

M i l l e r , M. 1984, On the perception of rhythm. — Journal of Phonetics 12, 75–83.

N i e m i n e n , T. 1996, Suomen kielen puherytmi. Pro gradu thesis. Department of Finnish and General Linguistics, University of Tampere.

N o o t e b o o m , S. G. 1972, Production and Perception of Vowel Duration. A study of durational properties of vowels in Dutch. Eindhoven: Philips Research Laboratories.

P i k e , K. L. 1945, The Intonation of American English. Ann Arbor: University of Michigan Press.

P o r t , R. F., D a l b y , J., O ' D e l l , M. 1987, Evidence for mora timing in Japanese. — Journal of the Acoustical Society of America 81, 1574–1585.

R o a c h , 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. London: Edward Arnold. Pp. 73–79.

W i i k , 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.