SPEECH RHYTHMS AS CYCLICAL ACTIVITY

Michael L. O'Dell & Tommi Nieminen

Department of Finnish and General Linguistics
University of Tampere

ABSTRACT

Utilizing a mathematical technique known as Average Phase Difference theory we derive some general predictions for interacting hierarchical speech rhythms such as syllable, stress group and breath group rhythms. We present additional data confirming various predictions of the coupled oscillator model.

1. INTRODUCTION

Speech exhibits various rhythms manifested by repetitions of many different kinds. Many of these form hierarchical cycles—faster, lower level cycles repeating within slower, higher level cycles. One example of this which has received some attention in the literature is the case of syllables within a stress cycle. The evidence which has been accumulating for various languages would appear to indicate that the relationship between such hierarchical rhythms very generally represents a compromise of sorts: the average period of the higher rhythm generally grows as it includes or synchronizes with more and more periods of the lower rhythm, but the period of the lower rhythm also shrinks to some extent. Thus very typically, for instance, average syllable duration diminishes somewhat while the time elapsing between consecutive stresses (interstress interval, or ISI) grows when the number of syllables per ISI increases.

Utilizing a mathematical technique known as Average Phase Difference theory (APD, see [4]) we have derived some general predictions for interacting hierarchical rhythms. The idea is that such speech rhythms can very generally be modeled as a set of oscillators each with its own eigenfrequency (at which it would oscillate in isolation), which influence each other so as to achieve synchrony, for instance the coupling might be such that one oscillator would complete one cycle in synchrony with 4 cycles of an oscillator corresponding to a lower rhythm. An important feature of the model is that the mutual influence of speech rhythms is handled in a continuous manner, allowing the application of dynamic systems theory. The details of the mathematical derivation can be found in [8] and [9]. Figure 1 shows the arrangement for two oscillators influencing each other. The eigenfrequencies are denoted by ω_1 and ω_2, ϕ is a generalized phase difference between the two oscillators, $H(\phi)$ is the coupling function and r is a parameter which expresses the relative strength of the coupling influence in opposite directions.

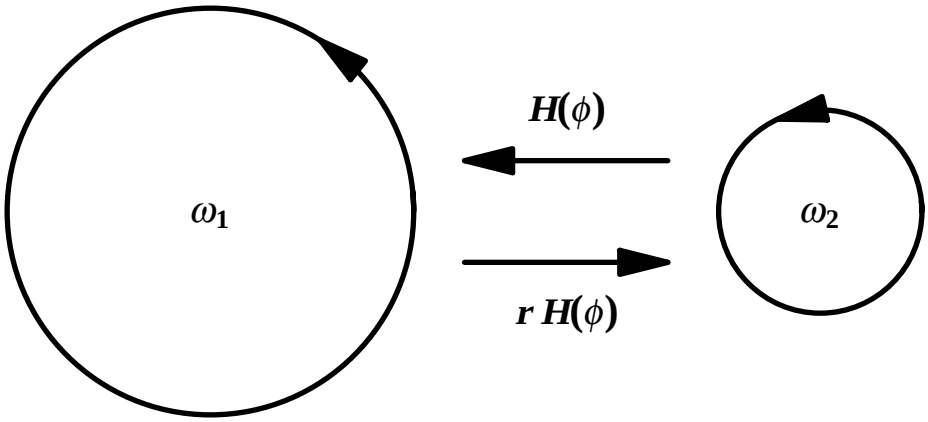

Figure 1. Model with two coupled oscillators.

2. COUPLED OSCILLATOR MODEL

2.1. Interpretation of parameters

One prediction of this model which is in close agreement with a growing body of empirical data (see eg. [2, 7, 8, 9] for a summary) is that the period (total duration) of a higher unit, such as an ISI, should on average be a linear function of the number of smaller units, such as syllables, contained in it. Eriksson [2] expresses this relation with the equation $I = a + bn$, where I is total duration of stress group, n is the number of syllables in the stress group, and a and b are empirically determined coefficients. The coupled oscillator model also predicts the constant term (ie. a) in this function should be greater than zero. Eriksson's coefficients a and b can be related to the parameters of the coupled oscillator model by rewriting the linear function in the form: $I = (br) + (b)n$, where r is the ratio of the two coupling influences as seen in Figure 1. Thus if r is greater than 1, the higher rhythm dominates in the sense that its influence over the lower rhythm is greater than the influence exerted on it by the lower rhythm. If r is less than one the opposite is true. The constant b is equal to $(r\,\omega_1 + \omega_2)^{-1}$ and thus depends on the eigenfrequencies of the two rhythms involved, which presumably also depend in turn on such factors as speaking rate. With the function in this form it is easily seen that the relative strength parameter r of the model can be estimated from the regression results by dividing the constant term by the slope coefficient. It so happens that this also provides a neat geometrical interpretation in the case of the two rhythm model: in a plot of total duration (ISI) as a function of n, the straight line representing the linear function will cross the n-axis at the point $n = -r$.

2.2. Other similar models

2.2.1. Saltzman & Byrd. Saltzman & Byrd [10, 11] have developed a model of coordination or relative timing of overlapping speech gestures which, although not directly concerned with hierachical rhythms, is nevertheless of relevance to our model because they also use coupled oscillators as a starting point. Their work is very much in the spirit of APD theory, since they utilize a coupling which is a

function of phase differences, and therefore do not distinguish between different absolute phases in the component oscillators. They point out that a system of oscillators coupled in this way need not be attracted to a single value of phase difference, but that it is perfectly possible to imagine such a system with many attractors or with a whole "phase window" (or range of phase differences) in which it tends to settle. This is important because, as they point out, empirical evidence has shown that phase differences between gestures may indeed vary widely, though they may be restricted to a definite phase window. Such considerations will be important for our model if and when it becomes possible to characterize the phases of e.g. syllable and stress group rhythms continuously throughout the cycle rather than dealing only with durations, i.e. the periods of the rhythms involved. As it stands, our conclusions are general enough that varying phase relations should not affect the (average) period of component oscillators.

2.2.2. Barbosa & Madureira. Another interesting study which is relevant to our model is the work by Barbosa & Madureira [1] on Brazilian Portuguese. Based on the observation that various lengthening phenomena at the boundaries of some units in Brazilian Portuguese (including stress group) seem to be manifested in a continuous manner, "spilling over" into adjacent units, they propose a hierarchical model of rhythm similar to ours, based on oscillators with continuous coupling. One important difference is that in their proposed model the coupling influences are restricted to operate in one direction instead of allowing the possibility of mutually coupled oscillators as in our model.

3. ADDITIONAL EXAMPLES

We next present some additional examples of the general phenomena of hierarchical rhythms in speech.

3.1. Additional Finnish speakers

As a part of ongoing research aimed at exploring the extent of rhythmic variation within a language as well as between languages we have partially analyzed recorded interviews with two additional speakers of Finnish. Both speakers were taken from an independent project at the University of Tampere to sample the speech of Finnish speakers in Tampere at approximately 20 year intervals. Eventually we hope to extent our study of speech rhythms to include many more speakers as well as speech from different time periods.

3.1.1. Speaker 3M3B (1997). Figure 2 shows the results of measurements of approximately 2 minutes from an interview recorded in 1997 with Speaker 3M3B, male, age 39, a native of Tampere. This data is very typical in that a) the relationship between the number of syllables in a stress group and the total duration of the stress group is approximately linear, and b) the regression line does not go through the origin (as it would on average if syllable durations were independent of each other), but rather crosses the ISI-axis at a positive value, ie. the regression

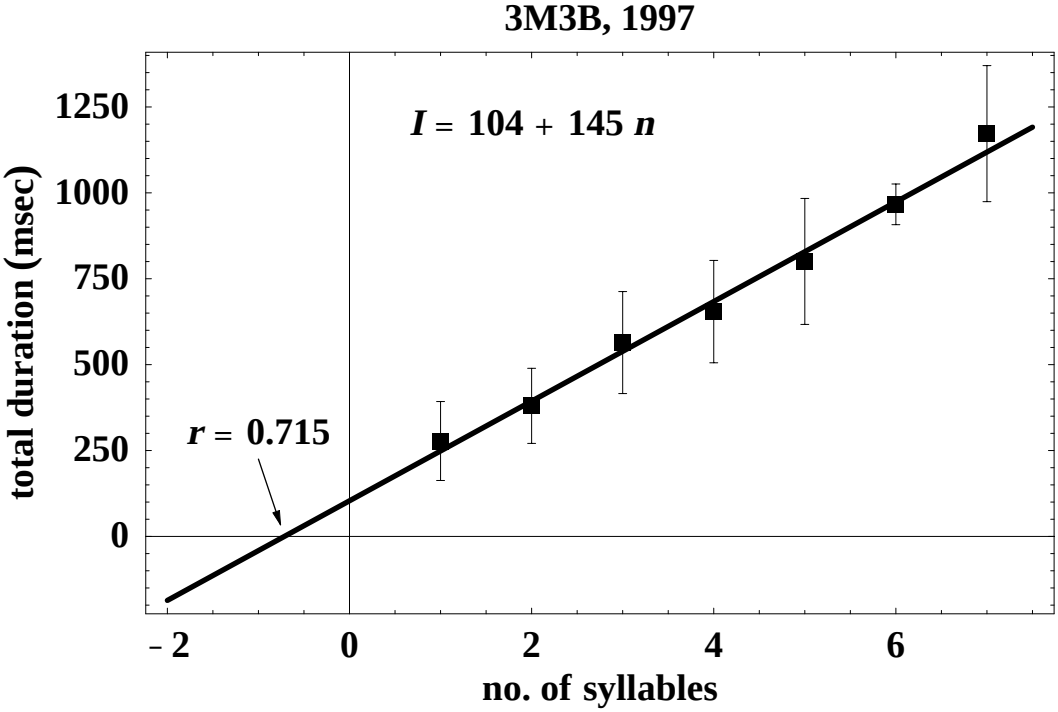

Figure 2. Regression line with means and standard deviations for speaker 3M3B.

equation includes a positive constant term (104 msec in this case). This is exactly what the oscillator model predicts. As pointed out above, the relative strength parameter of the model can be estimated from the regression line and in the present case gives a value of $r = 0.715$, as shown in Figure 2. This is very close to our value estimated earlier for Nieminen's Finnish data ($r = 0.81$ [9]).

It is not surprising that the regression shown in Figure 2 yields a high coefficient of determination, $R^2 = 0.729$, and the total regression model is highly significant, $F(1,65) = 174.608$, $p < 0.00001$. It is of considerable interest to evaluate the significance of the constant term as well, because if syllable durations are independent of the number of syllables in the stress group, we would expect the constant term to be zero on average. For this purpose we applied a one-tailed *t*-test using the standard error of the constant term estimated in the regression analysis, giving $t_{65} = 2.603$, $p = 0.0057$. It would appear very unlikely that such a high positive value would be obtained by chance. It might also be suspected that the positive constant which is characteristic of such data could somehow be due to the particular distribution of syllable durations. To test this possibility, we ran a Monte Carlo experiment on the present data, retaining the structure of stress groups, but rearranging the measured syllable durations in a random way, then recalculating the regression. One thousand trials were run in this way to see how uncommon it would be to obtain a constant term as high as for the real data. In two cases out of 1000, the constant term was as high as in the real data, providing a rough estimated significance of $p = 0.002$.

Another possible explanation for the positive constant term presents itself if we take into consideration the possibility that duration itself may be a cue to stressed

syllable, at least for the researcher measuring the stress groups. If a longer syllable duration leads more readily to classification as stressed, we would expect the constant term to be positive, since it would then represent the additional duration associated with stressed syllables. In the present data, the average duration for stressed syllables was 183 msec and 171 msec for unstressed syllables. While this points to a possible correlation of stress and duration, the difference, 13 msec, is much too small to explain the 104 msec constant term. It would appear that there is a real dependence of syllable rate on stress group size.

3.1.2. Speaker 3N3C (1997). Preliminary results for a second speaker gave very similar results. We present briefly our results for Speaker 3N3C, female, age 39 at time of interview (1997), also a native of Tampere. For this speaker we obtained a regression equation of $I = 97 + 140n$, giving an estimate for the relative strength of $r = 0.693$. This is remarkably close to the values for our other Finnish speakers. The coefficient of determination for this regression was higher than the previous speaker, $R^2 = 0.940$, but this may be in part due to the smaller number of cases evaluated. The total regression model was again highly significant, $F(1,39) = 610.413$, $p < 0.00001$. The positive constant term was also very significant as tested by one-tailed t-test, $t_{39} = 3.539$, $p = 0.0005$.

3.2. New Zealand English

Warren [12] measured durations of stress feet from newscasts representing British (BBC) English as well as various varieties of New Zealand English, including so called Maori English. It has been claimed that the rhythm of New Zealand English has been influenced by language contact with Maori, considered to have mora-timing. Using his Figure 2 to estimate the regression lines, we arrived at approximately $I = 0.548 + 0.810\ n$ for Maori English and $I = 1.162 + 0.592n$ for BBC English. Because the foot durations in this figure have been normalized (by dividing by the average syllable duration for each set of data), these coefficients are not directly comparable with other results, but scaling of the I-axis does not affect the n-axis intercept so the relative strength parameter can still be estimated as before by dividing the constant term by the slope. This gives approximate values of $r = 0.67$ for Maori English as opposed to $r = 1.96$ for BBC English, with the other varieties of New Zealand English intermediate between these two. The value for British English agrees remarkably well with the estimate $r = 1.97$ obtained with Eriksson's regression on Dauer's data for English [2]. The value for Maori English is similar to the estimate $r = 0.64$ obtained with Eriksson's regression on Dauer's data for Spanish [2], or indeed for values obtained by us for Finnish (see section 3.1., [8, 9]).

3.3. Examples from studies utilizing other rhythmic units

Of course there is nothing in our coupled oscillator model which would restrict it to syllable and stress group rhythms. We would therefore expect the same predictions (linear relationship of period to number of units, positive constant

term) to hold for cases in which other types of hierarchical rhythms are compared. In fact there are some such cases reported in the literature and they do appear to support this conclusion.

3.3.1. Swedish stress group vs. phoneme. In a study of rhythmical structures in text reading Fant *et al*. [3] presented data for duration of stress interval (i.e. ISI) in Swedish as a function of the number of phonemes (cf. their Figure 4). The regression they performed yielded the regression equation $T_n = 158 + 53n$ (in msec) and a high coefficient of determination, $R^2 = 0.88$, indicating a very linear relationship with a positive constant. A relative strength parameter of stress group over *phoneme* rhythm can be estimated on the basis of this data as $r = 2.98$.

3.3.2 Phrase and mora for Eskimo and Yoruba. Nagano-Madsen [6] presents in her Figure 6.6 on page 132 average mora duration as a function of the number of morae in the phrase (or utterance) for Eskimo (West Greenlandic) and Yoruba. In general the expected relation applies, that is, mora duration decreases as the number of mora in the phrase increases. By estimating the average durations from this figure, we were able to calculate regressions for Eskimo and Yoruba and derive estimates of the relative strength parameters. Our results are presented in Table 1. The coefficients of determination are quite high, but it must be kept in mind that these regressions are based on average values, not individual cases.

	relative strength parameter	**coefficient of determination**
Eskimo	$r = 2.27$	$R^2 = 0.90$
Yoruba	$r = 1.88$	$R^2 = 0.97$

Table 1. Results for Eskimo and Yoruba.

4. MULTI-LEVEL HIERARCHICAL RHYTHMS

An early work which included measured durations of Finnish words is Laurosela [5]. In an earlier article [9] we calculated a value of $r = 1.84$ for the relative strength parameter based on the Laurosela data. This is obviously very different from the other estimates of r obtained from Finnish data, especially since it is greater than one, implying dominance of the higher rhythm. As we conjectured in the earlier article, this may be in part due to the different circumstances of Laurosela's recordings: word lists rather than running speech. If we consider a breath group, or interval between pauses, to represent yet a higher level oscillator with its own eigenfrequency, then in addition to containing a varying number of syllables, each stress group in the Laurosela data will also be alone within its own breath group.

As pointed out in [8] and [9], the oscillator model can easily be extended to any number of hierachical rhythm levels, with a series of relative stength parameters, one for each pair of adjacent levels. For the expanded model the period of the rhythm at a certain level of the hierchy, say level k, turns out to be a linear

function of all the numbers of units at other levels which are synchronized with a single level k unit: $T_k = \ldots + c_{k-2}/N_{k-2,k} + c_{k-1}/N_{k-1,k} + c_k + c_{k+1}N_{k,k+1} + c_{k+2}N_{k,k+2} + \ldots$, where $N_{k-i,k}$ indicates the number of level k units contained within a single period of the level $(k - i)$ oscillator, and $N_{k,k+i}$ indicates the number of level $(k + i)$ units contained within a single period of the level k oscillator. For instance, taking into account breath group, stress group and syllable, duration of breath group would be a linear function of both the number of stress groups included as well the total number of syllables, with a positive constant term. In terms of the eigen-frequencies of the three oscillators (ω_1, ω_2, ω_3) and the two relative coupling strength parameters ($r_{1,2}$, $r_{2,3}$, see Figure 2), the equation can be rewritten $T_1 = c_1 + c_2N_{1,2} + c_3N_{1,3} = (qr_{1,2}\, r_{2,3}) + (qr_{2,3})N_{1,2} + (q)N_{1,3}$, where $q = (r_{1,2}\, r_{2,3}\, \omega_1 + r_{2,3}\, \omega_2 + \omega_3)^{-1}$. With the equation in this form it can be seen that the relative strength parameters $r_{1,2}$ and $r_{2,3}$ can be estimated from regression data as the ratios of successive coefficients, $r_{1,2} = c_1/c_2$ and $r_{2,3} = c_2/c_3$. An analogous relation holds true for larger models as well. If every breath group contains exactly one stress group, as in the Laurosela data, $N_{1,2} = 1$, the equation will reduce to $T_1 = q(r_1r_2 + r_2) + q(N_{1,3})$. This line will cross the n-axis at $N_{1,3} = N_{2,3} = -(r_{1,2}\, r_{2,3} + r_{2,3})$ instead of just $-r_{2,3}$. If this is interpreted as the relative strength parameter between stress rhythm and syllable only, the value obtained will on average be too large, which may in part explain the large value obtained for the Laurosela data.

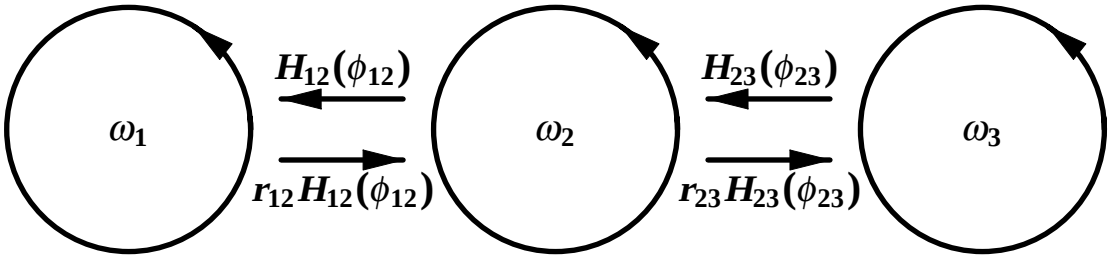

Figure 3. Coupled oscillator model with three component oscillators, eg. breath group, stress group and syllable.

Of course the effect of number of stress groups in a breath group ($N_{1,2}$) cannot be tested for the Laurosela data precisely because there is always exactly one. We returned to our Finnish data ([7] and section 3.1. above) to see what effect, if any, could be observed. To enhance comparability with other results based on the duration of stress group (ISI), we solve for stress rhythm period in our model, given that it is coupled both to a (faster) syllable rhythm as well as a (slower) breath group rhythm. The result is (using the same symbols as above) $T_2 = c_1/N_{1,2} + c_2 + c_3N_{2,3}$. Running a regression on Nieminen's data [7] gave the following equation (rounded to the nearest millisecond): ISI (i.e., T_2) = $117/N_{1,2} + 92 + 142N_{2,3}$. All coefficients were significantly greater than zero ($p < 0.0001$), indicating that number of stress groups within the same breath group did indeed have an effect on stress group duration (ISI). Relative strength parameters estimated from this regression give $r_{1,2} = 1.27$ and $r_{2,3} = 0.65$. The dominance of syllable over stress rhythm comes out about the same as before, and it would appear that breath group rhythm is slightly dominant over stress group as well

($r_{1,2} > 1$). For the purposes of comparison with the Laurosela data, we also calculate $r_{1,2}r_{2,3} + r_{2,3} = 1.47$, the expected n-axis intercept for a regression including only single stress breath groups. This is closer to the 1.84 obtained for the Laurosela data, although that figure is still a bit higher. Whether the remaining difference is due to differences in speaker, dialect, or the existence of still larger effects (utterance length?) is impossible to assess.

A regression analysis was performed on our data for speaker 3M3B as well, yielding very similar coefficient values: $T_2 = 100/N_{1,2} + 93 + 141N_{2,3}$. Relative strength parameters estimated from this regression give $r_{1,2} = 1.08$ and $r_{2,3} = 0.66$; $r_{1,2}r_{2,3} + r_{2,3} = 1.37$. In this case, however, only the c_2 and c_3 coefficients were significantly greater than zero, so the value for c_1 and the estimates based on it are highly unreliable. This is certainly due in part to the fact that the database is much smaller. Whether the effect of $N_{1,2}$ will prove significant in a larger sample is a question currently being pursued. In any case, the values obtained are highly similar, which in itself is suggestive.

5. CONCLUSIONS

As more data becomes available, the tendency predicted by the coupled oscillator model for durations to exhibit a non-trivial linear relationship to the number of component units is gaining confirmation.

The parameters of our model estimated on the basis of several Finnish speakers and different styles of speaking were fairly stable, suggesting that the model may provide a useful tool for describing various rhythmic aspects of specific languages as well as explaining general tendencies across languages. A certain amount of variation between speakers and between speaking styles is of course to be expected, and more work is needed to explore the extent of such variation.

A second finding was that fairly long range effects such as number of stresses per breath group can have a significant effect on timing. This has yet to be investigated for other languages, so it is not clear the extent to which such influence should be considered language specific. The similar results based on data from three different Finnish speakers, however, is suggestive of a robust effect for at least this language. At the same time, the possible existence of higher level effects should serve as a warning that the larger environment may need to be taken into careful consideration when making cross-language comparisons, or indeed any comparisons involving different sets of data.

List of references

[1] Barbosa, P. & Madureira, S. (1999) Toward a hierarchical model of rhythm production: Evidence from phrase stress domains in Brazilian Portuguese. In J. Ohala, Y. Hasegawa, M. Ohala, D. Granville & A. Bailey (eds.), Proceedings of the 14th International Congress of Phonetic Sciences, San Francisco, 1–7 August 1999. Linguistics Department, University of California, Berkeley. 297–300.

[2] Eriksson, A. (1991) Aspects of Swedish Speech Rhythm. University of Göteborg, Göteborg.

[3] Fant, G., Kruckenberg, A. & Nord, L. (1990) Acoustic correlates of rhythmical structures in text reading. In K. Wiik & I. Raimo (eds.), Nordic Prosody V: Papers from a Symposium, Turku. 70–86.

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

[5] Laurosela, J. (1922) Foneettinen tutkimus Etelä-Pohjanmaan murteesta. Suomalaisen Kirjallisuuden Seura, Helsinki.

[6] Nagano-Madsen, Y. (1992) Mora and Prosodic Coordination: A Phonetic Study of Japanese, Eskimo and Yoruba. Lund University Press.

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

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

[9] O'Dell, M. L. & Nieminen, T. (1999) Coupled oscillator model of speech rhythm. In J. Ohala, Y. Hasegawa, M. Ohala, D. Granville & A. Bailey (eds.), Proceedings of the 14th International Congress of Phonetic Sciences, San Francisco, 1–7 August 1999. Linguistics Department, University of California, Berkeley. 1075–1078.

[10] Saltzman, E. & Byrd, D. (1999) Dynamical simulations of a phase window model of relative timing. In J. Ohala, Y. Hasegawa, M. Ohala, D. Granville & A. Bailey (eds.), Proceedings of the 14th International Congress of Phonetic Sciences, San Francisco, 1–7 August 1999. Linguistics Department, University of California, Berkeley. 2275–2278.

[11] Saltzman, E. & Byrd, D. (in press) Task-dynamics of gestural timing: Phase windows and multifrequency rhythms. Human Movement Science.

[12] Warren, P. (1999) Timing properties of New Zealand English rhythm. In J. Ohala, Y. Hasegawa, M. Ohala, D. Granville & A. Bailey (eds.), Proceedings of the 14th International Congress of Phonetic Sciences, San Francisco, 1–7 August 1999. Linguistics Department, University of California, Berkeley. 1843–1846.