

COUPLED OSCILLATOR MODEL FOR SPEECH TIMING: OVERVIEW AND EXAMPLES

Michael L. O'Dell

Tommi Nieminen

1 Background

The present article is intended to sum up the past ten years of our research concerning the timing of speech and the development of a coupled oscillator model (COM) for speech timing. This enterprise started with our paper (O'Dell and Nieminen 1998) presented at the 1998 Finnic Phonetics symposium in Pärnu—the first of the series to be held in Estonia—and has since continued through several conferences and conference papers.

Our initial focus (O'Dell and Nieminen 1998, 1999) was on the so-called rhythm dichotomy (Pike 1945) and particularly Eriksson's (1991) re-interpretation of it. Eriksson had noted that in both stress-timed and syllable-timed languages, a simple linear relation holds between the duration of the stress group and the number of the syllables in it, the only difference being in the constant term of the linear regression equation (y -intercept). This intriguing observation provides simple mathematical tools with which to describe the classic rhythm dichotomy and we proposed the COM as a possible explanation of the observed mathematical relation.

Later, we have expanded the scope of the model and applied it to additional data and other types of timing phenomenon (O'Dell and Nieminen 2001, 2002a,b, 2006). Most recently we have been developing statistical tools to account for rhythmic variation in spontaneous speech (O'Dell et al. 2007, 2008).

In the following we first present, in Section 2, a brief introduction to the terminology and the basic concepts of the COM. We also discuss its relationship both to dynamical systems theory, this being the general (i.e. not speech-specific) point of view, and to the phonetic study of speech in particular. In later sections, applications of the model are presented mainly from our own studies but partly taken from studies conducted by other researchers in the field: Section 3 describes the interaction of speech units at different levels, Section 4 deals with intralevel effects, and Section 5 provides an example of applying the model to a specific case.

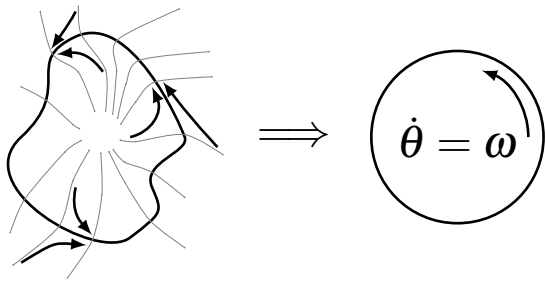


Figure 1: Simplification of a limit cycle

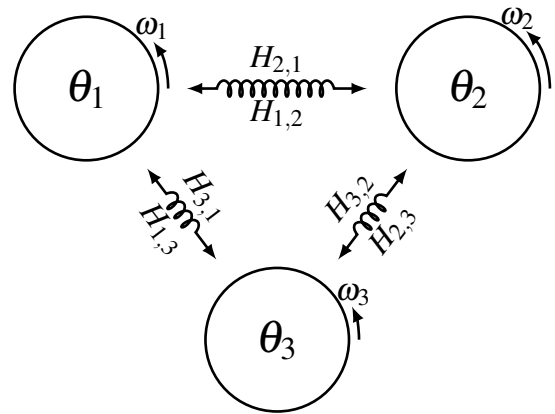


Figure 2: Example of coupled oscillators

2 Overview

Any process that tends to repeat itself regularly can generally be described as an **oscillator**. Speech examples of oscillators include such basic phenomena as syllables or vocal fold vibration during voicing. In dynamical systems theory an oscillator corresponds to a limit cycle attractor (a closed loop in state space which neighboring trajectories approach asymptotically, cf. Figure 1).

A limit cycle in a system gives rise to a natural frequency (ω , the inverse of the period T of the process on the limit cycle itself), as well as a natural phase (θ) for the process: any point on the limit cycle is associated with a set of points in state space (called an *isochron*, cf. the thin gray lines in the lefthand diagram of Figure 1) which converge as the process continues and can thus be considered to represent the same phase. It is often convenient to simplify description of the oscillator by retaining only the phase variable and reparameterizing if necessary so that the limit cycle is divided into phases equally throughout the period. The result is a single phase variable increasing at a constant rate ($\dot{\theta} = \omega$) as illustrated in the righthand diagram of Figure 1.

When the time course of a process is influenced by some other process, that influence is called **coupling**. Speech examples of this would include articulatory gestures influencing each other or dialogic turn taking. In dynamical systems theory the influence can be indicated by adding *coupling functions* $H_{ij}(\theta_i, \theta_j)$ to the rate equation (cf. also Figure 2):

$$\dot{\theta}_i = \omega_i + \dots + H_{ij}(\theta_i, \theta_j) + H_{ik}(\theta_i, \theta_k) + \dots \quad (1)$$

Average Phase Difference (APD) theory (Kopell 1988) greatly simplifies analysis by approximating the general coupling function, which is a function of the states (phases) of both oscillators, with a function of phase *difference* $\phi_{ij} = \theta_j - \theta_i$ (averaged over one cycle):

$$H_{ij}(\theta_i, \theta_j) \approx H_{ij}(\theta_j - \theta_i) = H_{ij}(\phi_{ij}) \quad (2)$$

In the case of mutually synchronizing influence between a pair of oscillators we assume that the

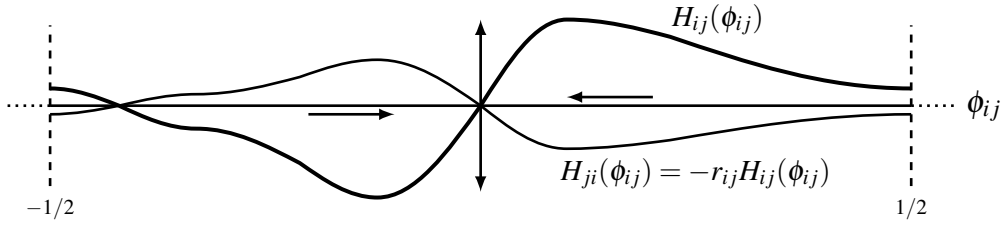


Figure 3: Example of coupling function pair

coupling functions in opposite directions are (approximately) identical in form, but opposite in sign and possibly different in magnitude. The difference in magnitude can then be characterized by a parameter r_{ij} we call relative coupling strength (cf. also Figure 3):

$$r_{ij} = -H_{ji}/H_{ij} \quad (3)$$

It is often useful to normalize the coupling functions over the whole system by associating a mass-like parameter m_i with each oscillator. An oscillator with greater ‘mass’ has more effect on other oscillators while being less affected itself. The coupling functions can then be expressed as

$$H_{ij} = \tilde{H}_{ij}/m_i, \quad H_{ji} = -\tilde{H}_{ij}/m_j \quad (4)$$

so that $\pm\tilde{H}_{ij}$ indicates the mutual coupling influence (a kind of ‘force’) between oscillators i and j , while the pair m_i and m_j determines relative coupling strength (with one additional constraint, eg. $\sum_i m_i = 1$).

In cases where a coupling function aims to synchronize a *varying number* n_{ij} of cycles of (subordinate) oscillator j with each cycle of (superordinate) oscillator i we may speak of **hierarchical coupling**. A speech example of this would be, for instance, the variable number of syllables within recurring (phrasal) stress groups. In dynamical systems theory this can be described using a *generalized* phase difference:

$$\phi_{ij}(n_{ij}) = \theta_j - n_{ij}\theta_i \quad (5)$$

Modeling durations. The COM uses the concept of *phase* to describe timing. In order to apply the model to data it is necessary to derive predictions in terms of *durations*, since empirically speaking we generally measure durations or intervals between boundaries observed in the speech signal. To this end, the following **fundamental oscillator equation for durations** derived using APD is extremely useful for relating the COM to measured durations:

$$T_i = \sum_{j=1}^K c_j n_{ij}, \quad c_j > 0 \quad (6)$$

Equation (6) says that T_i , the resultant (equilibrium) period (duration) of oscillator i in a system with K oscillators, is a linear function of the numbers n_{ij} of units synchronized with it. It must be remembered, however, that the mapping from phase configuration of the system to observed acoustic boundary may be complex. In particular, we cannot assume that a given boundary type always corresponds to the same phase of some oscillator. For example, so called intrinsic durational differences of segments could be the result of natural frequency differences, or they could be due to differences in coarticulation, so that the “same” boundary represents different phases of the oscillators.

Useful relations between the coefficients c_i in Equation (6) and parameters of the COM can be derived from

$$c_i = r_{ik} / \left(\sum_{j=1}^K r_{jk} \omega_j \right) = m_i / \left(\sum_{j=1}^K m_j \omega_j \right) \quad (7)$$

which shows that the coefficients are proportional to both the relative strength parameters r_{ik} and the oscillator mass parameters m_i . Since $\sum_{j=1}^K c_j \omega_j = 1$, and the natural (uncoupled) duration for n_{ij} cycles of oscillator j is just n_{ij} / ω_j , the period T_i can also be expressed as a weighted average of these uncoupled durations:

$$T_i = \sum_{j=1}^K \pi_j \left(\frac{n_{ij}}{\omega_j} \right), \quad \pi_j = c_j \omega_j \quad (8)$$

These relations can be summarized as $r_{ij} = c_i / c_j = m_i / m_j = (\pi_i / \pi_j) (\omega_j / \omega_i)$.

All in all, the COM offers four types of explanation, corresponding to four types of variable, for any observed difference in measured durations. The observed durations may stem from systems with (i) a different number of units (or cycles) n_{ij} at some level, (ii) different natural frequencies ω_i for some oscillators, (iii) different coupling strengths r_{ij} between some oscillators, or (iv) different phase relations (ie. varying coarticulation) ϕ_{ij} between oscillators.

3 *Interlevel coordination*

Rhythmic gradation. Perhaps one of the best known phenomena in speech rhythm (though the details have often been controversial), is the general tendency that as more subunits are included in some superordinate rhythm unit, the duration of that unit increases while the average duration of the subunits decreases. Paradoxically, although the phenomenon has been known for well over a century, terminology referring to it is anything but standard, and many researchers have avoided the use of a general term. We use the term *rhythmic gradation*, a translation of Sievers’ (1893) term *rhythmische Abstufung*.

The COM shows that rhythmic gradation is a consequence of synchronizing rhythms hierarchically. We therefore expect the phenomenon to be very widespread, probably extending to other types of rhythmic behavior as well as speech.

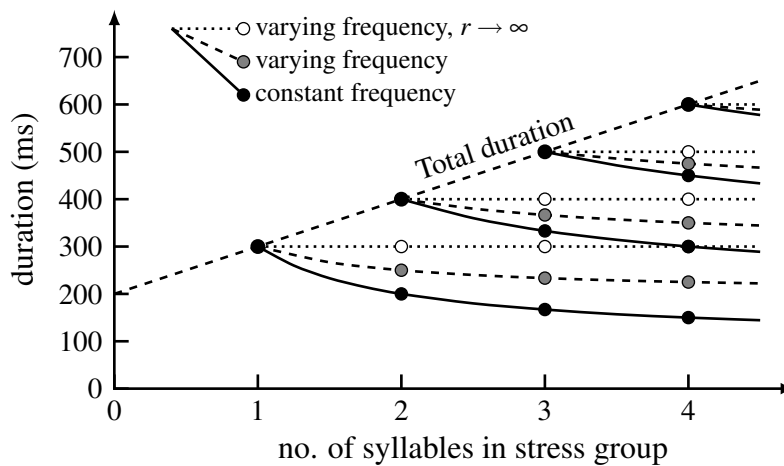


Figure 4: Examples of rhythmic gradation

Two types of equation have been used to model the rhythmic gradation effect for measured durations in speech: a linear function for total duration ($a + bn$, eg. Eriksson 1991; also implied by forms such as $d + D/n$ for component durations, eg. Lindblom 1968; Klatt 1976), or a negative exponential function for component durations ($a + be^{-cn}$, eg. Fónagy and Magdics 1960; this would seem to imply a relation of the form $an + bne^{-cn}$ for total duration). The difference between the two will often be difficult to assess empirically; however, the COM predicts a linear relationship, as can be seen in Equation (6).

Within a cycle duration may be “allocated” unevenly if frequencies or coupling strengths vary. Figure 4 illustrates various possible relations between durations and number of syllables in a stress group, given a model with two oscillators, one for syllable and one for stress group. The dashed line labeled “Total duration” illustrates the linear relationship predicted by the model (in this example $100n + 200$ ms, similar to the regression lines given by Eriksson 1991 for English and Thai).

If all parameters are assumed constant throughout the stress group, then syllable frequency will not vary and all syllables will be of equal duration (shown in Figure 4 by black filled circles connected with solid lines). By letting syllable frequency vary during the stress group (with first syllable slower) while simultaneously letting the relative coupling strength of the syllable increase without limit, an extreme case can be achieved with all of the ‘extra’ duration of the stress group concentrated in the first syllable (shown in Figure 4 by open circles connected with dotted lines).

Outside these cases of extreme centralization vs. complete dispersion of duration we find a whole range of possibilities (one example is shown in Figure 4 by filled gray circles connected with dashed lines). Saltzman et al. (2008) for example varied both frequency and coupling in order to fit the model to empirical data on rhythmic gradation in English.

Number of cycles n_{ij} as a dynamic variable. Barbosa (2001) used coupled oscillators to determine natural sounding stress positions (ie. n_{ij}) for speech synthesis. Based on this idea, O’Dell and Nieminen (2002a,b) modeled the probability in continuous speech of a stressed syllable (ie. new stress group boundary) as a function of the number of syllables since the previous stress. Of course in order to do this, at least some *a priori* differences in syllable “stressability” need to be taken into consideration (in this case syllables were lexically classified as stressable vs. unstressable). The conclusion reached was that synchronization of syllable and stress group rhythms does indeed affect the probability of a syllable being stressed. This result highlights the fact that hierarchical synchronization does not necessarily imply planning at a certain level, as has sometimes been supposed (Kim and Cole 2005). The duration data alone cannot tell us whether or to what extent n_{ij} , the number of j cycles corresponding to one i cycle, is planned independently, forcing durations patterns to conform. A statistical relation pointing to rhythmic gradation at some level is also compatible with system dynamics influencing choice of n_{ij} dynamically ‘on the fly.’

4 *Intralevel coordination*

Saltzman *et al.* (cf. 2008) have used coupled oscillators to derive timing for the gestures in a Task Dynamic–Articulatory Phonology setting.

Two things will happen as oscillator driven gestures are synchronized in a 1 : 1 manner. First, the component oscillators will find a common frequency (weighted average) if the coupling force is strong enough. This means, for instance, that if two consonant gestures have different natural frequencies, then we expect a difference in the durations of syllables containing those consonants.

Another effect concerns the *phase differences* of the component oscillators. Although oscillators have a preferential or natural frequency, they have no preferential phase (by definition), so that phase shifts are easily brought about by external forces such as coupling. In cases with only simple chain coupling of gestures (that is, unlike in Figure 2 above, the coupling graph contains no loops), any ‘target’ phase differences imposed by the coupling will always be achieved. However, in cases with loops in the coupling graph (called competitive coupling by Nam and Saltzman 2003), the set of target phase differences imposed by the coupling functions may be mutually incompatible. In such a case the phase differences of the system will settle in a compromise of the pairwise (target) phase differences.

The so-called ‘C-Center effect’ is a predicted consequence of competitive coupling (Nam and Saltzman 2003; Goldstein *et al.* 2007). In English and some other languages it appears that the vowel exhibits a fairly constant temporal relation to the center of the cluster of consonants in the syllable onset, that is, as consonants are added to the onset, earlier consonants start (even) earlier relative to the vowel while later consonants start later. This is exactly what is predicted in competitive coupling of gestures.

In the example shown in Figure 5 (based on Nam and Saltzman 2003), consonants are cou-

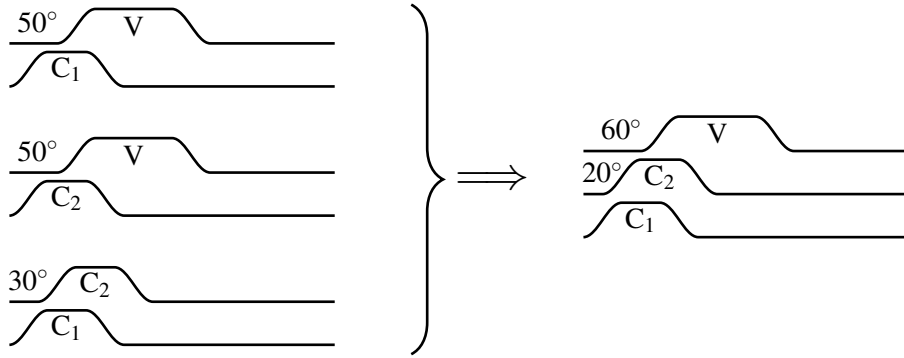


Figure 5: Illustration of the C-Center effect

pled to a following vowel with target phase difference 50° , and with no other couplings this will be the equilibrium phase difference. However, with two consonants in the onset, the consonants are also coupled to each other with a target phase difference of 30° (to avoid simultaneous production). The result predicted by the COM (confirmed in Nam and Saltzman 2003 using simulations) is a compromise set of phase differences forcing C_1 to start earlier (40° relative to V) and C_2 to start later (60° relative to V).

This analysis of gestures as syllable internal oscillators implies that the duration of the syllable will *not* be a linear function of the number of components (as would be the case with hierarchical coupling). Adding a component oscillator may slightly increase or decrease the composite frequency depending on the natural frequency of the added component. But in general it will not increase syllable duration while decreasing segment duration as in rhythmic gradation. This difference in behavior can be used as an empirical test to help sort component rhythms into hierarchical levels.

Interaction of hierarchical and intralevel coordination. To see how hierarchical and intralevel coordination interacts, it is helpful to view a set of oscillators L (considered as a single level) alternatively as a single oscillator unit which is coupled hierarchically to another level. If the parameters of the oscillators in L are known, one can calculate parameters for a single oscillator which is equivalent to L in terms of external influences:

$$m_L = \sum_{i \in L} n_{L,i} m_i \quad \text{or} \quad r_{L,j} = \sum_{i \in L} n_{L,i} r_{i,j}, \quad \text{and} \quad \omega_L = \frac{1}{m_L} \sum_{i \in L} m_i \omega_i \quad (9)$$

On the other hand, if parameters have been determined for a composite level oscillator, say the syllable, then Equations (9) give conditions for the internal relationships of component oscillators. Here $n_{L,i}$ refers to the number of oscillator i cycles corresponding to one cycle of the composite oscillator L . The composite frequency ω_L is a *weighted average* of the component frequencies, while the mass parameter m_L of the composite is the *sum* of the component masses

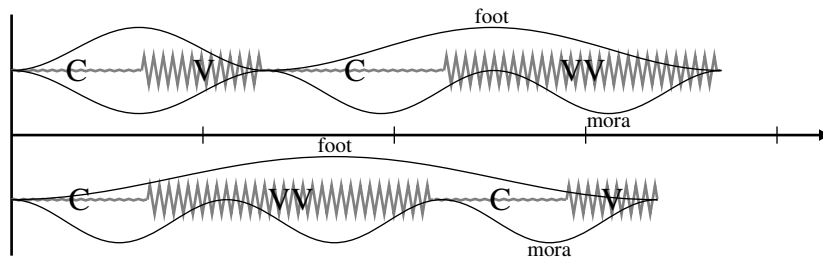


Figure 6: Schematic diagram of Finnish word types CVCVV and CVVCV

(or equivalently, since $r_{ij} = m_i/m_j$, the coupling strength of the composite relative to an external oscillator j , is the sum of component relative coupling strengths). One immediate consequence is obvious: as component oscillators are added to a level (e.g. consonants to a syllable), either the level must become more dominant, or the components become less dominant individually.

The phase θ_L of the composite oscillator can be chosen arbitrarily, although there may be esthetic reasons for associating zero phase with some particular configuration of component phases. Hierarchical coupling affects *phase differences* between oscillators, so it will also affect within level (intergestural) phase differences. This effect will in general be fairly small, although the exact amount depends on the form of the coupling functions. For instance, as syllables are added to a foot, we expect syllable internal coarticulation to increase somewhat.

5 An example: Why is Finnish CVVCV shorter than CVCVV?

As an exercise in applying the COM to detailed data, we consider the robust but little discussed tendency for the structure CVCVV in Finnish to have a longer measured duration than the structure CVVCV. The situation is shown schematically in Figure 6, based on Lehtonen's data (Lehtonen 1970). Other researchers have reported very similar data for these structures (eg. Suomi and Ylitalo 2004).

Number of units. As we have seen, under the COM, one possible explanation for different durations is a difference in number of units (cycles) for an oscillator at some level. Both word structures CVVCV and CVCVV have the same number of segments, the same number of syllables, and indeed the same number of morae. For this explanation to be applicable, CVCVV must contain a greater number of units at some other level, for example, a single 'foot' in (CVVCV) vs. two 'feet' in (CV)(CVV). This hypothesis actually gives a remarkably good fit due to the fact that its two main predictions are close to empirical data:

1. In CVVCV the first syllable (two morae) is approximately twice the duration of the second (one mora), indicating very little foot internal variation in mora frequency.

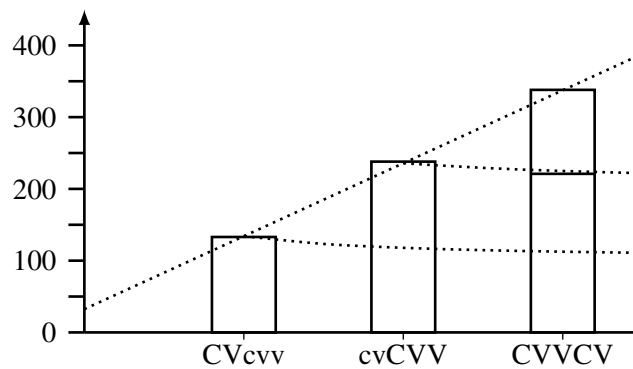


Figure 7: Durations of ‘feet’ with 1, 2 and 3 morae

2. The durations which under this hypothesis correspond to single feet with one mora (first syllable of CVCVV), two morae (second syllable of CVCVV) and three morae (all of CVVCV) are very nearly in a linear relationship with a positive constant (cf. Figure 7; The dotted lines show the fit with coefficients $c_{\text{FOOT}} = 32$ ms, $c_{\text{MORA}} = 102$ ms in Equation (6) above).

Another hypothesis is interesting because it corresponds to the number of syllables reconstructed for these word types in late Proto-Finnic¹: two ‘proto-syllables’ for (CVV)(CV) vs. three ‘proto-syllables’ for (CV)(CV)(V). Perhaps cycles corresponding to syllables of the earlier stage of the language still influence the rhythm of present day Finnish. However, this hypothesis fits the data worse than the previous one, because it makes two poor predictions:

1. The first syllable of CVCVV should be approximately the same duration as the last syllable of CVVCV (both one mora, one ‘proto-syllable’). Instead it is approximately 14% longer.
2. In CVCVV the second syllable (two morae, two ‘proto-syllables’) should be approximately twice the duration of the first (one mora, one ‘proto-syllable’). Instead it is only about 179% as long.

Varying oscillator parameters within the foot. The COM also has other parameters which have an effect on observed durations. If we wish to adhere to the traditional analysis of Finnish which would assign both CVCVV and CVVCV structures to a single foot, then within the

¹Cf. eg. Hakulinen (1961, 20): “... long vowels in any syllable other than the first are secondary. They are the result of the most recent sound changes in the purely Finnish period of development. Generally the change has been brought about by the disappearance of a medial consonant, whereupon two short vowels have come to be side by side and have combined to form one vowel. E.g. ... *kalaa* (part. sg. of *kala* fish) < **kalaδa*; ...” (emphasis in the original; δ indicates a voiced dental fricative, IPA [ð]).

coupled oscillator framework we must assume that some parameters vary during the course of a foot, presumably one or more oscillator frequencies or one or more coupling strengths. Such variation could easily be interpreted as a realization of “word stress” if desired.

It turns out that it is possible to get CVCVV longer than CVVCV by varying parameters within a foot. To get this result requires a syllable oscillator in the model in addition to the mora and foot oscillators postulated above. In terms of the fundamental duration equation (6), CVCVV longer than CVVCV requires that the average value of the (varying) syllable coefficient c_{SYLL} must be greater for the first mora than for the third. This corresponds reasonably well with the conclusion of Suomi and Ylitalo (2004) that in a Finnish foot the “first two mora are lengthened.” Note that if all other coefficients are held constant, then the first syllable (or mora) of CVCVV will also be longer than the last syllable of CVVCV, which actually is the case. The larger value of c_{SYLL} for the first mora could theoretically result from two different causes (or both):

1. Lower frequency for one or more oscillators in the first mora. However, this difference alone cannot give a good fit to the data, because it predicts that if the first syllable of CVCVV is longer than the second syllable of CVVCV (which as noted is indeed the case), then the first syllable of CVVCV should also be longer than the second syllable of CVCVV, whereas the reverse is true.
2. Higher relative coupling strength of the syllable oscillator during the first mora. In this case the data can be fit fairly closely by letting the relative coupling strength of the syllable fall almost to nil by the second mora. The situation can be restated as follows: Syllable timing is relatively strong at the very beginning of the foot but by the end of the first mora the syllable has lost almost all influence and mora timing has taken over. The effect is that the second syllable of CVCVV should be approximately twice as long as the second syllable of CVVCV (in Lehtonen’s data the ratio is 203%), but the influence of the syllable oscillator at the start somewhat reduces the difference between the first syllables of CVCVV and CVVCV (166% in Lehtonen’s data).

In sum, it appears that two hypotheses can give a reasonable fit to the data: different number of feet and varying syllable strength. This is not intended as a claim that one of these hypotheses must in some sense be correct, but rather as an illustration of the application of the COM to data. Of course, in a more thorough treatment, other possible structures must also be taken into account, such as the fact that CC *shortens* preceding long VV or preceding C but *lengthens* previous short V. Obviously there is also variation due to other factors such as speaking rate, different segments, intonation, etc.

6 Conclusion

Over the past ten years the COM has proved to be a very useful tool for modeling and understanding speech rhythms. Exploring the mathematical properties of the model has given insight

into a wide variety of timing phenomena unified in a single framework. The concepts developed here find application in other areas of speech timing as well, such as interspeaker turn taking coordination (cf. eg. Buder and Edrington 2008) and the coupling of production and perception rhythms.

There is still much work to do in developing this tool. One important area which we have only recently begun to investigate is the question of integrating variation and uncertainty into the model. Unfortunately space does not permit us to develop this theme here. The reader is referred to O'Dell et al. (2007, 2008).

References

- Barbosa, P. A. (2001). Generating duration from a cognitively plausible model of rhythm production. In *Proceedings of the Seventh European Conference on Speech Communication and Technology (Eurospeech 2001), Aalborg, Denmark, September 3–7*, volume 2, pages 967–970.
- Barbosa, P. A., Madureira, S., and Reis, C., editors (2008). *Speech Prosody 2008: Fourth Conference on Speech Prosody, Campinas, Brazil*.
- Buder, E. H. and Edrington, J. L. (2008). Conversational prosodic interactivity when one partner has aphasia. In Barbosa et al. (2008), pages 501–504.
- Eriksson, A. (1991). *Aspects of Swedish Speech Rhythm*. University of Göteborg, Göteborg.
- Fónagy, I. and Magdics, K. (1960). Speed of utterance in phrases of different lengths. *Language and Speech*, 3(4):179–192.
- Goldstein, L., Chitoran, I., and Selkirk, E. (2007). Syllable structure as coupled oscillator modes: Evidence from Georgian vs. Tashlhiyt Berber. In Trouvain and Barry (2007), pages 241–244.
- Hakulinen, L. (1961). *The Structure and Development of the Finnish Language*. Indiana University. (Transl. John Atkinson).
- Kim, H. and Cole, J. (2005). The stress foot as a unit of planned timing: Evidence from shortening in the prosodic phrase. In *INTERSPEECH 2005, Lisbon, Portugal*.
- Klatt, D. K. (1976). Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *Journal of the Acoustical Society of America*, 59:1208–1221.
- Kopell, N. (1988). Toward a theory of modelling central pattern generators. In Cohen, A. H., Rossignol, S., and Grillner, S., editors, *Neural Control of Rhythmic Movements in Vertebrates*, pages 369–413. New York: John Wiley and Sons.
- Lehtonen, J. (1970). *Aspects of Quantity in Standard Finnish*. Number VI in *Studia Philologica Jyväskyläensia*. University of Jyväskylä.
- Lindblom, B. E. F. (1968). Temporal organization of syllable production. *Quarterly Progress and Status Report*, 9(2-3):1–5.
- Nam, H. and Saltzman, E. (2003). A competitive, coupled oscillator model of syllable structure. In Solé, M. J., Recasens, D., and Romero, J., editors, *Proceedings of the 15th Interna-*

- tional Congress of Phonetic Sciences*, volume 3, pages 2253–2256. Universitat Autònoma de Barcelona, Spain.
- O'Dell, M., Lennes, M., and Nieminen, T. (2008). Hierarchical levels of rhythm in conversational speech. In Barbosa et al. (2008), pages 355–358.
- O'Dell, M., Lennes, M., Werner, S., and Nieminen, T. (2007). Looking for rhythms in conversational speech. In Trouvain and Barry (2007), pages 1201–1204.
- O'Dell, M. and Nieminen, T. (1998). Reasons for an underlying unity in rhythm dichotomy. *Linguistica Uralica*, 3:178–185.
- O'Dell, M. and Nieminen, T. (1999). Coupled oscillator model of speech rhythm. In Ohala, J., Hasegawa, Y., Ohala, M., Granville, D., and Bailey, A., editors, *Proceedings of the XIVth International Congress of Phonetic Sciences*, volume 2, pages 1075–1078. University of California, Berkeley.
- O'Dell, M. and Nieminen, T. (2001). Speech rhythms as cyclical activity. In Ojala, S. and Tuomainen, J., editors, *21. Fonetikan päivät Turku 4.–5.1.2001*, number 67 in Turun yliopiston suomalaisen ja yleisen kielitieteen laitoksen julkaisuja / Publications of the Department of Finnish and General Linguistics of the University of Turku, pages 159–168. (A corrected version of this article can be found at <http://papers.legisign.org/articles/odell-nieminen2001.pdf>.)
- O'Dell, M. and Nieminen, T. (2002a). How long is a stress group? *Cadernos de Estudos Lingüísticos*, 43:93–108.
- O'Dell, M. and Nieminen, T. (2002b). Rytmijakson pituus oskillaattorimallissa. In Korhonen, P., editor, *Fonetikan Päivät 2002 / The Phonetics Symposium 2002*, pages 195–204. Helsinki University of Technology, Laboratory of Acoustics and Audio Signal Processing.
- O'Dell, M. and Nieminen, T. (2006). Tahdin ajoitus suomessa oskillaattorimallin näkökulmasta. In Aulanko, R., Wahlberg, L., and Vainio, M., editors, *Fonetikan päivät 2006 / The Phonetics Symposium 2006*, number 53 in Helsingin yliopiston puhetieteiden laitoksen julkaisuja, pages 134–143.
- Pike, K. L. (1945). *The Intonation of American English*. University of Michigan Press, Ann Arbor.
- Saltzman, E., Nam, H., Krivokapic, J., and Goldstein, L. (2008). A task-dynamic toolkit for modeling the effects of prosodic structure on articulation. In Barbosa et al. (2008), pages 175–184.
- Sievers, E. (1893). *Grundzüge der Phonetik zur Einführung in das Studium der Lautlehre der indogermanischen Sprachen*. Leipzig: Breitkopf & Härtel, 4th edition.
- Suomi, K. and Ylitalo, R. (2004). On durational correlates of word stress in Finnish. *Journal of Phonetics*, 32(1):35–63.
- Trouvain, J. and Barry, W. J., editors (2007). *Proceedings of the 16th International Congress of Phonetic Sciences*. Universität des Saarlandes, Saarbrücken, Germany.