Implications of rhythmic discreteness in speech

Robert Port

Table of Contents

Implications of Rhythmic Discreteness in Speech **1**
Robert F. Port, Indiana University 1
I. Basic Experimental Evidence for Harmonic Timing Effect 2
II. Implications of the Harmonic Timing Effect 3
III. Predictions 9
IV. Concluding Discussion 10

Implications of Rhythmic Discreteness in Speech

Robert F. Port, Indiana University

In recent studies we have verified what laymen already understand, that human speech is easily and naturally spoken in a rhythmical way. But hearing a rhythmic speaking style and demonstrating rhythm objectively are quite different things. Our empirical verification of rhythmic speech is based on the important realization that vowel onsets (ie, approximately P−centers) are the most important event determining perceived speech rhythm (Allen, 1972; Morton et al 1976). That is, when speaking rhythmically, English speakers (and very likely speakers of other languages as well) adjust overall timing so that vowel onsets occur near certain priviledged temporal locations. This is important since it means that if we measure vowel onset locations, we do not need to pay much attention to other aspects of phonetic events to observe and characterize the rhythm of speech.

Although the term `rhythmic speech' can be used in many vague ways, we can define it here as describing speech that exhibits a tendency to locate prominent acoustic onsets at regular periodic intervals on one or more time scales. This definition is far more flexible and amenable to experimental evaluation than traditional descriptions in terms of, for example, `isochrony' (Abercrombie, 1967; Pike, 1943).

Although we do not yet have developmental data, these effects should surely be of interest to students of language development. It seems that some kind of rhythm is found in children's speech from well before first words. There is the (cyclical) reduplication of syllables in babbling and the observation that children differentiate the prosody of their mother's language from other languages shortly after birth (Mehler et al, 86; Jusczyk, 1997). The cognitive skills to be explored in this paper are ones that children acquire very early in life, so the adults we used as subjects can be assumed to already have significant experience in this regard −− even if they may be largely unaware of their own metrical skills.

Intuitive Performance

The simplest way to be convinced of the robustness of the effect I am talking about is to repeat a short English phrase out loud over and over. For example, if one repeats the phrase ``Buy the boy a cake'' repeatedly (it should be tried for at least 5 repetitions), one will most likely find, first, that the whole repetition cycle − from Buy to Buy − tends to be constant. And also that the word `cake' locates itself half way between the repetitions of the word `buy' (more precisely, it is the onset of the vowels that line themselves up this way). It is almost as though one says ``ONE TWO, ONE TWO, BUY the boy a CAKE, BUY the boy a CAKE, ...'' An alternative reading (especially

for faster tempos) is to a three−beat meter, as in `ONE TWO THREE, ONE TWO THREE, BUY the BOY a CAKE, BUY the BOY a cake," etc.

Actually there is a third way to repeat this phrase that can be found if one tries to leave a pause after the end of the phrase before repeating. Thus, one might say `BUY the boy a CAKE [PAUSE], BUY the boy a CAKE [PAUSE], BUY ..." Again this time it will be discovered that it is a 3−beat pattern although at a slower tempo. But `cake' falls on beat 2 (rather than on beat 3 as above) with a musical rest on the third beat. Production in any one of these patterns is very stable and consistent. If one tries to do some other pattern, it becomes quite difficult and keeps slipping toward one of the stable patterns, like 1/3 or 1/2.

I. Basic Experimental Evidence for Harmonic Timing Effect

Given this simple demonstration, how strong is the bias to locate these vowel onsets only near these 3 locations, 1/3, 1/2 and 2/3 of the way through the repetition cycle? Cummins and Port (1998) tested these preferences by presenting subjects with a two−tone metronome pattern with one tone, A, marking the beginning of the cycle (where the word `buy' would occur using this demonstration phrase) alternating with the other tone, B, at randomly distributed locations between 20% of the cycle and 80% of the cycle. The subjects' task was to repeat the phrase so that `buy' lines up with tone A and `cake' lines up at tone B (that is, so its vowel onset would be simultaneous with tone B onset). Of course, if we kept the repetition cycle, A−A, a constant duration, the subjects would have been forced to change their speaking rate up to a factor of 4 between the earliest (requiring the fastest tempo) and the latest (requiring the slowest tempo). So instead we made our A−A metronome interval vary in such a way that the interval from A to B was fixed and only the repetition cycle varied. This gave the speakers a constant amount of time for pronunciation of the text. We measured the location of onset of the final syllable and report it as a particular phase angle, in the range (0, 1), of the repetition cycle. Only a couple practice trials were employed with each subject.

The results are shown separately for 8 speakers in Figure 1. About half of the subjects had music training but the other half did not. Although the target phase angles for the onset of the final stressed syllable were distributed uniformly over the interval from 0.20 to 0.80 of the repetition cycle, the speakers actually located their onsets near only 3 locations in the cycle, 1/3 for all the early phase angle targets, 1/2 for targets near the middle of the cycle and 2/3 for all target phases later than about 0.55. Notice, however, that 2 of the 8 speakers could not seem to find the pattern that locates the final syllable at 2/3 of the cycle. (Neither of these two was a musician. Aside from this, the musicians and nonmusicians performed about the same.)

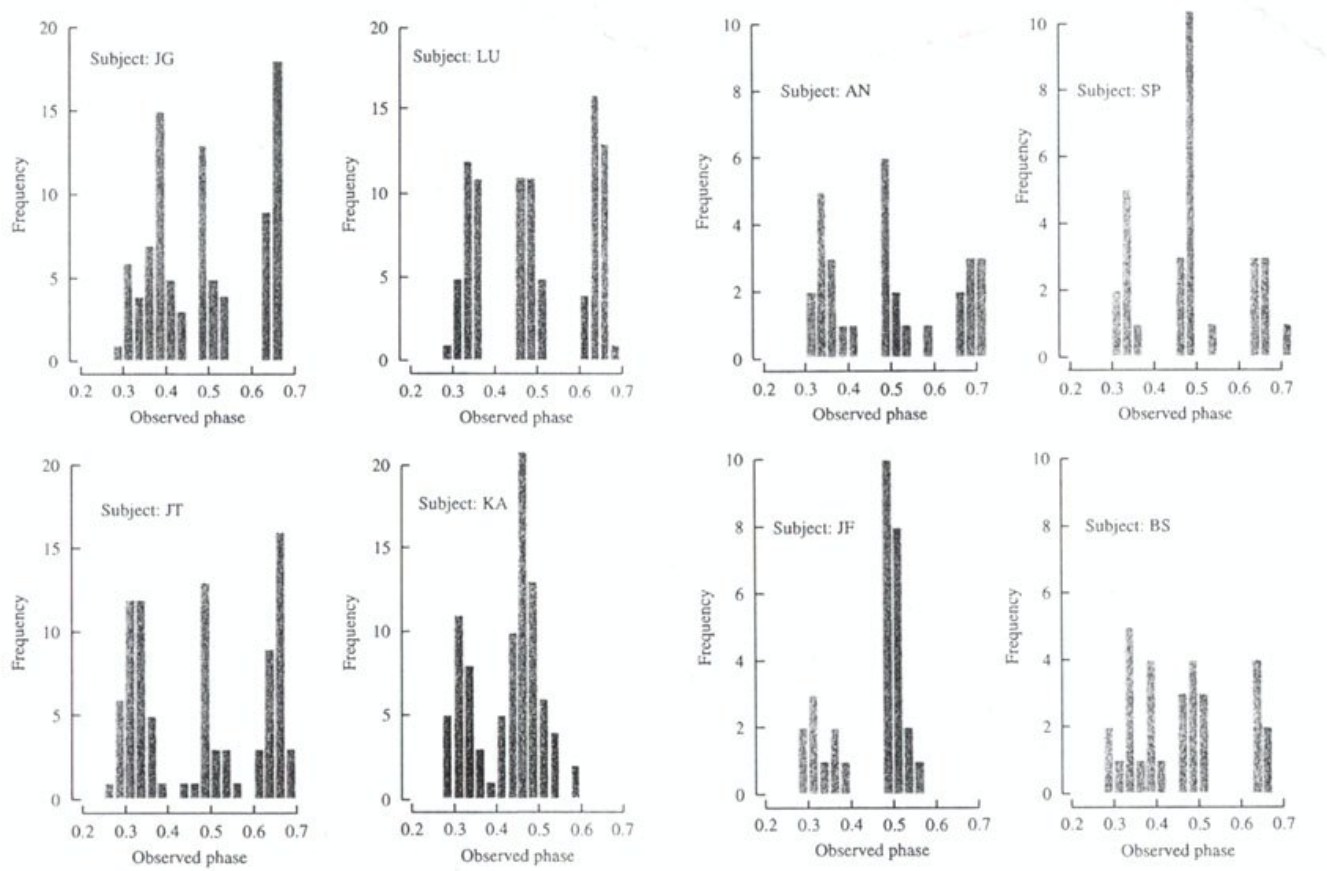

Figure 1: Density plots of observed phase for eight subjects from Cummins and Port (1998)

II. Implications of the Harmonic Timing Effect

These results are quite startling. They demonstrate that, without any training, speakers are strongly biassed to align stressed vowel onsets with the simplest harmonic fractions of a repetition cycle. By harmonic fractions, we mean the location of phase zero of an oscillator at a harmonic frequency of the repetition cycle, where the harmonic is phase locked with the repetition cycle.

In rather different terms, we have validated with appropriate time measurements the perceptual experience we had above of the rhythmicity of speech. Our subjects, and, indeed, all adults, are ready to exhibit this kind of behavior without special training on a moment's notice. Of course, rhythmic or periodic speech production can be observed, not just in this artificial task, but also when singing or chanting and or for brief periods during performance of familiar passages of prose.

Notice that this result demonstrates not merely periodic regularity in speech (which, after all, was supplied by our metronomic stimulus), but that there are NESTED periodic patterns. That is, there are

regular periods on two time scales: one at the repetition cycle rate and another either 2 or 3 times faster than the first but clearly phase locked to it. What kind of cognitive mechanism could account for these particular timing constraints?

Dynamical Interpretation of Meter

Consideration of these data have led to the formulation of two hypotheses to make sense of these phenomena. Neither is completely new. Both of these should be familiar from previous research.

When we find any unmistakably periodic behavior from an organism, one sensible theory is that something is oscillating to control that behavior. This normally implies that some interdependent parameters are being recursively updated (as if by a differential equation) that leads one parameter to rise and fall complementarily with another (McAuley and Kidd, 1998; Large and Jones, 199). An oscillating system can behave according to the equations without our knowing the degree to which the relevant parameters are mechanical or neural or cognitive. The mathematics of dynamical systems can still help us understand and make testable predictions about its behavior.

We can imagine an oscillator cycling such that every time the function reaches 1 (= 0) phase, it emits a pulse as in the top panel of Figure 2. The location of a pulse and the period of the cycle it initiates will adapt to the sequence of input pulses (McAuley and Kidd, 1998; Large and Jones, 1999). For the case of the nested periodicities, where the faster oscillator couples its phase zeros with the pulse of the slower oscillator, a more complex structure is required with at least 2 coupled oscillators (Large and Jones , Appendix).

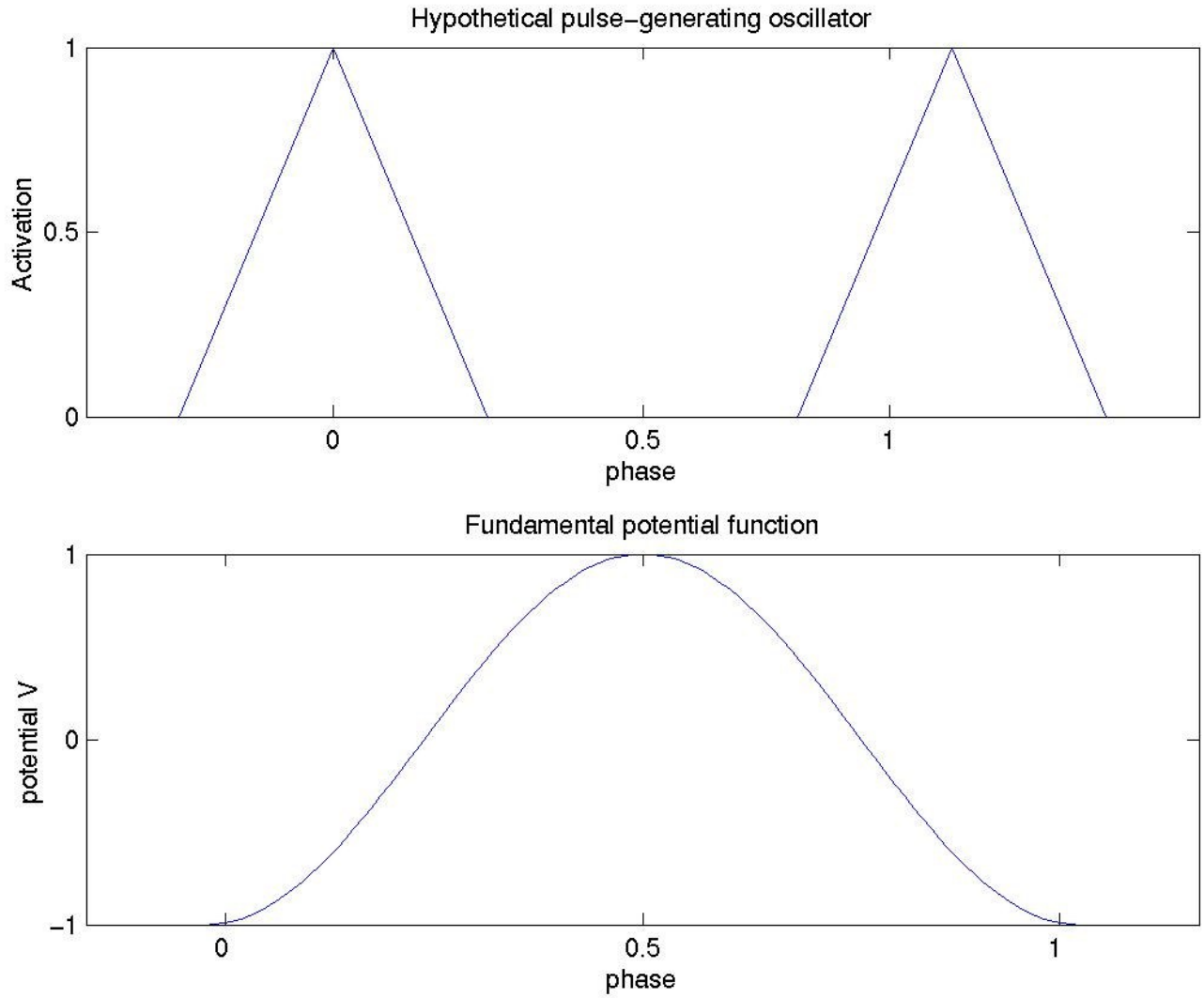

Figure 2: 1 pulse generating oscillator produces a potential function with one attractor. A faster pulse−generating oscillator couples with this fundamental pulse at phase 0 to produce complex periodicities.

So, our first hypothesis is:

H 1: Musical meter and the harmonic timing effect are set up in the nervous system by phase−coupled and frequency−coupled pulse−generating oscillators.

These oscillators tend to oscillate at different, but integer−ratio frequencies like 1:2 and 1:3. At these frequencies, if every second or every third pulse of the faster oscillator coincides with the pulse of the slower oscillator, then we have either a 2−beat or 3−beat meter respectively.

Now we need the second hypothesis. What is the significance of the

pulse for motor control? One might imagine it to be the moment of initiation of a movement. Actually, what is attracted to the pulse is the most perceptually salient event −− like the tapping sound of a finger (rather than, say, finger movement onset) or onset of a vowel (not the onset of the mouth opening gesture).

> H 2: Phase zero of any of these oscillations attracts perceptually prominent events (like vowel onsets or taps of a finger). The phase of the internal system is adjusted so that the perceptually salient event is synchronous with the oscillator pulse.

H2 accounts for why vowel onsets (especially stressed ones in the case of English) or finger taps tend to occur near the pulse of one oscillator or another, while H1 accounts for the underlying metrical structure itself.

Given a system of oscillators like this and a rule for locating attractors, we can propose to represent the multi−oscillator meter as a potential function for vowel onsets using the phase of the slowest oscillator (that is, the phrase repetition cycle) as a time scale. When the system has two oscillators at frequencies 1 and 2, the potential function should have attractors at both phi = 0 and phi = 0.5 (just like the potential function of Haken, Kelso and Bunz, 1986; Kelso, 1995). A useful first hypothesis is that the potential function is shaped like the sum of two inverted cosines, with one having a minimum at 0 (= 1) phase and the other having a minimum at both 0 and 0.5, as shown in Figure 3A.

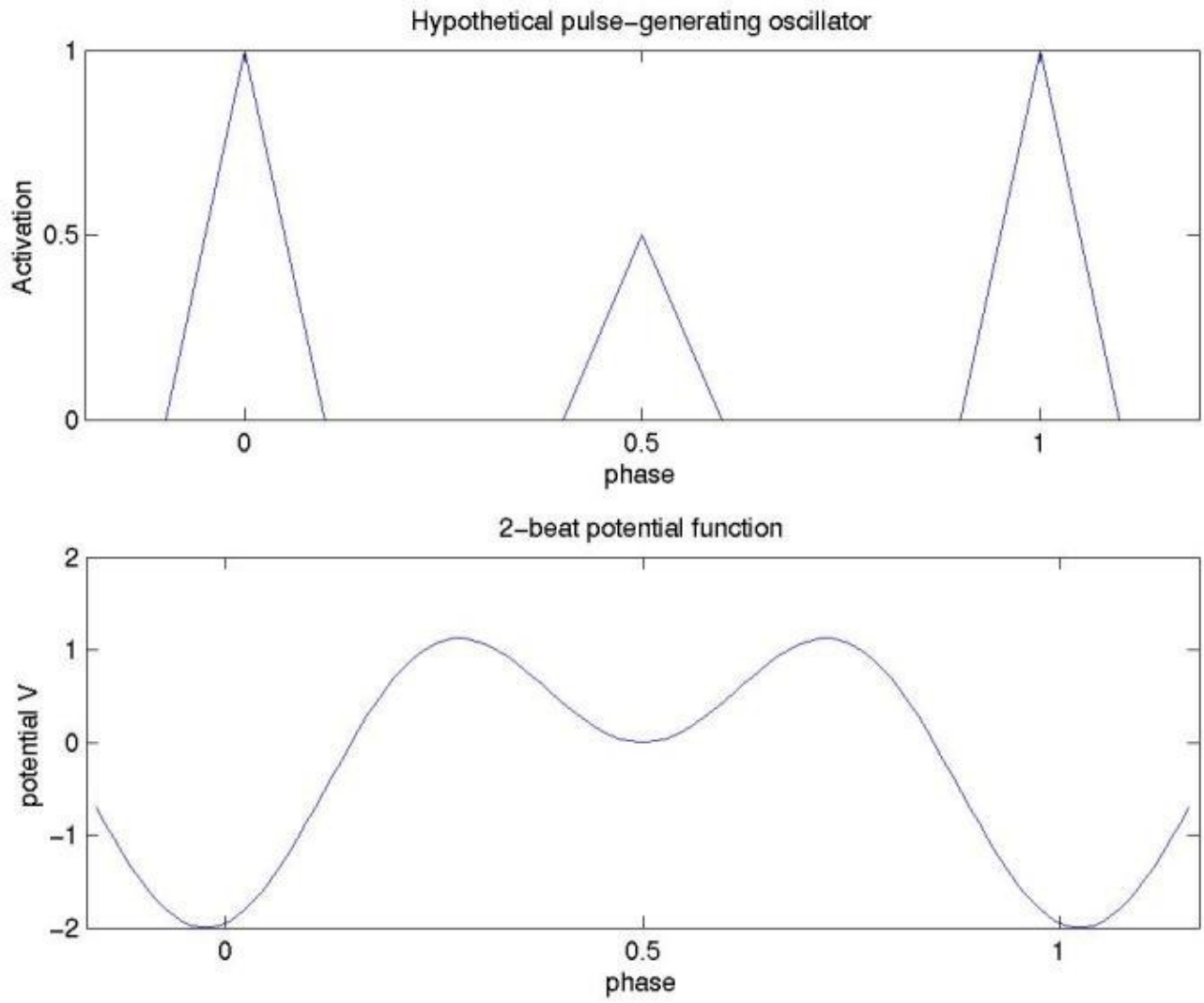

Figure 3A: 1:2 coupled pulse−generating oscillators create a potential function that attracts prominent onsets a 0 and 0.5 phase. The pulses in the top panel of this figure are of arbitrary shape and type.

****

Footnote: The equation for the shape of this potential function, V, would then be:

V(x) = − A cos phi (x) − B cos phi (2x)

The relative amplitude of A and B determines the degree to which the harmonic at 2x creates a stable attractor.

****

Since we also observe evidence of oscillators at the frequency ratio 1:3, we should similarly postulate a potential function with minima at phi = 0, 0.33 and 0.67 − at each location where the harmonic

waveform rises through its phase 0 on the assumption that the two oscillations are phase locked, as shown in Figure 3B. (Notice that Haken, Kelso and Bunz found no evidence of attractors here in their finger−wagging task.)

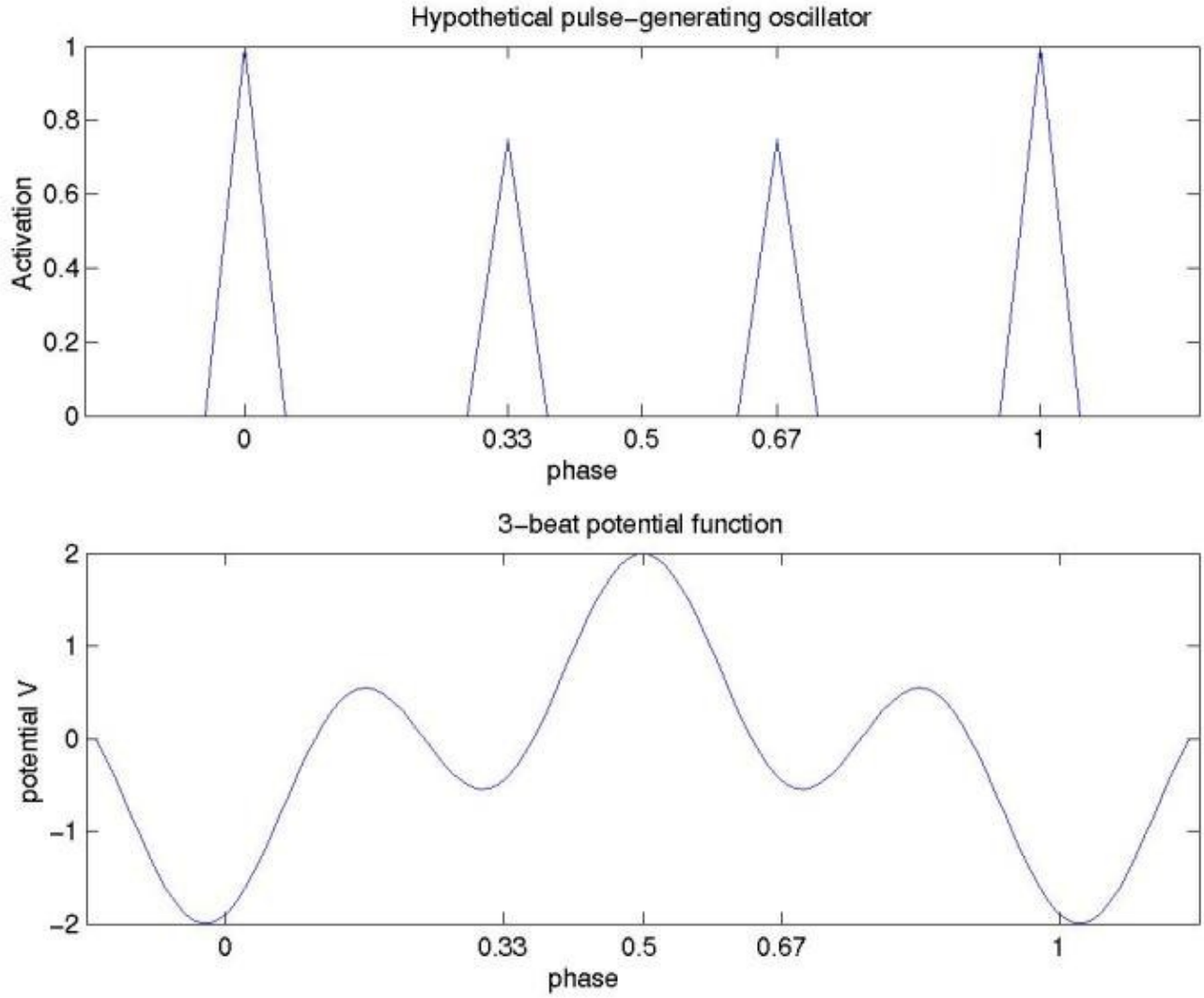

Figure 3B: 1:3 coupling of pulses generates a potential function with attractors at 0.33 and 0.67. In the top panel, the pulses are of arbitrary size and type.

Review of reasoning so far

So we observe evidence that certain temporal locations are preferred during periodic performance − as shown, eg, by the data in Figure 1. We interpret such data as reflecting the influence of temporal attractors at harmonic fractions of the main cycle. These are postulated to result from coupled oscillators (of some kind) in the frequency ratio 1:2 or 1:3 that generate a pulse once a cycle for all active oscillators that attract attention−grabbing events during production and is itself an attention grabber during perception (as

shown by Large and Jones, 1999).

Relevance for spontaneous speech? Of course, this display of strong temporal constraints was found while speakers were doing a very repetitive speech task. Still, these resonant behaviors probably still exert some influence on spontaneous speech as well, although we would expect the effects to increase under conditions where the speech text becomes more familiar, such as when it is memorized.

III. Predictions

There are some things we expect to be true of this oscillatory structure:

First, since the theory specifies attractors in terms of phase angle, we expect that at least for moderate changes in rate (that is, changes in the duration of the repetition cycle), the attractors should be unaffected in terms of phase but vary in direct proportion with cycle duration. This was verified, for example, in the Cummins and Port study (by varying the A−A tone interval over a range of over 10%) and in previous experiments.

Second, the attractors should vary in `strength' and their degree of attractiveness should be observable in the effects of perturbation on events near the attractor. That is, given a periodic perturbation of the system, any effect should be less prominent when the attractor is stronger (that is, when its potential well is deeper or has steeper sides).

Third, if we represent the attractor structure as a potential function along the phase positions over the range (0, 1) of the slowest oscillator, then phase zero should have the strongest attractor and the attractors created by harmonics of the repetition cycle (at various integer fractions of the longest cycle) should get weaker as their frequency increases − just as the harmonics of a plucked string have amplitudes that decrease as the frequency rises. Thus, an attractor at 1/2 should be less stable than the attractor at 0 or 1, and an attractor at 1/4 should be weaker than an attractor at 1/2, and so on. Such differences in attractor strength have been found in an experiment that compares perturbation of target syllables occurring at 1/3 of the cycle with the same text materials occurring at 1/4 (Port, et al, mspt 2002). As predicted, the attractor at 1/3 was stronger than that at 1/4.

Universals of Meter vs. Language Differences

Given these strong effects, the next question that arises is to what degree they are universal, as opposed to being characteristics of the `prosodic grammar' of this particular language. This issue was initially addressed by Tajima and Port (in press, 2002) who set up similar tasks for native speakers of Japanese and English to look for

similarity of response to the manipulated factors. Although much more exploration of this issue is required, the evidence suggest that there is something universal about, for example, a meter of 1:2. But there are also clear language−specific differences in speaker timing. The English speakers acted more as though there were attempting to regularize the inter−stress intervals while the Japanese speakers showed a greater tendency to regularize syllable durations (cf. Abercrombie, 1967; Pike, 1943).

Acquisition of Metrical Behavior

Since a straightforward method for observing rhythmic behavior in language is now available −− using the `speech cycling' method −− it is surely of interest to study the development of these metrical structures in children to see how these patterns are acquired in English and other speech communities. A simple test would be to ask a child to repeat a simple phrase, like `Give me a cookie', over and over. Perhaps initially children will exhibit less regularity in timing (that is, larger standard deviations) and exhibit weaker attraction to harmonic fractions of the repetition cycle than they will later in linguistic development. Alternatively, these metrical patterns might be observed fully formed almost as soon as multisyllabic utterances are possible. Whatever the case is, how early do language−specific differences in metrical preferences appear? If newborns can recognize the prosodic structure of their mother's speech, perhaps they already have the beginnings of a metrical model for their future native language.

A number of questions arise regarding the acquisition of metrically constrained speech, beyond the question of when the earliest evidence is found. Presumably, a meter with 2 coupled oscillators (eg, with frequencies f and 2f) appears later than a meter with only one level of periodicity. Is a 3−beat, waltz−like meter more difficult than a 2−beat meter? Another important issue is whether children may be MORE constrained by metrical constraints than more skilled speakers. It seems entirely possible that children may lean on regular metrical patterns as they learn to produce fluent, multiword utterances. Thus we might observe more regular timing in children's spontaneous speech than in adult speech.

IV. Concluding Discussion

In conclusion, then, primitive speech rhythms depend on oscillations that tend to fall into discretely distinct forms based on integer−ratio time intervals (like 1:2, 1:3, etc). Given additional experimental results, these temporal patterns seem to imply that:

1. The speech rhythms (tendencies to locate beats at periodic locations) result from (or are constrained by/ timed to be in accord with) cognitive oscillatory structures. We don't know what may be oscillating at these rates, but we infer something must be to account

for the phenomena.

2. Some aspects of speech rhythm are universal, and probably arise early in linguistic development, while others differ between languages,

3. Even nonperiodic spontaneous speech must be somewhat influenced by these dynamics, just as the mechanical resonance of, say, a limb will PARTIALLY account for whatever behavior may be imposed by the body on the limb is attached to. That is, we would expect to see some evidence of the limbs mechanical attractors in its overall behavior. It may be these resonances that account for intuitions of different rhythmic types between languages such as those proposed by Abercrombie and Pike.

4. If languages differ in their characteristic rhythmic behavior at the time scale of syllables and phrases, then a phonological grammar should probably be built on TOP of this sloshy, dynamical timing system −− one that can easily be set into periodic oscillations in a partly language−specific style. This sloshing system creates attractors for prominent events (such as syllable onsets and stressed syllable onsets) that appear and disappear like waves. These temporal constraints may provide a framework on which to `hang' individual phonological syllables and segments.

5. It seems likely that this kind of global temporal patterning could be acquired fairly early on in the process of language development. I hope that investigation of the developmental phenomena can get under way soon.

BIBLIOGRAPHY

Abercrombie, D. (1967). Elements of general phonetics. Aldine Pub. Co., Chicago.

Allen, G. (1972). The location of rhythmic stress beats in English: An experimental study I. Language and Speech, 15:72−−100.

Cummins, F. and Port, R.F. (1998). Rhythmic constraints on stress timing in English. Journal of Phonetics, 145−−171.

Haken, H., Kelso, J., and Bunz, H. (1985). A theoretical model of phase transitions in human hand movements. Biological Cybernetics}, 51:347−−356.

Jusczyk, Peter (1997) The Discovery of Spoken Language. MITP.

Kelso, S. (1995). Dynamic Patterns: The Self−Organization of Brain and Behavior. MIT Press, Cambridge, MA.

Large, E.W. and Jones, M.R. (1999). The dynamics of attending: How we track time varying events. Psychological Review, 106:119−−159.

McAuley, D. and Kidd, G. (1998). Effects of deviations from temporal expectations on tempo on discrimination of isochronous tone sequences. Journal of Experimental Psychology: Human Perception and Performance, 24:1786−1800.

Mehler, J., G. Lambertz, P. Jusczyk and C. Amiel−Tison (1986) Discrimination de la langue maternelle par le nouveau−ne. Contes Rends de l'Academie de sciences de Paris 303, 637−640.

Morton, J., Marcus, S., and Frankish, C. (1976). Perceptual centers (p−centers). Psychological Review, 83:405−408.

Pike, K. L. (1943). Phonetics. University of Michigan Press, Ann Arbor.

Port, R. and Leary, A. (2002, in press). Speech Timing and Linguistic Theory de Boeck University Press.

Port, R.F., Cummins, F., and McAuley, J.D. (1995). Naive time, temporal patterns and human audition. In Port, R.~F. and van Gelder, T., editors, Mind as Motion. MIT Press, Cambridge, MA.

Port, Robert, David Collins, Ken de Jong, Adam Leary and Deborah Burleson (2002). Temporal attractors in rhythmic speech. Submitted.

Tajima, K. and Port, R.F. (2002). Speech rhythm in English and Japanese. In Local, J., editor, Papers in Laboratory Phonology VI. Cambridge University Press.

van Gelder, T. and Port, R. (1995). It's about time: Overview of the dynamical approach to cognition. In Port, R. and van Gelder, T., editors, Mind as motion: Explorations in the dynamics of cognition}, pp1−43. Bradford Books/MIT Press.