Speech and Rhythmic Behavior*

Robert Port, Keiichi Tajima, and Fred Cummins
Dept. of Linguistics, Cognitive Science Program
Indiana University, Bloomington, Indiana 47405, USA

February 14, 1998

Abstract

Animals and humans exhibit many kinds of behavior where the frequencies of gestures are related by small integer ratios (like 1:1, 2:1 or 3:1). We show that speakers who repeat a short phrase to a metronome have a strong tendency to place the onsets of stressed syllables at temporal harmonic fractions of the metronome cycle (like 1/2, 1/3 and 2/3). Studies of errors by early language learners also show that some metrical patterns are easier than others. All these effects support a view of meter as an abstract dynamical system on the state space of two or more oscillators.

1 Introduction

It is a common observation that human speech is often rhythmically produced. One thinks of worksongs, nursery rhymes, auctioneer calls, group recitation of prayers and chants, marching songs, cheers at sport events, chants by train conductors and so forth. It is worth our time to wonder where such rhythmic performance comes from. It appears that typical speech rhythms vary from language to language, especially in ordinary prose. Chinese and French, Japanese and English all have different characteristic temporal structures (although language-specific rhythms must still exhibit some universals). If we restrict ourselves to a single language, can we find any relationship between the rhythmic structures observed during speech production and those employed in nonspeech motor behaviors? Does speech exhibit timing similarities to other kinds of rhythmic or periodic behavior by our species? People exhibit many cyclic behaviors: walking, waving an object, repetitive reaching, finger drumming, scratching an itch or dancing. We suspect that many properties of these skills will be found in speech as well.

To acknowledge such a similarity would imply a description of speech that is dynamical and employs continuous time. Of course, such descriptions are at odds with those typically employed in linguistics, where one is restricted to discrete and timeless segments exclusively. Phonological research in linguistics has generally treated "prosody" in terms of timeless symbols, like **S** and **w** ('strong' and 'weak'), where units of various sizes from syllables to whole phrases possess values of strong or weak (Chomsky and Halle, 1968; Liberman and Prince, 1977; Selkirk, 1984; Hayes, 1995). However, there is no continuous time dimension at all, no natural way to interpret strong and weak as having consequence for the timing of speech (Kelso, 1995; van Gelder and Port, 1995). Yet speech prosody does affect speech timing in many ways. All of these effects are likely to require dynamical models to account for them (see Port and van Gelder, 1995). Evidence for the effects of prosodic structure on the timing of speech is found starting from the earliest phrase constructions by children at two years of age, up into adulthood where our own research has been conducted.

We begin this essay with a discussion of some examples of "self-entrainment" of motor gestures in animal and human behavior. Second, we will suggest that the notion of *meter* should be defined as a special case of the self-entrainment of oscillators, and thus should be viewed as essentially a dynamical concept—an attractor structure of at least a pair of oscillators. Meters of this kind are characteristic of musical and

*To appear in G. J. P. Savelsbergh, H. van der Maas and P. C. L. van Geert (eds), *The Non-linear Analyses of Developmental Processes*, Royal Dutch Academy of Arts and Sciences, Amsterdam.

poetic traditions throughout the world. Third, we will show from simple experiments that speech too can be locked into simple metrical structures, in which some are simpler and more natural than others. Finally, we will present some evidence that the relative difficulty of various meters in adults is mirrored in speech errors made by children in early stages of language learning. Our conclusion is that some very simple rhythmic structures, like 1:1, 2:1 and 3:1 are widespread in human physical, as well as cognitive, behavior.

1.1 What is "Self-entrainment"?

There is a very important characteristic of the behavior of humans that seems to be largely overlooked as a general feature of cognitive systems. This is the fact that humans and animals typically exhibit *self-entrainment* in their physical activity. But what is self-entrainment? When one physically oscillating system *entrains* another, it means that the timing of repetitive motions by one oscillator influences the motions of the other oscillator such that they fall into a simple temporal relationship with each other—that is, the oscillators tend to perform their motions in the same amount of time or in half (or double) the time, or some other simple integer ratio of time.

Mathematically this can be modeled by minimally assuming a system of differential equations of the form:

$$\dot{x} = f(x, y)$$
$$\dot{y} = f(y, x)$$

where x and y are periodic and $\dot{x}$ and $\dot{y}$ are derivatives with respect to time. One well-studied example is the van der Pol equation (see Abraham and Shaw, 1983, or Glass and Mackey, 1988). When the behavior of each oscillator is affected by the state of the other, then they will couple, that is, they will tend to fall into stable regimes in which their frequencies are close to a small integer ratio, and where their instantaneous phases tend to be locked together.

By *self*-entrainment we simply mean that the two oscillators in question are parts of a single physical system, for example when a gesture by one part of the body tends to entrain gestures by other parts of the same body.

This property is well-known in the literature of motor control. Researchers on motor control have commented many times on its general importance and tried to encourage others to see its relevance to an understanding of the temporal structure of human cognition (e.g., Bernstein, 1967; Kelso, 1995; Turvey 1990). However, most other branches of cognitive science tend to underappreciate the significance of this phenomenon for an understanding of cognition as a whole. This reflects a lack of appreciation of the problem of time in general within cognitive science (see van Gelder and Port, 1995). If one takes the problem of the timing of events in cognition to be important, then any such simple but powerful temporal constraint immediately presents itself as potentially of great utility for understanding the production of language.

1.2 Self-entrainment in Everyday Activity

Research on motor behavior especially in humans, as well as in other animals like fish, shows that when one gesture is performed simultaneously with another gesture—even by a distant part of the body—the two gestures have a strong tendency to constrain each other. Glass and Mackey (1988) review many such cases from biology.

It seems that for some actions, several oscillators must be coordinated simply to do the action—like the full set of legs when walking (two for bipeds, six for most insects). But some cyclic gestures *could* in principle be completely independent—like walking while waving the hand or wagging the index fingers of each hand (or motion of the two pectoral fins of a goldfish). Yet they tend *not* to be independent—especially if the gestures continue for awhile. Gestures often tend to cycle in the ratio 1:1 or 1:n or m:n (where m and n are very small integers). For example, most joggers notice that during steady-state jogging, one's breathing tends to lock into a fixed relationship with the step cycle—with, say, two or three steps to each inhalation cycle, or perhaps three steps to two inhales (Bramble and Carrier, 1983). To the runner it *feels easier* and less effortful and it probably is (Diedrich and Warren, 1995).

Similar preferences for integer ratio gestures have been observed in the laboratory in various forms for over a hundred years (see references in Collier and Wright, 1995 and Treffner and Turvey, 1993). For example,

in one recent study (Treffner and Turvey, 1993, Experiment 3) subjects were asked to sit in a chair with their arms resting on the chair arms and then to swing a pendulum along each side of the chair. The pendula were pieces of dowel with various weights attached at one end, so that they had varying natural frequencies, that is, different frequencies at which they would swing most easily (e.g., if suspended from a fulcrum at some point along its length and swung). This frequency is quite close to the rate at which the subjects swing them with least effort. They studied how each pendulum was swung just by itself and also when the other arm had a different pendulum to swing, for various pendulum combinations. The results showed that when swinging the two pendula, subjects had a strong tendency, after some settling time, to entrain their arms to each other in simple ratios like 1:1, 2:1, 3:1 or 3:2—depending on differences in the natural frequency of the two pendula. Other, more complex ratios would occur sometimes (like 4 cycles on the left to 5 cycles on the right), but these were quite unstable and seemed to frequently slip over to simpler ratios, like either 3:2 or 2:1.

This phenomenon, where one arm entrains the other one, is not restricted just to cyclic activity like waving a pendulum or a finger. Just a single, one-time gesture, if performed by two hands, tends to exhibit entrainment. Kelso, Southard, and Goodman (1979) asked subjects to perform easy and hard reaching gestures with one or two arms, and to do so as quickly as they could without making many errors. In one case, they sat at a table and put their right index finger on a spot (with a touch sensor). Then on a signal, they reached to their right along the table to touch either a nearby target area with a large touch sensor or else a small, more distant target area. They found that subjects can perform the short, easy reach quite a bit faster than the harder, more distant reach, with either hand. This is no surprise, of course. But when they asked subjects to perform these reaches with both hands simultaneously, they found that the easy reach was strongly constrained by the harder one. Both gestures started and ended together and took just about the same amount of time. It seemed that each hand entrained the other in the ratio 1:1 for the duration of the reach. Kelso *et al.* interpret this as evidence about the attractor structure of the dynamical system that provides the mechanism for coordinating the two gestures. Coordinating the two reaches so that they are the same duration is apparently easier and more reliable than to allow (or force) each gesture to be independent of the other.

It is not claimed that this kind of self-entrainment is the *only* way the two arms can be moved. With enough practice and suitable motivation, subjects could presumably learn many possible relations between the arms. But it seems to be the most natural way. Apparently the most straightforward way for humans to perform a novel coordinated act is to find a very simple harmonic ratio for their durations, preferably 1:1, but possibly also 2:1.

Other examples of the strong tendency toward self-entrainment may be found in difficulties faced when learning to play musical instruments that use both hands—that is, instruments like guitar, piano, flute and drums (as opposed to trumpet). In these instruments, the novice must learn to control the phase relations of the two hands appropriately. To produce an extended roll on a bongo drum, for example, the two hands must be kept at opposite phase even at very high rates. For novice drummers the hands tend to slip over to zero phase lag (where they hit the drum head simultaneously). In the case of plucked string instruments, like the guitar, the fingers of the left hand must clamp into the fingerboard *just before* the plectrum strikes the string with the right hand. It seems to us that for a novice player, there is a tendency for the left hand finger to slap down on the string *simultaneously* with the plectrum strike across the string. Of course, you don't get a good sound in this case (because the plectrum excites the string while it is still partially damped by the soft flesh of the finger tip). If they keep practicing, learners eventually get the phase offset correctly and can then produce fast runs and arpeggios.

The piano is played in many different styles, and there may be different phase relationships that are typical of each genre. Compare playing a military march on the piano (where a feature of the style is that the left and right hands frequently strike the keys simultaneously) with playing boogie-woogie (where the left hand beats out a steady bass pattern on the beat while the right hand operates seemingly independently with comparatively few of its strokes being in phase with the left hand). Getting the knack of such a style of performance requires decoupling the two arms from each other in some sense—a challenging task for the learner. It seems that musical performance skills frequently involve careful control of self-entrainment by parts of the body.

Another characteristic of the self-entrainment of separate gestures is that any similarity between the

gestures tends to be automatically increased or exaggerated. For example, the old task of rubbing the tummy and patting one's head is notoriously tricky until one has practiced it a little. One reason it is hard is that, although each gesture by itself—the rotary gesture and the pat—is an easy and familiar skill, one discovers when performing them together, that they have very similar natural frequencies (at least if you are moving the whole arm at the elbow in each case). This similarity of natural period seems to lead the different hand gestures to interfere with each other. In contrast, for example, if you rest your arm on a table and tap one finger rapidly while rubbing the belly slowly, there doesn't appear to be much interference (although if you keep it up, they will still probably fall into a regular 1:n harmonic ratio with each other). Why does the degree of similarity of period between the competing gestures matter? It is probably because a one-to-one temporal relationship tends to encourage treatment of the cyclic events as the same event. The novice guitar player has the same difficulty; strokes with the left and right hands get merged or confused with each other. These familiar examples demonstrate the many cases of entrainment or mutual interference between oscillations of one body part and another part.

Finally, it is important to point out that the easiest ratio of self-entrainment other than 1:1 is 1:2, where one oscillator cycles twice for each cycle of the other. A three-beat pattern is probably the next easiest. For example, in a finger tapping experiment by Yamanishi, Kawato and Suzuki (1980), subjects were asked to tap a finger on, say, the left hand at a specified phase lag relative to the finger on the right hand. You might not be surprised that they were most accurate when the phase lag between the fingers was either at 0 or near 1/2. That is, aside from unison tapping, it was easiest to accurately perform the task when the left and right index fingers alternated evenly. These gestures can be interpreted as two oscillations at the same frequency that are in opposite phase. However, since in humans one finger always dominates the other (usually the right one), it is also possible that psychologically the subjects could entrain these motions with a 2:1 meter where the non-dominant finger is on the 'off-beat'. Of course, one might have thought that 1/3 and 2/3 would also be fairly easy, so that, as if in waltz time, the two hands would go Left-Right-[rest], Left-Right-[rest].... But their subjects did not find these waltz-time interpretation easy. These are certainly temporal ratios widely employed in musics of the world. However, of their 9 subjects none were able to accurately produce target phases near 1/3 or 2/3. Only the target phases of 1/2 or 0 were accurately performed.

Similarly, Haken, Kelso, and Bunz (1985) found that wagging the fingers in an "out-of-phase" relationship (that is, phase = 1/2) could be accurately performed at slower rates, whereas phase lags of 1/3 or 2/3 did not appear. So, an important observation is that phase lags of 1/2 (closely related to nested oscillator ratios of 2:1) are easier than 1/3 or 2/3 (implying a nested oscillator ratio of 3:1).

1.3 Perceptual Self-entrainment

For the cases mentioned above, the coupling might be due largely to the physical link between the physical oscillators: the legs and trunk in the jogging case, or the two arms in the pendulum data. That is, the observed coupling effect seems to be due entirely to physical forces—and thus arguably to have little relevance to theories of cognition. However, very similar effects are also observed where one of the oscillations involved has such extremely low stimulus energy that the coupling must be considered to be strictly "informational"—that is, made available by the auditory or visual system—and is not due to an interfering force. In one experiment (Schmidt, Carello and Turvey, 1990), subjects swung one leg from a seated position on the edge of a table. They were asked to watch another subject sitting next to them on the table and to swing their leg either in phase or out of phase with the other person at various frequencies. At slow rates, the subjects were able to keep their phase close to the assigned values of 0 or 0.5 with respect to each other. However, they showed a strong tendency to fall from the out-of-phase pattern (one leg forward, the other back) to the in-phase pattern when rate was increased. So, although nothing but visual information links the two systems, the behavior is exactly the same as the behavior we discussed when the novice tries to produce a trill on a bongo drum. It is also identical to what is observed in the well-studied laboratory task where subjects wag their two index fingers in phase and out of phase at various rates (Kelso, 1995).

Thus, it makes little difference whether the two limbs are in the same body (where physical forces can account for the coupling) or in different bodies (where only stimulus information could account for coupling). Apparently then, the entrainment phenomena include cases of entrainment between the visual system and motor control for limbs. From this perspective, the common tendency to tap our foot or nod our heads to

music is just another example of self-entrainment between the auditory system and the motor system.

The conclusion we draw from these experimental and anecdotal observations is that physical links between independent oscillators is certainly not the only mechanism and probably not even the primary mechanism to account for the widespread observation of mutual entrainment in humans and animals. It seems that the coupling of different kinds of oscillation at simple harmonic ratios is *a ubiquitous and intrinsic property of animal control systems*. If this is so, then we might expect to see broad exploitation of (or at least accommodation to) this property in other aspects of human cognition. In fact, it seems to us that in humans at least, there is evidence of entirely abstract, domain-independent temporal structures consisting of coupled oscillations—purely psychological or cognitive structures in time that do not depend on physical motion by any massive object.

1.4 What is Meter?

Thus far, we have looked at cases of stable action patterns based on entrainment of parts of the body with itself or with external oscillators. These suggest how neural systems can find stable structures that closely resemble simple physical dynamical systems. To us they suggest a novel way of defining meter, one that differs from the standard approaches.

The usual approach to meter is to define it in algebraic terms that seem to be directly implemented in the standard European musical notation. That is, for example, a waltz-like meter is defined simply as three "time units"—such as quarter notes or eighth notes, or more generally as simply beats. The time units of musical notation are postulated as *a priori* objects. They are represented most often in conventional musical notation by either quarter-notes (e.g., in 3/4 meter) or eighth notes (in 6/8 meter). And any particular meter is defined by statements like "A waltz has three beats per measure". It then follows from universal principles that the first one of the three will be the accented or special beat (Lerdahl and Jackendoff, 1983) . The rate or period of basic beats happens to be close to the motions that can be made by swinging a human finger, hand, arm or leg.

In linguistics too, meter has become a standard notion over the past 30 years. In both Metrical Phonology and in Prosodic Phonology, similar assumptions to those of music theory are made. A standard discrete-sized unit, usually a syllable, is assumed as the basis from which larger nested (or hierarchical) structures are constructed by stringing or nesting the basic units into feet, phonological phrases, etc (Liberman,1978; Hayes, 1995; Nespor and Vogel, 1986). For both theories the assumptions are about static objects, actual time applies as the symbols are implemented.

In contrast to these, we propose that meter be seen as just an instance of a stable structure in time. To understand meter, we should concentrate attention on the dynamical system that supports the integer-ratio time structure—i.e., on the vector field rather than the symbolic tokens themselves. We suppose that musical meter is a special case of self-entrainment (like the examples shown above) where both of the oscillators are internal to the brain. Or we might say that they are cognitive oscillators. The meters based on 2:1 and 3:1 seem to be found in most musical cultures. These simple ratios often are nested within each other to create even more complex meters like the typical 4/4 (two beats within two half-measures = 4:1) of a polka or the meter of the Cuban *bembé*, which is a cycle of 12 fast beats that is simultaneously divide into both three measures (of 4 beats each) and 4 measures (of 3 beats each) by different percussion voices.

Our interest here is not in musical meter, but only in the most basic and simple meters—ones that are simple enough that they might be generally influential in speech production. These are especially the 2:1 pattern and secondarily the 3:1 meter.

1.5 Self-entrainment in Speech Timing

Phoneticians and phonologists seek to specify the kind of information speakers and listeners employ in understanding and producing speech in various languages. Within linguistics the standard assumption is that this information has an essentially static structure consisting of segments and features, and hierarchical trees of such tokens. Although linguists have long sought support for *static* units of speech (Jakobson, Fant and Halle, 1952; Chomsky and Halle 1967; Ladefoged, 1972), there have been many difficulties with such attempts (e.g., Lisker and Abramson, 1971; Dorman, Raphael, and Liberman, 1979; Port and Dalby, 1982). Speech events always take place in time (unless they have been already written down on paper), yet

linguistics insists that temporal properties are not relevant to anything that could be called "the language itself".

However, various kinds of temporal units have been proposed for speech in various languages: inter-stress intervals in English (Martin, 1972), the mora in Japanese (Port, Dalby, and O'Dell, 1987; Han, 1994), the bisyllabic foot in Finnish and Estonian (Lehiste, 1990), etc. The notorious difficulty with these units is that they do not seem to be quite as regular as one might hope. Attempts to find isochronous stresses in English (as predicted by Abercrombie, 1967) have not been judged to be very successful (Lehiste, 1977; Dauer, 1983). Of course, what counts as success will necessarily depend on what method is employed to make measurements and on the computational theory that is presupposed. A clock measuring in milliseconds is probably not an appropriate mechanism (Port, Cummins, McAuley, 1995). But can self-entrainment be observed in speech? If so, then perhaps an alternative "clock" for speech timing can be created.

A second place to look for entrainment would be in performance of songs or poetry (see, for example, Boomsliter and Creel, 1979). Human cultures seem invariably to create rhythmical genres of artistic speech. Notice that since we are looking for temporal effects, only actual performances or recordings of them will provide relevant data, not the kind of orthographic or phonetic descriptions of performance customarily used by linguists. To look seriously at temporal events requires differentiating many different styles of speech, since any piece of text can normally be pronounced by native speakers with a wide range of possible rhythmic and intonational styles. The linguistic competence of speech rhythm cannot be explored with mere symbolic transcriptions, but intrinsically requires realtime analysis methods.

Aside from music and poetry recordings, a simple speaking task that is easily amenable to laboratory work might be to create an artificial speech style that encourages an interaction between the natural temporal rhythm of speech and some other periodic event. To propose an extreme case, we might ask subjects to pound a large hammer into a pillow while repeating some phrase: *"He POUNDed the hammer, He POUNDed the hammer..."*. Or they might be marched along a treadmill while counting *"Hup Two Three Four"* or some other text. If we found speech timing to be easily entrained to strenuous nonspeech actions in cases like these, no one would be surprised. But what if one takes a more gentle approach and simply asks subjects to repeat a particular phrase over and over at least 5 times? *"Pick any phrase you want. Pick any phrase you want. Pick any phrase you want ...."* For example, if subjects hear a metronome that periodically signals them to start producing some piece of text, then we will find that they will indeed put the first stressed syllable on the metronome. The surprising result is that the second stressed syllable will locate at harmonic fractions of the phrase onset. The periodic perceptual pattern produced by the metronome signal might draw into entrainment any potential periodicities within the text. We sought to design experiments that would see if speech would entrain to harmonics.

If it turns out that this kind of speech self-entrainment occurs easily, then we might conclude that normal speech timing probably is based on mechanisms that are similar in fundamental ways to oscillatory systems. This will have implications for basic theories about speech control. For example, some theories will be less compatible with these results than others. In particular, theories of speech motor control that emphasize executive, top-down, turn-on/turn-off models (e.g., Chomsky and Halle, 1968; Levelt, 1989) are less compatible than models based on, say, fixed-point attractors and damped oscillators (e.g., Kelso, Saltzman, and Tuller, 1986; Browman and Goldstein, 1995).

1.6 Speech Cycling and Beats

In the past couple years, we have been exploring what we call the Speech Cycling task, where subjects repeat phrases over and over, with their timing stabilized by a simple metronome (Port, Cummins, and Gasser, 1996; Cummins and Port, 1997; Tajima, 1997). Measurements of the location of one or more prominent syllable onsets, or 'beats', can be made and interpreted as phase angles relative to the repetition cycle of the text fragment (Cummins and Port, 1997). The basic task can be elaborated in various ways and various kinds of text fragments can be cycled in any language we wish. We suspect that versions of this class of tasks will prove useful for research on many phenomena in phonetics, phonology and probably other psychological issues.

The fundamental idea is to get speech to entrain to an external oscillator, thus stabilizing the performance in time and providing a standard time scale (that is, phase angle) so that temporal constraints can be clearly

observed. There is also an opportunity for experimental perturbation (e.g., of rate, etc) to explore the dynamical aspects of speech production (cf., Haken, Kelso, and Bunz, 1985; Kugler and Turvey, 1987; Treffner and Turvey, 1993).

The first two experiments reported here are early explorations of these phenomena. They demonstrate some simple and striking properties of this task. Experiments 3A and 3B gets into details of preferences for certain prosodic patterns over others—in particular, for *"DUM da DUM da"* over *"DUM da da DUM da da"*. But first let us look at our early exploratory experiments with the speech cycling task. Although our later experiments are better controlled, these early ones demonstrate the robustness of the basic effects.

2 Experiment 1: Rate Change for Simple Phrase

2.1 Methods

The subject for this experiment was a male college-aged native speaker of American English. The subject was asked to repeat a short sentence triggered by a short beep, a 50-ms, 500-Hz sine wave presented at a comfortable level over a loudspeaker. The phrase was *"talk about the game"*. The speed of the metronome was varied in two ways. In the Increasing Rate condition, the metronome rate was changed in 11 steps from slow to fast, from 3 sec down to 460 msec. For each step the period was decreased to 75% of the previous period down to 1 sec. Below that they were decreased to only 87.5% with each step. In the Decreasing Rate condition, tempo was decreased in the same steps from fast to slow.

In this experiment we explored the use of minimal instructions. The subject was asked to "say this phrase in time with the metronome beeps". He was told that if the metronome got too fast for him, then he should just slow down (that is, drop back to taking two metronome cycles for each repetition). He was not instructed to align any particular part of the phrase with anything else, nor was he given any practice or reinforcement for this task.

For each rate, the subject listened to the metronome for about 10 pulses and then jumped in to repeat the target phrase 10 times while listening to the beeps. Then the subject paused and waited until the next metronome rate was produced and then repeated the task. Each set of 10 repetitions at a single rate is referred to as a trial. In total, the subject produced 110 repetitions of the phrase in each of the Increasing and Decreasing rate conditions. All experimental productions were recorded directly onto computer disk.

Since the onset of a vowel is an acoustically prominent event—and one that is known to produce a strong burst of neural activity in the auditory nerve (see Delgutte, 1996), and because it has the advantage of occurring in most syllables, we measured the location of the vowel onset in each syllable by hand from computer-generated waveform displays.

2.2 Results

We report here only the data on the onset of the vowel in the initial word *"talk"* and in the nuclear-stressed word *"game"* which began the second foot in our test phrase. The location of the onset of *"game"* was then computed as a phase angle with respect to the phrase repetition cycle. This was done by taking the time interval between *"talk"* and *"game"*) and dividing this by the interval between the two successive phrase onsets (that is, from *"talk"* to *"talk"*. Because a cycle cannot be determined for the very last repetition of a trial, only 9 phase measurements were obtained from the 10 repetitions at each rate in each condition.

As an illustration of subject performance, Figure 1 displays the sequence of phase angles of the onset of the final, stressed word in the phrase as they occurred through a single run of trials from the slowest to the fastest rate. Beginning on the left, we show the first trial of 9 productions when the metronome was set to 3 sec. It can be seen that the onset of *"game"* occurs at a phase angle of about 0.2 (0.2×3 sec = 150 msec after onset of *"take"*). As the rate is increased, the phase angles get higher, up through the first three or four trials. Then the phase remains around 0.45 to 0.48 for trials 4, 5, 6 and even 7. trial 8 is highly variable, as the subject slips later and later. Then in trials 9, 10 and 11, the subject takes two metronome cycles for each text cycle and the onsets again drop back to the region just below 0.5 again. Thus, it appears that for quite a bit of the range, the phase angle hovers just below 0.5.

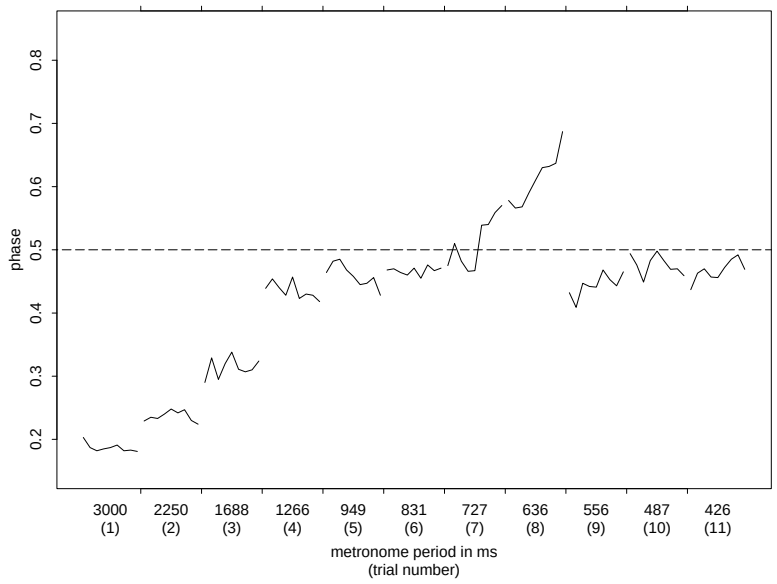

Figure 1. The phase angle of the onset of *"game"* as produced in the Increasing rate condition of Experiment 1. For each trial (that is, metronome rate), there is a series of 9 sequentially produced data points. Note that the speaker spends considerable time in the region of phase just below 0.5.

To see the preference for this phase clearly, we collapsed the data across the two presentation conditions (Increasing and Decreasing rates) and display a frequency histogram of the phase of the produced phase as a function of metronome rate in Figure 2. This display shows a very strong bias in favor of phase close to, but just below, 0.5. From these data it is not clear whether there might be other modes near 0.3 and 0.6, but other experiments in our laboratory demonstrate that modes can often be found near 1/3 and 2/3 as well (Cummins and Port, 1997).

Of course, the preferred phase in both Figures 1 and 2 is not exactly 0.5. It appears to be closer to 0.46. We should point out, however, that when listening to it with our "musician's ears" (all three authors are amateur musicians), we can only say that it *sounds* impressionistically like the syllable is "on a beat" which is half way through the phrase repetition cycle. That is, if asked to use musical notation to write down the speaker's speech rhythm, we would be very confident in starting this word at the second beat of a two-beat measure. The question why the apparent preferred phase is at 0.46 rather than 0.5 cannot be properly treated here. First, we have not fully validated our choice of measurement point, the definition of a beat. As is well known from the "P-center" research (Morton, 1976; Scott, 1993; de Jong, 1994), the location of the "perceptual pulse" for these words may lie at some distance from the actual vowel onsets that were measured here. Other factors might also play a role, including the presence of aspiration in *"talk"*but not *"game"*.

If we examine the relationship between the onset beat of *"talk"*, that is the phrase onset, and the metronome pulse, we find that except at the slowest rates, *"talk"* occurs very close to the metronome pulse. Thus, a phase of 0.5 with respect to *"talk"*–*"talk"* is also very close to 0.5 with respect to the metronome pulses.

Altogether, these results show that this subject, with minimal instructions, exhibits a tendency to place a stressed syllable at music-like phase angles. This subject shows a tendency to time the word onsets of this utterance at a *harmonic* of the repetition cycle—in ratio 2:1. Thus the results are quite in accord with the self-entrainment hypothesis for speech. With no instructions to do anything but repeat a phrase in time with a metronome, the subject aligned the two stressed syllables of the phrase *"talk about the game"* as though they were the beats of a two-beat musical measure phase-locked with the metronome.

The next step is to verify this result with other speakers and with a different text fragment.

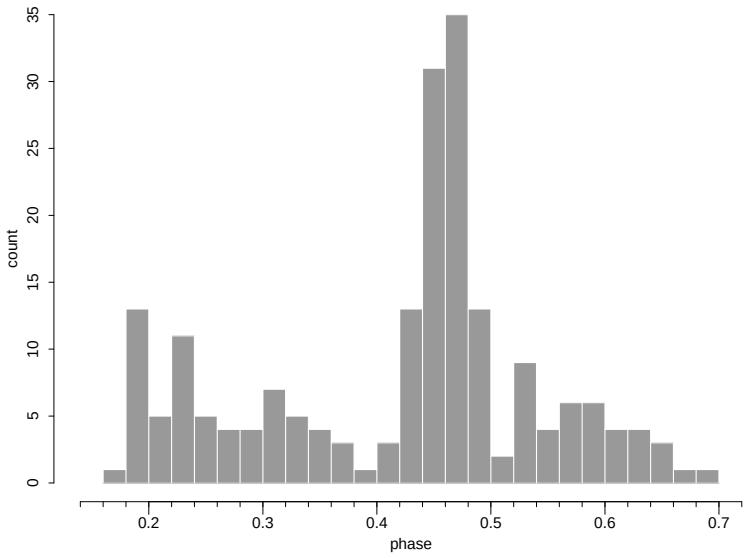

Figure 2. A frequency histogram of the phase of *"game"* onset across both the Increasing and Decreasing rate conditions of Experiment 1. A strong preference for phases just below 0.5 is evident.

3 Experiment 2: Replication with Different Speaker

In this variant of the previous pilot experiment, we employed the the phrase *"beat about the bush"* (similar in prosodic structure to *"talk about the game"*) with a different speaker. Again a male volunteer was given a repetition rate with the metronome signal. The tempos were presented in three orders, Increasing rate, Decreasing rate and Randomly-ordered rates. The subject was told to "pronounce the phrase once for each beep of the metronome." The metronome periods varied over the range from 2 sec to 565 msec and were reduced, in the Increasing rate condition, in 10% decrements for a total of 13 steps. The same stimuli were presented in the opposite order for the Decreasing rate condition, and in random order for the Randomly-ordered condition. The temporal location of the onset of *"bush"* was measured initially in msec and then transformed into a phase angle with respect to the onset of successive productions of *"beat"*. In this case, we employed an automatic "beat extractor" that measured beat location (quite close to vowel onset) by seeking points of maximum energy increase in a rectified, smoothed energy measure over the formant frequency region (250–2500 Hz) of the speech signal (see Cummins, 1995).

The basic results, across the three rate conditions, are shown as a histogram of observed phases for the onset of *"bush"* in Figure 3. This subject showed a preference for placing the onset of *"bush"* at phase of either 0.47 (near 1/2) or at 0.62 (near 2/3). Again, listening as musicians, the tokens with a phase near 0.5 sounded like the word was starting at 1/2 and the ones above 0.6 sounded like they were pronounced to a waltz time. That is, they sound like *"bush"* is occurring on the third beat of a three-beat pattern. The effect of the rate condition (Increasing, Decreasing, Randomly-ordered) was inconclusive in these data, although a hysteresis effect (where the system tends to remain in the same attractor as the control parameter, rate, is changed) has been found in other experiments using variants of this task (Cummins, 1997).

The strongly bimodal distribution of Figure 3 confirms the observation from Experiment 1 that, when speaking along with a periodic signal, these speakers exhibit a preference to locate stress onsets at certain integral fractions of the phrase repetition period. This appears to be true across a wide range of rates and across a variety of experimental details.

This experiment confirms our impressionistic sense that the pattern of rhythms that were observed in Experiment 1 would be exhibited by any English phrase with a similar prosodic structure. That is, the specific words employed are not relevant to the effect. *"Beat about the bush", "take a pack of cards", "slip between*

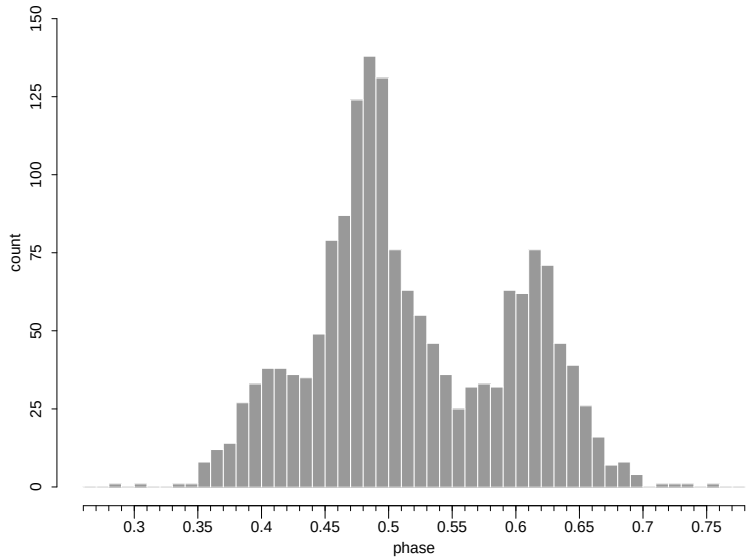

Figure 3. A frequency histogram of the phase of *"bush"* in experiment 2 across the Increasing, Decreasing and Randomly-ordered rate conditions. A strong preference for phases just below 0.5 and just above 0.6 is evident.

the sheets", etc, will all exhibit similar timing preferences, while *"forget to eat your potatoes"* or *"in the morning I'll take you in to work"* cannot be expected to exhibit the same preferences. Impressionistically, we might say that our speakers tended to self-entrain when producing these phrases. Indeed, the entire abstract 2:1 metrical structure seems be almost a "readymade", a system that is fundamentally independe of the language,that seems to attract the onsets of stressed syllables toward phases of, especially, zero or one half.

4 Harmonic Timing Effect

The main results of these simple experiments have also been verified in other experiments in our lab. For example, in another variant of the Speech Cycling task (Cummins, 1995, Cummins and Port, 1997), subjects were asked to repeat the phrase *"take a pack of cards"* while listening to a periodic signal. But in this case, the signal was not a simple metronome beep, but rather a voice in synthetic speech saying both *"take"* and *"cards"*. Thus the phase of the second foot onset was specified as a target phase angle and the repetition rate was kept constant (cf. Yamanishi, Kawato and Suzuki, 1980). The target phase for *cards* ranged from 0.3 to 0.65. Again here, an automatic beat extractor was employed to locate syllable onsets in subjects' productions. Although the target phase angle for *"cards"* varied uniformly over a wide range (this time they were randomly chosen phases across that range), subjects showed a strong tendency to bunch their produced phases for the onset of *"cards"* near certain preferred values. The preferred values were clearly 1/2, 1/3 and 2/3—the three largest harmonic fractions of 1. These and other experiments (see Cummins and Port, 1997) encourage confidence that the harmonic timing effect in speech cycling tasks is highly robust. In fact, the effect is not limited to English speakers, but is also found among Japanese speakers (Tajima, 1997).

Accordingly we conclude that the Harmonic Timing Effect is a remarkably strong set of constraints on the timing of speech under conditions that encourage settling. Such conditions include our speech cycling task, as well as chanting, poetry reading, singing while working together, group recitation, etc. Harmonic timing behavior appears to be an instance of entrainment of speech to a *meter*, that is, to a temporal structure of harmonically related and phase-locked oscillators. Such a structure is essentially a vector field on the state space of the two oscillators that renders phase-locked frequency ratios of 2:1 and 3:1, etc, quite stable

(Abraham and Shaw, 1983; Glass and Mackey, 1988). Apparently 2:1 is more stable than 3:1.

Speech timing at the gross time scale (of the phrase and foot) appears to have some fundamental similarities to other kinds of motor behavior. One of the most striking of these similarities is the tendency to get locked into synchronization with one or another meter.

Our expectation is that related effects will be found in the speech of any natural language (see, for example, Tajima, 1997). It is possible that oscillator entrainment is one of the primary methods for global timing control in complex activity by animals. Self-entrainment may be the most basic and easily observable evidence of this method of timing. It seems likely to us that such properties will turn up in many other aspects of cognition and language.

The next set of experiments offers another replication of the effects but this time using a little more variety of text and several more speakers but with fewer repetitions. In addition to looking at the initial and nuclear stresses, we will look carefully at the timing of *another* prominent syllable, the phrase-medial stressed syllable.

5 Experiment 3A: More Prosodic Patterns

Experiments 3A and 3B replicate and extend the findings of the first two experiments, using a larger sample of speakers and a greater variety of prosodic patterns.

5.1 Methods

In Experiment 3A, four speakers repeated the phrase *"give the dog a bone"*. Since only small hysteresis effects were found across the increasing, decreasing, and random presentation conditions, the tempos were presented only in increasing rate, from slow to fast. The metronome period was set to an initial period of 2200 ms on the first trial, and was shortened by 7% on successive trials. The speakers stopped when the metronome got too fast for them (which usually occurred around 700 ms). On each trial, the speakers listened to the first four metronome beeps, then repeated the phrase along with the beeps for eight times. Rather than telling them to simply repeat the phrase "in time with the beeps", we instructed the speakers to "align the beginning of each repetition with each successive beep". The speakers found this easy to do, and this ensured that the productions were comparable across speakers and repetitions.

From each trial of eight repetitions, the first, second, and last repetitions were omitted; the first two were omitted in order to exclude transient effects. Syllable onsets from the remaining tokens were marked using our semi-automatic beat extractor. On average, each speaker did 16 trials, from slow to fast rate; this yielded a total of roughly 320 tokens for analysis (= 4 speakers $\times$ 16 trials $\times$ 5 repetitions). We then computed the phase angle of the onsets of the syllables *"dog"* and *"bone"* with respect to the onset of successive productions of the phrase-initial syllable *"give"*.

Figure 4 shows a histogram of observed phases of the onset of the word *"bone"*. Data are collapsed across all speakers and rates. The figure shows its largest mode around 0.47 or 0.48, quite close to 1/2. Although there is an apparent mode near 0.3, it is not clear which other modes aside from 1/2 should be taken seriously given the relatively few observations here. Again, listening as musicians, the tokens near the mode at 1/2 sounded as if the word *"bone"* was halfway between successive repetitions of *"give"*; that is, it sounded like it was produced on the second beat of a two-beat measure.

In addition to looking at how stressed syllables are timed relative to the overall repetition cycle, it may also be informative to see how syllables are timed with respect to other syllables *within the same repetition*. For example, we calculated the *phrase-internal phase angle* of the medial stressed syllable *"dog"* relative to the interval between the first and the last syllables (*"give"* and *"bone"*) of each repetition. This should tell us how syllables are timed relative to each other within individual repetitions, and whether or not speakers prefer certain phase relations over others.

Figure 5 is based on the same set of productions as Figure 4, but here, instead of showing the *external phase* of *"bone"*, that is the phase relative to the two adjacent productions of *"give"* (nearly the same as the metronome cycle), the phase of *"dog"* is shown relative to the first and last syllables of the phrase. This figure shows a large mode near 0.5, with very small variance. This indicates that there is a strong preference for speakers to place the medial stressed syllable of this phrase halfway between the first and last stressed

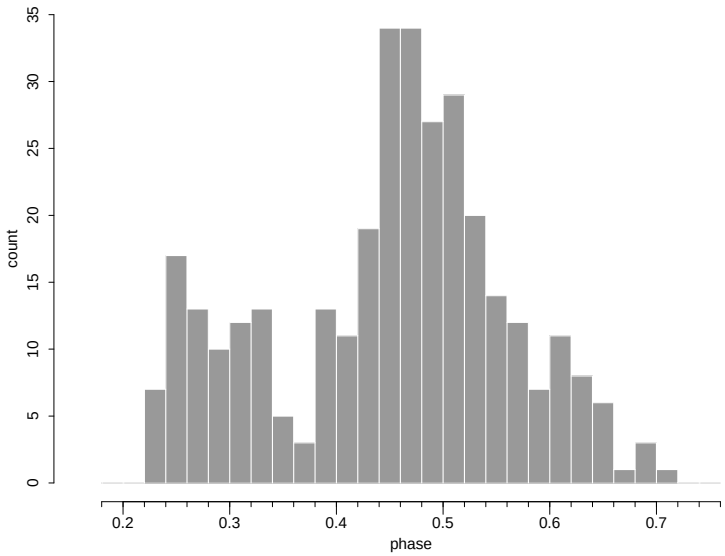

Figure 4. A frequency histogram of the phase of *"bone"* onset across all metronome rates and across the 4 speakers from Experiment 3A. A strong preference for phases just below 0.5 is evident.

syllables. (In fact, this same pattern was also found for *"talk about the game"* and *"beat about the bush"* from Experiments 1 and 2, where the onset of the syllable *"-bout"* was invariably placed halfway between the first and last syllables.)

Why are the syllables so tightly constrained in phase? The prosodic structure of the whole sentence may contribute significantly to the phrase-internal stability of this phrase. The entrainment of the foot within the prosodic phrase (which, in turn is nested within the phrase repetition cycle) may account for small phrase-internal variance. That is, the high consistency of timing may result from self-entrainment among linguistic units at several hierarchical levels, syllables, stress-feet and prosodic phrases. In these sentences, there are two syllables to each of the first two feet and four feet to the prosodic phrase (where the last foot is silent, like a musical rest). All fit into one phrase repetition cycle. The self-entrainment may be enhanced here because the phrase *"give the dog a bone"* exhibits a pattern of uniform alternating stress between strong and weak syllables. In the case of a text fragment that does not have simple alternating stress, stability of timing might be more difficult to obtain.

6 Experiment 3B: More Complex Prosodic Structure

To explore these issues, we turn to Experiment 3B, in which speakers repeated phrases with different patterns of the strong and weak syllables, not only ones that alternate. Experiment 3B is part of a larger investigation on cross-linguistic comparisons of speech rhythm (Tajima, 1997), but here we will look closely at data from two typical English phrases, shown in Table I. These two phrases, *"Betty forgot the bag"* and *"bake the beans in a den"*, each contain three stressed syllables. Unlike *"give the dog a bone"*, the number of intervening unstressed syllables is different. Thus Phrase 1 has the pattern **SwwSwS** (where **S** is strong and **w** is weak), while Phrase 2 has the pattern **SwSwwS**. In terms of traditional metrics, a foot of the form **Sw** is called a *trochee* and **Sww** is called a *dactyl*. We might predict then that, in contrast to uniform trochaic feet, a dactylic foot would tend to throw the speech rhythm off somewhat from the simple alternating, trochaic pattern.

The phrases were repeated by four speakers, some of whom also served in Experiment 3A. The procedure was the same in Experiment 3B as in Experiment 3A, except that the metronome was set to an initial period

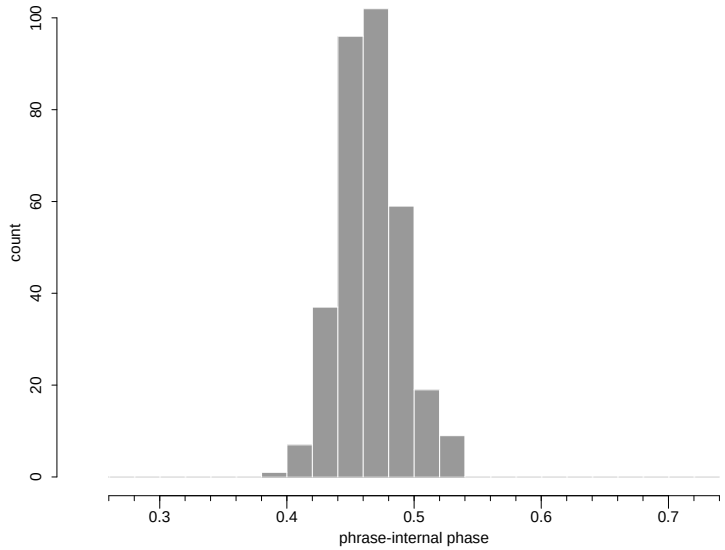

Figure 5. A frequency histogram of the phrase-internal phase of *"dog"*—i.e., the onset of *"dog"* relative to the interval between *"give"* and *"bone"*. A very strong preference for phases just below 0.5 is evident.

		Measurement	
	Phrase	External phase	Internal phase
1.	BETty forGOT the BAG	BAG relative to BET-... BET-	GOT relative to BET-... BAG
2.	BAKE the BEANS in a DEN	DEN relative to BAKE... BAKE	BEANS relative to BAKE... DEN

Table I. Phrases used in Experiment 3B, taken from Tajima (1997). External and internal phases were calculated for each phrase as shown above.

of about 1500 ms—rather than 2200 ms—with a slightly longer period for longer phrases.

If phrases with a perfectly alternating stress pattern are especially stable, then given that the new phrases in Table I deviate from such a pattern, we expect that the internal timing structure of these phrases would not be as invariant as we saw in Experiment 3A. That is, if we use the same analysis technique as for *"give the dog a bone"*, and plot a histogram of the phase angle of the medial stressed syllable relative to the first and last syllables of the phrase, then we might expect to see more spread in the distribution of data points.

Figure 6 shows the result of such an analysis. Each phrase was analyzed separately, but the data are collapsed across all speakers and tempos. The histograms in Figure 6 look qualitatively different from the one in Figure 5 since there appears to be considerably more spread in the distribution than in the previous figure. This suggests that the internal rhythm of the phrase—defined as the relative timing of the three stressed syllables—was, as predicted, less invariant. Instead, depending on the token (and on the speaker), the medial stressed syllable appears to have had a target value at an earlier or later phase. In that sense, phrases in Experiment 3B were less stable than the phrase Experiment 3A, a result consistent with our expectation here. So apparently the medial stressed syllable does not always have its target internal phase at 1/2.

Looking closer we can also draw from Figure 6 the observation that several preferred phases are evident. In the two panels, there appear to be modes near 0.33, 0.41, 0.56 and 0.65. This suggests that the "imperfectly"

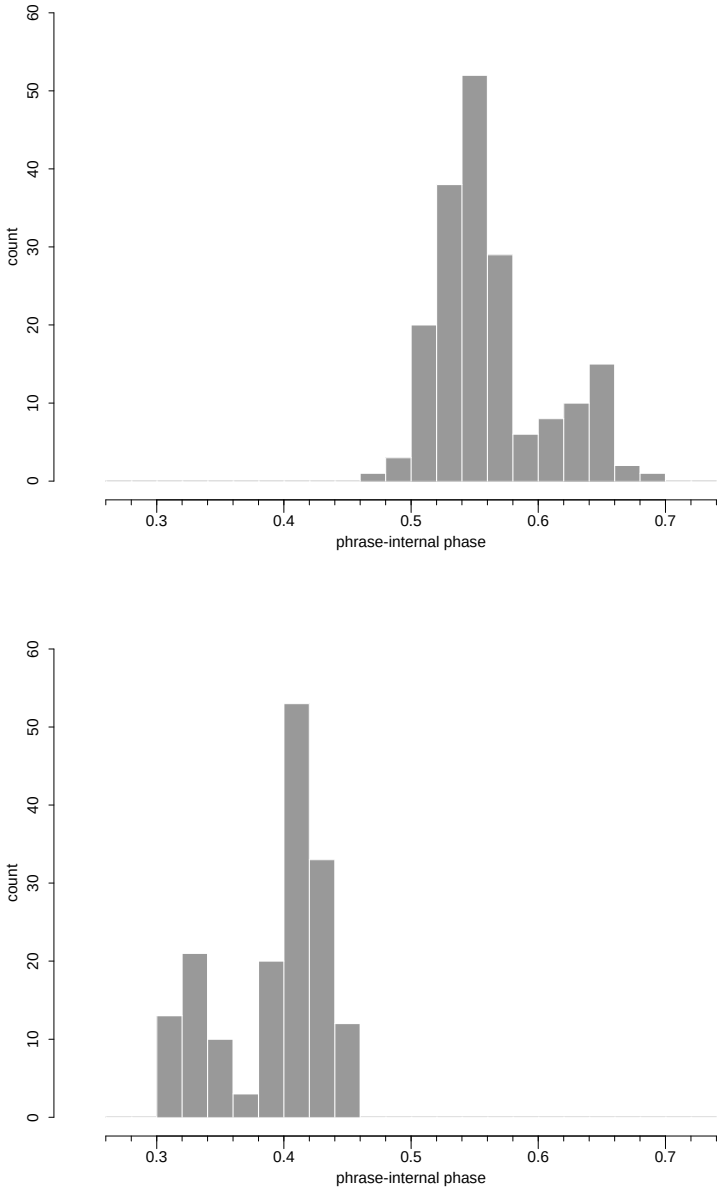

Figure 6. Frequency histogram of the phrase-internal phase of the medial stressed syllable *"-got"* in *"Betty forgot the bag"* (top), and *"beans"* in *"bake the beans in a den"* (bottom). Each panel reveals multiple preferred internal phases for the medial stressed syllable.

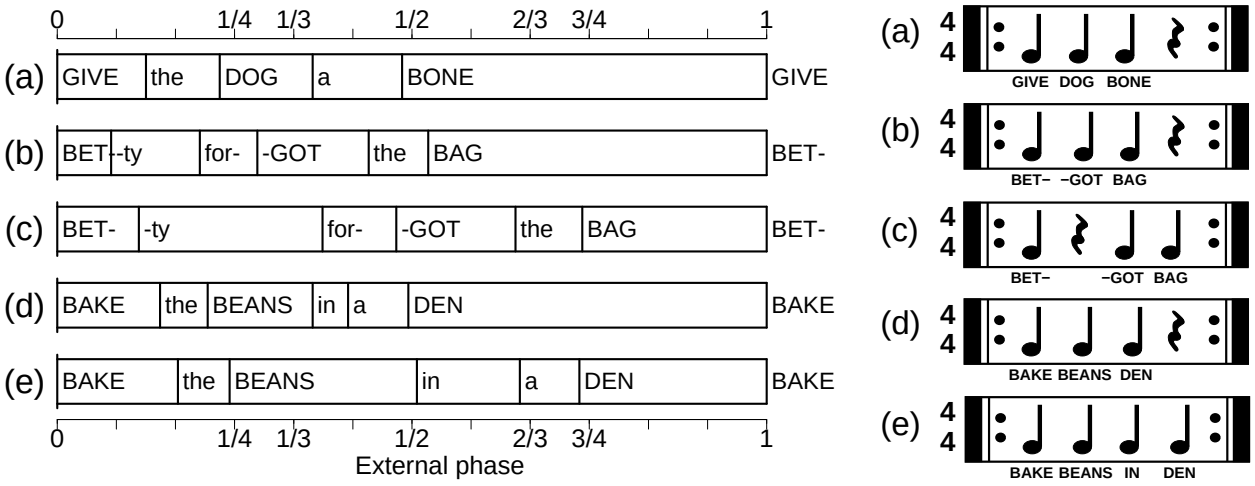

Figure 7. Timing pattern of each qualitatively distinct rhythm found for individual phrases in Experiments 3A (in panel a) and 3B (in panels b–e). The scale at the top and bottom divides phase into 12 equal parts. Syllable onsets were obtained from the beat extraction procedure (Cummins and Port, 1997).

alternating stress pattern of the phrase does not simply result in a weakening of overall stability. Instead, it apparently yields several stable modes, some of which are weaker in their degree of stability than in the unimodal case in the previous experiment. For example, the medial stressed syllable *"-got"* in *"Betty forgot the bag"* (top panel of Figure 6) showed two preferred internal phases: a value slightly above 1/2 and a value near 2/3. This is in contrast to what we observed for *"give the dog a bone"* (Experiment 2), where all four speakers produced the same internal rhythm for the phrase.

When we informally listen to trials whose internal phases are close to each of the modes in Figure 6 (8 repetitions each), we find that the different histogram modes represent qualitatively distinct rhythms. In fact, these qualitative differences illustrated in this figure could be easily described using musical notation (although we acknowledge that musical skills on the part of the three coauthors could bias our perception). Figure 7 illustrates the timing pattern of each mode in Figures 5 and 6. The left half of the figure plots sample productions from each mode. It displays the timing of syllable onsets as determined by our beat extraction procedure. The right half of the figure provides musical notations for the five distinct rhythmic modes.

The topmost phrase *"give the dog a bone"* (Figure 7(a)) is from Experiment 3A. As shown here, the medial stressed syllable *"dog"* is very close to halfway between *"give"* and *"bone"*. Listening as musicians, the three stressed syllables indeed sounded isochronous.

For *Betty forgot the bag*, however, two impressionistically distinct rhythms were identified. In the first pattern (Figure 7(b)), the three stressed syllables (*"Bet-"*, *"-got"*, and *"bag"*) sounded as if they were the first three beats of a four-beat pattern and a rest on four. In the second pattern (Figure 7(c)), the interval between *"Bet-"* and *"-got"* was twice as long as that between *"-got"* and *"bag"*. That is, *"-got"* and *"bag"* were on the third and fourth beats of a four-beat pattern, with an apparent "musical rest" on the second beat. Similarly, for *"bake the beans in a den"*, we also found two distinct rhythms. In the first pattern (Figure 7(d)), the three stressed syllables (*"bake"*, *"beans"* and *"den"*) sounded isochronous (although *"beans"* sometimes sounded like it came a bit early). In the second pattern (Figure 7(e)), the function word *"in"* was produced with an apparent pitch accent (based on auditory impression), and the result sounded as if it had four stressed syllables in an isochronous series (*"bake"*, *"beans"*, *"in"*, and *"den"*).

The timing patterns shown in Figure 7 appear to support the hypothesis that there is entrainment at many levels of description, not only between the metronome itself and the production of stressed syllables. In particular, there seems to be a kind of mutual entrainment between the production of individual syllables and the location of the "major beats" of the repetition cycle. For the patterns seen in Figure 7, we could

assume that each repetition cycle contains four major beats, much as in a 4/4 time signature in music (as suggested by the musical notations in Figure 7). For all five patterns shown, the strong and weak syllables do not seem to be timed arbitrarily. Instead, the syllables are timed so as to respect the major beats of the repetition cycle. Even the "unstable" phrases, *"Betty forgot the bag"* and *"bake the beans in a den"*, did not just show token-to-token continuous variation, but rather show evidence of distinct modal patterns that employ four equally spaced stresses. In *"Betty forgot the bag"* (Figure 7(c)), the isochronous pattern was achieved by inserting a pause (like a musical rest) between the two unstressed syllables *"-ty"* and *"for-"*, while in *"bake the beans in a den"* (Figure 7(e)) this was done by adding a pitch accent on the function word *"in"*.

The multiple distinct rhythms found for each phrase in Figure 7 suggests that there is indeed more inter- and intra-speaker variability in the repetition of phrases like *"Betty forgot the bag"* than in phrases like *"give the dog a bone"*. A pattern that does not alternate between strong and weak appears to be less intuitive and more complex for English speakers to produce. The preference for certain alternating, "eurhythmic" patterns has been pointed out by metrical phonologists as well (Hayes, 1984; Selkirk, 1984). In general, given an alternating phrase like *"give the dog a bone"*, most English speakers have no difficulty finding a comfortable and stable way of repeating it. In fact, as shown in our data from Experiment 3A, speakers found the task so intuitively clear that they did not settle on any other rhythm than the straightforward isochronous rhythm illustrated in Figure 7(a). In contrast, the phrases used in Experiment 3B, which contained two weak elements in a row, deviated from a perfectly alternating pattern, making them less intuitive and less straightforward resulting in more than one solution to the problem of fitting them into the phrase repetition task.

Thus, there is a relationship between stress pattern of a sentence as determined by the choice of lexical items and the temporal stability of the phrase when it is artificially cycled. The trochee-dominated pattern **SwSwS** is more temporally stable than a pattern that alternates dactyls with trochees like **SwwSwS**. This observation may be interpreted as a consequence of the relative ease of finding a production pattern compatible with mutual entrainment between the foot and the syllable. It seems that a 2:1 entrainment ratio between feet and syllables yields more temporally stable productions (given the speech cycling task) for English speakers than a 3:1 ratio does.

The relative difficulty of a 3:1 ratio compared to 2:1 is further corroborated by looking more closely at Figure 7 panels (b) vs. (c) and (d) vs. (e). Comparing the two alternative rhythmic modes, it is clear that both "fixes" to the **Sww** portions of the text convert them into forms that fit a 2:1 (or 4:1, that is, 2:1 within 2:1) harmonic structure.

7 Development of Speech Production

Now we finally address the last issue of this paper. Do these apparent rhythmic constraints on speech production have consequences for language development? We expect they would. That is, we should expect that children too, given a speech repetition task similar to that used above, would exhibit preferences for harmonic timing. And we might expect that English-learning children would find metrical patterns that easily aligned with 2:1, like **SwSwSw** easier or preferable to patterns that require a 3:1 metrical structure, like **SwwSww**.

Although systematic exploration of these issues in the speech of young children has yet to be done, there is some evidence that **Sww** feet are more difficult than **Sw** feet. Recent experiments by LouAnn Gerken (1996) suggest the possibility that such metrical constraints can actually influence performance on a simple five-word speech imitation task. Gerken asked children to produce simple sentences and noted the effect of phonological and metrical details on their ability to perform the task. It turns out that the metrical structure of the sentences had a large effect on performance (cf. also Wijnen, Krikhaar and den Os, 1994).

For example in one experiment, children averaging around 24 months of age played with small plastic models of animals and people with personal names. The experimenter asked the children to imitate saying sentences like *"Bill kicks the pig"* and *"Bill catches the pig"*. The experimenter noted the frequency with which the article *"the"* was correctly repeated. In the case of *"Bill kicks the pig"*, the article was repeated in 84% of the cases, whereas for *"Bill catches the pig"*, the article was correctly repeated only 52% of the time.

The critical difference between the two types of sentences was the prosodic structure. In the first case,

there is a single weak syllable between the two stressed ones, *kick* and *pig*, whereas in the second case, there are two weak syllables. Gerken's data are compatible with the view that 2:1 rhythms are easier than 3:1 rhythms for two-year-old language learners as well as for adults.

Thus it seems that in early stages of language acquisition, the rhythmic structure of phrases can have a large influence on children's ability to imitate them accurately. We would expect that in a children's version of the speech cycling task we would observe preferences similar to those observed for adults; that is, trochaic feet should be easier and less variable than dactylic feet.

8 Conclusions

This essay has touched on a broad range of issues dealing with periodic behavior and speech. But all of this contributes to a single story—one which we admit to be somewhat speculative. The question can be put this way: Why do speakers in the speech cycling task find it so easy to produce these phrases in such a way that the location of especially prominent syllables tends to be harmonic fractions of the phrase repetition cycle?

The account we offer is that this phenomenon, the Harmonic Timing Effect, is based on a kind of entrainment of linguistic "feet" with the phrase repetition cycle. This entrainment appears to resemble, on one hand, very general examples of entrainment found in rather simple physical systems and in animal behavior. And, on the other hand, it illustrates the kind of meter that underlies systems of music in many human cultural traditions. Entrainment is a typical behavior of systems of coupled oscillators. Of course, more specific mathematical models need to be developed to account for how meters can control speech production, but the general class of appropriate models for meter itself is clear. It must be a model involving coupled oscillators.

It is familiar that two oscillators that influence each other will tend to mutually lock their phases and find harmonic-ratio frequencies. It is familiar because such behavior is completely predictable from the mathematics of coupled oscillators and because behavior like this has been observed in many concrete physical instances: cricket chirps, muscle cells excited by periodic electrical pulses, breathing and running, swinging arms and legs, etc. All these cases show evidence of strong attractors at harmonic ratios. We have seen that the same effects can also be observed in human action in the relation between limbs and other body parts, especially during repetitive motions. The term self-entrainment was used to refer to cases of such coupling between parts of a single body.

Meter seems to be a class of universal abstract structures in time. A simple meter is a kind of minimal temporal object. They can be interpreted as particular cases of self-entrainment, where the (minimally) two oscillators are implemented as "cognitive oscillators". They lock each other into a fixed spatiotemporal pattern. The meter 2:1 is the easiest and most stable. Such meters underlie many genres of speech production (chant, poetry, song) as well as many musical systems in cultures of the world. Recent research on children's speech suggests that simple meters also underlie the early productions of 2-year-old language learners. It is likely that work along this line will lead to important insights into language acquisition.

Speech cycling is a task that we have found to be conducive to the entrainment of language to meter. Thus we find our speakers adjusting the timing of speech somehow, so as to line up the onsets of stressed syllables with metrically preferred points in time. So, the meter itself "selects" particular instants in real time and "demands" that important events happen at those points. Our speakers responded by adjusting their timing so that important events—the onsets of stressed syllables—*would* happen then.

It seems to us that the tendency to exhibit this behavior is evidence of the "attractiveness" of meter to speech. Speech is attracted (speaking informally) to it, but still is constrained by the serially ordered speech gestures themselves. The potential space of possible meters and styles for resolving conflicts between the meter and the words is very large. In fact, various members of our species are continually, on a daily basis, developing novel ways of producing speech to meter—of finding new ways to sing and compose poems.

Altogether then, we have found some rather robust effects in the preferred timing of utterances that have several stressed syllables. When a short sentence is repeated, the second stress in the phrase is strongly attracted to the middle of the phrase repetition cycle (i.e., phase = 1/2). A second pair of attractive phase angles are at 1/3 and 2/3. One consequence of the influence of these metrical preferences is that some patterns of strong and weak syllables can be cycled easily and reliably, while others (e.g., those containing a **Sww** foot) are less stable and tend to find variable alignments with the major beats of the cycle.

We are convinced that pursuing the links between self-entrainment, meter, and speech is crucial to understanding many longstanding issues about rhythm and language use, ranging from cross-linguistic variations in rhythm, to the effects of rhythm on children's language development.

References

Abercrombie, D. (1967). *Elements of General Phonetics.* Aldine Publishing Company, Chicago, IL.

Abraham, R. and Shaw, C. (1983). *Dynamics, The Geometry of Behavior, Part 1.* Aerial Press, Santa Cruz, California.

Bernstein, N. (1967). *The Coordination and Regulation of Movements.* Pergamon Press, London.

Boomsliter, P. C. and Creel, W. (1977). The secret springs: Housman's outline on metrical rhythm and language. *Language and Style*, 10(4).

Bramble, D. M. and Carrier, D. R. (1983). Running and breathing in mammals. *Science*, 219:251–256.

Browman, C. P. and Goldstein, L. (1995). Dynamics and articulatory phonology. In Port, R. F. and van Gelder, T., editors, *Mind as Motion: Explorations in the Dynamics of Cognition*, pages 175–193. MIT Press, Cambridge, MA.

Chomsky, N. and Halle, M. (1968). *The Sound Pattern of English.* Harper and Row, New York.

Collier, G. L. and Wright, C. E. (1995). Temporal rescaling of simple and complex ratios in rhythmic tapping. *Journal of Experimental Psychology: Human Perception and Performance*, 21:602–627.

Cummins, F. (1995). Identification of rhythmic forms of speech production. *Journal of the Acoustical Society of America*, 98(5):2894.

Cummins, F. (1997). *Rhythmic Coordination in English Speech: An Experimental Study.* Doctoral dissertation, Indiana University, Bloomington, IN.

Cummins, F. and Port, R. F. (1997). Rhythmic constraints on stress timing in English. Submitted.

Dauer, R. M. (1983). Stress-timing and syllable-timing reanalyzed. *Journal of Phonetics*, 11:51–62.

de Jong, K. J. (1994). The correlation of P-center adjustments with articulatory and acoustic events. *Perception & Psychophysics*, 15.

Delgutte, B. (1996). Auditory neural processing of speech. In Hardcastle, W. J. and Laver, J., editors, *Handbook of Phonetic Sciences.* Blackwell, Oxford.

Diedrich, F. J. and Warren, W. H. (1995). Why change gaits? Dynamics of the walk-run transition. *Journal of Experimental Psychology: Human Perception and Performance*, 21:183–202.

Dorman, M., Raphael, L., and Liberman, A. (1979). Some experiments on the sound of silence in phonetic perception. *Journal of the Acoustical Society of America*, 65:1518–32.

Gerken, L. (1996). Prosodic structure in young children's language production. *Language*, 72:683–712.

Glass, L. and Mackey, M. (1988). *From Clocks to Chaos: The Rhythms of Life.* Princeton University Press, Princeton, NJ.

Haken, H., Kelso, J. A. S., and Bunz, H. (1985). A theoretical model of phase transitions in human hand movement. *Biological Cybernetics*, 51:347–356.

Han, M. S. (1994). Acoustic manifestations of mora timing in Japanese. *Journal of the Acoustical Society of America*, 96:73–82.

Hayes, B. (1984). The phonology of rhythm in English. *Linguistic Inquiry*, 15:33–74.

Hayes, B. (1995). *Metrical Stress Theory: Principles and Case Studies.* The University of Chicago Press, Chicago.

Jakobson, R., Fant, C. G. M., and Halle, M. (1952). *Preliminaries to Speech Analysis.* MIT Press, Cambridge, MA.

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

Kelso, J. A. S., Southard, D., and Goodman, D. (1979). On the nature of human interlimb coordination. *Science*, 203:1029–1031.

Kelso, J. S., Saltzman, E., and Tuller, B. (1986). The dynamical perspective in speech production: Data and theory. *Journal of Phonetics*, 14:29–60.

Kugler, P. N. and Turvey, M. (1987). *Information, Natural Law, and Self-Assembly of Rhythmic Movement.* Erlbaum, Hillsdale, N.J.

Ladefoged, P. (1972). *A Course in Phonetics.* Harcourt Brace Jovanovich, Fort Worth, TX.

Lehiste, I. (1977). Isochrony reconsidered. *Journal of Phonetics*, 5:253–263.

Lehiste, I. (1990). Phonetic investigation of metrical structure in orally produced poetry. *Journal of Phonetics*, 18:123–133.

Lerdahl, F. and Jackendoff, R. (1983). *A Generative Theory of Tonal Music.* MIT Press, Cambridge, MA.

Levelt, W. J. M. (1989). *Speaking: From Intention to Articulation.* MIT Press, Cambridge, MA.

Liberman, M. (1978). Modeling of duration patterns in reiterant speech. In Sankoff, D., editor, *Linguistic Variation: Models and Methods*, pages 127–138. Academic Press, New York.

Liberman, M. and Prince, A. (1977). On stress and linguistic rhythm. *Linguistic Inquiry*, 8:249–336.

Lisker, L. and Abramson, A. (1971). Distinctive features and laryngeal control. *Language*, 44:767–785.

Martin, J. G. (1972). Rhythmic (hierarchical) versus serial structure in speech and other behavior. *Psychological Review*, 79(6):487–509.

Morton, J., Martin, S. M., and Frankish, C. (1976). Perceptual centers (p-centers). *Psychological Review*, 83:405–408.

Nespor, M. and Vogel, I. (1986). *Prosodic Phonology.* Foris, Dordrecht.

Port, R., Cummins, F., and Gasser, M. (1996). A dynamic approach to rhythm in language: Toward a temporal phonology. In *Proceedings of the Chicago Linguistic Society.* University of Chicago.

Port, R., Cummins, F., and McAuley, D. (1995). Naive time, temporal patterns, and human audition. In Port, R. F. and van Gelder, T., editors, *Mind as Motion: Explorations in the Dynamics of Cognition*, pages 339–437. MIT Press, Cambridge, MA.

Port, R. and van Gelder, T., editors (1995). *Mind as motion: Explorations in the dynamics of cognition.* Bradford Books/MIT Press.

Port, R. F. and Dalby, J. (1982). Consonant/vowel ratio as a cue for voicing in English. *Perception & Psychophysics*, 32:141–152.

Port, R. F., Dalby, J., and O'Dell, M. (1987). Evidence for mora timing in Japanese. *Journal of the Acoustical Society of America*, 81:1574–1564.

Schmidt, R. C., Carello, C., and Turvey, . M. T. (1990). Phase transition and critical fluctuations in the visual coordination of rhythmic movements between people. *Journal of Experimental Psychology: Human Perception and Performance*, 16:227–247.

Scott, S. K. (1993). *P-centres in Speech: An Acoustic Analysis.* Doctoral dissertation, University College London.

Selkirk, E. O. (1984). *Phonology and Syntax: The Relation between Sound and Structure.* MIT Press, Cambridge, MA.

Tajima, K. (1997). Cross-linguistic speech rhythm in a phrase repetition task. *Journal of the Acoustical Society of America*, 101:3128.

Treffner, P. J. and Turvey, M. T. (1993). Resonance constraints on rhythmic movement. *Journal of Experimental Psychology: Human Perception and Performance*, 19(6):1221–1237.

Turvey, M. T. (1990). Coordination. *American Psychologist*, 45:938–953.

van Gelder, T. and Port, R. F. (1995). It's about time: An overview of the dynamical approach to cognition. In Port, R. F. and van Gelder, T., editors, *Mind as Motion: Explorations in the Dynamics of Cognition.* MIT Press, Cambridge, MA.

Wijnen, F., Krikhaar, E., and Den Os, E. (1994). The (non)realization of unstressed elements in children's utterances: Evidence for a rhythmic constraint. *Journal of Child Language*, 21:59–83.

Yamanishi, J., Kawato, M., and Suzuki, R. (1980). Two coupled oscillators as a model for the coordinated finger tapping by both hands. *Biological Cybernetics*, 37:219–225.