Available online at www.sciencedirect.com

SCIENCE DIRECT®

Journal of Phonetics 31 (2003) 599–611

Journal of Phonetics

www.elsevier.com/locate/phonetics

Meter and speech

Robert F. Port

Departments of Linguistics and Cognitive Science, Indiana University, Bloomington, IN, USA

Received 28 November 2002; received in revised form 18 July 2003; accepted 1 August 2003

Abstract

Speech is easily produced with regular periodic patterns—as if spoken to a metronome. If we ask what it is that is periodically spaced, the answer is a perceptual 'beat' that occurs near the onset of vowels (especially stressed ones). Surprisingly, when periodically produced speech is studied it exhibits attractors at harmonic fractions (especially halves and thirds) of the basic periodicity. It is shown that the Haken–Kelso–Bunz model provides conceptual tools to account for the frequency histogram of acoustic beats in the speech. Why might there be attractors at periodically spaced phase angles? It is hypothesized that there are neural oscillations producing a pulse on every cycle, and that these pulses act as attractors for the beats at the onsets of syllables. Presumably these periodic time locations are generated by the same physiological mechanism as the periodic attentional pulse studied for some years by Jones (Psychol. Rev. 96 (1989) 459; Psychol. Rev. 106 (1999) 119). We propose that neurocognitive oscillators produce periodic pulses that apparently do several things: (1) they attract perceptual attention; (2) they influence the motor system (e.g., when producing speech) by biasing motor timing so that perceptually salient events line up in time close to the neurocognitive pulses. The consequent pattern of integer-ratio timings in music and speech is called *meter*. Speakers can control the degree to which they allow these metrical vector fields to constrain their timing.

1. Introduction

As Peter Jusczyk observed, children learn prosodic structure very early in the language acquisition process (Jusczyk, 1997; Mehler et al., 1988). Other data show that prebabbling infants notice deviations from regular timing of perceptual centers or P-centers (which are temporally close to vowel onsets, Fowler, Smith, & Tassinary, 1986). These observations suggest that children may be able to both produce and perceive simple periodic patterns in speech well before producing their first words. Frequently speakers produce speech in a periodic way, sometimes by coupling their speech production to another speaker or to a metronomic pattern, e.g., when chanting or declaiming. This essay reviews some relevant phenomena and proposes general

0095-4470/$ - see front matter © 2003 Elsevier Ltd. All rights reserved.
doi:10.1016/j.wocn.2003.08.001

theoretical mechanisms to account for these behaviors. Since the mechanisms are very simple, we might expect them to appear fairly early in the development of speech.

We will review some experimental observations on the kind of event that most often recurs periodically and some properties of periodic speech, then sketch some basic ideas about how periodic patterns in speech might arise.

1.1. Finding a beat in speech

One of the most important discoveries about periodicity in speech has been known for some time although its importance for global aspects of speech timing may have been underestimated. George Allen (1972, 1975) showed that if English speakers are asked to align a finger tap with a word, they will line it up close to the onset of the vowel in the stressed syllable of the word. This implies that there is a perceptually salient acoustic event at these time points in speech. Subsequent research on 'perceptual centers' or 'P-centers' was able to refine the notion of the 'beat' associated with prominent syllables by showing that large initial clusters tend to move the P-center temporally to the left (into the consonant cluster, e.g., in *skate* vs. *ate*) while final clusters can move the P-center somewhat to the right (into the vowel, as in *baa* vs. *banks*, Morton, Marcus, & Frankish, 1976; Marcus, 1981; Pompino-Marschall, 1989). These perturbations, however, are small (5–15 ms) relative to the repetition cycle (typically about 500 ms). Apparently, the beat location can be approximated automatically (Scott, 1993) by measuring the amount of energy in lower frequencies (between 200 and 800 Hz), smoothing sufficiently and then looking for large energy onsets which are prominently encoded in the auditory nerve (Delgutte & Kiang, 1984). When speakers attempt to produce a series of regular events with their speech, these observations about beats and pulses imply they will regularize the spacing of vowel onsets, especially stressed vowels (at least for English). This clarifies the question of what it is that is periodic, and thus what 'periodically produced speech' might mean. And since other aspects of the signal play only a fairly small role, these findings encourage the use of automatic measurement methods that simulate the beat locating aspects of auditory performance.

1.2. The harmonic timing effect

Aside from the P-center work, there has been other research on simple periodic speech phenomena. The case of subjects cyclically repeating a short phrase has been shown to lead to the *harmonic timing effect*. A number of studies have shown that when speakers repeat a short piece of text many times, they exhibit a strong preference for locating prominent (e.g., stressed) syllable onsets at simple harmonic fractions of the repetition cycle (Port, Tajima, & Cummins, 1996; Cummins & Port, 1998; Tajima, 1998). For example, Cummins and Port (1998) presented English-speaking subjects with a two-tone metronome pattern. Tone A marked the beginning of each cycle and alternated with tone B that was randomly located at phase angles between 0.20 and 0.80 of the A–A cycle. The subjects' task was to repeat a phrase like '*Dig for a duck*' so that the first stressed word lines up with tone A and the final stress lines up with tone B.[1] The location of

---

[1] The interval from A to B was fixed and only the repetition period (A–A) varied, giving the speakers a constant amount of time for pronunciation of the text to partially equalize articulatory difficulty across conditions.

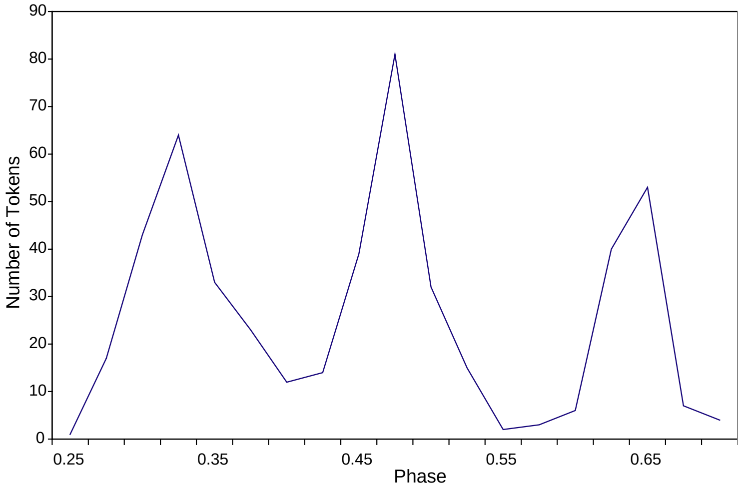

Fig. 1. Final-syllable onset distribution in terms of phase angle, pooled across 8 speakers (from Cummins & Port, 1998). Subjects were asked to read a 4-syllable phrase like *Dig for a duck* in time with alternating metronome tones, placing the first syllable on the first tone and the last syllable on the second one. The phase lag between the first and second tone varied in a random uniform distribution from 0.20 to 0.80. The total number of tokens in the histogram is 8640 in bins of 1/40 of the phase circle of which these 19 had tokens.

onset of the final syllable, *duck*, was measured as a particular phase angle between 0 and 1 (the beginning of the next cycle). The frequency histogram of performance for all speakers is shown in Fig. 1. Although the target phase angles for the onset of the final syllable were distributed uniformly over the interval from 0.20 to 0.80 of the repetition cycle, the speakers did not reproduce anything resembling the flat input distribution but were strongly biased to locate their onsets near just 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 most target phases later than about 0.57. This bias is called the *harmonic timing effect* because locations like 1/2 and 1/3 would be the phase-zero pulses for (phase-locked) harmonic frequencies of the fundamental. Similar results have now been observed in a number of experiments and the phenomenon can easily be demonstrated to oneself (by repeating a 4–6 syllable phrase and noticing where the phrase-final stress occurs when the pattern is stably repeated).

Although this experiment employed only English speakers, one might expect that other languages should at least have a bias to pay special attention to vowel onsets and to favor low-frequency harmonics whenever nested meters are constructed. There is some data directly comparing English and Japanese in a similar task (Tajima & Port, 2003). The speakers of the two languages adjusted to perturbing influences on timing in language-specific ways, but the data clearly showed that speakers of Japanese were paying attention to the vowel onsets in this task, just as much as the English speakers.

Notice that the speech results demonstrate not merely regularity at the frequency of the metronome, but also at higher frequencies. There is periodicity on two time scales: one at the repetition cycle rate and another either 2 or 3 times faster than the metronome but phase-locked to it (so the phase-zero pulse is actually two simultaneous pulses). What kind of cognitive mechanism could account for these particular constraints on speech timing is the primary issue we are concerned with here.

These experiments show that when there is periodicity at one level, there may sometimes be periodicity at a harmonic of that frequency. This feature of motor temporal behavior is not restricted to speech, but can be observed in simple limb movements as well.

2. Non-speech periodic behavior and the HKB model

It may be appropriate to compare the harmonic timing phenomenon to oscillatory finger motion as illustrated in Kelso's finger-wagging task. Kelso (1984) had subjects oscillate one finger on each hand to the left and right. When the phase relationship of the fingers is such that they simultaneously move toward and away from the midline (described as 0 phase), performance is easiest. Most phase relationships between the fingers are very unstable although, at a slow enough tempo, the fingers can simultaneously move left and then to the right (called 180° or anti-phase) with stability. Haken, Kelso and colleagues developed a generally well-supported and very simple model for the interaction of the two fingers (Haken, Kelso, & Bunz, 1985; Kelso, Scholz, & Schöner, 1986; Schöner, Haken, & Kelso, 1986).

It is possible that the same model, known as the HKB model, might describe the production of syllable onsets governed by our auditory metronome sequence. One reason to expect this is possible is the results of Tuller and Kelso (1990) who had subjects pronounce monosyllables like [pi] and [ip] to a rate-controlled metronome. They reasoned that a CV syllable would begin at zero phase while a VC syllable would begin half way through the cycle and their data supported the prediction. To the degree that producing periodic speech with an auditory metronome is like visual coupling to the wagging of a limb, we should expect HKB to describe the phenomena. In fact, Schmidt, Carello, and Turvey (1990) showed that if two people each swing one of their legs sitting on the edge of a table, they find it much easier to swing in phase than out of phase. These results support many specific predictions of the HKB model while demonstrating crossmodal coupling using vision.

It may be that rhythmically produced speech is partly analogous to finger wagging. Both are periodic and have similar frequency ranges. Finger wagging differs, however, in one important respect. The reciprocal motion of the finger (or other limb) means that there are two unique points of motionlessness in each complete cycle, exactly a half-cycle apart. When subjects repeat a short piece of text, however, it is the first stressed syllable of the text that defines the beginning of the repetition cycle. It will depend on the text and on the speaking style where other salient perceptual events (like the stressed syllable onset in an accented word) might occur.

2.1. The HKB model

This model is based on various inanimate physical systems (Schöner et al., 1986; Scholz, Kelso, & Schöner, 1987), proposing the simplest mathematical function that accommodates qualitative features of the data. A dynamical system can be characterized as a *vector field* on a *state space*. Here the state space is the relative phase of two oscillators (the two fingers or limbs) where $\varphi_R = 0$ is taken to be the fingers moving simultaneously toward the midline and then away, and $\varphi_R = 0.5$

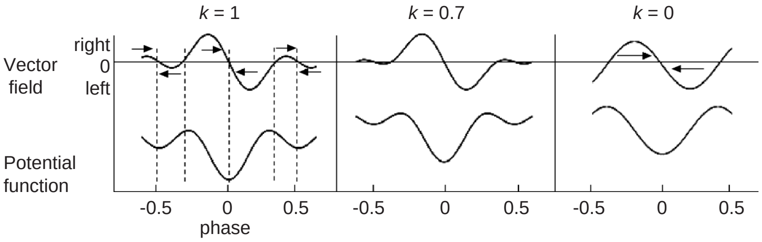

Fig. 2. Vector field and attractor landscapes for the HKB model for 3 decreasing values of k from left to right. The upper panels show the vector fields where positive values mean motion to the right around the phase circle (with the rate of motion proportional to amplitude) and negative values mean motion to the left. Thus, zero crossings that are positive to the left and negative to the right identify stable fixed points while zeros that have a negative vector field to the left and positive to the right are unstable fixed points. The lower panels show corresponding potential functions, V, the integral of the vector field. Here valley minima represent stable fixed points and dome peaks are unstable ones. Thus in the left pair of panels, where $k = 1$, there is a stable attractor at $\varphi_R = 0$ and another attractor, less attractive, at $\varphi_R = +0.5$ (repeated at $\varphi_R = -0.5$). The middle panel shows a smaller value of k and a correspondingly weaker attractor at one half, while the right panel shows $k = 0$ where there is a repellor at $\varphi_R = \pm 0.5$.

is when both fingers move left and both move right. Then a vector field is postulated along the circular axis of φ_R,[2] the relative phase, showing for each value of φ_R the direction and rate of phase change, as shown in the upper left panel of Fig. 2 where $k = 1$. The equation for the vector field is

$$\mathrm{d}\varphi_R/\mathrm{d}t = \sin\varphi_R - k\sin 2\varphi_R.$$

Places where the vector field is 0 are fixed points, but the vector field representation does not differentiate the stable fixed points from the unstable ones very well. So the integral of the vector field, the potential function, V, is useful since stable fixed points appear as valleys and unstable ones as domes. The corresponding potential function is

$$V(\varphi_R) = -\cos\varphi_R - k\cos 2\varphi_R$$

shown in the lower curves of Fig. 2. The $\cos 2\varphi$ term produces an attractor at 0.5 (as well as at 0) while the $\cos\varphi$ produces an attractor only at 0. The parameter k affects the amplitude of the $\cos 2\varphi$ term relative to $\cos\varphi$.

$V(\varphi_R)$ gives a set of attractor landscapes for different values of the control parameter k. In a potential function display, the attractive states are in the wells where any deviation results in return to the center of the well, while repellor states, values of relative phase that are unstable, are the peaks. The sign of the slope specifies the direction of phase change and with the rate proportional to the absolute slope of the potential function: steeper slopes mean faster change of relative phase for a given perturbation. When k is large enough, there is a potential well at $\varphi_R = 0.5$ as well as one at 0, but when k is small the well at 0.5 disappears and the only stable phase angle is at 0 phase. The two attractors correspond to phase-locking at anti-phase (± 0.5)

---

[2] Relative phase can be looked at either as a line or as a circle depending on one's purposes. One oscillator can slip by any amount, a half cycle, a whole cycle or n cycles relative to another—thus relative phase seems to be a line. But at any moment, looking at the system, 1.5 cycles locates you in the same position as 0.5 cycles—thus it is also like a circle.

and in-phase ($\pm 1 = 0$). But a small k turns the system with 2 stable modes into a system with one stable phase relationship.

2.2. HKB predictions

As suggested above, this model makes several testable predictions when the control parameter (cycling rate) is increased. First, there is *slipping to in-phase*. If the rate is slowly increased from slow to fast, the parameter k will get smaller. When the target phase is 0, the attractor there remains strong. However, if the target phase is 0.5, then subjects should eventually be forced to slip over to the in-phase relationship as the attractor disappears. The cycle rate at which the attractor at 0.5 becomes unstable, called the critical point, will vary with the subject and with practice at the task. (Of course, if the rate is too fast, then even 0 will fail as an attractor and the model no longer applies). Second, there is *critical slowing down*. Because the slopes of the attractor basin become shallower at faster rates near the critical point, the model predicts slower movement of instantaneous phase back to the basin center. So given a single perturbation (such as mechanical hindering of motor performance or an entraining stimulus event that is too early or too late) when operating near the critical point, the recovery of phase back to the preperturbation phase should take longer than it does at slower rates.

Finally, there should be *critical fluctuations*. Assuming there is some amount of intrinsic noise, the weakening stability of the anti-phase attractor should be reflected in an increase of fluctuations of relative phase as the rate approaches the critical point. The phase should 'wobble' more in this situation even if the internal noise is constant. All of these predictions were demonstrated for variants of the finger-wagging task (Haken et al., 1985; Kelso et al., 1986; Kelso & Jeka, 1992). Analogous predictions might apply to speech.

Some forms of rhythmic speech exhibit behavior that resembles in fundamental ways the oscillatory limb-motion tasks to which the HKB model has been applied. The oscillatory limb motions may provide insight into the task of perceptual-motor coupling, despite the fact that the finger-wagging task alone might suggest dependence on the mechanical properties of limbs. But this behavior is just as easily observed in situations where only perceptual coupling can provide the physiological basis for the phenomena (cf. Schmidt et al., 1990). Together these results suggest that an extension of this model might be appropriate for several kinds of rhythmic behavior, including the speech cycling task (Cummins & Port, 1998; Tajima & Port, 2003; Port et al., in preparation). The model would then be viewed as an archetypal structure that can influence (or couple) many systems. It coordinates motor events with perceptual events.

The system involved in these perceptuo-motor tasks must involve coupling between perception and motor behavior as well as between the internal oscillators (at f and $2f$) for metrical patterns. Coupling is essential in order for different events to be constrained to occur simultaneously. It implies ability to adjust instantaneous phase and/or oscillation rate. A number of computational mechanisms have been explored to accomplish this varying in their degree of neural plausibility (Large, 1994; McAuley, 1995; Eck, 2002).

The proposal here is that the HKB-type attractor layout in relative time may be, in fact, a generic metrical structure that can be exploited by many motor and perceptuo-motor tasks. In particular, it is suggested that it reflects the activity of two internal, coupled oscillators at a frequency ratio of 1:2. Because of the coupling, they are constrained to fire when in phase with each other when possible. And a phase-zero pulse by either oscillator is an attractor for any point in a gesture where a perceptually salient event occurs. The mechanism described here is probably the same as the periodic pulses of attention studied by Jones (Jones & Boltz, 1989; Large & Jones, 1999).

What do we know about these oscillators? First of all, they must adapt within a few cycles to change of rate when stimulated. All the periodic phenomena discussed so far are rate-invariant. The basic behaviors look the same across many rates of oscillation (which is why the invariants are defined in relative phase rather than milliseconds). Second, it appears they are very responsive to perceptually salient events, such as acoustic onsets and visual extrema. Third, the range of rates involved seems to represent the range of motor oscillation frequencies of bodily movements from about 4–5 finger taps/s to the 4–5 s period of, e.g., swaying the trunk at the hip.

3. Meter and periodicity

Although linguists and students of poetics often describe meter in terms of serial patterns of strong and weak syllables, the most intuitively natural notion of meter seems to be that of music where it is based on periodic structures in continuous time.[3] The simplest possible meter has a single periodicity but musical meters often employ several oscillators constrained to integer-ratio frequencies and constrained to be phase-locked. A graphic representation of several levels of meter is shown in Fig. 3.

If a speaker reads a list of words or numbers, or repeats a short word, there may be periodicity on only one level. This can be modeled with a single oscillator assuming that its pulse is an attractor for auditory beats (that is, for P-centers). Then speech timing is warped in a way that permits the stressed (or pitch accented) vowel onsets to settle close to certain of the periodic series of time points (McAuley & Kidd, 1998). This is described as 'Level 1' meter in Fig. 3 (and in Large, Fink, & Kelso, 2002). The other meters in Fig. 3 are common in Western music as well as in other musical traditions.

3.1. What is meter?

The account of harmonic timing proposed here depends on several coupled oscillators. These metrical cognitive structures have also been described by Large and Jones (Large & Jones, 1999; Large, 2000; Large & Palmer, 2002; Large et al., 2002). The state space of two oscillators is a torus. Because of tight coupling between the oscillators, the practical state space is constrained to certain paths around the torus—those in which one oscillator cycles either 2 or 3 times while the other cycles just once. Each oscillator cycles such that every time the function reaches phase 1 (or zero), it exhibits a distinctive state characterized by both greater perceptual attention and also those events in motor gestures that create perceptual salience. If there is an external periodic input, then the phase and frequency of the next cycle of the internal system will adapt to the input. In cases where there is a nested periodicity, the faster oscillator couples its phase zeros with the pulse of the slower oscillator and a more complex nested structure is created with 2 or more

---

[3] Liberman (1978) clearly differentiated music-like meter based on ratio-scale time from meter based on serial-order, the time scale assumed within linguistic theories. The first is naturally notated with musical notation or rational numbers while the other can be notated with strings of symbol tokens like phonemes, words or *strong* and *weak*.

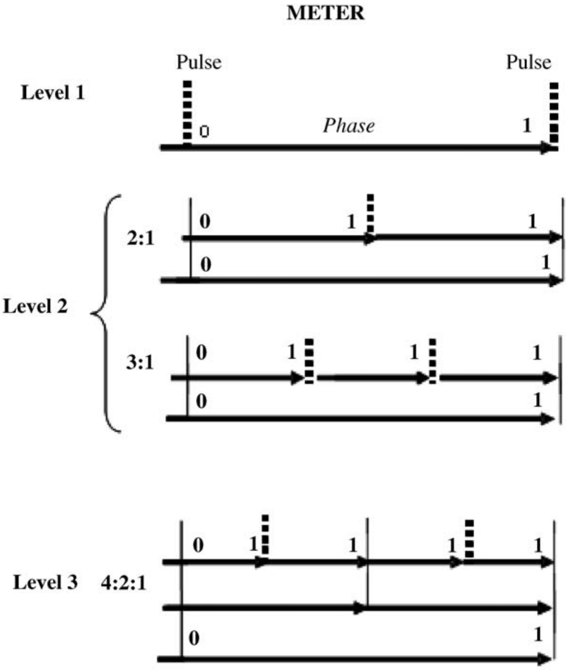

Fig. 3. Some basic types of meter. Level 1 has a single periodic pulse. Level 2 has two nested oscillations, typically in the ratio 2:1 or 3:1. Level 3 nesting has 3 simultaneous oscillations. The solid vertical lines show phase-locking relationships and dashed lines are phase zeros that are not phase-locked to others. Level 2 meters could, in principle, be $n:1$ but n rarely exceeds 5.

coupled oscillators shown in the lower panels of Fig. 3. The oscillators create a pattern of time locations toward which certain speech gestures are attracted—the ones that correspond to perceptually salient events.

Thus, for example, when performing the speech cycling task employed in the Cummins and Port (1998) study, the metronome (and the oscillatory response to it) produced an attractor for the initial syllable of the phrase (e.g., *dig*). The harmonic oscillator at $2f$ or $3f$ produced the target temporal locations for the final stressed syllable (e.g., *duck*). Now, we might ask whether these temporal approximations of the randomly located target tones in that experiment were due to 'rounding off' by perceptual oscillators (so that the pulses of the complex metronome employed were heard as occurring at thirds or the half) or whether they were perceived more accurately than this but the oscillations employed for controlling the timing of speech production were doing the rounding. Unfortunately, the experiments conducted so far cannot answer this question. It seems likely, however, that the effect occurred initially during the perception process.

A story along this line seems to depend on two hypotheses to account for the basic meters found in harmonic timing. The first hypothesis is:

Hypothesis 1. *Musical meter and the harmonic timing effect reflect a system of phase- and frequency-coupled* ***oscillators that generate pulses****—discrete periodic events.*

These oscillators tend to oscillate at low integer-ratio frequencies like 1:2 and 1:3 as shown in the upper panel of Fig. 4. Here the pulses are schematized as rectangular pulses and the slower pulse series is suggested to be stronger than the faster series. Large (2000) develops further mathematical description of such systems.

We also need a second hypothesis regarding the significance of the pulse for motor control. One might at first imagine that this pulse would be the moment of initiation of a movement, i.e., the moment when it is 'triggered'. Actually, what is attracted to the pulse is the most perceptually salient event, such as the tapping sound of a finger or onset of a vowel (not, say, the onset of the mouth opening gesture). Movement onsets must therefore precede the pulse by an appropriate amount.

Hypothesis 2. ***Phase zero*** *of any of these oscillations* ***attracts perceptually prominent motor 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.*

Hypothesis 2 accounts for why vowel onsets (especially stressed ones in the case of English) or finger taps tend to occur near the pulse of some oscillator, and why gestures like moving a finger in a circle present no clear line-up point with a metronomic pattern like music.

Given a system of oscillators like this, we propose representing a multi-oscillator meter as a potential function using the phase of the slowest oscillator (that is, the phrase repetition cycle) as a pattern-internal time scale. When the system has two oscillators at frequencies 1 and 2, the potential function should have attractors at both $\varphi = 0$ and 0.5 (just like the potential function of Haken et al., 1985; 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 minima at both 0 and 0.5, as shown in Fig. 4A.

Since our speech experiments also show evidence of oscillations at the frequency ratio 3:1, we extend the HKB model by postulating an alternative potential function with minima at $\varphi = 0$, 0.33 and 0.67 at each location where the harmonic at $3f$ passes through phase 0 on the assumption that the two oscillations are phase-locked, as shown in Fig. 4B. Notice that Haken, Kelso and

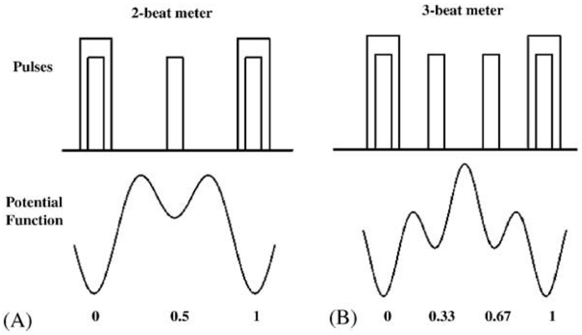

Fig. 4. Potential functions over phase suitable for (A) a 2-beat meter and (B) a 3-beat meter. Each is shown with pulses that are proposed to generate the potential functions. The troughs in potential are attractor locations and the peaks are repellors. The first curve is given by $V(x) = -\cos\varphi(x) - k\cos\varphi(2x)$, where $k = 1$, the same as Haken et al. (1985). The second potential function is given by $V(x) = -\cos\varphi(x) - k\cos\varphi(3x)$ where $k = 1$.

Bunz found no evidence of attractors here in their finger-wagging task. This may result from the fact that the finger-wagging gesture is a *reciprocal* motion that exhibits a moment of zero velocity at both ends of the gesture separated by visual blur. This property may produce a bias for a harmonic at $2f$ (and possibly $4f$, cf. Smethurst & Carson, 2001) and a bias against a harmonic at $3f$. The speech task employed here can be executed so as to place a distinctive event at either one half or at one-third or two-thirds. It seems that other attractors could also be observed with appropriate training or with speakers of other languages.

3.2. Some predictions: experimental and developmental

The general model proposed here makes many predictions that can be explored in future research. There are some things we expect to be true of these sometimes complex oscillatory structures.

First, since the theory specifies attractors in terms of phase angle, we expect that at least for moderate changes in rate, the attractors should be unaffected in terms of phase but rather change location in direct proportion with cycle duration. This is obviously true of music and has also been confirmed for speech cycling tasks, for example, in the Cummins and Port study (by varying the repetition cycle duration over a range of over 10%) and in other previous experiments (Port et al., 1996, in preparation).

Second, the attractors may vary in 'strength', that is, in the degree of attractiveness of the fixed point. These differences should be measurable in the effects of perturbation on events near the attractor. That is, given a periodic perturbation of the system, any effect should be less pronounced when the attractor is stronger (that is, when its potential well is deeper or has steeper slopes).

Third, phase zero should always have the strongest attractor and the addition of harmonics should only strengthen the attractor at zero. Further, the attractors created by harmonics of the repetition cycle (at integer fractions of the longest cycle) should get weaker at larger multiples of the first harmonic. The reason for this may be related to the decreasing amplitude of the harmonics of a plucked string as the frequency increases. Thus, an attractor at 1/2 may be expected to be less stable than the attractor at 0 and 1/3 less stable than 1/2, and so on. (This prediction is partially confirmed by Port et al. (in preparation), by showing that an attractor at 1/3 results in tighter clustering of stop release times than the same text pronounced with the stop location at 1/4.)

Fourth, the emphasis here on the role of the perceptual consequences of speech or other gestures makes many concrete predictions. For example, experimental manipulation over the time lag before the acoustic effect of a finger motion or a speech sound should lead to reorganization of gesture timing. Similarly, gestures that lack a prominent perceptual point-like effect (like moving a finger in a circle) should align more poorly with a metronome.

Finally, given the simplicity of some metrical cognitive structures, we should expect that some aspects of these behaviors will turn out to be acquired early in life. And some of those acquired early should tend to be universal, rather than characteristic of the prosodic grammar of just a particular language. The universality issue has been initially engaged by Tajima, Zawaydeh, and Kitahara (1999) and Tajima and Port (2003). The latter set up similar tasks for native speakers of Japanese and English to look for similarity of response to the manipulated perturbing factors.

Their evidence suggests that, despite major differences in the temporal structure of English and Japanese, at least the simplest possible nested meter, 2:1, seems fundamental in both languages. (The English speakers acted more as though they were attempting to regularize the inter-stress intervals while the Japanese speakers showed a greater tendency to regularize the spacing of all vowel onsets, cf. Abercrombie, 1967; Pike, 1943.)

Returning to the issue of developmental aspects of speech rhythm, there is much evidence that infants (Bertoncini, 1993; Mehler, Dupoux, Nazzi, & Dehaene-Lambertz, 1996) and early speakers are quite sensitive to rhythmic aspects of speech prosody, at least, if by 'rhythm' we mean patterns of strong and weak syllables (since appropriate measures of timing have not been made). For example, English-speaking 2-year-olds can imitate strong–weak feet (i.e., trochaic words like *paper*) better than iambs (e.g., *between*) or dactyls (e.g., *elephant*) (Gerken, 1996). If a trochaic *Sw* word can be identified with a '2-beat' foot and a dactylic *Sww* word can be identified with a '3-beat foot', then the greater simplicity of 2-beat vs. 3-beat pattern could account for Gerken's observation that 2-year-olds who are asked to repeat spoken sentences like *Bill kicks the pig* and *Bill catches the pig* are much more likely to drop the article *the* in the second sentence (52% vs. 84% respectively, Gerken, 1996). The second one requires production of a 3-beat foot which may be prosodically more challenging for 2-year-olds.

Of course, such tasks are still not very close to the temporally defined periodicity focused upon in this paper, but the speech cycling task seems like it should be relatively easy for children. Certainly cyclic repetition of phrases by young speakers seems to occur often quite spontaneously in prespeech and early linguistic vocalizations. What is still required is for temporal data to be collected from language learners to see (a) how early clear periodic timing can be observed in their production, (b) at what age language-specific metrical differences appear, and (c) whether variants of the speech cycling task will turn out to be useful as diagnostics of other speech-production difficulties.

4. Conclusion

The issue explored in this essay is that, in many situations, speakers will exhibit periodic location of salient events like vowel onsets. For English, this is especially true of syllables with pitch accent or stress. It is proposed that this periodic behavior reflects periodic attractors in relative phase that are generated by one or more internal oscillators producing pulses that are sometimes coupled to external periodicities. These oscillations can be described as neurocognitive because they represent major neural patterns somewhere in the brain and are cognitive because they may be time-locked to events across many different modalities—audition, cyclic attention, speech, limb motion, etc.—in order to solve a wide range of problems in complex motor coordination and in the prediction of environmental events. The metrical structure of music, poetry and chant is proposed to reflect essentially the same oscillatory system.

The coupling of oscillators at integer-ratio frequencies accounts for the preference for stressed syllables being located at harmonic fractions of a repetition cycle, since, by hypothesis, the phase zero of any of these oscillators attracts vowel onsets and other aspects of attention as well, as demonstrated in many experiments by Mari Jones (e.g., Jones & Boltz, 1989). The negative cosines of the HKB potential functions represent these attractors. The present author is not in a position to speculate on what may be oscillating in the nervous system corresponding to these functions.

One important issue is whether the theory proposed here is of relevance for spontaneous ordinary speech or if it is relevant only for specialized periodic or rhythmical speech such as is employed in chant or song or in repetitive laboratory speech tasks. My hunch is that these attractors remain relevant in all speech although their effects are most clearly seen in simple repetitive speech. This expectation is like expecting that the mechanical resonance of a limb will partially account for the statistical distribution of behaviors one observes in the limb.

Finally, 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—a system that can easily be set into periodic oscillation in a partly language-specific style. This system will create attractors in time for salient events. We suspect these temporal constraints, probably acquired early in life, provide a framework on which to hang smaller phonological units like syllables and segments.

Acknowledgements

Thanks to Adam Leary for help with figures and to Sarah Hawkins and Noel Nguyen for helpful comments on earlier drafts.

References

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

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

Allen, G. (1975). Speech rhythm: Its relation to performance universals and articulatory timing. *Journal of Phonetics*, *3*, 75–86.

Bertoncini, J. (1993). Infants perception of speech units: Primary representational capacities. In S. de Boysson-Bardie, P. de Schoenen, P. Jusczyk, J. MacNeilage, & J. Morton (Eds.), *Developmental neurocognition: Speech and face processing in the first year of life*. Dordrecht: Kluwer.

Cummins, F., & Port, R. (1998). Rhythmic constraints on speech timing. *Journal of Phonetics*, *26*, 145–171.

Delgutte, B., & Kiang, N. (1984). Speech coding in the auditory nerve: I. Vowel-like sounds. *Journal of Acoustical Society of America*, *75*, 866–878.

Eck, D. (2002). Finding downbeats with a relaxation oscillator. *Psychological Research*, *66*, 18–25.

Fowler, C. A., Smith, M. R., & Tassinary, L. G. (1986). Perception of syllable timing by prebabbling infants. *Journal of the Acoustical Society of America*, *79*, 814–825.

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

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

Jones, M. R., & Boltz, M. (1989,). Dynamic attending and responses to time. *Psychological Review*, *96*, 459–491.

Jusczyk, P. (1997). *The discovery of spoken language*. Cambridge, MA: MIT Press.

Kelso, J. A. S. (1984). Phase transitions and critical behavior in human bimanual coordination. *American Journal of Physiology*, *15*, R1000–R1004.

Kelso, S. (1995). *Dynamic patterns: The self-organization of brain and behavior*. Cambridge, MA: MIT Press.

Kelso, J. A. S., & Jeka, J. J. (1992). Symmetry breaking dynamics of human multilimb coordination. *Journal of Experimental Psychology: Human Perception and Performance*, *18*, 645–668.

Kelso, J. A. S., Scholz, J. P., & Schöner, G. (1986). Non-equilibrium phase transitions in coordinated biological motion: Critical fluctuations. *Physics Letters A*, *118*, 279–284.

Large, E. W. (1994). *Dynamical representation of musical structure*. Unpublished dissertation, Computer and Information Science Department, Ohio State University.

Large, E. W. (2000). On synchronizing movements to music. *Human Movement Science*, *19*, 527–566.

Large, E., Fink, P., & Kelso, J. A. S. (2002). Tracking simple and complex sequences. *Psychological Research*, *66*, 3–17.

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

Large, E. W., & Palmer, C. (2002). Perceiving temporal regularity in music. *Cognitive Science*, *26*, 1–37.

Liberman, M. Y. (1978). *The intonational system of English*. Bloomington, IN: Indiana University Linguistics Club.

Marcus, S. (1981). Acoustic determinants of perceptual center (P-center) location. *Perception and Psychophysics*, *30*, 247–256.

McAuley, D. (1995). *Perception of time as phase: Toward an adaptive-oscillator model of rhythmic pattern processing*. Unpublished Ph.D. dissertation, Departments of Computer Science and Cognitive Science, Indiana University.

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

Mehler, J., Dupoux, E., Nazzi, T., & Dehaene-Lambertz, G. (1996). Coping with linguistic diversity: the infant's point of view. In J. L. Morgan, & K. Demuth (Eds.), *Signal to syntax*. Mahwah, NJ: Erlbaum.

Mehler, J., Jusczyk, P. W., Lambertz, G., Halsted, H., Bertoncini, J., & Amiel-Tison, C. (1988). A precursor of language acquisition in young infants. *Cognition*, *29*, 144–178.

Morton, J., Marcus, S., & Frankish, C. (1976). Perceptual centers (P-centers). *Psychological Review*, *83*, 405–408.

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

Pompino-Marschall, B. (1989). On the psychoacoustic nature of the P-center phenomenon. *Journal of Phonetics*, *17*, 175–192.

Port, R. F., de Jong, K., Kitahara, M., Collins, D., Leary, A., & Burleson, D. Temporal attractors in rhythmic speech (in preparation).

Port, R. F., Tajima, K., & Cummins, F. (1996). Self-entrainment in animal behavior and human speech. *Online proceedings of the 1996 midwest artificial intelligence and cognitive science conference*, Indiana University, Bloomington, IN. http://www.cs.indiana.edu/event/maics96/proceedings.html

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

Scholz, J. P., Kelso, J. A. S., & Schöner, G. (1987). Non-equilibrium phase transitions in coordinated biological motion: Critical slowing down and switching time. *Physics Letters A*, *123*, 390–394.

Schöner, G., Haken, H., & Kelso, J. A. S. (1986). A stochastic theory of phase transitions in human hand movement. *Biological Cybernetics*, *53*, 442–452.

Scott, S. K. (1993). *P-centers in speech: An acoustic analysis*. Ph.D. thesis, University College London.

Smethurst, C., & Carson, R. (2001). The acquisition of movement skills: Practice enhances the dynamic stability of bimanual coordination. *Human Movement Science*, *20*, 499–529.

Tajima, K. (1998). *Speech rhythm in English and Japanese: Experiments in speech cycling*. Ph.D. thesis, Linguistics Department, Indiana University.

Tajima, K., & Port, R. F. (2003). Speech rhythm in English and Japanese. In J. Local, R. Ogden, & R. Temple (Eds.), *Phonetic interpretation: Papers in laboratory phonology VI* (pp. 317–334). Cambridge, UK: Cambridge University Press.

Tajima, K., Zawaydeh, B., & Kitahara, M. (1999). A comparative study of speech rhythm in Arabic, English and Japanese. In: J. J. Ohala, Y. Hasegawa, M. Ohala, D. Granville, A. C. Bailey (Eds.), *Fourteenth international congress of phonetic sciences*, San Francisco (pp. 285–288).

Tuller, B., & Kelso, S. (1990). Phase transitions in speech production and their perceptual consequences. In M. Jeannerod (Ed.), *Attention and performance VIII* (pp. 429–451). London: Academic Press.