

Rhythm as an Affordance for the Entrainment of Movement

Fred Cummins
School of Computer Science and Informatics
University College Dublin
Belfield
Dublin 4
Ireland

tel: +353 1 716 2902

fax: +353 1 269 7262

email: fred.cummins@ucd.ie

Short title: Rhythm as Affordance

Abstract

A general account of rhythm in human behaviour is provided, according to which rhythm inheres in the affordance that a signal provides for the entrainment of movement on the part of a perceiver. This generic account is supported by an explication of the central concepts of affordance and entrainment. When viewed in this light, rhythm appears as the correct explanandum to account for coordinated behavior in a wide variety of situations, including such core senses as dance and the production of music. Speech may appear to be only marginally rhythmical under such an account, but several experimental studies reveal that speech, too, has the potential to entrain movement.

1. Introduction

It is mid-afternoon at one of the many Summer music festivals. Amid the mud and tents, the fire jugglers and circus acts, is a large marquee, inside which a hundred people are dancing silently. There is ambient noise, but it bears no relation to the movement of the dancers. This is a silent disco, in which all participants wear wireless headphones. Each dancer is fully immersed in the performance, bouncing and gyrating rhythmically, partaking in a hi-tech manner in an activity that is as old as human culture. They dance happily in time with the music, but the spectator is left entirely out of the loop. It is an engaging idea, and by all accounts quite fun, both to watch and to take part in. But something is odd. Their movements do not match. No sense of an integrated whole arises, as one would normally expect from the movements of dancers on a conventional dance floor. The silent disco often adds this further twist: two DJs play tunes at the same time, typically with quite different tempi, and the participants can listen to either of the channels. Now it becomes clear that there are two entirely distinct group activities going on in parallel. There are, in fact, two separate collective organizations manifest here, thoroughly intertwined spatially, but distinguished by the organizing forces of two different rhythmic sources. Rhythm acts as the glue to organize each of the two distinct groups into a coordinated unity. Each person belongs to one or other of the two groups. Widely dispersed dancers listening to the same channel can be seen to be mutually coordinated, while spatially adjacent dancers listening to different channels maintain no stable relationship. In this example, we see the power of rhythm to constrain and coordinate movement, creating a whole out of disparate, otherwise autonomous, parts.

This paper seeks to re-situate the study of rhythm as it applies to speech. The account presented here is intended to be continuous with our understanding of rhythm in diverse forms of activity, both animal and human. Rhythm, it will be argued, is not a property of a signal. It can not be found simply by looking at an acoustic wave, a visual stimulus, or a set of numbers. Rather, rhythm will be viewed as an affordance for movement, allowing the coordination of action with a stimulus. This approach to rhythm demands that we take embodiment seriously, as there can be no effective rhythm in the absence of the potential for physical movement. It will further seek to characterize the function of rhythm as a coordinative mechanism, capable of generating stability in the temporal organization of the behavior of multiple independent entities. With this perspective, speech rhythm appears as a special case, continuous with all forms of rhythmic expression, but with some quite distinctive characteristics.

This account of rhythm will not satisfy some in the field. Many phenomena that are conventionally interpreted as 'rhythmic' by phoneticians, and especially phonologists, will not fit into the mould being crafted here. Descriptions of metrical structures that distinguish languages and language types do not typically refer to movement, or even individual people. For some such accounts, the term is of relatively slight importance. It is perfectly possible to describe structural properties of sequences of syllables without couching the account in terms of rhythm. I believe it would be better to do so. As speech scientists and linguists, we have an obligation, rarely acknowledged, to ensure that findings from our field are interpretable by scholars from disparate backgrounds. To this end, it is desirable that terms in common currency, with clear sets of associations, are not subverted by technical, narrowly constrained usage. It is often necessary to co-opt a common concept and apply it within a specialist field, but that professional act ought to be done with one eye over one's shoulder, ensuring that insights from detailed study are once more communicable back to a wider scientific readership.

The history of the study of speech rhythm is laid out in detail elsewhere in this issue, and need not be repeated here. It is sufficient to note that the phonetic treatment of rhythm has changed in focus of late. Sketching that history very crudely, we might observe that a long-lasting focus on isochrony as a means of characterizing rhythm (Lehiste, 1977), and, through that, of establishing a rhythmic typology across languages (Dauer, 1983), has, of late, given way to a focus on the statistical properties of the signal as expressed in a variety of metrics (Ramus et al., 1999; Grabe and Low, 2002). These metrics, it is hoped, may do what the search for isochrony never did, which is provide

a way of capturing impressionistic differences in the perceived rhythmic characteristics of many languages. On the view to be adopted here, rhythm does not inhere in any language, but in the act of speaking. The present proposal is thus very much in the spirit of previous attempts to relate rhythm in speaking to rhythm in other aspects of human behaviour (Allen, 1975; Martin, 1972; Lashley, 1951).

The discipline of phonetics is traditionally described as having three related but separate areas of study: the production of speech, the speech signal itself, and the perception of speech. All of the recent innovations in characterizing rhythm in speech have in common that they derive their indices (nPVI, %V, etc) from the speech signal (Grabe and Low, 2002; Ramus et al., 1999; Galves et al., 2002; Gibbon and Gut, 2001). They are thus accounts of signal structure. What the traditional description of phonetics misses entirely is the bald fact that speech is a purposive behavior, linking and coordinating the behavior and mutual understanding of two or more people. It is to this coordinative function of speech that this account will turn. This will ultimately require the consideration of rhythm within all three traditional domains: from production, through the signal, to perception and further action.

The introductory description of a silent disco points to the fundamental role of rhythm in organizing movement. Rhythm allows that coordination and it facilitates it. When a group of people dance to a DJ's tune, the rhythm is not in the acoustic signal. It is not in the movement of the people. It is that which links the two. This demands a strong definition:

Rhythm is an affordance for the entrainment of movement.

In order to tease out the implications of this claim, it will be necessary to introduce and clarify the notions of both affordance and entrainment. These are not familiar terms to many phoneticians, and each may be read with a shallow or a more profound sense. An attempt will be made to distinguish both readings below. In discussing the core definition of rhythm, no reference to speech is necessary. Rhythm is a ubiquitous phenomenon found throughout nature, and central to many human activities. A principled account of rhythm must do useful work in a variety of domains, including movement *per se*, dance, and music. Once that work has been done, we can then turn back to speech, and see how rhythm in speech appears in light of this stance. If this account is to be of real use, the resulting understanding of rhythm in speech ought to do service both to phoneticians

studying the phenomenon, and to those who seek to understand speech rhythm from other disciplines. In this way, the phonetic study of rhythm may become continuous with current accounts of human behaviour across the board.

2. Affordances

The term 'affordance' is conventionally traced to James J. Gibson, and his programmatic approach to perception and action, *Ecological Psychology* (Gibson, 1979). The notion appears simple at its core, and yet upon closer examination, it has the potential to reveal a radically altered view of the relation between an organism and its environment. Those interested in recent philosophical discussion about the notion of affordance are referred to a special issue of *Ecological Psychology* (15(2), 2003). For present purposes, we will adopt the definition provided by Chemero (2003), according to which affordances are "relations between the abilities of organisms and features of the environment". This seems fairly straightforward, and can be illustrated with a simple example. A cup is graspable. Grasping is one of the things that I, as an actor, can do with it. Graspability is not to be considered an inherent property of the cup, however. To an ant, or one without hands, or one with hands much larger or smaller than mine, the cup is not graspable. Its graspability lies in the relation between the cup and the action possibilities it affords me.

Affordance, in this sense, has been enthusiastically taken up within the design community (Norman, 1999). However, the concept can be misapplied egregiously if care is not taken. A proximate cause of error is to think that the affordance is just one among many properties of the cup. My cup is hard, porcelain, decorated, and graspable. This is fundamentally at odds with the recognition that an affordance is a relation between an object and the capacity for action of a specific organism.

So far so good. We can begin to understand the claim that rhythm is an affordance by recognizing that we are talking about the relation between the capacity for movement of a specific person and the structure (temporal or otherwise) of a specific signal. The second part of the claim, that the affordance is for the entrainment of movement will be discussed in the next section.

The notion of affordance brings with it some profound implications, if it is taken seriously and integrated within a naturalistic view of organisms. To illustrate this, the example of a single cell

swimming within a nutrient medium may serve well. For a lone cell in a medium, movement is based on chemotaxis. A gradient in the concentration of a chemical is expressed at the cellular membrane, thereby providing information specific to the constitution of the cell. A toxin will elicit movement down the gradient. Nutrients elicit movement upwards. In this yoking of perception and action, the cell is exploiting ‘information’ which is informative *only by virtue of the cell’s own constitution*. Chemicals are only toxic or nutritious by virtue of the properties of an organism. One aspect of the environment is thus differentiated by the cell in terms predicated upon its own constitution *qua* cell.

In the course of evolution, perception/action systems evolved. With each change along the way, some information *of relevance to the organism* is selectively recognized. Of the many, essentially infinite, number of discriminations that might be made in an environment, perception/action systems necessarily pick out just those discriminations that are pertinent to the functioning of the organism. These functions, of course, are specifically those that support the stable persistence of an organism with its organizational closure, and the reproduction thereof (Maturana and Varela, 1987). A startling metaphysical implication of this recognition is that the entire phenomenological world an organism encounters is drawn in terms of affordances that exist, or have existed somewhere its entire phylogenetic lineage. The world an ant encounters is fundamentally incommensurable with the world encountered by a human (Nagel, 2004). Metaphysics is not the order of the day here, of course. But this serves to underscore the important claim being made here that rhythm is not something that arrives on the stage of life along with the cultural vehicles of music or speech. Rather, as an affordance, it is continuous with the total field of action possibilities of an organism in its environment.

A nice illustration of the reality of the affordance concept and its implications for perception/action relations mediated by nervous systems is provided in a recent neurophysiological study (Lin et al., 2007). It has been discovered that certain cells in mouse hippocampus become more active when the information in the environment is such that the mouse could, if it desired, make a nest there. Now the potential for building a nest is a very high level property of an environment, and it would be impossible to say which specific and dissociable properties of a visual/haptic/olfactory scene distinguish one which affords nesting from one which does not, but the mouse’s nervous system and body are fine-tuned to just those properties of the environment that are of importance to the mouse. (Unfortunately, a wonderful finding like this in neuroscience is interpreted as demonstrating that the

mouse possesses some kind of abstract, conceptual, encoding. That interpretation misses the direct nature of the fit between animal and environmental information completely.)

So much, then, for the notion of affordance. The second substantial part of the definition of rhythm provided above is specific about the affordance that rhythm provides. Rhythm allows the entrainment of movement. To substantiate that claim, entrainment needs some explication, and again a relatively unproblematic interpretation will be provided, along with a deeper account that carries profound implications.

3. Entrainment

Christian Huygens, polymath and inventor of the pendulum clock, is reputed to have made the first systematic observation of the mutual entrainment of two independent systems (Bennett, 2002). Recovering at home from an illness, he noticed that two pendulum clocks mounted in a single housing case inevitably generated non-independent pendular motions. Specifically, the two pendula consistently maintained an anti-phase relationship, whereby one pendulum would reach the mid point of its cycle, just as the other was initiating its own cycle. The relationship was found to be insensitive to minor perturbation, such as introducing a transient delay in one pendulum. After perturbation, the pendula rapidly settled back into a regular anti-phase relation, or, as Huygens described it, an “odd sympathy”.

Entrainment may be characterized as the yoking together of two oscillatory systems such that their periods of oscillation become related. For the pendula in the above example, the two clocks have identical endogenous frequencies (eigenfrequencies), and the result of entrainment is the alignment of their phases. This is thus pure phase entrainment. For such simple interacting systems, in-phase (zero phase difference) and anti-phase (one half cycle difference) are typically stable forms of organization to which all other initial configurations will tend over time. They thus represent attractor states of the combined system. Phase is a circular variable, and several conventions exist for describing it. Here, a convention of describing phase as ranging from 0 to 1 will be adopted. In-phase and anti-phase relations thus represent relative phase differences of 0 and 0.5, respectively. For a concise introduction to the basics of dynamical systems theory on which this account is based, the reader is referred to Norton (1995).

Two periodic systems may find stable coordination in regimes other than the the 1:1 frequency ratio illustrated by the clock pendula. The motion of the planet Mercury represents the interaction of two distinct oscillations: revolution and rotation. Unlike the Earth's moon, where these two are entrained with identical periods, Mercury exhibits a stable relationship between the two of 2:3. Depending on the difference in eigenfrequencies of two systems and the strength of the coupling force between them, many other possible attractors exist. Those which are expressed by the ratio of large integers are less stable than those composed of simple ratios, with a ratio of 1:1 being the most stable.

Mathematically, two interacting systems may be described as autonomous dynamical systems yoked by a coupling term. For example, a well studied model of bimanual coordination developed by Kelso and colleagues captures the dependency relations obtaining between two hands when they are constrained to oscillate at the same frequency (Haken et al., 1985; Kelso, 1995). Within this model (hereafter, the HKB model), each hand is modelled by a specific form of self-driven oscillator, and the form of the coupling between the oscillators is expressed by distinct terms in the equations describing the combined system. Figure 1 (left panel) illustrates the composite system of two oscillating hands. In each equation, the terms describing the endogenous dynamic of the hand are gathered on the left of the equality, while the coupling terms are collected on the right. The specific functions used can be found in Haken et al., (1985). In this simple system, the relationship is modeled as perfectly symmetrical, which is a simplification addressed in later versions (Carson et al., 1994).

[INSERT FIG 1 ABOUT HERE]

Coupling introduces a constraint on the behaviour of each hand. The motion of one hand is no longer completely independent of the phase and velocity of the other. This very important characteristic of coupled systems generally has as a consequence that the resulting composite system is effectively of a lower dimension than the aggregate of the components. Equivalently, the resulting system can be described with fewer bits of information (Collier and Burch, 2000). This is illustrated in the right half of Fig 1, in which a potential function, $V(\phi)$ describes the stability of the relative phase (i.e. the phase difference between the two hands) of the combined system. The troughs in the potential function correspond to stable attractors, and if the system is primed to start at some other relative phase, the potential function describes its subsequent evolution towards one

or other attractor. One of the strengths of the HKB model is that the simpler potential function describing the stability properties of the system as a whole can be derived directly from the equations that describe each oscillator, together with an appropriate form for the coupling terms.

There are many aspects to this system that are of great interest, such as the dependence of the shape of the potential function on the rate of oscillation, and the concomitant changes in the attractor layout, but for the present purposes, it serves mainly to illustrate that the coupling of two systems generates a composite that has its own, simpler, structure. The reduction in complexity that is attendant on entrainment is the source of much of the order observed in nature (Collier and Burch, 1998). It is aptly illustrated by the obvious distinction between the two groups of dancers described in the opening of this article. Each group is an autonomous system, constituted by entrainment between dancers and music, and dancers and other dancers.

But the world is not a clockwork toy. Most motion we engage in is not strictly periodic. The study of speech rhythm, in particular, has had great difficulty in bridging the gap between models that assume simple periodicities, and the complex and sometimes confusing patterns of articulator motions and speech signal properties that result. Several attempts have been made to develop models of speech rhythm that assume underlying oscillators (O'Dell and Nieminen, 1999; Barbosa, 2002). Even allowing rapid coupling among these oscillators, it is hard to capture the aperiodicity of real speech.

Entrainment may be of more use than it is sometimes given credit for. The above examples of clocks and hands all describe long-term, stable forms of coordination. Mathematically, these are only accurate descriptions of physical systems if the system is allowed to evolve from some starting condition and then to settle within the basin of one attractor or another. Before the state of the system lies on an attractor, its motion is described as a transient. But even during the transient, the motions of two entraining systems are no longer independent. As long as some form of coupling is possible between two systems, the resulting composite system will be more constrained than the two constituents considered separately. There will be an attendant reduction in the effective dimensionality of the system as a whole. This coupling need not even be continuous. A point wise injection of energy at some phase of one system into another may result in stable coordination across the systems, as seen for example in the pushing of a child on a swing. The coordination made possible by entrainment is by no means restricted to the asymptotic behaviour of perfect oscillators.

Some further cases in which entrainment leads to a measure of coordination among systems might usefully be mentioned here. Weakly coupled periodic systems, or coupled systems with a great difference in eigenfrequencies may exhibit *relative coordination*, as first described by von Holst (Kelso, 1995). In relative coordination, there is a tension between the intrinsic dynamic structure of each of the two systems and the coupling force that links them. As these compete with each other, there will be a tendency to adopt specific phase and frequency relations, and that tendency will wax and wane. Thus there will be a greater statistical likelihood of specific forms of coordination, without the finality and simplicity of complete entrainment.

Secondly, periodic attractors are not the only stable form of asymptotic behavior in second order dynamical systems. Chaotic attractors can result from the interaction of even relatively simple system variables, producing stable behavior which is nonetheless unpredictable over non-negligible time spans. Yet two chaotic attractors, if coupled, may give rise to harmonic oscillation in their combined evolution (Collier and Burch, 1998).

Finally, lest it be thought that entrainment is only about pendula in lockstep, even a weak coupling between dynamical systems can, under certain circumstances, produce significant changes in the response of one of the two systems if the internal dynamics of the system are appropriate. This has been proposed as the mechanism by which smells are differentiated in rabbit olfaction, whereby the unstimulated system lies on a complex chaotic orbit, but with just a very small perturbation, initiated by the detection of one or other pheromone, the entire system is driven into one or other harmonic orbit of the many quasi-periodic orbits that collectively constitute the chaotic attractor (Skarda and Freeman, 1987). For such a small coupling factor to have such a large influence, the dynamics of the nervous system of the rabbit must, of course, be exquisitely tuned to the informationally-relevant characteristics of the olfactory environment. This kind of tuning appears to be precisely what evolution is good at, and brings us back to the notion of an effective fit between the organization and capabilities of an organism and its attendant discriminatory and action possibilities in an environment.

4. Rhythm (not in speech)

Putting these pieces together, rhythm appears as the glue that underlies the synchronization of players in an orchestra, that forms whole groups of coherent motion out of disparate disco dancers, and that harnesses any number of disparate autonomous individuals into a self-organized collective that admits of a simpler description than the mere collective motion of the individuals. Examples of rhythmic organization abound in nature. Male fiddler crabs will attempt to woo a female by waving their over-large appendage (Backwell et al., 1998). When a group of males vie for the attention of a single female, each is aware of the motion of the neighbouring suitors. Amusingly, the organization that results is a synchronous waving of all males, in which no one individual can be distinguished by their individual waving. Another well known example is the synchronization of the flashing of fireflies in certain South-East Asian swamps (Buck, 1988). Both of these are limiting cases, in which the regularity of motion is sustained, such that a stable collective pattern can form as the system settles onto a periodic attractor.

But relative coordination, and the constantly changing influence of weak coupling may be seen in the swarming behavior of birds and bees, in the group coherence of buffalo herds and the shoaling of fish. Here there is no sustained periodicity, but there is still an obvious organizing effect that arises from the combination of endogenous motion and an exogenous coupling factor derived from the perception of the movement of nearby con-specifics (Deutsch, 1996). Entrainment is evident, though there is no sustained periodicity.

A final mention might be made of rather obvious entrainment of an animal to a periodic signal of human origin. A recent study by Patel has convincingly demonstrated that a cockatoo can synchronize its movements with the underlying beat of a piece of music (Patel et al., 2008), and track the beat as tempo changes.

All these examples beg the question of whether speech might be considered to be rhythmic at all by the definition adopted here. If we take the affordance for the entrainment of movement to be central to the concept of rhythm, is speech rhythmic at all?

5. Rhythm in speech

At first blush, it may appear that the constrained definition of rhythm provided here excludes speech almost totally, and that rhythm so considered is simply not relevant to the production or

perception of speech. There are a few cases where speech is used to bring about the overt entrainment of movement, but they are few and far between. When a crowd of protesters chant a repeated slogan and pump the air with their fists, we see a clear role for speech in organizing the collective motion of the individuals. On a smaller scale, entrainment within a single person may be seen when a passionate rhetorician pounds upon the table in emphasis, an example of self-entrainment (Port et al., 1996). Speech is also simply part of the fabric of music in rapping, scat singing, chanting and similar. Again these are clear examples of the role of speech in facilitating the entrainment of movement, but they are all special cases, and in some cases, are only to be considered ‘speech’ by a liberal definition. They all exhibit manifest periodicities, or recurrent invariant intervals, maintained over several seconds or longer.

5.1 Speech cycling

An experimental vehicle was introduced in Cummins and Port (1998) that was designed to explore the similarities in coordinative principles underlying speech production and periodic, regular oscillation of the limbs. In speech cycling, a subject repeats a short phrase, such as “big for a duck” in time with an external periodic marker or metronome. As the phrase is repeated, it has been observed that the internal timing of each phrase becomes coordinated with the overall phrase repetition cycle. In particular, the stressed syllable onset (of “duck” in this example) is found to lie at $1/3$, $1/2$ or $2/3$ of the containing phrasal cycle. Intermediate phases of this stressed onset can not be maintained with any stability. The harmonic relationship of the stress foot and the phrasal cycle indicate a simple entrainment among oscillatory processes at asymptote. This is of a kind with the 1:1 entrainment found in the oscillating hand paradigm of Kelso and co-workers (Kelso, 1995), or the 1:n stability found in other tapping tasks (Summers et al., 1993).

[Insert Figure 2 about here]

The speech cycling task is an avowedly artificial, constrained speaking situation, bearing little resemblance to the ebb and flow of conversation. As was noted above, the notion of entrainment is not restricted to asymptotic behaviour of strictly periodic processes, but is also evident in the reciprocal influence any distinct autonomous systems may have on each other through a coupling means. This can be seen in speech, as illustrated by another experimental laboratory task, that of speaking in synchrony with a co-speaker.

5.2: Synchronous speech

Choral speaking in large groups is a familiar activity found in prayer, chanting, and the recitation of oaths, among others. In all these cases, the resulting speech tends to have a highly stylized, over-practiced prosody, quite unlike a fresh, spontaneous rendition of the underlying text. In the synchronous speech task, a new text is given to two or more participants. After familiarizing themselves with the text, the subjects are given a start signal, with the instruction to read the text in close synchrony with their co-speakers (Cummins, 2002). A remarkable fact about this task is that almost all subjects can adhere to the task instructions immediately, and the resulting parallel speech channels are found to be highly synchronized. Median asynchronies of about 40 ms. are typical (Cummins, 2003), although this rises to about 60 ms. at phrase onsets. Practice at the task does not typically generate any improvement, unless the text is kept constant, suggesting that the performance of subjects depends on the coordinative strategies and abilities employed when speaking normally. The resulting synchronous speech does not have the exaggerated prosodic form typical of choral speaking.

In this setting, the speech of each speaker is an external stimulus that serves to modulate the endogenously produced speech of the other speaker(s). One way of interpreting this performance is to regard each speaker as an autonomous dynamic system, and to view the task as imposing a coupling relation between them, yielding sustained mutual entrainment. One support for this analytical approach comes from the fact that pairs of speakers rarely, if ever, exhibit a leader-follower relationship. In over 80 dyads examined to date, no single clear case of asymmetry has been observed. Further support for this approach is garnered from the consistency of the synchrony relation. An alternative account couched in terms of monitoring and error correction seems simply inappropriate in accounting for sustained synchrony.

If the performance of two synchronous speakers is accurately described as coupling among dynamical systems, there must be a basis for that coupling. That is, there must be some means for one system to influence the other. Presumably this information is to be found primarily in the speech signal, but a cautionary note is in order here. The principle of affordance expresses a fit between the information available in perception to an organism, and the action possibilities available to that organism. There is no part of this description that requires that the physical

description of that information be simple in the sense of readily decomposable by analysis into “basic” physical variables. Indeed, the example of the affordance of nesting for mice that was provided above is a useful reference point here, as it would be very difficult indeed to analytically decompose the information that specifies that the environment affords nesting. This point has been nicely illustrated by Runeson (1977), who points out that

“Biological evolution might have arrived quite easily at solutions which require the utmost of capacity and sophistication of a human mind for their basic principles of operation to be understood” (Runeson, 1977, p. 174)

Recently, a technique for the continuous quantitative assessment of synchrony in parallel speech signals was developed (Cummins, 2008). Using this tool, a series of stimuli were prepared derived from the synchronous reading of a text by a single subject. These stimuli selectively retained some information of the original recording, and discarded the rest. By asking subjects to attempt to synchronize with these modified recordings, it was possible to inquire about the relative importance of fundamental frequency, amplitude envelope modulation, and intelligibility in synchronization. Synchronization was unaffected if the model speech was resynthesized with a flat fundamental frequency. Likewise, merely low pass filtering the model speech with a cut off at 500 Hz did not reduce synchrony. Indeed, any modification to the speech that left gross temporal structure and intelligibility more or less intact seemed to support synchronization. However the retention of the slow modulation of the amplitude envelope in signal correlated noise did not support synchronization. In a second set of stimuli, the amplitude envelope was used to modulate a sustained vowel created by replicating a single pitch period excised from a natural utterance. Synchrony was better for these model utterances than for amplitude modulated noise. Synchrony was further improved when the sustained vowel was resynthesized with a fundamental frequency matching the original utterance. Collectively, these results suggest that no single simple physical variable can be implicated as the basis for coupling among speakers. Amplitude envelope, fundamental frequency and voice quality (vowel versus noise) were all found to contribute to some degree to the observed synchronization of speaker with model utterance. If rhythm is viewed as an affordance for the entrainment of movement, this is not terribly surprising, as the auditory and movement systems have a long, complex, and contingent phylogenetic history resulting in an exquisite match between perceptual ability and movement proclivity, which must be considered as twin sides of a single perception/action system, rather than two distinct systems.

5.3: Further evidence

Synchronization provides one source of evidence for the view that speakers can effectively couple their speech production, but the difficulty of uncovering the basis for that coupling makes formal modeling of the resulting system difficult. There is, however, ample evidence from neuroscience that the perception of speech from one source serves to directly couple the motor speech production system of a listener, as would be predicted by the entrainment account of rhythm in speech. This evidence comes from TMS studies in which the evoked muscle potentials of speech articulators were found to be increased relative to baseline when subjects simply listened to the speech of others (Fadiga et al., 2005). In one study, motor evoked potentials from listeners' tongue muscles were elevated relative to baseline when listening to speech generated using substantial tongue movement (Fadiga et al., 2002). In another study, listening to speech or watching lip-rounding movements of speakers generated enhanced motor evoked potentials in the orbicularis oris muscle (Strafella, 2000). In both these studies, the simple act of listening to speech automatically caused an increased activation in those pathways responsible for producing speech sounds. These results are of a kind with a greater number of studies demonstrating that there is a tight coupling in time between the subsystem responsible for carrying out motor movements of the hands and the observation (visual or auditory) of similar movements or acts in others (Fadiga et al., 2005; Gangitano et al., 2004). To quote from a review by Fadiga et al. (2005):

when we observe another individual acting we strongly 'resonate' with his or her action. In other words, our motor system simulates underthreshold the observed action in a strictly congruent fashion. The involved muscles are the same as those used in the observed action and their activation is temporally strictly coupled with the dynamics of the observed action. (Fadiga et al., 2005, p. 213)

This, of course, is entirely consistent with the fundamental assertion of the Motor Theory of Speech Perception, which is that neural representations underlying speech perception are the same as those underlying speech production (Liberman and Mattingly, 1985).

6. Conclusion

The very concept of rhythm employed here bears little relation to any property of the speech signal conventionally described as rhythmic in the phonetics literature. Phoneticians have perhaps sometimes taken their lead from phonologists in starting out to describe ‘rhythm’ as one dimension along which languages (English, Polish, etc) may be characterized and classified. This may be a revealing way to look at language, but it does some violence to the sense with which the term ‘rhythm’ is used elsewhere.

If, on the other hand, rhythm is understood as an affordance for the entrainment of movement, we arrive at an understanding of the role of temporal structure within speech that is continuous with our understanding of rhythm in nature. Rhythm organizes, coordinates, and unifies disparate systems. It is a basic coordinative principle of nature found in ensemble behaviors of animals, and in the structured relations among parts within a single organism. Speech is a specifically human activity. But it is also coordinated movement, and the principles underlying the fine coordination of speech movement are not essentially distinct from those underlying all coordinated movement.

Acknowledgements

This work has been supported by the Science Foundation Ireland through Principal Investigator Grant Number 04/IN3/I568 to the Author.

References

- Allen, G. D.: Speech rhythm: its relation to performance universals and articulatory timing. *Journal of Phonetics*, **3**:75–86 (1975).
- Backwell, P., Jennions, M., Passmore, N., and Christy, J.: Synchronized courtship in fiddler crabs. *Nature*, **391**:31–32. (1998).
- Barbosa, P. A.: Explaining cross-linguistic rhythmic variability via a coupled-oscillator

model of rhythm production. In *Proceedings of Prosody 2002*. (2002).

Bennett, M., Schatz, M., Rockwood, H., and Wiesenfeld, K.: Huygens's clocks. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, **458**(2019):563–579. (2002).

Buck, J.: Synchronous Rhythmic Flashing of Fireflies. II. *The Quarterly Review of Biology*, **63**(3): 265. (1988).

Carson, R. G., Byblow, W. D., and Goodman, D.: The dynamical substructure of bimanual coordination. In *Interlimb Coordination: Neural, Dynamical and Cognitive Constraints*, Chapter 16, pages 319–337. (Academic Press, 1994).

Chemero, A.: An outline of a theory of affordances. *Ecological Psychology*, **15**(2):181–195. (2003).

Collier, J. and Burch, M.: Order From Rhythmic Entrainment and the Origin of Levels Through Dissipation. *Symmetry: Culture and Science*, **9**:2–4. (1998).

Collier, J. and Burch, M.: Symmetry, Levels and Entrainment. *Proceedings of the International Society for Systems Sciences*. **1**. (2000).

Cummins, F.: On synchronous speech. *Acoustic Research Letters Online*, **3**(1):7–11. (2002).

Cummins, F.: Practice and performance in speech produced synchronously. *Journal of Phonetics*, **31**(2):139–148. (2003).

Cummins, F.: Rhythm as entrainment: The case of synchronous speech. *Journal of Phonetics*. In Press. (2008).

Cummins, F. and Port, R. F.: Rhythmic constraints on stress timing in English. *Journal of Phonetics*, **26**(2):145–171. (1998).

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

- Deutsch, A.: Orientation-Induced Pattern Formation: Swarm Dynamics in a Lattice-Gas Automaton Model. *International Journal of Bifurcation and Chaos*, **6**:1735–1752. (1996).
- Fadiga, L., Craighero, L., Buccino, G., and Rizzolatti, G.: Speech listening specifically modulates the excitability of tongue muscles: a TMS study. *European Journal of Neuroscience*, **15**(2):399–402. (2002).
- Fadiga, L., Craighero, L., and Olivier, E. Human motor cortex excitability during the perception of others action. *Current Opinion in Neurobiology*, **15**(2):213–218. (2005).
- Galves, A., Garcia, J., Duarte, D., and Galves, C.: Sonority as a basis for rhythmic class discrimination. In *Proceedings of Prosody 2002*. (2002).
- Gibbon, D. and Gut, U.: Measuring speech rhythm. In *Seventh European Conference on Speech Communication and Technology*. ISCA. (2001).
- Gibson, J. J.: *The Ecological Approach to Visual Perception*. (Houghton Mifflin, Boston, 1979).
- Grabe, E. and Low, E.: Durational Variability in Speech and the Rhythm Class Hypothesis. *Laboratory Phonology 7*. (2002).
- Haken, H., Kelso, J. A. S., and Bunz, H.: A theoretical model of phase transitions in human hand movement. *Biological Cybernetics*, **51**:347–356. (1985).
- Kelso, J. A. S.: *Dynamic Patterns*. (MIT Press, Cambridge, MA., 1995).
- Lashley, K. S.: The problem of serial order in behavior. In Jefress, L. A., editor, *Cerebral Mechanisms in Behavior*, pages 112–136. (John Wiley and Sons, New York, NY., 1951).
- Lehiste, I.: Isochrony reconsidered. *Journal of Phonetics*, **5**:253–263. (1977).
- Lieberman, A. M. and Mattingly, I. G.: The motor theory of speech perception revised. *Cognition*, **21**:1–36. (1985).
- Lin, L., Chen, G., Kuang, H., Wang, D., and Tsien, J.: Neural encoding of the concept of

nest in the mouse brain. *Proceedings of the National Academy of Sciences*, **104**(14):6066. (2007).

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

Maturana, H. and Varela, F.: *The Tree of Knowledge: The Biological Roots of Human Understanding*. (New Science Library: Distributed in the United State by Random House, Boston, 1987).

Nagel, T.: What is it like to be a bat? *Philosophy: Basic Readings*, **83**. (2004).

Norman, D.: Affordance, conventions, and design. *Interactions*, **6**(3):38–41. (1999).

Norton, A.: Dynamics: an introduction. In Port, R. F. and van Gelder, T., editors, *Mind as Motion: Explorations in the Dynamics of Cognition*, Chapter 1, pages 45–68. (Bradford Books/MIT Press, Cambridge, MA., 1995).

O'Dell, M. L. and Nieminen, T.: Coupled oscillator model of speech rhythm. In *Proceedings of the International Congress of Phonetic Sciences*, San Francisco. (1999).

Patel, A., Iversen, J., Bregman, M., Schulz, I., and Schulz, C.: Investigating the human-specificity of synchronization to music. In et al., M. A., editor, *Proceedings of the International Music Perception and Cognition Conference (ICMPC10)*. Sapporo: Japan / Adelaide: Causal Productions. (2008).

Port, R. F., Tajima, K., and Cummins, F.: Self-entrainment in animal behavior and human speech. *Online Proceedings of the 1996 Midwest Artificial Intelligence and Cognitive Science Conference*. URL: <http://www.cs.indiana.edu/event/maics96/proceedings.html>. (1996).

Ramus, F., Nespors, M., and Mehler, J.: Correlates of linguistic rhythm in the speech signal. *Cognition*, **73**(3):265–292. (1999).

Runeson, S.: On the possibility of 'smart' perceptual mechanisms. *Scandinavian Journal of Psychology*, **18**:172–179. (1977).

Skarda, C. A. and Freeman, W. J.: How brains make chaos in order to make sense of the world. *Behavioral and Brain Sciences*, **10**(2):161–195. (1987).

Strafella, A. and Paus, T.: Modulation of cortical excitability during action observation: a transcranial magnetic stimulation study. *NeuroReport*, **11**(10):2289. (2000).

Summers, J. J., Rosebaum, D. A., Burns, B. D., and Ford, S. K.: Production of polyrhythms. *Journal of Experimental Psychology: Human Perception and Performance*, **19**(2):416–428. (1993).

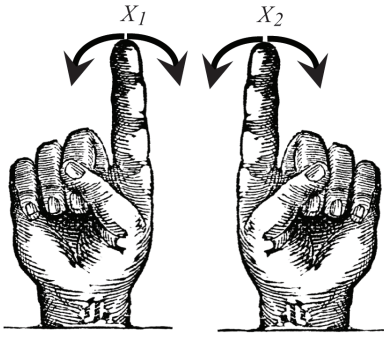
Figure Captions

Figure 1:

Left: Two fingers constrained to oscillate with identical frequencies are modelled as coupled oscillators. The autonomous dynamic of the finger is on the LHS of the equation, the coupling term on the right. Right: The coupled system can be described in simpler terms using a potential function that captures the relative stability as a function of the relative phase, i.e. the phase difference between the two hands. This potential function can be derived from the more complex set of equations describing each oscillator separately.

Figure 2:

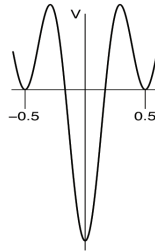
Left: histograms of observed phases of the medial stressed syllable onset within the phrase repetition cycle for four subjects. Target phases provided by a two-tone metronome were drawn from a uniform distribution from 0.3 to 0.7. Right: Musical characterization of the three modes observed in the subject data. Each mode corresponds to a simple integral nesting of a stress foot within the phrase repetition cycle. Data from Cummins and Port (1998).



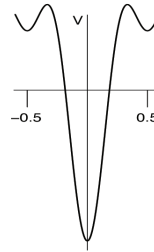
$$\ddot{x}_1 + f(x_1, \dot{x}_1) = G(x_1, \dot{x}_1; x_2, \dot{x}_2)$$

$$\ddot{x}_2 + f(x_2, \dot{x}_2) = G(x_2, \dot{x}_2; x_1, \dot{x}_1)$$

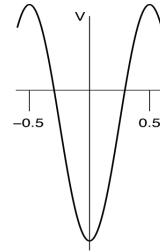
$$V = -a \cos \phi - b \cos 2\phi$$



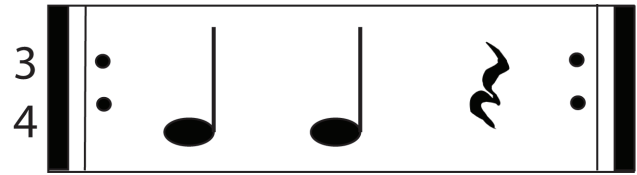
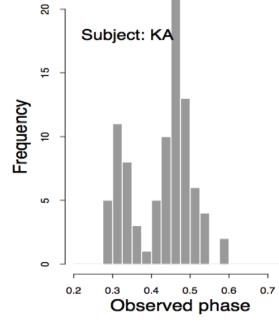
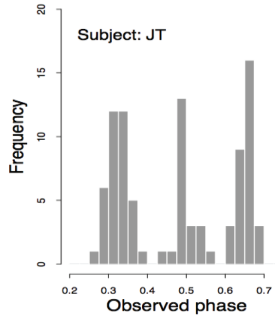
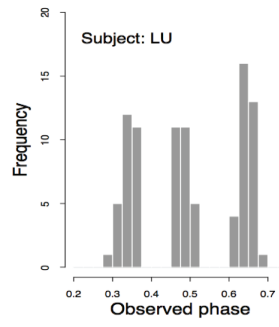
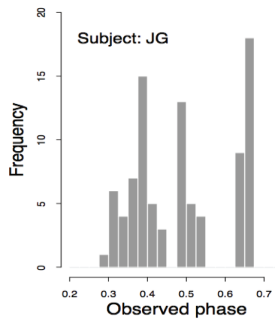
$$b/a = 1$$



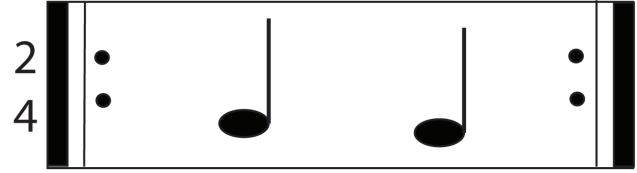
$$b/a = 0.5$$



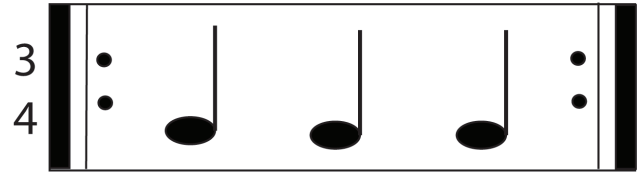
$$b/a = 0$$



Big... duck



Big... duck



Big for a duck