

Theories and Models of Accent

John Goldsmith

University of Chicago

The problem

- How can we tell whether we are studying the phonological equivalent of the long division algorithm? That particular (heuristic-led, stochastic) algorithm was a momentous invention, and it undoubtedly changed world history – but it offers little or no insight into the workings of the number system, and is no longer a significant part of our world, even though we still use division all the time, using a different algorithm that hardly anyone knows.

This is inexorably linked to the question of the reality of our work as linguists, and to the autonomy of linguistics as a discipline. We may ask: who should worry if linguists' models don't look like the models of the other cognitive sciences (or other sciences *tout court*) – them or us?

It is a source of concern that the level of description of our contemporary phonological models is the level of analysis that is most convenient for smart phonologists – while the level of analysis preferred by most learning algorithms is (a) lower; (b) quantitative in character; (c) one at which computation can be understood as optimization using local search techniques. (Optimality theory shares with generative phonology in every way the placement of generalizations at the level of analysis convenient for human observers; avoidance of what Smolensky 1986 calls the subsymbolic level is the great irony of current OT.)

- *Quantitative* character of analysis: Since the rise of generative phonology, and with the notable exception of SPE's treatment of English stress (utilizing a 0-4 levels of stress, with automatic downshifting of stress during cyclic stress reassignment), the watchword has been:
 - *if numbers are involved, it's phonetics; if it's phonetics, it's not phonology*. This maxim can be made exceptionless by restricting it to non-integers:
 - *if numbers other than whole numbers are involved, then it's phonetics, and it's not phonology*.

Among the apparent violations of the first, but not the second, maxim are: (1) column height on the metrical grid (for accent, uniformly; for moraicity, Hayes 1995 among a few others). (2) Measure of constraint-violation in OT (Kirchner).

The irony runs deeper. Generative theory is riddled with quantitative measures of complexity, from underspecification theory and phonological markedness considerations (both informal versions of information theory) and its intellectual grandfather, phonemic theory, all the way to the generative evaluation metric, which proposes a theory of grammar selection based on minimal length of grammar. Largely unbeknownst to linguists, sophisticated work in this area, both theoretical and applied, has continued for decades (from Kolmogorov and Shannon in the 1930s and 1940s down to Rissanen's minimum description length approach (1989), and others in the 1990s).

- *Digression on OT use of continuous variables*: OT's fundamental claim is that "In a variety of clear cases where there is a strength asymmetry between two conflicting constraints, no amount of success on the weaker constraint can compensate for failure on the stronger one." (Prince and Smolensky 1997 *Science* 275: 1604.) But that claim can easily be made invulnerable to test if we

choose to reinterpret violations as "evidence" that constraints should be broken up into pieces. For example, in Kenstowicz's classic Chukchee epenthesis case, things work by the principle: "Epenthesize when you need to avoid a consonant cluster inside a syllable, but don't epenthesize *inside* a morpheme unless you have no way out." That is, using strictly ranked constraints, *No Coda Clusters* is the highest ranked constraint, *Align Morpheme to Syllable* is the next ranked constraint, and *No Coda* is ranked next. Clearly the right move is to make Coda weight a constraint whose value is numerical, and which can outweigh Alignment when its violation is "2 coda consonants" or greater. As Smolensky himself notes (ibid., 1608), "[t]hat strict domination governs grammatical constraint interaction is not currently explained by principles of neural computation"; as he doesn't say, neural models suggest a harmonic phonology as in Goldsmith (1990, 1993), itself explicitly based on Smolensky's harmony theory (Smolensky 1986).

- We want a model in which search -- both for grammar and for individual utterance -- is near-local, and in the ideal case is a matter of evolution into an attractor of a dynamical system. (Contrast this with search in OT, which can be described as: "Search for the representation R which has the smallest decimal (or larger radix as necessary) expansion $0.a_1a_2a_3a_4$," where a_1 is the number of violations by representation (R) of constraint 1, a_2 is the number of violations of constraint 2, etc. Call that mapping V, whereby a representation R is mapped to such a decimal expansion; as far as I know, no knowledge of the local topography of the map from R to V(R) at a given place is helpful in locating the V's global minimum. Smolensky (op. cit.) implies (as I read him) that just such a Lyapunov function (or "harmony function") exists for a neural network implementation of OT (and he contrasts P & S's model to others by virtue of its *not* being an "informal" implementation), but how to develop a Lyapunov function whose local (or global) minima corresponds to those of V(R) remains a mystery.

Dynamic Computational networks (with Gary Larson)

1. Motivations:

- 1.1. Integration of rules and representations: a continuation of the unfinished task of enriching representations; explicate the notion of stress clash.
- 1.2. Make available continuous-variable learning algorithms.
- 1.3. Deal with cyclicity non-derivationally.
- 1.4. Integrate accent and sonority.
- 1.5. Do away with irregular and representationally clumsy use of numerical values (Continuous column constraint, moraic mismatch (Hayes 1995)). .

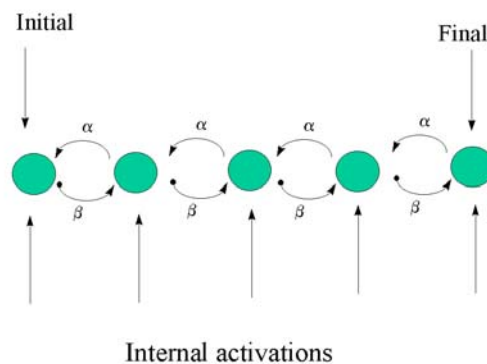
2. Model

We model each syllable with one node; each node has an activation level (which can be positive or negative). The activation level of each unit is the sum of four terms:

1. Positional activation – by virtue of position in a word (initial, final, penult)
2. Inherent activation – in quantity sensitive systems, heavy syllables get extra activation
3. Bias – in some models, we assign a constant (non-zero) activation to all units
4. Local effect – activation from left- and right-hand neighbors. Activation from the right-hand neighbor is that neighbor's activation times α ; activation from the left-hand neighbor is that neighbor's activation times β .

1. Limitations

1. Values of the derived activation remained largely uninterpreted.
2. The units appeared to be hard-wired: how do we deal with insertion and deletion?
3. The waves are epiphenomenal, and rhythm is not explained.
4. No natural way to deal with consecutive stresses: stress clash avoidance is built-in too strongly.
5. No way to link lengthening to stress clash avoidance.



Harmonic OSCILLATORS: some background

Oscillators are at home in the material world:

4. There are two empirical traditions studying the implementation of rhythm, one (evolutionarily) higher and one lower. The higher is summarized in Scott Kelso, *Dynamic Patterns* (MIT Press 1995), itself heavily indebted to the work of Hermann Haken; the lower is more informed and interested in problems that can be defined at the cellular level, or in terms of networks of countable numbers of neurons.

4.1. A typical presentation of the problem of rhythm:

Rhythmic phenomena abound in biology at all levels of analysis (refs.). At the behavioral level, rhythms can be observed in which the entire body or parts of the body move in a cyclic, repetitive fashion. Running, swimming, flying, breathing, chewing, and grooming are the most cited examples, but rhythmic behaviors need not be so stereotypic, a fact attested to by the musical accomplishments of humans. Although they are commonplace, behavioral rhythmicities are not well understood. To date, the efforts of neuroscience to explain behavior such as locomotion have been shaped in part by the issue of whether the behavior is generated centrally, by properties intrinsic to the central nervous system, or whether it is achieved by sensory feedback from moving parts of the body (refs.) Conventional wisdom and the weight of evidence favor the view of central pattern generation: a repetitive rhythmic output is said to be the product of a neural "oscillator" conceived either as a single pacemaker neural (ref.) or as a network of neurons (ref.) The pioneering work of von Holst (ref), an early proponent of central-pattern generators, addressed the problem of the coupling of two or more neural oscillators...

Glass and Mackey (op.cit): Periodic stimulation of spontaneously oscillating physiological rhythms has powerful effects on the intrinsic rhythm. As the frequency and amplitude of the periodic stimulus are varied, a variety of different coupling patterns are set up between the stimulus and the spontaneous oscillator. In some situations the spontaneous rhythm is entrained or phase locked to the forcing stimulus so that for each N cycles of the stimulus there are M cycles of the spontaneous rhythm, and the spontaneous oscillations occurs at fixed phase (or phases) of the periodic stimulus ($N:M$ phase locking)...The stable zones of the phase locking most commonly observed correspond to low-order ratios between the number of cycles of the forcing stimulus and the intrinsic rhythm (i.e., 2:1, 3:2, 1:1, 2:3, 1:2).(119-123).

Both of these traditions are of interest to us, and both are in agreement in viewing the system as a dynamical system, that is, as a set of differential or difference equations which trace the evolution of a system in an n -dimensional space, where a variable can be identified as time (either discrete time, in the case of difference equations, or continuous time, in the case of differential equations).

There are two distinct families of models of coupled oscillators – one in which the coupling is continuous throughout the phases of both oscillators, and one in which one oscillator "kicks" the other when its phase is 0. Both families of dynamical systems have large zones of regular behavior as well as large zones of chaotic behavior. We will focus on the first class here (in part, because what we want is to explore systems where coupling of low rational numbers – especially 1:1, 2:1, and 3:1 -- emerges out of a broad range of parameter settings, rather than to build that in).

In the simplest cases, systems of multiple oscillators can be viewed as sets of oscillators θ_i , where $d/dt \theta_i(t) = \omega_i + H_i(\theta_1, \theta_2, \dots, \theta_n)$. (1)

For the very simplest case, we could make two additional assumptions: all the effects can be understood as the sum of pairs of interacting oscillators; and oscillators interact on the basis of the difference of their phase. That is,

$$H_i = \sum_j h_{ij} (\theta_i - \theta_j) \quad (2)$$

So, what are the h_{ij} ? The H_i are periodic (every 2π , they come 'round again), so we consider the Fourier expansion (i.e., model it with sines – cosines are unnecessary because $h_{ij}(0) = 0$).

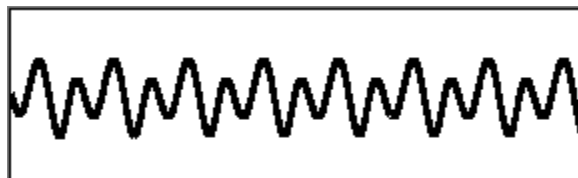
In short, $d/dt \theta_i(t) = \omega_i + \sum_j a_{ij} \sin (\theta_i - \theta_j)$ (3)

A model like this allows for excellent (1:1) phase-locking between oscillators whose inherent frequency (ω_i) are quite different.

Linguistic Model of prosodic rhythm

6. Back to linguistics. How can we think about phonological problems as dynamical systems? What do we get by trying to do so? What we gain is two things, possibly three: first, the possibility of more rich, powerful models, if the continuous variables involved can be mapped to observables; second, we may be able to *explain* qualitative (discrete) behaviors if we can show that they correspond to attractors, or basins of attraction, in the phase space of such a dynamical system (and third, we may be able to improve channels of communication with our sister disciplines).

I will assume that there are three oscillators available (mora, syllable, and foot). Each have an inherent period, of increasing magnitude, from mora to foot. I would like to show that the metrical grid is a simplified idealization of such a dynamical system.



Sum of two waves, and ...

x x x x x

x x x x x x x x x Metrical sketch of that graph:

Equation (3), as it stands, deals only with 1:1 phase locking, but it can be generalized to the more general case of M:N phase locking (M, N integers) by *dropping* the assumption that each oscillator is a "pure tone" (that is, of constant angular velocity) and assuming only that it is periodic.

Now, saying that the oscillators are periodic implies that they can be modeled with a Fourier series, which is to say (ignoring the cosine term as before – remember this from phonetics?).

$$\theta_i(t) = \sum_n a_n \sin(n\omega_i t) \quad (4)$$

The a_n form a series in which only the first few terms make any difference to us: we'll get enough mileage for now just looking at the first two. This means that (3) isn't quite general enough; in the case of two coupled oscillators, each oscillator will have two terms to describe it:

$$\theta_1(t) = a_{1,1} \sin \omega_1 t + a_{1,2} \sin 2\omega_1 t \quad (5)$$

$$\theta_2(t) = a_{2,1} \sin \omega_2 t + a_{2,2} \sin 2\omega_2 t$$

and so the new form of (2) is a bit more complicated: instead of there being just 2 parameters specifying the coupling between oscillator 1 and 2, there are 4:

$$H_i = \text{some function of } (\theta_1, \theta_2) =$$

$$\text{some function of } (a_{1,1} \sin \omega_1 t + a_{1,2} \sin 2\omega_1 t,$$

$$a_{2,1} \sin \omega_2 t + a_{2,2} \sin 2\omega_2 t)$$

and we'll actually assume it's ...

= a linear combination of

$$(\sin \omega_1 t - \sin \omega_2 t) \text{ and } (\sin \omega_1 t - \sin 2\omega_2 t) \text{ and } (\sin 2\omega_1 t - \sin 2\omega_2 t) \quad (6)$$

[but the first and third can be collapsed to a single term].

So when this is all done, what we end up with is a pair of oscillators whose frequency is determined by the following equation, and without loss of generality we can consider a system with two degrees of freedom: we fix ω_2 as 1, letting ω_1 vary; and α can take on any value between 0 and 1 (see note 2).

$$d/dt \theta_1(t) = \omega_1 - \alpha \sin(\theta_1 - \theta_2) - (1 - \alpha) \sin(\theta_1 - 2\theta_2) \quad (7)$$

$$d/dt \theta_2(t) = \omega_2 + \alpha \sin(\theta_1 - \theta_2) + (1 - \alpha) \sin(\theta_1 - 2\theta_2)$$

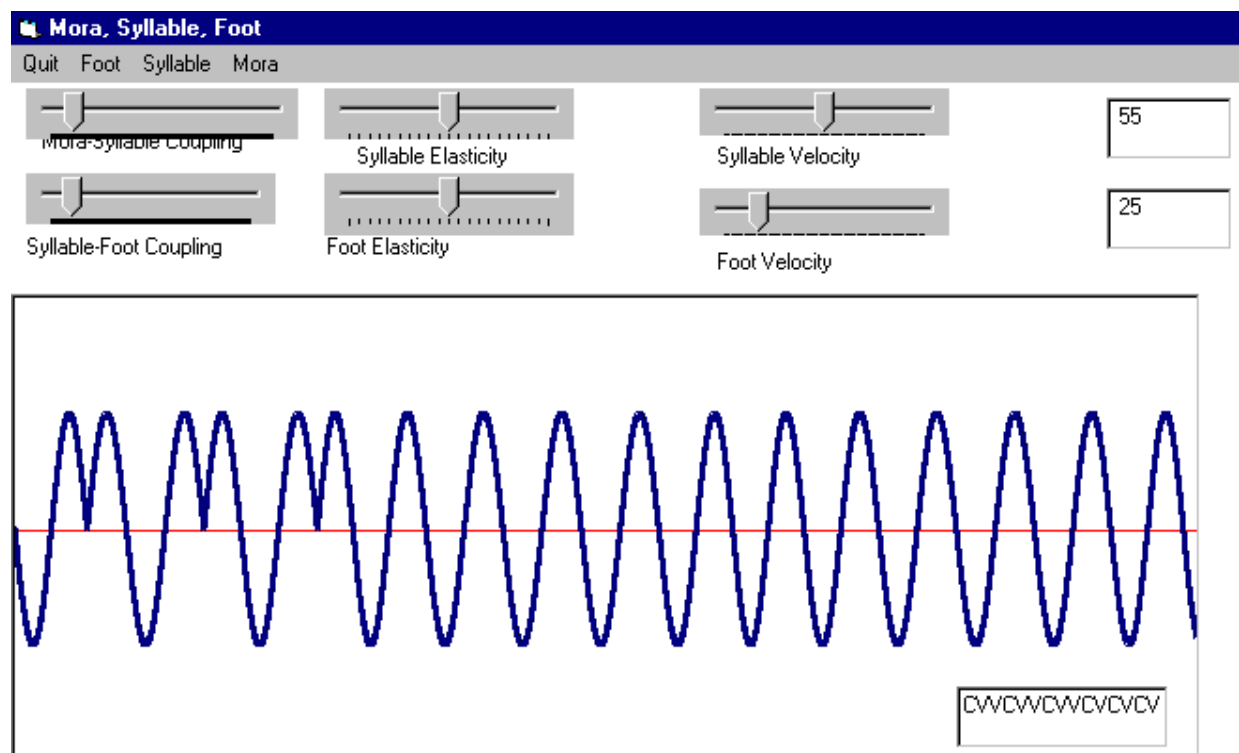
What we would like to find is this:

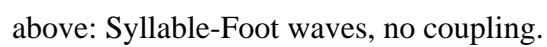
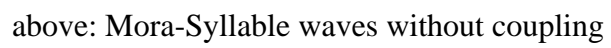
a setting for these parameters whereby the system would easily move from 1:1 to 2:1 phase-locking and back again on the basis of the linguistic context.

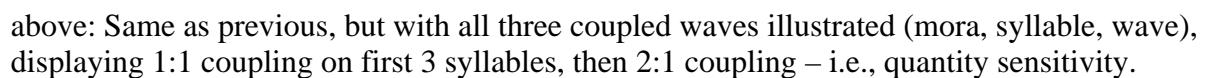
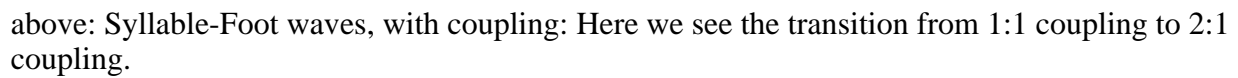
This highlights an important point: a system of coupled oscillators is a dynamic system with attractors, unlike the model of dynamic computational networks. *The model will contribute to our understanding of linguistic problems if its attractors can be shown to correspond to common linguistic patterns.*

8. Linguistic concerns

7. Quantity-Sensitive stress systems: with proper parametric settings, we *can* establish a triple of oscillators (mora, syllable, foot) which forms a dynamical system, and which has two attractors, one in which the foot:syllable ratio is 1:2 and another in which the ratio is 1:1. **The system chooses 2:1 when the syllables are light, and 1:1 when they are heavy – i.e., it displays the core quantity-sensitive accent placement algorithm.** This is a remarkable result, I think. We want to replace the hand-crafted process of assigning constituency on the basis of factors that we eyeball with dynamical systems whose attractors just happen to have the appropriate integral phase-locking relationships.







Phonology can be a sophisticated descriptive domain, playing by its own rules. To *some* degree, it must be so. But mainstream theoretical linguistics has fallen behind in the cooperative venture that we call the cognitive sciences in some important respects. (The most important respect is a cavalier attitude towards learning algorithms, and an unwillingness to let work on learning algorithm drive linguistic theory – as opposed to be perfectly willing to let linguistic theory drive learning theory, which inevitably leads one to a learning theory that looks like an expert system, and hence devoid of serious interest.)

Phonology must thus continue to develop its own domain-internal descriptive tools. But both foundational concerns and the ability to translate linguistic models into systems that can be linked to other cognitive systems requires us to make the effort to understand the mathematical underpinnings of linguistic systems and to be willing to be open to the possibility that successful phonological modeling is not reducible to convenient generalizations that are easily grasped at the phonologist's most comfortable level of analysis.