Emergence of syllable structure from a coupled oscillator model of intergestural timing

Louis Goldstein
Department of Linguistics
Yale University
and
Haskins Laboratories

Phonology: combinatorial system of speech units

- Phonology
  - Discrete, context-independent speech units recombine to create the word-forms of language.
  - What are the primitive units ?
  - What is the glue that holds them together in word-forms ?
- Articulatory Phonology
  - Goal is to attempt to find answers to these questions
  - Both phonological and physical properties emerge lawfully from a common representation.

Articulatory phonology: units

(Browman & Goldstein, 1992; 1995a)

- Act of speaking can be decomposed into atomic units of vocal tract constriction action, or *gestures*.
- Properties
  - Macroscopic. Gestures are discrete and can function as units of information (contrast and combination).
  - Microscopic. Continuous, context-dependent motion of articulators and sound unfolds lawfully from pattern of temporally overlapping gestures.

Articulatory Phonology: glue

- What is the glue that holds gestural atoms together in the appropriate patterns?
- Answer should account for observed regularities of gestural combination (properties syllable structure).

Syllable Structure: regularities of gestural combination

- Macroscopic (Phonological)
  - Onsets and rimes exhibit relatively free combination in most languages.
  - Other combinatorial possibilities are typically more limited:
    - Nuclei and codas
    - Cs wihin onsets and and within codas
  - CV syllables are unmarked.
- Microscopic (Physical)
  - Relative timing of consonants in an onset cluster is more stable (less variable) than in a coda cluster.
  - Timing of consonants to the vowel varies as additional consonants are added to an onset, but not to a coda (in English).

Outline

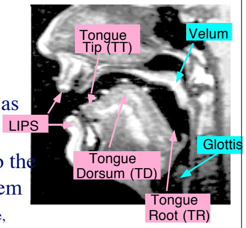
1. Gestures as discrete units
2. Coupling model of planning intergestural timing
3. Coupling model and syllable structure

What makes gestures discrete?

- distinct organs
- within-organ differentiation into distinct modes
- abstract (task) dynamical description

Organ independence

- Gestures control independent constricting devices, or organs.
  Organs = Articulators of phonological theory (Halle, 1983)
- Gestures of distinct organs count as discrete differences.
- Even neonates show sensitivity to the partitioning of the oro-facial system into distinct organs (Meltzoff & Moore, 1977).



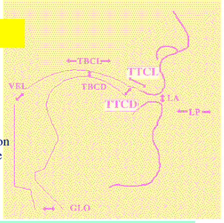
Discrete differentiation of within-organ action

- Gestures of a given organ can be differentiated by the degree and location of the constriction goal.

Constriction Goal Parameters

tick
sick
thick

LP	lip protrusion
LA	lip aperture
TTCL	tongue tip constrict location
TTCD	tongue tip constrict degree
TBCL	tongue body constrict location
TBCD	tongue body constrict degree
VEL	velic aperture
GLO	glottal aperture



- Possible to model the emergence of discrete regions of continua through self-organization in systems of agents that attune to one another (e.g. de Boer, 2000; Goldstein, 2003; Oudeyer, 2003) .

Between- vs. within-organ differentiation

- View predicts that systematic differentiation of an organ's constriction goals are acquired later than systematic use of distinct organs themselves. (Studdert-Kennedy, 2002; Goldstein, 2003).
- Infant must attune to the environment to develop within-organ modes.
- Preliminary support using perception of infant productions
  - Goldstein (2003), Son (in prep)

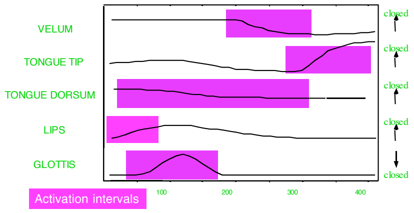
Discreteness in time: Dynamical systems

- Articulators move continuously in time during a constriction action.
- Where is the discrete unit in this continuous change?

Differential Equations with fixed parameter values

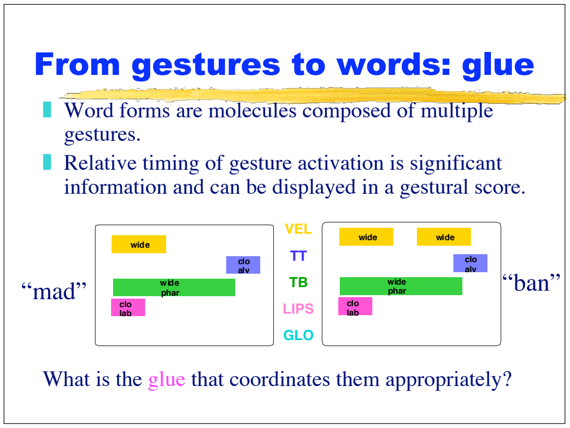
Give rise to continuous motion over time

"pan"



Task Dynamics (Saltzman, 1985; 1995)

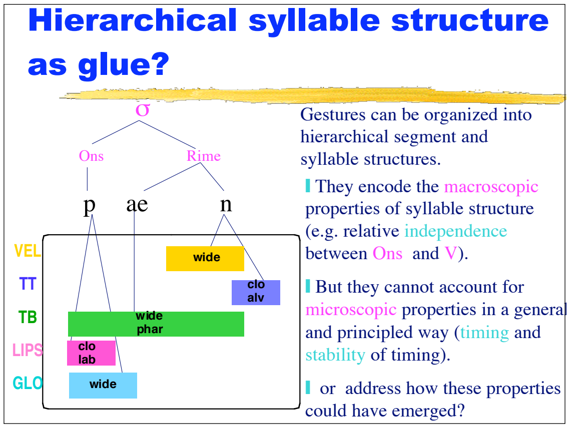
- Constriction formation can be modeled as a (time)-invariant dynamical system that achieves a goal (task):
  - e.g., LA (distance between the lips) is task goal variable.
- Form of continuous motion over time emerges from the dynamical specification of active gestures.
- Context-dependence emerges from temporal overlap of invariant dynamical units
  - Invariant dynamics at the task level shapes the time-varying, context-dependent dynamics at lower levels of the system (articulators and muscles).



From gestures to words: glue

- Word forms are molecules composed of multiple gestures.
- Relative timing of gesture activation is significant information and can be displayed in a gestural score.

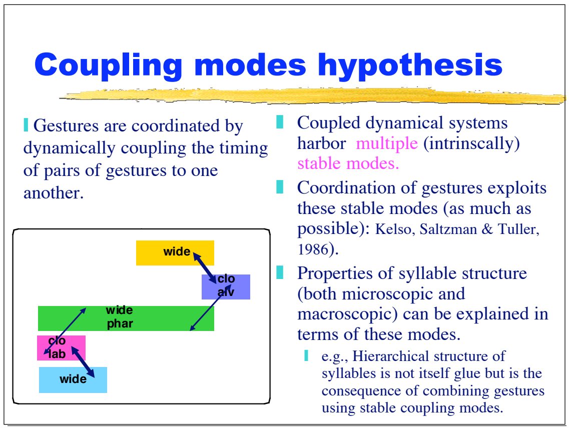
What is the glue that coordinates them appropriately?



Hierarchical syllable structure as glue?

Gestures can be organized into hierarchical segment and syllable structures.

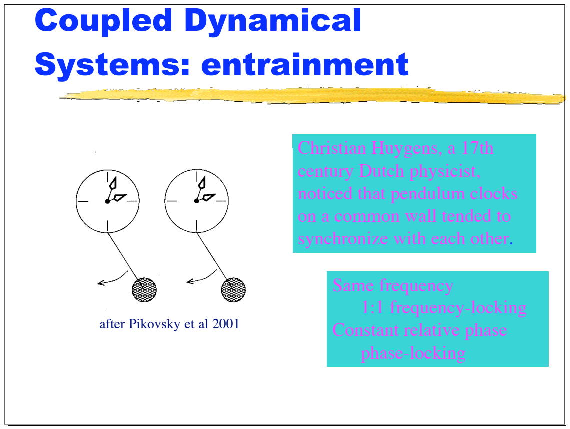
- They encode the macroscopic properties of syllable structure (e.g. relative independence between Ons and V).
- But they cannot account for microscopic properties in a general and principled way (timing and stability of timing).
- or address how these properties could have emerged?



Coupling modes hypothesis

- Gestures are coordinated by dynamically coupling the timing of pairs of gestures to one another.

- Coupled dynamical systems harbor multiple (intrinscally) stable modes.
- Coordination of gestures exploits these stable modes (as much as possible): Kelso, Saltzman & Tuller, 1986).
- Properties of syllable structure (both microscopic and macroscopic) can be explained in terms of these modes.
  - e.g., Hierarchical structure of syllables is not itself glue but is the consequence of combining gestures using stable coupling modes.

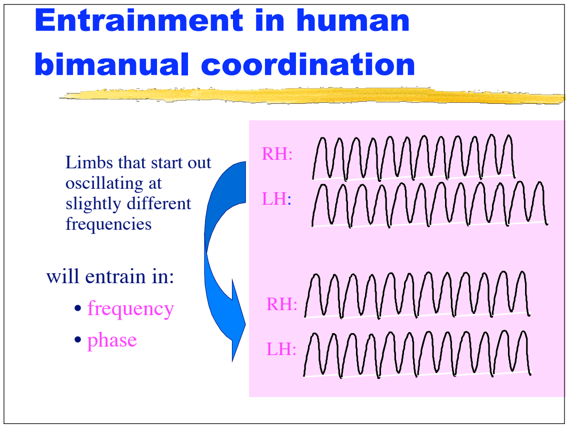


Coupled Dynamical Systems: entrainment

after Pikovsky et al 2001

Christian Huygens, a 17th century Dutch physicist, noticed that pendulum clocks on a common wall tended to synchronize with each other.

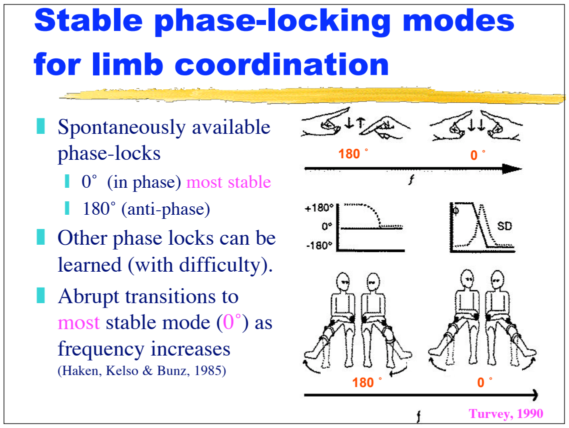
Same frequency
1:1 frequency-locking
Constant relative phase
phase-locking



Entrainment in human bimanual coordination

Limbs that start out oscillating at slightly different frequencies

will entrain in:
- frequency
- phase



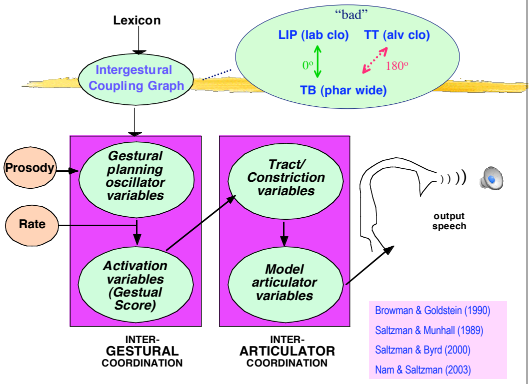
Stable phase-locking modes for limb coordination

- Spontaneously available phase-locks
  - 0° (in phase) most stable
  - 180° (anti-phase)
- Other phase locks can be learned (with difficulty).
- Abrupt transitions to most stable mode (0°) as frequency increases (Haken, Kelso & Bunz, 1985)

Turvey, 1990

Planning intergestural timing
(Nam, Saltzman & Goldstein)

- Planning can be modeled as kind of internal repetition.
  - Each gesture corresponds to an oscillator.
  - Oscillators are coupled pair-wise to one another (according to a coupling graph) so as to achieve a target relative phase.
  - During (internal) repetition, coupling causes oscillators to settle at stable relative phases (Saltzman & Byrd, 2000).
  - Final relative phases can be used to trigger gestural activation (as shown in the gestural score).
- Coupling graph for an utterance
  - specifies how pairs of gestures are coupled to one another (target relative phases).
  - Properties of syllable structure emerge as consequences of this graph.

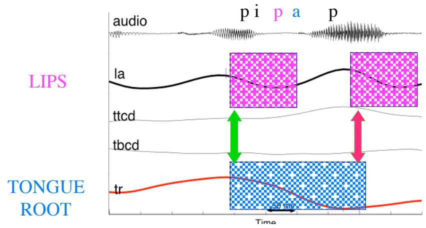


Browman & Goldstein (1990)
Saltzman & Munhall (1989)
Saltzman & Byrd (2000)
Nam & Saltzman (2003)

Modes in Coupling Graphs: C and V gestures

- If a consonant (C) gesture and a vowel (V) gesture are to be coordinated in an intrinsically stable mode, there are just two possibilities:
  - in-phase
    - hypothesized for C-V (onset relation) most stable
  - anti-phase
    - hypothesized for V-C (coda relation)
- Distinct C-V and V-C modes have been hypothesized has far back as Stetson (1951)
  [more recently, Tuller & Kelso, 1991; DeJong (2001)]
  - Here implications are followed for a theory of syllable structure

Evidence for C-V and V-C modes



C and V gestures (of CV) are in phase

V and C gestures (of VC) are anti-phase

Explaining combinatorial properties of syllables

- Hypothesis: Combinatorial freedom of gestures is possible just where intergestural coordination exploits the most stable mode of coupling.
  - As long as gestures are coupled in the most stable mode, any gesture can be combined with any other.
  - With less stable (or non-intrinsically stable modes), specific phasings may have to be learned, so free combination is less likely.

Predictions

- Onset C gestures should combine freely with V gestures, (which can explain free combinatoriality of onsets and rimes).
- Coda C gestures are in a less stable mode with Vs, and therefore there should be increased dependency between V and final C.
- Within-onset and within-coda consonant coordination may employ non-intrinisically stable modes.
  - specific couplings must be learned
  - acquired late
  - typically small numbers of combinations

C and V gesture valences

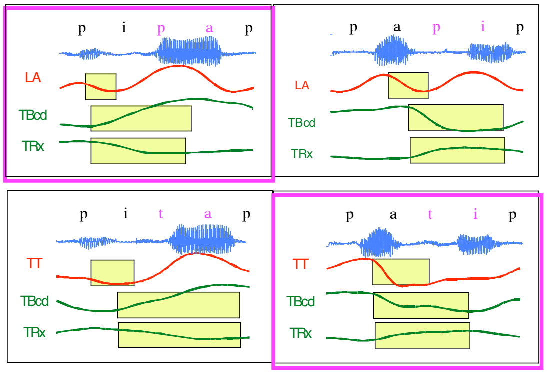
- C and V gestures are differentiated by
  - degree of constriction (V is wider)
  - dynamic stiffness (V takes longer to get to target)
  - activation interval (V still active after C released)
- Nature of these differences is such that C and V gestures can be in phase (at onset) and still be both be recoverable by listeners (Mattingly, 1981).
- These gestural properties, together with the stability of in-phase coupling gives rise to valence of C and V gestures -- they combine freely with each other in C-V structures.

Biases in CV combinations

- Grammatically, onset C and V combine freely in many languages (e.g., English).
- However, MacNeilage and Davis (2000) have found there are statistical biases in C-V combinations in the lexicons in a sample of 10 languages
  - Combinations occurring with greater than chance frequency:
    - Coronals with front Vs
    - Labials with central Vs
    - Dorsals with back Vs
  - McNeilage and Davis find the basis for these patterns in the earliest "syllables" produced by infants.
    - They hypothesize that infants are only oscillating their jaws.

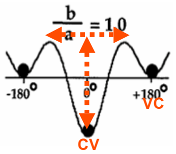
Alternative: gestural synchony and articultory constraint

- Some problems with jaw oscillation only theory for infants:
  - Preferred patterns occur more frequently than expected by chance, but many other combinations also are produced.
  - Adult languages show similar trends, but we know adults do more than oscillate the jaw -- C and V can be independent.
- Alternative Hypothesis:
  - While gestures in CV are hypotheiszed to be triggered synchronously, some CV combinations do not afford articulatory synchrony between C and V gestures, due to intrinsic constraints of the gestures themselves (e.g., Recasens, Solé) or their recoverability.
  - The most frequent combinations are those in which the articulatory synchrony matches synchrony in gestual triggering.



Specific model of modes and additional predictions

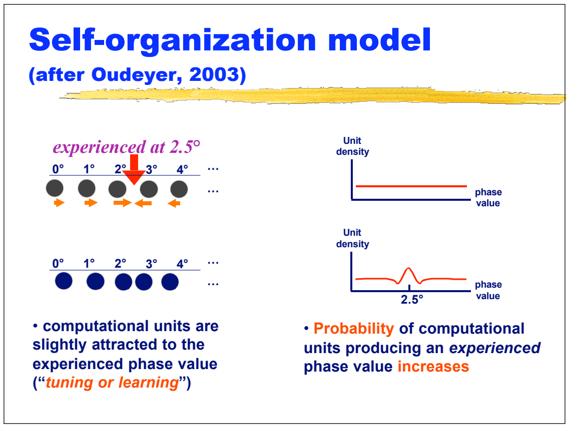
- A potential function has been found to characterize qualitative features of coupled oscillatory systems (Haken, Kelso & Bunz, 1985).
  - two local minima (0˚, 180˚)
  - $V(\Phi) = -a \cos(\Phi) - b \cos(2\Phi)$
  - modeled results of many experiments on interlimb coordination
  - in-phase attractor is wider and deeper



- Predictions
  - Shorter planning time for CV than VC syllables
  - Earlier acquisition of CV than VC syllables

Acquisition of CV vs. VC

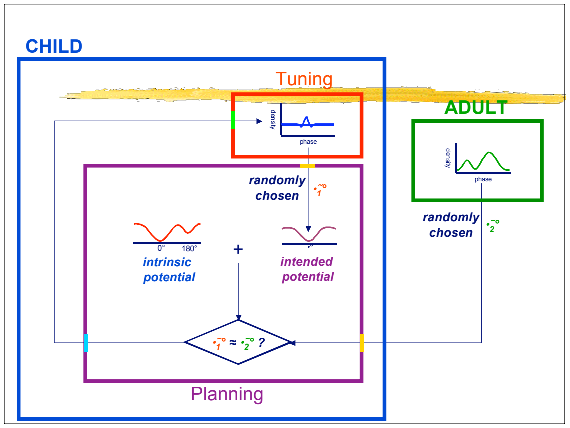
- Infants develop CV syllables before VC (in all languages).
- Self-organization model for phase leaning that incorporates HKB coupling function (Nam).

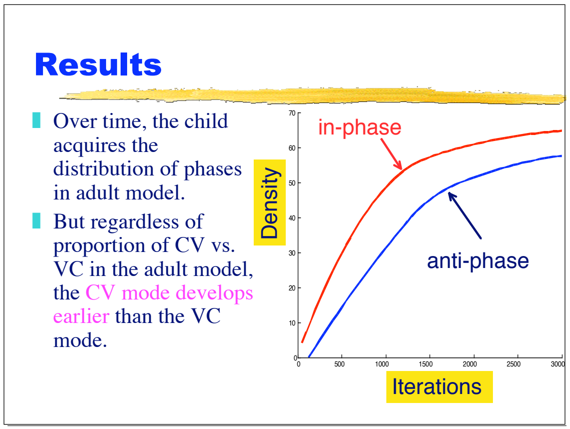


Self-organization model

(after Oudeyer, 2003)

- computational units are slightly attracted to the experienced phase value (“*tuning or learning*”)
- Probability of computational units producing an *experienced* phase value increases



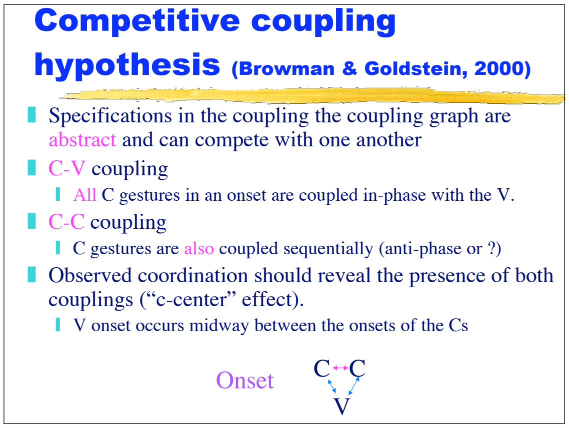


Results

- Over time, the child acquires the distribution of phases in adult model.
- But regardless of proportion of CV vs. VC in the adult model, the CV mode develops earlier than the VC mode.

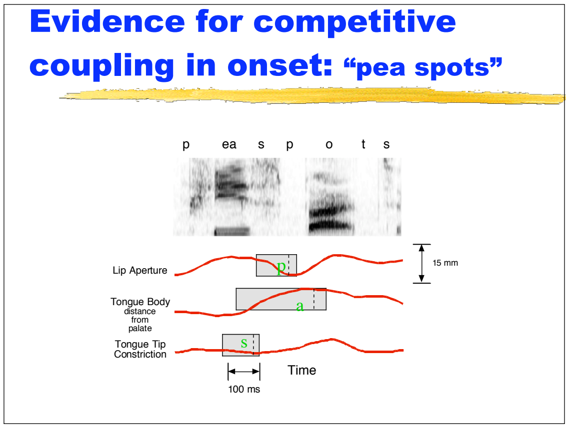
Onsets composed of multiple gestures

- If onset is defined by an in-phase relation between C gesture and V, then all onset C gestures should be synchronous with V (and therefore with each other).
- Combinations of a clo or crit gesture (stop or fricative) with a wider gesture allow recoverability of both gestures even when synchronized.
  - This result is a segment (e.g., nasal or aspirated stop).
- Combinations of multiple clo or crit gestures present recoverability problems if synchronous.
  - Gestures must be at least partially sequential (cluster).
  - What makes them all part of the onset?



Competitive coupling hypothesis (Browman & Goldstein, 2000)

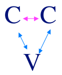
- Specifications in the coupling the coupling graph are abstract and can compete with one another
- C-V coupling
  - All C gestures in an onset are coupled in-phase with the V.
- C-C coupling
  - C gestures are also coupled sequentially (anti-phase or ?)
- Observed coordination should reveal the presence of both couplings (“c-center” effect).
  - V onset occurs midway between the onsets of the Cs



Evidence for competitive coupling in onset: “pea spots”

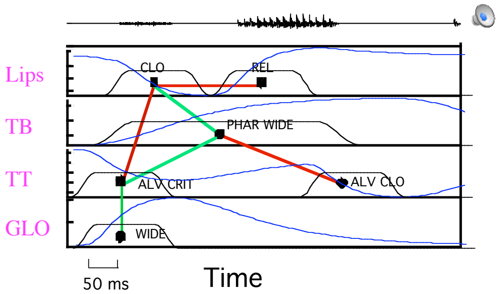
Competitive coupling model

- Generalization of phase planning model to accomodate multiple, competing ψ (Nam & Saltzman, 2003).
- Coupling graph hypothesized for onsets:



- Oscillators still settle into stable patterns (but ψ will not be all be achieved if they are in competition).
- Output phasing consistent with C-center is obtained

Example: /spæt/



Clusters in onset vs. coda

- Onset is in-phase relation between C gestures and V
- Coda is anti-phase relation:
  - between V and C
  - among C gestures in coda
  - Only weak attraction expected of multiple Cs to anti-phase relation to V.

Onset vs. coda: microscopic consequences

- Coda clusters do not regularly show C-center (Byrd, 1995; Honorof & Browman, 1995)
  - Predicted by coupling graph--no coupling between C_2 and V.
- Timing between C gestures is more stable in onset clusters than in coda clusters (Byrd, 1996).
  - Adding noise to intergestural coupling model results in more C-C variability in codas than in onsets, due to multiplicity of specified couplings in onset.

Language-particular coupling grammars

- Differences in topology of coupling graphs
  - Modes provide preferences, but ultimately, coupling graphs must be learned.
  - Different V-C coupling in VC light vs. heavy
  - Different coupling of oral constrictions and velum in coda.
- Language differences in coupling graphs could be modeled as resulting from different constraint rankings:
  - Gafos (2002)
  - Nam (2004)
    - max (zero-coordination) [in-phase]
    - min (NON-zero-coordination) [other phase targets]

Language differences in coupling strength

- In a competitive model, coupling strengths (potential well depth) can differ for different links.
- Language differences in relative coupling strength:
  - Georgian initial clusters (Chitoran, Goldstein & Byrd, 2002)
    - more separation in time than English clusters (C-C > C-V)
    - more separation in back-to-front order than front-to-back.
  - May yield qualitative differences, depending on nature of competitive model
    - linear vs. non-linear (strict dominance)

Summary and Prospects

- A competitive, coupled oscillator model for planning intergestural timing may be able to account for several microscopic and macroscropic properties related to syllable structure.
- Future Directions:
  - Modelling of multisyllabic utterances
  - Development of an explicit model that takes account of intrinsic articulatory constraints in modulating relative timing of gestures

References

Browman, C. P., & Goldstein, L. (1988). Some notes on syllable structure in articulatory phonology. *Phonetica, 45*, 140-155.
Browman, C. P., & Goldstein, L. (1992). Articulatory phonology: An overview. *Phonetica, 49*(3-4), 155-180.
Browman, C. & Goldstein, L. (1995a). Dynamics and articulatory phonology. In R. Port & T. van Gelder (Eds.) *Mind as motion; Explorations in the dynamics of cognition.*. (pp. 175-193). Cambridge, MA: MIT Press.
Browman, C. & Goldstein, L. (1995b). Gestural syllable position effects in American English. In F. Bell-Berti & L. Raphael (Eds.) *Producing speech: Contemporary issues*. (pp. 19-33). NY: American Institute of Physics.
Browman, C. P., & Goldstein, L. (2000). Competing constraints on intergestural coordination and self-organization of phonological structures. *Bulletin de la Communication Parlée*, 5, 25-34.
Byrd, D. (1995). C-centers revisited. *Phonetica, 52*, 285-306.
Byrd, D. (1996). Influences on articulatory timing in consonant sequences. *Journal of Phonetics, 24*, 209-244.
Chitoran, I., Goldstein, L. & Byrd, D. (2002) Gestural overlap and recoverability: Articulatory evidence from Georgian. Gussenhoven, C. & Warner, N. (eds). *Papers in Laboratory Phonology,* 7 (pp.419-447). Berlin; Mouton deGruyter.
deBoer, B. (2000). Self-organization in vowel systems. *Journal of Phonetics, 28*, 441-465.
deJong, K. (2001).Rate induced re-syllabification revisited. *Language and Speech, 44*. 229 – 259.
deJong, K., Lim, B., & Nagao, K. (2004). *Language & Speech*, *47*, 241-266.
Goldstein, L. (2003). Emergence of discrete gestures. In Solé, M.J., Recasens, D., and Romero, J. Proceedings of the 15th International Congress of Phonetic Sciences, pp. 85-88.
Goldstein, L., & Fowler, C. (2003). Articulatory phonology: a phonology for public language use. In Meyer, A. & Schiller, N., *Phonetics and Phonology in Language Comprehension and Production: Differences and Similarities* (pp. 159-207). New York: Mouton.
Halle, M. (1983). On distinctive features and their articulatory implementation. *Natural Language and Linguistic Theory, 1*, 91-105.

Haken, H., Kelso, J. A. S., & Bunz, H. (1985). A theoretical model of phase transitions in human hand movements. Biological Cybernetics, 51, 347-356.
Honorof, D. N., & Browman, C. P. (1995). The center or edge: How are consonant clusters organized with respect to the vowel? In K. Elenius & P. Branderud (Eds.), *Proceedings of the XIIIth International Congress of Phonetic Sciences.*, (Vol. 3, pp. 552-555). Stockholm: KTH and Stockholm University.
Kelso, J. A. S., Saltzman, E. L., & Tuller, B. (1986). The dynamical perspective on speech production: Data and theory. *Journal of Phonetics, 14*(1), 29-59.
MacNeilage, P.F. and Davis, B.L., 2000a, Origin of the internal structure of word forms, *Science*, 288, 527-531.
Mattingly, I. G. (1981). Phonetic representation and speech synthesis by rule. In T. Myers, J. Laver, & J. Anderson (Eds.), *The cognitive representation of speech*, (pp. 415-420). Amsterdam: North Holland.
Meltzoff, A.N. and Moore, M.K. (1977).Imitation of facial and manual gestures by human infants. *Science, 198,* 75-78.
Nam, H. (2004). Syllable-level intergestural timing model: split-gesture dynamics focusing on positional asymmetry and moraic structure. Paper presented at LabPhon 9, University of Illinois, Champagne-Urbana, IL.
Nam, H. and Saltzman, E. (2003). A competitive, coupled oscillator model of syllable structure. In Solé, M.J., Recasens, D., and Romero*, J. Proceedings of the 15th International Congress of Phonetic Sciences*.
Oudeyer, P.-Y. (2003). L'auto-organisation de la Parole. Unpublished PhD, University of Paris VI.
Oudeyer, P.-Y., (in press). The self-organization of speech sounds. *Journal of Theoretical Biology*.
Pikovsky, A.,Rosenblum, M. & Kurths, J. (2001).*Synchronization: A universal concept in non-linear science*. Cambridge: Cambridge University Press.

Saltzman, E. (1985). Task dynamic coordination of the speech articulators: A preliminary model. In H. Heuer & C. Fromm (Eds.), *Experimental brain research*, (pp. 129-144). New York: Springer-Verlag.
Saltzman, E. (1995). Dynamics and coordinate systems in skilled sensorimotor activity. In Port, R. and van Gelder, T., *Mind as Motion* (pp. 150-173). Cambridge, MA: MIT Press.
Saltzman. E. and Byrd,D. (2000). Task-dynamics of gesturalt iming: Phase windows and multifrequency rhythms," *Human Movement Science*, *19*, 499-526.
Saltzman, E., Löfqvist, A., Kay, B., Kinsella-Shaw, J., and Rubin, P.,(1998). Dynamics of intergestural timing: A perturbation study of lip-larynx coordination. *Experimental Brain Research* 123: 412-424.
Saltzman, E. L., and Munhall, K. G. (1989). A dynamical approach to gestural patterning in speech production. *Ecological Psychology*, 1, 333-382.
Stetson, R. H. (1951). Motor phonetics. 2nd ed. Amsterdam: North-Holland.
Studdert-Kennedy, M. (2002). Mirror neurons, vocal imitation, and the evolution of particulate speech. In Gallese, V. & Stamenov, M. (eds). *Mirror Neurons and the Evolution of Brain and Language*. Amerstam: John Benjamins.
Studdert-Kennedy, M. and Goldstein, L. (2003) Launching language: Gestural origin of discrete infinity. Christiansen, M. & Kirby, S. (eds.) *Language evolution: The States of the Art*. (pp. 235-254). Oxford: Oxford University Press.
Tuller, B., & Kelso J. A.S. (1990). Phase transitions in speech production and their perceptual consequences.In M. Jeannerod (Ed.), *Attention and performance XIII*. Hillsdale, NJ: Erlbaum.
Tuller, B. & Kelso, J.A.S. (1991). The production and perception of syllable structure. *Journal of Speech and Hearing Research*, *34*, 501-508.
Turvey, M.T. (1990). Coordination. *American Psychologist, 45*, 938-953.

Thank You!