Articulatory Phonology

Chapter for *Routledge Handbook of Phonological Theory*

Nancy Hall, California State University Long Beach

1. INTRODUCTION

The central premise of Articulatory Phonology (AP) is that the representational units of phonology correspond to speech production events. Whereas most phonological theories assume that speakers mentally represent a word in terms of features or segments, AP uses a very different set of representations: articulatory *gestures*, and the *coordination structure* that determines their relative timing. Gestures act both as units of contrast and as units of speech production, essentially erasing the traditional distinction between phonology and phonetics.

AP has played a large role in the trend towards “laboratory phonology.” Developed in large part at Haskins Laboratories, through the work of Catherine Browman, Louis Goldstein, and colleagues (e.g. Browman and Goldstein 1986, 1988, 1989, 1990, 1992a, 1992b; Byrd 1995, 1996; Nam and Saltzman 2003; Nam et al. 2004, 2006, 2009), the model has been developed and tested through extensive articulatory phonetics research, as well as computational simulations of speech production.

This chapter is laid out as follows: section 2 introduces the basic mechanics of the theory, including the gestural representational system, the computational system that produces and interprets gestural representations, and the types of articulatory data that proposed representations are often based on. AP work on speech errors is reviewed as a case study. Section 3 reviews AP analyses of a variety of phonological processes, both categorical and non-categorical. Section 4 gives an overview of AP work on syllable structure, particularly the coupling model, an important recent development which attempts to explain onset-coda assymmetries as results of different gestural coupling relations. Section 5 covers several current trends in AP research, including work on modelling phonological acquisition, morphological structure, tone and intonation.

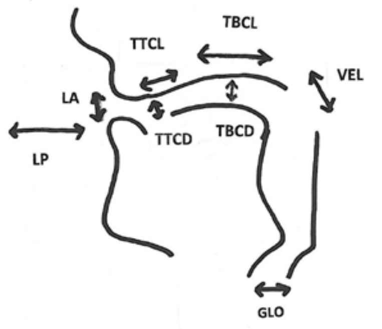
2. GESTURES AS PHONOLOGICAL PRIMITIVES

In AP, the basic units of phonological representation are not features or segments, but articulatory gestures. A gesture can be thought of as a task, a goal to be achieved through articulatory movements. Typical tasks in speech production might include “form a closure with the lips,” “spread the vocal folds,” or “position the tongue body close to but not touching the velum.”

Formally, articulatory goals are defined in terms of “tract variables” (Browman & Goldstein 1989). The most commonly used are those below. Each tract variable refers to a region of the vocal tract. Some of the goals specify a degree of constriction; other specify a location of constriction.

(1) **Tract variables**

lip aperture (LA)
lip protrusion (LP or PRO)
tongue tip constriction location (TTCL)
tongue tip constriction degree (TTCD)
tongue body constriction location (TBCL)
tongue body constriction degree (TBCD)
velic aperture (VEL)
glottal aperture (GLO)



Each of these variables can take a range of values (Browman & Goldstein 1989: 209), as shown below.

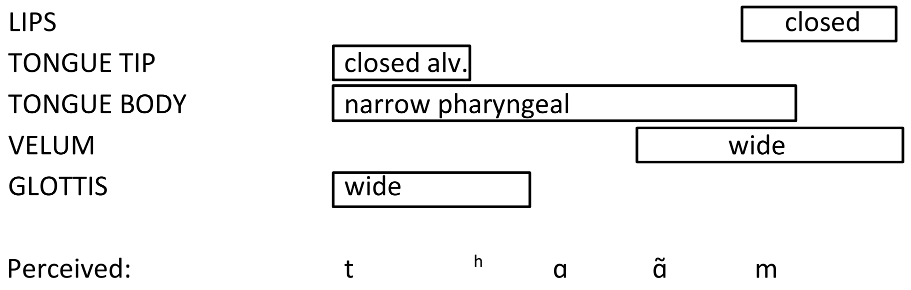
(2) Constriction degree values: closed critical narrow mid wide
Constriction location values: protruded labial dental alveolar postalveolar palatal velar uvular pharyngeal

Segments have no formal role in most AP work; they are regarded as epiphenomenal. Typically, what would be considered a segment in other frameworks corresponds to several gestures in AP. A transcription of [t], for example, would correspond to the gestures "GLO wide" (for voicelessness), and "TT alveolar closed." Since they refer to the same articulator, the variables of TTCD and TTCL must be specified together, as must TBCL and TBCD, and LA and LP.

The set of tract variables will probably expand as more sounds are modeled. For example, Browman & Goldstein (1989: 228) and Proctor (2010: 93) suggest that a tract variable for tongue tip closure orientation (TTCO) would be useful for capturing the difference between apicals and laminals, and a tract variable for cross-sectional shape of the tongue body would help capture the difference between laterals and centrals. Goldstein (1994: 238) notes that a Tongue Root (TR) variable will be needed for gutturals, but that its articulator set has not been explicitly modeled.

Gestures have a duration in time, and overlap with one another. This overlap is represented in a "gestural score." Below is an example of a possible gestural score for the word *Tom* [tʰãm]. [All gestural scores in this chapter should be understood as schematic and not necessarily to scale; some details may be inferred, and others have been simplified to illustrate the points at hand. For precise temporal and spatial data, please refer to the original sources cited.]

(3) Gestural score of [tʰɑ̃m] (horizontal axis = time)



Three gestures begin at about the same time: the tongue tip closure and glottal opening of the [t]; and the tongue body constriction of the [ɑ] (note that low vowels are considered pharyngeal constrictions). The idea of ordering is quite different in AP compared to most other theories. In a segmental representation of *Tom*, we would say that [t] comes before [a], but here the gestures associated with [t] simply *end* earlier than the tongue body gesture associated with [ɑ]. The tongue body gesture is itself overlapped by the velum opening gesture, and since the velum lowers before the lips close, the end of the vowel is nasalized.

Note that the glottis is opened for voicelessness (and creates aspiration by staying open after the tongue tip constriction ends). There is no gesture for voicing; a glottal aperture that produces voicing is assumed to be the default state (Browman and Goldstein 1989: 239). The velum is assumed to be raised by default and requires a lowering gesture to produce nasals (as shown above), but once lowered it requires a raising gesture to return to closure. Browman & Goldstein (1986: 242) note that “the decision to treat velic opening and closing as two separate gestures, as compared with the glottal and oral gestures that incorporate both opening and closing, is based on the fact that each velic gesture may act as a word-level phenomenon, so that the velum can possibly be held in either a closed or an open position indefinitely.”

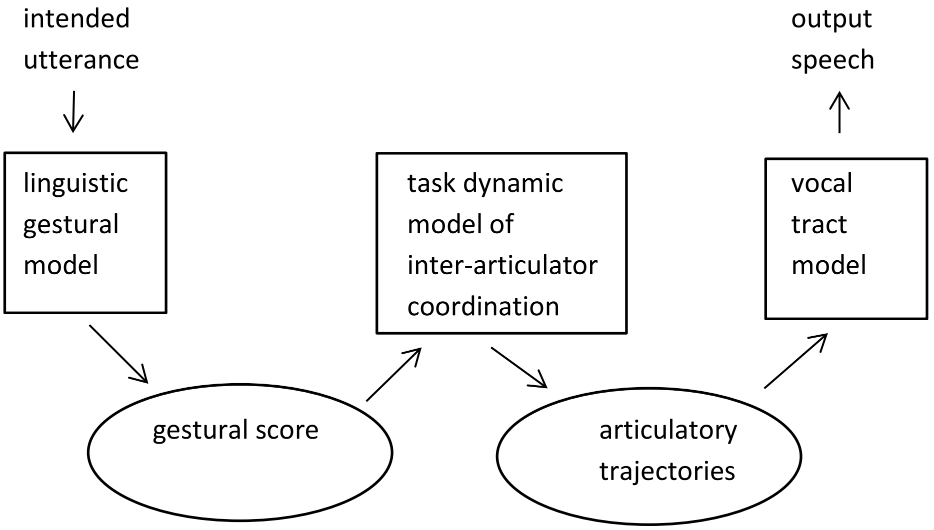
It is important to understand that a gesture, in AP terms, does not refer to the articulatory movements themselves. It specifies a goal, not a means of achieving the goal. The same gesture may cause different movements in different contexts, and possibly even involve different articulators, depending on factors such as what position the articulators start in. For example, imagine how you would achieve the task “tongue-tip alveolar closure” (for a [t] or [d], say), starting with the tongue in position for [i], as in “beet.” This would involve a small upward movement of the tongue, and likely no movement of the jaw, since the jaw would already be fairly high. The movements involved would be different than if the tongue started in the low back position of [ɑ], as in “bot” --- in that case, the tongue would need to move farther and from a different direction, and probably the jaw would raise. Yet the abstract closure gesture would be identical in both cases. Incidentally, if such a gesture was activated when the tongue was already creating an alveolar closure---perhaps for a preceding [n], as in *bent*---then the closure task would actually be accomplished without any movement. It would simply cause the tongue to stay in place longer.

By a similar token, the presence of a gesture does not necessarily mean the gestural target is fully achieved. In fast speech, a closure gesture might not result in a complete stoppage of airflow. Yet again, the abstract gesture would be the same.

2.1. Computational modeling of gestures

Gestural scores are part of a larger computational model of speech production, whose basic structure is shown below (Browman and Goldstein 1990:342). An intended utterance provides input to a linguistic gestural model. This model determines the coordination of gestures required for a particular utterance, and produces a gestural score representing the results.

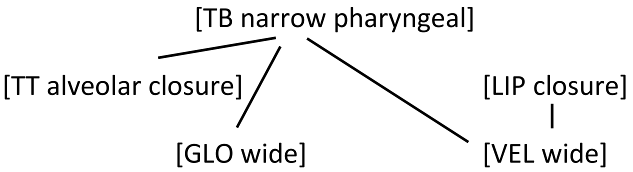
(4)



after Browman & Goldstein (1990: 342)

Within the linguistic gestural model, certain pairs of gestures are temporally coordinated with respect to one another. For example, the word [tʰãm] might, hypothetically, start with formal coordination relations between the pairs of gestures linked by lines below:

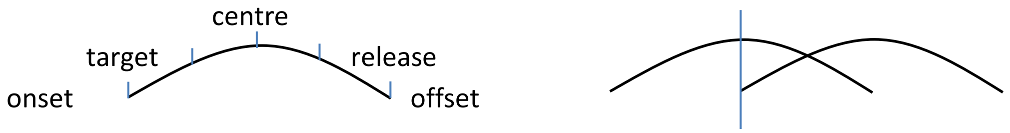
(5)



Determining which gestures are actually coordinated is an empirical question, answerable only through articulatory studies. Coordinated pairs of gestures are recognized largely through showing relatively stable timing with respect to one another. If a speaker produces a word repeatedly, especially with some variation in speech rate, the relative timing of coordinated gestures (like LIP and VEL in the hypothetical example above) will be more consistent than the relative timing of non-coordinated gestures (like LIP and TB).

Although various details of the model may vary from proposal to proposal, one major shift deserves special mention. Earlier and later AP models assume a very different mechanism for determining the level of overlap between a pair of coordinated gestures. The earlier approach, as sketched in Browman & Goldstein (1992b: 161), was for the linguistic gestural model to specify an alignment between two gestures' internal "landmarks." Landmarks were stages in the articulation of a gesture. This approach was perhaps most fully developed in Gafos (2002)'s analysis of consonant clusters in Moroccan Arabic. In this grammar, each gesture's landmarks included an ONSET, when the articulators first came under active control; a TARGET, when the desired constriction was to be reached; a CENTER, at the midpoint of the constriction; a RELEASE, when movement away from the target began, and an OFFSET, when the articulator ceased to be under active control. When coordinating two gestures, the linguistic gestural model could specify a particular alignment such as CENTER = ONSET (shown below right, where the vertical line indicates the point of alignment).

(6) Internal structure of a gesture Sample alignment CENTER = ONSET



(horizontal axis = time; vertical axis = gesture's activation level)

This approach had the advantage of being computationally explicit, but the disadvantage of allowing too many types of coordination. It has largely been supplanted by the coupled oscillator model (Saltzman & Byrd 2000, Nam & Saltzmann 2003, Nam 2007, Nam et al. 2009), which is discussed further in Section 4. However, familiarity with the older model is still useful for reading AP literature.

To convert gestural scores to articulatory movements, AP uses a model of task dynamics. Task dynamics is a model of the control of skilled movements, originally developed to describe non-speech movements such as reaching (Bernstein 1967; see Hawkins 1992 for an overview of applications to speech). Based on a gestural score, as well as a parameter for speech rate, the dynamical equations of the task dynamic model determine the actual trajectories that articulators would take. In many cases, overlapping gestures put competing demands on an articulator. In this situation, the task dynamic model "blends" the gestures, creating an articulatory path that compromises between the two goals.

As an implementation of task dynamics, AP researchers generally use TaDA (Task Dynamic Application, available freely at http://www.haskins.yale.edu/TaDA_download/). Developed by Hosung Nam and colleagues (Nam et al. 2004), one of TaDA's functions is to convert a gestural score to a set of vocal tract shapes. These articulatory trajectories in turn provide input to the vocal tract model HLsyn (Hanson & Stevens 2002), which converts them to an acoustic output.

Using this model, then, a researcher can simulate the articulatory and acoustic output of a hypothetical gestural representation. Such simulations play a large role in AP research. Typically, simulations are run for the purposes of comparison with articulatory (and sometimes acoustic) records of actual speech. A proposed gestural representation is judged as successful to the extent that the simulation and the actual speech match.

2.2. Methods: Articulatory data collection

Although a gesture is not itself a physical movement, physical movements are the best evidence for inferring gestural structures. For this reason, AP depends heavily on techniques for measuring articulatory motion, such as electromagnetic articulography (EMA/EMMA), ultrasound, x-ray microbeam, and real-time MRI. Even when articulatory records are obtainable, it is not always straightforward to detect the underlying gestural structure; as Gick et al. (2006: 69) comment: "the criteria for determining under what circumstances an observed physical event should be considered phonologically real have been vague in the previous literature on AP (and essentially absent from most other prominent models of phonology)." There is now a body of work aimed at developing algorithms for extracting gestural scores from articulatory speech records (for example, Ramanarayanan et al. 2013).

It is unfortunately much more difficult to infer gestural structures from acoustic records. Although a few articulatory events do have clear acoustic correlates (such as the achievement of a stop closure after a vowel), many do not. For example, it's often impossible to identify the beginning or end of a gesture from the acoustic record, because usually it is masked by other gestures that overlap it. In a word like *tee* [t^{h}i], the vowel gesture begins at some point during the stop closure, but it's hard to say exactly when. This is a problem, because recent work suggest that the beginnings of gestures are important anchor points in the control of coordination. So although studies do sometimes compare the results of simulations to acoustic records, articulatory records are considered far preferable.

One disadvantage of the reliance on articulatory data is that such data can be expensive and labor-intensive to collect and process. For this reason, AP studies tend to be based on a small number of subjects. Languages are quite often described on the basis of a single speaker, and a study with five speakers can be considered large in AP terms. The number of languages described to date is likewise small. Typological studies of dozens of languages, which play an important role in approaches such as Optimality Theory, are impractical in articulatory research. Lack of access to expensive equipment can also be a barrier for new researchers, although it should be noted that some seminal work in AP has been based on existing articulatory corpora rather than purpose-collected data. Several such corpora are now publicly available, such as the x-ray speech database of Munhall et al. (1995).

2.3. Case study: articulatory study of speech errors

As an example of how a phenomenon can appear qualitatively different when described from an articulatory as opposed to acoustic viewpoint, consider the recent AP work on speech errors (Pouplier and Goldstein 2005, Goldstein et al. 2007). Most studies of speech errors are based on impressionistic transcriptions, and these studies have converged on some apparently robust generalizations: for example, most speech errors are said to involve moving or substituting segments (as opposed to features), and these errors are said to almost never violate phonotactic rules of the speaker's language.

So, for example, *tariffs and barriers* might be mangled to *bariffs and terriers*, but would not become *tbariffs and btarriers*, in a language that does not allow [tb] or [bt] onsets.

However, AP studies such as Goldstein et al. (2007) have shown through EMA that the articulatory reality is different. When speakers were given a tongue-twisting exercise like repeating *top cop* over and over, they would slip and produce tokens that sounded like *cop cop or top top*. Yet in the articulatory movement traces, it was evident that the [k]s or [t]s produced in error were not like normal stops. Often, people seemed to produce both gestures at once, with a simultaneous velar and alveolar closure gesture (not necessarily fully achieved). Pouplier and Goldstein (2005) show that such errors are hard to hear correctly; listeners tend to either miss them or hear them as segmental substitutions. Similarly, Goldstein et al. (2007) also found that speech errors might involve only one of the gestures associated with a segment. For example, /m/ has both a bilabial closure gesture and a velar lowering gesture. When speech errors occurred in a phrase like *kim kid*, sometimes only one of the /m/'s gestures would move to the /d/. The velum might lower slightly during the tongue tip gesture, yet without lip movement.

These studies offer a strong challenge to the conclusions of non-articulatory studies: they suggest that many speech errors involve the movement of gestures rather than whole segments, and that the result does not have to conform to the language's phonotactics. Needless to say, this result is also highly consistent with the AP claim that gestures rather than segments are the basic units of speech.

3. REPRESENTATIONS OF PHONOLOGICAL PHENOMENA

The gestural representation system of AP means that phonological phenomena such as alternations must also be described in gestural terms. In some cases, these phenomena are analysed quite differently in AP than in segment and feature-based theories.

3.1. Categorical and non-categorical processes

One advantage of gestural representations is their ability to capture the differences between fast and slow speech without fundamentally restructuring the utterance. In theories based on segmental representations, casual speech is often described as characterized by the deletion or substitution of segments (see Browman and Goldstein 1990: 359 for numerous examples). Segments may acoustically disappear: a phrase like *he looked past me* might sound like [hi lʊk pæs mi], with /t/s eliding between two consonants; a word like *support* may sound like [sport], with elision of the schwa. Other segments lenite: an intervocalic /b/ as in *about* might be pronounced as a fricative [β]. Both nasal and oral stops tend to assimilate in place to following stop, so that phrases like *fat cat* may sound like [fækkæt]. In segment-based frameworks, such changes must be analysed (and transcribed) as categorical changes, governed by rules such as t → Ø / C__C. Yet this flies against evidence, both from articulatory studies and speaker intuitions, that at least some of these changes are actually gradient. The lenited segments resulting from fast-speech assimilation or lenition are not necessarily identical to regular, lexical occurrences of the (apparently) same segments.

Browman & Goldstein (1990) propose that no gestures are deleted in fast speech, nor do gestures change their tract variable values (such as LIP closure). Rather, fast speech causes reductions in the magnitude of gestures and increases in the relative overlap of gestures. Both of these changes can affect the acoustic output, by causing gestures to not reach their targets, or by hiding one gesture behind others. In a famous example, the authors identified a token in a corpus of X-ray films of speech where an English speaker pronounced *perfect memory* with the [t] acoustically absent. Yet the x-ray record showed that the tongue tip gesture was still executed. It was simply inaudible because it was completely overlapped by the closures of the preceding [k] and following [m].

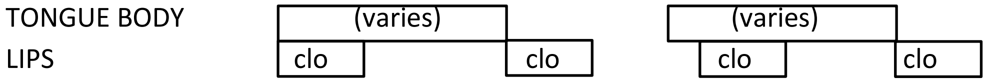
The AP approach does not assume, however, that every assimilation, lenition, etc. is necessarily a result of gestural overlap. Rather, it allows a better description of the difference between categorical and non-categorical changes. For example, Zsiga (1995) compares two processes in which /s/ palatalizes to [ʃ]. In words like *confession*, the (arguable) underlying /s/ that is pronounced in the stem *confess* obligatorily palatalizes, producing [kənfɛʃn̩]. Zsiga shows that this type of derived [ʃ] is indistinguishable from lexical [ʃ]. Yet a different picture emerges for the optional, casual speech phenomenon in which phrases like *press you* are pronounced like [prɛʃu]. Zsiga shows that this [ʃ] is different both from lexical [ʃ] and from the [ʃ] in *confession*. The degree of casual speech palatalization is variable, and some tokens are s-like at the onset of the fricative, yet ʃ-like by the end. Zsiga proposes that the palatalization in *press you* is caused purely by gestural overlap. When the tongue tip gesture of /s/ overlaps the tongue body gesture of /j/, the blending of the two gestures in the task dynamic model causes the tongue to retract to a more /ʃ/-like position. The palatalization in *confession*, on the other hand, involves some categorical alternation. In Zsiga's model, *confession* involves feature-spreading; her approach is unusual among AP theorists in giving a formal role to features. Another approach would be to assume that *confess* and *confession* underlyingly have different TT gestures.

3.2. Presence or absence of gestures: the case of schwa

Another area in which there has been considerable examination of categorical vs. non-categorical gestural changes is vowel alternations, particularly involving schwa. Schwa-vowels present an interesting ambiguity, because it is possible to produce an acoustic schwa without specifying a vocalic tongue-body target. This is partly because the tongue position for schwa is similar to the tongue's resting position, to which it returns during periods when it is not under active control.

To see what gestural scores could in principle underlie a schwa, Browman & Goldstein (1992a: 51-54) simulate the production of [pVpəpVp] sequences (where the gestural targets of the Vs vary), using gestural scores like that below. They find that during the gap between the second and third lip closures, where there is no active tongue body gesture, the tongue body dips towards a schwa-like position and a perceptible schwa appears in the acoustic output.

(7) gestural score of simulated [pVpəpVp]



Similarly, Gick & Wilson (2006) show that the acoustic schwa in words like *fire* [fajəɹ] does not necessarily reflect a schwa-target. The tongue has to pass through a schwa-like configuration on the way from the high front target of /j/ to the low back target of /ɹ/.

It does not appear that all schwas in real speech are targetless, however. Analyzing x-ray data of a English speaker producing sequences like [pipəpipə], Browman & Goldstein (1992a: 51) detect tongue movement towards a possible target associated with the medial schwa, and conclude that the data "argue against the strongest form of the hypothesis that schwa has no tongue target."

Nevertheless, the possibility of producing schwa without active control raises the question of whether targetless schwas occur in natural language. Gafos (2002) argues that such a schwa occurs in final CC clusters in Moroccan Arabic. Words ending in a heterorganic CC cluster have an audible schwa in slow speech but not fast speech: for example, the participle of "write" can be pronounced [kat^{ə}b] or [katb]. Using simulations, Gafos shows that a timing relation of CENTER = ONSET for the oral gestures (such as tongue tip alveolar closure and lip closure in /katb/) produces an audible release at slower rates of speech but no release at higher rates. On the other hand, final clusters of identical consonants have a schwa at all rates of speech, as in [wlas^{ə}s] "swollen gland". Gafos shows that a timing relation of OFFSET = ONSET produces a consistent audible schwa at all speech rates. He argues that this timing relationship reflects a principle of avoiding overlap between identical gestures, a type of gestural Obligatory Contour Principle.

Hall (2006) argues that targetless vowels show different phonological behaviors than vowels that correspond to a tongue body gesture. In a typological study (based on transcriptions), she identifies vowels, described as epenthetic, that have characteristics typical of a targetless vowel. These vowels have qualities that can be explained without positing a distinct gesture (either schwa, or influenced by the qualities of overlapping vowel or consonant gestures); they tend to be optional and disappear at fast speech rates; and they occur in heterorganic clusters, which are more prone to having an acoustic release between the consonants. She argues that vowels with these characteristics also tend to act phonologically invisible: for example, they do not count as a syllable in the stress system or for minimal word requirements; they are ignored in language games; and they fail to trigger phonological processes such as spirantization of a following stop. Speakers may be unaware that the vowels are even present. Furthermore, such vowels tend to occur in CC clusters that are cross-linguistically unmarked, and hence unlikely candidates for phonological repair.

Davidson & Stone (2003) show that targetless schwas may also occur in second language speech. When English speakers are asked to read pseudo-Slavic forms like *zgomu*, they often insert a schwa in a non-native consonant cluster, producing what sounds like [zəgomu]. Yet when these productions are studied by ultrasound, the tongue body position turns out to be different than that in real English words like *succumb* [səkʌm], where a schwa occurs between consonants that have the same TT and TB targets as /zg/. The articulatory trajectory of the tongue in [zəgomu] is consistent with the lack of an articulatory target for the schwa. This suggests that the acoustic schwa may be only a result of low gestural overlap between the consonants, which in turn is probably caused by speakers' lack of experience with coordinating consonant pairs that do not occur in their native language.

These studies illustrate how a common phonological topic like vowel epenthesis is seen differently in AP: the central question is what gestural structure underlies the (acoustic) vowels. The answer to this question may turn out to determine other aspects of the vowels' phonological patterning.

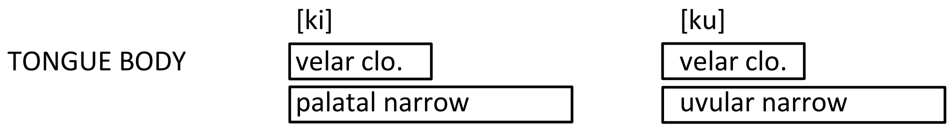
3.3. Capturing contrasts and allophony

The topic of contrast is another that is seen differently in AP than in most theories. In AP, contrasts are modeled in terms of gestural specifications, or coordination of gestures; the substance of the contrast is the same as the substance of phonetic realization. There is no precise equivalent for the concepts of 'phoneme', 'allophone', or 'feature'.

The methodology of studying contrasts in AP is also non-traditional. For example, Proctor et al. (2010) set out to construct a gestural theory of coronal contrasts in Wubuy. This Australian language has four coronal stops, transcribed as [t ʈ t̺ c]. In traditional, feature-based approaches to describing these contrasts, features would be posited based partly on phonetic descriptions, and partly on phonological patterns such as neutralization in particular environments. In Proctor et al.'s gestural approach, on the other hand, the first step was to collect EMA data of three speakers producing each stop between two vowels. Next, the researchers simulated a hypothetical gestural alignment using TaDA, and compared the results to the EMA data. Where discrepancies were found, in tongue shape or tongue trajectories, the model was iteratively adjusted and re-run to minimize the differences. Phonological distribution patterns played no role in the argumentation concerning the sounds' representation (although distribution does, of course, require explanation in AP). The result of the analysis is not a set of features, but a set of aperture, location and coordination settings for tongue body and tongue tip constrictions.

What is usually called allophony can result from more than one cause, in the AP approach. One cause is gestural blending, as described earlier in the task dynamic model. For example, suppose a velar stop gesture overlaps a vowel gesture. Since both gestures involve the same articulator, the tongue body, the task dynamic model must blend the two. The result is that the actual location of constriction will be different in sequences like [ki] and [ku]: a front vowel will pull the tongue body forward, creating a more forward constriction. The result would usually be described as allophony of the /k/, but at an abstract gestural level, there is no difference between the closure gesture in [ki] and [ku] (Saltzmann & Munhall 1989).

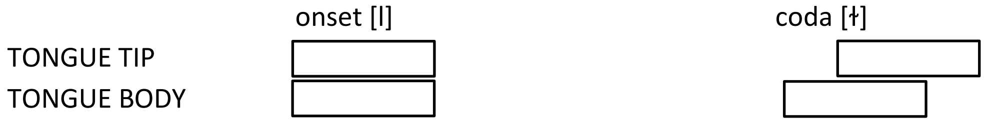
(8)



Of course, languages differ in the extent of CV coarticulation they display. For example, an English velar is only slightly affected by a following vowel, but a Navajo velar is dramatically affected: Navajo /x/ has allophones as divergent as [w] before [o] and [ç] before [i]. Iskarous et al. (2012: 7) propose to model this cross-linguistic variation in coarticulation levels by allowing language-specific settings for blending parameters, which essentially designate the relative strength of conflicting gestural targets. For example, if a language assigns the TBCL of the velar closure gesture a stronger weight than the TBCL of the vowel gesture, then the vowel will have minimal effect on the constriction location of the closure, as in English. If the vowel gesture has a stronger weight, then extreme consonant coarticulation will result, as in Navajo.

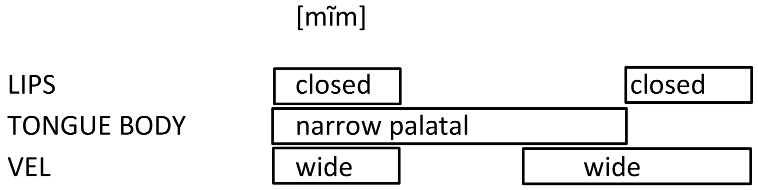
Some allophony, however, is not merely a matter of blending, but reflects differences in the magnitude or coordination of sets of gestures when they occur in different positions within a word or syllable. For example, English /l/ is described as having a “clear” quality in onset position and a “dark,” velarized quality in coda position. This is more than a blending effect; the gestures that correspond to /l/ have a different timing relation in different positions. English /l/ involves both a front tongue tip constriction and a back tongue body constriction. In onset position, the two gestures begin about simultaneously and the tongue tip gesture is strong, but in coda position, the tongue body gesture precedes the tongue tip gesture and the tongue tip gesture is relatively weak (Sproat & Fujimura 1993, Krakow 1999).

(9)



This turns out to be not an isolated fact about /l/, but a more general pattern in the organization of English syllables. For example, the velic opening and oral closure gestures of a nasal also tend to be simultaneous in onset position, but the velic opening precedes the oral closure in coda position, causing nasalization of the preceding vowel.

(10)



As Krakow (1999) points out, /l/ allophony has nothing to do with vowel nasalization in a featural view. But when represented gesturally, there is a clear parallelism: onsets are characterized by simultaneous production of gestures, and by strength of the oral gesture; codas are characterized by sequential production, with the oral gestures weaker and later. Findings like this have led to a strong focus on the role of syllable structure in gestural organization.

4. THE COUPLING MODEL OF SYLLABLE STRUCTURE

The recent AP focus on syllable structure has led to a new conception of the principles underlying gestural coordination. The older linguistic gestural model, where pairs of gestures were assigned relations like CENTER = ONSET (Gafos 2002), has been replaced by a theory in which there are only two kinds of gestural coupling (Saltzman & Byrd 2000, Nam & Saltzmann 2003, Nam 2007, Nam et al. 2009).

As mentioned before, AP draws on a body of work on the coordination of skilled motion, such as limb oscillation (Turvey 1990). In this work, skilled motions are modelled as being similar to critically damped oscillators. An oscillator is a system that displays a periodic movement, like a pendulum, or a spring with a weight attached. “Critical damping” means that the oscillator slows down as it approaches the target. For a real-life analogy, think of the springs between a car’s chassis and frame: if you push on the bumper and then release it, the springs return it directly to its equilibrium position, but shock absorbers critically damp the springs so that the bumper won’t bounce. The dynamical equations that describe this kind of motion are similar to what the task-dynamic model uses to describe speech motions. Once a task is activated (such as “tongue tip alveolar closure”), the articulator(s) start moving towards the target as if being controlled by a spring, but slow down on approach as if damped.

Each gesture is modeled as being controlled by a nonlinear planning oscillator, or “clock.” If we imagine a clock hand travelling through a 360° rotation, the beginning of the gesture occurs when the hand is at 0° phase and the end of the gesture at 360° phase.

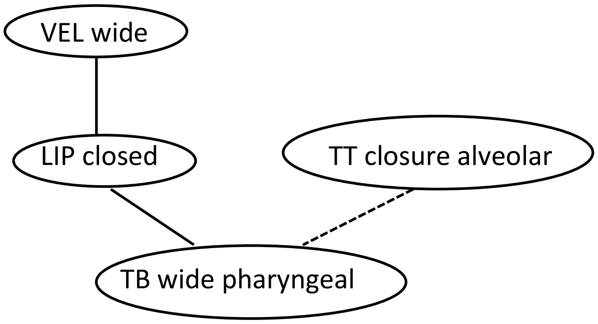
Oscillators can affect each others’ movement if they are “coupled.” This observation goes back to 1665, when Dutch physicist Christiaan Huygens noticed that the two pendulum clocks on his mantelpiece always beat in unison, and would return to this unison even if he deliberately disrupted their timing. He deduced that the clocks were subtly affecting each other through vibrations transmitted through the mantelpiece. Since then, physicists have shown that coupled oscillators tend to stabilize in one of two “normal modes”: in-phase timing, in which the oscillations are parallel (i.e., two pendulums swinging right and left at the same time), or anti-phase timing, in which the oscillations are opposite (i.e., one pendulum starting left when the other starts right). The process of gravitating towards these stable modes is called entrainment or mode locking.

Biophysicists have argued that entrainment is seen in the coordination of skilled motions as well (Haken et al. 1985). If you try to repeatedly do two movements at the same time (for example, tap your two index fingers on a table), you will tend to coordinate them either in-phase, by tapping the fingers simultaneously, or anti-phase, by alternating taps. Of these two modes, in-phase coordination is easier and more stable; the coordinated movements are very consistent in their relative timing and the rhythm is resistant to change. As a task gets harder, for example by speeding up the rate of finger-tapping, people tend to spontaneously switch from anti-phase to in-phase coordination.

Any “phase-lock” other than 0° or 180° relative phase is fairly difficult to maintain. People do accomplish more complex phasings when they learn skills like drumming or juggling, but these typically require considerable practice and often instruction.

It is hypothesized that speech evolved to use intrinsically stable modes of coordination whenever possible (Goldstein et al. 2006). Recent work (Saltzman & Byrd 2000, Nam & Saltzmann 2003, Nam 2007, Nam et al. 2009) has pursued the hypothesis that *all* gestural coordination can be captured with just two phasing relations. If two gestures have a controlled timing relation, then they are coupled either in-phase or anti-phase. The input to the gestural model consists of a "coupling graph" specifying which gestures are coupled and how. For example, the graph below shows the coupling structure that Goldstein et al. (2007) hypothesize for English [mæd]. Solid lines indicate in-phase coupling (meaning that the gestures would ideally begin simultaneously); the dotted line indicates anti-phase coupling. This is not the only conceivable coupling graph for this word, of course, and whether it is the correct one is a question to be settled empirically.

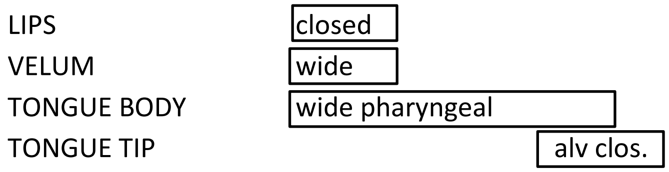
(11) Coupling graph of English *mad*, after Goldstein et al. (2007)



Once a coupling graph is established, it must be converted to a gestural score. This is done through an "intergestural level," which is implemented in TaDA. As described in Nam (2007), this level consists of a planning process that determines the most stable coordination of the entire gestural constellation. This is accomplished through a planning simulation in which each oscillator is started at an arbitrary phase of its clock (for example, VEL might begin at 40°, LIP at 65°, etc.), and all the gestures are set to oscillate repeatedly. At first, their relative timing changes on each repetition, as they are gradually pulled away from their random initial phasing relations towards the in-phase or anti-phase relations designated in the coupling graph. But eventually they settle into a stable pattern of relative timing, which stops changing from one repetition to the next. This stable timing pattern, plus a speech rate parameter, is the basis of the gestural score. (Again, this planning process takes the place of the earlier approach in which the gestural score was determined by explicit rules like CENTER = ONSET).

In the case of *mad*, the gestural score will be something like that below. This is a relatively simple case, because none of the in-phase or anti-phase coupling are in competition with one another (a problem that will be discussed further below).

(12) Gestural score of English *mad*



These planning simulations have more than one role in theory: they identify the most stable timing pattern and output a gestural score, but they can also be used to compare the stability of different coupling arrangements. This is tested through adding a “noise factor” such as random variation in speech rate (Nam & Saltzmann 2003: 2254), and seeing how much this disrupts the gestures’ relative timing. Simulations can also yield a measure of stabilization time, or how long it takes for the stable pattern to emerge. For example, Nam (2007: 497) carried out simulations in which a consonant was considered to have separate gestures for its closure and its release. He found a faster stabilization time for the consonant’s closure-release phasing in CV syllables than in VC. Faster stabilization time is assumed to correlate with faster planning time, from the speaker’s point of view, and it is assumed that sequences with faster planning time will be preferred because they are easier to produce.

4.1. Coupling and syllable structure

There are several reasons to think that in-phase and anti-phase coupling may define the difference between syllable onsets and syllable codas.

First, this fits with the results of many articulatory studies (Browman & Goldstein 1988, Honorof & Browman 1995, Marin 2013, Pastätter & Pouplier 2014). It has been observed that gestures in a syllable onset tend to start about simultaneously, both with one another and with the vowel gesture. Short lags of up to 50 ms. or so are common, but the numbers trend toward zero, as would be expected if they are in-phase. This is seen above in (12), where the TB gesture of the vowel begins around the same time as the VEL wide gesture and the LIP closure gesture. Coda gestures, on the other hand, start partway through the vowel gesture, consistent with an anti-phase relation, and if there are multiple gestures in the coda they tend to spread out rather than be produced simultaneously.

Second, onsets in real speech tend to show less variability in their timing than codas (Byrd 1996). This fits with the finding that in-phase timing is typically more stable than anti-phase (Haken et al. 1985, Goldstein et al. 2006). Nam & Saltzmann (2003) show through simulations that adding a noise factor causes greater variability in codas than in onsets.

One intriguing implication of this approach is that it offers a new possible explanation of the well-known typological generalization that onsets are cross-linguistically preferred over codas. If onsets reflect in-phase coordination and in-phase coordination is easier (as studies outside linguistics propose), then it is not surprising that all languages allow onsets while many ban codas. It may also help explain why codas are typically acquired later by children; why the inventory and frequency of codas is typically lower than that of onsets; why onset-nucleus combinations are very free while nucleus-coda combinations are often constrained; and why VC#V sequences are frequently resyllabified to V.CV. As Nam (2007: 489) observes, these patterns can all be captured with the generalization that languages prefer to maximize synchronous (in-phase) coupling, while minimizing asynchronous (anti-phase) coupling.

It should be noted that this theory does not explain all aspects of syllable typology: for example, it does not explain why languages disfavour onsetless syllables. Nor is this the only functional advantage proposed for CV syllables. Ohala (1996), for example, makes the case that onsets are easier to hear. These explanations are not mutually exclusive, of course; they may be mutually reinforcing.

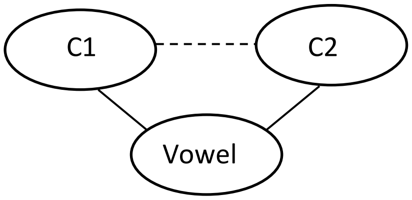
4.2. Complex onsets: the c-centre effect

The examples of coupling shown above were relatively simple, in the sense that none of the couplings were in competition with one another. But a more complicated situation arises when gestures have mutually conflicting coupling relations, as happens with certain combinations of multiple onset or multiple coda gestures.

As noted above, onset gestures are hypothesized to have an in-phase relation to the vowel. Having multiple gestures in-phase with the vowel is not a problem as long as those gestures can be produced simultaneously while still being perceptually recoverable. This is generally the case with the gestures that make up what is traditionally considered a segment, such as the tongue body and tongue tip gestures of an /l/, or the glottal opening and lip closure of a /p/.

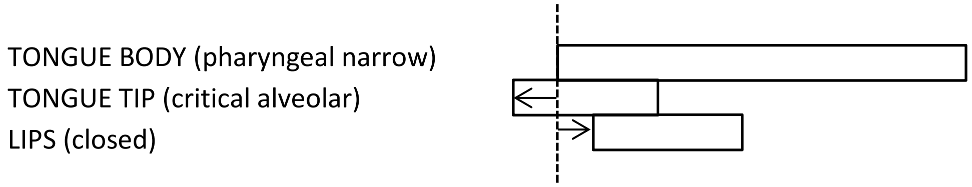
But in other cases, two onset gestures would not be recoverable if they were produced simultaneously. For example, if an onset contains a tongue-tip critical gesture and a lip closure gesture (as in *spa*), producing them simultaneously would cause the tongue-tip gesture to be acoustically masked by the lip closure. For both gestures to be recoverable, they must be in anti-phase relation to one another. This is shown below: the two consonantal gestures are coupled anti-phase with one another, but both are coupled in-phase with the vowel, since both are in onset position.

(13) Coupling relations in CCV onset



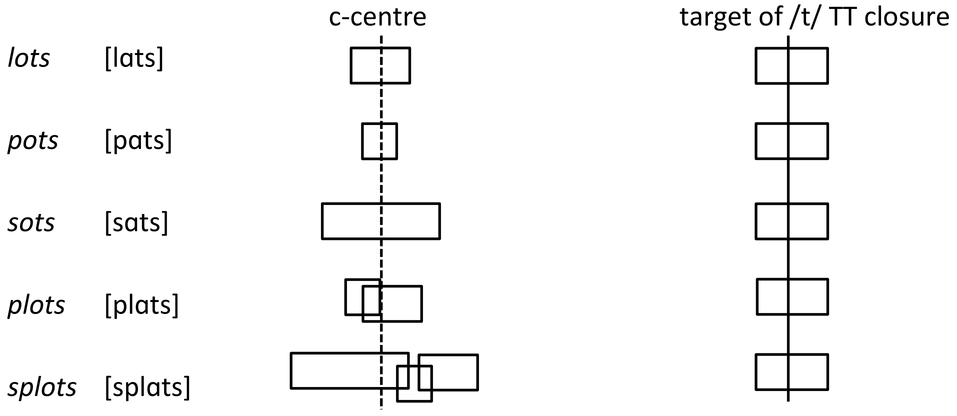
This coupling graph presents a problem: it is not possible for all three coupling relationships to achieve their target phasing. If both onset gestures are perfectly in-phase with the vowel, they cannot be anti-phase with one another. However, Nam & Saltzman (2003) show that the planning simulation still does arrive at a stable timing pattern, which is a compromise between the desired phasings. As shown below, the first consonant begins before the vowel, and the second consonant begins after the vowel. Neither onset gesture has exactly its preferred timing with respect to the vowel, as represented by the dotted line; each is shifted by about an equal distance, as represented by the arrows.

(14) Predicted timing of complex onsets; competitive coupled oscillator model



This is in fact what happens in real speech as well, at least in some languages. A series of studies of English (Browman & Goldstein 1988, Honorof and Browman 1995; Byrd 1995, among others) have shown that there is a stable relationship between the centre of the entire onset (whether it consists of one, two, or three consonants) and the rest of the syllable. This is known as the c-centre effect. For example, Browman & Goldstein (1988) compared x-ray microbeam records of an English speaker articulating words like *lots, pots, plots,* and *splots*. They measured the difference between each onset gesture and the target achievement of coda tongue tip raising of /t/. The coda gesture was chosen as an anchor because it was easier to identify than the target achievement of the vowel. The results are shown schematically below. With singleton onsets, there was a relatively consistent distance between the centre of the oral gestures and the coda; with complex onsets, the temporal midpoint of the whole onset fell at around the same point in time that the midpoint of a singleton onset would occupy. This is known as the c-centre effect, where "c-centre" (shown as a dotted line below) refers to the collective midpoint of the onset oral gestures. Computing the distance between other anchors, such as the right or left edge of the onset to the coda, yielded higher standard deviations.

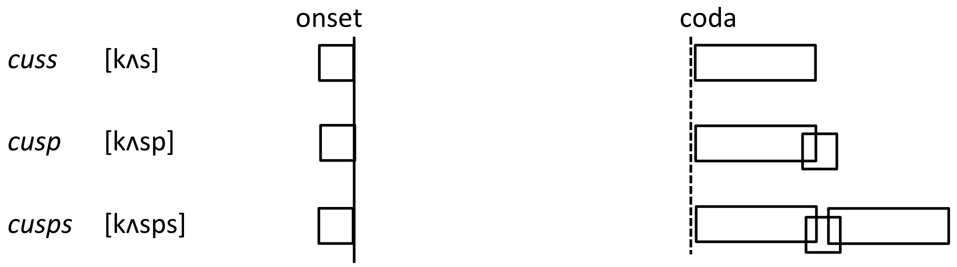
(15) Timing of oral gestures in onsets (Browman & Goldstein 1988)



The c-centre effect held only for consonants that formed an onset. In a phrase like *piece plots* [pis plɑts], the tongue tip critical gestures of the first /s/ did not participate in the c-centre effect with respect to the following syllable.

Codas do not appear to participate in c-centre effects. Several studies of English syllables (Browman & Goldstein 1988, Honorof & Browman 1995) find that in codas, there is a stable relationship between the left edge of the first coda consonant (shown as a dotted line below) and the rest of the syllable. As more consonants are added to the coda, the syllable simply becomes longer; the first coda consonant does not change its timing relative to the onset and vowel. This suggests that coda consonants are coupled anti-phase with one another, and that only the first coda consonant is coupled with the vowel.

(16) Timing of coda clusters (based on Honorof & Browman 1995)



However, Byrd (1995)'s study of five English speakers found some individual variation in the global timing of syllables, suggesting that not all speakers of a dialect necessarily use the same gestural organization.

It is an open question how many languages show this asymmetry in the timing of onsets and codas. The c-centre effect has been found for onsets in French (Kühnert et. al 2006), Italian (Hermes et al. 2013), and Georgian (Goldstein et al. 2007), but not for Slovak (Pouplier & Beňuš 2011: 18). Kochetov 2006 finds an onset-coda timing asymmetry in Russian, but it is different than the pattern in English. In some languages there seems to be variation depending on the type of cluster involved: Marin (2013) finds the c-centre effect in Romanian for sibilant-initial onset clusters, but not stop-sibilant onset clusters. Pastätter & Pouplier (2014) find similar patterns for sibilant-initial and sibilant-final onset clusters in Polish. They suggest that sibilants may be resistant to overlap with the vowel, and that this could disrupt the c-centre effect for specific clusters. Both Romanian and Polish codas show sequential organization similar to that of English codas.

Of course, some cross-linguistic variability in the organization of gestures is not surprising, given that phonologists have long argued that languages differ in how they syllabify similar strings of sounds. Given the relatively small number of languages examined to date and the centrality of this topic for understanding gestural timing, this is likely to remain a central area of research in AP for the near future.

4.3. Physical study of syllable structure

Phonologists do not always agree on the syllabification rules of particular languages. One intriguing implication of the AP approach is that disputes about syllable structure could be settled empirically, through articulatory data. If we hypothesize that the traditional notion of "complex onset" refers to consonantal gestures that participate in a c-centre effect, then this structural pattern can be detected experimentally. Under this view, cross-linguistic differences in the timing of CCV sequences, discussed in the previous section, would be equivalent to cross-linguistic differences in the syllabification of such sequences. Of course, it remains to be shown that the c-centre effect consistently correlates with traditional, distributional diagnostics of complex onsethood. Yet several early results support the idea that it may.

In Moroccan Arabic, for example, there is controversy over the syllabification of consonant clusters in words such as /kra/ 'rent'. While some phonologists assume that /kr/ is a complex onset, there is evidence (especially from oral poetic meter) that only the /r/ is an onset consonant, and the /k/ has some other status. Proposals vary as to whether it is a "minor syllable," a syllable nucleus, or is licensed by a mora (see Shaw et al. 2009 for background). Shaw et al. (2009) studied the articulation of such sequences using EMA, comparing the results to simulations using TaDA. They were particularly interested in patterns of temporal stability, which is generally strongest within syllables. The timing patterns found for prevocalic CC clusters were most consistent with the hypothesis that C1 was a syllable nucleus, rather than a complex onset.

Goldstein et al. (2007) compared CV, CCV, and CCCV sequences in Georgian and Tashlhiyt Berber, using EMA. They found Georgian shows the c-centre effect, while Berber does not. In Berber, words like /mun/, /s-mun/ and /t-s-mun/ ('accompany', 'cause-accompany', '3fs-cause-accompany') all had the same relative timing of /m/ to /u/. Georgian is traditionally analysed as having complex onsets, while Berber is usually analysed a having only single-C onsets, so the phonetic findings accord with other evidence that these languages organize sounds differently.

These techniques can also be used to compare the organization of different gestural clusters within one language, as in a recent study of "impure s" in Italian. There are various arguments, both distributional and psycholinguistic, that Italian word-initial /sC/ clusters are different than other Italian CC clusters. For example, they condition a special allomorph of the definite article*: il sale, il premio* but *lo studente*. Using EMA, Hermes et al. (2013) show that the c-centre effect holds for initial clusters such as /pr/, but not for /sp/. In /prima/, the /r/ shifts rightward compared to /rima/, but in /spina/, the /p/ has the same timing as in the name /pina/. This finding fits with other evidence that /s/ is not part of the syllable onset.

4.4. Moraic structure

The coupling model may shed light on another long-standing puzzle about syllable structure: why do coda consonants contribute to syllable weight in some languages but not others, and why do onset consonants never contribute to weight? One possibility is that the phonological patterning associated with "moraic" codas relates to a kind of timing relation. Nam (2007) attempts to model the difference between moraic and non-moraic codas. In his approach, every oral constriction involves a coupling of two gestures: a closure gesture and a release gesture. Nam hypothesizes that in languages where coda consonants add weight to the syllable, the vowel is coupled only with the closure gesture of the coda. In languages where coda consonants do not add weight, the vowel is coupled with both the closure and release gestures. The multiple couplings increase overlap between the vowel and coda, resulting in a shorter syllable. This proposal has yet to be rigorously tested against a range of languages, but offers an interesting hypothesis as to how gestural timing could relate to traditional notions of syllable weight.

5. NEW DIRECTIONS FOR THE COUPLING MODEL

5.1. Modelling phonological acquisition

Although the main application of the coupling model has been to understanding patterns of gestural organization cross-linguistically, several researchers have argued that the model also makes predictions about the acquisition and processing of phonological structure. As noted above, the model proposes that gestural scores are produced through a planning simulation, in which gestures oscillate repeatedly until they settle into a stable pattern of coordination ("entrainment"). Different gestural coupling structures require different numbers of oscillations to reach this stable phasing. It has been suggested that the time required for entrainment of a particular structure is a prediction both of how difficult speakers find it to acquire the structure, and how long it takes them to plan the production of the structure.

Nam et al. (2009) simulates acquisition of syllable structure in a Hebbian learning model in which a child agent tunes its initially random phase representations to match the perceived relative phase in adult productions. The adult's productions were varied across languages (for example, there are more tokens of codas in some simulations than others), to simulate the environment of languages with different frequencies of particular syllable structures. It was found that the child's CV phasing always stabilized faster than VC phasing. The lag is greatest in simulations where the adult produces more CV tokens, but strikingly, it persists even if the adult produces more VC tokens than CV tokens. This suggests that the greater ease of learning in-phase coordination can overcome even a paucity of such tokens in the environment. However, when VCC and CCV structures are added to the simulation, after acquisition of CV and VC, the child agent is quicker to master VCC than CCV. This is counterintuitive based on the idea that codas are "marked," but it follows from the fact that (in the simulation) VCC has a simpler phasing structure than CCV. CCV involves two in-phase couplings that are in competition; VCC involves two non-competitive anti-phase couplings. The simulation accords with reports that children have been found to acquire complex codas before complex onsets in some languages (see references in Nam et al. 2009: 2).

AP may also help explain why certain CV combinations are favoured in acquisition, and more frequent in the adult lexicon (Goldstein et al. 2006, Giulivi et al. 2011). During the babbling stage, children tend to produce CV syllables where the overlapping gestures are mechanically independent, like the lip and tongue body gestures in /ba/, or involve constrictions in similar locations, such as the two tongue body gestures in /gu/. Giulivi et al. (2011) use TaDA simulations to identify the most "synergistic" CV combinations, where synergy means that the final tongue body configuration for the C and V are similar. They argue that this measure of synergy predicts how easy it will be to produce the C and V in-phase.

5.2. Gestural coordination and morphological structure

The coupling model may offer a new approach to the phonology of morpheme boundaries. There are some indications that gestures belonging to a single lexical entry are coordinated in a different way than gestures that belong to different morphemes.

Cho (2001) used EPG to study morphologically simplex and compound words in Korean. He found that a sequence such as [ti] showed more variability in the relative timing of the oral gestures associated with

/t/ and /i/ when it was heteromorphemic, as in /mat-i/ ‘the oldest’, than when it was mono-morphemic, as in /mati/ ‘knot’. A similar difference is found between lexicalized and non-lexicalized compounds. Cho proposes that this is because the timing relations are lexically specified in ‘knot’, where the /t/ and /i/ gestures are part of single lexical entry. In Cho’s Optimality Theoretic (OT) analysis, the lexically specified timing relation is stronger because it is protected by IDENT constraints.

Nam et al. (2003), in a non-OT analysis, demonstrate through simulations that this difference variability could follow from the different coupling patterns below. When an intervocalic C is not with the following V, the two will show more inconsistent timing than if they had a specified coupling relation.

(17) Coupling within and across morphemes

/ati/ (same morpheme) at-i (different morphemes)

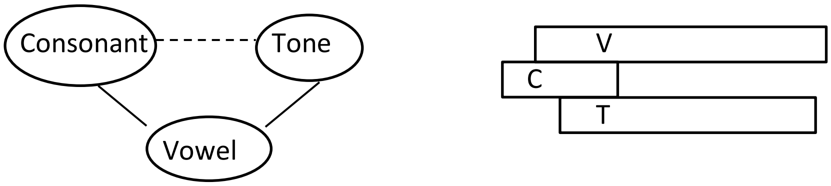
It is possible that even some categorical morphophonological alternations could be re-analysed as effects of coordination. Goldstein (2011) reanalyses English past tense allomorphy, a classic case of apparent segmental alternations, in terms of gestural coordination. He suggests that the three reported past tense allomorphs (-t, -d, - ɨd) actually consist of the same gestures, namely a TT closure and release, plus a VEL closure to prevent nasality. Through simulations, he shows that a natural-sounding output can be achieved for words like *nabbed* and *napped* [næbd, næpt] by coupling the TT release gesture of the suffix to the release gesture of the preceding consonant. In *napped*, the glottal opening gesture associated with the /p/ inhibits voicing on the suffix (to a lesser extent, the same happens with *nabbed* simply due to the length of the closure). Under this proposal, there is no phonological alternation and no allomorphy in such words: whether the suffix sounds like [-t] or [-d], it consists of the same gestures in the same coupling relations. As for the [-ɨd] variant, Goldstein proposes that the suffix still consists of the same gestures (with no vowel gesture for the [ɨ]), but they participate in a different coordination relation with less overlap between C’s. This creates the percept of a transitional targetless vocoid (as discussed relative to schwa in Section 3.2). EMA and real-time MRI studies confirm that tongue body shapes during the [-ɨd] suffix are consistent with lack of TB target (Smorodinsky 2001, Lammert et al. 2014). Goldstein’s simulations show that if the tongue tip closure of the suffix is coupled with stem-final closure gesture, no vocoid occurs; but if the tongue tip closure of the suffix is coupled to the stem-final release gesture, a vocoid does occur.

5.3. Coupling models of tone and intonation

Recently, the coupling model has been extended to account for tone and intonation. This area was pioneered by Gao (2009)’s work on Mandarin tone, which argues that tonal movements can be described as gestures, and that tonal gestures engage in phasing relations with one another as well as with vocalic and consonantal gestures. Just as a typical C or V gesture is a task of reaching a constriction target, a tonal gesture is a task of reaching a tonal target. Unlike other constriction gestures, the presence of a tonal gesture can be read directly from the acoustic record. A High tone gesture, for example, begins at the onset of a pitch rise, and ends when the highest point is achieved. In giving tones a duration, this approach differs from autosegmental-metrical theory (Goldsmith 1990), in which tones are thought of as dimensionless points, with intervening time periods filled in by interpolation.

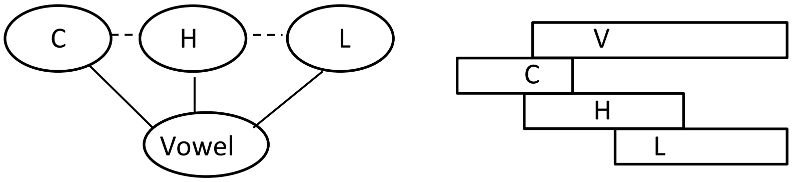
Gao (2009) (as described in Mücke et al. 2012) proposes that in a CV syllable with a single lexical Tone (T), such as Mandarin Tone 1 (High) and Tone 3 (Low), the T gesture behaves essentially like an additional onset C gesture. The C and T gestures are coordinated anti-phase with one another, but both in phase with V gesture. This causes a shift in alignment exactly analogous to the C-centre effect, so that the gestures are actually activated in the order C, V, T. The C and T gestures each begin about 50 ms from the V gesture (whose onset is identified from articulatory records as usual).

(18) Coupling relations in lexical tone on a CV syllable, after Gao (2009)



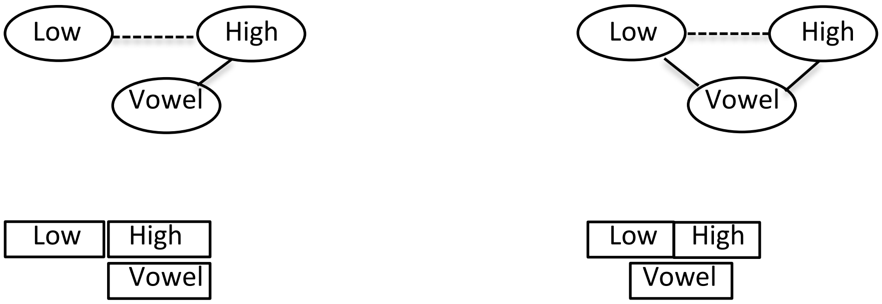
For complex tones, like Mandarin tone 4 (High-Low), the same principle applies. The gestures of C, H, and L are coupled anti-phase to one another, but in-phase to the vowel. The H gesture begins simultaneously with the V gesture, while C is pushed earlier and V later.

(19) Coupling relations in complex lexical tone



Gestural phasing can also be used to analyse the coordination of intonational tones. Mücke et al. (2012) offer a gestural analysis of intonational rises in Catalan and German stressed syllables. They found that in Catalan, the C, V and T gestures start about simultaneously --- there was no C-centre effect like that Gao (2009) found in Mandarin. In German, C and V begin together but the rise starts much later. Modelling with TaDA shows that the difference between German and Catalan can be captured by different couplings: in both languages, L and H tones are anti-phase with one another and H is in-phase with V; in German L is also in-phase with V; while in Catalan there is no coupling of L and V. The graphs below show the gestural scores predicted by each coupling relation.

(20) Coupling relations and gestural scores in intonational tone, after Mücke et al. (2012)


(a) Catalan (b) German

Cross-linguistic differences in the realization of intonational rises may also result from different tonal compositions. Niemann et al. (2011) argue that while rises in German reflect two gestures, low and high, as shown above, rises in Italian reflect only a single high gesture.

Mücke et al. (2012) propose an interesting hypothesis: they suggest that effects like that in Mandarin, where a tone participates in the C-centre effect, are only likely occur in lexical tone systems, where coupling between tone and non-tone gestures is represented lexically. Non-lexical tones, like those of German and Catalan, are unlikely to affect within-syllable coupling relations.

6. SUMMARY

Over the past thirty years, the AP approach has been applied to an increasingly wide range of problems in sound structure. Although the number of languages studied still remains small, and many topics such as morphology and intonation are only beginning to receive attention, recent work in these areas show promise for the development of a more comprehensive model of speech, including cross-linguistic variation. AP research holds itself to unusually rigorous empirical standards, generally demanding that analyses be based on precise articulatory records and computationally explicit simulations of speech production. It is unique among phonological frameworks in the extent to which it draws on, and participates in, a wider tradition of work on biomechanics. A biophysicist wandering into a linguistics conference would not recognize the abstract entities posited in most phonological frameworks (moras, archiphonemes, faithfulness constraints, etc.), but s/he would understand what it means to model speech movements as coupled oscillators.

7. FURTHER READING

Browman, C. P., & Goldstein, L. (1990). Tiers in articulatory phonology, with some implications for casual speech. *Papers in laboratory phonology I: Between the grammar and physics of speech*, 341-376.

One of the earlier works on Articulatory Phonology, this article introduces basic concepts, and demonstrates how casual speech processes can be described in terms of changes in gestural magnitude or overlap between gestures.

Gafos, A. I. (2002). A grammar of gestural coordination. *Natural Language & Linguistic Theory*, *20*(2), 269-337.

This article shows how Articulatory Phonology representations can be used within an Optimality Theoretic (OT) grammar. The grammar sketched focusses on the coordination of CC sequences in Moroccan Arabic.

Nam, H., Goldstein, L., & Saltzman, E. (2009). Self-organization of syllable structure: A coupled oscillator model. *Approaches to phonological complexity*, 299-328.

This article typifies the more recent Articulatory Phonology approach to syllable structure, arguing that CV syllables are unmarked because they result from in-phase coupling of C and V gestures. It provides a good introduction to the concept of coupled oscillators.

8. RELATED TOPICS

Chapter 23:	Laboratory phonology
Chapter 24:	Phonetically grounded phonology

REFERENCES

Bernstein, N. A. (1967). *The co-ordination and regulation of movements*. London: Pergamon.

Browman, C. P., & Goldstein, L. (1992b). Articulatory phonology: An overview. *Phonetica* 49(3-4), 155-180.

Browman, C. P., & Goldstein, L. (1992a). ‘Targetless’ schwa: An articulatory analysis. *Papers in Laboratory Phonology II: Gesture, segment, prosody*, 26-56.

Browman, C. P., & Goldstein, L. (1990). Tiers in articulatory phonology, with some implications for casual speech. *Papers in Laboratory Phonology I: Between the grammar and physics of speech*, 341-376.

Browman, C. P., & Goldstein, L. (1989). Articulatory gestures as phonological units. *Phonology*, *6*(02), 201-251.

Browman, C. P., & Goldstein, L. (1988). Some notes on syllable structure in articulatory phonology. *Phonetica*, *45*(2-4), 140-155.

Browman, C. P., & Goldstein, L. M. (1986). Towards an articulatory phonology. *Phonology*, *3*(01), 219-252.

Byrd, D. (1996). Influences on articulatory timing in consonant sequences. *Journal of Phonetics, 24*(2), 209-244.

Byrd, D. (1995). C-centers revisited. *Phonetica*, *52*(4), 285-306.

Cho, T. (2001). Effects of morpheme boundaries on intergestural timing: Evidence from Korean. *Phonetica*, *58*(3), 129-162.

Davidson, L., & Stone, M. (2003). Epenthesis versus gestural mistiming in consonant cluster production: an ultrasound study. In *Proc. of the West Coast Conference on Formal Linguistics* (Vol. 22, pp. 165-178).

Gao, M. (2009). Gestural coordination among vowel, consonant and tone gestures in Mandarin Chinese. *Chinese Journal of Phonetics*, *2*, 43-50.

Gafos, A. I. (2002). A grammar of gestural coordination. *Natural Language & Linguistic Theory*, *20*(2), 269-337.

Gick, B., & Wilson, I. (2006). Excrescent schwa and vowel laxing: Cross-linguistic responses to conflicting articulatory targets. *Laboratory Phonology*, *8*, 635-659.

Gick, B., Campbell, F., Oh, S., & Tamburri-Watt, L. (2006). Toward universals in the gestural organization of syllables: A cross-linguistic study of liquids. *Journal of Phonetics*, *34*(1), 49-72.

Goldsmith, J. A. (1990). *Autosegmental and metrical phonology*. Blackwell.

Goldstein, L. (2011). Back to the past tense in English. *Representing language: Essays in honor of Judith Aissen*, 69-88.

Goldstein, L. (1994). Possible articulatory bases for the class of guttural consonants. *Phonological Structure and Phonetic Form: Papers in Laboratory Phonology III*, 234-241.

Goldstein, L., Pouplier, M., Chen, L., Saltzman, E., & Byrd, D. (2007). Dynamic action units slip in speech production errors. *Cognition*, *103*(3), 386-412.

Goldstein, L., Chitoran, I., & Selkirk, E. (2007, August). Syllable structure as coupled oscillator modes: evidence from Georgian vs. Tashlhiyt Berber. In *Proceedings of the XVIth international congress of phonetic sciences* (pp. 241-244).

Goldstein, L., Byrd, D., & Saltzman, E. (2006). The role of vocal tract gestural action units in understanding the evolution of phonology. *Action to language via the mirror neuron system*, 215-249.

Haken, H., Kelso, J. S., & Bunz, H. (1985). A theoretical model of phase transitions in human hand movements. *Biological cybernetics*, *51*(5), 347-356.

Hall, N. (2006). Cross-linguistic patterns of vowel intrusion. *Phonology, 23*(03), 387-429.

Hanson, H. M., & Stevens, K. N. (2002). A quasiarticulatory approach to controlling acoustic source parameters in a Klatt-type formant synthesizer using HLsyn. *The Journal of the Acoustical Society of America, 112*(3), 1158-1182.

Hawkins, S. (1992). An introduction to task dynamics. *Papers in Laboratory Phonology: Gesture, Segment, Prosody, 2,* 9-25.

Hermes, A., Mücke, D., & Grice, M. (2013). Gestural coordination of Italian word-initial clusters: the case of 'impure s'. *Phonology, 30*(01), 1-25.

Honorof, D. N., & Browman, C. P. (1995). The centre or edge: How are consonant clusters organized with respect to the vowel. In *Proceedings of the XIIIth international congress of phonetic sciences* (Vol. 3, pp. 552-555).

Iskarous, K., McDonough, J., Whalen, D.H. (2012). A gestural account of the velar fricative in Navajo. *Laboratory Phonology*, 3(1), pp. 195-210.

Kochetov, A. (2006). Syllable position effects and gestural organization: Evidence from Russian.

Krakow, R. A. (1999). Physiological organization of syllables: a review. *Journal of Phonetics, 27*(1), 23-54.

Kühnert, B., Hoole, P., & Mooshammer, C. (2006). Gestural overlap and C-centre in selected French consonant clusters. In *Proceedings of the 7th Speech Production Seminar*.

Lammert, A., Goldstein, L., Ramanarayanan, V., & Narayanan, S. (2014). Gestural control in the English past-tense suffix: an articulatory study using real-time MRI. *Phonetica*, 71(4), 229-248.

Marin, S. (2013). The temporal organization of complex onsets and codas in Romanian: A gestural approach. *Journal of Phonetics*, *41*(3), 211-227.

Mücke, D., Nam, H., Hermes, A., & Goldstein, L. M. (2012). Coupling of tone and constriction gestures in pitch accents. In: P. Hoole, L. Bombien, M. Pouplier, C. Mooshammer, & B. Kühnert (Eds.), *Consonant clusters and structural complexity* (pp. 205–230). Berlin: Mouton de Gruyter.

Munhall, K.G., Vatikiotis-Bateson, E., & Tohkura, Y. (1995). X-ray Film database for speech research. *Journal of the Acoustical Society of America, 98, 1222-1224.*

Nam, H., & Saltzman, E. (2003). A competitive, coupled oscillator model of syllable structure. In *Proceedings of the 15th International Congress of Phonetic Sciences* (Vol. 1, No. 2253-2256).

Nam, H., Goldstein, L., Saltzman, E., & Byrd, D. (2004). TaDA: An enhanced, portable Task Dynamics model in MATLAB. *The Journal of the Acoustical Society of America*, *115*(5), 2430-2430.

Nam, H. (2007). Syllable-level intergestural timing model: Split-gesture dynamics focusing on positional asymmetry and moraic structure. *Laboratory Phonology*, *9*, 483-506.

Nam, H., Goldstein, L., Browman, C., Rubin, P., Proctor, M., & Saltzman, E. (2006). *TADA (TAsk Dynamics Application) manual*.

Nam, H., Goldstein, L., & Saltzman, E. (2009). Self-organization of syllable structure: A coupled oscillator model. In F. Pellegrino, E. Marsico, I. Chitoran, & C. Coupé (Eds.), *Approaches to phonological complexity* (pp. 299-328). Berlin: Mouton de Gruyter.

Niemann, H., Mücke, D., Nam, H., Goldstein, L., & Grice, M. (2011). Tones as gestures: The case of Italian and German. *Proceedings of ICPhS XVII*, 1486-1489.

Ohala, J. J. (1996). Speech perception is hearing sounds, not tongues. *The Journal of the Acoustical Society of America*, *99*(3), 1718-1725.

Pastätter, M., & Pouplier, M. (2014). The temporal coordination of Polish onset and coda clusters containing sibilants. In *Proceedings of the 10th International Seminar on Speech Production, Cologne, Germany, 5-8 May 2014*.

Pouplier, M., & Beňuš, Š. (2011). On the phonetic status of syllabic consonants: Evidence from Slovak. *Laboratory Phonology*, *2*(2), 243-273.

Pouplier, M., & Goldstein, L. (2005). Asymmetries in the perception of speech production errors. *Journal of Phonetics*, *33*(1), 47-75.

Proctor, M., Bundgaard-Nielsen, R., Best, C., Goldstein, L., Kroos, C., & Harvey, M. (2010, December). Articulatory modelling of coronal stop contrasts in Wubuy. In *Proceedings of the 13th Australasian international conference on speech science and technology* (pp. 90-93).

Ramanarayanan, V., Goldstein, L., & Narayanan, S. S. (2013). Spatio-temporal articulatory movement primitives during speech production: Extraction, interpretation, and validation. *The Journal of the Acoustical Society of America*, *134*(2), 1378-1394.

Saltzman, E., & Byrd, D. (2000). Task-dynamics of gestural timing: Phase windows and multifrequency rhythms. *Human Movement Science*, *19*(4), 499-526.

Saltzman, E. L., & Munhall, K. G. (1989). A dynamical approach to gestural patterning in speech production. *Ecological Psychology*, *1*(4), 333-382.

Shaw, J., Gafos, A. I., Hoole, P., & Zeroual, C. (2009). Syllabification in Moroccan Arabic: evidence from patterns of temporal stability in articulation. *Phonology*, 26, 187-215.

Smorodinsky, I. (2001). Schwas with and without active gestural control. *The Journal of the Acoustical Society of America*, 109(5), 2446-2446.

Sproat, R., & Fujimura, O. (1993). Allophonic variation in English /l/ and its implications for phonetic implementation. *Journal of Phonetics, 21*(3), 291-311.

Turvey, M. T. (1990). Coordination. *American psychologist*, *45*(8), 938.

Zsiga, E. C. (1995). An acoustic and electropalatographic study of lexical and postlexical palatalization in American English. *Phonology and phonetic evidence: Papers in Laboratory Phonology IV*, *4*, 282.

Word count: 11384