In Victor H.Yngve and Zdzislaw Wasik (eds.). (2004) *Hard-Science Linguistics,* Continuum, pp. 67-86.

Articulatory Events are Given in Advance

Douglas N. Honorof

1. Introduction

Since his early work on turn-taking (1970), Yngve has been asking how people communicate rather than how people 'use language' to communicate. Yngve redefines the problem in these terms because he is convinced that we have inherited a flawed rhetoric for talking about people as communicators – a rhetoric dependent upon ancient, but, ultimately, unworkable assumptions about linguistic objects. Yngve also questions the scientific adequacy of standard discovery procedures used in investigating people as communicators. Specifically, he discourages bottom-up approaches that begin with phonetics and end with pragmatics, challenging us, instead, to begin from careful observation of exchanges between people and work our way down only when a lower level of structure suggests itself. In discouraging bottom-up approaches, Yngve redirects our attention from 'linguistic communities' and the micro-level grammatical constructs their ideal speaker-hearers are widely held to share, toward actual individual people and their properties as communicators.

Yngve's shift of focus from language to people is perhaps best understood in historical context. Yngve (1996, Chapter 3) recounts how, during the early years of twentieth-century structuralist linguistics, real-world objects – the acoustics and physiology of acts of speech – came to be viewed as belonging, in Bloomfield's terms, to other sciences. Differences in speech behavior among individuals fared no better under Bloomfield's fundamental assumption: '... in every speech-community some utterances are alike in form and meaning' (Bloomfield 1933:78). In spite of much rhetoric about the 'scientific' advances linguistics was making, the structuralists robbed

the field of its primary tie to the observable world by excluding the study of acts of speaking and individual variation from linguistics. More recent structuralist models stemming from Chomsky's early work in transformational-generative grammar have perpetuated the divorce of 'performance' from 'competence', the former being very nearly neglected in practice and held to lie outside the realm of linguistics proper. Thus language, a 'logical domain' construct based on the arbitrary and field-specific assumptions of the grammatical-semiotic and normative-grammatical traditions, has come to be mistaken for a legitimate object of scientific inquiry. As modern linguists, we have busied ourselves defining and redefining the objects of our study, but, in Yngve's view, failed to recognize that science does not study objects of its own creation. Rather, he insists, science studies objects given in advance. Within Yngve's Human Linguistics, unlike the syntactic constituents, words, phonemes, and other units of traditional grammar, people are the real-world objects that exist in advance of our observation of them, and therefore constitute more suitable objects for scientific investigation.

I agree with Yngve that, in adopting the methods of formal logic and creating grammars that do not attempt to model real-world communicative behavior (and that therefore cannot be tested against behavior), we introduce a noticeable level of circularity into linguistic theory. Our inability to test hypotheses empirically has led Yngve to recommend that we reject all *ad hoc* grammatical building blocks. In this connection, I am especially intrigued by Yngve's rejection of the segmentation of utterances. He very rightly points out that segmentation is 'not inherent in the sound waves' (1996:32). While I make no claim to know more than the average phonologist about how speech is parsed into words, phrases, and sentences by the listener, like Yngve, I remain as yet unconvinced that phonemes are real units in the physical world (or perhaps even real in cognition; but see below). However, Yngve may be missing a key point here. Although researchers have met with considerable difficulty in attempting to segment sound pressure waveforms, laboratory phonologists around the world have met with considerable success in decomposing signals derived from the articulatory movements of people engaged in the act of talking. This being the case, I believe that it is possible to build a model of phonology that conforms to – and can be tested against – real-world patterns of human speech articulation. This is exactly what my colleagues at Haskins Laboratories have been doing in recent decades. I have been privileged to take part in some of that research. In the present

chapter, I take the reader on a brief tour of this work. Along our way, we will consider the scientific status of our work and the nature of the objects that we study.

2 Gestural events in the real world

2.1 Coordinative structures

Scholars have attempted to parameterize speech into individually manipulable features for as long as there have been phonologists. Although a feature-based approach to speech is indeed ancient in origin, the parameterization effort received a major boost when linguistic anthropologists and their colleagues began transcribing the speech of unfamiliar peoples into alphabetic notation for taxonomic, lexicographic, and pedagogical purposes. Even today, descriptivists outnumber theoreticians, though their work is no longer as well represented in the leading linguistics journals. Interest in features received another boost when linguists turned their attention to computational modeling of speech using symbols that could be entered from a standard keyboard, for example, for projects in machine translation, automated speech recognition, automatic speaker verification, speech synthesis, etc.

In recent decades, phonologists have emphasized the structuralist interest in contrast as a key to finding universal patterns in phonological (cognitive) systems. Most have not been particularly interested in the physics of speech except insofar as it can corroborate independently motivated theoretical stances. The nature/nurture controversy has driven the quest to find universals in the handsome collection of phonological data we have amassed as a community of scholars, which has further driven an interest in parameterization.

A number of my colleagues at Haskins Laboratories have been rethinking the problem of parameterization of speech in recent years. After observing the physiological properties of people in the act of moving their vocal organs, they have adopted a viewpoint that originated in work on visual perception within the framework of specificity theory which has evolved into the ecological approach to perception (Gibson and Gibson 1955; Gibson 1979; Gibson and Pick 2000). Ecological psychologists proceed from the understanding that, like other animals, we live in environments that we have to know about in order to function effectively. To know about our environment, we must perceive whatever is in it – at least whatever

in it exists at roughly the human scale. These theorists hold that we learn to perceive the layouts (that is, permanently arranged surfaces), objects, and events (that is, movements and actions of objects) in our environment without constructing intermediary mental representations of them. In this sense, perception is believed to be direct.

The notion that we visually perceive objects rather than patterns of light may not strike the reader as odd. However, some objects move or are moved and therein generate sounds. When they do, the ecological psychologist says that we hear objects in motion, as well. The assertion that we hear objects in motion rather than patterns of sound often strikes us as counterintuitive. I believe our discomfort with the notion that we directly perceive the sources of structure in sound waves follows from our biological proclivity to trust what we see over and above what we hear. We are visual believers and auditory skeptics by nature. Even so, if the reader will consider the parallels between optical and auditory signals, the parallels in perception will make sense.

For example, consider a windchime. Ecological psychologists will argue that, even when the chime is still, we perceive (that is, see) it directly, not a prototype of it, or a cognitive representation of it, or cues that suggest it, or codes from which we must infer the presence and properties of the chime. The situation is no different in the realm of perception by ear. Sound is normally generated by an action or collision that sets objects into motion. The windchime is an object given in advance that generates a sound when struck by wind in space over time. Such dynamic (spatio-temporal) events are as clearly given in advance as the objects involved. We perceive the windchime directly by eye and we perceive its movement by eye and ear just as directly. The movement of the chime's colliding parts structures the patterns of light and sound that reach the human retina and eardrum. These patterns of light and sound contain abundant information that specifies the properties of the source of the disturbance. Thus the proximal stimuli (patterned light and sound waves) convey information about distal events (the causal source of the patterns of light and sound) to the visual and auditory perceptual systems respectively.

An ecological approach to speech perception begs the question 'What is out there when we speak?'. What events structure the air? The answer must begin with production. Speech is generated by the coordinated movement of human objects: the articulators of the talker's vocal tract: tongues, lips, jaw, etc. Such speech events are abundant in our ecological niche. However, as early as the 1950s, investigators were noting difficulties inherent in segmenting wave-

forms and in synthesizing invariant phonetic percepts across contexts on the basis of acoustics. (For discussions of early findings within a Motor Theoretic perspective, see work by Liberman, *et al.* 1967 and Liberman and Mattingly 1985.) Evidence that the listener can perceive articulatory stabilities was also emerging. Thus those among my colleagues at Haskins who work within the Direct Realist theory of speech perception have concluded that patterns of coordination in vocal organ movement are exactly the real-world events that are out there (see Fowler 1991; Surprenant and Goldstein 1998). Articulators can be seen even when people are not moving them, but patterns of coordinated movement among articulators in space over time (also known as coordinative structures) can be seen and heard.

The specific attributes of the coordinative structures relevant to speech (that is, speech events or articulatory gestures) have been formalized in what has become the gestural theory of speech production (see Saltzman 1986; Browman and Goldstein 1995). Over the past two decades, my colleagues at Haskins and elsewhere have tested gesture-based hypotheses and amassed considerable evidence suggesting that gestures are indeed the public, real-world, task-directed, spatio-temporal events of production and perception.

The Haskins group has developed a model of gestures – in our view, a model of real-world events, not purely a model of theoretical events. The gestural model successfully predicts the spatial and temporal properties of gestures observed when the coupled articulators of real people work together synergistically to bring structures in the vocal tract into approximation (at present, along the two-dimensions of the midsagittal plane only). In most cases, the computational component of the gestural model plots to a computer screen the coupled articulator trajectories that correspond to the synergistic constricting action of two or more articulators. For example, in outputting a bilabial closure gesture, the model plots lip aperture curves involving the coordinated action of the upper lip, lower lip, and jaw. The model does not simply produce the spatio-temporal trajectories of the individual articulators. Therefore, the model can be used to test predictions about events against the events themselves.

As indicated by the example of the bilabial closure gesture just given, it is important for the reader to bear in mind always that gestures differ from raw movement curves. In the gestural model, two gestural events are deemed equivalent if they share the same spatio-temporal target, even though they may actually be achieved by different relative contributions of the articulators involved. In

the example given above, with a goal of achieving a target such as bilabial closure, in one instance the jaw might contribute more than the lower lip and the upper lip remain nearly still, but in another instance, the upper lip might do nearly all the work. The stability is in the goal. Instance-by-instance differences in gestural movement curves observed in space over time can occur as a consequence of differing degrees of overlap between neighboring gestures (coarticulation) due to different speaking styles or rates. Or, two articulator trajectories for a unitary gesture might vary because the gestures are produced in differing gestural contexts. Differences might even occur simply because observed combinatorial gaps among gestures make speech so redundant that undershot gestural targets are, in many cases, recoverable by the listener. Even though the gestural movement curves may vary from instance to instance, the gesture is identical to both talker and listener. The same cannot be said of purported acoustic correlates of, or cues for, phonemes, for which stable acoustic patterns have been difficult to identify in speech records, and for which complex rules of phonetic realization have been even more difficult to formulate. In gestural terms, there is neither derivation, nor generation, nor implementation. There are no abstract underlying units that must be realized – the gestures are at once units of perception, action, and cognition. They are always present during the act of talking and listening. Gestural movement curves lawfully produce acoustic consequences, but we seek invariant patterns in the synergistic behavior of real-world articulators.

The Haskins computational model has been applied successfully in testing numerous hypotheses about gestures for well over a decade. The program uses task dynamics (Saltzman and Munhall 1989) to model gestures. While much remains to be understood about the physiology behind coordinated human movement, task dynamics does appear to do a reasonably good job of approximating complex control of the anatomical structures believed to be most directly relevant for speech. The mathematics are beyond the scope of the present chapter, but see Hawkins (1992) for an accessible introduction to the equations involved. Although the mathematics are somewhat complex, conceptually, the model is fairly straightforward. Here is how it works.

Gestures have their own internal equations modeled after the workings of a critically damped point-attractor system. Gestures are not sequenced. Rather, a certain point within one gesture (corresponding, say, to achievement of target) is phased temporally with

respect to a certain point within another gesture (corresponding, say, to release of target). The researcher wanting to test a prediction about gestural organization lays out the predicted gestures in a non-linear fashion on a multi-tiered grid (minimally, one tier for velic gestures, one for oral gestures, and one for laryngeal gestures). This grid is known as a gestural score. Once the user has composed a gestural score, the score is run through the computational model where gestural movement curves are generated. Under an analysis-by-synthesis strategy, these movement curves can be compared with actual curves derived from individual-articulator trajectories collected, for example, as articulometer 'subjects' (real people) talk with transducers affixed to flesh points along the midline of the vocal tract (Perkell *et al.* 1992). Finally, the movement curves may be input to an articulatory synthesizer for generation of sound that can be played for listeners in naturalness or perceptual tests. Articulatory synthesis can also be used in a more exploratory way. The configurable articulatory synthesizer is essentially a midsagittal talking head. The head's two-dimensional articulators can be manipulated geometrically on the computer monitor, and cross-sectional vocal tract area functions can be computed and acoustic signals generated for the listener.

At this relatively early stage of development, the model may be overly simplistic in the details of the way it specifies the internal dynamics of gestures, as may be the patterns of intergestural coordination that we specify (see Mattingly 1990). For instance, it is possible that some parameters of off-midline articulation about which we know very little are actually important in the formation of gestural events. It is also possible that we have constrained our initial observations of speaker behavior too narrowly even in the mid-line and have therefore missed relevant aspects of vocal tract constrictions. It is also possible that our theory is essentially right-headed, but that our mathematical-computational model will need to be tweaked in order to produce correct output with respect to a particular prediction. However, no matter how far we are from having captured an accurate picture of gestures and gestural organization, it is my belief that gestural events do exist in the real world, and that we will get better and better at modeling them. It is also possible that we are mistaken in asserting that gestures are events for the speaker and hearer, but it seems to me extremely unlikely that gestures do not exist at all.

I have used 'gestures' to refer to real-world events and to phonological units in our theoretical representation of those events. The scientific justification for our work as a theory of the real world

and not a theory of theories lies in our ability to test the articulatory output of the computational model against gestural movement curves derived from articulatory events in the real world. In the next section, I describe the nature of the data we consider in deriving such gestural movement curves.

2.2 What count as data

Within mainstream phonology, new theories are minted and fall almost instantly into wide circulation on average every decade. In the 70s we had generative phonology. In the 80s we had autosegmental and metrical phonology, which evolved into a very popular nonlinear framework: feature geometry. In the 90s we saw the adoption of optimality theory and of related logical-domain theories involving constraints on output. Each theory emerges out of the insufficiency of an older model to deal with a particular set of data elegantly. However, with each round of new theories, the frenzy that follows has been the same: reanalyze everything (or at least everything of interest within the new model). We dig through old papers, extract the data that were once well 'explained' and apply the new model to the data hoping to discover whether the new model is indeed just powerful enough.

Articulatory Phonology, the arm of the gestural school that concerns itself with universal and nonuniversal patterns in gestural organization, has followed a different path. While it is true that Articulatory Phonology also arose to account for a specific set of problems (especially postlexical assimilations, epenthesis, and elisions in casual speech), its proponents have generally addressed novel questions on the basis of novel data. This is as it must be. Traditional phonological 'data' are simply transcriptions of acts of speech. Transcriptions are highly problematic. First, they are observer-dependent. Second, transcriptions impose phonemic (that is, letter-sized) units on streams of speech. We have yet to find firm evidence in acoustics or physiology for a phonemic unit of analysis. Interestingly, even the universality of the phoneme as a 'mental' object has been called into question by what little data there are on phonemic awareness among illiterates and among adults literate only in nonalphabetic orthographies. To be fair, a lack of phonemic awareness does not necessarily mean that phonemes are not 'psychologically real'; it may be the case that people who do not read alphabetic orthography are simply unaware of units that they 'know' in cognition. (See Read, *et al.* 1986; Adrian *et al.* 1995.)

Nongestural, phoneme-based theories require complex, *ad hoc* rules of phonetic interpretation (interpolation, translation, instantiation, realization, implementation) to predict the phonetic quality of segments in output and to imbue segments extrinsically with temporal information so that they can be realized in production (Fowler 1980). It has long been a working assumption of standard structuralist grammar that such rules for phonetic realization of static segments can be written successfully, though few researchers have bothered. If solid evidence for a unit akin to the phoneme does emerge some day, it may be entirely possible to reconstruct something like (but not identical to) the notion of the phoneme as a constellation of gestures. Goldstein and Fowler have suggested the chemical term 'ion' for such a possible set of bonded gestures (Goldstein and Fowler, in press). Ions would be available to recombine with other ions into a large number of compounds (words or syllables). Such a gesture-based treatment of the segment would not require rules of phonetic interpretation – a definite plus. However, the key point is that, if it ever becomes necessary to posit an ion to account for apparently phonemic behavior, we would, nevertheless, not be justified in counting phonemic transcriptions as data. Goldstein's argument is that, just as the evidence for ions in chemistry comes from empirical investigation of chemicals, not from our intuition that sodium carbonate and calcium carbonate have something in common, so the evidence for gestural 'ions' in speech would have to come from real-world investigation of speech. The difference in the nature of the data in Articulatory Phonology versus traditional structuralist phonology cannot be overemphasized, particularly given that the Haskins work on casual speech, and more recently on speech errors, reveals how transcriptions can be systematically misleading as to the actual real-world properties of some speech events (Pouplier 2003). Clearly, even if we do someday find ourselves adopting a phoneme-like unit as a construct, we will not simply turn to digging up old transcription 'data' and reanalyzing them without collecting articulatory data. Gesture theory considers only real-world movement as data.

While, for us, data are observed in the real world, they are not necessarily observed under as natural a circumstance as are some of the data Yngve has collected. In other words, although we study people in the act of talking, we do not generally study people in the act of communicating spontaneously. Our movement data are collected in the laboratory, though our experimental designs do not exactly simulate what Yngve would call linkages. In our designs,

talkers are usually asked to produce multiple repetitions of gestural constellations, often by reading aloud. These constellations might correspond to units of analysis from traditional grammar (which are rather easy to elicit from literate talkers) or they might be constellations that the subject has never before produced (nonsense utterances). The repeated acts of speech (no matter how variable from repetition to repetition) form the basis of discrete, replicable, observer-independent spatial and temporal measurements, for example at articulator velocity peaks or zeros, at peaks in articulator acceleration, or at extrema in articulator displacement. These measurements are subsequently subjected to conventional statistical testing. Here, as in all hard sciences, statistics provide a basis for testing the null hypothesis, and for drawing conclusions about an individual's behavior that may, in principle, with a large enough sampling of the population, generalize to the group.

2.3. Physiological data acquisition

Some gestures can be seen with the naked eye – those produced in the anterior regions of the vocal tract – but most gestures are hidden from view. However, even gestures that we cannot easily see can be studied and measured with the right tools. The measurements made depend on the data collection device used. Gesturalists have tended to acquire laboratories full of unusual instruments for use in physiology experiments. Collectively, we have used video cameras, SELSPOT optical motion analysis systems (Innovision Systems, Warren, MI/USA), the velotrace (Horiguchi and Bell-Berti 1987), magnetic resonance imaging, ultrasound imaging, cineradiography (where local jurisdiction allows), electromyography, point-source tracking (x-ray microbeam [Nadler *et al.* 1987]), and, especially, electromagnetic articulometry (e.g., Perkell *et al.* 1992, etc.). There are also laboratory techniques for measuring laryngeal activity such as electroglottography (see Scherer *et al.* 1988) and laryngoscopy (direct and technology-assisted) and subglottal coordination indirectly (Pneumotach Mask [Glottal Enterprises, Syracuse, NY/USA]) or somewhat more directly (Respitrace [Ambulatory Monitoring Inc., Ardsley, NY/USA]). Some of the foregoing techniques and devices are borrowed from, or inspired by, clinical practice. A few have even been designed specifically for purposes of speech research. A further step removed from movement, but sometimes also instructive, are contact patterns between articulators such as can be measured through palatography. Wandering even further from the direct

observation of movement, we routinely examine the lawful acoustic consequences of gestures. Finally, the effects of gestures on listeners can be measured – listeners themselves being real-world objects.

3. Slow but steady progress

3.1 Coarticulation and individual variation

Considering the relative youth of the gestural endeavor, the techniques we have available to us are truly impressive, and, yet, they are nevertheless crude in comparison to the task of studying the complexities of speech physiology. Perhaps that is why we collect so many new devices; the old ones do not suit our purposes. Because our instruments are so crude, we often come close to wrongly accepting the null hypothesis. There certainly are repeating patterns in speech to be found, but those patterns appear to be hidden in a mire of coarticulation – necessary 'noise' that our crude instruments alone cannot always see through. In order to give our data collection devices a leg up, we often must collect pilot data several times before we have analyzable data – data that allow us to tease apart the effects of gestural co-production well enough to find evidence of an individual gesture. Some combinations of gestures obscure each other in output so badly that we must simply give up and ask another question.

Even once we have a well-designed stimulus set, we are often faced with the hairy problem of inter-speaker variation. As happy as we are that we can identify the voice of a familiar caller on the telephone, in the laboratory idiosyncratic differences in speech habits can make it difficult to draw conclusions quickly. Some of these individual talker differences in behavior can arise from genuine idiolectal (that is, 'personality') differences. We try to screen talkers for membership in homogeneous populations with respect to at least the style of speech the experimental instructions imply. However, in my experience in the laboratory, the extent of individual variation even among well-screened subjects is surprisingly high. Collecting data from many subjects often helps idiosyncrasies to 'wash' in the statistics, but physiological data collection and analysis are time-consuming and costly. Conventional funding structures simply do not encourage it. Even when we do have a large number of subjects, variation remains the rule. Yngve shows tremendous clarity in joining the sociolinguists to call into question the logical-domain notion of the ideal speaker-

hearer. Perhaps we must content ourselves to study the individual in multiple linkages more often than we study the group until our technologies speed up the research process.

3.2 Convergence

Although studying gestures does present special practical problems that slow down the entire enterprise, we are making genuine progress – producing occasional results that may even outlive our own productivity as researchers. Even some of the longstanding puzzles of traditional grammar have been addressed very elegantly by articulatory phonologists (Browman and Goldstein 1991). Chances are, intuitions about how people communicate cannot all be wrong. It is therefore not surprising that gesture-based findings and logical-domain phonology have converged on occasion, and that ideas found in literature authored by traditional phonologists have provided fodder for successful gestural investigation. Given that traditional grammar is not a hard science, convergence with it does not argue for or against the rightheadedness of a gestural approach.

However, there are other areas of convergence between the Haskins work on gestures in speech and work in other sciences that is very encouraging. Because the gestural model is rooted in ecological and task-dynamic approaches to the production and perception of human movement in general (locomotion, grasping, etc.), convergence there does meaningfully corroborate our findings. In addition, we see parallels between our findings on the gestural organization of speech and findings on the spontaneous emergence of order and complexity in other self-organizing systems in nature – systems that make potentially unlimited use of finite, discrete units in building larger structures (physics, chemistry, genetics, etc.). For example, without necessarily committing to any particular units of traditional grammar, Haskins researchers have begun to consider how a finite set of discrete gestural units might be organized into larger stable configurations such as consonant-clusters and syllables (Browman and Goldstein 1988; Honorof and Browman 1995; Studdert-Kennedy and Goldstein 2003). Such larger, stable configurations of gestures may be structured by real-world functional constraints. For example, some such patterns may naturally emerge from the competing requirements that a) speech events be sequential in order to be recoverable by the listener, and, b) speech events overlap in order to hasten the flow of information through parallel transmission (Mattingly 1981; Browman and Goldstein 2000). Any

convergence between the findings of gestural research and findings in other physical-domain fields (biology, physics, etc.) only serve to shore up the status of gestural work as a hard science (Ohala, 1990).

4. Does gestural research weigh in as a hard science?

4.1 Yngve's two criteria for weighing hypotheses or theories

I have introduced the reader to a theory in which spatio-temporal events in the vocal tract are held to be real-world events in production and perception. Let us now consider how well the gestural approach holds up to the two standard criteria of acceptance of hypotheses or theories in hard science laid out by Yngve (1996:99–100).

> Criterion 1. Theory driven, hypothesis-generated predictions 'pass tests against the real world by means of careful observations and experiments'.

We do subject our predictions about gestures to careful experimental testing. We expend considerable effort refining and calibrating our data collection devices, screening subjects, and presenting tasks to them in ways that do not prejudice behavior. We collect a large number of data points, measure them, and analyze them statistically. We reject hypotheses that do not stand up to testing.

> Criterion 2. Observational and experimental results are reproducible when questioned.

Our measurements are, whenever possible, automated, which makes them observer-independent. Even where algorithmic measurement is not possible, very strict measurement criteria are followed and published, allowing for replication by other research groups. There is not much funding for studies that aim solely to replicate results of other researchers, and the work we do often makes use of instrumentation that exists at very few other laboratories, but replication and extension of our studies certainly can be undertaken, and sometimes are. In any case, there are less costly devices on the market that can be used for confirming our articulometric findings using slightly different designs. Confirmation through similar means is even better than confirmation by replication, after all; true replication can duplicate methodological error.

4.2 Yngve's four assumptions underlying scientific work

Yngve also lays out the four, time-tested, commonsense presuppositions of hard science that we must take on faith (1996:101–02). Let us now examine whether we have made only the same four assumptions, or introduced any special assumptions.

> Assumption 1. 'There actually is a real world out there to be studied.'

Our procedures are based upon this ontological assumption. Gestures, though partly conventional (learned), are indeed out there. They are produced and perceived because they exist, and are not just convenient fictions. Gestural events take place independent of our theories and observation.

> Assumption 2. 'The real world is coherent so we have a chance of finding out something about it.'

Because we study human individuals as well as gestural events, and because individual behavior can be difficult to constrain even in simulated communicative situations in the laboratory, the world sometimes seems a little more chaotic to the experimenter than it actually is, especially when our sample size is small. This is a matter of frustration specifically because we share the regularity assumption with other scientists.

> Assumption 3. 'We can reach valid conclusions by reasoning from valid premises ... We can trust our ability to calculate predictions from our theories for comparison with the real world.'

We assert that gestures are real-world events (dynamical objects) produced by talkers and perceived by listeners. The mathematical definition of gestures under task-dynamic modeling allows us to calculate predicted movement curves that can be compared with actual movement curves obtained from talkers in the laboratory. So far, so good. However, there is a third element to the model: gestures are held to be units of production, perception, and phonology. This is where, at first glance, it might seem that we inch close to the edge of the hard science-traditional grammar border. We accept the rationality assumption, but, having tested our predictions against the real world, step back from the communicating individual and go on to ask questions about formal phonological patterning among the units of analysis themselves. The gestural model itself together with the phonological patterns we arrive at through informal observation

of speech and by reading the writings of scholars of traditional grammar inspire new hypotheses. We then subject predictions so derived to further laboratory testing. In this sense, we allow formalist work to inspire our prediction-generating process, but our commitment to behavioral data forces us to test our predictions against the real world. Our experiments rarely produce entirely unambiguous results, but even in the traditional hard sciences, this is to be expected.

Because gestural events exist, it is reasonable to assume that people have conscious or subconscious knowledge of them. We do not rely on 'native-speaker intuitions' to investigate that knowledge, however. The gestural units of which people have knowledge are spatial and temporal, so our knowledge of them is reflected directly in dynamic behavior. Even when we are looking at phonology, our predictions are tested in the laboratory.

Yngve warns us against assuming blindly that units such as phrases, words, phonemes, etc. exist and that people use them to communicate. He notes that such grammatical units belong to the logical domain until proved otherwise. However, we believe we have found strong evidence for gestural units in the physical domain, in particular in acoustic and, especially, physiological speech records. Thus we are on sound scientific footing in asserting that talkers and listeners learn to use such coordinative structures to communicate.

Once we have admitted to consideration the gesture as a spatio-temporal 'object' (that is, an event) that can be used by people in accomplishing a task, we are inclined to ask how such events are learned by the child and how they might have evolved. These questions bring us to the point where we can more meaningfully contribute to the nature/nurture dialog. Given that gestural units are subject to physical constraints, phonological learning need not necessarily imply language-specificity. To be sure, some gestural patterns are used by some groups of talkers who are able to communicate with each other (that is, who share some phonological knowledge), and not by others. But other gestural patterns may turn out to be universal. Gestural universals are bound to follow from the functional demands the real world places on the evolution of gestural communication over time (Studdert-Kennedy and Goldstein, in press), and in the individual user attuning to the environment (Goldstein and Fowler, in press). This is not to say that the environment always structures the human organism. Some real-world demands on gestural communication may emerge from our own species-specific auditory and neurological anatomy and physiol-

ogy, in which case we may be structuring the environment of the learner of spoken gestural systems of communication by placing constraints on the evolution of gestural events.

Yngve has criticized prominent linguists for their skin-deep allegiance to hard-science linguistics – an allegiance born of a desire to appear scientific, but lacking in the commitment to build models that can be subjected to external validation. The idea is that linguists are not honest in admitting that they are logicians, and that they go so far as to borrow scientific rhetoric to argue points based on intuition or on purely logical assumptions about language. At Haskins Laboratories, if anything, we suffer from the opposite type of confusion of identity. We actually borrow logical-domain rhetoric to talk about hard science. Although we test our theories in the physical domain, we are constructing a phonological theory that resembles, in some respects, the soft science of the traditional logical-domain structuralist. Furthermore, we report our results in mainstream linguistics journals using many terms borrowed directly from the traditional study of language. Doing so allows us to engage the larger Linguistics community, and to benefit from the insights of its great minds, even though we may make different assumptions about what count as data, how predictions may be generated, and what counts as a good test of a model.

Given that our explanations tend to be very tightly constrained by the real world, our peers in traditional linguistics often think of us as functionalists. In my view, form and function are related, but the forms themselves are also of interest; we treat phonetics *and phonology* in a unified manner. Doing so may, in fact, make us *structuralists*, but clearly we are structuralists with a difference. The Haskins work involves units of analysis that are at once theoretical constructs and mathematical predictors of real-world, gestural movement curves – not special-purpose, theoretical objects from the logical domain. Our units are given in advance, but can also serve as playthings for logicians.

> Assumption 4. 'Observed effects flow from immediate real-world causes.'

We observe movement curves and infer gestural organization. Clearly, we accept the causality assumption.

5. Summary

The present chapter does not aim to enlist support for gesture-based

work over Human Linguistics or vice-versa. Rather, I have simply described my personal perspective on a laboratory-based research program that shares important features with Human Linguistics. The gesturalist approach I describe meets the two criteria and four assumptions of hard science set forth in Yngve's 1996 book. Our work has proved to be slow going at times, but nearly always profitable.

At the crossroads of behaviorism and structuralism, we sidestep the mind-body and performance-competence dichotomies. To my way of thinking, the ideal linguistics uses hard-science methodology to discover events in the real world that also help structure human perception and cognition. If linguistic events occur in the physical domain and we are able to perceive them, it only makes sense that learners should use them to build cognitive structure. In working both top-down and bottom-up, my colleagues have found ample empirical evidence for just such a real-world (spatio-temporal) object given in advance – the gesture. These gestural units of perception and action were discovered by studying the behavior of objects given in advance – people.

Nevertheless, we gesturalists sometimes frame our arguments in terms borrowed from traditional structuralist grammar. At times, structuralist techniques are even borrowed to help us manipulate variables in the laboratory as we attempt to simulate measurable and quantifiable communicative behavior within individuals.

Quantifiable communicative behavior shared by groups may be another matter. In this connection, Yngve rejects the traditional notion of the ideal speaker-listener – a notion summed up by Chomsky as follows:

> Linguistic theory is concerned primarily with an ideal speaker-listener, in a completely homogeneous speech-community, who knows its language perfectly ... This seems to me to have been the position of the founders of modern general linguistics, and no cogent reason for modifying it has been offered. [1965:3–4]

Perhaps Chomsky intends this statement as an idealization meant to simplify the linguist's job in eliciting the grammar from informants, not as a literal endorsement of the notion that all members of a speech-community share exactly the same grammar. In any case, Yngve would encourage us to study the individual as an individual or as a member of more than one communicating community. Through my own gesture-based research and through my reading of variationist literature, like Yngve, I have come to question the notion

of the ideal speaker-listener. In my case, I do so entirely without glee. I wish my experimental subjects were more alike in their properties as communicators. Cross-speaker similarities in behavior would make interpretation of experimental results much tidier. In any case, having found a real-world event that is at once serviceable as a unit of production, perception, and phonology – the gesture – it would certainly be very comforting to find that individuals who routinely communicate with each other share at least minimal elements of a gesture-based system of communication. After having served on the front lines of speech physiology research, I will not easily be persuaded that neatly bounded, homogeneous speech communities exist, but I certainly hope that our work produces a clearer picture of the sorts of communicative behaviors that are shared between people who sometimes talk with each other.

Acknowledgements

I acknowledge the support of NIH Grant DC-03782 to Haskins Laboratories from which I received funding during the preparation of the present chapter. I thank Carol Fowler, Louis Goldstein, and Vic Yngve for very helpful comments.

References

Adrian, J. A., Alegria, J., and Morais, J. (1995), 'Metaphonological abilities of Spanish illiterate adults'. *International Journal of Psychology*, 30 (3), 329–353.

Bloomfield, L. (1933), *Language*. New York: Holt.

Browman, C. and Goldstein, L. (1988), 'Some notes on syllable structure in articulatory phonology'. *Phonetica*, 45, 140–155.

Browman, C. and Goldstein, L. (1991), 'Gestural structures: Distinctiveness, phonological processes, and historical change', in I. G. Mattingly and M. Studdert-Kennedy, *Modularity and the Motor Theory of Speech Perception*. Hillsdale, NJ: Lawrence Erlbaum, pp. 313–338.

Browman, C. and Goldstein, L. (1995), 'Dynamics and articulatory phonology', in R. F. Port and T. van Gelder, *Mind as Motion: Explorations in the Dynamics of Cognition*. Cambridge, MA: MIT Press, pp. 175–193.

Browman, C. and Goldstein, L. (2000), 'Competing constraints on intergestural coordination and self-organization of phonological structures'. *Bulletin de la Communication Parlée*, 5, 25–34.

Chomsky, N. (1965), *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.

Fowler, C. A. (1991), 'Auditory perception is not special: We see the world, we feel the world, we hear the world'. *Journal of the Acoustical Society of America*, 89 (6), 2910–2915.

Fowler, C. A. (1980), 'Coarticulation and theories of extrinsic timing'. *Journal of Phonetics*, 8, 113–133.

Gibson, E. J. and Pick, A. D. (2000), *An Ecological Approach to Perceptual Learning and Development*. Oxford: Oxford University Press.

Gibson, J. J. (1979), *The Ecological Approach to Visual Perception*. Boston: Houghton Mifflin.

Gibson, J. J. and Gibson, E. J. (1955), 'Perceptual learning: Differentiation or enrichment?' *Psychological Review*, 62, 32–41.

Goldstein, L. and Fowler, C. A. (in press), 'Articulatory Phonology: A phonology for public language use', in A. S. Meyer and N. O. Schiller, *Phonetics and Phonology in Language Comprehension and Production: Differences and Similarities*. Mouton de Gruyter.

Hawkins, S. (1992), 'An introduction to task dynamics', in G. J. Docherty and D. R. Ladd, *Papers in Laboratory Phonology 2*. Cambridge: Cambridge University Press, pp. 9–25.

Honorof, D. N. and Browman, C. P. (1995), 'The center or edge: How are consonant clusters organized with respect to the vowel?', in K. Elenius and P. Branderud, *Proceedings of the XIIIth International Congress of Phonetics Sciences, 3*, Stockholm, Sweden, pp. 552–555.

Horiguchi, S. and Bell-Berti, F. (1987), 'The velotrace: A device for monitoring velar position'. *Cleft Palate Journal*, 24 (2), 104–111.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., and Studdert-Kennedy, M. (1967), 'Perception of the speech code'. *Psychological Review*, 74, 431–461.

Liberman, A. M. and Mattingly, I. G. (1985), 'The motor theory of speech perception revisited'. *Cognition*, 21, 1–36.

Mattingly, I. G. (1981), 'Phonetic representation and speech synthesis by rule', in T. Myers, J. Laver, and J. Anderson, *The Cognitive Representation of Speech*. Amsterdam: North Holland, pp. 415–420.

Mattingly, I. G. (1990), 'The global character of phonetic gestures'. *Journal of Phonetics*, 18, 445–452.

Ohala, J. J. (1990), 'There is no interface between phonology and phonetics: A personal view'. *Journal of Phonetics*, 18, 153–171.

Nadler, R. D., Abbs, J. H., and Sujimura, O. (1987), 'Speech movement research using the new x-ray microbeam system', in *Proceedings of the XIth International Congress of Phonetic Sciences (1)*. Tallinn, Estonia: Academy of Sciences of the Estonian S.S.R. Institute of Language and Literature, pp. 221–224.

Perkell, J. S., Cohen, M. H., Svirsky, M. A., Matthies, M. L., Garabieta I., and Jackson, M. (1992), 'Electromagnetic midsagittal articulometer (EMMA) systems for transducing speech articulatory movements'. *Journal of the Acoustical Society of America*, 92, 3078–3096.

Pouplier, M. (2003), *Units of Phonological Encoding: Empirical Evidence*. Doctoral dissertation, Department of Linguistics, Yale University. (To become available via free download from ProQuest Digital Dissertations, http://wwwlib.umi.com/dissertations/gateway.)

Read, C., Zhang, Y.-F., Nie H.-Y., and Ding, B.-Q. (1986), 'The ability to manipulate speech sounds depends on knowing alphabetic writing'. *Cognition*, 24, 31–44.

Saltzman, E. (1986), 'Task dynamic coordination of the speech articulators: A preliminary model'. *Experimental Brain Research*, 15, 129–144.

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

Scherer, R. C., Druker, D. G., and Titze, I. R. (1988), 'Electroglottography and direct measurement of vocal fold contact area', in O. Fujimura, *Vocal Physiology: Voice Production, Mechanisms and Functions*. New York: Raven Press, pp. 279–91.

Studdert-Kennedy, M. and Goldstein, L. (2003), 'Launching language: The gestural origin of discrete infinity', in M. H. Christiansen and S. Kirby, *Language Evolution*. Oxford: Oxford University Press, pp. 235–254.

Surprenant, A. M. and Goldstein, L. (1998), 'The perception of speech gestures'. *Journal of the Acoustical Society of America*, 104 (1), 518–29.

Yngve, V. H. (1970), 'On getting a word in edgewise', in M. A. Campbell *et al.*, *Papers from the Sixth Regional Meeting, CLS 6*, Chicago: Chicago Linguistic Society, pp. 567–78.

Yngve, V. H. (1996), *From Grammar to Science: New Foundations for General Linguistics*. Amsterdam/Philadelphia: John Benjamins.