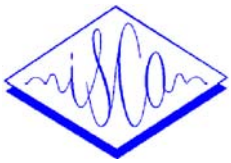

6th European Conference on
Speech Communication and Technology
(EUROSPEECH'99)
Budapest, Hungary, September 5-9, 1999

HOW SPEECH WORKS
QUESTIONS AND PRELIMINARY ANSWERS

Björn Lindblom
Stockholm University
&
The University of Texas at Austin

ABSTRACT

In this presentation I attempt to discuss research from a wide spectrum of topics all relevant to the study of speech communication. It is my hope that placing more narrowly focused work in a common, broader context will produce a kind of symbiosis effect that will help us see the individual subtopics in perhaps new and insightful ways. Admittedly, such an undertaking might seem ambitious and even ill-advised. However, as our field is put under increasing pressures to deliver clinical and technological applications of greater real-life relevance, it does appear useful to try to assess what we know and do not know about speech processes. Also, as is sometimes the case, the whole tends to be more than the sum of its parts.

FROM LAB SPEECH TO UNSCRIPTED CONVERSATION

For many years phoneticians have tried to understand speech by investigating carefully designed speech samples, so-called 'laboratory speech'. Recently, several studies of spontaneous, unscripted speech have been reported (SPOSS 98). It is seems fair to say that the new findings have clearly indicated that laboratory speech tends to drastically undersample the signal variability exhibited by natural speech patterns. Patterns of extensive reduction and elaboration are typically observed and are likely to challenge many of our most cherished theoretical ideas about how the production and perception of speech work.

AN ARTICULATORY BALLET

However, the study of laboratory speech is indispensible to a deeper understanding the basic processes of speech production. Some recent work in our Stockholm laboratory has produced a set of X-ray recordings which I will show excerpts of in my talk. This method imposes severe constraints on the duration of the experiemental sessions. Nonetheless, the elegance and complexity of how articulators go about realizing the phonetic structure of the test words is readily apparent. To formulate some of the questions that arise in the analysis of these data we will review our work on quantitative models of coarticulation and how articulatory movements are coordinated.

'HÖRBARE MIMIK' AND 'LA VIVE VOIX' (IVAN FÓNAGY)

Our vocal system is a dynamic expressive instrument which is continually recalibrated in response to our state of mind and according to the demands of the speaking situation. Such modulations can produce marked changes in the acoustic signal. For instance, when speaking to an infant affectionately, in many cultures talkers may change their 'normal' pronunciation by superimposing strong nasalization and lip roudning. Speaking angrily we may clench our teeth thereby modifying speech sounds in complex but basically systematic ways. Are such transforms easier to describe in articulatory rather than acoustic terms? If so, what are the implications for 'gesturalist' vs 'auditorist' theories of speech?

THE PHONETICIAN'S KEY PROBLEM

Clearly, what is linguistically and psychologically the "same" utterance can be realized physically in innumerable and, at first glance, unpredictable ways. Speaker identity, speaking style, motor constraints and channel noise all contribute to this complexity. The key problem in theoretical and applied areas of speech research has been, and continues to be, the massive variability of the speech signal. How do our brains generate and cope with this variation? How does it solve the so-called *invariance* problem? And how do children learn to talk from this incomplete, noisy and richly varied input? How do they manage to find the phonology and the intricate grammatical system of the native language behind it?

SIGNAL-INDEPENDENT INFORMATION AND SPEECH DEVELOPMENT

Several theoretical frameworks have been proposed to deal with these basic issues. One approach is to view linguistic units as gestural or auditory *invariants* deeply embedded in a signal subserving many non-linguistic functions. The hypothesis is that listeners are capable of extracting those either gestural or auditory invariants from the speech wave.

Another school of thought views speech as *adaptively* organized. The physical implementation of a phonetic form varies along an 'hyper-hypo continuum' between under- and overarticulation in accordance with the listener's access to 'signal-independent' information. On this view the task of the signal is to supply not

invariants, but just enough information to disambiguate the input and to make lexical access successful.

Explaining speech variability in this way presupposes a way of quantifying the listener's 'need for information'. This is an incompletely understood topic which merits a great deal of attention. To illustrate one possible apporach to this problem work on the *Neighborhood Activation Model* (Paul Luce) will be discussed.

EXEMPLAR-BASED PHONETIC LEARNING

The 'adaptive' H&H account of phonetic variation encounters another major difficulty when applied to speech development. We know that children are borne with sophisticated cognitive and perceptual mechanisms, but also that they need time to become fully calibrated by their linguistic experience. How much 'signal-independent information' are they able to apply to the ambient language? Does infant-directed speech solve the invariance problem for the child? The available evidence suggests that Baby Talk is prosodically lively and rich in emotive coloring. Such transformations would not seem to provide the child with nice and clean prototypical information on linguistic categories.

A recent trend in work on speech perception and phonetic learning suggests a way out of this problem: Exemplar-based learning. As pointed out by proponents of this paradigm, traditional speech perception accounts assume representations (e.g., phoneme-sized units) to be simple (context-free invariants). To derive such units from the speech signal calls for complex mapping operations capable of solving the invariance problem. Mechanisms of this type have been proposed – e.g., the 'phonetic module' of the motor theory, the 'smart mechanisms' of direct realism and the 'top-down' processes (reconstructive rules, inference making and hypothesis testing) of cognitively oriented approaches. How do those processes operate on-line in adult listeners and how do they develop? Our understanding of these issues is currently far from complete.

Exemplar-based models adopt the opposite perspective assuming representation to be complex and mapping to be simple. Categories form as emergent products of cumulative phonetic experience because that experience is patterned in complex but principled ways

ENERGETICS AND SPEECH MOVEMENTS

The claim that movements are shaped by a minimum energy criterion rests solidly on a large body of physiological studies. A great deal of quantitative data is available on how much energy various species use during locomotor tasks. This information is presented by plotting the amount of energy that the subject expends against traveling speed. An example is the data on horses walking, trotting and galloping. When the energy used is expressed per unit distance traveled, the measurements tend to be U-shaped and have distinct minima. Significantly, these minima are found at speeds that subjects spontaneously adopt when moving freely and unconstrained by the speed of a laboratory treadmill.

Experimental biologists interpret such findings to suggest that locomotion (human walking and running, birds flying, fish swimming etc) is shaped by a criterion of 'minimum-energy expenditure'. The phonetician is naturally asks: What is the relevance of energetics to speech? Are speech movements and whole body movements similarly organized? The energy costs of speech movements are likely to be small in comparison with those of locomotion. Therefore our brains might say: 'Be my guest! Do anything you like'. On the other hand, evolution being a tinkerer and also a miser, it would be inclined to exercise parsimony also in motor control. Thus the same rules ought to apply for small as for big movements. Until measurements of speech energy costs are made, we should of course keep an open mind on those issues. Why should speech be different?

BEHAVIORAL STATUS OF LINGUISTIC UNITS

Traditionally , linguistics deals with the problem of the behavioral variety of speech patterns by assigning the study of the signal to phonetics and the experimental sciences. Phonology investigates postulated sound structure, that is the abstract entities and processes assumed to underlie speech behavior and by definition independent of performance and language use. In this way, the study of speech sounds is split into two: Phonology becomes digital, and phonetics analog.

It is interesting to examine this traditional division of labor from the viewpoint of speech development. If children's phonetic input is indeed massively variable, how do they discover the hidden structure of speech? Linguistic analysis provides the phonetician with discrete and invariant units as a starting point for organizing observations and for specifying continuous and variable behavioral correlates. This situation has given rise to two major issues: the invariance and the psychological reality problems. For many many decades now they have been the primary concern of the experimentalist rather than the linguist.

CONCLUSIONS

Do the standard operational procedures of linguistics produce units that are descriptively effective but epiphenomenal from a behavioral standpoint? It is instructive to broaden the perspective and ask: Where do discrete/invariant units come from in development and evolution? Are they innately prespecified, as envisioned by proponents of Chomskyan Universal Grammar? Or do they unfold during development as adaptive emergents? In all honesty, we do not know. However, adding that question to our agenda may open up opportunities for re-examining some of our most fundamental issues and might eventually lead to a resolution of the invariance and psycholigcal reality problems as well as bring about a greater degree of unification of all the diverse areas of inquiry that make up the discipline of speech communication.