ADAPTIVE VARIABILITY AND ABSOLUTE CONSTANCY IN SPEECH SIGNALS: TWO THEMES IN THE QUEST FOR PHONETIC INVARIANCE

Björn Lindblom

Department of Linguistics, University of Stockholm, Sweden

ABSTRACT

Our topic is the classical problem of reconciling the physical and linguistic descriptions of speech: the invariance issue. Evidence is first presented indicating the possibility of defining phonetic invariance at the articulatory, acoustic or auditory levels of the speech signal. However, as we broaden the scope of our review, we find that attempts to define phonetic invariance in terms of absolute physical constancies tend to lose ground to theories that recognize signal variability as an essentially systematic and adaptive consequence of the informational mutuality of natural speaker-listener interactions. We reach this conclusion not only by examining experimental data on on-line speech processes but also by analyzing typological evidence on how the phonetic structure of consonant systems varies with inventory size in a lawful manner.

INTRODUCTION

Traditionally the problem of invariance in phonetics can be said to be that of proposing physical descriptions of linguistic entities that have the characteristic of remaining invariant across the large range of contexts that the communicatively successful real-life speech acts present to us.

Many of us share the conviction that taking steps towards the solution of this problem will be crucial if we are to acquire a deeper theoretical understanding of the behavior of speakers and listeners as well as develop more advanced systems for speech-based man-machine communication (Perkell & Klatt 1986).

The present paper will attempt to address some of the questions that we typically encounter in the search for invariance. We shall do so by summarizing research undertaken mostly in our own laboratory in Stockholm. Although thus deliberately limiting the scope of our review we hope that the issues raised will nevertheless be of sufficient interest to stimulate general discussion.

IS PHONETIC INVARIANCE ARTICULATORY?

A few decades ago phoneticians began to interpret phonetic events by comparing articulators to highly damped oscillatory systems. More recently, such models have acquired an important role within the framework of action theory (Kelso, Saltzman and Tuller 1986). In the sixties it was hoped that a lot of the variability that speech signals typically exhibit - e.g. reductions and vowel-consonant coarticulation (Ohman 1967) - could be explained in terms of the spatial and temporal overlap of adjacent "motor commands" (MacNeilage 1970). Articulatory movements were seen as sluggish responses to an underlying forcing function which was assumed to change, usually in a step-wise fashion, at the initiation of every new phoneme (Henke 1966). Owing to variations in say stress or speaking tempo, different contexts would give rise to differences in timing for a given sequence of phoneme commands. Articulatory and acoustic goals would not always be reached, the so-called 'undershoot' phenomenon (Stevens and House 1963). But since such undershoot appeared to be lawfully related to the duration and context of the gestures (Lindblom 1963), the underlying articulatory "targets" of any given phoneme - 'die Lautabsicht' - would nevertheless, it was maintained, remain invariant. Accordingly, at that time it seemed possible to argue that phonetic invariance might be articulatory.

Duration-dependent undershoot still seems to to be a phonetically valid notion for biomechanical reasons. But it is clearly not as inevitable a phenomenon as was first thought. Current experimental information indicates that in fast speech articulatory and acoustic goals can be attained despite short segment durations (cf Engstrand 1987, Gay 1978, Kuehn and Moll 1976). Furthermore undershoot has been observed in unstressed Swedish vowels that exhibit long durations owing to 'final lengthening' (Nord 1986). Such deviations from simple duration-dependence appear to highlight the reorganizational abilities of the speech production system. One way of resolving the problem posed by these somewhat contradictory results might be obtained if it were shown that when instructed to speak fast, subjects have a tendency to 'overarticulate', thus avoiding undershoot to some extent, whereas when destressing they are more prone to "underarticulate" (cf discussion below of hypo- and hyper-speech). The demonstration of language-specific patterns of vowel reduction
In summary, the original observations of "undershoot" carried the implication that the intonational rise was not found in the intonational lilt. Evidence of articulatory movement. Phonetic invariance was found to be a property of the organism from a dynamic perspective with a view to reanalyzing many of the traditional notions of "articulatory" representation in extant speech production models. Their writings convey the expectation that many articulatory processes are "articulatory gestures" and that these "articulation models" will simply fall out as consequences of the dynamic properties intrinsic to the speech motor system. In the terminology of Kelin, Saltzman and Tuller (1964, 55) "... both time and timing are deemed to be intrinsic consequences of the system's dynamical organization. Methodologically, action theory is commendable since, being committed to interpreting phonetic phenomena as fortuitous (intrinsically) consequences rather than as controlled (extrinsically) aspects of speaker's (articulatory) behavior, it guarantees a maximally thorough examination of speech production processes. However, it is difficult to see how, applying the action theoretic framework to the data on compensatory behavior, one could possibly avoid postulating some sort of 'temporal target' representation which is (i) extrinsically structured and (ii) responsible for executing the gestures and which is (iii) responsible for extrinsically tuning their dynamics. Speech production is a highly versatile process and sometimes appears to be time-consuming.

The plasticity of the speech motor system is often maximized by overarticulation, and the segments in one of the normal or the non-normal jaw positions and whether they would be able to produce the gestures in the normal conditions.

The results of the investigation, in the bite-block words, deviated systematically but very little from the normal patterns and concluded that the non-normal articulations do not compensate for the non-normal articulations. The results are in agreement with those reported earlier by Hillel, Rett and Ams (1973). Moreover, they are completely analogous to the previous demonstrations that naive speakers are capable of producing articulatory gestures that are normal at the first glance in spite of an unnatural jaw opening imposed by the use of a "bite-block."
18 PHONIC INvariance Auditory?  

The notion that a perceptual effect exists is a rather curious parallel to the theory of independent variables, such as the Psychoneurotic Effect, in that it is the presence of an independent variable that determines whether or not a perceptual change occurs. However, it is not clear what the perceptual effect is or whether it is related to the Psychoneurotic Effect. The perceptual effect is not well defined, and it is not clear what the independent variable is. It is also not clear whether the perceptual effect is related to the Psychoneurotic Effect.

IMPLICATIONS OF SPEAKING Style: The Hyponomous Dimension  

Everyday experience indicates that speaking style is a highly variable process. We are capable of varying our style of speech from fast to slow, soft to loud, casual to clear, intimate to public. We speak in different ways to different audiences, based on our own speaker and the computer's voice. And we vary our pronunciation as a function of the social situation we are in. We vary our speaking style in laboratory interactions (Labov 1972).

There are at least three principal types of phonetic invariance: articulatory, acoustic, and auditory invariance. The first two are generally considered to be perceptual or experimental variables, while the third is perceptual invariance in the sense of variance in the acoustic properties of the stimulus. This invariance is shown in her attempts to produce different responses. Kingsbury and Krull (1968) have demonstrated that the location of the perceptual axis can be changed by varying the frequency of the words. Lacerda (1986) has demonstrated that the location of the perceptual axis can also be changed by varying the frequency of the words. However, this is not the case when the words are spoken in a different language. According to Lacerda (1986), we can characterize one part of the perceptual axis as being the axis of frequency of the words. This axis is frequency of the words, and it is used to compare the perceptual axis of French words with the perceptual axis of English words. The French words are characterized by a higher frequency of the words, and the English words are characterized by a lower frequency of the words.

This is not the case when the words are spoken in a different language. According to Lacerda (1986), we can characterize one part of the perceptual axis as being the axis of frequency of the words. This axis is frequency of the words, and it is used to compare the perceptual axis of French words with the perceptual axis of English words. The French words are characterized by a higher frequency of the words, and the English words are characterized by a lower frequency of the words. This is a significant finding, as it suggests that the perceptual axis can be changed by varying the frequency of the words, and it suggests that the perceptual axis is not fixed and can be changed by varying the frequency of the words.

SPEECH UNDERSTANDING: (IN)DEPENDENCE OF BILINGUALISM  

At the Department of Romance Languages at Stockholm University, the proficiency of proficient native Swedish students is in understanding spoken French. However, the proficiency of French speaking students is not necessarily the same as the proficiency of Swedish speaking students. This is because the proficiency of French speaking students is not necessarily the same as the proficiency of Swedish speaking students. The proficiency of French speaking students is not necessarily the same as the proficiency of Swedish speaking students. The proficiency of French speaking students is not necessarily the same as the proficiency of Swedish speaking students. The proficiency of French speaking students is not necessarily the same as the proficiency of Swedish speaking students.
important role in the perception of speech. The experimental data on production indicates that the behavior of the speech motor system is shaped primarily by two forces - plasticity (listener-oriented reorganization, or information-driven simplification) - which interact on a short-term basis so as to generate signals that may be 'rich' or 'poor' in extraneous physical information. The evidence on perception has identified two major sources of information: signal-dependent and signal-independent processes and suggests that on a short-term basis perception arises from the latter (i.e., 'context') modulating the former in an analogously 'rich' or 'poor' manner. The possibility of modulating the logical possibilities of these conceptual simplifications is shown in the diagram of the enclosed figure. This is not a very rigorous scheme but seems useful, at least analogically, in contrasting some of the ideas currently entertained in phonetics (cf. Lande, Phonetics, January issue 1966). This graph states that for speech to be intelligible the sum of explicit physical information and signal-independent information must be above a threshold, that is the degree to which the signal provides a constant the x and y-values of specific speech sounds fail. Right on that line. Points above the line are associated with what might be termed 'over- clearness' and those below it with 'unintelligible' speech.

It appears reasonable to assume that in the real-life situations utterances can vary tremendously with respect to how socially and communicatively acceptable they are. For our present purposes let us focus on speech situations in which a successful real-life speaker-listener interaction and assume that production patterns are not clustered near and above the straight line; what would such a result imply? It would mean that the differences in the amounts of information contributed by the two parties on the one hand and 'context' on the other. When speakers come close to the silent line it would indicate first of all that they are capable of varying their speech output in a plastic way (i.e. evidence on hypo-hypersonic modes and other instances of reorganization of speech motor control) and secondly that, while perhaps not being perfect 'mind-readers', they are at least capable of adapting their speech to the short-term fluctuations in the listener's access to 'context' or signal-independent information (i.e. experimental documentation of numerous cases showing that listeners are in fact capable of successfully coping with highly context-dependent and coarticulated speech stimuli). The possibility of such complementarity in real speech emerges also from speech perception and was recently illustrated by Hume and Hunnicutt (1965) as well as from Lieberman's (1969) work.

If we hypothesize that this strategy - let us call it the STRATEGY OF ADAPTIVE VARIABILITY - comes near the way real speakers actually behave when they are communicatively successful and obtainable, then a natural way of resolving some of the paradoxes that would arise for economy (i.e. over-use) is to assume that the expected utility of ecological signals is to exhibit absolute physical invariance. The proposed way of thinking about the issue does not of course rule out finding physical speech sound invariance in restricted domains of observed data but it does explain why our quest for a general concept of phonetic invariance has largely been unsuccessful. And, in a pessimistic vein, it predicts in fact that it will continue to be so.

Our reasoning leads us back to a consideration of a question raised in his 1970 review of the inventory issue: '...the essence of the speech production process is not an innate ideational code invariant central signals, but an elegantly controlled motor flexibility of response to the demand for a relatively constant end (p 104)'.
idea we can say that a larger paradigm goes with a RICHER signal inventory. The other side of the coin is of course the smaller paradigm that is attributed to word endings goes with a FOUNDER signal inventory. In both cases the presence of assimilation and coarticulation should vary inversely with the size of the paradigm. A study of this sort could be carried out on a hypothetical "J" in" complex since it shows more than one elaborated source feature (respiration). Logically a six-consonant system could use the ejective set for its larger inventory whereas the smaller systems are not able to use the [j] inventory. Relative to the "category" or "elaborated" series (cf. the notion of "implicative hierarchy" in Kripke's terminology), the claim we make is accordingly that we see a positive correlation between paradigm size and the number of elements that a sound pattern selects from a dimension of "articulatory complexity".

The validity of our analysis naturally hinges on the success with which we can give non-circular, independently motivated definitions of "articulatory complexity". When it comes to the details of the analysis that problem is a topic for futurequantitative phonetic theory. For the moment we believe that the major trends are rather gross ones in the data that can be convincingly demonstrated by the force of the examples. They provide us with the following generalization: Small consonant paradigms involve "unharvested" phonetic systems, large paradigms involve "harvested" phonetic systems. That is of course exactly what Nootboom's hypothesis predicts and it is taken as evidence against an explanation for why seven consonant systems do not show inventories like the following (Holyoak 1980):

\[ d \, 44 \, t \, 2 \, r \, 4 \]

We take this present typological data of consonant systems as providing strong evidence in favor of (a) language structure evolving as an adaptation to the speaker-listener interactions and for (b) the correctness of a theory of Acoustic Variability as an account of those processes.

REFERENCES


