

THE RELATIONS BETWEEN SPEECH PRODUCTION AND LEVELS  
OF REPRESENTATION: INTRODUCTORY REMARKS

Björn Lindblom

Department of Linguistics  
University of Texas at Austin

ABSTRACT

This session can be seen as complementary to a previously organized symposium on **Speech processes in the light of action theory and event perception** [1]. It is aimed at examining the role of perceptual processes in the motor control of "phonetic gestures".

1. THE INVARIANCE ISSUE: IN-PRINCIPLE SOLUTIONS.

One of the classical problems in phonetics is the difficulty of specifying physical invariants corresponding to linguistic categories [2]. There are currently several favored research paradigms that, implicitly or explicitly, take a stance on that issue and can be said to offer programs for the in-principle resolution of it. They seem to fall in either of two categories.

Theme 1: **Phonetic invariance is in the signal.** According to this approach phonetic invariance for linguistic categories will ultimately be established once we learn to look at the signal in the right way and to make the right kind of measurements be they articulatory, acoustic or auditory. The tacit hope is that one day

discoveries will be made that render Hockett's often quoted Easter Egg metaphor inappropriate [3]. Recent formulations of the Motor Theory [4], Direct Realism [5], Coordinative Structures [6], the Quantal Theory [7] and the notion of "Icebergs" [8] appear to come close to this approach.

Theme 2: **Phonetic invariance is not in the signal.** This alternative links the variability of speech signals to the adaptive organization of speech. According to this view, the listener's short-term demands for explicit signal information do not stay constant during and across utterances. Thus the lack of signal invariance is seen to arise as a consequence of the talker's tacit recognition of variations in short-term perceptual and situational demands and his/her adaptive response to them. The so-called H&H theory exemplifies this type of reasoning [9].

2. IMPLICATIONS FOR SPEECH MOTOR CONTROL

What are the implications of these alternatives with respect to the task of the speaker? What are the parameters that (s)he actively controls?

Our answers to those questions are closely tied to the assumptions we make about the nature of speech perception. For instance, if speech perception is assumed to be based on the extraction of higher-order signal invariants - be they gestural, acoustic or auditory - then speech motor control must be seen as aimed at producing those gestural, acoustic or auditory invariants.

If, on the other hand, it is not based on signal invariance, we must envision control parameters in a different way. Let us suppose that the role of the signal is to supplement knowledge already available to the listener (in short- and long-term memory) and that its purpose is to discriminate among competing candidates in the listener's lexicon. Hence the speaker's task is to control the phonetic discriminability - rather than the invariance - of signal attributes. In other words, the talker should generate signals that are sufficiently rich to facilitate correct identification.

3. CONSTANCY AND ADAPTIVE TUNING OF PHONETIC GESTURES

Within the time limits of this symposium it is not possible to do justice to all the paradigms that are currently explored in phonetics and that bear on the invariance issue. The selection of the present contributions was deliberately made so as to promote a discussion of issues related to Theme 2. The justification for that decision is that so far, although neither particularly novel nor

counter-intuitive, theme 2 seems to represent the scientifically less traveled research avenue.

Some of the issues that follow from theme 2, can be stated as follows: What is the status of "phonetic gestures"? Are they the theoretical primitives from which explanations of on-line phonetic variability are to be derived? Are they the ultimate control units of speech motor control? Or are "phonetic gestures" in no way prime constructs but themselves derivable from the dynamic tug-of-war between production and perception demands? If so, to what extent are they tuned to meet perceptual demands? If perceptually motivated tuning of gestures can indeed be demonstrated, what is the extent of such transforms? In other words, what is the scope of cooperative behavior in speaker-listener interaction (cf Nootboom)? If, as argued by Kohler, the varying degrees of reduction that speech exhibits in response to situational and perceptual conditions, how do we describe those listener-oriented conditions in a quantitative and language-independent way (cf Diehl)? And in a similarly language-independent manner how do we quantify processes of reduction? Does Articulatory Phonology (Browman and Goldstein) present a way of addressing those questions?

4. EXISTENCE AND SIGNIFICANCE "CLEAR SPEECH"

In the present context I would like to draw attention to a recent study of "clear speech" by Seung-Jae Moon of the University of Texas at

Austin [10]. Moon made measurements of vowel and consonant formant patterns in the stressed /w\_l/-syllables of mono-, bi- and tri-syllabic words. These test items were chosen so as to maximize locus-to-target distances, maintain stress constant and produce variations in the duration of the stressed vowel. The idea behind this design was to induce "undershoot" effects in the vowel formant patterns. Two speaking styles were investigated: citation forms and "clear", overarticulated speech. His results, which will be summarized in a paper contributed to this conference, indicate that "clear speech" tokens are not simply louder citation forms but involve reorganization of acoustic patterns and the underlying articulatory gestures. Acoustically, this reorganization takes the form of removing contextual effects of the /w\_l/ environment, that is reducing undershoot, and shifting "clear speech" formant patterns closer to null-context reference values for the various vowel categories. Intelligibility tests showed that "clear speech" is more resistant to noise than citation forms.

I am going to allow myself to generalize from Moon's results regarding them as a correct and generally valid description of "clear speech". By way of conclusion, let me raise the following questions for our session: Why should there be such a thing as "clear speech", that is a style of speech that apparently has different acoustic properties from those of

neutral citation-form speech? Why do speakers bother to change their pronunciation when attempting to clarify? Given that they do change their articulatory patterns, what does such behavior imply about the organization of speaker-listener interactions? And what does the very existence of "clear speech" imply about perceptual processing and the invariance issue?

#### 5. REFERENCES

- [1] LINDBLOM B (ed) (1986): Speech processes in the light of action theory and event perception, theme issue, Journal of Phonetics 14:1, 1-196.
- [2] PERKELL J and KLATT D (1986): Invariance and variability in speech processes, LEA:Hillsdale, N J.
- [3] HOCKETT C F (1958): A manual of phonology, Bloomington, Indiana:Indiana University Press.
- [4] LIBERMAN A M and Mattingly I G (1985): "The Motor Theory of Speech Perception Revised", Cognition 21:1-36.
- [5] FOWLER C A (1986): "An Event Approach to the Study of Speech Perception from a Direct-Realist Perspective", Journal of Phonetics 14:1, 3-28.
- [6] KELSO J A S, SALTZMAN E L and TULLER B (1986): "The dynamical perspective on speech production: Data and theory", Journal of Phonetics 14:1, 29-59.
- [7] STEVENS K N (1989): "On the quantal nature of speech", Journal of Phonetics 17:1/2, 3-45.
- [8] FUJIMURA O (1990): "Articulatory perspectives of speech organization", in Hardcastle W J and Marchal A (eds): Speech Production and

Speech Modeling, Kluwer Publishers, Dordrecht.

[9] LINDBLOM B (1990): "Explaining phonetic variation: A sketch of the H&H theory", 403-439 in Hardcastle W J and Marchal A (eds): Speech Production and Speech Modeling, Kluwer Publishers, Dordrecht.

[10] MOON S-J (1991): An Acoustic & Perceptual Study of Formant Undershoot in Clear & Citation-Form Speech, Ph.D. diss. University of Texas at Austin.

#### 6. ACKNOWLEDGEMENTS

This research was supported by a grant from the Advanced Research Program of the Texas Board of Coordination and grant No. BNS-9011894 from the NSF.