I. Introduction

To the physiologist who would study language in terms of the interests represented at this symposium, the most obvious linguistic processes—the selection of words to convey meaning and the arrangement of words in sentences—must seem far removed from familiar concepts and methods. Surely, he would prefer to study processes that are physiologically more accessible, but are yet linguistic. We believe that the production and perception of speech, in the narrow sense, is one such process; we suggest, therefore, that the physiologist might do well to start there. The questions we would have him ask can be put very simply: How does a speaker convert the phonetic units—the consonants and vowels—to a stream of sound? On hearing that stream, how does a listener recover the phonetic units?

For the purposes of this paper, we can do as well with one of those questions as with both; in principle, either one will do. We have chosen to deal primarily with the second—the one about speech perception—because we find more data there that speak to the points we want to make.

In the received view, speech perception in our narrow sense is thought to be neither very linguistic nor very interesting. Language is commonly supposed to be structured in levels (syntactic, phonological, phonetic) that represent successive recodings of the information. Each of these levels...
consists of simple units (words, phones) of some kind, organized into successively larger and more complex units (phrases, sentences; syllables, breath groups). It has been the business of the linguist to describe the rules by which the units are organized at each level, and to discover the code by which they are converted from one level to the next.

But the recoding is usually assumed to end with the phones, the empty units that lie at the lowest level of the whole linguistic structure; it is not supposed to include the process we want to talk about here: the conversion from phone to sound and from sound to phone.\(^1\) The traditional view of this conversion is that it is by means of a simple alphabet, each phone being represented by a unit sound. On that view the relation between phone and sound is trivial and outside language proper; the linguist is interested in the sound alphabet only as a concrete base in which he can, when necessary, anchor his abstract assumptions.

An alternative view, which seems to us not very different in its consequences, is that the relationship between phone and sound, while indeed part of language, has a character too irregular to invite systematic description. This would appear to be the position of Chomsky and Miller (1963, p. 318) and Chomsky and Halle (1968, pp. 293–295). Though these investigators assume that a universal phonetics underlies speech production, they believe that speech perception depends not only on the acoustic properties of the signal but also on “the hearer’s knowledge of the language as well as on a host of extra-grammatical factors,” so that the perception and the sound cannot be rigorously related. They regard speech perception as a “heuristic” process in which a hypothesis about the speaker’s utterance is suggested by a sampling of cues in the speech signal. That hypothesis is then tested and corrected by reference to all levels of the grammar—a task that can be carried out by successive iterations of an analysis-by-synthesis procedure—until a plausible reconstruction of the string of phones has been arrived at. But if it is sufficient to have recourse to higher grammatical levels in order to carry out a process that is now viewed as unruly rather than simple, then the process of speech perception must still be considered uninteresting. For an explanation of speech perception, the linguist is, in effect, diverted from serious consideration of speech and directed back to the higher linguistic levels.

Neither do these easy assumptions about speech pose problems of any

\(^1\) The term “phone” is customarily used to mean the way a phonological segment is realized in a particular context. If one assumes that the perceptual, articulatory, and acoustic domains of speech are in one-to-one correspondence, then the phonological realization can be described equivalently as an acoustic event (or class), an articulatory event (or class), or a perceptual unit. But, as will be seen, we believe that these domains of speech are not in one-to-one correspondence. We will therefore use the term “phone” to mean the abstract unit which is the output of the phonology. This unit is, in effect, the elementary perceptual unit of speech. Our primary concern, in this paper, is to explore the relationship between the phone, so defined, and its acoustic representation.