# ON THE NECESSITY OF DISTINGUISHING BETWEEN SPEAKING AND LISTENING

HARLAN LANE*

When DeCordemoy wrote in 1668: "Whensoever any sound agitates the Brain, there flow immediately spirits towards the muscles of the Larynx, which duely dispose them to form a sound altogether like that which was just now striking the Brain...",[1] he might have been the first scientist to naively conflate the processes of speaking and listening, but he certainly was not the last. It is evident that speaking and listening are related but distinct processes, and that an adequate theory of language behavior must take these facts into account. Nevertheless, most theories give weight to the presumptive similarities between uttering speech and listening to it, while giving little weight to their differences. Conflationist theories are prominent in such diverse areas as studies of speech communication, at both segmental and suprasegmental levels, automatic speech recognition, foreign language learning, and child language.

With respect to the first domain, Twaddell wrote, in 1952: "...The hearer matches the acoustic stimuli he receives against his own habits of muscular speech action, and identifies the incoming speech sound as corresponding to this or that of his own speech articulations. At both ends of a speech transmission, it is the muscular activity, not the acoustic character, which dominates the identification". Hockett reasoned similarly in his Manual of Phonology (1955), Liberman in his review of research on speech perception (1957), Delattre in his survey of the acoustic correlates of consonants and vowels (1958): in short, speech is perceived by reference to articulation.

At the suprasegmental level, Schmitt (1924), Jespersen (1932), D. Jones (1950), Stetson (1951), Ladefoged (1958, 59, 63), Fonagy (1958, 65), Laziczius (1959), Lehiste & Peterson (1959) have all favored a "motor theory" of stress perception, effectively agreeing with S. Jones when he wrote in 1932: „Accent is *sui generis* depending for its perception on the kinesthetic sense... The listener refers what he hears to how he would say it. Thus he translates exteroceptor into proprioceptor sensations, the kinesthetic memory serving as stimulus." Galunov & Chistovich

* Center for Research on Language & Language Behavior, Ann Arbor, Michigan, U.S.A. and Institut de Phonétique de Paris, France.

[1] Because of space limitations, references are available from author.

(1966) and Liberman (1967) express the corresponding position with regard to the perception of pitch and intonation.

Analysis-by-synthesis is a strategy for automatic speech recognition that may simulate the human perceptual process according to Stevens (1960). It is also a stage in second-language learning according to Hockett (1955). Finally, Allport (1924), Liberman et al. (1961), and Kozhevnikov & Chistovich (1965) see it as a component of language development in the child; the latter write: "In the process of imitation [by the child] are provided conditions that are favorable to the forming of conditioned-reflex correlations between groups of sound signals and complexes of articulatory motions... It is assumed that these conditioned reflex correlations play an important part also in the process of speech recognition by an adult."

Opposing the conflation of speaking and listening are, in the first place, the findings of developmental and clinical studies by Lenneberg (1962), MacNeilage, Rootes & Chase (1967), and Fuller (1967); "Neither babbling, imitation nor articulate speech is necessary for the understanding of the natural language", Fuller concluded. Fant (1964), Jakobson & Halle (1956), and Jakobson in his book on child language (1942), all emphasize the relative independence of speaking and listening during language development. A similar view concerning *second* language acquisition is held by Jakobson & Halle (1956), and by Sapon (1965). Morton & Broadbent doubt that the "homunculus is really necessary" in their 1965 paper on passive versus active recognition models, as does Fant (1963, 66).

Further evidence for the functional independence of speaking and listening comes from studies of the perception of the suprasegmental characteristics of speech. Ladefoged (1959) has suggested that we judge loudness in terms of our own vocal level, and Warren (1962) has suggested that we judge our own vocal level in terms of loudness but, in fact, the autophonic scale (the scale of the speaker's perception of his own voice) is not the reception scale, and the reception scale is not the autophonic scale (Lane, 1961, 62, 63).

At the segmental level, the inference that speaking mediates listening is experimentally based on evidence for "categorical perception" of speech sounds. However, noises and visual patterns can be perceived categorically, too (Lane, 1965, 66). In order to illustrate how closely recent findings for color perception (Lane & Kopp, 1967) match those for speech and how, therefore, they constrain the interpretation of the speech results, we may substitute color terms for speech terms in an article on the motor theory of speech perception (Liberman et al., 1963): "Although the [colors] lie on a [visual] continuum, the perception is essentially discontinuous. Because of the discrimination peaks at the [color-class] boundaries, the incoming [colors] are [seen] categorically and they are, therefore, quickly and accurately sorted into the appropriate [color class]. "Although this article's *description* of categorical perception turns out to apply to color as to speech sounds, the same may not be said concerning its *interpretation* of categorical perception: "What kind of mechanism underlies the categorical perception of the [colors]. The answer seems to us... that the perception



of [color] is tightly linked to the feed-back from the speaker's own articulatory movements... In time, these movements... come to mediate between the [color] and its ultimate perception."

## DISCUSSION

*Singh:*

It may be pointed out that not only the consonants are perceived categorically but also the vowels. The results of a recent experiment conducted by Singh and Morehead (reported at the last Acoustical Society meeting in New York) showed sharply dividing spectral patterns in perceiving the consonants (p t k) as well as the vowels (a i u); especially the frequency spectra that facilitated and deteriorated the identification of vowels [i] and [u] were separated, in the lower-end of the spectrum, quite categorically.