GENERAL DISCUSSION OF SESSION 6:
PRIMARY SPEECH PERCEPTS

Chairman: A. Cohen

The title of this session suggests the search for perceptual units, which at one time in the past had already been designated as a wild goose chase. It seems clear by now that what we do in processing the incoming acoustic speech signal is highly task dependent. This goes for ordinary speech communication situations and all the more so in laboratory bound tasks.

The main issue at this session was the relation between, on the one hand, perceptual processing units as such, which are primarily stimulus bound, and can be considered to be the result of sensory transduction, i.e. auditory sensations, and, on the other hand, computational units which presuppose patterning of higher order mental constructs, such as are required, e.g., for the easy access of items in the mental lexicon. This apparent dichotomy can also be formulated in terms of proximal and distal objects.

Both aspects can be approached by mediation through the generally accepted notions derived from a linguistic analysis, giving rise to features, phonemic or allophonic segments, syllables, words, and even phrases. The smaller the chunk in focus the more stimulus bound the searcher’s stance turns out to be.

The outcome of the discussion was that there are no ready to hand primary percepts, although there seemed to be a consensus that linguistic units certainly help to focus the various components that make up the highly complex speech signal. Strong claims were made for the allophone as an ordering principle particularly from the point of view of those studying the relations between dynamic spectral characteristics and the auditory domain, whereas others claimed the study of word recognition to be the most rewarding approach for getting to grips with the primary task of speech understanding. Thus we can distinguish between an analytical and a more global approach; these go hand in hand with the two main characteristics inherent in the speech signal, one leading to segmentation of the time axis by dint of which linguistically induced contrastive events can be discerned, and the other being a continuous scanning with the emphasis on the identification of the components making for the coherent shape of the speech signal. In the latter sense acoustic similarities help in localising and focusing on the speaking source. It is mainly through this coherence in natural speech, often lacking in synthesised versions, that, in conformity with the laws of form, comparable to Gestalt phenomena in visual processing, an interpretation can be given to the auditory input.
The paper by Jusczyk stressed the importance of studying the ontogenetic buildup of the perceptual discriminating faculty in focusing on minimal consonant and vowel distinction in neonates and very young infants. The object is to determine the relation between the perceptual characteristics of each of these two classes of segments in conjunction with those of the syllable. Moreover, a language specific shaping is not ruled out and constitutes a combined research object with Mehler, who found evidence for a different handling of syllabification between speaker of different languages, such as French and English, in a number of experimental tasks involving reaction time measurements.

Remez, using a more global approach, concentrated on continuities in the speech signal which, as such, help in organising the overall patterning from which segmental information can be removed and replaced by sinusoids without too great detriment to intelligibility.

**PROPOSITION**

One may wonder, at this juncture, whether a fresh phenomenology of the speech chain is in order. This might help in setting up research objectives for the near future in which auditory sensation as the most peripheral input to the speech processing mechanism is also studied in its more global aspects and made to match better with the quintessential character of connected speech as a continuum. Any subsequent segmentation must be seen for what it is most of the time, i.e. derived from a linguistic analysis. While such a segmentation seems inevitable, it should be acknowledged that the choice of level of analysis is arbitrary or due to the experimenter's taste.

If we are interested in the organ of hearing we are inclined to come up with a different choice of units than if the aim is to go in for high quality synthesis in which perceptual criteria are invoked to obtain optimal intelligibility, naturalness, speaker identity, and variation.

Our conclusion may be that there is no such thing as The psychophysics of speech perception, yet the combined efforts of experts from a variety of disciplines are needed to make the study of speech perception a rewarding pursuit. In this pursuit the chase for perceptual units should be called off for the time being until the speech science community has sufficiently assimilated its awareness of the impact of linguistic propensities. In the meantime, contributions from the linguistic points of view may help in deciding which features in the speech chain are due to universal and which to more language specific properties.

It is my belief, as a linguist by training, that though linguistics in the past may have been a bane to speech studies at times, we should turn that into a boon, if properly dimensioned, in the future.