ABSTRACT

Speech perception is argued to be essentially a top-down process coming down stepwise from higher levels to lower ones, the higher levels being characterizable, from the phonological point of view, in terms of their prosodic features.

In Donald Norman's words, "it is usually thought that the analysis of speech requires levels of abstraction. For example, speech sounds are transformed into phonemes, and phonemes into words"/9, p.388/.
The analysis of this kind, nicknamed 'bottom-up', until very recently was accepted almost universally. The only alternative was presented by the one-time influential analysis-by-synthesis model which typically did not make use of the notion 'level of abstraction'. In other words, the predominant views link the very notion of levels to the more traditional 'bottom-up' model, while its top-down counterpart, the analysis-by-synthesis model, is not thought to require the notion. The reason seems to be rooted in a rather narrow understanding of what levels of abstraction may be like: it is believed by many that at the outset of the process of speech perception man deals with the phonetically 'richest', i.e. the least abstract, characteristics of the incoming signal, the latter being step-by-step 'refined' so as to dispose of communicatively irrelevant details.

Yet the psychology of vision /II/ as well as the still earlier ideas of N.A. Bernstein /I/ suggest a valuable hint to the contrary. It is argued that at the first stages of visual perception man deals with highly generalized, and therefore abstract, features of the object to be perceived. Such features are sufficient to get a 'rough idea' of what is being seen, i.e. to assign the object to a very broad class of entities. If the actual setting is informative enough to provide ground for (speech/visual) choice within the class, the object is recognized with all its relevant details without further analyzing its actual physical characteristics. If not, its lower-level features, less abstract and more specific, are to be brought into consideration until the information is made sufficient to identify the object.

It may be seen from the above very sketchy exposition that visual perception exhibits a clearly top-down character. On the other hand, it is hardly in line with the analysis-by-synthesis model -- at least not beyond the anticipation routines common to all perception strategies. It seems to be of primary importance that the perceptual process evolves as a stepwise progressing from a more abstract representation of an object to a more specific (concrete) one. That means, at the same time, that the process is hierarchical in nature. Levels of abstraction are also levels of control where the output of a higher level largely constrains and, for that matter, controls the functioning of the lower one.

If auditory perception is presumed to be essentially parallel to visual perception, then we may accordingly seek similar stages in processing the sound information. One of the crucial problems is singling out particular sound features to be assigned to higher levels of speech perception. Since in processing the speech signal the listener aims from the very outset at grasping its meaning, the features sought should be applicable to big speech chunks as possible. Clearly such features are most likely to be prosodic (suprasegmental), i.e. pertaining to intonation, stress (accent) or tone.

One possible method of investigating the relative role of suprasegmentals and segmentals (syllables, vowels, consonants) in speech perception is artificial distortion of certain acoustic parameters responsible for realization of particular segmentals or suprasegmentals, which gives an opportunity to see their contribution to the process. For instance, Price and Levitt /10/ report that insufficient prosodic information makes the /θ/ - /ð/ distinction unstable. They speak of pro-
body as of "an aid in initial parsing of a sentence" (p. 302). In our view, such data support the thesis that prosody plays a role in sentence prosody as compared to that of segments and tones which consists in the following: syllables are classified as being in and that this classification is carried out by the listener. The identification of segments is carried out by the listener under the influence of the tonic information, the listener finds it possible to determine the presence of a noun or a verb, for example.

The perception of tone is an important aspect of speech perception. However, it is not as well understood as pitch or stress. Tone is a pitch-related property of speech that is used in the perception of meaning and in the identification of words. Tone is also used in the perception of sentence boundaries and in the perception of intonation.

In our study, we sought to determine the role of prosody in the perception of tone. We conducted a series of experiments in which we presented listeners with speech containing tones that varied in pitch and duration. We then asked the listeners to identify the tone of each word in the speech.

The results of our experiments showed that listeners were able to identify tones with a high degree of accuracy. This indicates that prosody plays a role in the perception of tone.

In conclusion, our study suggests that prosody plays a role in the perception of tone. Further research is needed to determine the exact nature of this role and to understand how it interacts with other aspects of speech perception such as pitch and stress.
existence of modules operating in parallel. Besides, if the initial hypothesis about a word or, say, phrase is rejected as violating some regularities of mapping prosodic structures onto segmental ones the process is started anew — thus acquiring a shuttle-like character in its functioning.

REFERENCES