PERCEPTION OF STATIONARY FORMANT PATTERNS[1]

Hartmut Traunmüller[2]

The object of my research study was the role of intrinsic factors in perceiving stationary speech sounds. Features that are acoustically present within the segment in question are here called intrinsic factors, as distinguished from extrinsic (contextual) factors. As stimuli simple synthetic vowels with one or two stationary formants were used.

The point of departure of my discussion is not the acoustic signal but the representation of sounds along the basilar membrane, where the Bark scale of tonality z(F) is applicable. It has been presupposed that formants and pitch are the decisive features of the phonetic identity of sounds. The question to be answered was to what extent the positions of - and the distances between - the relevant parts of the sound representation are used in phonetic perception.

If distances are most essential, there seems to be no need to suppose a normalizing process in order to explain perceptual invariance for isolated vowels produced by children, women, and men, despite overlap in formant-frequency data of different phonemes, since these age- and sex-conditioned differences on the whole are reflected as a uniform displacement of

---

1) Summary of my paper entitled "Einige Aspekte der Wahrnehmung quasistationärer Vokale" in PILUS 32, p. 8-13, Stockholm Univ., oct. 1976.

2) Department of Speech Communication, Royal Institute of Technology (KTH), S-100 44 Stockholm 70, Sweden.

the sound representation along the basilar membrane. But there is a less than average displacement found in F_1 of non-open vowels produced by men.

One-formant vowels covering almost the whole range where phonetic interpreting is possible, were generated in steps of at most one Bark for F_o and the formant. For two-formant vowels F_1 was maintained somewhat above F_o, while F_2 on the one side, and F_o, F_1 on the other side were generated as above mentioned. F_o of the stimuli was non-stationary.

The vowels were identified by two Austrian and two Swedish subjects. In the dialect of the Austrian subjects there occur five distinctive degrees of opening among vowels, which results in good resolution among one-formant vowels, while in the dialect of the Swedish subjects there occur four maximally closed vowels, which results in good resolution among the 'closed' two-formant vowels.

The results showed:

1) that there is a lack of agreement between the vowel identifications of the different subjects, particularly with regard to those stimuli that were not perceived as natural vowels because of too high $z(F_o)$ or $z(F_1)$.

2) that the distance between F_1 and F_o $(z(F_1)-z(F_o))$ is decisive for the identity of one-formant vowels as long as the distance is less than approximately 6 Bark, otherwise the identity is determined mainly by $z(F_1)$.

3) that the phoneme boundaries between [i] - [y] - [ʉ] in two-formant vowels are determined mainly by $z(F_2)$. This cannot explain the phonetic similarity of /y/ produced by different speakers.

4) that one-formant vowels are identified not only as back vowels with corresponding F_1. Instead of the expected sequence [u o ɔ ɑ a] the following sequence [u o ɔ œ (ɶ) æ] was obtained from three of four subjects.

Result 1) points to the possibility that different persons may have different perceptual strategies. Results 2) and 3) are in accordance with the hypothesis that the phonetic identity primarily is determined by the distances between formants with a small mutual distance, including the voice fundamental (e.g. F_1, F_o or F_4, F_3, F_2 in front vowels). In addition, formant positions function as substitutes if adjacent formants are not perceived. No satisfactory explanation for result 4) has been found. It is, however, evident that this result cannot be brought to agreement with the idea that vowels are perceived by means of simple matching of their total spectra against memorized patterns.