PERCEPTION OF THE FIRST FORMANT CONSIDERING THE HARMONIC STRUCTURE IN THE SPECTRUM

Rolf Carlson and Björn Granström

Perception of the first formant is normally thought of as an estimation of the peak in the trasfer function of the vocal tract. But how could a human estimate this peak when most times there is no energy at that special frequency because of the harmonic structure of the voice source?

Three different hypotheses about the perception of the first formant could be proposed.

a) The listener can reconstruct the peak of the envelope from the perceived harmonics irrespective if there is energy at the peak or not.

b) The listener computes the weighted mean of the excitation in different low frequency areas along the basilar membrane.

c) The listener selects the largest peak of the auditory excitation pattern. At medium and high pitch a single harmonic is picked out.

In an identification test using isolated vowels with different F1 and F0 values we searched for the phoneme-boundary between [i] and [e]. The result showed that hypothesis c) had to be rejected. We also used different weighting techniques in order to get a parameter representing the most important frequency (MIF) in the F1 domain. Only one of the methods tried could be accepted, taking the result of the identification test into account, namely the weighted mean of the two loudest harmonics in the loudness (sone) space. This method gave an MIF quite close to F1.

Reference:

Carlson R., Fant G., and Granström B. 1973. "Two-formant models, pitch, and vowel perception", to be presented at the Symposium on Auditory Analysis and Perception of Speech. Leningrad, Aug. 1973.