TWO-FORMANT MODEL OF THE ACOUSTIC DESCRIPTION OF SPEECH ARTICULATION

N. Degtyaryov

Institute Eng. Cybern. Ac. of Sc. BSSR

ABSTRACT

The present report suggests and constitutes a hierarchical model for obtaining an acoustic description of speech articulation. The primary description of the formant model is suggested. A secondary, more precise description of the submodels is suggested. A two—formant model of speech articulation can be given by the parameters of the first three-four voice formants /1,2/. That is why each submodel can be described by means of its particular, different from the other submodels, set of significant formant characteristics. The necessity of consideration, while analysing, of these structural speech signal characteristics produced brings us to the following important conclusions. Firstly, the complete model of speech signal formant analysis must be structural and must include the submodels of speech sounds formation manners. Secondly, during the formant analysis process there must be a controlled commutation of submodels correlating with the nature of articulation manner of an analysed speech sound. Obviously such information can be obtained only on the means of phonetic context hypothesizing.

1. INTRODUCTION

In spite of the great efforts made, the problem of reliability in automatic extracting of formant parameters from speech signal, is far from being solved. The situation makes us think over some new approaches to the problem of formant speech analysis. The extracting methods of formant analysis of speech signals are based, as a rule, on a non-complete model of speech signals generation, the latter extracts amplitudes and frequencies of the first three-four voice formants /1,2/. That is why the most stable results of analysis can be obtained only on the segments, matching with the given model. The reasons and character of the mistakes found here (loss of the 3rd and 4th formants because of their low level as compared with noises, loss of the 2nd formant because of its shunting during nasalization or low resolving power of the spectrum analyser, etc./2/), demonstrate structural limitedness of the formant analysis models used as the result of that different characteristics of speech signal for different speech sounds, as regard the manner of their formation, are not taken into consideration. This brings us to the problem of an acoustic speech articulation description which considers structural characteristics of a formant model of speech signal generation.

2. STRUCTURAL MODEL OF SPEECH SIGNAL GENERATION

The universal theory of speech generation /3,4/ suggests an acoustic or equivalent electric submodels for each manner of speech sounds formation. The structure of each of the submodels is specific as it reflects articulation (speech organs shape, place and type of the excitation source) of one particular manner. That's why each submodel can be described by means of its particular, different from the other submodels, set of significant formant characteristics. The necessity of consideration, while analysing, of these structural speech signal characteristics produced brings us to the following important conclusions. Firstly, the complete model of speech signal formant analysis must be structural and must include the submodels of speech sounds formation manners. Secondly, during the formant analysis process there must be a controlled commutation of submodels correlating with the nature of articulation manner of an analysed speech sound. Obviously such information can be obtained only on the means of phonetic context hypothesizing.

3. HIERARCHY MODEL OF SPEECH ARTICULATION ACOUSTIC DESCRIPTION

The main conclusion we've come to, is that we can solve the problems of formant analysis of a speech signal only on the basis of a complete structural analysis model by means of synthesis using the information on the current phonetic context, hypothesized from the upper levels of a perception model. Some information on the problem can be found in /1,5/. Besides, the attempts to make a more or less complete mathematical speech signal model also result in a controlled structure /6/. According to our conception we must accept that the articulation of a formant is conditioned by a definite phonetic context. So detailed formant description can be given in symbolic and parametric representation and this is secondary by nature. Then there must be a certain generalized, but unconditioned and in this sense primary acoustic description of articulation, which forms an initial stage in the process of speech analysis and recognition.

4. ACOUSTIC DESCRIPTION OF ARTICULATION BY MEANS OF GROUP FORMANTS CHARACTERISTICS

To create a system of parameters of the primary articulation description, let's consider some general characteristics of speech sounds that can be found on the spectrum envelope. It happens so that we can extract four types of the spectrum envelope; they reflect the main formant characteristics of speech sounds. During the formant analysis process there must be a controlled commutation of submodels correlating with the nature of articulation manner of an analysed speech sound. Obviously such information can be obtained only on the means of phonetic context hypothesizing.
The number of counts, above the threshold $k \max a_j$, here is the physical meaning of the parameters evaluated by (I)-(3); they express the integral amplitude, frequency and band values of the spectral counts $a_j$, representing this formant group. The main qualities of the suggested parameters system and the algorithm of their extraction (separation) are: 1) the possibility of separating the two first formants even in the case of their mutual (reciprocal) masking /5/; 2) the possibility of reflecting different formant characteristics (See Fig. 1b and 1c) without separating the upper formants, that is, of course, the most difficult problem; 3) equal efficiency for reflecting formant characteristics of different, from the point of view of their manner of articulation, sounds.

Fig. 2 presents A,F parameters tracks for the words “adln” and “’s’se”.

5. CONCLUSION
The suggested system of A,F,B parameters of the speech articulation primary description is a good basis for the upper level analysis of the speech recognition model. Firstly, the parameters precisely reflect the speech spectrum formant characteristics. Secondly, they meet the demands of the linear model of parameter approximation /6/, which is a way towards the solution of the speech recognition problems. Thirdly, the suggested model provides the basis for describing some topologic invariants and, thus, contributes to the solution of the multispeaker recognition problems.

6. REFERENCES