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 consistutes a hierarchical model for obtaining an acoustic description of speech articulation. The primary description of speech articulation can be given by the parametres of two-formant model spectrum. A secondary, more precise articulation description is achieved by means of formant parametres of submodels, conditioned by different manners of speech sounds formation. A two-formant model of acoustic speech articulation description is suggested.

1. INTRODUCTION

In spite of the great efforts made, the problem of reliability in automatic extracting of formant parametres 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 analisis 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 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 be means of phonetic context hypothesizing.

3. HIERARCHY MODEL OF SPEECH ARTICULATION ACOUSTIC DESCRIPTION

Thus, 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, hypothesizen 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 mathematic speech signal model also result in a controlled structure /6/.According to our conception we must accept that the notion of a formant is conditioned by a definite phonetic context. So detailed formant description can be given in symbolic and parametric representation and thus 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 parametres of the primary articulation description, let's cosider some general characteristics of speech formants 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 articulation (See Fig.1.) In /5/ a system of integral parametres A, F, B reflecting the formant characteristics of the speech spectrum envelope is presented. There is also an evaluation algorithm of the parametres. The main point of the algorithm that in two spectral regions with the adaptive boundary separating them, the formant groups are described by moments from spectrum counts,grouping around the maximum one:

1) $A = \sum_{1}^{m} a_\nu / m$

2) $F = \sum_{1}^{m} f_\nu a_\nu / \sum_{1}^{m} a_\nu$

3) $B = \sum_{1}^{m} (f_\nu - f_{\nu-1}) a_\nu / \max_j a_j$

a_j – counts of the instantaneous power spectrum on frequencies f_j ; $j = \overline{1,n}$;

$\nu = \arg Z_j$; $Z_j = \mathrm{sign}(a_j - h \max a_j)$; $0 < h < 1$; m – the number of counts, above the threshold $h \max a_j$. Here is the physical meaning of the parametres evaluated by (I)-(3): they express the integral amplitude, frequency and band values of the spectral counts a_ν , representing this formant group. The main qualities of the suggested parametres system and the algorithm of their extraction (separation) are: I)the possibility of separating of 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 parametre tracks for the words "ə'din" and "ʃə'se".

5. CONCLUSION

The suggested system of A,F,B parametres of the speech articulation primary description is a good basis for the upper level analysis of the speech recognition model. Firstly, the parametres precisely reflect the speech spectrum formant characteristics. Secondly, they meet the demands of the linear model of parametre 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

/1/ КОРЕС ,G.(1986) "Formant tracking using hidden Markov models and vector quantization", IEEE Trans. on acoust. speech and sygnal process. Vol.AISP-34,No 4, 709-729.

/2/ БУХТИЛОВ Л., ЛОБАНОВ Б. (1988)"Алгоритм оценки формантных частот", Автоматическое распознавание слуховых образов (АРСО-I4).Ч.I: Материалы докладов и сообщений. Каунас, I0-II.

/3/ FANT, G.(1960),"Acoustic theory of speech production", The Hague:Mouton & Co.

/4/ FLANAGAN, G. (1965) "Speech analysis, synthesis and perception", Springer, Berlin-New-York.

/5/ ДЕГТЯРЕВ, Н. (I988) "Двухформантная аппроксимация спектров речи", Автоматическое распознавание слуховых образов (АРСО-I4). Ч.I.: Материалы докладов и сообщений. Каунас, I2-I3.

/6/ ВИНЦЮК, Т. (I982) "О математических моделях речевого сигнала, используемых в распознавании речи", Автоматичнское распознавание слуховых образов (АРСО-I2): Тезисы докладов и сообщений. Киев, 34-37.

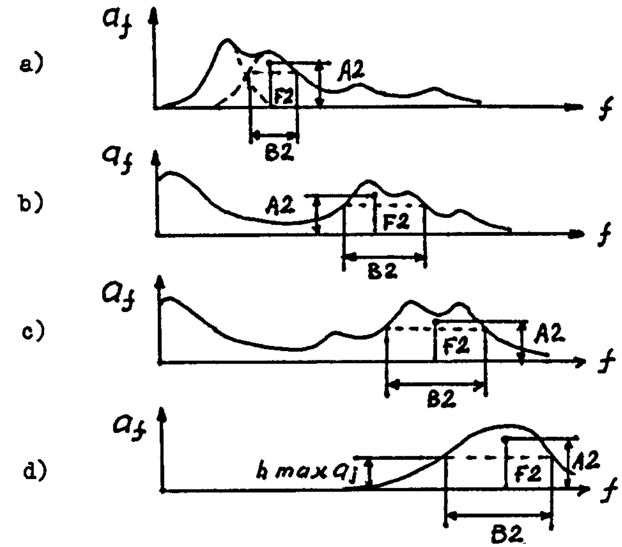

Fig.1. Formant groups and their generalised parametres for a) compact and aspirative; b) diffuse; c) nasal and voiced fricative; d) unvoiced fricative sounds of speech.

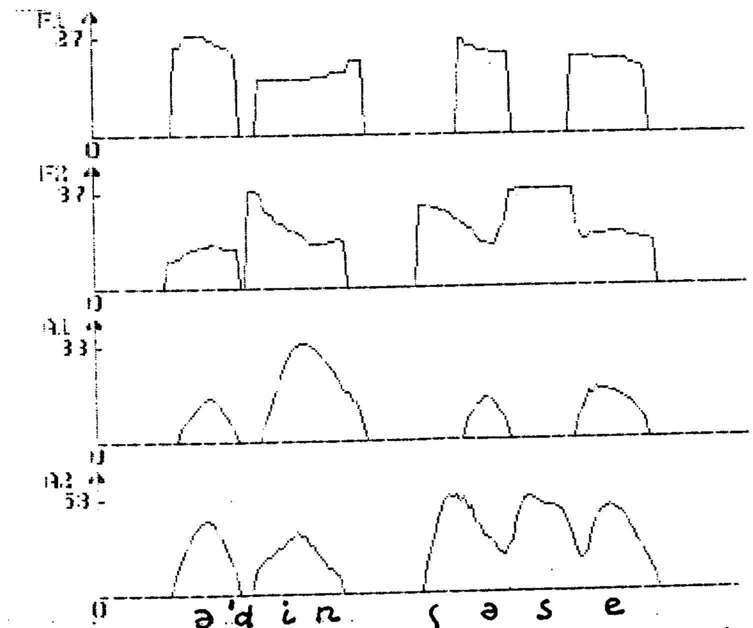

Fig.2. A,F parametre tracks for the words "ə'din" and "ʃə'se".