THE STUDY OF AUDITORY DETECTION OF THE JUMP OF FORMANT FREQUENCY AND AMPLITUDE AS A CONSONANT

V.V. Lyublinskaya, E.I. Stolyarova, S.Y.Zhukov

Pavlov Institute of Physiology of the Academy of Sciences
of the USSR, Leningrad, 195164

ABSTRACT

The perception of rapid changes (jumps) of formant frequency and amplitude in the spectrum of synthesized vowel was studied (in the experiments).

The boundaries of these changes associated with the consonants of different phonetic qualities were determined. Auditory images of studied stimuli in the form of space-time distribution of the responses of the detectors of amplitude irregularities in analyser frequency channels were received on the model. The character of the model representation of acoustic transitions from consonant to vowel was revealed.

INTRODUCTION

The subject of the study is auditory representation of "acoustic events" inherent in combinations of consonant and vowel phonemes in current speech. On dynamic spectrograms in these points one can observe rapid changes of formant frequency and amplitude, as well as those of the envelope amplitude.

It is known that the result of auditory analysis of formant transitions is used by man as phoneme determiners of diphones. As for automatic analysis of transitions it is known to be a difficult task, in particular, the formant frequency determination. For this reason it seems useful to apply the well-known principles of auditory processing for the analysis of transitions in speech signals.

According to some neurophysiological research the neurones in auditory system respond in a special way to the rapid changes in amplitude or spectral characteristics that occur in speech. The neurones which respond only to the positive or negative amplitude jumps (on- and off-responses) have been described in many papers/4/.

The simulations of such reactions was realized by the functional model of auditory determination of the amplitude irregularities (ADAI) /3,5/. It includes a model of peripheral spectral analyser (the "cochlea") and the system of the envelope processing in every frequency channel.

Positive and negative markers strictly localised in time are the responses to the respectively amplitude increase and decrease in the channels. The signal is represented in the ADAI model as the space-time distribution of the positive and negative markers in the channels. It was assumed that the markers might be used to form the segmentation function of speech flow and to sample the spectral information /3/. For this purpose, it is necessary to assume the integration of similar markers over the frequency channels. At the same time, on-and off-responses to narrow frequency signals may be strictly localized in frequency scale. This was also confirmed by the psychoacoustic data /1/. The narrow time and frequency localization of these reactions assumes the formation of space-time distributions as the response to the formant transitions.

This work was aimed to find out the possibility to use the responses of the ADAI model for the analysis of such acoustic events as the formant frequency and amplitude jumps. The present research has been inspired by the well-known fact that the jump of the formant frequency or the amplitude jump along the vowel-like segment of the signal is identified as a consonant and the whole signal as the syllable CV or VC depending on the direction of the jump /2,3/. The jump value determines the phoneme quality of the consonant. When the jumps are relatively large the stimulus is perceived as [m]V or [n]V, when the jumps are smaller - as [l]V /2/. The present research comprised 2 stages. Psychoacoustic experiments with synthesized vowels were carried out during the first stage. They were devised to determine the physical value of the jump of formant frequency and amplitude /ΔF1 orΔA1/, when they were identified as the consonants [m] or [l] . The stimuli with the studied characteristics were analised in the functional model of peripheral spectral analyser and in ADAI model during the second stage.

PERCEPTION EXPERIMENTS.

Synthesized two-formant vowels (192 ms-24 pitch periods, 8 ms each) were used in experiments. The parallel formant analog synthesizer generated the stimuli. The variations of stimuli parameters were realized in 2 ways, as shown in figure 1. The parameters $F1_c$ and $A1_c$ of the first segment (the "consonant" segment, 64 ms) were controled. The second formant was constant and it was 10 dB less than the level of the first one. The set of experiments has been done, each test included only one type of stimuli. The values of F1 and F2 of the synthesized vowels are shown in the Table.

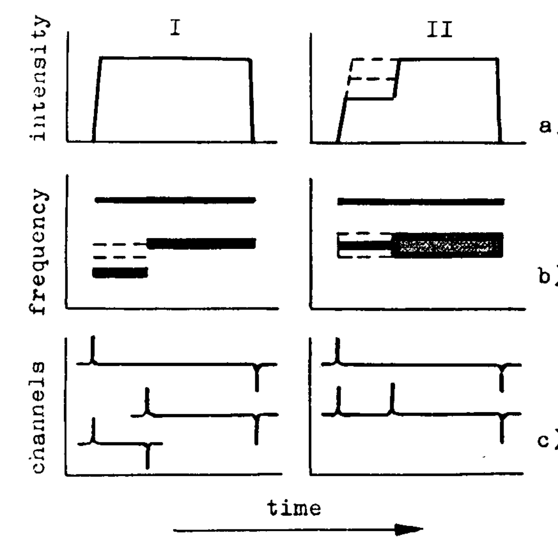

Fig.1. Structure of the stimuli in experiments: a) the amplitude envelope; b) the formant trackes; c) the markers in the channels of the ADAI model.

Table
The parameters of the stimuli and experimental data.

VOWEL	F2 (Hz)	$F1_V$ (Hz)	I ΔF(Hz) [m]	I ΔF(Hz) [l]	II ΔA(dB) [l]	II ΔA(dB) [m]	SUBJ.
u	625	400	120	60	4.4	11.1	1
			80	50	3.4	7.6	2
ɨ	1480	400	120	60	5.7	14.0	1
			100	60	4.8	11.2	2
i	2250	400	120	50	6.7	13.7	1
			100	60	4.3	12.9	2
e	1800	440	140	60	5.7	13.5	1
			130	50	4.3	11.9	2
o	780	535	175	55	4.3	9.6	1
			165	95	3.8	8.6	2
ε	1665	585	205	75	5.7	10.9	1
			205	115	3.8	11.5	2
a	1100	900	500	160	5.0	5.5	1
			500	210	3.8	9.0	2

Two methods - adjustment and identification - were applied in experiments. In the first case the subject controled the values of $F1_c$ and $A1_c$ to achieve the perception of the stimulus as [m] V or [l] V.

Results of the adjustment were registered by the experimenter.

According to the second method the sets of stimuli were presented to the subject, $\Delta F1$ and $\Delta A1$ being varied within definite limits.

Two subjects participated in adjustment experiments and five subjects - in identification experiments.

The results of the first type experiments are shown in the Table where the average values of $\Delta F1$ and $\Delta A1$ are indicated as the responses of each subject, when the stimuli were determined as the [l] V or [m] V syllables. The identification experiments data are analogous and therefore not described here.

The main properties of the perception of the jumps of formant frequency and amplitude are the following:

1. The perception of F1-jump depends on the quality of the vowel. The higher $F1_v$, the larger the jump $\Delta F1$, percepted as the consonant must be.
2. No regular dependence on $F1_v$ in the perception of the A1-jump is revealed.
3. The common feature inherent in perception of both frequency and amplitude jumps is revealed. The larger jump was identified as an [m], the smaller one as an [l] .

MODEL RESPONSES TO THE STIMULI

The sets of stimuli phonetically identified as V, [l] V, [m] V according to $\Delta F1$ and $\Delta A1$, were chosen for the analysis in the model. The markers distribution at the moment of the jump was examined for each stimulus. Two modes of operation were possible depending on the threshold value of markers generator: at the threshold of the detection of amplitude irregularity in the signal or at the threshold of the detection of the consonant while changing the amplitude of the signal.

Under the first condition the responses of the ADAI model were distributed on a wide frequency ranges. Under the second condition the markers were obtained in narrow frequency ranges near F1 and F2. We calculated and compared the number of channels were the markers could be registered at the moment of the jump. The number of marked channels correlate with the value of $\Delta F1$ or $\Delta A1$ under both conditions.

The patterns of the markers distributions were different for the frequency and amplitude jumps: only positive markers near $F1_v$ were registered for the amplitude jump, at the same time, the positive markers near $F1_v$, as well the negative ones near $F1_c$ were obtained for the frequency jump.

DISCUSSION AND CONCLUSION

The ADAI model reveals cues for the distinction between frequency and amplitude jumps, on the one hand, and allows to estimate the values of these both jumps according to the results of these experiments. We hope that the model features can help to describe the formant transitions in speech signals. The experimental data don't allow to make a conclusion about the information used by man for the phonetic interpretation of the frequency jumps. Whether he uses the time-frequency distribution of on- and off-responses only or he follows also the formant tracks. Possibly, both processings are necessary to provide the effective auditory perception of speech signals.

REFERENCES

I. Акустика речи и слуха: Сборник научных работ/Ред.- Л.А.Чистович./ - Л.:Наука, I986. - I44 с.
2. Жуков С.Я. Жукова М.Г. Восприятие слога в зависимости от его формантной структуры. - Физиология человека, I978, т.4, с.220-224.
3. Физиология речи. Восприятие речи человеком./ Чистович Л.А., Веннов А.В., Гранстрем М.П. и др./ Л., I976, 386 с. (Руководство по физиологии).
4. Delgutte B. Some correlates of phonetic distinctions at the level of the auditory nerve.- In: The representation of speech in the peripheral auditory system. Amsterdam, 1982, p.43-60.
5. Koshevnikov V.A., Stoljarova E.I. Segmentation of speech by a model of auditory system. - Symposium Franco-Sovietigue sur la parole, Lannion, 1983, p.95-105.