PERCEPTION OF SYNTHETIC VOWEL STIMULI WITH CHANGING ONSET F2-TRANSITION BY RUSSIAN AND FINNISH SUBJECTS

*V. Kouznetsov**, *A. Iivonen***
*Moscow Linguistic University, Moscow, Russia
**Helsinki University, Helsinki, Finland

ABSTRACT

The aim of the present paper is to investigate the perceptual role of the direction of the onset F2-transition in determining the phonetic vowel quality and to study language specific aspects of auditory analysis.

INTRODUCTION

Contemporary theories on vowel perception may be roughly divided into two major classes: those that consider information conveyed by onset and offset formant transitions essential to vowel identification (dynamic-specification models), and those assuming that all necessary information is contained at the vowel nucleus (target models).

Russian language provides a good opportunity to test these hypotheses due to the specific allophonic variation of stressed vowels in the context of palatalized and nonpalatalized consonants. Figure 1 depicts five russian vowel phonemes pronounced in the symmetrical contexts of palatalized and nonpalatalized fricative [s]. Formant frequencies were measured at the middle of the vowels [1]. Apostrophes surrounding vowels indicate palatalized environment. The plot reveals that several pairs of allophones can not be discriminated using F-pattern at the vowel nucleus, for example: ['U'-ɨ, ɨ-'O']. None the less, listeners rarely confuse these vowels. This is explained not only by the fact that the vowels occur in different consonant environments but by the differences of transition sections as well: at the onset of ['U'] F2 glides down, while at the onset of [ɨ] (Russian symbol for this vowel is [ы]) it rises up or stays level. Attempts to assess the perceptual significance of the direction of F2 transition in recognition of ['U] and [ɨ], using forced-choice identification of synthetic stimuli, produced contradictory results [2,3].

In the present study the same problem is addressed, employing different synthesizer and experimental procedures. To discover traces of language specific behavior the experiments reported below were conducted on russian and finnish subjects.

EXPERIMENT 1

In this experiment a modified ABX procedure was used to test the listeners ability to discriminate stimuli on the basis of the direction of F2-transitions.

METHOD

Stimuli. Test stimuli were synthesized using Klatt software synthesizer. Frequency and bandwidth of F1, F3, and F4 were kept constant: F1 = 300 Hz, B1 = 50 Hz; F3 = 2300 Hz, B3 = 200 Hz; F4 = 3300 Hz, B4 = 250 Hz. The frequency of the F2 was changed from 700 to 1900 Hz in 100 Hz step. The stimuli had either steady F2 or an onset F2-transition of $\pm$ 200 Hz or $\pm$100 Hz was added to them. Stimulus duration was 150 ms, transition duration - 55 ms. A falling F0-contour was used: 127-100 Hz. The voice amplitude increased from 55 to 60 db during the first 10 ms and fell down to 31 db on the last 35 ms. The stimuli were sampled at 10 kHz via a 10-bit D/A converter. A spectrogram of two tokens is presented in Figure 4.

Two experimental tapes were recorded using a speech processing tool-kit ISA (designed by R. Toivonen). Each tape contained 84 randomly ordered triads composed of stimuli with F2 transition either of $\pm$ 200 or $\pm$ 100 Hz. A triad consisted of stimuli differing only in direction of F2 transition. The third triad element was also realized by a vowel with an appropriate steady F2 (Z-element). Thus there were three types of triads: ABA, ABB, and ABZ. Both temporal orders of the first two elements were used in triad construction. The ISIs were 500 ms within triads, 2 s between triads and 5 s after each group of 15 triads.

Subjects and Procedure. Russian listeners were 14 students (19-20 years old) who took up an introductory course in phonetics. Ten finnish subjects took part in the experiments, their age varied from 23 to 52 years. Presentation was over loudspeakers in a quite room. The task was to identify the third vowel in a triad as A or B by circling the appropriate response on an answer sheet.

The tape composed of the stimuli with $\Delta F2$ = 200 Hz was presented first. It was preceded by 9 stimuli for familiarization and 10 practice triads. All perception test, including paired comparison of the Experiment 2, were conducted in one session that lasted about 45 minutes.

RESULTS

In order to find out whether the group of subjects was able to discriminate between stimuli A and B a χ^2-test was applied to each type of triads separately. Under the null hypothesis (no discrimination) the subject responses would split evenly between A and B and the expected frequency would be equal to half the number of subjects in a group.

The best discrimination performance of russian listeners was on ABB triads with $\Delta F2$ = 200 Hz: the probability of correctly rejecting the null hypothesis was 0.93 (χ^2 = 39.3 with 28 d.f.). On 19 (out of 29) ABB triads not less than 9 subjects identified the third stimulus correctly. Only in one case the listeners majority made a mistake.For the other two types of triads the level of significance was well above 0.10.

On ABA triads the russian listeners responded consistently (distribution of responses was at least 9/5 or 5/9) in 14 trials out of 28, but in half of them their judgement was incorrect. The same degree of response consistency was reached 14 times on the ABZ triads where Z was identified as B (in 10 cases B was realized by a stimulus with rising F2-transition).

The discrimination of stimuli with $\Delta F2$ = 100 Hz was more poor, but the general pattern of the responses was the same as described above.

The data of the finnish listeners is not reported here for it requires special treatment since the subjects were allowed not to take any decision in case of doubt.

EXPERIMENT 2

Trying to assess in this experiment the role of the direction of F2-transition in identification of ['U] and [ɨ] we did not consider it correct to ask russian and finnish listeners to perform the task of recognition because [ɨ] is an alien sound to Finnish language while ['U] is "unknown" to naive russian listener as a distinct vowel category for it does not occur in isolation and therefore its phonetic quality must be abstracted from the context. The perceptual importance of the direction of F2-transition may be studied by asking listeners to judge (dis)similarity between stimuli with rising and falling F2-transition and the two standard stimuli with steady F2 that are most similar to ['U] and [ɨ].

METHOD

In this experiment the same set of stimuli was used as in Experiment 1.

The stimulus with a steady F2 = 1700 Hz was used as a token of ['U]. Its formant frequencies are quite close to those of Finnish short [y] (see Figure 1. [4]). The second standard stimuli had F2 = 1400 Hz and as it was revealed by pilot observations its phonetic categorization was uncertain. If a rising F2-transition was added to it, it was definitely perceived as [ɨ].

Every stimulus with F2-transition was paired with the two standards in both orders. 10 times each standard was paired with itself. Separate tapes, each containing 130 pairs, were created for stimuli having $\Delta F2$ = 200 and 100 Hz. The stimuli with greater F2-transition were presented first. All the other experimental conditions were the same as in Experiment 1.

Procedure. The subjects were asked to judge each pair for its dissimilarity on a five-point scale on which 1 was considered most similar and 5 least similar.

RESULTS

Figures 2 and 3 show the medians of the dissimilarity judgements for each pair of stimulus and standard (regardless of their order) plotted against F2 values at the stimulus steady-state. Minimums of the curves point at the F2 of an appropriate standard. Examination of the Figures leads to the following conclusions. For both groups of listeners the main factor affecting similarity estimate is the difference between stimulus F2 value at the steady-state and that of the standard. The effect of the direction of F2-transition if $\Delta F2$ = 100 Hz is negligible. When $\Delta F2$ = 200 Hz and stimulus F2 was lower than that of the standard, stimuli with falling F2-transition were judged more similar to both standards than stimuli with rising F2. The relationship was reversed when stimulus F2 was higher than that of the standard. It seems, that the listeners based their judgement of similarity not only on comparison of the physical properties of the sounds but on categorical decisions as well: in Figure 3 stimuli having F2 ≤ 1000~1100 Hz were estimated equally similar to the standards though the difference in F2 continued to grow.

From the results reported and discussed above it is apparent that in the present experimental paradigm no evidence was obtained supporting the hypothesis that russian listener had some language specific rules for categorical interpretation of the direction of F2-transition.

ACKNOWLEDGEMENT

We are indebted to our students J. Välikangas, K. Laasonen and T. Sertun for their help in preparation of the experimental tapes, carrying out the experiments and a part of data analysis. Due to the telecommunication problems the major part of the responsibility for the paper contents lies with the first author.

REFERENCES

[1] Kouznetsov, V., and Ott, A. (1987), "Spectral properties of russian stressed vowels in the context of palatalized and nonpalatalized consonants", *Proc. of XIth ICPhS,* v.3, pp. 117-120.

[2] Kouznetsov, V. (1991), "A cross-linguistic study of perception of vowel stimuli with changing F2" (in Russian), *Proc. of the XVIth All Union seminar on automatic recognition of auditory images,* pp. 160-162.

[3] Lublinskaya, L. (1993), "Application of phonetic criteria in speech psycho-acoustics" (in Russian), In: *Issues in phonetics I* (eds. T. Nikolaieva et al.), pp. 274-287.

[4] Iivonen, A., and Laukkanen, A. (1993), "Explanations for the qualitative variation of finnish vowels", *Studies in logopedics and phonetics,* v.4, pp. 29-54.

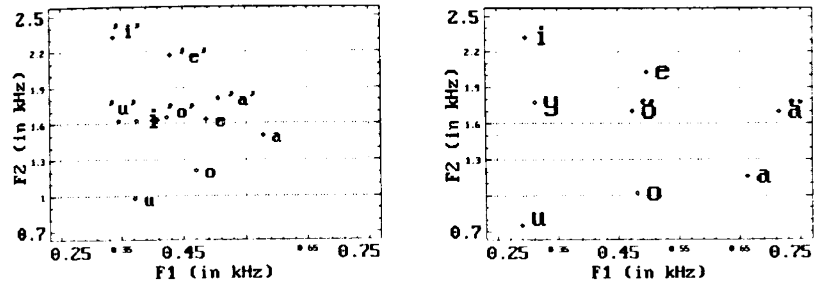

Figure 1. Vowel formant space. Left - russian stressed vowels in the context of palatalized and plain consonants. Right - finnish stressed vowels (data is averaged across long and short vowels).

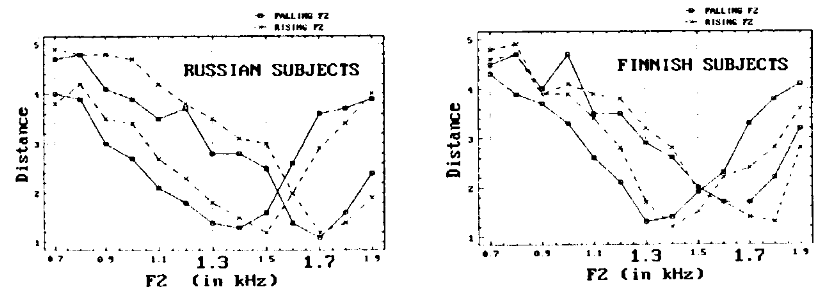

Figure 2. Median distances between two standards (steady F2 = 1400/1700 Hz) and stimuli with F2-transition of 200 Hz. Left - russian data, right - finnish data.

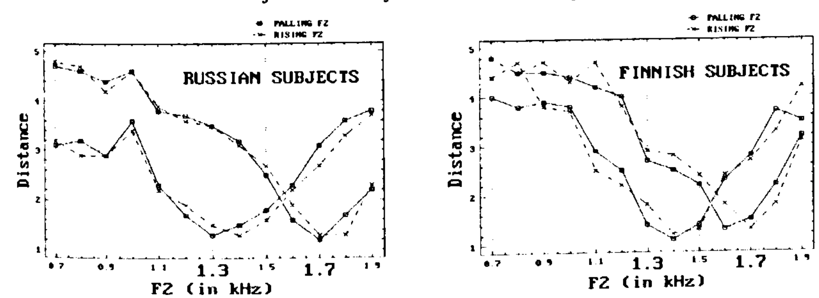

Figure 3. Median distances between two standards (steady F2 = 1400/1700 Hz) and stimuli with F2-transition of 100 Hz. Left - russian data, right - finnish data.

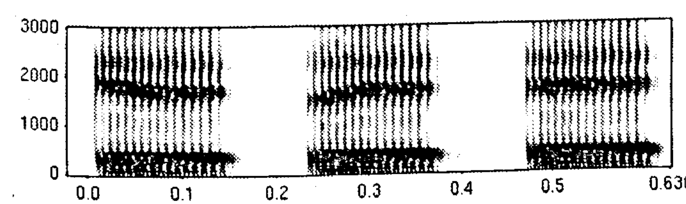

Figure 4. Spectrogram of test stimuli.