FORMANT TRANSITIONS AND RELEASE BURSTS AS PERCEPTUAL CUES FOR RUSSIAN VOICELESS PLOSIVES

V.N. LOZHKIN

Department of Applied Linguistics
Moscow State University
Moscow, USSR

ABSTRACT

The present study attempts to investigate the significance of the release burst and the formant transitions in the perception of Russian voiceless plosives by native listeners. The method involved deleting of releases in some consonants, it resulted in worsening recognition of these sounds - 34% for initial plosives, 60% and 70% for intervocalic and final plosives respectively. Thus, it is stated that release segments carry important information bearing on the place of articulation of Russian stops. The results of the study are in agreement with those obtained on the material of English and Hungarian stops and defy the prevailing significance of CV-transitions in voiceless plosives recognition.

INTRODUCTION

Although the history of experimental studies stimulated by the development of the 'Visible Speech' Sound Spectrograph dates back to the 40 ies, it is hardly possible to say that the distinctive features of phonemes are fully investigated. It is not surprising since the acoustic features of sounds in fluent speech can vary dramatically due to the context, the speaker's peculiarities, the mode of articulation, etc. Besides, the speech signal is highly redundant and possesses a great variety of distinctive features.

Quite a number of works have been dedicated to the distinctive features of plosives. It is a stated fact that information about the place of articulation of a stop can be found in the formant transition of adjacent vowels as well as in the stop burst. The relative significance of release and transition in the stop identification, however, is to be further investigated.

This problem is relevant for the systems of automatic speech recognition and high-quality speech synthesis. Another important problem concerns the search for invariant (i.e. independent of a context) features of the place of articulation/I/ /2/.

There exist at least four estimates of the relative significance of the release and transition cues.

On the one hand, under the influence of the studies carried out in Haskins Laboratories in the 50s on the material of synthesized syllables the view of the dominant role of the CV-transitions for the place of articulation identification has been adopted(The bulk of the results that became classical can be found in /3/). On the other hand, there are many indications that a formant transition might be of a smaller importance since the crucial information about the place is in the release. Such evidence has been obtained in some early studies of human speech/4-7/. One can argue that the transition and release cues are functionally equivalent, the former may be dominant in one case and the latter in another/8/. Finally, one can also argue that it is not correct to oppose transition and release since they may become inseparable in the case of a prevocalic position of a plosive/9/. It seems that all the approaches are sufficiently grounded(comparative analysis of various approaches is to be done elsewhere).

Keeping in mind that the results obtained in one language are not necessarily relevant for others and since much of the available information concerns English consonants we attempted to investigate the problem on the material of the Russian initial, intervocalic and final voiceless plosives. The relative significance of the transition and release cues in the perception of the place of articulation is discussed here. The problem is of particular interest since palatalized stops are very characteristic of the Russian language(we could mention only few articles dealing with the subject/I0-I2/)

METHOD

The model of a voiceless plosive is used according to which the four segments can bear information about the place of articulation of an intervocalic plosive: I)the segment of VC- or final transition; 2)the closure;3)the release after an abrupt closure breaks;4)the segment of a CV- or initial transition. The VOT was taken

for a release end. In a more detailed model a release is further divided into: a) a starting impulse;b)frication;c)aspiration(cf./I3/). A certain amount of residual noise may add to the voiced beginnig of a following vowel/IO/. Since these peculiarities of the Russian language are not phonemically relevant they are not considered here.

To investigate the relative significance of the transition and release cues for prevocalic, intervocalic and postvocalic plosives some meaningless successions or 'non-words' (=nonsense words) of the $C_1VC_2VC_3$-type were chosen, the consonant being the same in the one case (e.g. 'papap', 'totot', etc) and different in the other (e.g. 'patak', 'kopot', etc). The vowel has been taken out of the set (a,o,u,i,e), the second syllable of each non-word was stressed.

30 non-words were tape recorded by two male speakers. The instruction to the speakers was to utter the stimuli distinctly without changing the quality of vowels. The interval between the stimuli was 5 seconds.

The tape rings were made of the copies of the original recordings that underwent segmentation by means of the low-noise electronic separator described elsewhere /I4/.

A release for one of the plosives (initial, central or final) has been deleted in every non-word by means of the separator. The fragments of non-words with a release deleted were used as test stimuli. Thus the relative significance of the CV-transition, VCV-transitions and VC-transitions in place identification was studied.

The deletion procedure has been controlled aurally and by means of the oscilloscope. The test stimuli have been recorded on the second tape-recorder. Each fragment was recorded three times on a tape ring. The triads of the identical stimuli were separated by the pure monotone markers. The presentation rate of the test stimuli which depended on the ring length and the tape speed was about 3.8 seconds.

Non-words with release deleted were mixed with undamaged non-words and thus presented to ten listeners (students, laboratory assistants etc.) without hearing loss. Most of the listeners were experienced in listening to articulatory tests and synthesized speech patterns. The signal was fed into the head-phones in a quiet room. Each listener could adjust the volume in his head-phones.

The instruction given to the listeners was as follows: "You will hear non-words of the CVCVC-type, where C is any of the plosives – /p/,/p'/,/t/,/t'/,/k/,/k'/. Each non-word is repeated 3 times. After listening to a triad you are to write it down in ordinary letters ('papap' for example). If you detect a distorted (damaged) consonant please underline it as shown below: 'patat', or 'kutuk', or 'kakak'*)

Notice: the soft /k'/ may occur in a final position that are not typical for the Russian speech, e.g. petek' "

The instruction was presented by the experimenter orally and then its printed text was distributed among the listeners. The nature of the damage was not revealed to the listeners as well as the consonants of a non-word being the same $C_1=C_2=C_3$ or different $C_1\neq C_2\neq C_3$. Having learned the instruction the listeners began to listen to the test for a few minutes and then to listen and fix their judgements on special forms. Each listener listened to every test 3 times with a few days intervals.

A test consisted of I20 randomized non-words out of which 30 were not damaged, while 90 contained a stop with a deleted release. Thus a test contained in all 270 undamaged and 90 damaged stops. There were two non-word lists with different stimuli order, the first list was read by speaker L, and the second by speaker S.

RESULTS

The results of these tests are confusion matrices. A sample of such a matrix is given in the table below. The right and wrong judgements for the /p/, /t/,/k/ stimuli (with release deleted) preceeding /a/ vowel in the first hearing session (speaker L) are presented in the table.

Table

uttered \ perceived	p	t	k	–	p	t	k	–	p	t	k	–
p	I4		4	2	I7	2		I	20			
t	II	4	3	2	6	I4			3	I5		2
k	I3	4	2	I	I2	2	4	2	5	5	9	I
	38	8	9	5	35	I8	4	3	28	20	9	3

The judgements are summerized for ten listeners in every matrix. The "–"-sign stands for refusal. The left matrix corresponds to the initial position of a stop, the second and the third – to the central and the final positions, respectively.

It can be seen that in two presentations of a damaged initial /p/ (non-words 'papap' and 'patak'), I4 out of 20 judgements were correct and 4 were substitutions of /k/ for the initial /p/. There were also 2 refusals.It follows from the same matrix that the initial /t/ (in 'tatat' and 'takap') was given only 4 correct judgements, II responses were misjudgements for /p/, 3 – for /k/ and there were also two refusals. The numerals in the bottom row of the first matrix show the number of judgements of the /p/, /t/,/k/ or "–" -type in all 60 presentations of the initial voiceless plosives preceding /a/ (speaker L). There were 38 /p/-judgements, 8 – /t/, 9 – /k/, 5 – "–"; there were also 20 correct judgements (the sum along the main diagonal).The second and the third matrices are constructed in a similar way. There are 30 tables of the kind (2 speakers x 5 vowels x 3 presentations).For the sake of brevity all tables are not presented here.

For getting more reliable results the data for every speaker (3 presentations) were combined. One can further combine the data for both speakers, data concerning different vowels, etc. to answer the question – in what way a release deletion may influence the plosive recognition in general, when it is independent of the vowel context and the place of articulation. Figure I(a,b) shows the major result of the experiment. The graphs show that the place of articulation of damaged stops for central plosives is recognized better than for initial ones and for final plosives better than for central ones. At first, we assumed that greater intelligibility of central and final plosives on the second syllable being stressed. The analysis of non-words with the first stressed syllable revealed a similar tendency (graph L^*, fig. I,a). (For the sake of brevity more detailed data of the relative significance of transition and release in stop recognition is omitted here).

Fig. 1, a

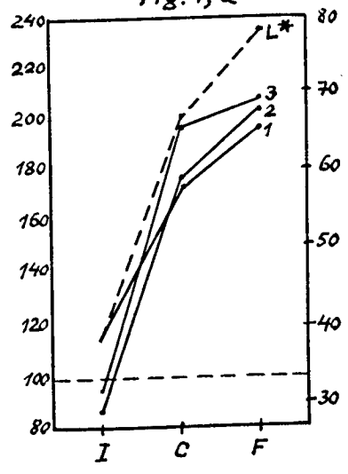
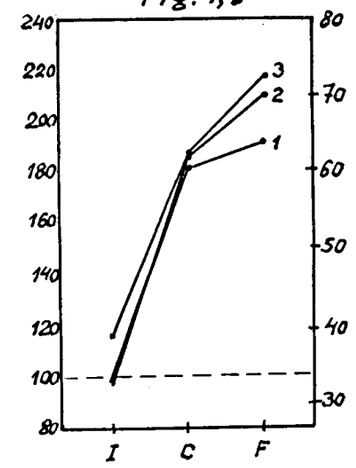


Fig. 1, b



Graphs a and b show the percentage of correct judgements as a function of the position of the damaged plosive. a – data for the speaker L; b – data for the speaker S. 'I' stands for the initial position of the plosive, 'C' – for its central, and 'F' – for its final position. The absolute numbers of correct judgements is plotted along the ordinate on the left, while the percentage of these judgements is to be found on the right. Curves I, 2, 3 correspond to the first, second and third listening to the non-words with the stress on the second syllable. Curve L^* shows the analogous results for the non-words with the first syllable stressed. Horisontal dash-line corresponds to guessing level.

Thus, the results of the tests show that in human speech a stop release carries very important information and its deletion may result in worsening recognizability of the place of articulation of voiceless plosives, especially of the initial ones. In the last case the recognition does not exceed the guessing level (~34%),*) while the information conveyed by VC-transition may provide higher recognizability for intervocalic plosives (60%) and for final plosives (~70%).

DISCUSSION

The results of the present study agree with the results of the other authors. Thus in /I5/ the relative significance of transition and release in the identificat-

*) It is necessary to mention that listeners detected damaged stops very poorly, marking the right consonants and missing the damaged ones.

*) It is not quite correct to mention the guessing level since the listeners judgements in this case were not mere guessig: most of them were judgements of /p/-type. The explanation might be that the release for /p/ is very short and faint so listeners may misjudge any voiceless plosive as /p/ when the release is absent.

ion of the place of English voiceless stops preceding /i/,/a/,/u/ vowels has been studied. It has been found that the voiced segment of the initial transition is neither a sufficient nor necessary cue for the place identification in the initial position, and it is the release that accounts for a correct consonant identification.

Similar conclusions were made for the Hungarian voiceless stops in VCV-syllables/I6/. The release was judged to be the most informative among the segments of closure, release and transitions due to /7/, where the distribution of stop cues has been studied.

A few studies of Russian voiceless stops have been described in /I0-I2/. It was found that in the final position of a stop the release is more important cue than the transition(the method consisted in transplanting bursts from one context into another)/I2/. The significance of bursts and final transitions in final positions of stops has been studied in detail in /I0/ on the CVC-syllables. It was shown that in most cases an isolated burst of final stops was sufficient for the identification of place. Final transitions can also carry information about the place of articulation.

In the article /II/, however, VC-transitions were not found significant in the perception of plosives.(A more detaled comparative analysis is required to explain the disparity between /II/ and the present study).

The question remains - which of the two transitions, CV or VC, is more informative in the place identification of the plosives? The results of the present study suggest that VC-transitions is more informative than CV-transitions.Similar conclusions can be found in /I7/, where the role of CV- and VC-transitions in the place of articulation of English stops and fricatives in syllables with neutral /ə/ in natural speech was determined.VC-transitions proved to be more informative than CV-transitions especially for voiceless stops: VC-transitions accounted for 92% of their intelligibility whereas CV-transitions accounted for only 32%. The corresponding values for voiced plosives were 92% and 7I%. According to the authors the reason for VC-transitions being more informative is their better physical manifestation. The conclusion has been substantiated by inverse listening data in particular.

SUMMARY

The results of the present study provide further evidence of the relative significance of transition and release as perceptual cues for Russian voiceless plosives. The regularities which have been found may prove useful in developing more efficient automatic recognition systems and high-quality speech synthesis.

REFERENCES

I.K.N.Stevens,S.E.Blumstein. Invariant cues for place of articulation in stop consonants. J.Acoust.Soc.Am. I978,v.64, N5,I358-I368.

2.D.Kewley-Port,D.B.Pisoni,M.Studdert-Kennedy. Perception of static and dynamic acoustic cues to place of articulation in initial stop consonants. J.Acoust.Soc.Am., I983,v.73,N5,I779-I793.

3.J.L.Flanagan.Speech analysis,synthesis and perception.Springer Verlag,I965

4.M.Halle,G.W.Hughes,J.-P.A.Radley. Acoustic properties of stop consonants. J.Acoust.Soc.Am.,I957,v.29,NI,I07-II6.

5.A.Malècot.The role of releases in the identification of released final stops. Language,I958,v.34,N3(part I), 370-380.

6.W.S.-Y. Wang.Transition and release as perceptual cues for final plosives.J. Speech Hear.Res.,I959,v.2,NI,66-73.

7.P.W.Guelke,E.D.Smith. Distribution of information in stop consonants. Proc. IEE,I963,v.II0,N2,680-688.

8.M.F.Dorman,M.Studdert-Kennedy,L.J. Raphael. Stop consonant recognition:release bursts and formant transitions as a functionally equivalent, context-dependent cues. Perception and Psychophysics, I977,v.22,N2,I09-I23.

9.S.E.Blumstein,K.N.Stevens. Perceptual invariance and onset spectra for stop consonants in different vowel environment. J.Acoust.Soc.Am.,I980,v.67,N2,648-662.

I0. Дукельский Н.И. Принципы сегментации речевого потока. М.-Л. изд-во АН СССР. I962.

II. Люблинская В.В. Распознавание артикуляторных признаков смычных согласных по переходу от гласного к согласному. Акуст.журнал,I966,т.I2,вып.2,2I3-22I.

I2.Охарева Н. Г. Взрыв как основной показатель места образования конечных глухих взрывных согласных."Исследования моделей речеобразования и речевосприятия" Л.I98I,с.I40-I48.

I3.G.Fant.Stops in CV-syllables.Speech Transmission Laboratory.Quarterly Progress and Status Report.STL-QPSR 4/I969, I-25.

I4.Речь.Артикуляция и восприятие.М.-Л. Наука,I965,с.62.

I5.C.LaRivière,H.Winitz,E.Herriman. Vocalic transitions in the perception of the voiceless initial stops.J.Acoust.Soc. Am.,I975,v.57,N2,470-475.

I6.K.Vicsi.The most relevant acoustical micro segment and its duration necessary for the recognition of unvoiced stops.Acustica,I98I,v.48,NI,53-58.

I7.D.J.Sharf,T.Hemeyer.Identification of place of consonant articulation from vowel formant transitions.J.Acoust.Soc. Am.,I972,v.5I,N2(part 2),652-658.