A RE-EXAMINATION OF STOP IDENTIFICATION FROM VOCALIC TRANSITIONS: LANGUAGE-SPECIFIC VS. PERCEPTUALLY UNIVERSAL FACTORS (EVIDENCE FROM RUSSIAN)

Vladimir Kouznetsov
Moscow State Linguistics University, Russia

ABSTRACT

Previous studies of relative importance of initial (CV) and final (VC) vocalic transitions as cues for identification of deleted stops have revealed the role of various factors that may be roughly classified as linguistic and perceptual ones. The present paper describes two experiments designed to investigate the interplay of these factors, using Russian speech as the source of the test stimuli. Results of the first experiment concerned with identification of voiceless palatalized and nonpalatalized stops in initial, final and intervocalic positions have shown high significance of all linguistic factors and some of their two-factor interactions. Perceptual dominance of VC transition was evident only in the case of palatalized stops. In the second experiment the method of conflicting transitions was applied to investigate perceptual weight of the formant transitions for recognition of intervocalic stops. The major finding of the experiment was that no asymmetry between initial and final transitions was established either for palatalized or nonpalatalized stops. As a rule, linguistic factors outweighed perceptual ones. But in those cases where the former were neutralized the effect of perceptual factors was apparent.

1. INTRODUCTION

Relative perceptual weight of the CV and VC vocalic transitions as cues for identification place of articulation (PA) of deleted stops has been examined for several decades now.

It was convincingly demonstrated that perceptual significance of formant transitions is mainly conditioned by such factors as language specific coarticulation processes and linguistic experience of the listener. Thus, for example, in English VC transitions provide better cues for identification of stops than CV transitions [5] while in Japanese the distribution of information load is reversed [2] and in Dutch both transitions provide equal cues for stop identification [6]. On the other hand, the findings of a number of experiments, using for identification of intervocalic stops stimuli with conflicting transitions[1], unambiguously indicate that the observed superiority of CV transition should be attributed to perceptual factors (recency effect and backward masking, for instance). In a cross-linguistic study American and Japanese subjects identified intervocalic [b, d, g] in more than 80% of cases using information from conflicting CV transitions irrespective of the playback direction of the stimuli [1]. It was concluded that the CV dominance effect is due not to a physical asymmetry arising from speech production factors but rather to perceptual reasons.

It was interesting to study the interaction of linguistic and perceptual factors in Russian speech, because there are two types of consonants (palatalized and nonpalatalized) and it is well known that CV transitions for palatalized consonants bear little information on their PA as the formant pattern at the beginning of the vowel has uniformly characteristics close to the vowel [i].

The data on perceptual importance of vowel transitions in Russian is rather few, fragmentary and conflicting. According to Lublinskaja [4] the VC transition does not provide enough information for the recognition of the following stop. But Lozhkin and Savinsky [3] obtained the highest correct identification scores (about 70%) for the same transition, while the score for CV transition came close to chance.

The goal of the present study was to collect basic data for Russian on perceptual significance of vocalic context (VCONT) for identification of deleted stops and to assess the role of linguistic and perceptual factors. Keeping in line with the research objectives, two experiments were conducted. The first one was concerned with assessment of perceptual significance of vocalic transitions for identification of voiceless palatalized and nonpalatalized stops in initial, final and intervocalic positions. In the second experiment the relative importance of the formant transitions for recognition of intervocalic stop was studied using stimuli with conflicting transitions. Both experiments made use of the same speech material.

2. EXPERIMENT I

2.1. Method

2.1.1. Speech material. It consisted of 54 V1CV2 structures in which vowels were realized by [i, a, u] in all possible combinations in initial and final positions while C = [p, t, k, p^j, t^j, k^j]. V2 was stressed.

2.1.2. Speakers. Two male speakers were recorded in a sound treated room using a microphone AKG C 1000S and a DAT-recorder SONY TCD-D7. The frequency of sampling was 48 kHz and the size of sampling - 16-bit.

2.1.3. Stimuli. Three types of stimuli were made up from the recorded utterances: V1(CV2), (V1C)V2 and V1(C)V2 (the brackets surround the deleted segment). For all V1(C)V2 stimuli the duration of the pause was uniformly set to 170 ms (average duration of the consonants of both speakers). This was done to prevent the listener from using closure duration as a cue for PA recognition. For each speaker 6 test series were constructed (3 types of stimuli x 2 types of stops (palatalized/nonpalatalized)). In each series 27 different stimuli were repeated 5 times in random order. Each stimulus was presented once and the duration of ISI interval was 2 sec. At the beginning of the first series listeners were made familiar with the stimuli and the response forms by practicing on 20 examples of the test stimuli. The remainder of the series were preceded by a sample set of 10 stimuli.

2.1.4. Subjects. Ten normal-hearing students (4 males among them) participated in the listening tests. The age of the subjects ranged from 18 to 22 years.

2.1.5. Procedures. The series of stimuli were administered in the following order: V1(C)V2, then V1(CV2) or (V1C)V2. The presentation order of the test series of different speakers, of stimuli with palatalized and nonpalatalized stops was balanced across- and within-subjects. During test session listeners heard the stimuli only of one speaker. The stimuli were presented through headphones Sennheiser HD 459 II in a sound-treated room. The subjects were asked to indicate which of the three stops they heard (a 3-AFC classification task).

2.2. RESULTS. The results of a fully factorial ANOVA (2x3x3x3x3x2-design) with feature palatalized/nonpalatalized (P/NP), PA, VCONT, V1, V2, and SPEAKER as fixed factors, performed on the percentage error identification scores of the three types of stimuli, showed that all main effects and some interactions were highly significant ($p<0.0000$). The most important main effects were (in descending order) P/NP, PA, VCONT, and the interaction P/NPxPA.

Table 1 presents averaged percentage error scores of identification of palatalized and nonpalatalized stops as a function of VCONT, PA and vowel quality. The values marked by star correspond to chance recognition. For V1CV2 context the indicated vowel is related to V2 if the consonant is nonpalatalized, and to V1 if it is palatalized. This vowel specification is determined by the relative weight of V1 and V2 as factors in variance models.

The table 1 shows that [p] was practically perfectly recognized from vowels [i, a], but the performance drastically degraded in the case of [u] and identification of the initial stop reaches chance level. As compared to [p] error score of [p^j] increases for [i, a] but diminishes for [u] especially in the intervocalic position, though identification of initial [p^j] is also at chance level.

Consonant [t] is recognized very reliably. The identification of [t^j] drops sharply in all cases except for the combination [ut^j].

There is a striking difference in the recognition scores for [k]: the error score is not greater than 5% in case of [a], but for vowels [i, u] it is on the average around 35%. Initial [k] could not be identified from transitional segment of [i]. Identification of [k^j] from formant transitions of vowels [a, u] significantly degrades, coming closer to chance. Remarkably, the recognition of initial and final [k^j] from vocalic transitions of [i] improves as compared to the identification of [k] in these positions.

From the analyses of the tabulated data it is clear that mechanical averaging across linguistic variables is unjustifiable. Yet in some cases generalization is plausible. It should be noted that consonants [p^j] and [t] are more reliably recognized than the other palatalized and nonpalatalized stops. The identification of both types of velar plosive is the poorest. On the average PA of nonpalatalized stops is identified better than that of palatalized ones. Vowel quality produces greater influence on the identification of nonpalatalized consonants than on palatalized ones. Recognition of all nonpalatalized consonants, including [k], is almost perfect from formant transitions of [a].

A small advantage of the final transition over the initial one in the case of nonpalatalized stops becomes quite substantial for the palatalized consonants. In most cases the identification of palatalized stops from CV transitions is close to chance. In general, the information load of vocalic context for the intervocalic consonant is comparable to that of VC transition. But for [t^j] and [k] the tendency is reversed and what is more in opposite directions: for recognition of [t^j] intervocalic context provides poorer cues than VC transition while integration of unreliable information of final and initial transitions of [i] and [u] increases identification of intervocalic [k].

3. EXPERIMENT II

3.1. METHOD

3.1.1. Stimuli. The test stimuli were constructed from the same speech material that was recorded in Experiment I. Symmetric vocalic context for [p, t, k, p^j, t^j, k^j] was made up of [a, i, u] vowels with conflicting transitions. As in Experiment I the intervocalic closure duration was 170 ms for all stimuli. Conflicting contexts were formed separately for palatalized and nonpalatalized consonants. Thus, the stimulus set consisted of 36 items. For each speaker a separate test series was constructed. Each stimulus occurred 3 times in a random order in a series.

3.1.2. Subjects and procedures. Ten normal-hearing students (others than those employed in Experiment I) participated in the listening tests. They were instructed to write down what they heard using Russian orthography (an open classification task). The experiment was run under computer control. Each stimulus was presented two times with 250 ms ISI. Then the listener could playback the stimulus as many times as he wanted. The next stimulus was presented when the listener pressed a key of the key-board.

3.2. RESULTS. Subjects responses were classified into several categories. The first category comprises responses in which two consonants were recognized. In its turn this category was divided into two subcategories: one for the whole set of the responses and the other for the correct identification of stop cluster. The response was judged to be correct if the PA of the identified consonants was identical to that of the stimulus stops. The second category consisted of the responses identifying a single stop. Here the number of subcategories was determined by the number of possible stops. The last category comprised the responses in which the subjects were unable to identify intervocalic segment.

In the three-way Tables 2 and 3 the percentage distribution of the responses is presented for nonpalatalized and palatalized consonants of both speakers. The column titled ‘Stimuli cons’ indicates the consonant cluster of the stimulus, while its vowel context is specified, for example, by the column title ‘Vowel context i-i'. Title abbreviation 'Correct' designates correct responses.

Let us start with the examination of the two-consonant identifications. Unlike the previous research (see, for example, [1]), the findings of the present study do not support the view that CV formant transition is perceptually more prominent than VC transition. It was established that Russian listeners were able to use the information of both transitions and to identify a consonant cluster.

Statistically significantly (from here on statistical significance implies at least a 5% level of significance) the identification of a cluster of nonpalatalized stops is better than that of the palatalized ones. The perception of the latter is significantly influenced by the cluster type: the combination [$t^j p^j$] had the highest intelligibility. Identification of nonpalatalized consonants is significantly affected by the quality of the surrounding vowels. The highest recognition scores were obtained from [a] context.

Analyses of the single-consonant responses reveals that the subjects gave no preference to either formant transition identifying both types of the consonants. Their choice between conflicting transitions was easily predicted from the data on consonant "identifiability" obtained in Experiment I. Thus, the fact that 80% of responses for stimulus [utpu] were [t], while for [itpi] near 70% of responses were [p], agrees well with the finding that [p] is poorly recognized from [u] and [t] - from [i] (see Table 1). Similar explanation could be provided for perceptual superiority of the final transition for stimulus [ukpu] and of the initial transition for [ikpi]. From this it may be concluded that identification of the intervocalic consonant from conflicting transitions is primarily conditioned by linguistic factors. It does not follow from this conclusion that perceptual factors play no role in consonant identification. One of the findings of the Experiment I is that all nonpalatalized consonants are most reliably identified from the vocalic transitions of vowel [a]. When this vowel formed conflicting context for intervocalic nonpalatalized stops, identification of one consonant was always based on the CV transition (see Table 2). It could be assumed that in this case where linguistic variables are to a certain extent 'neutralized', the process of identification is more strongly governed by perceptual factors.

Practically complete absence in the identification of palatalized stops of [k^j] responses and apparent dominance of [p^j] responses are consistent with the data of the Experiment I on "identifiability" of these plosives. This could be interpreted as another example of the prevalence of the linguistic factors over perceptual ones.

Significantly rarely the identified stop is not cued by any of the transitions. This 'strange' consonant response is more often used in recognition of palatalized stops. In this case the role of 'strange' stop is regularly assumed by [p^j]. The same consonant is more frequently identified from the conflicting transitions. The reverse tendency is characteristic to [k^j, k] stops. The tendency reaches the level of significance only in the case of the palatalized stop.

The comparison of the response distribution in the columns with the common heading 'One cons' reveals that for palatalized consonants (Table 3) in all vocalic contexts and for nonpalatalized stops in the [i-i] context (Table 2) there is a very similar pattern of percentage distributions. This observation is supported by significant correlations established among the patterns. This finding could be considered as an indication that phonetic quality of the vocalic context is of minor importance in these cases.

4. CONCLUSION

This study of the identification of place of articulation of deleted voiceless stops from vowel formant transitions showed that:

- The relevance of linguistic factors is much greater than that of perceptual ones. In cases where linguistic variables are to a certain extent 'neutralized', the process of identification is more strongly governed by perceptual factors.
- From the analyses of the data obtained it is clear that mechanical averaging across linguistic variables is unjustifiable.
- Consonants [p^j] and [t] are more reliably recognized than the other palatalized and nonpalatalized stops. The identification of both types of velar plosive is the poorest.
- Vowel quality produces greater influence on the identification of nonpalatalized consonants than on palatalized ones.
- A small advantage of the final transition over the initial one in the case of nonpalatalized stops becomes quite substantial for the palatalized consonants.
- In general, the information load of vocalic context for the intervocalic consonant is comparable to that of VC transition.
- Russian listeners are able to use the information of both conflicting transitions and to identify a consonant cluster.

NOTES

1. Conflicting transitions are constructed by pairing vowels with formant transitions cueing different place of articulation of the intervocalic consonant.

REFERENCES

[1] Fujimura O., Macchi M.J. and Streeter L.A. 1978. Perception of stop consonants with conflicting transitional cues: a cross-linguistic study. *Language and Speech*, 21, 337-346.

[2] Kashino M., Wieringen A. van and Pols L.C.W. 1992. Cross-language differences in the identification of intervocalic stop consonants by Japanese and Dutch listeners. *Proc. of Intern. Conf. on Spoken Language Processing'92*. Banff,. vol. 2, 1079-1082.

[3] Lozhkin V.N., Savinskij V.G. 1989. Distinctive features of stops. *Speech informatics*,.55-63. (in Russian)

[4] Lublinskaja V.V. 1966. Recognition of acoustic cues of stops from VC transition. *Acoust. Journal*. 12 (2), 213-221.(in Russian)

[5] Ohde R., Sharf D.J. 1981. Stop identification from vocalic transitions plus vowel segments of CV and VC syllables: a follow-up study. J. *of the Acoust. Soc. of Am.*., 69 (1), 297-300.

[6] Van Wieringen A. 1995. *Perceiving dynamic speechlike sounds.* Amsterdam.

	i			a			u		
	V1C	**V1CV2**	**CV2**	**V1C**	**V1CV2**	**CV2**	**V1C**	**V1CV2**	**CV2**
p	6	2	1	2	1	1	22	48	64*
$\mathbf{p^j}$	6	8	26	8	11	10	19	13	57*
t	2	6	11	3	0	3	1	0	4
$\mathbf{t^j}$	15	31	59*	28	39	49	1	18	37
k	42	18	52*	4	4	5	39	17	41
$\mathbf{k^j}$	22	27	44	50	30	87*	89*	69*	63*
Means	16	13	32	16	14	26	29	28	44

Table 1. Averaged error scores of identification (in %) of palatalized and nonpalatalized stops as a function of VCONT, PA and vowel quality. Percentage errors are starred where correct identification was not better than chance at 0.05.

Vowel context: **a-a**

Stimuli cons	Two cons		One cons			Non identified
	All	Correct	**p**	**t**	**k**	
kp	78.6	76.7	15.0	0.0	6.7	0.0
kt	58.3	55.0	0.0	31.7	10.0	0.0
pk	39.0	35.6	8.5	0.0	52.5	0.0
pt	50.0	40.0	3.3	46.7	0.0	0.0
tk	38.3	31.7	0.0	13.3	46.7	0.0
tp	63.3	50.0	28.3	3.3	1.7	0.0

Vowel context: **u-u**

Stimuli cons	Two cons		One cons			Non identified
	All	Correct	**p**	**t**	**k**	
kp	10.3	5.2	0.0	0.0	53.4	36.2
kt	20.0	8.3	0.0	60.0	13.3	6.7
pk	1.7	0.0	35.6	5.1	47.5	10.2
pt	6.8	5.1	16.9	76.3	0.0	0.0
tk	30.0	13.3	1.7	66.7	0.0	1.7
tp	20.3	1.7	1.7	78.0	0.0	0.0

Vowel context: **i-i**

Stimuli cons	Two cons		One cons			Non identified
	All	Correct	**p**	**t**	**k**	
kp	30.0	15.7	68.6	0.0	0.0	1.4
kt	23.3	15.0	11.7	55.0	10.0	0.0
pk	13.3	0.0	75.0	11.7	0.0	0.0
pt	20.0	20.0	48.3	31.7	0.0	0.0
tk	8.5	0.0	20.3	71.2	0.0	0.0
tp	33.3	26.1	66.7	0.0	0.0	0.0

Table 2. Results of identification of nonpalatalized stops (percent response)

Vowel context: **a-a**

Stimuli cons	Two cons		One cons			Non identified
	All	Correct	$\mathbf{p^j}$	$\mathbf{t^j}$	$\mathbf{k^j}$	
$\mathbf{k^jp^j}$	13.3	1.7	45.0	26.7	15.0	0.0
$\mathbf{k^jt^j}$	4.4	0.0	42.6	45.6	5.9	1.5
$\mathbf{p^jk^j}$	5.0	0.0	95.0	0.0	0.0	0.0
$\mathbf{p^jt^j}$	8.3	5.0	90.0	1.7	0.0	0.0
$\mathbf{t^jk^j}$	23.3	0.0	38.3	38.3	0.0	0.0
$\mathbf{t^jp^j}$	21.4	20.0	20.0	20.0	0.0	0.0

Vowel context: **u-u**

Stimuli cons	Two cons		One cons			Non identified
	All	Correct	$\mathbf{p^j}$	$\mathbf{t^j}$	$\mathbf{k^j}$	
$\mathbf{k^jp^j}$	15.0	0.0	55.0	21.7	0.0	8.3
$\mathbf{k^jt^j}$	6.7	1.7	43.3	41.7	3.3	5.0
$\mathbf{p^jk^j}$	8.5	0.0	66.1	13.6	8.5	3.4
$\mathbf{p^jt^j}$	6.7	1.7	53.3	35.0	3.3	1.7
$\mathbf{t^jk^j}$	26.7	5.0	20.0	48.3	5.0	0.0
$\mathbf{t^jp^j}$	20.0	20.0	26.7	48.3	3.3	1.7

Vowel context: **i-i**

Stimuli cons	Two cons		One cons			Non identified
	All	Correct	$\mathbf{p^j}$	$\mathbf{t^j}$	$\mathbf{k^j}$	
$\mathbf{k^jp^j}$	13.6	13.6	71.2	1.7	6.8	6.8
$\mathbf{k^jt^j}$	13.6	0.0	33.9	15.3	13.6	23.7
$\mathbf{p^jk^j}$	3.4	0.0	91.4	1.7	3.4	0.0
$\mathbf{p^jt^j}$	0.0	0.0	96.7	1.7	0.0	1.7
$\mathbf{t^jk^j}$	10.2	0.0	11.9	45.8	8.5	23.7
$\mathbf{t^jp^j}$	30.5	18.6	50.8	13.6	0.0	5.1

Table 3. Results of identification of palatalized stops (percent response)