TEMPORAL CUES IN THE PERCEPTION OF THE VOICING CONTRAST IN RUSSIAN

S. M. E. Barry

Department of Psychology and Speech Pathology,
Manchester Polytechnic

ABSTRACT

The perceptual role of several temporal characteristics as cues to the voicing contrast in Russian is investigated. The following parameters were examined: duration of the closure, duration of the preceding vowel and duration of voicing during the closure. In identification experiments, all three factors contributed to cue the contrast. The duration of closure voicing was an important cue. Consonant and vowel duration contributed either when they were co-operating with the voicing cue or when the voicing cue was ambiguous. In addition, the results support a model where the absolute duration of voicing, rather than voicing duration relative to closure duration, is perceptually relevant.

1. INTRODUCTION

The relevance of temporal cues to the voicing contrast has been widely investigated, both with regard to English (closure voicing and consonant duration [5], vowel duration [8], the consonant/vowel ratio [2, 7], consonant and vowel duration as independent cues [6]), and, for example, French [3] and German [4]. This paper extends the investigation to Russian, testing temporal differences which were found in production [1]. It was found that, as in English, voiced consonants were shorter than their voiceless counterparts and the vowel preceding the consonant longer. The present perception experiments were designed to investigate whether these differences could be perceptual cues to the contrast, and if so how they interact with a third cue: the duration of voicing during the closure.

2. METHOD

2.1 Stimuli

Two pairs of tokens were selected from recordings previously analysed [1]. The tokens were /rota/ ['rɔtə] (military company) and /roda/ ['rɔdə] (sort, gen.sg.). A pair was taken from two female speakers' recordings (referred to below as Set 1 and Set 2 respectively). In Set 1, the /roda/ token contained voicing at the /d/ burst, as the stop was fully voiced; this cycle at the burst was not cut out in the editing (see below); in Set 2 voicing at the burst was absent, voicing during the closure having ceased before the burst.

The vowel and stop durations (in msec) of the tokens are as follows:

	Set 1		Set 2	
	stop	vowel	stop	vowel
/roda/	65	168	81	165
/rota/	113	145	114	144
difference	+48	–23	+33	–21

note: stop duration refers here to the duration of the hold phase only.

The four original tokens (sampled at 20kHz) were digitally edited as follows:
a) /t/ in /rota/ was shortened: to 65 msec in Set 1 and to 81 msec in Set 2.
b) /d/ in /roda/ was lengthened: to 113 msec in Set 1 and 116 msec in Set 2.
c) /o/ in /rota/ was lengthened by repeating a section of the steady state vowel: to 169 msec in Set 1 and 165 msec in Set 2. The stop duration was kept as in the original token.
d) /o/ in /roda/ was shortened by cutting out a section of the steady state vowel: to 144 msec in Set 1 and 143 msec in Set 2. The stop duration was kept as in the original token.

In each of these eight stimuli, therefore, one cue has been introduced which conflicts with the other information present in the token. A comparison of responses to the edited stimulus and to the unedited token will show the effect of this alteration.

To test the interaction of stop or vowel duration and voicing during the closure, in each of the above 8 stimuli, together with the 4 original tokens, the closure voicing was edited out in stages, in the case of /roda/ stimuli, or added in in stages, in the case of /rota/ stimuli. 12 series of stimuli were thereby created, each with a gradually increasing duration of voicing during the closure.

The series resulting from an original /rota/ token are referred to below as /rota/ stimuli, those from a /roda/ token as /roda/ stimuli.

2.2 Procedure

There were 10 presentations of each of the 78 stimuli. The experiment was in two parts: Set 1 and Set 2 stimuli being presented separately. Within each part the presentations were randomised, and preceded by 10 other words recorded by the same speaker. The experiment was carried out in Moscow, with 37 subjects who were native Russian speakers aged 17-40 resident in Moscow (18 female, 19 male). The subjects heard the stimuli played on a cassette recorder in quiet surroundings, 27 with headphones, 10 without. The format was a forced choice identification task, and subjects wrote their responses on prepared answer sheets.

3. RESULTS AND DISCUSSION

3.1 Voicing during the closure

Figures 1 and 2 present the mean responses to the /rota/ stimuli Set 1 and 2 and the /roda/ stimuli Set 2. It can be seen that voicing during the stop is an important cue. Full voicing leads to approximately 100% /d/ responses. Absence of voicing leads to 100% /t/ responses in the case of /rota/ stimuli, and in the case of /roda/ stimuli to 47.7%, 60% and 80.6% /t/ responses, depending on stop and vowel duration. Full voicing is, not unexpectedly, a sufficient cue to override all others, including in the case of /rota/ stimuli all other cues to a /t/ present in the original token. The absence of voicing does increase /t/ responses but is not, at least in the case of this particular token, a sufficient cue to a /t/.

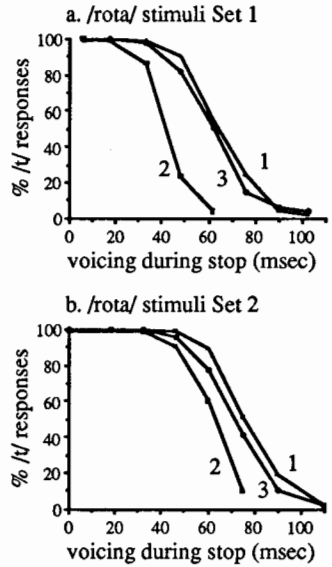

Figure 1: Responses to /rota/ stimuli with 1=original stop and vowel duration, 2=shortened stop, and 3=lengthened vowel.

22 of the 37 subjects gave a 100% /d/ response to the /roda/ Set 1 stimuli, even when the consonant or vowel duration had been altered and there was no voicing during the closure. Only two subjects gave a 50% or more /t/ response to any stimulus. The most likely characteristic leading to this almost total /d/ response is the presence of voicing at the burst.

For one subject only, there was an 80-100% /t/ response to all /rota/ stimuli, even with consonant or vowel duration changed. For this subject, at least in the case of these particular tokens, closure

voicing is not a sufficient cue to a /d/ and is being overridden by another cue or cues to a /t/ in the token. His results are excluded from the mean results for the /rota/ stimuli presented below.

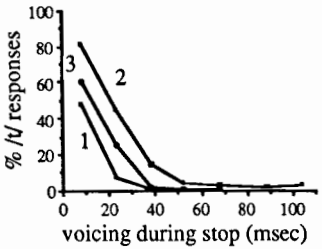

Figure 2: Responses to /roda/ stimuli Set 2 with 1=original stop and vowel duration, 2=lengthened stop and 3=shortened vowel.

3.2 Vowel duration

Figures 1a and b show a small decrease in /t/ responses to /rota/ stimuli with a lengthened vowel, where closure voicing is ambiguous. There is a small difference in the 50% crossovers between the two labelling functions: of 3.1 msec (Set 1) and 4.8 msec (Set 2). In both cases t-tests for paired samples show this difference to be significant (Set 1: $t = -3.40$, $p = 0.002$; Set 2: $t = -6.62$, $p < 0.001$).

Figure 2 shows that in /roda/ Set 2 stimuli the shortened vowel has an effect on responses where the closure voicing has been almost completely edited out. The combination of vowel duration and absence of voicing is a stronger cue than the absence of voicing alone, although not a sufficient cue to completely override others in the signal. In the case of the stimuli with 8 msec closure voicing, the difference in /t/ responses is 12.6%, which a paired samples t-test showed to be significant ($t = 4.89$, $p < 0.001$).

It is possible that the difference in responses is small because the amount by which the vowel duration had been altered was small. The alterations were this size, however, in order to be representative of the differences found in production.

3.3 Stop duration

In the case of /rota/ stimuli, the change to a /d/ percept occurs with less closure voicing when the stop is shortened than when it is not (figures 1a and b). The difference in the 50% crossover points was 23.2 msec (Set 1) and 14.1 msec (Set 2), which in paired samples t-tests were significant (Set 1 $t = -19.45$, $p < 0.001$; Set 2 $t = -12.62$, $p < 0.001$).

A possible interpretation of these results is that the shorter consonant duration is an additional cue to a /d/. However, as shortening the stop automatically increases the proportion of the stop which is voiced (for a given absolute duration of voicing), it is possible that the difference in response is due rather to the listeners using the voiced proportion of the stop as a cue, and not the absolute duration of voicing [3]. To test this, the responses were analysed as a function of the proportion of the closure which was voiced (figure 3).

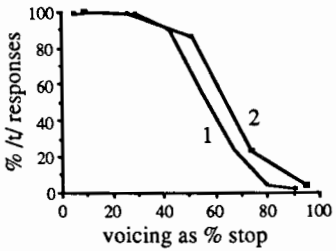

Figure 3: Responses to /rota/ Set 1 stimuli as a function of the percentage of the stop which is voiced (1=original stop duration 2=shortened stop).

The change to a /d/ response is now later (that is, when there is a greater voiced quotient) when the stop is shorter than when it is the original length. This was the case for Set 2 stimuli also. The difference between the 50% crossover points was 7.1% in Set 1 and 10.0% in Set 2, and was significant (Set 1 $t = 6.49$, $p < 0.001$; Set 2 $t = 9.54$, $p < 0.001$).

If the voiced quotient was the sole factor determining the response, the two labelling curves would be identical. This significant difference between the curves is unlikely to be due to the shorter consonant duration itself contributing, as the difference is in the opposite direction to that expected from production findings: here, a shorter stop appears to require a

greater voiced quotient to be perceived as a /d/. If, however, it is the absolute duration of the voicing, and not the voiced quotient, which is relevant in the perception of the contrast, the difference in figure 3 would be explained: a stimulus with the shorter stop will have a greater voiced quotient than a stimulus with equal absolute duration of voicing but a longer stop.

If it is the absolute duration of voicing which is perceptually relevant, figure 3 would not contradict the hypothesis that a decreased stop duration is contributing to a /d/ response: there would be no direct evidence as to whether stop duration is relevant or not.

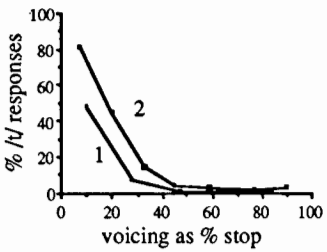

Figure 4: Responses to /roda/ Set 2 stimuli as a function of the percentage of the stop which is voiced (1=original stop duration, 2=lengthened stop).

If it is correct to conclude that the absolute duration of voicing is a cue, and not the voiced quotient, an explanation for the difference in the curves in figures 1a and b (voicing in msec) is that the difference shows the contribution of the decreased stop duration to the perception of a /d/.

This conclusion is supported by responses to the /roda/ stimuli (figures 2 and 4). Whether closure voicing is analysed in terms of proportion of closure duration (figure 4) or absolute duration (figure 2), a longer stop leads to more /t/ responses when there is little voicing during the closure. The difference between the two labelling functions in /t/ responses to stimuli with 8 msec voicing (32.6%) was significant in a paired samples t-test ($t = 6.99$, $p < 0.001$). The difference between the labelling functions when 10% of the closure is voiced is also significant (difference 24.4%, $t = 5.86$, $p < 0.001$). Stop duration is therefore contributing to cue a /t/, however voicing is analysed.

4. CONCLUSION

The duration of closure voicing is for most of the above subjects an important cue to the voicing contrast in Russian, although not in all cases does it override other cues present in the signal. The analysis supports a model in which the absolute duration of voicing is relevant and not the duration of voicing in proportion to that of the stop.

Stop and preceding vowel duration are additional cues to the voicing contrast in Russian. They have been analysed in this paper as independent factors; further analysis will investigate whether or not the results support a quotient model in which consonant to vowel ratio is the relevant perceptual cue.

5. REFERENCES

[1] BARRY, S.M.E. (1988), "Temporal aspects of the devoicing of word final obstruents in Russian", *Speech '88: Proc. 7th FASE Symposium*, 81-88.

[2] DENES, P. (1955), "Effect of Duration on the Perception of Voicing", *J. Acoust. Soc. Amer.*, *27*, 761-764.

[3] VAN DOMMELEN, W. (1983), "Parameter Interaction in the Perception of French Plosives", *Phonetica*, *40*, 32-62.

[4] KOHLER, K.J. (1979), "Dimensions in the Perception of Fortis and Lenis Plosives", *Phonetica*, *36*, 332-343.

[5] LISKER, L. (1978), "Rapid vs. Rabid: A Catalogue of Acoustic Features That May Cue the Distinction", *Haskins Lab. SR-54*, 127-132.

[6] MASSARO, D.W. & COHEN, M.M. (1983), "Consonant/vowel ratio: An improbable cue in speech", *Perception & Psychophysics*, *33*, 501-505.

[7] PORT, R.F. & DALBY, J. (1982), "Consonant/vowel ratio as a cue for voicing in English", *Perception & Psychophysics*, *32*, 141-152.

[8] RAPHAEL, L.J. (1972), "Preceding Vowel Duration as a Cue to the Perception of the Voicing Characteristic of Word-Final Consonants in American English", *J. Acoust. Soc. Amer.*, *51*, 1296-1303.