ON THE PERCEPTION OF VOICING IN WORD-FINAL STOPS IN GERMAN

Peter M. Janker and Hans Georg Piroth

ZAS - Research Centre for General Linguistics, Berlin, Germany

ABSTRACT

The perceptual capability of recovering underlying voicing in devoiced stops was investigated. In an identification experiment the standard minimal word pair for final devoicing in German - <Rat> (advice) vs. <Rad> (wheel) - was used. Items cut from utterances of the simplex words and from compounds were presented in isolation. A correlation analysis revealed that the number of “Rad”-answers to stimuli from compounds significantly increased - contrary to expectation - for shorter vowels and longer occlusions. Additionally, listeners also relied on release durations in their judgment when simplex items were presented.

1. INTRODUCTION

In phonological theories it is generally assumed that a devoicing rule operates on syllable-final German obstruents [15, 17, 16, 6]. But, there is an ongoing discussion on whether this process results in a complete or in an only partial phonetic neutralization. If this devoicing rule has as an output a phonetic representation that includes at least some not or not completely neutralized feature values, then listeners should be able to detect cues for distinguishing between voiceless and devoiced obstruents, even if there is no contextual information from semantic or pragmatic levels.

Besides a variety of acoustic analyses in this field on detectable residuals of underlying voicing in the signal [10, 5, 11, 12, 7, 2], a few investigators have also addressed the question of whether or not measured differences could provide cues sufficient to preserve the voicing contrast also in perception. Dinnsen [3] as well as Charles-Luce [1] already remark that possible acoustic residuals need not necessarily contribute to the perceptual decision. Dinnsen and Charles-Luce [4] reported gradual differences in the extent to which Catalan speakers produced final devoicing which - as mentioned by Słowiaczek and Dinnsen [13] - could also be differentiated by Catalan hearers. After Słowiaczek and Dinnsen had also found individual and systematic variations in the productions of Polish speakers [13], Słowiaczek and Szymanska [14] executed a perception experiment with Polish and English native listeners exposed to the material obtained in the former production study. Either group was able to identify both voicing categories to about 60% correct. This is compatible with the results found earlier by Port and O’Dell [11] for German with native listeners in a similar investigation. Port and Crawford [12] yielded a higher identification rate (69% on the average) for German with word stimuli partly taken from utterances contrasting the voiceless and devoiced words pairwise in one sentence.

The present investigation was designed to replicate and extend the findings on the perceptual status of German final obstruent devoicing by using speech samples that (as we suggest) allow a more reliable base for the interpretation of the perceptual processing of the variations occurring in “devoicable” final consonants.

2. MATERIAL

The speech data were taken from the material recorded for the production study of Piroth et al. [8], that was continued afterwards to account for the variety of dialectal impacts on the realization of Standard German [9]. This corpus contains about 5000 Standard German utterances of words with all possible final obstruents (phonologically voiced and voiceless) as well as nasals and liquids (always phonologically voiced) as control items not affectable by final devoicing. These were embedded in sentence frames repeatedly produced by six speakers of different dialectal regions of Germany. To achieve different contextual positions two variants of sentence frames and four of word types were chosen to occur in the corpus (Tab. 1).

| Sentence 1: | “Ich sage _____ nochmal.” (I say _____ again.) |
| Sentence 2: | “Ich sage _____ .” (I say _____.) |
| Word type 1: | [+cons] intervocalic |
| Word type 2: | [+cons] morpheme-final, voiceless continuation |
| Word type 3: | [+cons] morpheme-final, voiced continuation |
| Word type 4: | [+cons] word-final |

Table 1. Sentence frames and word types with examples.

This leads to five contextual positions within which the target words were produced: 1 to 3 (word types 1 to 3 embedded in sentence frame 1), 4 and 5 (word type 4 in frames 1 and 2).

For the special purpose of the present experiment, all instances containing the standard - minimal pair - example of <Rat> vs. <Rad> and the corresponding compounds were selected from the corpus. In contrast to other investigations, speakers remained unaware of the systematic relations between obstruents - voiceless vs. (de)voiced - in the material because of the large number of filler sentences - with final nasals and liquids - included. In addition, this procedure also generates a large number of <Rat> and <Rad> utterances in different pronunciations, distributed randomly according to the individual speech behavior characteristics of the speakers, but systematically according to the contextual structure of the corpus.
From the sentence utterances (see Tab. 1 above) the acoustic segments corresponding to <Rad> or <Rat> were cut in the waveform display of the signal under auditory and visual control from the beginning of the r to the end of the obstruct, except for position 1 items, as they are not within the realm of the devoicing rule.

In this way, 72 utterances each of <Rat> and <Rad> became available for presentation.

3. PROCEDURE
The signal files of the utterances were grouped by wordform - compounds (pos23), simplex words (pos45) - and speaker. This was done to account for the different overall durations of compounds and simplex words, the latter to allow the listeners to adjust for the speaker’s pronunciation characteristics, as they would do in a normal hearing situation.

After digital to analogue conversion (Data Translation DT2821) the monophonic signals were low-pass filtered (Behringer PEQ 305, cut-off 7 kHz, 12 dB/dec), arranged for an identification task and presented binaurally via closed headphones (Sony MDR CD 550 with Sony F535R amplifier) six times in random order under computer control (CSRE 4.0). Randomisation was changed for every listener and repetition.

Ten native listeners were instructed to decide whether the word presented meant <Rat> (advice) as in <Amtsrat> (councillor) or <Rad> (wheel) as in <Fahrtrad> (bicycle) by pressing a marked response key on the computer keypad. Since no context information was provided, listeners could only rely on acoustic signal features in the stimuli to make their decision.

Thus 4320 responses were obtained for either test (pos23 and pos45).

4. RESULTS
For simplex words (pos45) listeners responded ‘Rat’ to 59.0% of the stimuli, for compounds their response ratio was inverted yielding 47.6% ‘Rad’ responses. The rating was correct for 54.5% of the stimuli in pos45 (significantly different from chance with Chi-square = 35.2, p<.001) and for 59.9% in pos23 (Chi-square = 169.6, p<.001). For pos45 <Rat> was identified clearer (31.7%), as opposed to <Rad> (22.8%) showing a highly significant bias in favor of <Rat> responses (Chi-square = 63.5, p<.001). In contrary, for pos23 correctness to <Rad> stimuli was slightly higher (31.1%) than to <Rat> stimuli (28.8%). This difference was significant on the 5% level only (Chi-square = 4.0, p = .045). Figure 1 gives the counts for all responses.

Response frequencies were additionally submitted to a correlation analysis to test the hypothesis that listeners group their identification decisions according to the ranges of acoustic feature values of the stimuli (as provided by the database of the acoustic analysis produced for Piroth and Janker [9]). The results are given in Tab. 2.

5. DISCUSSION
Assuming that final devoicing is complete, no difference between devoiced and voiceless items should be found. The investigations reported above found differences in the production of devoiced and voiceless items. Some of them tested whether or not these systematically sampled production differences were discriminable and perceptually meaningful. For the present experiment items cut from devoiced and voiceless utterances were correctly assigned to voiced and voiceless categories significantly above chance level. This suggests that the voiceless-voiceless distinction is at least partly maintained. The percentage of correct answers is in approximately the same range as found in the literature [11, 12, 14] and for pos45 occurs also the reported bias for voiceless items. Contrary to the findings in the literature, pos23 showed a tendency for a bias to ‘voiced’ responses. For some stimuli a correlation of response behavior to nucleus duration (positive), closure duration (negative), duration of voicing into closure (positive) and of aspiration (negative) was reported [11]. In the present experiment, no influence of voicing into closure on the listener could be encountered. The release correlates with responses only in pos45, due to the fact that for items cut from compounds (pos23) the release is truncated by the following consonant and therefore in general very short, thus not providing a durational cue for voicelessness. Also contrary to other researchers’ results, there is a negative correlation for
nucleus duration and a positive one for closure duration. These effects are similar for pos23 and pos45 and therefore not influenced by the unavailability of the release as a marker in pos23. Nevertheless, vowel and release durations do not correlate for either pos45 or pos23.

The observed differences to the reported findings might be due to the different stimulus material used. The items of the cited experiment were selected from a contrastive elicitation sample, while the items used here were taken from a “random” sample. The speakers had reported that they remained unaware of word
contrasts hidden in the material, therefore we assume that the
distribution of acoustic feature values approached those of every
day speaking behavior. Since all utterances of the word pair
<Rad>/<Rat> were included into the test the listeners were
exposed to the whole variety of the material.

We therefore suggest that listeners confronted with such a
variety of utterances employ assessment strategies during the
categorization process which are different from those of
participants in earlier experiments. Short vowel durations and
short release durations seem to have been strong cues for <Rad>
responses, as can be seen from Figure 4: Only items with vowel
durations below approximately 220ms and release durations
about 110ms are associated to high <Rad> response values.

6. CONCLUSION

Interpreted in terms of preference rules for decision strategies
applied by listeners in ambiguous cases one can summarize that
(1) shorter vowel durations and longer occlusions function as
indicators for ‘voiced’ and lead to an enhancement of ‘Rad’
responses. (2) Longer releases are mostly due to the aspiration
noise following bursts of voiceless final stops. Since the releases
of the items taken from compounds are truncated by the
following consonants and therefore in general very short, those
stimuli fail to provide this durational clue for the distinction.

ACKNOWLEDGMENT

We wish to thank Lieselotte Schiefer from the Institute of Phonetics at
Munich University who headed the team that initiated our research on
final devoicing during a course given by her and one of the authors.

REFERENCES

[10] Port, R., Mitleb, F. M. and O’Dell, M 1981. Neutralization of
syllable-final voicing in German is incomplete. Journal of the Acoustical Society of America, 70, S10.
voicing in German. Journal of Phonetics, 13, 455-471.

Figure 4. <Rad> responses for compounds, split by underlying
<Rad> or <Rat>, in relation to the duration of the vowel, the
occlusion and the release.