PERCEPTUAL EFFECTS OF ASSIMILATION-INDUCED VIOLATION OF FINAL DEVOICING IN DUTCH

Cecile Kuijpers[1], Wilma van Donselaar[1], Anne Cutler

Max Planck Institute for Psycholinguistics
Nijmegen, The Netherlands
C.Kuijpers@ped.kun.nl

ABSTRACT

Voice assimilation in Dutch is an optional phonological rule which changes the surface forms of words and in doing so may violate the otherwise obligatory phonological rule of syllable-final devoicing. We report two experiments examining the influence of voice assimilation on phoneme processing, in lexical compound words and in noun-verb phrases. Processing was not impaired in appropriate assimilation contexts across morpheme boundaries, but was impaired when devoicing was violated (a) in an inappropriate (non-assimilatory) context, or (b) across a syntactic boundary.

1.   INTRODUCTION

In fluent speech numerous phonological adjustment processes occur, within as well as across words. Some of these processes are obligatory whereas others are optional. If a phonological rule is obligatory, there is one acceptable 'standard' realization. Presumably, there is also one standard lexical representation. If a phonological process is optional, both the standard realization and the phonological variant are equally acceptable.

An example of an obligatory rule in Dutch phonology is the devoicing of syllable-final obstruents. Speakers are not free to pronounce Dutch words like 'bed' (*bed*) or 'kwab' (*lobe*) with a voiced final consonant. An example of an optional rule in Dutch is schwa epenthesis in words with final clusters, e.g. of liquids and non-coronal obstruents [1]. Words like 'melk' (/mɛlk/ -*milk*-) and 'dorp' (/dɔrp/ -*village-)* can be realized as /melək/ and /dɔrəp/, respectively. A speaker is free to use the standard form or the phonological variant, and listeners have to take into account that a word can be realized in different ways.

The goal of the present study was to examine the effect on speech processing of phonological variation due to voice assimilation. Phonological variation leads to different surface forms. The question is how these different surface forms can be mapped onto lexical representations. One possibility is that there are various lexical representations for words that have phonological variants. Another possibility is that phonological rules are applied to the incoming phonetic string.

1.1. Phonological variation

A recent series of perception studies has investigated optional schwa epenthesis in Dutch [2, 3]. Results indicated that phonological variants with schwa epenthesis (e.g., /melək/) do not have separate lexical representations; instead, phonological rules must be applied to derive the variant forms. Furthermore, schwa epenthesis actually facilitates processing: words like 'melk' are easier to recognise with than without epenthesis. This occurs because addition of the epenthetic vowel makes the consonant cluster more perceptible than it is in the standard forms with consonant clusters. Epenthesis of schwa after /l/, for instance, results in a "clear" /l/ as opposed to the "dark" /l/ in a syllable-final cluster [4]. Enhanced perceptability of consonants before vowels than before other consonants would also explain why phonological variants of words with optional schwa deletion (e.g., /refra:t/) are processed less rapidly than their standard realizations retaining the schwa (/refəra:t/) [3].

Assimilation means that phonemes are altered due to the influence of adjacent phonemes. A number of studies have investigated assimilation in Dutch, Japanese, and English [5,6,7,8]. Place assimilation in Dutch is optional, affects the coronal nasal /n/, and is regressive (the effect goes from right-to-left). In a phoneme monitoring experiment listeners had to detect the phoneme /n/ in 'groeN Boek' vs. 'groeM Boek'. It was found that assimilation had an inhibitory effect on phoneme recognition. The detection of the word-initial phoneme (e.g., /b/) was not influenced by the preceding phoneme [5].

In Japanese, place assimilation is obligatory within words. In the word 'toNbo' the moraic nasal always takes the labial feature of the subsequent /b/ (toMbo), and in 'roNgo' the velar feature of /g/ (roNGgo). In a cross-linguistic study, Otake *et al*. [6] found that Japanese listeners detected the nasal N, which had variable realizations, equally fast in different phonetic contexts. Dutch listeners, on the other hand, were sensitive to the phonetic contexts. With the postnasal consonant, Japanese listeners' responses were affected by contextual match (shorter latencies to /b/ in toMbo than in toNbo) but Dutch listeners' responses were not. The cross-linguistic results indicate that Japanese listeners exploit their native phonology in processing speech segments, and that the use of anticipatory cues to place assimilation is not just a prelexical, low-level acoustic effect.

Cross-modal priming and phoneme monitoring studies have also revealed inhibitory effects of phonological context inconsistent with assimilated forms [7]. The material used in these experiments consisted of noun-verb combinations. Listeners identified the phoneme /p/ equally fast in the standard form 'kiD Pulling' and the assimilated form 'kiB Pulling', but phoneme detection was slowed in inappropriate contexts (e.g., /g/ in kiB Grabbing). Thus it appears that the phonological context is used to assess the validity of the surface change. No evidence was found of predictive use of potentially assimilated segments. The authors of this study concluded that assimilated forms are treated by the perceptual system in essentially the same way as standard unassimilated forms.

More recently, Gow [8] has shown that listeners do use assimilation to anticipate upcoming context. In a phoneme detection experiment stops (e.g. /b/) were recognised more rapidly after word-final segments with contextually appropriate assimilation (e.g., 'teM buns') than after unmodified segments ('teN Buns'). As in previous work, stops after inappropriate alterations ('teNG Buns') elicited longer monitoring latencies. According to Gow [8], assimilatory modification is used to anticipate following context and is a perceptually enriching process. Long latencies in an inappropriate context reflect violation of anticipation and disruption of speech processing.

In all the studies discussed above, place assimilation has been examined, across words in sentences. In the languages under study, with the exception of Japanese, place assimilation is an optional rule. Another optional rule is regressive voice assimilation in Dutch (RVA). In RVA, a voiceless obstruent becomes voiced before a following voiced stop /b/ or /d/ [1]. The word-medial consonant cluster can be realized in its standard form ('kaasboer' *cheese monger* and 'asbest' *asbestos*) or in the assimilated variant with voiced medial clusters ('kaazboer' and 'azbest'). Although RVA itself is optional, it is closely linked to the obligatory Dutch rule of Final Devoicing. In Dutch, voiced segments in syllable-final position only occur as a result of voice assimilation. In the present study, we ask whether listeners use the presence of syllable-final voicing to anticipate upcoming voiced phonemes in different contexts.

2. EXPERIMENT 1

The first experiment asks whether assimilatory information facilitates recognition of a following phoneme. Our prediction is that listeners will be faster in detecting /b/ in 'kaazboer' than in 'kaasboer'. Because of the final devoicing rule, the presence of a syllable-final voiced segment must point to an upcoming voiced consonant. It is predicted that a voiceless consonant is detected less rapidly after a syllable-final voiced segment (e.g. /z/ in 'kaazplank' vs. /s/ in 'kaasplank' *cheese plank*).

2.1. Method

2.1.1. Material

Forty word pairs were selected, a total of eighty words. These were all frequent compound words in Dutch, each consisting of a concatenation of two nouns. The first noun was identical in both words of a pair; the second noun had either a voiced or a voiceless initial consonant. The syllable-final consonants are referred to as C1, and the syllable-initial consonants as C2. The listener had to monitor for the target phoneme C2. In twenty word pairs C1 was voiceless according to the orthography, in the other twenty word pairs C1 was orthographically voiced (e.g., /d/ in 'tijDBom', 'tijDPerk'). C1 and C2 were never identical and never occurred more than once in a compound word. Another forty cross-spliced items were used (see below).

A trained male speaker read all material. The words with orthographically voiced C2 were always realized with voice assimilation (e.g., 'ZB') and the words with orthographically voiceless C2 with a voiceless cluster (e.g., 'SB'). The words were digitized at 16 kHz. For each word pair, two additional items were created by cross-splicing ('kaaSBoer', 'kaaZPlank').

The cross splicing was performed using both visual and auditory information from the speech waveform. For each compound the first syllable was separated from the second one and combined with the second syllable of the paired item. Cross-spliced utterances sounded as natural as unspliced utterances. In two conditions the consonants were matched for the feature 'voice' (e.g., 'kaaZBoer', 'kaaSPlank'), in the other two conditions the consonants were not matched ('kaaSBoer' with a legal sequence /sb/, 'kaaZPlank' with an illegal sequence /zp/). There were 160 items, 40 items in each condition.

2.1.2. Fillers

A total of 260 filler words were added to the material. The filler material consisted of both compound words (e.g., 'theekop', *tea cup*) and monomorphemic words (e.g., 'krediet', *credit*). In 210 items the target phoneme was /p/, /t/, /s/, /f/, or /d/. It occurred either in word-initial, word-medial, or word-final position. Fifty items did not contain the target phoneme. Each target phoneme occurred an equal number of times in initial, medial, and final position. Twenty representative practice trials were presented.

2.1.3. Participants

Forty-eight native speakers of Dutch, students at the University of Nijmegen, took part in the experiment for a small payment.

2.1.4. Procedure

Four different materials sets were used. The two original and two cross-spliced trials for each item pair were counterbalanced across sets according to a Latin square design. Each set was made up of 300 items (test and filler). All conditions of one item appeared in the same position. Subjects were allocated in random order to one of the four sets. They were instructed to press a button as soon as they detected the target sound in the word (go-no go response). Because some phonemes might be easily confused, for instance /f/-/s/ and /p/-/t/, the target specifications were presented both visually (a letter) and auditorily before each trial. The response window was 2500 ms and reaction times were measured from the onset of the closure of the target (C2). The experiment took 35 minutess and.was carried out in three blocks. Subjects heard the stimuli over closed headphones. Stimulus presentation and response collection were controlled by the NESU software.

2.2. Results and discussion

The two types of items ('kaasboer' and 'tijdbom'). exhibited a similar reaction time pattern and were pooled for further analyses. Because of experimental errors only18 items remained in each set of words. In each condition, the hit rates exceeded 95%, even in the condition with the illegal sequence of consonants (e.g., 'kaaZPlank'). The reaction time averages are given in Figure 1.

There was a significant effect of C1 on the detection of the following consonant (C2). Voiceless C1 led to faster detection of C2 than voiced C1 ([F1(1,47)=5.97, p<.02; F2(1,34)=3.35, p=.07]). There were virtually no reaction time differences between items with voiced C2. However, items with voiceless C2 exhibited large differences (see Figure 1). The interaction between C1 and C2 was highly significant ([F_1(1,47)=16.99, p<.001; F_2(1,34)=11.31, p<.01]). On average, subjects detected

the voiced targets faster than the voiceless targets (462 ms vs. 514 ms). This difference (as in 'kaaSBoer' vs. 'kaaSPlank') was significant in both subjects and items analyses [$F_1(1,47)=35.64$, $p<0.001$; $F_2(1,17)=20.89$, $p<0.001$]. Effects were mainly due to the illegal voiced-voiceless condition. Therefore, the conditions were subjected to a post-hoc Newman-Keuls analysis ($p=0.05$). The illegal condition ('kaaZPlank') differed significantly from all other conditions, but no other differences between conditions were significant.
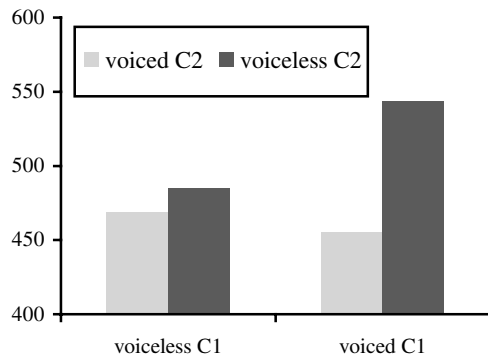

Figure 1: Reaction times in ms for target C2. Data refer to 'kaasboer', 'kaasplank' and 'kaazboer', 'kaazplank'.

No evidence was found that listeners use assimilatory modification to anticipate the following context in compound words. This is consistent with previous findings [5,6]. The prediction that the detection of targets would be delayed in phonologically illegal clusters was confirmed. Thus lexical processing is considerably delayed if the obligatory final devoicing rule is violated by an inappropriate (non-assimilatory) context.

3. EXPERIMENT 2

Although no anticipatory effect was found in Experiment 1, it might be that listeners do use voice assimilation to anticipate lexical processing in a different context. Perhaps assimilation is not used to disambiguate within compound words but to disambiguate compound nouns from noun-verb phrases. Several production studies (e.g. [9]) showed that RVA occurs more frequently within compound words than between words in noun-verb phrases. In Experiment 2 we test whether monitoring for phonemes in contextually appropriate assimilation is slower in noun-verb phrases than in compound words. This would suggest that word-final devoicing can be overruled by voice assimilation being more plausible within compound words than within noun-verb phrases.

3.1. Method

3.1.1. Material

Twenty-eight compound words were selected. The second part of the compound word started with a voiced stop (e.g. /b/ in 'kaaSBoer'). In sixteen items the word-final consonant C1 was orthographically voiceless (e.g., kaaSboer), in twelve items the word-final obstruent was orthographically voiced (e.g., klaDblok). Each compound word had a noun-verb phrase as its counterpart. The phonemic sequence was identical up to and including the C1C2 cluster. For instance, the compound word 'kaasboer' (*cheese monger*) was paired with the noun-verb phrase 'kaas bakt' (*cheese bakes*). It was not possible to have identical second vowels, because this considerably reduced the number of acceptable noun-verb combinations.

Short sentences were constructed (6-7 words preceding the target). These were acceptable in both conditions and realized with the same intonation contour. The sentences were cut off after the second word (e.g. 'Peter zegt dat hij de kaasboer/kaas bakt'…*Peter says that he the cheese monger /cheese bakes)*. Subjects had to monitor for the second obstruent (C2) in the cluster, either /d/ (7) or /b/ (21).

The same male speaker read all material. The compound words were realized with a voiced C1 ('kaaZboer') and the words in the noun-verb phrase were realized with a voiceless C1 ('kaaS bakt'). The speaker was instructed to finish the sentences with a dummy ending in order to use a natural intonation. There was no pause between the noun and the verb. The sentences were digitized at 16 kHz. For each pair two additional test items were created by cross splicing, resulting in items such as 'kaaSboer' and 'kaaZ bakt'. Cross splicing was done in the same way as in Experiment 1.

3.1.2. Fillers

A total of 92 filler items were added to the stimulus items, also embedded in truncated sentences. Sixty filler items contained a target phoneme which was either /d/ (23), /b/ (9), /s/ (30) or /k/ (30). The filler items were compounds (23), adverbs/composite verbs (23), adjectives/simple verbs (23), or simple nouns (23). Each target phoneme occurred an equal number of times in initial, medial, and final position. Thirty-two items did not contain a target. The target phoneme occurred either at the beginning of the first syllable or at the beginning of the second syllable. Twelve sentences had the same length as the test sentences and the item always occurred at the end of the sentence. Eighty sentences were longer than the test sentences (max 10 words). In half of the sentences the filler item occurred at the beginning of the sentence, in the other half at the end. The filler items were always followed by one or more words (conjunctions, prepositions, or adverbs). Twelve practice trials were presented at the start of the experiment.

3.1.3. Participants

Forty-eight subjects from the same population were tested. None had participated in the previous experiment.

3.1.4. Procedure

Four different sets were constructed, each containing 120 carrier sentences. The original and cross-spliced items were counterbalanced across sets according to a Latin square design. Subjects were allocated in random order to one of the four sets. As before, the target was presented both visually and auditorily and the sentences only auditorily. There were twelve randomly embedded filler sentences that subjects had to finish. In these cases the sentence "Maak nu de zin af" (*Please finish the sentence*) appeared on the screen after they had given their response. The completion task was intended to force the subjects to pay attention to the semantic content of the

sentence. The response window was 2500 ms after the onset of the carrier word. Reaction times were measured from the onset of the (voiced) target phoneme. Twelve practice items preceded the main experiment; testing took approximately 20 minutes.

3.2. Results and discussion

Two items were discarded because of experimental error. In each condition the hit rates exceeded 80% and did not show any significant effects. In the compound context, phoneme detection (e.g., /b/) was equally fast in the phonological variant ('kaaZBoer') as in the standard form ('kaaSBoer'). This is a replication of the results of Experiment 1. In the noun-verb context, the effect of C1 was significant in the subjects and items analysis $[F_1(1,47)=10.79; p<0.01; F_2(1,25)=3.96; p=0.05]$. The voiced target C2 was detected significantly faster in a non-assimilatory context. Thus, a homogeneously voiced cluster is more likely to be associated with a compound word than with a noun-verb phrase. Reaction times showed a main effect of context $[F_1(1,47)=98.50, p<0.001; F_2(1,25)=12.73, p<0.01]$. Phoneme detection was faster in compound words (mean 495 ms) than in noun-verb phrases (mean 584 ms). Further acoustic measurements showed that the reaction time difference was not due to durational differences (mean difference 14 ms) between the target-bearing words (e.g., 'boer' vs. 'bakt').
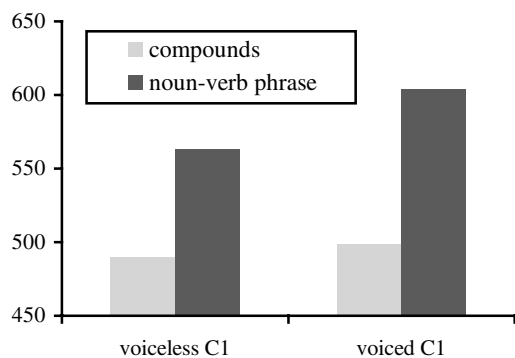

Figure 2: Reaction times in ms for target C2. Data refer to 'kaasboer', 'kaas bakt' and 'kaazboer', 'kaaz bakt'.

The results show that listeners are sensitive to the presence or absence of voice assimilation in hypothesizing the type of linguistic boundary. It seems that assimilation tends to render a sequence of monomorphemic words less likely than a bimorphemic compound word.

4.   CONCLUSIONS

In the first experiment we gained evidence that voice assimilation does not lead to anticipatory processing in lexical compound words, and that voiced final segments are not assumed to occur outside an assimilatory context. In bimorphemic lexical compound words it was found that C2 was processed equally fast in 'kaazboer' as in 'kaasboer' or 'kaasplank', but significantly slower in 'kaazplank'. This confirms that processing is hindered by violations of obligatory constraints (final voicing without assimilation), but is neither facilitated nor hindered by the application of a rule which produces a phonological variant (see also [5,6,7]).

The second experiment, however, shows that the phonological variant actually hinders processing in a specific context. Voice assimilation occurring across two monomorphemic lexical words, and thus separated by a word boundary ('kaaz bakt' – 'kaas bakt'), is less likely to overrule final devoicing relative to bimorphemic compound words. These results do not correspond to findings from previous studies [5,7,8,10] of assimilation across words. All these studies, however, investigated assimilation of place, which is an optional rule not challenging any obligatory phonological rule such as final devoicing. Listeners are thus able to extract information from the violation of the otherwise obligatory rule, and the information consists not of positive prediction but of exclusion: if syllable-final devoicing is violated, then (a) a voiceless segment should not follow, and (b) a minor (or major) syntactic boundary is unlikely to follow. Note that our results are fully interpretable within competition models of spoken word recognition, in which effects of number of competitors are known to appear [11]. In compound words the probability of a unique lexical entry at C2 is higher in lexical compound words than in noun-verb combinations, leading to less competition, and, in turn, faster monitoring latencies.

5.   REFERENCES

[1]   Booij, G., The Phonology of Dutch. Clarendon Press, Oxford , 1995.

[2]   Van Donselaar, W., Kuijpers, C. and Cutler, A., "Facilitatory effects of vowel epenthesis on word processing in Dutch", *J. of Mem. and Lang.,* 41: 59-77, 1999.

[3]   Kuijpers, C., van Donselaar, W. and Cutler, A., "Phonological variation: Epenthesis and deletion of schwa in Dutch", *Proc. of the Fourth ICSLP*, Vol. 1: 149-152, Philadelphia, 1996.

[4]   Warner, N.L., Jongman, A., Cutler, A. and Mücke, D., "The phonological status of Dutch epenthetic schwa", *Phonology*, 18: 387-420, 2001.

[5]   Koster, C.J., Word Recognition in Foreign and Native Language, Foris Publications, Dordrecht, 1987.

[6]   Otake, T., Yoneyama, K., Cutler, A. and van der Lugt, A., "The representation of Japanese moraic nasals", *J. Acoust. Soc. of Am.*, 100(6): 3831-3842, 1996.

[7]   Gaskell, M.G. and Marslen-Wilson, W.D., "Mechanisms of phonological inference in speech perception", J. Exp. Psych.: Human Percept. and Perf., 24(2): 380-396, 1998.

[8]   Gow, D.W., "Assimilation and anticipation in continuous spoken word recognition", *J. Mem. and Lang.,* 45: 133-159, 2001.

[9]   Slis, I.H., "Assimilation of voice in Dutch as a function of stress, word boundaries, and sex of the speaker and listener", *J. of Phon.*, 14: 311-326, 1986.

[10]   Gow, D.W., "Does English coronal place assimilation create lexical ambiguity?", *J. Exp. Psych.: Human Percept. and Perf.*, 28: 163-179, 2002.

[11]   Norris, D., McQueen, J.M. and Cutler, A., "Competition and segmentation in spoken word recognition", *J. Exp. Psych.: Learning, Mem. and Cogn.,* 21: 1209-1228, 1995.

[1]   CK is now at the Department of Special Education, University of Nijmegen; WvD is now at the Netherlands Organisation for Scientific Research (NWO), The Hague.