

# The functionality of incomplete neutralization in Dutch: The case of past-tense formation

1

Mirjam Ernestus and Harald Baayen  
University of Nijmegen and Max Planck Institute for Psycholinguistics  
Wundtlaan 1, 6525 XD Nijmegen, The Netherlands.  
mirjam.ernestus@mpi.nl, harald.baayen@mpi.nl

## Abstract

In many languages, underlyingly voiced obstruents are realized as voiceless in word-final position (Final Devoicing). Previous research has shown that this neutralization of the underlying [voice]-specification may be phonetically incomplete, with underlyingly voiced obstruents being realized as slightly voiced. Listeners are sensitive to incomplete neutralization, since they can tell apart above chance level the phonetic realizations of words underlyingly ending in voiced and voiceless obstruents. This study discusses to what extent incomplete neutralization is functional. It reports a series of experiments showing that incomplete neutralization can be induced in Dutch just by the way words are spelled, and that no minimal word pairs are required for incomplete neutralization to emerge. In addition, these experiments also show that Dutch listeners take advantage of incomplete neutralization, even when there are no task-specific reasons for them to do so. Incomplete neutralization appears to be a subphonemic cue to past-tense formation, and has a more substantial role in language processing than has been assumed thus far. Finally, our data show that Dutch listeners and speakers dynamically adapt their production and interpretation of the acoustic signal to the voicing properties of the orthographic and acoustic forms encountered previously in the experiment<sup>1</sup>.

## 1. Introduction

In many languages, including Catalan, Dutch, German, Polish, and Russian, underlyingly voiced obstruents are realized as voiceless in word-final or syllable-final position (see, e.g., Kenstowicz 1994:75). This Final Devoicing results in the phonological neutralization of underlyingly voiced and voiceless obstruents. Example (1) illustrates Final Devoicing in Dutch. The verbs [vɛrveidən] *verwijden* and [vɛrveitən] *verwijten* differ phonologically only in the [voice]-specification of the stem-final alveolar plosive preceding the infinitive suffix [ən] *en*, see (1a). The [voice]-specifications of these plosives are distinctive, and the plosives must therefore be specified underlyingly as /d/ and /t/ respectively. When the stems occur word-finally, the plosives are syllable-final, and they are both realized as [t] (unless regressive voice assimilation applies). The underlying phonological difference is neutralized, as shown in (1b).

(1) a.	<i>verwijden</i>	[vɛrveidən]	‘widen-inf’
	<i>verwijten</i>	[vɛrveitən]	‘reproach-inf’
b.	<i>verwijd</i>	[vɛrveit]	‘widen’
	<i>verwijt</i>	[vɛrveit]	‘reproach’

Previous research has shown that Final Devoicing does not always result in complete neutralization at the phonetic level (e.g., Dinnsen and Charles-Luce 1984; Port and O’Dell 1985; Slowiaczek and Dinnsen 1985; Port and Crawford 1989; Warner et al. in press). Neutralized obstruents that are underlyingly voiced tend to have more acoustic characteristics of voiced obstruents than neutralized obstruents that are underlyingly voiceless. They tend to be shorter, to be realized with vocal fold vibration during a longer period, and to be preceded by longer vowels. In what follows, we will refer to voiceless obstruents that possess some acoustic characteristics of voiced obstruents as slightly voiced obstruents.

Although the acoustic differences between neutralized underlyingly voiced and voiceless obstruents are generally small, listeners are able to take advantage of these differences. They assign above chance level the correct spelling to the members of minimal word pairs that differ from each other only in the underlying [voice]-specification of their final obstruents, which is reflected in their spellings (e.g., Port and O’Dell 1985; Port and Crawford 1989; Warner et al. in press). The question arises as to which extent incomplete neutralization plays a role in normal speech processing. In the spelling tasks, listeners may be forced to base their choices on acoustic features they would not rely on in normal processing circumstances. The finding that listeners take advantage of incomplete neutralization may be an artificial task effect (Warner et al. in press).

This paper addresses the question of the extent to which the interpretation of incomplete neutralization is functional. We tested whether listeners take advantage of incomplete neutralization also when their task does not force them to do so.

The language under investigation is Dutch. Incomplete neutralization in this language has been reported by Warner et al. (in press) and by Ernestus and Baayen (in press). Warner et al. (in press) carried out a production experiment with 15 native speakers of Dutch, reading 20 minimal word pairs. They observed a significant difference in vowel duration with vowels preceding underlyingly voiced obstruents being on average 3.5 ms longer than vowels preceding underlyingly voiceless obstruents. In addition, they also found a difference in burst duration in the predicted direction (underlyingly voiceless plosives had longer bursts), both for the words with phonologically long and the words with phonologically short vowels (see their Table 2), but the difference reached significance only for the words with long vowels. In the study by Ernestus and Baayen (in press), underlyingly voiced and voiceless plosives differed in the duration of their release noise, which includes both the plosive’s burst and the following period of aspiration. This study was based on 29 words ending in underlyingly voiceless plosives and 30 words ending in underlyingly voiced plosives, realized by a single speaker. In Dutch, incomplete neutralization is apparently mainly evidenced by vowel duration and the duration of the burst or release noise<sup>2</sup>.

We carried out two perception experiments in which native speakers of Dutch listened

to the stems of pseudo verbs. The stems were all realized with voiceless final obstruents,<sup>3</sup> obeying Final Devoicing. We asked the participants to write down the past-tense forms of the verbs. In Dutch, past-tense forms are created by suffixing [tə] *te* or [də] *de* to the verb stem. The suffix [tə] is added when the stem underlyingly ends in a voiceless obstruent, while [də] is added elsewhere (e.g. Booij 1995). The participants in our past-tense formation experiments, therefore, had to interpret the final neutralized obstruents as underlyingly voiced or voiceless. We showed in Ernestus and Baayen (2003) that participants base their interpretation on the bias of the gangs of phonologically similar words. That is, participants tend to add *te* to a pseudo verb when most phonologically similar words underlyingly end in voiceless obstruents, and they add *de* when most phonologically similar words underlyingly end in voiced obstruents. In the present study, we investigate whether participants also base their interpretation of stem final obstruents on acoustic subphonemic cues to underlying voicing resulting from incomplete neutralization. If so, they should more often add *de* to a verb stem when its final obstruent is realized as slightly voiced than when it is realized as completely voiceless.

One way to obtain pseudo words with slightly voiced obstruents is to create these words out of words with completely voiceless obstruents by changing those acoustic characteristics that are different for underlyingly voiced and voiceless neutralized obstruents in Dutch. Another method is suggested by previous research which shows that orthography may induce or enforce incomplete neutralization (Fourakis and Iverson 1984; Port and Crawford 1989). We opted for this latter method. It may not only provide us with slightly voiced obstruents, but also with additional information about how slight voicing is realized in Dutch. We asked a native speaker of Dutch to read aloud pseudo verb stems that were spelled with voiceless final obstruents (e.g., *daup*) and with voiced final obstruents (*daub*).

## 2. Experiment I

*Materials.* We compiled two lists of pseudo verb stems, in which each stem is preceded by the pronoun *ik* [ik] ‘I’. In these lists, the verb stems function as first person singular present tense forms. The first list follows the design of the experiment reported in Ernestus and Baayen (2003), and contains monosyllabic pseudo verbs ending in bilabial or alveolar plosives, or alveolar, labiodental, or velar fricatives. All final obstruents, except the velar fricatives, are spelled as voiceless in this list, and we therefore refer to this list as List -Voice. We will call the second list List Slightly Voiced. This list is identical to List -Voice, except that the 14 final bilabial plosives following sonorant segments are spelled with the letter *p* in List -Voice and with *b* in List Slightly Voiced, the 28 final alveolar plosives following sonorant segments are spelled with *t* in List -Voice and with *d* in List Slightly Voiced, and, finally, the 16 final velar fricatives are spelled with *g* in List -Voice and with *ch* in List Slightly Voiced. The remaining 115 words are spelled identically in both lists, and will be referred to as the words with constant spelling. They are all spelled with final graphemes representing voiceless obstruents, since, according to the Dutch spelling conventions, they cannot end in graphemes representing voiced obstruents. List -Voice contained an additional nineteen words, which we will not consider here. Table 1 summarizes the materials.

Table 1: Summary of the materials used in Experiments I and II, with the following abbreviations: “Son” for “sonorant segment”, “P” for “bilabial plosive”, and “T” for “alveolar plosive”. Column “Word type” indicates whether the words end in a sonorant segment and a plosive, in a velar fricative, or are words with constant spelling. Columns “List -Voice” and “List Slightly Voiced” show the spelling of the final obstruents in these lists.

Words ending in	Word type	Number of words	List -Voice	List Slightly Voiced
Son+P	Son+Plosive	14	<i>p</i>	<i>b</i>
Son+T	Son+Plosive	28	<i>t</i>	<i>d</i>
Velar fricative	Velar fricative	16	<i>g</i>	<i>ch</i>
Obstruent+P	Constant spelling	3	<i>p</i>	<i>p</i>
Obstruent+T	Constant spelling	34	<i>t</i>	<i>t</i>
Alveolar fricative	Constant spelling	57	<i>s</i>	<i>s</i>
Labiodental fricative	Constant spelling	21	<i>f</i>	<i>f</i>

We expected that all final obstruents would be realized as voiceless, because of Final Devoicing. In addition, we also expected that *b*, *d*, and *g* would be realized with more characteristics of fully voiced obstruents than *p*, *t* and *ch*, since *b*, *d*, and *g* represent underlyingly voiced obstruents in Dutch, whereas *p*, *t* and *ch* represent underlyingly voiceless obstruents. That is, we expected that vowels preceding *b*, *d*, and *g* would be longer than the vowels before *p*, *t* and *ch*, and that *b* and *d* would be realized with shorter release noises than *p* and *t*. Finally, we expected that the obstruents that were spelled identically in the two lists might be more voiced in List Slightly Voiced than in List -Voice. If *b* and *d* would indeed be slightly voiced in List Slightly Voiced, this might lead to automation of the articulatory gestures for slight voicing. As a consequence, also the final obstruents of the words with constant spelling, which do not carry information in their orthography about underlying voicing, might be slightly voiced. Conversely, List -Voice may lead to automation of the articulatory gestures for voicelessness, and the final obstruents of the words with constant spellings might be realized as completely voiceless in this list.

*Procedure.* We asked a female native speaker of Dutch to read aloud the two lists. There was an interval of an hour between the recording of List -Voice and that of List Slightly Voiced. Our speaker declared afterwards that she had not noticed that the two lists were nearly identical.

We recorded her speech on a DAT (BASF master 94) by means of an Aiwa HD S100 recorder and a Sony microphone ECM MS957 in a sound attenuated room. We digitized the recorded speech at a sampling rate of 48 kHz.

*Results and discussion.* We first investigated whether the words spelled with voiced final graphemes had been realized with more voicing than the corresponding words spelled with

voiceless graphemes by measuring the durations of the vowels. In case a vowel was followed<sup>5</sup> by a sonorant consonant, we included the duration of this sonorant in our measurement, since it is in general impossible to determine exactly where the vowel ends in the acoustic signal and where the following sonorant consonant begins. We defined the beginning of the vowel as the beginning of the regular wave form with the characteristics of the vowel, and the end of the vowel (plus sonorant consonant) as the (sudden) end of the regular wave form. In addition, we also measured the durations of the final velar fricatives, the durations of the closures of the final plosives, and the release noises of these plosives, which include the bursts of the plosives and the following period of aspiration (see Figure 1 for two wave forms of release noises). We took the closure of the plosive to end at the sudden increase in amplitude at the beginning of the burst, and we assumed that the fricative and the release noise of the plosive ended where the amplitude of the wave form was nearly identical to that of the background noise.

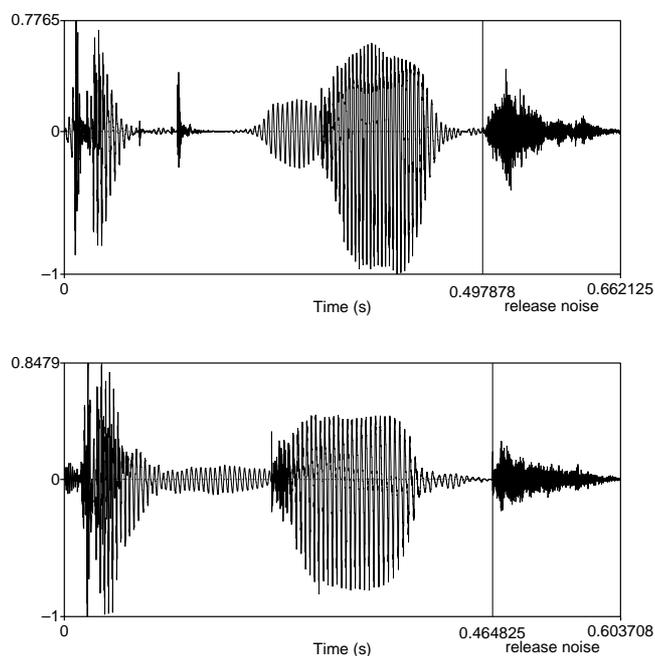


Figure 1: The wave forms of the phrases *ik duut* (upper panel) and *ik duud* (lower panel), with the vertical line in the wave forms indicating the beginning of the release noise.

We found a significant difference between the realizations of the graphemes *b* and *d* on the one hand and the graphemes *p* and *t* on the other hand. The graphemes *b* and *d* were on average realized with shorter release noises (mean: 96 ms) than *p* and *t* (mean: 119 ms, paired  $t(42) = 4.14$ ,  $p < 0.001$ ; all t-tests in this study are two-tailed). We observed no significant difference between the closure durations of the plosives (paired  $t(42) = 0.06$ ,  $p = 0.957$ ), nor between the durations of the preceding vowel plus sonorant consonant sequences (paired  $t(42) = 0.31$ ,  $p = 0.762$ ). This implies that the difference in release noise does not simply

result from a difference in speech rate. Shorter release noises are characteristic of voiced<sup>6</sup> obstruents (Slis and Cohen 1969), and our results therefore provide additional evidence that incomplete neutralization in Dutch may be evidenced by the duration of the release noise. In addition, we conclude that spelling can induce incomplete neutralization in Dutch, even if the speaker does not read minimal word pairs.

We found no significant differences between the realizations of *g* and *ch*. These fricatives did not differ in duration ( $t(15) = -0.99$ ,  $p = 0.343$ ), nor did the preceding vowel plus sonorant consonant sequence ( $t(15) = 1.52$ ,  $p = 0.149$ ). The absence of a significant difference may be due to the low number of final velar fricatives (16) in our data set. In addition, it may be due to the weak status of the voiced velar fricative in Dutch. The distinction between the voiced and the voiceless velar fricative is only functional in intervocalic positions, and only for speakers of particular variants of Standard Dutch.

We then investigated whether the words with constant spelling had been realized with different amounts of final voicing in the two lists. We measured the durations of the vowels, and the duration of the codas. If the words with constant spelling are more voiced in List Slightly Voiced than in List -Voice, we expect that their vowels are longer in List Slightly Voiced. With respect to the durations of the codas, although voiced obstruents tend to be shorter than voiceless obstruents in intervocalic position, we do not expect that the codas of the words with constant spelling are shorter in List Slightly Voiced than in List -Voice. Ernestus and Baayen (in press) found that the period of vocal fold vibration is an important cue to the voicing of final obstruents in Dutch, as it is for the voicing of obstruent clusters in intervocalic positions (Van den Berg 1988). The period of vocal fold vibration can be longer if the coda itself is longer. Hence, our speaker may have lengthened the codas in List Slightly Voiced in order to have them sound as more voiced. We also measured the periods of vocal fold vibration in the obstruent clusters, on the basis of the spectrograms of the sound files. We expect them to be longer in List Slightly Voiced than in List -Voice.

First, we compared the words with constant spelling with respect to the durations of their vowels, by means of a linear mixed effects model (LME, Pinheiro and Bates 2000) with the pseudo word as error stratum and with the duration of the vowel as the dependent variable. We took as independent variables List (List -Voice, List Slightly Voice) and the qualitative length of the vowel, which is short for [i y u ε ɪ ɔ α ʏ] and long for the other Dutch vowels, and reflects the fact that in the strong positions of prosodic feet these other Dutch vowels are phonetically long (Rietveld, Kerkhoff, and Gussenhoven 1999). This model did not provide evidence for differences in vowel duration between the two lists.

An LME model with duration of the coda as dependent variable, the pseudo word as error stratum, and List, Qualitative length of the vowel, and Type of coda (single fricative, fricative-fricative, plosive-fricative, plosive-plosive, fricative-plosive) as independent variables revealed main effects for Type of coda ( $F(4,105) = 51.62$ ,  $p < 0.001$ ) and List ( $F(1,112) = 5.20$ ,  $p = 0.025$ ), and an interaction of Type of coda by Qualitative length of the vowel ( $F(5,105) = 2.76$ ,  $p < 0.022$ ), after removal of two outliers. The different types of codas had average durations of 312.0 ms, 388.6 ms, 399.6 ms, 314.5 ms and 351.1 ms, respectively. Single fricatives and fricative-fricative clusters were longer after qualitatively long vowels than after

short vowels (on average 2.5 ms and 19.9 ms longer, respectively), whereas fricative-plosive<sup>7</sup>, plosive-fricative, and plosive-plosive clusters were on average longer after qualitatively short vowels (33.9 ms, 7.5 ms, and 35.1 ms longer, respectively). More importantly for our research question, codas were on average 10.9 ms shorter in List -Voice than in List Slightly Voiced. Codas may be longer in List Slightly Voiced because our speaker lengthened the codas in this list in order to lengthen the periods of vocal fold vibration. This hypothesis is supported by the analysis of the periods of vocal fold vibration in the obstruent clusters to which we turn now.

We analyzed also the periods of vocal fold vibration during the final obstruent clusters with an LME model with the pseudo word as error stratum. We discarded 11 items (out of 146) that were realized with no vocal fold vibration whatsoever. The removal of these outliers improved the normality of the dependent variable, which led to an improvement in the normality of the model's residuals. We took List, Type of coda, Duration of the coda, and Qualitative length of the preceding vowel as independent variables. The model revealed main effects for all factors, except the duration of the coda (List:  $F(1,54) = 87.42$ ,  $p < 0.001$ ; Type of coda:  $F(3,68) = 30.49$ ,  $p < 0.001$ ; Qualitative length of the vowel:  $F(1,68) = 12.11$ ,  $p < 0.001$ ). In addition, there was an interaction of List by Type of coda ( $F(3,54) = 10.29$ ,  $p < 0.001$ ), which is illustrated in Table 2. This table gives for each list the mean length of the period of vocal fold vibration during the different types of obstruent clusters. The differences between the lists are significant for the plosive-fricative clusters (paired  $t(19) = -5.36$ ,  $p < 0.001$ ), the plosive-plosive clusters (paired  $t(11) = -7.29$ ,  $p < 0.001$ ), and the fricative-plosive clusters (paired  $t(24) = -2.85$ ,  $p = 0.009$ ). The periods of vocal fold vibration are longer in List Slightly Voiced than in List -Voice, which suggests that the words of List Slightly Voiced have been realized with more final voicing than the words of List -Voice. The LME model also revealed an interaction of the duration of the coda by the qualitative length of the vowel ( $F(1,54) = 4.86$ ,  $p = 0.032$ ). Longer codas were realized with longer periods of vocal fold vibration after qualitatively long vowels. Finally, we also found an interaction of the duration of the coda by List ( $F(2,54) = 4.78$ ,  $p = 0.012$ ). Longer codas are realized with longer periods of vocal fold vibration, especially in List Slightly Voiced. This interaction supports the hypothesis that our speaker has lengthened the codas in List Slightly Voiced in order to lengthen the periods of vocal fold vibration, which may make these codas sound as more voiced. We conclude that the words with constant spelling were realized with more voicing in List Slightly Voiced than in List -Voice.

Experiment I shows that in Dutch the spelling may affect how a speaker realizes the fine acoustic details of final obstruents. The next experiment addresses the question of whether Dutch listeners might be sensitive to these acoustic characteristics in a past-tense production task. In this experiment, participants listened to the pseudo words of either List -Voice or List Slightly Voiced, and were asked to write down the past-tense forms of these verbs. To do so, they had to interpret the final obstruents as underlyingly voiced or voiceless. This task does not require sensitivity to incomplete neutralization, since the structure of the rime of the words is a crucial determinant for the underlying voicing (Ernestus and Baayen

Table 2: The mean length of the period of vocal fold vibration in milliseconds during the four types of final obstruent clusters in the words with constant spelling in the two lists.

Obstruent cluster	List -Voice	List Slightly Voiced
Fricative + Fricative	11.143	11.200
Plosive + Fricative	17.120	34.875
Plosive + Plosive	21.008	40.225
Fricativ + Plosive	10.396	15.496

2003). If listeners nevertheless make use of the subtle acoustic characteristics of the words for classifying the final obstruents as voiced or voiceless underlyingly, then we may expect them to interpret the final obstruents that are slightly voiced to be underlyingly voiced more often than the obstruents that were realized as completely voiceless. This would allow us to conclude that listeners take advantage of incomplete neutralization, even when there are no intrinsic reasons for them to do so.

In Experiment II, participants listened either to List -Voice or List Slightly Voiced. That is, they heard either no slightly voiced obstruents or a high number of such obstruents. We might expect only a small effect of slight voicing in a list with many slightly voiced and few voiceless final obstruents. Speech is highly variable, and listeners can interpret subtle characteristics of the acoustic signal, such as segment duration, only by comparing them to those of other parts of the signal. Thus, listeners interpret vowels as long or short by reference to the durations of preceding segments (Nootboom 1979; Nootboom and Doodeman 1980; Kemps et al. 2002). In a list with relatively many slightly voiced obstruents, slight voicing is the norm, and will not be interpreted as providing much information about the underlying voice specification of the final obstruent. This means that if we nevertheless find a significant difference in the responses to List -Voice and List Slightly Voiced, we can conclude that the interpretation of incomplete neutralization is a strong intrinsic feature of Dutch.

### 3. Experiment II

*Participants.* Fifty-six native speakers of Dutch, most of them students at Nijmegen University, were paid to participate in the experiment. Twenty-eight participants listened to the phrases of List -Voice, and twenty-eight participants listened to the phrases of List Slightly Voiced.

*Materials.* We used the phrases recorded for Experiment I. These phrases were presented in one of three random orders to the participants, with four intervening breaks. The actual test phrases were preceded by eleven practice phrases with existing verbs, and twenty practice phrases with pseudo verbs, which had also been recorded by the speaker from Experiment I.

*Procedure.* The participants listened through closed head-phones (Sony MDR-55) to the phrases, and wrote down as accurately as possible the past-tense forms of the pseudo verbs.

The experiment was self-paced. Participants heard a new phrase only after they had indicated that they were ready by pressing a button.

*Results and discussion.* In the great majority of trials, participants wrote past-tense forms ending in *te* or *de*. Some past-tense forms, however, were created by vowel alternation. For instance, a number of participants produced *ties* as the past-tense form for [tas], and *bast* as the past-tense form for [bɪst]. This type of past-tense formation seems to be independent of the underlying [voice]-specification of the stem-final obstruent, and we therefore did not take these forms into account in the analysis. We also discarded responses ending in *bte*, *pde*, *dte*, *tde*, *zte*, *vte*, *gte*, and *chde*, which are illegal according to the Dutch spelling conventions. In these responses, the grapheme representing the stem-final obstruent indicates an underlying [voice]-specification that is opposite to the specification indicated by the form of the past-tense suffix, and we do not know whether the participants interpreted the stem-final obstruent as underlying voiced or voiceless. Finally, we also discarded past-tense forms the stems of which do not completely correspond to the presented stimuli. The participants probably misunderstood the words in these trials. In total, we discarded 8.58% of the produced past-tense forms (831 out of 9688).

We calculated the log of the ratio of the resulting number of *de* responses to the resulting number of *te* responses for each verb in each list. This logit is our dependent variable.

We first investigated whether the participants determined their responses on the basis of phonological gang bias, as did the participants in the study by Ernestus and Baayen (2003). Ernestus and Baayen (2003) determined the relevant phonologically gangs by means of a Classification Tree (Breiman et al. 1984). This tree had as its input all 1697 lexical stems that end in an obstruent and are present in the CELEX lexical data base (Baayen, Piepenbrock, and Gulikers 1995). It grouped these stems into 11 gangs of words, such that words ending in the same rime are grouped together, and rimes with a similar preference for an underlyingly voiced specification for the final obstruent form one gang. The gang bias for a given word is the proportion of words in its gang underlyingly ending in a voiced obstruent.

Recall that the two lists contained three types of words. First, they contained words ending in a sonorant segment plus a bilabial or alveolar plosive. This plosive was slightly voiced in List Slightly Voiced. Second, the lists contained words ending in a velar fricative, which was equally voiced in the two lists. Third, the lists contained words with constant spelling, whose final obstruents were more voiced in List Slightly Voiced than in List -Voice.

We analyzed the logits using a linear model with Phonological gang bias, List (List -Voice and List Slightly Voiced), and Type of word as independent variables. We found main effects for all these independent variables (Phonological gang bias:  $F(1,339) = 539.79$ ,  $p < 0.001$ ; List:  $F(1,339) = 11.67$ ,  $p < 0.001$ ; Type of word:  $F(2,511) = 46.33$ ,  $p < 0.001$ ). In addition, we also found an interaction of Phonological gang bias by Type of word ( $F(2,339) = 5.90$ ,  $p = 0.003$ ).

As Phonological gang bias is not properly normally distributed, we also carried out non-parametric tests in order to ascertain an effect of this factor. First, Spearman correlation tests for List -Voice and List Slightly Voiced separately confirm that Phonological gang bias is

Table 3: The mean proportion of *de* responses in List -Voice (Experiment II), List Slightly Voiced (Experiment II), and List Mix (Experiment III) for the words ending in sonorant segments plus plosives, the words ending in velar fricatives, and the words with constant spelling.

Word type	List -Voice	List Slightly Voiced	List Mix
Sonorant + Plosive	0.072	0.104	0.137
Velar fricative	0.770	0.779	0.719
Constant Spelling	0.216	0.278	0.251

a good predictor of the participants' responses (List -Voice:  $r_s = 0.63$ ,  $S = 320439$ ,  $p < 0.001$ ; List Slightly Voiced:  $r_s = 0.58$ ,  $S = 365893$ ,  $p < 0.001$ ). Second, a Regression Tree (Breiman et al. 1984) also supports Phonological gang bias as a predictor of the participants' responses. We conclude that listeners based their choice between *te* and *de* at least in part on the phonological gangs. The participants did not have to take advantage of the subtle cues to voicing in the acoustic signal to do this task.

The question of interest to the present study is whether the participants nevertheless took advantage of incomplete neutralization. We therefore fitted a linear model to the data with the logit of the *de* and *te* responses as dependent variable, and with List and Type of word as independent variables. List ( $F(1,170) = 31.11$ ,  $p < 0.001$ ) and Type of word ( $F(2,170) = 60.10$ ,  $p < 0.001$ ) emerged as significant predictors. There was no interaction of List by Type of word. Table 3 gives the mean proportions of *de* responses for the three types of words in the two lists.

The words ending in sonorant segments plus plosives received more *de* responses from the listeners to List Slightly Voiced than from the listeners to List -Voice. This is in line with our finding in Experiment I that these words were realized with more voiced-like plosives in List Slightly Voiced than in List -Voice. Listeners clearly use the subphonemic cues in the signal when choosing the underlying [voice]-specifications for the final obstruents of pseudo words. This is not a trivial result, since listeners also base their choice on the phonological gangs of the words, and therefore do not need to take advantage of incomplete neutralization. Apparently, listeners also rely on incomplete neutralization when there are no compelling reasons for them to do so. In Dutch, incomplete neutralization emerges as a subphonemic cue to past-tense formation.

The words with constant spelling also received more *de* responses from the listeners to List Slightly Voiced than from the listeners to List -Voice. This result shows that longer periods of vocal fold vibration lead listeners to interpret final obstruents as voiced more often. Moreover, it shows that listeners interpret incomplete neutralization even when different interpretations do not map onto different spellings.

Finally, the words ending in velar fricatives received approximately similar responses in List -Voice and List Slightly Voiced. This suggests that these words were equally voiced in the two lists, in line with the absence of systematic acoustic differences in Experiment I.

In Experiment II, listeners heard either no slightly voiced obstruents or a high proportion<sup>11</sup> of slightly voiced obstruents. In Experiment III, we investigated the effect of the proportion of slightly voiced obstruents in the list. We constructed a new list, List Mix, by taking the words from List Slightly Voiced, but replacing the acoustic realizations of the words with constant spelling with the acoustic realizations of these same words from List -Voice. The only slightly voiced obstruents in List Mix are thus the plosives following sonorant segments. We expected that these slightly voiced plosives would be more often interpreted as underlyingly voiced in List Mix than in List Slightly Voiced, since slight voicing is more exceptional in List Mix, and should be more informative.

#### 4. Experiment III

*Participants.* Thirty native speakers of Dutch, who had not participated in Experiment II, took part in the experiment. Most of them were students at Nijmegen University. They were paid for their participation.

*Materials.* The materials are identical to List Slightly Voiced, except that the realizations of the words with constant spelling were replaced by the realizations from List -Voice. We refer to this new list as List Mix.

*Procedure.* The procedure was identical to that of Experiment II.

*Results and discussion.* We discarded the same types of responses as we did in Experiment II, which represent 9.21% of all responses.

We analyzed the data of this experiment together with the data of Experiment II, with the log of the ratio of the number of *de* responses to the number of *te* responses as the dependent variable. A linear model with List and Type of word as independent variables revealed a main effect of List ( $F(2,340) = 15.71, p < 0.001$ ) and Type of word ( $F(2,170) = 50.33, p < 0.001$ ). In addition, there was an interaction of List by Type of word ( $F(4,340) = 5.57, p < 0.001$ ). This interaction is illustrated in Table 3, which gives the mean proportion of *de* responses for the three types of words in the three lists.

The words ending in sonorants plus plosives elicited more *de* responses in List Mix than in List Slightly Voiced (paired  $t(41) = -2.517, p = 0.016$ ), even though their acoustic characteristics were exactly identical in these two lists. This difference must be due to the words with constant spelling, since List Slightly Voiced and List Mix only differ with respect to the acoustic realizations of just these words. The words with constant spelling had more final voicing in List Slightly Voiced than in List -Voice and List Mix. We conclude that more voiced realizations of the words with constant spelling diminishes the likelihood of voiced interpretations for the slightly voiced plosives following sonorant segments. This points to a list bias effect on the part of the listeners. Listeners attach more value to slight voicing if slight voicing is not the norm.

As can be seen in Table 3, the words with constant spelling elicited fewer *de* responses in List Mix than in List Slightly Voiced (paired  $t(114) = -4.78, p < 0.001$ ). In contrast, the

percentage of *de* responses for these words does not differ significantly between List Mix and List -Voice (paired  $t(114) = -1.63$ ,  $p = 0.105$ ). Recall that Lists -Voice and Mix contained exactly the same acoustic realizations of the words with constant spelling, and that in List Slightly Voiced these words had been realized with more final voicing. Hence, the listeners' reactions to the completely voiceless realizations of these words do not provide evidence for a list effect that is due to the listeners. Their interpretations for the completely voiceless realizations seem to be based mainly on the acoustic signal for these realizations themselves.

Finally, we turn to the words ending in velar fricatives. They show the same pattern over the three lists as the words with constant spelling. A linear model excluding the words ending in sonorant segments plus plosives shows no interaction between Type of word and List. Note, however, that the number of words ending in velar fricatives is relatively small, and the absence of a significant difference between the words ending in velar fricatives and the words with constant spelling is therefore not very informative.

In conclusion, the responses to the three lists indicate that slightly voiced obstruents are interpreted as underlyingly voiced more often when the list contains fewer slightly voiced obstruents. In other words, slight voicing is considered to be more informative if it is not the norm.

## 5. General discussion

This study investigates whether listeners take advantage of incomplete neutralization even when they have no intrinsic reasons for doing so. In Experiment I, a speaker recorded pseudo words, and we analyzed various acoustic characteristics of the realizations that may be relevant to the perception of final voicing. In Experiments II and III, we presented the realizations to listeners, and we investigated whether listeners take advantage of the incomplete neutralization in the acoustic signal also if they are performing a task for which this is not necessary.

Our speaker recorded two lists containing the same pseudo verbs ending in obstruents. The lists differed mainly in that nearly all words were spelled with voiceless final graphemes in List -Voice, while the words ending in sonorant segments plus plosives were spelled with voiced final graphemes in List Slightly Voiced. Although our speaker was not aware of the fact that the words in the two lists formed minimal and identical word pairs, she realized the plosives represented as voiced with shorter release noises than the plosives represented as voiceless. The experiment therefore provides additional evidence that incomplete neutralization in Dutch may be evidenced by the release noise. In addition, it shows that incomplete neutralization may be induced by orthography. Finally, this speaker realizes words with incomplete neutralization, even when she does not know that she is realizing minimal word pairs. Apparently, minimal pairs are not necessary for speakers to realize words with incomplete neutralization.

A large number of words were spelled identically in the two lists. These words with constant spelling were realized with longer periods of vocal fold vibration in List Slightly Voiced than in List -Voice. This shows that our speaker varied the voicing of the final obstruents as a function of the voicing of the other final obstruents in the list. This speaker

related transfer effect probably results from the automation of the articulatory gestures<sup>13</sup> resulting in slightly voiced or completely voiceless obstruents.

In Experiment II, we presented the lists to two different groups of listeners. These listeners were asked to create the past-tense forms for the pseudo verbs, for which they had to interpret the final obstruents as underlyingly voiced or voiceless. Previous research (Ernestus and Baayen 2003) has shown that participants base their interpretation at least partly on the existing words belonging to the same phonological gangs as the pseudo verbs. Thus, participants do not need to take into account the subphonemic cues in the acoustic signal to underlying voicing in order to perform this task. Both groups of participants indeed based their interpretation of the final obstruents on the phonological gangs. In addition, listeners presented with slightly voiced obstruents more often opted for voiced interpretations. This shows that listeners take advantage of incomplete neutralization not only in spelling tasks that they can only perform by interpreting the incomplete neutralization. They also take advantage of incomplete neutralization when there are no intrinsic reasons for them to do so. In Dutch, incomplete neutralization emerges as a subphonemic cue for past-tense formation.

In Experiment II, listeners heard no (List -Voice) or many (List Slightly Voiced) slightly voiced obstruents. We expected that the high percentage of slightly voiced obstruents in List Slightly Voiced would attenuate the effect of slight voicing. Listeners interpret subtle characteristics of the acoustic signal by reference to the surrounding segments in the signal. If slight voicing is the norm, listeners might be less willing to base their interpretation of the final obstruent on slight voicing. In Experiment III, a new group of participants listened to a new list, List Mix, which is identical to List Slightly Voiced, except that the words with constant spelling were taken from List -Voice. The only slightly voiced obstruents in List Mix are therefore the plosives following sonorants. These plosives elicited more *de* responses in List Mix than in List Slightly Voiced. This shows that slightly voiced plosives are indeed interpreted as underlyingly voiced more often when there are fewer other slightly voiced obstruents in the list. This listener related transfer effect probably reflects the way in which speakers dynamically adapt to their conversation partner's speech habits or dialect.

The final obstruents of the words with constant spelling were interpreted as underlyingly voiced more often in List Slightly Voiced than in both List -Voice and List Mix, which is in accordance with their acoustic characteristics in these lists. The absence of a difference between List -Voice and List Mix suggests that listeners interpret completely voiceless obstruents directly on the basis of the acoustic signal without much reference to context.

Our finding that spelling induces incomplete neutralization does not mean that incomplete neutralization is simply an artefact of orthography. Incomplete neutralization has also been reported for Catalan minimal word pairs that do not reflect the difference in the underlying [voice]-specification of the final obstruents in their spelling (Dinnsen and Charles-Luce 1984). In addition, our finding that listeners take advantage of incomplete neutralization even when different interpretations do not map onto different spellings (that is, for the words with constant spelling) provides further evidence that incomplete neutralization is not a simple artefact of orthography.

One way to incorporate incomplete neutralization into the grammar is to implement it by

means of low-level phonetic rules. Speakers would realize final voiced graphemes as slightly voiced, because seeing a voiced grapheme would activate the low-level phonetic implementation rules for slight final voicing. Listeners would be affected by incomplete neutralization when interpreting a pseudo word, because incomplete neutralization would help them to build an underlying representation. When incomplete neutralization is considered as a result of phonetic implementation rules, the standard view of grammar has to be revised. Final Devoicing is a phonological rule, and, under the standard view, it would consequently apply before phonetic incomplete neutralization. As a consequence, phonetic incomplete neutralization processes would be unable to refer to the underlying distinction between voiced and voiceless obstruents, and cannot be applied. For instance, a word like /*verveid*/ (see 1) would be transformed into /*verveit*/, before it is passed to the phonetic component. The phonetic component, consequently, has no access to the underlying /*d*/, and we cannot explain why /*verveid*/ may be realized with slight final voicing. Several solutions have been suggested. For instance, Dinnsen and Charles-Luce (1984) and Slowiaczek and Dinnsen (1985) suggest that phonetic implementation rules may apply before phonological rules, and Port and O'Dell (1985) suggest that Final Devoicing and incomplete neutralization together form one implementation rule.

We would like to offer another solution. The mental lexicon contains auditory and visual form representations for a great many words, including inflected words (e.g., Baayen, Dijkstra, and Schreuder 1997; Alegre and Gordon 1999; Sandra, Frisson, and Daems 1999; Bertram, Schreuder, and Baayen 2000; Bybee 2001; Baayen et al. 2002; Baayen et al. 2003). Thus, with respect to the stems *verwijd* and *verwijt*, the lexicon of a Dutch speaker probably contains representations such as those listed in (2a) and (2b), respectively. The lexical storage of inflectional forms makes abstract underlying forms superfluous. The form *verwijd* needs not be stored as /*verveid*/, but can be stored as /*verveit*/, directly reflecting its pronunciation. The fact that this stem has a final [d] in the infinitive and plural forms is accounted for by the representations of these forms themselves.

- (2) a. [verveit] *verwijd* ‘widen’  
 [verveidən] *verwijden* ‘to widen’  
 [verveidən] *verwijden*, ‘widen’ (plural)  
 etc.
- b. [verveit] *verwijt* ‘reproach’  
 [verveitən] *verwijten* ‘to reproach’  
 [verveitən] *verwijten*, ‘reproach’ (plural)  
 etc.

Given lexical representations that directly reflect pronunciation, the production and interpretation of incomplete neutralization may well be primarily a lexical effect, in two ways. First, we think that the lexical representations of words contain information concerning the probability that the final obstruent is realized as slightly voiced (see, e.g., Bybee 2001 and Pierrehumbert 2003, who also assume that variability is represented in the lexical representations). The stochastic representation of the final plosive has a shorter expected release

noise duration in *verwijd* than in *verwijt*. Both speakers and listeners might well make use of this stochastic information. 15

Independently of this possibility, we also think that incomplete neutralization is affected by lexical analogy. Lexical analogy can explain the production of existing words with incomplete neutralization. When speakers have to realize *verwijd*, they activate [vɛrveit] as well as the words inflectionally related to *verwijd*, such as [vɛrveidən] (see also Bybee 2001). These inflectionally related words contain [d], and they support the realization of the final obstruent of *verwijd* as [d]. This may enhance the probability that the final obstruent is realized as slightly voiced.

The stored inflectionally related word forms probably also affect the listener's interpretation of incomplete neutralization for existing words. When listeners perceive a word, the auditory representations that best match the word are activated, including nearest neighbors such as inflectional variants (e.g., Bybee 1985). The activation levels of these neighbors are co-determined by the size of the mismatch between the acoustic signal and the auditory representations of the words. A neighbor with a voiced obstruent is more activated by a word realized with a slightly voiced obstruent than by a word realized with a completely voiceless obstruent. The final activation levels of the neighbors co-determine the interpretation of the acoustic signal. Thus, upon hearing [vɛrveit], the words in (2) all become activated to some degree. If the final obstruent of [vɛrveit] is slightly voiced, the inflectionally related words of *verwijd* are co-activated more than if this obstruent is completely voiceless. Hence, a slightly voiced realization makes the interpretation of the word as 'widen' more likely. If, conversely, the final obstruent is realized as completely voiceless, the meaning 'reproach' becomes more probable.

Lexical analogy can also explain the production and interpretation of incomplete neutralization for pseudo words, studied in this paper. If a speaker reads a pseudo word, the representations of existing words with similar spellings are probably activated. These orthographic neighbors may affect the speaker's realization of the pseudo word, with each activated neighbor affecting especially that part of the pseudo word to which it is similar. Thus, the reading of *dijp* may activate the Dutch words *dal* and *rijp*, and the phonetic representation of *dal* may affect especially the realization of the *d* of the pseudo word, while the phonetic representation of *rijp* affects the rime of *dijp*. If the final obstruent of the pseudo word is represented as voiced, the orthographic neighbors ending in the same orthographic rime tend to be realized with slight final voicing. As a consequence, speakers may realize also the pseudo word with slight final voicing. If, in contrast, the final obstruent is represented as voiceless, the activated orthographic neighbors ending in the same rime tend to be realized without slight voicing, and speakers realize the final obstruent of the pseudo word as completely voiceless.

As to the interpretation of incomplete neutralization in pseudo words, when a listener hears a pseudo verb with slight final voicing, especially the nearest phonological neighbors with similar subphonemic characteristics are co-activated. These neighbors tend to have inflectional paradigms supporting *de* as allomorph rather than *te*. Hence, the likelihood that the listener selects *de* increases compared to the case when the listener hears a pseudo verb

without slight voicing.

16

We have phrased our interpretation using the metaphors of localist spreading activation modeling. We believe, however, that our data can be incorporated straightforwardly in non-localist theories based on statistical language models or non-localist artificial neural networks, if so desired. Also these approaches can easily account for the fact that incomplete neutralization is part and parcel of the grammar of Dutch.

## References

- Alegre, Maria and Peter Gordon 1999 Frequency effects and the representational status of regular inflections. *Journal of Memory and Language* 40: 41–61.
- Baayen, R. Harald, Ton Dijkstra and Rob Schreuder 1997 Singulars and plurals in Dutch: Evidence for a parallel dual route model. *Journal of Memory and Language* 37: 94–117.
- Baayen, R. Harald, James McQueen, Ton Dijkstra and Rob Schreuder 2003 Frequency effects in regular inflectional morphology: Revisiting Dutch plurals. In: R. Harald Baayen and Rob Schreuder (eds.), *Morphological structure in language processing*, 355–390. Berlin: Mouton de Gruyter.
- Baayen, R. Harald, Richard Piepenbrock, and Leon Gulikers 1995 *The CELEX lexical database (CD-ROM)*. University of Pennsylvania, Philadelphia, PA: Linguistic Data Consortium.
- Baayen, R. Harald, Rob Schreuder, Nivja H. De Jong and Andrea Krott 2002 Dutch inflection: the rules that prove the exception. In: Sieb Nooteboom, Fred Weerman and Fred Wijnen (eds.), *Storage and Computation in the Language Faculty*, 61–92. Dordrecht: Kluwer Academic Publishers.
- Baumann, Monika 1995 *The production of syllables in connected speech*. Ph.D. dissertation, Max-Planck Institute for Psycholinguistics, Nijmegen.
- Bertram, Raymond, Rob Schreuder and R. Harald Baayen 2000 The balance of storage and computation in morphological processing: The role of word formation type, affixal homonymy, and productivity. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 26: 419–511.
- Geert E. Booij 1995 *The phonology of Dutch*. Oxford: Clarendon Press.

Breiman, Leo, Jerome H. Friedman, Richard A. Olshen and Charles J. Stone 1984  
*Classification and Regression Trees*. Belmont, California: Wadsworth International Group.

Bybee, Joan L. 1985 *Morphology: A study of the relation between meaning and form*.  
Amsterdam: Benjamins.

Bybee, Joan L. 2001 *Phonology and language use*. Cambridge: Cambridge University  
Press.

Dinnsen, Daniel A. and Jan Charles-Luce 1984 Phonological neutralization, phonetic  
implementation and individual differences. *Journal of Phonetics* 12: 49–60.

Ernestus, Mirjam and R. Harald Baayen 2003 Predicting the unpredictable: Inter-  
preting neutralized segments in Dutch. *Language* 79(1): 5–38.

Ernestus, Mirjam and R. Harald Baayen in press Intraparadigmatic effects on the  
preception of voice. In Erik Jan Van der Torre and Jeroen Van de Weijer (eds.), *Voicing in  
Dutch*. Submitted to the publisher.

Fourakis, Marios and Gregory K. Iverson 1984 On the ‘incomplete neutralization’ of  
German final obstruents. *Phonetica* 41: 140–149.

Jongman, Allard, Joan A. Sereno, Marianne Raaijmakers and Aditi Lahiri 1992 The  
phonological representation of [voice] in speech perception. *Language and Speech* 35: 137–  
152.

Kemps, Rachèl, Mirjam Ernestus, Rob Schreuder and R. Harald Baayen 2002 Sub-  
segmental and suprasegmental cues for morphological complexity: The case of Dutch noun  
plurals. Manuscript Max-Planck Institute for Psycholinguistics, Nijmegen.

Kenstowicz, Michael 1994 *Phonology in generative grammar*. Cambridge MA & Ox-  
ford UK: Blackwell.

Nooteboom, Sieb G. 1979 Perceptual adjustment to speech rate: A case of backward  
perceptual normalization. In: H. V. Deighton-Van Witsen (ed.), *Anniversaries in phonetics:  
Studia gratulatoria dedicated to Hendrik Mol*, 276–287. Amsterdam: Institute of Phonetic  
Sciences Amsterdam.

Nooteboom, Sieb G. and G. J. N. Doodeman 1980 A note on speech rate. *IPO Annual  
Progress report*: 58–60.

Pierrehumbert, Janet B. 2003 Probabilistic phonology: Discrimination and robust-

ness. In: Rens Bod, Jennifer Hay and Stefanie Jannedy (eds.), *Probability theory in linguistics*<sup>18</sup>, 177–228. Cambridge, MA: The MIT Press.

Pinheiro, José C. and Douglas M. Bates 2000 *Mixed-effects models in S and S-PLUS*. (Statistics and Computing.) New York: Springer.

Port, Robert and Penny Crawford 1989 Incomplete neutralization and pragmatics in German. *Journal of Phonetics* 17: 257–282.

Port, Robert F. and Michael L. O'Dell 1985 Neutralization of syllable-final voicing in German. *Journal of Phonetics* 13: 455–471.

Rietveld, Toni, Joop Kerkhoffs, and Carlos Gussenhoven 1999 Prosodic structure and vowel duration in Dutch. In: John Ohala, Yoko Hasegawa, Manjari Ohala, Daniel Granville and Ashlee C. Baily (eds.), *Proceedings of the 14th International Congress of Phonetic Sciences, San Francisco, 1-7 August 1999*, 463–466. Berkeley: Linguistic Department, University of California.

Sandra, Dominiek, Steven Frisson and Frans Daems 1999 Why simple verb forms can be so difficult to spell: The influence of homophone frequency and distance in Dutch. *Brain and Language* 68 (1/2): 277–283.

Slis, I H. and A. Cohen 1969 On the complex regulating the voiced-voiceless distinction ii. *Language and Speech* 12: 80–102; 137–155.

Slowiaczek, Louisa A. and Daniel A. Dinnsen 1985 On the neutralizing status of Polish word-final devoicing. *Journal of Phonetics* 13: 325–341.

Van den Berg, R, J, H. 1988 The perception of voicing in Dutch two-obstruent sequences. Ph.D. dissertation, Catholic University Nijmegen, Nijmegen.

Warner, Natasha, Allard Jongman, Joan Sereno and Rachèl Kemps in press Sub-phonemic durational differences in production and perception. *Journal of Phonetics*.

## Notes

<sup>1</sup> We would like to thank Mirjam Broersma, Taehong Cho, Anne Cutler, Delphine Dahan, Carlos Gussenhoven, Terrance Neary, and Tania Zamuner for stimulating discussion and comments, and Doug Whalen and two anonymous reviewers for their constructive criticism.

<sup>2</sup> Interestingly, Jongman et al. (1992) and Baumann (1995) failed to find incomplete<sup>19</sup> neutralization in Dutch. Jongman et al. (1992) investigated ten minimal word pairs, while Baumann (1995) restricted her study to twelve word pairs. Possibly these studies did not detect incomplete neutralization because the effects are too subtle in Dutch to surface in a small number of items, as is also suggested by Warner et al. (in press).