Available online at www.sciencedirect.com

SCIENCE DIRECT®

Journal of Phonetics 32 (2004) 455–491

Journal of Phonetics

www.elsevier.com/locate/phonetics

Acoustical and perceptual analysis of the voicing distinction in Dutch initial plosives: the role of prevoicing

Petra M. van Alphen*, Roel Smits

Max Planck Institute for Psycholinguistics, P.O. Box 310, 6500 AH Nijmegen, The Netherlands

Received 6 November 2002; received in revised form 22 April 2004; accepted 17 May 2004

Abstract

Three experiments investigated the voicing distinction in Dutch initial labial and alveolar plosives. The difference between voiced and voiceless Dutch plosives is generally described in terms of the presence or absence of prevoicing (negative voice onset time). Experiment 1 showed, however, that prevoicing was absent in 25% of voiced plosive productions across 10 speakers. The production of prevoicing was influenced by place of articulation of the plosive, by whether the plosive occurred in a consonant cluster or not, and by speaker sex. Experiment 2 was a detailed acoustic analysis of the voicing distinction, which identified several acoustic correlates of voicing. Prevoicing appeared to be by far the best predictor. Perceptual classification data revealed that prevoicing was indeed the strongest cue that listeners use when classifying plosives as voiced or voiceless. In the cases where prevoicing was absent, other acoustic cues influenced classification, such that some of these tokens were still perceived as being voiced. These secondary cues were different for the two places of articulation. We discuss the paradox raised by these findings: although prevoicing is the most reliable cue to the voicing distinction for listeners, it is not reliably produced by speakers.

1. Introduction

In phonetic research, the term 'acoustic correlate' is often used to indicate an acoustic property which covaries with a phonemic distinction. A large body of phonetic research has been devoted to identifying such acoustic correlates for a number of phonetic distinctions. This research has

*Corresponding author. Present address: Department of Psychology, University of Amsterdam, Roetersstraat 15, 1018 WB Amsterdam, The Netherlands. Tel.: +31-20-5256808; fax: +31-20-6391656.

E-mail address: p.m.vanalphen@uva.nl (P.M. van Alphen).

0095-4470/$ - see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.wocn.2004.05.001

shown that each phonemic distinction has several acoustic correlates. Subsequent perceptual experiments, employing synthetic stimuli in which one or more of these correlates were systematically varied, have shown that listeners are sensitive to many or all of these correlates when recognizing phonemes. An acoustic correlate which influences the perception of a phonemic distinction is often referred to as an 'acoustic cue' to that distinction.

The present study focuses on the voicing distinction in Dutch initial plosives, that is, the phonological distinction between [+voice] and [−voice]. In particular, we aimed to identify the most important acoustic correlates of voicing in Dutch initial plosives, to establish which of these correlates are theoretically most reliable for recognizing voicing and to determine which of the correlates are actually the strongest cues in listeners' categorizations. Our use of natural speech stimuli enabled us to study voicing perception when the full array of acoustic cues was present.

The phonological distinction between [+voice] and [−voice] in plosives has been one of the most intensively studied distinctions. Most languages contrast these two phonemic classes (which we will refer to as *voiced* and *voiceless* plosives), but the phonetic realization of this phonological distinction varies among languages. The moment that the vocal folds start vibrating relative to the moment of the release of the closure, the so-called voice onset time (VOT) plays an important role in these different acoustic realization. The notion VOT was introduced by Lisker and Abramson (1964), who measured the VOT of plosive production in 11 languages. They concluded that, across languages, three different VOT categories could be distinguished. The first category of plosives had a negative VOT, that is, they were produced with voicing during the closure. The second category of plosives had a slightly positive VOT; these plosives were produced with little or no aspiration. The third class had a clear positive VOT; these plosives were produced with aspiration. Given these three VOT categories, any language could thus in principle employ a three-way voicing distinction. There are, however only a few languages, for example Thai, which contrast these three voicing categories (fully voiced, voiceless unaspirated and voiceless aspirated) in plosives. Most languages have a two-way voicing distinction, which is implemented by two adjacent modes, one of which is associated with the phonologically voiced and the other with the phonologically voiceless plosive. A study by Keating, Linker, and Huffman (1983), in which 51 different languages were surveyed, showed that the voiceless unaspirated category is the most common category; it is used in almost all these languages. The two other categories, the fully voiced and voiceless aspirated category, appear equally often as the voicing category contrasting with the voiceless unaspirated category. Furthermore, Keating et al. (1983) observed that in many languages the use of these different VOT categories varies as a function of the position in a word at which the plosive occurs. In the present study, we will focus on plosives in initial position of words spoken in isolation.

The way in which the voicing distinction is implemented phonetically is different in Dutch than in most other Germanic languages. While most Germanic languages such as Danish, English, and German contrast voiceless unaspirated and voiceless aspirated plosives in initial position (Keating, 1984), Dutch does not. Dutch belongs to the group of languages, including for example Arabic, Bulgarian, French, Japanese, Polish, Russian, and Spanish, which contrasts voiced and voiceless unaspirated plosives (Keating, 1984; Lisker & Abramson, 1964). That is, the initial voiced plosives are produced with a negative VOT, which we will refer to as *prevoicing*, and the initial voiceless plosives are produced with little or no aspiration. In Dutch, there are three voiceless plosive categories, namely [p], [t], and [k], but there are only two voiced plosive categories, namely [b] and [d]. The voiced velar plosive [g] is marginally present in Dutch as it only occurs in loan words (e.g., *goal*). Therefore, the voicing distinction in the velar plosives was not included in the present study.

Prevoicing is the production of vocal fold vibration during the closure phase of a plosive in initial position. Vocal fold vibration can only occur when certain physiological and aerodynamic conditions are met (van den Berg, 1958). First, the vocal folds must be adducted and tensed. Second, a sufficient transglottal pressure gradient is needed to result in enough positive airflow through the glottis to support vibration. The second condition is relatively hard to meet in the case of the closure of a plosive, since all outgoing pathways are closed. As a consequence, the air flowing through the glottis accumulates in the oral cavity, causing oral pressure to approach subglottal pressure (Ohala, 1983). This process will be delayed if the volume above the glottis increases. Therefore, expansion of the vocal tract volume will facilitate the production of voicing during closure. A part of the expansion can be achieved by active enlargement of the supraglottal cavity, namely by lowering the larynx, raising the soft palate, advancing the tongue root, or drawing the tongue dorsum and blade down (Westbury, 1983; Stevens, 1998). In addition to active enlargement, supraglottal volume can also be expanded passively due to the raised intraoral pressure, provided that the walls of the supraglottal cavity are lax (Rothenberg, 1968; Stevens, 1998; Svirsky, Stevens, Matthies, Manzella, Perkell, & Wilhelms-Tricarico, 1997). Although it is difficult to differentiate changes in vocal tract size resulting from active and passive expansion (Westbury, 1983), it is generally assumed that both mechanisms play a role in the production of prevoicing.

We would argue that the production of prevoicing, which is conditional on maintaining sufficient transglottal pressure difference while the vocal tract is closed, is relatively easily disrupted when the capacity for passive or active expansion of the vocal tract is reduced, for example, when the place of articulation is more posterior. This argument receives further support from several studies which have shown that children who acquire languages that contrast voiced plosives with prevoicing and voiceless unaspirated plosives master the adult pattern later than children who acquire languages that contrast voiceless unaspirated and voiceless aspirated plosives (e.g., Allen, 1985; Kewley-Port & Preston, 1974; Konefal & Fokes, 1981; Macken & Barton, 1980). The late acquisition of prevoicing may be due to the relatively small vocal tract size in children (Rothman, Koenig, & Lucero, 2002). The relatively small vocal tract reduces the capacity for expansion of the supraglottal cavity. In addition, one has argued that the late acquisition may be due to the complexity of the articulatory gestures which are demanded for the expansion of the vocal tract (Kewley-Port & Preston, 1974). It is important to bear in mind, however, the fact that prevoicing is used in such a considerable number of languages which implies that it is by no means too difficult or uneconomic to produce (Westbury & Keating, 1986).

Studies which have investigated the occurrence of prevoicing suggest that prevoicing is rarely absent in initial voiced plosives (Keating, Mikos, & Ganong, 1981 on Polish; Yeni-Komshian, Caramazza, & Preston, 1977 on Lebanese Arabic; Caramazza & Yeni-Komshian, 1974 on European French). Only one study, on Canadian French (Caramazza & Yeni-Komshian, 1974), has found a substantial degree of overlap between the VOT distributions of voiced and voiceless plosives; no less than 58% of the voiced tokens in that sample ($N = 90$) were produced without prevoicing, while all voiceless plosives were produced without aspiration. Caramazza and Yeni-Komshian argued that in Canadian French the VOT values are shifting as a result of the influence of Canadian English. There are, however, no studies which systematically investigated the occurrence of prevoicing in Dutch. One of the goals of the present study was therefore to gain more insight into the way in which prevoicing varies in Dutch. Several factors were included which could have affected the occurrence of prevoicing and its duration.

The most complete study of the acoustics and perception of voicing distinction in Dutch plosives was conducted by Slis and Cohen (1969). Apart from prevoicing, they measured several additional acoustic properties that were known to play a role in the voicing contrast. They did not always describe, however, the details of their elicitation and measurement procedures. They also investigated the influence of several acoustic properties on the perception of voicing using synthetic speech. They varied each acoustic correlate separately (or maximally two at a time). In this way an overview was given of the way in which each of the acoustic properties influenced listeners' perception of voicing when all the other cues were kept constant. This analysis did not show, however, how all these acoustic properties vary together in natural speech nor which of the acoustic correlates are relied on most strongly by listeners. A full understanding of the phonetics of voicing requires an analysis of the variability in natural utterances of voiced and voiceless plosives, and an analysis of how listeners deal with that variability. The present study sought to provide such analyses. Because this is the first large-scale study of the above issues in Dutch, we focused on voicing in plosives in initial position in words spoken in isolation. We note that sentence context may influence the phonetic realization of the voicing distinction (e.g., Lisker & Abramson, 1964).

This study consists of three experiments. Experiment 1 was designed to investigate variation in the production of prevoicing in Dutch plosives and whether the presence or absence of prevoicing and the duration of prevoicing is influenced by a number of potentially relevant factors. Experiment 2 is a detailed acoustic analysis of the voicing distinction in Dutch plosives. Several acoustic properties in addition to prevoicing were measured and analyzed in order to find out which of these properties were correlates of the voicing distinction. Subsequently, classification tree analyses were used to indicate which of these acoustic correlates would serve as the most reliable cues for correct recognition of voicing. Experiment 3 investigated how the tokens of Experiment 2 were perceived by listeners and asked which of the acoustic properties identified in Experiment 2 are relied on most strongly by listeners when they identify plosives as voiced or voiceless. Together, the three experiments provide a detailed analysis of the production and perception of the voicing distinction in Dutch initial plosives with particular emphasis on the role of prevoicing.

2. Experiment 1

Although it is generally assumed that the presence of prevoicing is one of the major attributes of the voiced–voiceless distinction in Dutch plosives, there are to our knowledge no published studies that actually report acoustic measurements on prevoicing in Dutch other than the study conducted by Lisker and Abramson (1964) and the study by van Dommelen (1983). All Dutch tokens of /b/ and /d/ analyzed by Lisker and Abramson were produced with a voice lead. However, their measurements were based on the production of only one speaker. Furthermore, the way in which the speech was elicited was not described. The items may have been presented without fillers, in which case the speaker might have been aware of the type of distinction under investigation. This could have stimulated him to hyper-articulate, which may have resulted in more prevoicing than may normally occur in Dutch. The VOT values reported by van Dommelen showed that prevoicing was sometimes absent in initial voiced plosives, but these values were based on only a few different words.

The purpose of the first part of the present study was therefore to conduct a systematic and large-scale study of prevoicing variation in Dutch voiced plosives. In particular, we aimed to find out whether voiced initial plosives were consistently produced with prevoicing and whether the presence or absence of prevoicing and its duration varied as a function of several factors. The influence of the sex of the speaker was investigated as well as the influence of two segmental and two lexical factors: the place of articulation of the plosive (labial versus alveolar); the phoneme following the plosive (vowel versus consonant); the lexical status of the carrier stimulus (word versus nonword); and the competitor environment of the carrier stimulus, that is, whether changing the first voiced plosive into its voiceless counterpart resulted in a word or a nonword (competitor versus no competitor). The effect of two of these factors, namely the sex of the speaker and the place of articulation, were also investigated in an experiment on the occurrence of prevoicing in English (Smith, 1978). Although English is one of the languages which contrasts voiceless unaspirated and voiceless aspirated plosives, some English speakers do occasionally produce phonologically voiced plosives with prevoicing (e.g., Docherty, 1992). Note, however, that prevoicing is in general not important for the voicing distinction in English. To ensure robustness of our results, we recorded several speakers.

For two of the five factors, namely the sex of the speaker and the place of articulation of the plosive, we had clear predictions. The prediction for the speakers' sex is based on differences in vocal tract size between men and women. The volume of the vocal tract is smaller in female than in male speakers (e.g., Stevens, 1998). Assuming equal volume velocity through the glottis, oral pressure will tend to rise more quickly in females than in males, which makes it harder to produce prevoicing. Smith (1978) indeed found that in English, prevoicing was less often produced by female speakers than by male speakers. In line with these findings, we therefore predicted a smaller proportion of prevoiced tokens in female speakers in comparison to male speakers.

As described earlier, one phenomenon that helps to maintain sufficient transglottal pressure is passive enlargement of the oral cavity due to raised intraoral pressure. For alveolar plosives, the pharyngeal walls and part of the soft palate can yield to expansion of the oral cavity, while for labial plosives these surfaces plus all of the tongue surface and parts of the cheek can participate in the expansion (Houde, 1968; Rothenberg, 1968). Furthermore, the freedom to actively expand the vocal tract through movements of the tongue body is expected to be smaller for /d/ than for /b/, since in the former case the tongue is already involved in maintaining the closure. The oral cavity can thus be expanded more during the production of labial plosives than during the production of alveolar plosives. Consequently, oral pressure tends to rise less quickly during the production of labial plosives than during the production of alveolar plosives. According to this account, the production of prevoicing was expected to be easier for labials than for alveolars. In line with this, the study by Smith (1978) showed that in English, place of articulation affected both the duration of prevoicing and the occurrence of prevoicing in the predicted direction. Therefore, Dutch labial plosives were also expected to be produced more often with prevoicing and with longer prevoicing than alveolar plosives.

The other segmental factor, namely the phoneme that followed the plosive, was included to test for possible differences between items in which the plosive was followed by a vowel and items in which the plosive was followed by a consonant. It is likely that the anticipatory coarticulation of the following phoneme affects vocal tract size and the degree to which the vocal tract size can be expanded. Smith (1978) for example found that the height of the following vowel had an influence on both the proportion of prevoiced tokens and on the duration of prevoicing. Although it is difficult to make detailed predictions for different consonants and vowels without the use of articulatory measurements, we included this factor in order to test whether the following phoneme influenced prevoicing production.

In addition to these two segmental factors, two lexical factors were included. It is possible that the lexical competitor environment of the carrier stimulus influences the production of prevoicing. Speakers might speak more carefully when producing words starting with a voiced plosive when there is a voiceless word competitor (because this reduces the chance that the voiced plosive will be mistakenly perceived as voiceless) than when there is no voiceless word competitor. The influence of the existence of a voiceless word competitor on prevoicing was therefore tested. Furthermore, the lexical status of the items themselves was included as a factor.

2.1. Method

2.1.1. Materials

Sixty-four items beginning with voiced plosives were selected. They were all monosyllabic. In the materials the following factors were varied: the place of articulation of the plosive (labial versus alveolar); the phoneme following the plosive (vowel versus consonant); the lexical status of the item (word versus nonword); and the competitor environment of the item, that is, whether changing the first voiced plosive into its voiceless counterpart resulted in a word or a nonword (competitor versus no competitor). The vowels that followed the plosives were: /a/, /ɑ/, /o/, /ɔ/, /i/, /ɪ/, /ɛ/, /ɛɪ/, /œy/, and /ø/. The consonants that followed the plosives were: /l/, /r/, or /ʋ/. All four factors were fully crossed, resulting in 16 conditions. Each condition contained four items. Table 1 shows the full design and an example of each combination of factors. The full set of materials is listed in Appendix A.

In addition to the 64 test items there were 456 fillers, resulting in a list of 520 items. The group of fillers contained both mono- and bisyllabic words and nonwords. The fillers were added to prevent the participants' attention from being drawn to the stimuli starting with voiced plosives. Some of the filler items served as test items for Experiment 2. Approximately one-third of the items on the list started with a voiced plosive.

2.1.2. Participants

Participants were students from the Max Planck Institute subject pool. There were five male and five female speakers. All of them were native speakers of Dutch and fluent readers. They were paid for their participation.

2.1.3. Recordings

Participants were seated in a sound-proof booth and were asked to read the items on the list out loud in front of a microphone, which was placed approximately 30 cm from the mouth. The items

Table 1
Full design of Experiment 1

Place of articulation	Lexical status	Following phoneme	Competitor environment	Item (competitor)	
Labial	Nonword	Vowel	No competitor	baag —	
			Competitor	bijn —	(pijn) (*pain*)
		Consonant	No competitor	bleep —	
			Competitor	bluim —	(pluim) (*feather*)
	Word	Vowel	No competitor	biels *sleeper*	
			Competitor	boot *boat*	(poot) (*paw*)
		Consonant	No competitor	brood *bread*	
			Competitor	bril *glasses*	(pril) (*young*)
Alveolar	Nonword	Vowel	No competitor	daaf —	
			Competitor	daart —	(taart) (*pie*)
		Consonant	No competitor	dwomp —	
			Competitor	draan —	(traan) (*tear*)
	Word	Vowel	No competitor	deur *door*	
			Competitor	duin *dune*	(tuin) (*garden*)
		Consonant	No competitor	dwars *diagonally*	
			Competitor	drol *turd*	(trol) (*troll*)

Each combination of factors contained four items.

were presented without any context and participants were instructed to read the items one by one, separated by a pause, in a clear and natural way. If they made a mistake they could read the word again. Recordings were made onto digital audio tape (sampling rate of 48 kHz with 16-bit resolution). After applying an anti-alias filter, the utterances were redigitized at a sample rate of 16 kHz.

2.1.4. Measurements

For each token the duration of prevoicing was measured. The beginning of the prevoicing was defined as the point in time at which evidence of vocal fold vibration could be detected. Any clearly visually detectable period, no matter how small in amplitude, was accepted as part of

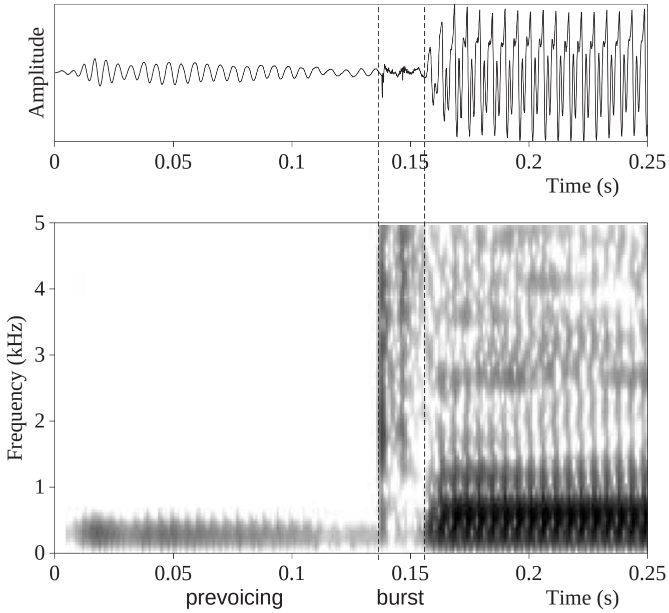

Fig. 1. Waveform and spectrogram of the first phoneme and part of the second phoneme of the Dutch word /bo:t/ (*boat*). The dashed lines indicate the end of the interval of prevoicing and the onset and offset of the burst.

voicing. The end of the prevoicing was defined as the point in time at which the noise of the release burst started, visible as a sudden peak in the waveform. Only when it was not completely clear where the prevoicing or the plosive release started, a wide-band spectrogram was used to locate the point in time where there was a sudden presence of aperiodic wide-band energy. Fig. 1 shows an example of an utterance starting with prevoicing. We found three different prevoicing patterns: no prevoicing; voicing interrupted by the plosive release; or voicing continued during the release, in which case the release was visible as a short-term turbulent structure of low energy on top of the voicing pulses. Below we do not distinguish between the two latter patterns.

2.2. *Results and discussion*

Fig. 2 shows the percentage of prevoiced tokens per speaker. There was considerable variation between subjects: some speakers produced prevoicing at the beginning of each voiced plosive, while other speakers only did so for some of the items. One speaker produced only 38% of the items with prevoicing. Overall 75% of the tokens were produced with prevoicing.

First, the influence of the speakers' sex on the proportion of prevoiced tokens (see Fig. 2) and on the duration of prevoicing was examined. To investigate the influence of this factor on the proportion of prevoiced tokens, a logistic regression (LR) analysis was performed with prevoicing (present or absent) as the dependent variable and sex (male or female) as factor. The LR model with sex as factor plus a constant yielded a deviance of $G^2 = 675$ (residual d.f. = 635), which was a significant improvement over the model consisting of only a constant ($G^2 = 711$; residual

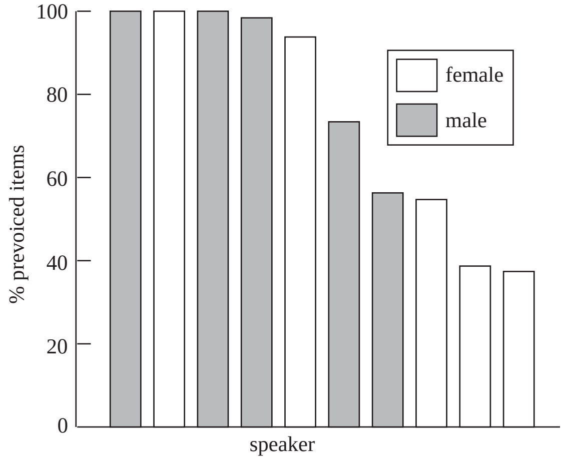

Fig. 2. Experiment 1: percentage of prevoiced items plotted separately for each speaker in rank order.

Table 2
Experiment 1 percentage of prevoiced tokens for all 10 speakers and mean prevoicing duration (in ms) of the prevoiced tokens of the five most frequent prevoicers

Factor	Level of factor	% Prevoiced tokens	Prevoicing duration	(S.D.)
Place of articulation	Labial	78.9	112.9	(32.2)
	Alveolar	71.8	104.1	(23.0)
Following phoneme	Vowel	85.5	117.5	(29.2)
	Consonant	65.3	99.5	(24.2)
Lexical status	Nonword	76.8	109.7	(26.4)
	Word	73.9	107.3	(30.1)
Competitor environment	No competitor	73.9	105.5	(28.0)
	Competitor	76.8	111.5	(28.3)

d.f. = 636). The coefficient for the sex of the speaker was significantly different from zero ($B = -1.2$, $p < 0.0001$). As predicted, male speakers produced more tokens with prevoicing than female speakers did (86% versus 65%, respectively). To investigate the influence of sex on prevoicing duration a one-way analysis of variance (ANOVA) on the prevoicing duration of only the prevoiced tokens was performed. The difference in prevoicing duration between males (109 ms) and females (89 ms) was not significant.

Second, the influence of the four factors (place of articulation, lexical status, following phoneme and competitor environment) on the proportion of prevoicing and the prevoicing duration was examined. The mean percentages of prevoiced tokens calculated separately for each of the four factors are shown in column 3 of Table 2. As before, a logistic regression analysis with prevoicing (present or absent) as the dependent variable was performed. This time there were four independent variables: place of articulation (labial or alveolar), lexical status (word or nonword), following phoneme (vowel or consonant) and competitor environment (no competitor or competitor). The LR model with these four factors plus constant ($G^2 = 669$; residual d.f. = 632) was significantly better than the model with only a constant ($G^2 = 711$; residual d.f. = 636). Of the four factors, only two were significant. These were place of articulation ($B = 0.21$, $p < 0.05$) and following phoneme ($B = -0.58$, $p < 0.0001$). Labial plosives were more often produced with prevoicing than alveolar plosives, and plosives followed by a vowel were more often produced with prevoicing than plosives followed by a consonant. The two lexical factors (lexical status and the competitor environment of the carrier stimulus) did not have a significant effect on the presence or absence of prevoicing.

Subsequently, we focused on the tokens with prevoicing to find out which of the four factors had an influence on the duration of prevoicing of these tokens. Since some speakers produced too few tokens with prevoicing to conduct a four-way repeated measures analysis of variance, we selected the five strongest prevoicers, that is, the speakers who produced more than 90% of the items with prevoicing. Only tokens produced with prevoicing were included in the analysis. Column 4 of Table 2 shows the mean duration of the prevoicing for the four factors separately, collapsed over these five frequent prevoicers. Following phoneme was the only factor showing a significant main effect: $F1(1,4) = 63.6$, $p < 0.001$; $F2(1,4) = 24.8$, $p < 0.001$. The duration of the prevoicing was longer for plosives followed by a vowel (118 ms) than for plosives followed by a consonant (99 ms). The effect of the place of articulation was significant in the items analysis: $F2(1,4) = 6.01$, $p < 0.05$, but did not reach significance in the subjects analysis: $F1(1,4) = 4.95$, $p = 0.09$. There were no significant effects of the lexical factors. There was, however, a significant three-way interaction of following phoneme, lexical status and word competitor: $F1(1,4) = 16.2$, $p < 0.05$; $F2(1,4) = 7.06$, $p < 0.05$, but a post hoc Tukey honestly significance test showed that there were no significant pairwise differences in the items analysis.

In summary, we found much variation in prevoicing of initial Dutch voiced plosives among speakers. Overall, about 75% of the tokens were prevoiced. Some speakers always produced prevoicing, but others did so in less than half of the cases. Female speakers produced fewer tokens with prevoicing than male speakers. There was however no sex difference in prevoicing duration. Of the four explored factors, only the two segmental factors had a significant effect on the percentage of prevoiced tokens. When the place of articulation of the initial voiced plosive was labial, tokens were more often prevoiced than when the place of articulation was alveolar. Furthermore, prevoicing was omitted more often when the following phoneme was a consonant than when it was a vowel. The following phoneme also had an effect on the duration of the prevoicing: the duration was shorter for plosives followed by a consonant than for plosives followed by a vowel. There was no effect of either the lexical status of the stimulus or the lexical competitor environment on prevoicing production.

Our study confirms the finding by Lisker and Abramson (1964) that some speakers realize all voiced plosives with prevoicing. It also shows, however, that other speakers do not always produce prevoicing. Overall, 25% of all voiced plosives were produced without prevoicing.

We predicted that both the sex of the speaker and the place of articulation of the plosive would affect prevoicing production. These predictions were confirmed. Male speakers produced prevoicing more often than females did and labial plosives were more often produced with prevoicing than alveolar plosives. The effect of the sex of the speaker is probably due to differences in the size of the vocal tract between men and women. Men tend to have larger vocal tracts than women (e.g., Stevens, 1998) and therefore the supraglottal pressure rises less quickly in the former. This makes it easier to produce prevoicing. The difference in the duration of prevoicing which was present was, however, not significant.

The effect of the place of articulation on the occurrence of prevoicing can be explained by differences in the size of the surface of the vocal tract walls which can participate in the passive expansion. Since labial plosives are produced more anteriorly than alveolar plosives, the surface of tissue which can be pushed outward as a result of the raised oral pressure is larger for labials than for alveolars. In line with this, labials were more often produced with prevoicing than alveolars. There was no effect on the duration of the prevoicing.

In addition to these two predicted effects there was also an effect of the following phoneme; prevoicing was more often produced and, if present, longer when the plosive was followed by a vowel than when it was followed by a consonant. The group of following consonants consisted of three different phonemes: /r/ after /b/ and /d/, /l/ after /b/, and /ʋ/ after /d/. Although in some cases, the following consonant may result in a smaller size of the oral cavity, for example when the plosive /b/ is followed by /l/ in comparison to when it is followed by an /a:/, this explanation would not hold for all consonants and vowels which were used in this study. Furthermore, when we studied the occurrence and duration of prevoicing in the vowels separately, no effect of vowel height was found. Based on the above, it is very unlikely that the observed difference in prevoicing between plosives followed by a vowel and plosives followed by a consonant is only caused by a difference in the volume of the oral cavity. We therefore propose that the degree to which the vocal tract can be expanded (passively or actively) plays a role in these findings. Articulatory measurements should be obtained in order to find a detailed explanation for this effect.

Finally, the two lexical factors appeared to have no influence on prevoicing production. The finding that nonwords and words did not differ in the production of prevoicing show that nonwords are not hyper-articulated in the sense of more reliable prevoicing. Furthermore, the absence of an effect of the competitor environment indicates that it is not the case that listeners articulate more carefully to avoid activation of a voiceless word competitor.

Given that a quarter of the voiced plosives were produced without prevoicing, the question emerges whether these plosives are still perceived as voiced. Is the production of prevoicing essential for the plosives to be perceived as voiced, or are other acoustic cues present and strong enough to evoke a voiced percept? To answer these questions, first a detailed acoustic analysis of the productions of voiced and voiceless plosives was conducted in Experiment 2. Several potential acoustic cues were measured and analyzed. A classification tree analysis was performed to investigate which of the measured cues would be the most reliable for categorization of the voiced–voiceless distinction. Experiment 3 was designed to find out whether listeners identified the produced tokens as voiced or voiceless and which of the measured cues influenced identification most strongly.

3. Experiment 2

Based on the study by Slis and Cohen (1969), and the information on the voicing distinction in other languages (mainly English), the following six measurements were selected for the purpose of the present study: duration of prevoicing, duration of the burst, power of the burst, spectral center of gravity (SCG) of the burst, F_0 immediately after burst offset, and F_0 movement into the vowel.

3.1. Method

3.1.1. Materials

From the complete collection of 520 tokens produced by 10 different speakers (Experiment 1) 48 item pairs were selected, which partly overlapped with the token set used in Experiment 1. These pairs differed only in the voicing of the initial plosive, in order to obtain the same variation in segmental context in both groups (voiced and voiceless). Note, however, that these items were not produced as pairs, but as single items in random order among many fillers. Half of the pairs started with labial plosives (/b/ or /p/) and the other half started with alveolar plosives (/d/ or /t/). Half of the pairs started with a consonant cluster (/b/ or /p/ followed by an /r/ or /l/, and /d/ or /t/ followed by /r/ or /ʋ/) and the other half of the pairs started with a plosive followed by a vowel. One-third of the pairs were nonword–word pairs, i.e., the voiced counterpart of the pair was a nonword and the voiceless counterpart a word, for example, bluim–PLUIM (plume); one-third of the pairs were word–nonword pairs, for example, BRAAM (blackberry)–praam; one-third of the pairs were word–word pairs, for example, BAARS (perch)–PAARS (purple). The complete set of items is given in Appendix B.

3.1.2. Measurement procedures

For each item produced by each speaker the six cues that were expected to signal the voiced–voiceless distinction were measured. Below, first the relevant references in Dutch are given, followed by a description of how the measurements were performed for each cue:

1. *Duration of prevoicing*: As already mentioned, Lisker and Abramson (1964) found that all tokens of voiced Dutch plosives produced by one speaker were prevoiced. Experiment 1, however, showed that only 75% of the tokens starting with a voiced plosive were produced with prevoicing. In a perceptual experiment using synthetic CV stimuli that varied only in VOT, Slis and Cohen (1969) found that voiced judgements correlated with a voice lead and voiceless ones with a voice lag. The methods of measuring the duration of the prevoicing were the same as in Experiment 1 (see method of Experiment 1). No prevoicing was expected in the productions of the voiceless plosives.

2. *Duration of the burst*: Slis and Cohen (1969) reported that the noise burst duration of Dutch plosives was on average 15 ms shorter for voiced plosives than for voiceless plosives. The difference in burst duration may be explained by the spatially more extended contact at constriction for voiceless plosives in comparison to voiced plosives (e.g., Cho & Ladefoged, 1999; Yoshioka, Murase, & Uematsu, 1996). Ernestus (2000) measured burst durations of 649 Dutch plosives in medial position and showed that expert listeners tended to classify plosives with short bursts durations as voiced and plosives with long burst durations as voiceless.

The onset of the burst was defined as the point in time at which the closure was released (see Experiment 1). The definition for the offset of the burst varied with the following phoneme. When the following phoneme was a vowel or an /l/, the offset of the burst was defined as the point at which higher formants were first visible in the spectrogram (see Fig. 1 for an example). When the following consonant was a /ʋ/ the offset of the burst was defined at the point where the

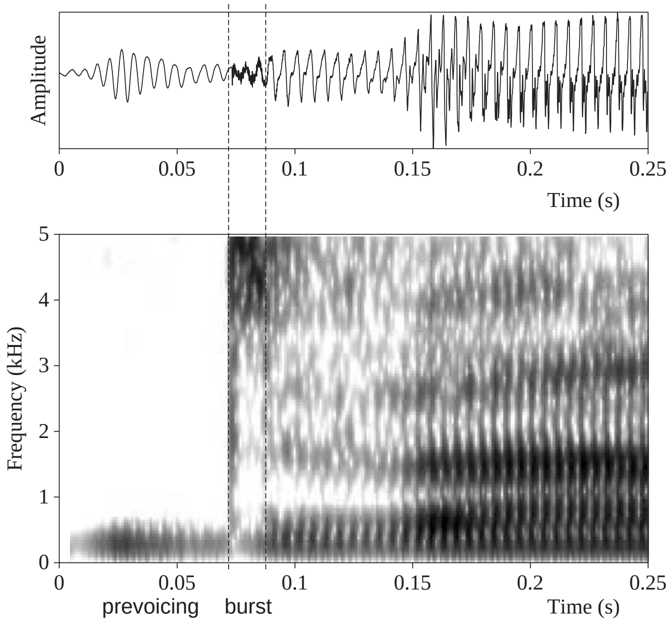

Fig. 3. Waveform and spectrogram of the first two phoneme and part of the vowel of the Dutch nonword /dʋaːlf/. The dashed lines indicate the end of the interval of prevoicing and the onset and offset of the burst.

spectrogram showed a sudden change in spectral composition of the noise (see Fig. 3). When the plosive was followed by an /r/ the labeling of the end of the burst depended on the way the /r/ was produced. In the cases where it was produced as an retroflex approximant [ɹ] , the onset of higher formants served as an indication of the offset of the burst. In the cases were it was produced as an uvular trill [R] or alveolar trill [r] or as an uvular fricative [ʁ], the change in the structure of the noise served as indication for the burst offset (see Fig. 4). In many of these latter cases, the trill or frication was preceded by a short schwa. The moment at which the higher formants of the schwa were visible in the spectrogram were then taken as the offset of the burst.

The burst included the following two acoustic events, as described by Stevens (1993): a brief transient as the air that has been compressed in the vocal tract discharges through the opening constriction, followed by frication noise, which is caused by rapid airflow through the constriction. Dutch voiceless plosives have little or no aspiration. In the cases where there was aspiration, we included the aspiration in the burst. The duration of the burst was expected to be longer for voiceless plosives than for voiced plosives. Note that for voiceless plosives, the duration of the burst reflects the positive VOT.

3. *Power of the burst above 500 Hz*: Slis and Cohen (1969) found that the amplitude of the voiceless noise burst was about 50% higher than the amplitude of the voiced noise burst. Possible causes for this difference mentioned in the literature are higher oral pressure behind the constriction and/or spatially more extended closure for voiceless plosives (e.g., Yoshioka et al., 1996).

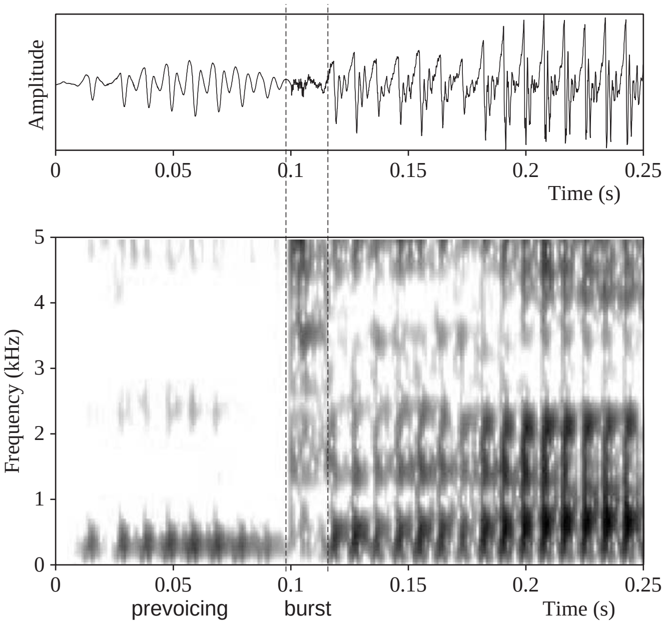

Fig. 4. Waveform and spectrogram of the first two phonemes and part of the vowel of the Dutch word /drɑp/ (*dregs*). The dashed lines indicate the end of the interval of prevoicing and the onset and offset of the burst.

The burst (as described under 2) was first high-pass filtered at a cutoff frequency of 500 Hz. Then the spectral power was calculated by taking the logarithm of the mean sum of squares of all sample points. Energy under 500 Hz was filtered out to exclude the energy generated by any vocal fold vibration during or immediately after the release of the closure. The spectral power was expected to be higher for bursts of voiceless plosives than for voiced plosives.

4. *SCG of the burst*: The SCG, or first spectral moment, has been used to describe the difference in place of articulation of fricatives and plosives (e.g., Forrest, Weismer, Milenkovic, & Dougall, 1988), since it is used as an acoustic measure for the size of the front cavity; the smaller the front cavity, the higher the SCG. We propose that the SCG is also an appropriate measure for the difference between voiced and voiceless plosives for several reasons. First, we predict that the SCG is influenced by the presence of voicing in the burst. When there is voicing, there is more energy in the lower frequencies which will shift the gravity to lower frequencies compared to when there is no voicing. Therefore, we expected the SCG to be lower in voiced plosives than in voiceless plosives. Second, Cho, Jun, and Ladefoged (2002) remark in a study of alveolar fricatives that the SCG might also reflect the velocity of the jet of air; a higher subglottal pressure results in a jet of air with a higher velocity, which will result in a higher SCG. Voiceless plosives are expected to be produced with a higher velocity of the air jet than voiced plosives, which would result in higher SCG for the voiceless plosives than for the voiced plosives. Finally, our intuition (as native speakers of Dutch) is that the place of articulation of /d/ is slightly different from that of /t/. The voiceless counterpart seems to be produced more frontally than the voiced one. This suggests that the front cavity is smaller for /t/ than for /d/, resulting in higher SCG for the voiceless alveolar plosive. If this is indeed true, we should find that the difference in SCG between

voiced and voiceless plosives is larger in the case of alveolar plosives than in the case of labial plosives.

To calculate the SCG, the burst was first filtered into 32 frequency bins with widths of 250 Hz, except for the first bin which was high-pass filtered at a cutoff of 50 Hz (to remove any spurious low-frequency components) resulting in a range from 50 to 250 Hz. For each filter the power of the filtered signal was calculated. Next, the 32 center frequencies were multiplied by the corresponding powers, summed together and divided by the sum of the powers, resulting in the SCG.

5. *Absolute* F_0 *and* F_0 *difference*: Many studies have reported a higher fundamental frequency (F_0) of the vowel adjacent to a voiceless plosive than of the vowel adjacent to a voiced plosive in English (House & Fairbanks, 1953; Kingston & Diehl, 1994; Lehiste & Peterson, 1961; Mohr, 1971; Löfqvist, 1978; Umeda, 1981). For Dutch, Slis and Cohen (1969) reported a difference of 6 Hz between the maximum frequency in the vowel after voiceless consonants and the maximum frequency after voiced consonants. A possible cause for the difference in F_0 is the lowering of the larynx during the production of voiced plosives in order to obtain sufficient transglottal pressure to produce vocal fold vibration. Lowering of the larynx can cause a downward tilt of the cricoid cartilage, which causes a shortening and hence slackening and thickening of the vocal folds (Honda, Hirai, & Kusakawa, 1993), resulting in a lower F_0. This explanation suggests that the F_0 pattern is directly related to the moment at which the vocal fold vibration starts. Ohde (1984) reported, however, that in English a high F_0 was found for both voiceless aspirated (in initial position) and unaspirated plosives (after a /s/ in initial position), although the VOTs of those groups were very different. These data suggest that F_0 differences are a product of articulations that are controlled independent of the timing of the glottal articulations to produce voicing (see also Kingston & Diehl, 1994). In sum, the relationship between voicing and F_0 patterns remain controversial.

Haggard, Ambler, and Callow (1970) demonstrated that stimuli were consistently perceived as/b/ when synthesized with a low-rising F_0 contour, but as /p/ with a high-falling contour. Further perception data by Haggard, Summerfield, and Roberts (1981) suggest that the actual cue is the onset frequency rather than the F_0 movement into the vowel. Many other studies have examined the influence of F_0 differences on the perception of the voicing distinction in plosives (e.g., Ohde, 1984; Kohler, 1985; Whalen, Abramson, Lisker, & Mody, 1993), but the underlying perceptual mechanisms remain largely unknown. Both the absolute F_0 value immediately after the plosive and the F_0 movement into the vowel were therefore included as potential cues in the present study.

F_0 was estimated for each frame of 10 ms of the vowel (or consonant plus vowel) that followed the initial plosive, using an algorithm called RAPT (Talkin, 1995), which estimates the fundamental frequency from the normalized cross-correlation function using dynamic programming. Subsequently, the mean F_0 was calculated for the first two voiced frames, resulting in a measure for the absolute F_0 immediately after burst offset. This absolute F_0 was expected to be lower for voiced plosives than for voiceless plosives.

To obtain a measure of F_0 change, the F_0 immediately after burst offset (see above) was subtracted from the F_0 in the middle of the vowel in cases where the plosive was followed by a vowel or from the F_0 in the middle of the consonant plus the following vowel in the cases where the plosive was followed by a consonant. Thus, a positive F_0 difference corresponded to a rising

F_0 pattern and a negative F_0 difference corresponded to a falling F_0 pattern. The F_0 in the middle was defined as the mean F_0 for the middle two or three frames (depending on whether the total number of frames was even or odd). The two or three middle frames were not allowed to overlap with the two frames used to estimate the F_0 immediately after burst offset. Note that, therefore the vowel (or second consonant plus vowel) was required to have a duration of at least 60 ms (six frames). For the tokens which did not meet this constraint, no F_0 difference was calculated.

3.2. Results and discussion

The measures in the current experiment were selected in order to describe the voiced–voiceless distinction. However, on the basis of previous literature, we expected that most of these measures would also vary with place of articulation. Therefore, in addition to the voicing category of the plosive, the place of articulation of the plosive was included in the data analyses. In some cases not all measurements could be obtained, for example, when the recording was affected by any external noise or when the vowel was too short to calculate the F_0 difference, or when it was not clear where a particular segment started or ended. Overall, 12 labials tokens and 18 alveolar tokens had to be excluded from further analyses.

The distributions of each measure are plotted separately for /b/ versus /p/ and /d/ versus /t/ in Figs. 5–10. Table 3 shows the mean values for each measure separately for the four plosives. To find out whether these measures were good correlates of the voicing distinction, a multivariate analysis of variance was conducted with the six measures as dependent variables and voicing category (voiced versus voiceless) and place of articulation (labial versus alveolar) as factors. Table 4 indicates the significant main effects and interactions. These effects will be discussed for each measure separately.

The duration of prevoicing was longer for voiced than for voiceless items. In fact, as expected, none of the voiceless items were produced with prevoicing. Therefore, only the histograms for the

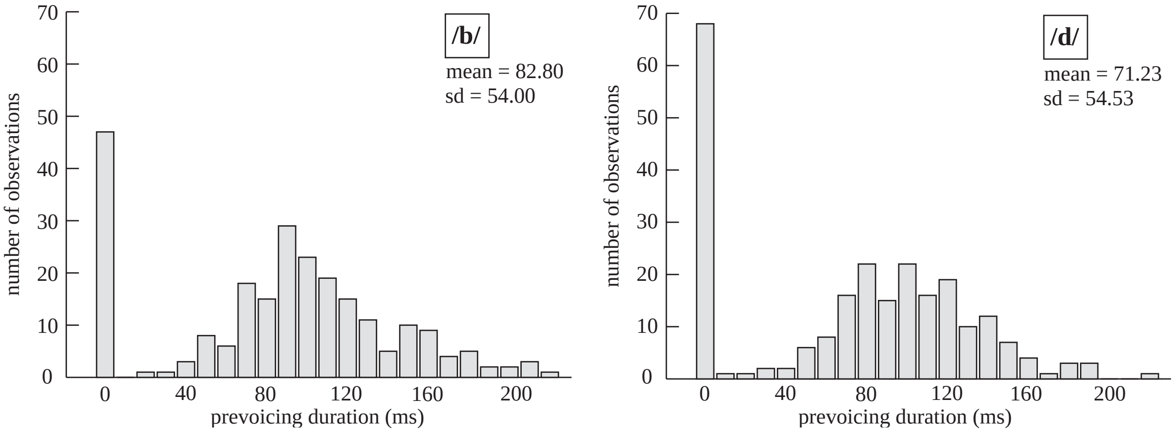

Fig. 5. Experiment 2: histograms of the prevoicing as produced by 10 different speakers, plotted separately for place of articulation (left versus right). Only the voiced categories are plotted; all tokens of the voiceless category were produced without prevoicing. The numbers on the *x*-axis represent the upper limits of each bin.

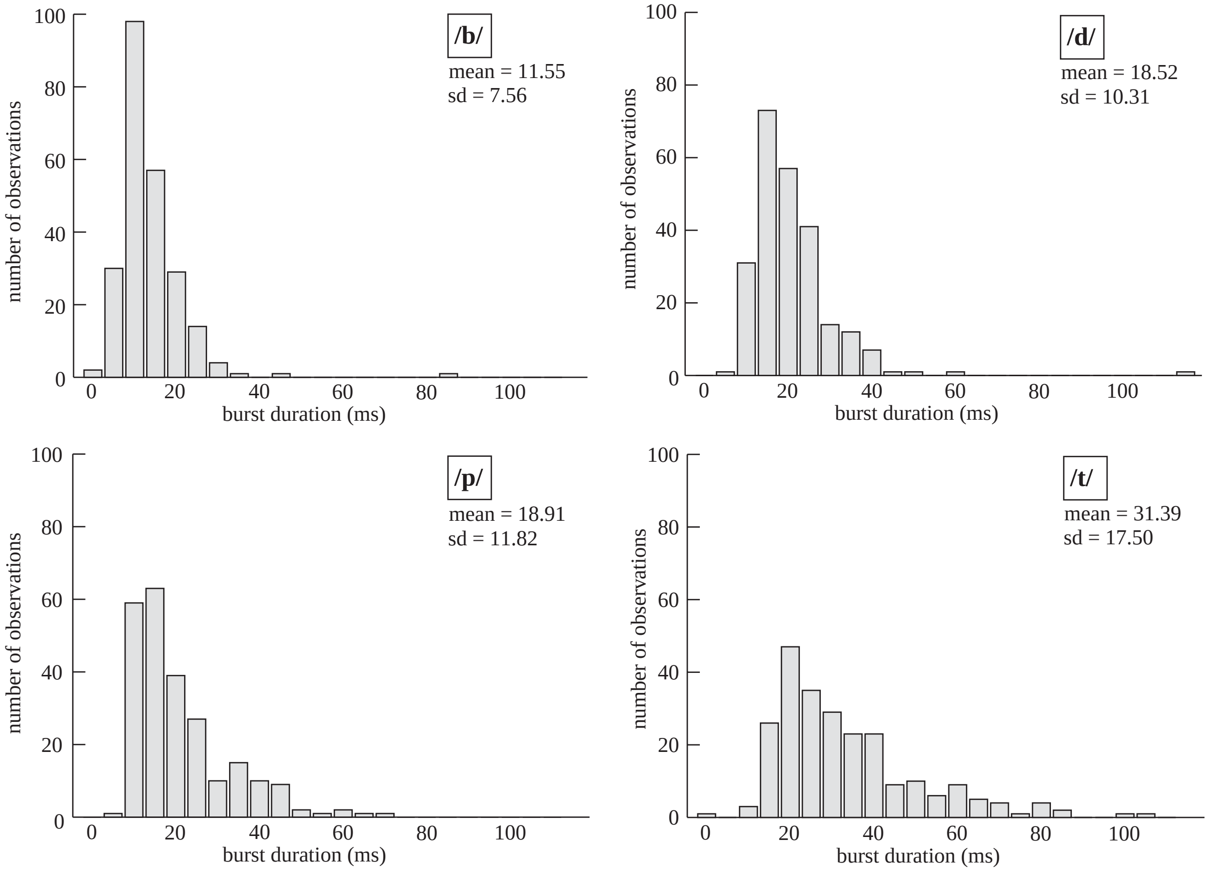

Fig. 6. Experiment 2: histograms of the burst durations as produced by 10 different speakers, plotted separately for place of articulation (left versus right) and voicing category (top versus bottom). The numbers on the *x*-axis represent the upper limits of each bin.

voiced plosives were plotted (see Fig. 5). Overall, 76% of the voiced tokens were produced with prevoicing.

The duration of the burst was longer for voiceless plosives than for voiced plosives (Fig. 6). The difference between the burst duration of voiced and voiceless plosives was approximately 10 ms. For the voiceless plosives the duration of the burst reflects the positive VOT. The mean VOT values for the voiceless plosives were almost twice as high as the VOT values for Dutch reported by Lisker and Abramson (1964), but were similar to the VOT values reported by Flege and Eefting (1987). Place of articulation also had an effect on burst duration: bursts were longer for alveolars than for labials. This is in line with findings in the literature for Dutch initial plosives (Smits, 1995).

The power of the burst above 500 Hz was higher for voiceless plosives than for voiced plosives (Fig. 7). The difference between voiced and voiceless plosives was about 3 dB. The power of the burst was also higher for alveolar plosives than for labial plosives.

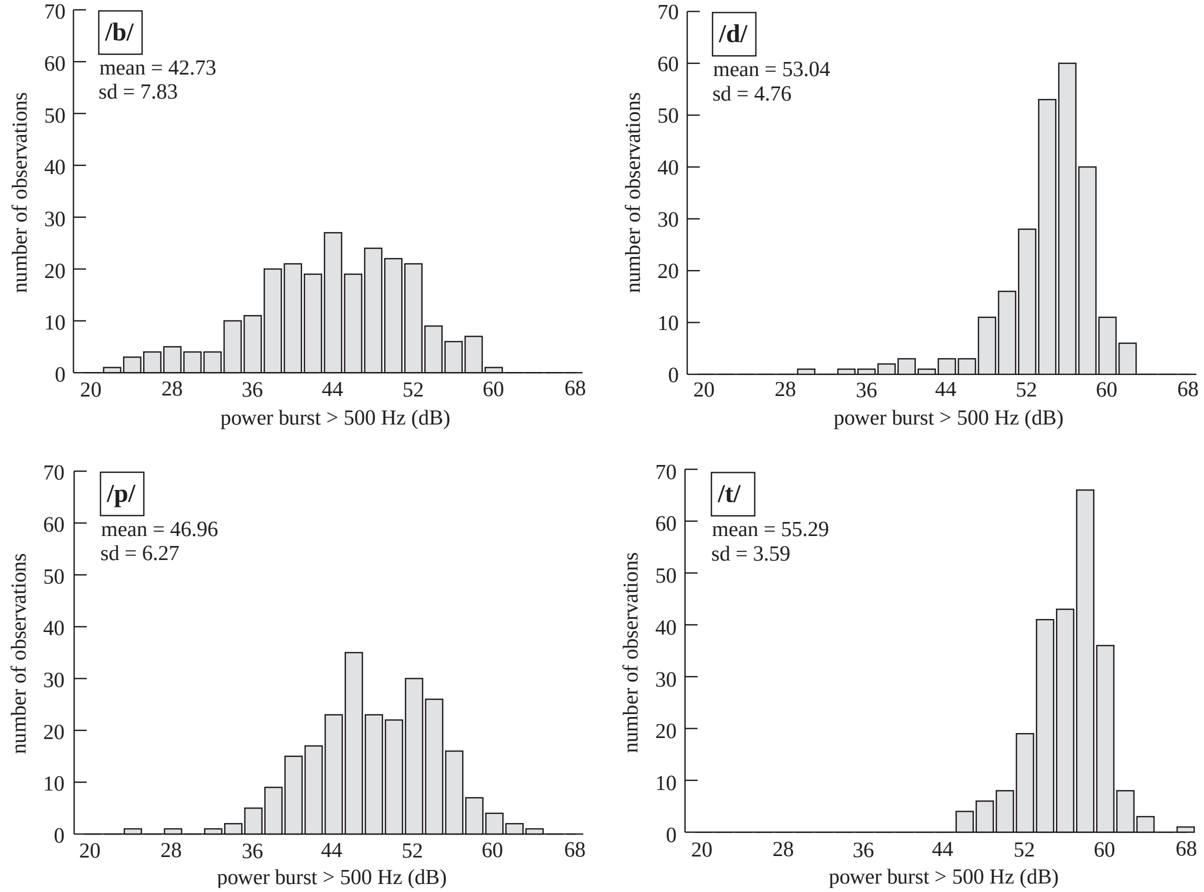

Fig. 7. Experiment 2: histograms of the power in the burst as produced by 10 different speakers, plotted separately for place of articulation (left versus right) and voicing category (top versus bottom). The numbers on the *x*-axis represent the upper limits of each bin.

Fig. 8 shows that the distributions of the SCG are different depending on the place of articulation. The mean SCG was higher for alveolars (2.8 kHz) than for labials (1.0 kHz), as has been described in the literature (e.g., Forrest et al., 1988). As predicted, the SCG appeared also to be influenced by the voicing category of the plosive: it was higher for voiceless plosives than for voiced plosives.

In addition to the two main effects, there was an interaction between voicing and place of articulation: the difference in the SCG between voiced and voiceless was considerably larger for alveolars (1.40 kHz) than for labials (0.32 kHz). At first sight, it may seem that this interaction can be explained by the fact that labial plosives are more often produced with prevoicing than alveolar plosives. One might predict that when more tokens are produced with prevoicing also more of the bursts would be voiced, which would result in lower SCG values for prevoiced tokens. However, the asymmetry in the presence of prevoicing in labials and alveolars would result in the opposite pattern, namely in a larger difference in SCG between voiced and voiceless labials than between voiced and voiceless alveolars. We also discussed how the SCG is related to the front cavity. The

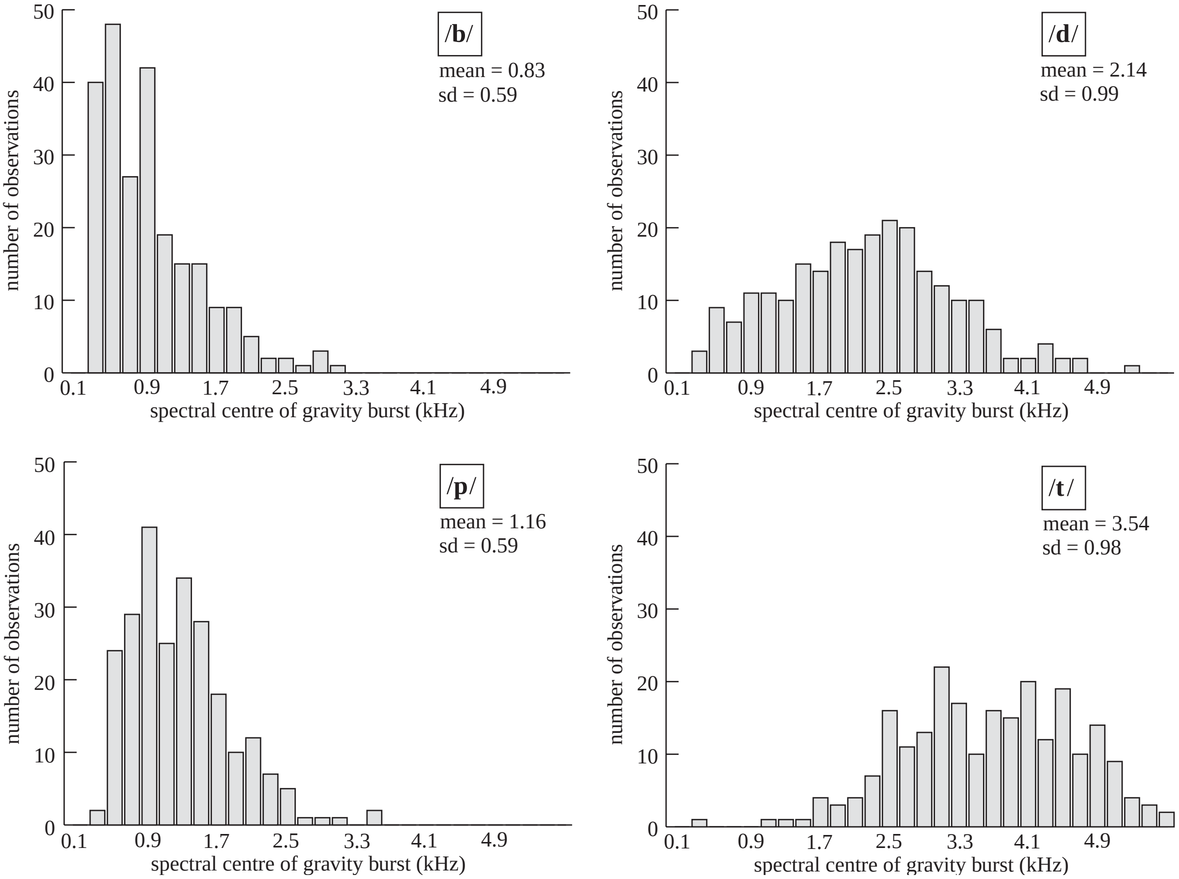

Fig. 8. Experiment 2: histograms of the SCG of the burst as produced by 10 different speakers, plotted separately for place of articulation (left versus right) and voicing category (top versus bottom). The numbers on the *x*-axis represent the upper limits of each bin.

present observed interaction is in agreement with our intuition that voiced alveolar plosives are produced slightly more posteriorly than the voiceless counterpart. The front cavity would be larger for /d/ than for /t/, resulting in a larger SCG difference between voiced and voiceless plosives in the case of alveolar plosives than in the case of labial plosives. Of course, this hypothesis will have to be correlated with independent articulatory evidence.

One possibility is that this difference in the place of articulation, and thus the size of the front cavity, is a by-product of the downwards displacement of the tongue body for the production of a voiced plosive (Svirsky et al., 1997). This downward displacement will enlarge the vocal tract, such that the air pressure rises less quickly and makes it therefore easier to produce prevoicing (Westbury, 1983). The downward displacement of the tongue body would also take place during the production of labial voiced plosives. In the cases of labials, however, displacement of the tongue does not affect the place of articulation, which is realized with the lips. Therefore, the size of the oral cavity will differ between voiced and voiceless labial plosives, but not the size of the

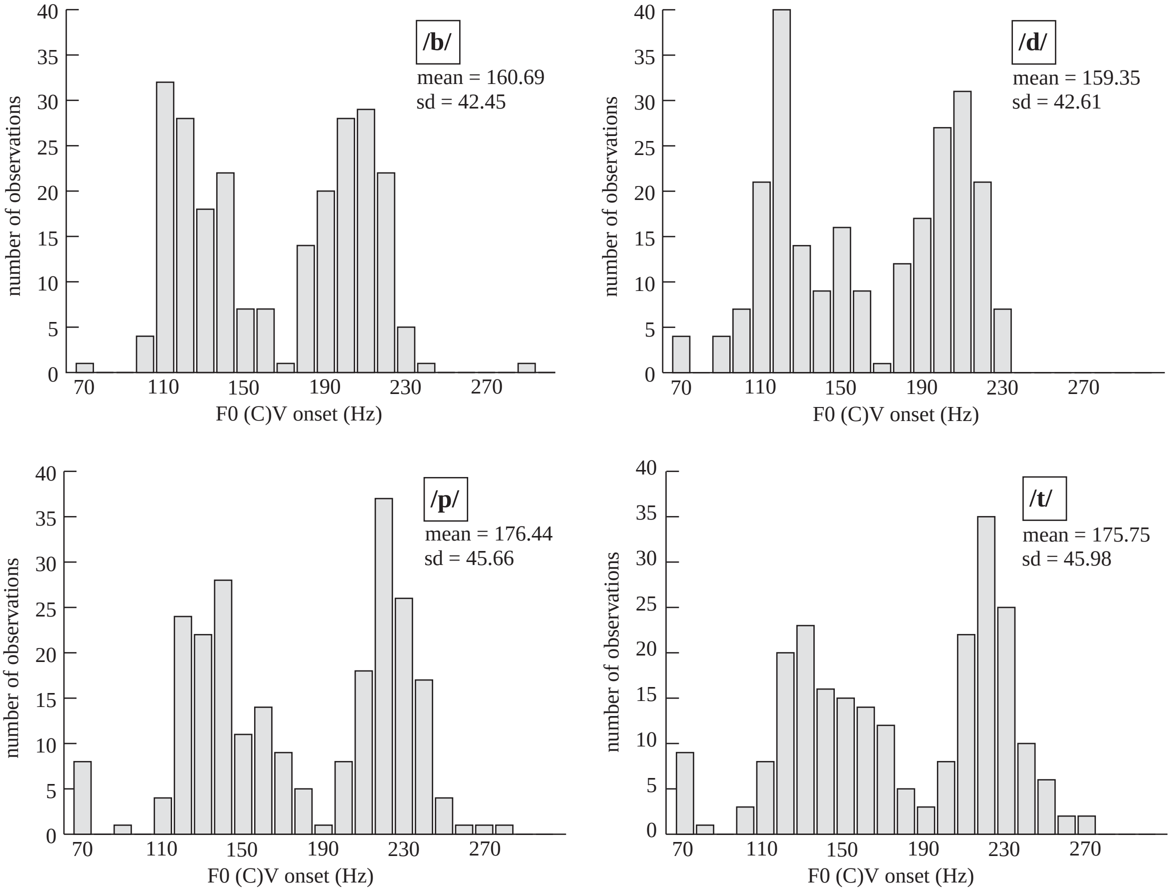

Fig. 9. Experiment 2: histograms of the F_0 at $(C)V$ onset as produced by 10 different speakers, plotted separately for place of articulation (left versus right) and voicing category (top versus bottom). The numbers on the x-axis represent the upper limits of each bin.

frontal cavity. One of the reviewers (Keating) suggested another possible alternative explanation, namely that the difference in place of articulation may be the product of a planned enhancing strategy. Backing the place of articulation for the /d/ will lower the SCG, thus mimicking the effect of prevoicing.

Fig. 9 shows that the distribution of absolute F_0 immediately after burst offset is bimodal. This is caused by the difference in F_0 between male and female speakers. Although the difference between the mean F_0 of voiced and voiceless plosives was in the same range of 10–15% found by Ohde (1984) and by Kingston and Diehl (1994), this difference was not significant. This may have been caused by the large standard deviations due to the bimodality.

The mean F_0 difference (Fig. 10) was positive for tokens starting with voiced plosives, consistent with a rising F_0, while it was negative for tokens starting with a voiceless plosive, consistent with a falling F_0. The difference between these means was significant. There also was an

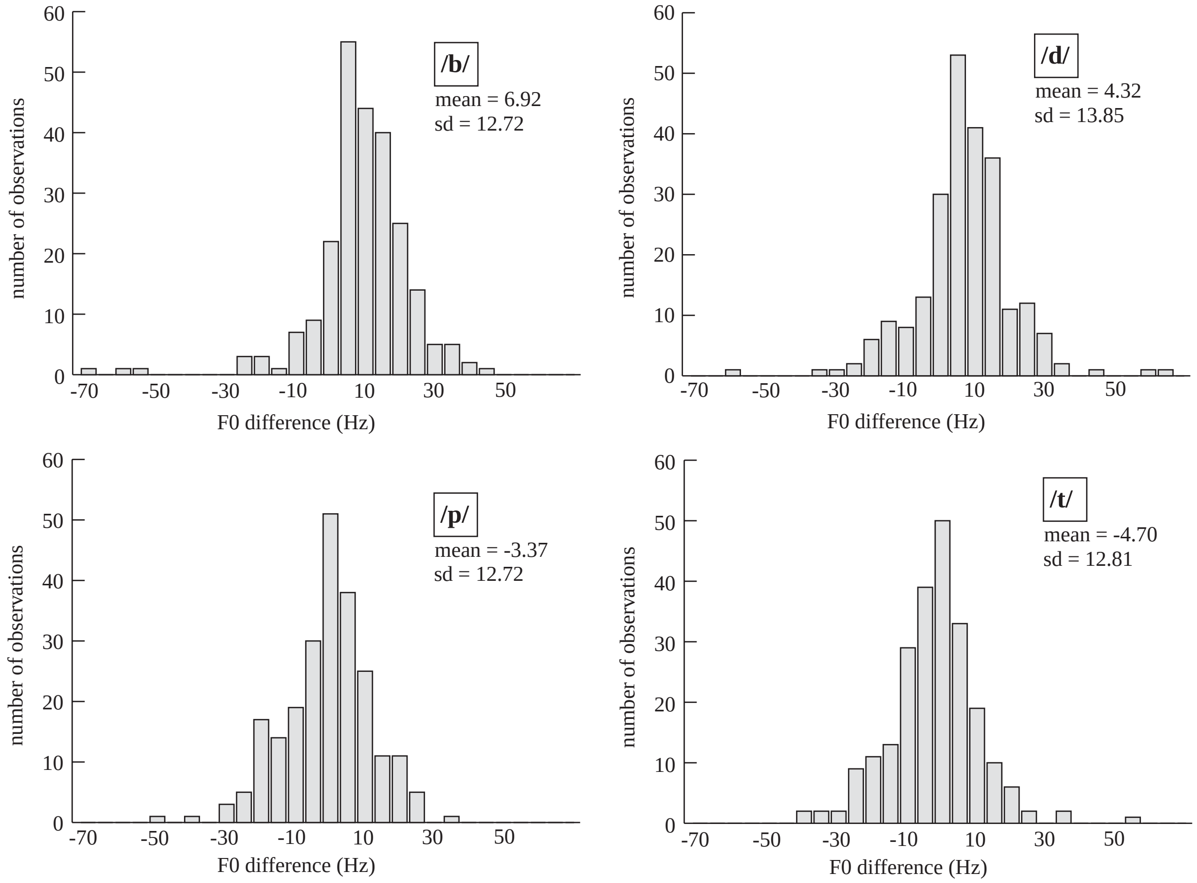

Fig. 10. Experiment 2: histograms of the F_0 difference as produced by 10 different speakers, plotted separately for place of articulation (left versus right) and voicing category (top versus bottom). The numbers on the *x*-axis represent the upper limits of each bin.

effect of place of articulation on the F_0 difference: the F_0 difference was larger for labials than for alveolars (1.84 versus −0.13).

Taken together, the results show that the means for all tested acoustic properties showed the predicted patterns and that all, except for the F_0 at the offset of the burst, differed significantly between voiced and voiceless plosives. Furthermore, most of the measures differed between labial and alveolar plosives. For the SCG, an interaction was found between voicing category and place of articulation. This suggests that there is a difference in the place of articulation between /d/ and /t/, but not between /b/ and /p/.

The acoustic analyses reveal that there are several acoustic properties which correlate with the voiced–voiceless distinction in Dutch plosives. The analyses do not show, however, which of these acoustic properties are most useful for correct recognition of the voicing feature. One would predict that listeners' phoneme identification would be influenced most by the cues which lead to the highest recognition scores. The obvious analyses to examine the relative strengths of the

Table 3
Experiment 2: mean values for each of the six measures separately for the two places of articulation and two voicing categories and the four combinations of these two factors

Measure	Place of articulation		Voicing		Mean
			Voiced	Voiceless	
Prevoicing duration (ms)	Labial		82.8 (54.0)	0.0 (0.0)	41.1
	Alveolar		71.2 (54.5)	0.0 (0.0)	35.6
		Mean	77.0	0.0	
Burst duration (ms)	Labial		11.6 (7.6)	18.9 (11.8)	15.3
	Alveolar		18.5 (10.3)	31.4 (17.5)	24.9
		Mean	15.1	25.1	
Power of burst (dB)	Labial		42.7 (7.8)	47.0 (6.3)	44.9
	Alveolar		53.0 (4.8)	55.3 (3.6)	54.2
		Mean	47.9	51.1	
SCG (kHz)	Labial		0.83 (0.59)	1.16 (0.59)	1.00
	Alveolar		2.14 (0.99)	3.54 (0.98)	2.83
		Mean	1.49	2.34	
F_0 at burst offset (Hz)	Labial		160.7 (42.5)	176.5 (45.7)	168.5
	Alveolar		159.4 (42.6)	175.8 (46.0)	167.5
		Mean	160.0	176.1	
F_0 difference (Hz)	Labial		6.9 (12.7)	−3.4 (12.7)	1.8
	Alveolar		4.3 (13.8)	−4.7 (12.8)	−0.1
		Mean	5.6	−4.0	

The number between parentheses indicate the sd.

various acoustic properties for recognition are linear discriminant analysis or logistic regression analysis. These analyses are inappropriate for our data set, however, since most of the predictor variables were highly skewed or multi-modal. Moreover, we wanted to add a categorical predictor, namely whether the following phoneme was a vowel or a consonant. The suitable analysis for this type of data is a classification and regression tree (CART) analysis (Breiman, Friedman, Olshen, & Stone, 1984). CART is a nonparametric analysis, i.e., no assumptions are made regarding the underlying distributions of the predictor variables. Furthermore, tree-based models are better in managing complex interactions that may exist in the data than the traditional methods.

Tree-based models operate by recursively partitioning a dataset in two (i.e., a binary split). Each split is based on the value of a single predictor variable. The choice of the predictor variable and its value for each split is based on an exhaustive search of all possible divisions of the data. The aim of each split is either to maximize the homogeneity of the groups in the case of nominal or ordinal responses (classification) or to best separate low and high values in the case of continuous response variables (regression). The algorithm continues splitting the subsets of data

Table 4
Experiment 2: significant effects of the multivariate ANOVAs (subject and item analyses) for the six measures

	Voicing	Place of articulation	Voicing × place of articulation
Prevoicing duration (ms)	$F1(1,36)=92.53$ $p<0.0001$ $F2(1,92)=623.48$ $p<0.0001$	$F1(1,36)=0.54$ Not significant $F2(1,92)=3.42$ Not significant	$F1(1,36)=0.54$ Not significant $F2(1,92)=3.42$ Not significant
Burst duration (ms)	$F1(1,36)=23.63$ $p<0.001$ $F2(1,92)=47.90$ $p<0.0001$	$F1(1,36)=21.83$ $p<0.0001$ $F2(1,92)=44.25$ $p<0.0001$	$F1(1,36)=1.72$ Not significant $F2(1,92)=3.57$ Not significant
Power of burst (dB)	$F1(1,36)=11.85$ $p<0.0001$ $F2(1,92)=63.20$ $p<0.001$	$F1(1,36)=97.57$ $p<0.0001$ $F2(1,92)=523.66$ $p<0.0001$	$F1(1,36)=1.08$ Not significant $F2(1,92)=5.95$ $p<0.05$
SCG (kHz)	$F1(1,36)=52.09$ $p<0.0001$ $F2(1,92)=155.91$ $p<0.0001$	$F1(1,36)=238.07$ $p<0.0001$ $F2(1,92)=713.77$ $p<0.0001$	$F1(1,36)=20.35$ $p<0.0001$ $F2(1,92)=60.28$ $p<0.0001$
F_0 at burst offset (Hz)	$F1(1,36)=1.42$ Not significant $F2(1,92)=158.71$ $p<0.0001$	$F1(1,36)=0.00$ Not significant $F2(1,92)=0.66$ Not significant	$F1(1,36)=0.00$ Not significant $F2(1,92)=0.08$ Not significant
F_0 difference (Hz)	$F1(1,36)=16.57$ $p<0.0001$ $F2(1,92)=110.38$ $p<0.0001$	$F1(1,36)=0.73$ Not significant $F2(1,92)=4.39$ $p<\ 0.05$	$F1(1,36)=0.06$ Not significant $F2(1,92)=0.52$ Not significant

(the nodes) until they are maximally homogeneous or contain too few observations. Finally, the constructed tree is pruned using cross-validation, that is, the tree is simplified without sacrificing goodness-of-fit.

In the CART analysis, we determined which of the acoustic correlates best predicted membership of the classes of voiced or voiceless plosives. The response variable that the analysis attempted to predict was the intention of the speaker to produce a voiced or voiceless plosive. Since the response variable was categorical, a classification tree analysis was conducted. All six acoustic cues were used as predictor variables. In addition to these numerical predictors, one categorical predictor was added, namely whether the plosive was followed by a vowel or consonant. Since the acoustic properties of voiced and voiceless plosives differed by place of articulation, separate classification tree analysis were conducted for labial and alveolar plosives. All tokens for which one or more of the measurements could not be made were excluded from the analyses (14 labials and 18 alveolars).

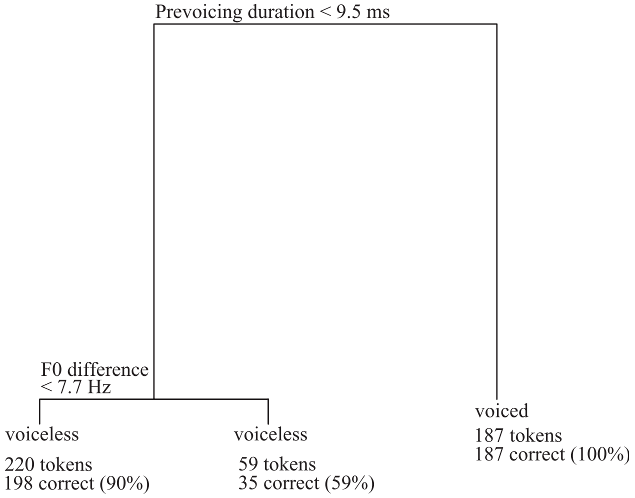

Fig. 11. Experiment 2: CART analysis of the labial productions of the speakers. Tokens satisfying the rule printed at the top of each split followed the left branch. Final nodes are labeled according to the plurality rule. Below each label the total number of tokens falling in that final node and the number of correctly classified tokens is given.

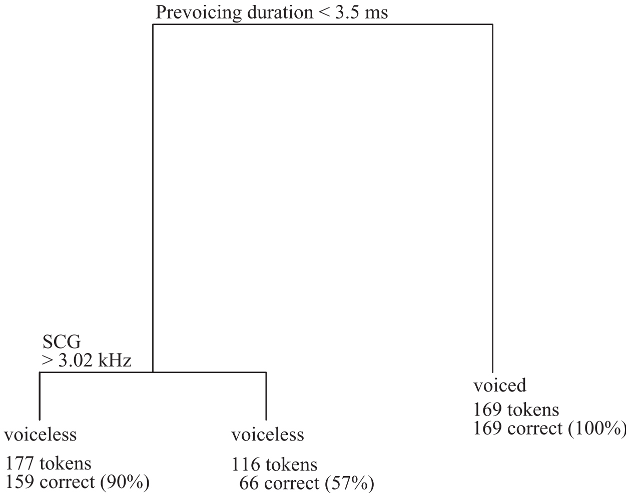

Fig. 12. Experiment 2: CART analysis of the alveolar productions of the speakers. Tokens satisfying the rule printed at the top of each split followed the left branch. Final nodes are labeled according to the plurality rule. Below each label the total number of tokens falling in that final node and the number of correctly classified tokens is given.

The resulting cost-complexity pruned classification trees are shown in Figs. 11 and 12. Each tree consists of a root node containing all tokens. This node is split based on a simple rule. Tokens satisfying the rule printed at the top of each split followed the left branch. The vertical length of each branch reflects the relevance of each factor, that is, the reduction in heterogeneity in each node. Each terminal node is labeled "voiced" or "voiceless" according to the plurality rule, i.e., according to the most represented class in that group of tokens. Below each label the number of tokens falling in that particular final node is given. The number of correctly classified tokens is printed underneath, which gives an indication of the goodness of fit of the tree.

The overall structure of the two trees is very similar. First the data is divided into two large groups on the basis of prevoicing duration, and then a small part of the data was further subdivided into two smaller groups. For the labial plosives, tokens produced with more than 9.5 ms of prevoicing followed the right branch and fell into a terminal node labeled "voiced". All tokens produced with more than 9.5 ms of prevoicing were intended by the speaker as being voiced plosives. The same holds for the alveolar plosives which were first split based on a prevoicing duration of 3.5 ms.[1] Since the cutoff values of 3.5 and 9.5 ms are barely enough to contain one period of voicing, we discuss the split according to whether the prevoicing duration was smaller or larger than 3.5 or 9.5 ms in terms of whether there was any prevoicing present or not.

The subset of labial productions without prevoicing was divided on the basis of the F_0 difference, while the subset of alveolar plosives was divided on the basis of the SCG. For the labials, 90% of the tokens without prevoicing and a small F_0 movement (<7.7 Hz) was intended as being voiceless. Although the majority of the labials without prevoicing and a larger F_0 movement (that is, a clear rising F_0 pattern) was still intended as being voiceless, this proportion was now only 59%. In other words, on the basis of the acoustic measures, the CART analysis essentially distinguishes three groups of plosives: clearly voiced ones, clearly voiceless ones, and an uncertain category of which a narrow majority is voiceless. For the alveolars, 90% of the tokens without prevoicing and a high SCG (>3.02 kHz) were intended as being voiceless, while only 57% of the tokens without prevoicing and a lower SCG were intended as being voiceless. Again, the CART analysis thus finds three categories, one of which is uncertain. Overall 90% of the labial plosives and 85% of the alveolar plosives were correctly classified by the CART analysis.

Although the overall structure of both trees is similar, the proportion of the first main split and the second smaller split is different for the two places of articulation. For the labials, the vertical length of the branches of the prevoicing split is very long and the branches of the second split very small. The same is true for the alveolars, but in comparison to the labials the branches of the prevoicing split are somewhat smaller and the branches of the second split are somewhat longer.

In summary, our analysis of the production data indicated that there are several acoustic correlates to the voiced–voiceless distinction in Dutch. The classification tree analysis showed that the duration of prevoicing is by far the most reliable predictor of voicing, for both labial and alveolar plosives. All the tokens produced with prevoicing were intended as being voiced. For labials, the F_0 movement was the second most reliable predictor of voicing, while for alveolars this was the SCG. The fact that the CART analyses selected different acoustic cues to split the labial plosives without prevoicing and the alveolar plosives without prevoicing strengthens our claim that the acoustic realization of the voicing distinction differs for the two places of articulation. The strength of these acoustic cues (F_0 movement and SCG) is however small in comparison to the strength of prevoicing.

[1] The values of 9.5 and 3.5 ms is somewhat arbitrary. The aim of each split is to best separate low and high values. In order to divide the group into tokens with prevoicing and tokens without prevoicing, the cutoff value should be between 0 and a positive value (the smallest prevoicing duration larger than 0). Apparently, the algorithm settled on 9.5 ms for the labial plosives and on 3.5 ms for the alveolar plosives.

Experiment 2 examined which acoustic properties signal the voicing distinction and which of these acoustic correlates were the most reliable predictors of the voicing distinction as produced by speakers, but did not examine how the produced tokens were perceived by listeners. One would predict that there would be a good correspondence between the intended voicing category and the perceived voicing category, since the productions were produced in isolation and under laboratory conditions. Nevertheless, the results remain somewhat ambiguous. On the one hand, the presence or absence of prevoicing is the most reliable predictor of voicing, while on the other hand, a quarter of the voiced plosives were produced without prevoicing. The question therefore arises how voiced plosives that are produced without prevoicing are perceived by listeners. Is prevoicing indeed the strongest cue to the perception of the voicing distinction, as one would expect on the basis of the earlier acoustic analyses? It need not to be true that listeners weight various cues in the same way as an automatic classifier would. What counts as important to the listener will of course depend on how the signal is processed by the peripheral auditory system and by the sensitivities of the speech perception system.

The purpose of Experiment 3 was therefore twofold. First, we wanted to find out how the productions of Experiment 2 were perceived by listeners. In particular, we were interested in the perceived voicing of the voiced plosives without prevoicing. Second, we wanted to explore which of the acoustic cues influenced the perception of the listener most strongly and whether these cues corresponded to the cues which appeared to best describe the voicing distinction as produced by the speakers.

4. Experiment 3

4.1. Method

4.1.1. Materials

The materials for the perception experiment were based on the 48 voiced–voiceless pairs which were each produced by 10 speakers and analyzed in Experiment 2. Only the initial portions of each token, up to the middle of the vowel, were presented, so as to prevent the listeners from using lexical information. To avoid creating abrupt amplitude changes, the offset of each fragment was ramped down to zero within a time-window of 10 ms.

4.1.2. Procedure

The fragments were presented binaurally over headphones at a comfortable listening level. The materials were blocked by place of articulation and speaker. This resulted in 20 blocks of 96 items. The /b/–/p/ blocks and /d/–/t/ blocks were alternated. The speakers were randomized across blocks. The items within a block were randomized with the constraint that the items belonging to one pair were never presented consecutively. Different listeners were presented with different randomizations. They were tested in sound-proof booths and were instructed to categorize the first sound of the fragment that they heard as /b/ or /p/ for half of the blocks and as /d/ or /t/ for the other half of the blocks. Before each block started, the two phoneme categories for that particular block appeared on the screen and stayed there during the entire block. Listeners were asked to make their decision by pressing one of two buttons of a response box which corresponded to the phonemes that appeared on the screen. They had to respond within 1.5 s. If they failed to do so, the response was not recorded. This occurred in 135 cases in total (0.9%).

4.1.3. Participants

Sixteen volunteers from the Max Planck Institute participant pool were paid to take part in this experiment. All were native speakers of Dutch and none reported any hearing loss. None had taken part in Experiment 1.

4.2. Results and discussion

The results of the perception experiment showed that 8.32% of the 15,225 responses did not correspond to the phoneme category written on the list from which the speakers read the items. Inspection of the mismatches showed that there were more mismatching responses to voiced plosives than to voiceless plosives (9.8% for the voiced plosives and 6.8% for the voiceless plosives) and more mismatching responses to labial plosives than to alveolar plosives (8.7% and 7.9% respectively). A binary logistic regression analysis with the number of matching versus mismatching responses as dependent variable and voicing category and place of articulation as independent variables showed that both factors had a significant effect on the number of mismatching responses: Wald(1) = 56.2, $p < 0.0001$ (voicing), Wald(1) = 10.4, $p < 0.01$ (place of articulation). The interaction between voicing and place of articulation was also significant (Wald(1) = 33.3, $p < 0.0001$). The difference between the number of mismatching responses to voiced and voiceless plosives was larger for the labials than for the alveolars (5.9% for /p/ and 11.6% for /b/ versus 7.8% for /t/ and 8.1% for /d/): Wald(1) = 33.3, $p < 0.0001$.

To find out whether the presence of prevoicing had an influence on the number of mismatching responses, the same logistic regression analysis was conducted on only the voiced plosives. This time place of articulation and presence of prevoicing were the dependent variables. The effect of prevoicing was significant: Wald(1) = 176.2, $p < 0.0001$. There were considerably more mismatches to voiced plosives produced without prevoicing than to voiced plosives produced with prevoicing (36.6% versus 1.2%). There was also a significant interaction between prevoicing and place of articulation: Wald(1) = 12.8, $p < 0.001$. The difference between the percentage of mismatching responses for plosives produced without and with prevoicing was larger for labials than for alveolars: 53.6% (without prevoicing) versus 1.3% (with prevoicing) for the labials, and 25.7% (without prevoicing) versus 1.1% (with prevoicing) for the alveolars.

The identification responses showed that most plosives were perceived as belonging to the voicing category which was intended by the speaker. Although overall the proportion of mismatching responses was very small, analyses showed that there was a difference between the four plosives. The proportion of mismatching responses was largest for the labial voiced plosives, especially for the labial voiced plosives produced without prevoicing. Half of these plosives were perceived as being voiceless. Of the alveolar voiced plosives without prevoicing, a quarter of the tokens were perceived as being voiceless. This suggests that for alveolars the secondary cues are stronger than for labials.

As in Experiment 2, two separate CART analyses were conducted for the two places of articulation. This time the purpose was to find out which of the acoustic cues in the signal could best describe the perception of the voicing distinction. In contrast to Experiment 2, where the

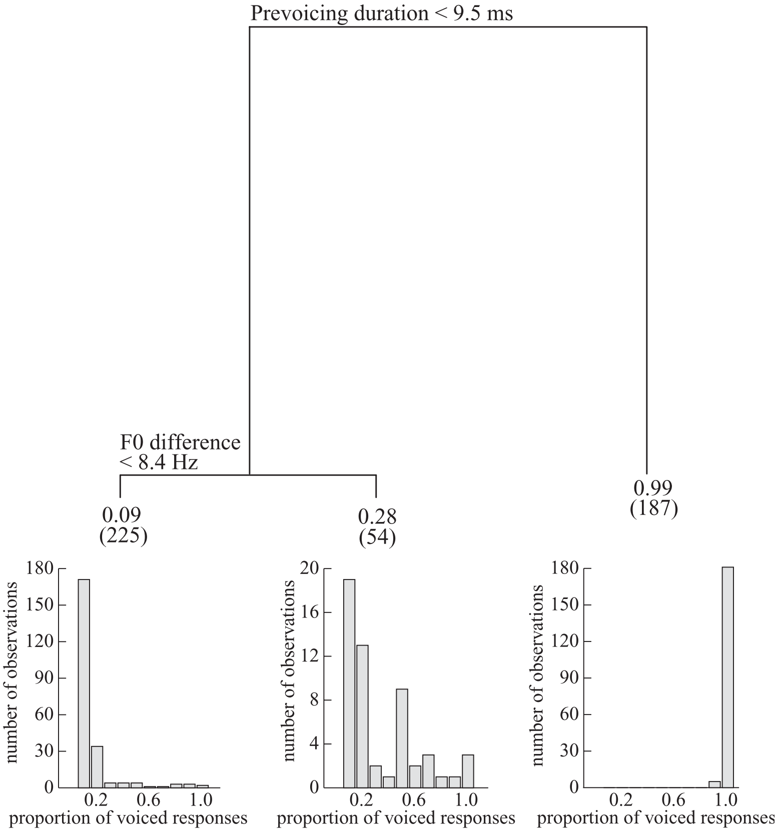

Fig. 13. Experiment 3: CART analysis of the proportion of voiced responses for the labial plosives. Tokens satisfying the rule printed at the top of each split follow the left branch. For each of the final nodes the mean of the proportion of voiced responses for the tokens falling in that final node is printed underneath. The number within parentheses indicate the total number of tokens falling in that final node. The histograms show the distributions of the mean proportions of voiced responses for the tokens falling in that final node.

response variable was the voicing category intended by the speaker, the response variable was now the proportion of voiced responses for each token. Since the response variable was now continuous, instead of categorical as in Experiment 2, a regression (rather than a classification) tree analysis was performed. The predictor variables of the regression tree analysis were identical to the predictor variables that were used in the classification tree analysis (Experiment 2): duration of prevoicing, duration of burst, spectral power of the burst above 500 Hz, SCG of the burst, absolute F_0 immediately after burst offset, F_0 difference, and following phoneme. As before, all tokens with missing measurements were excluded from the analyses (14 bilabials and 18 alveolars).

The two resulting cost-complexity pruned regression trees are shown in Figs. 13 (labials) and 14 (alveolars). The numbers printed directly below each node indicate the mean proportions of voiced responses for all tokens that fell into that final node. Below the mean proportion a small histogram shows the distribution of the proportions of voiced responses for the tokens. As before, the numbers between brackets indicate the numbers of tokens that fell in each final node.

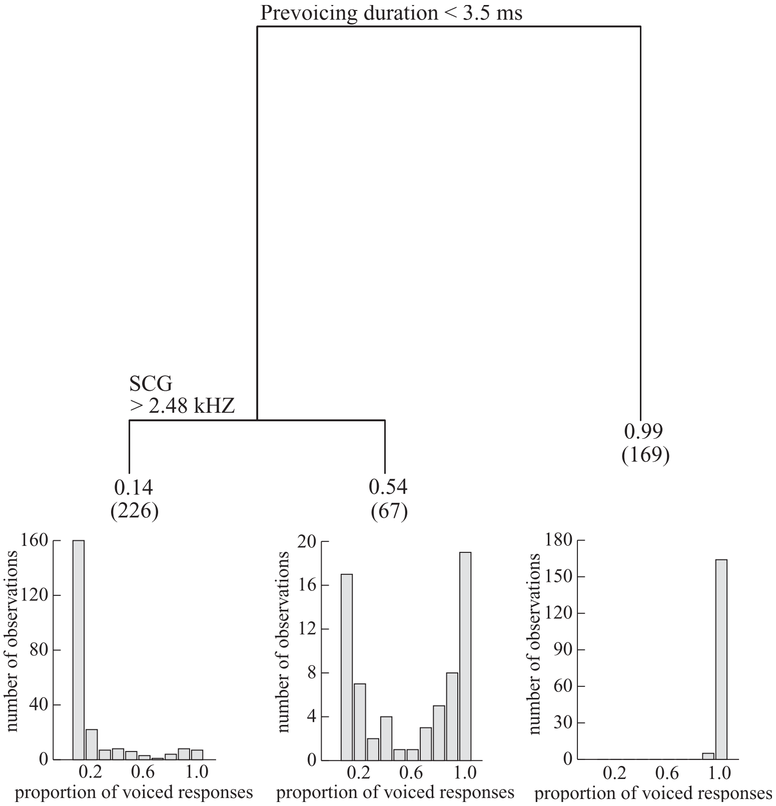

Fig. 14. Experiment 3: CART analysis of the proportion of voiced responses for the alveolar plosives. Tokens satisfying the rule printed at the top of each split follow the left branch. For each of the final nodes the mean of the proportion voiced responses for the tokens falling in that final node is printed underneath. The number within parentheses indicate the total number of tokens falling in that final node. The histograms show the distributions of the mean proportions voiced responses for the tokens falling in that final node.

The two regression trees (Figs. 13 and 14) are very similar to the corresponding classification trees of Experiment (Figs. 11 and 12). Again, for both the labials and the alveolars, the main split was based on prevoicing duration (9.5 and 3.5 ms, respectively, as before). The plosives produced with prevoicing were consistently perceived as being voiced (mean proportion of voiced responses was 0.99 for labials and alveolars), which is also shown in the histograms below the final nodes. The tokens produced without prevoicing were subdivided into two groups. As in Experiment 2, the labials were split on the basis of the F_0 movement, while the alveolars were split based on the SCG. Both splits divided the plosives without prevoicing into a relatively large group of plosives for which the mean proportion of voiced responses was small (0.09 for the labials and 0.14 for the alveolars), and into a relatively small group of plosives for which the voicing percept remained more ambiguous (0.28 for the labials and 0.54 for the alveolars). The histograms below these final nodes indicate that for the larger two groups most tokens were indeed perceived as being voiceless while for the two smaller groups there was a lot of variation between the tokens. Most of the alveolar plosives without prevoicing and a small SCG were consistently labeled as voiced or voiceless. There were only a few tokens which were ambiguous. The histogram for the labial plosives without prevoicing and with an F_0 difference larger than 8.4 Hz show a different pattern. Some of the tokens were consistently perceived as being voiceless, while only a few tokens were consistently perceived as being voiced. In addition, there were some tokens which appeared to be fully ambiguous. As we found in the earlier analyses, the strength of the prevoicing cue was much larger than that of the "secondary cues" F_0 difference and SCG. Nevertheless, prevoicing was less dominant for alveolars than for labials. This tallies well with our finding that listeners are more likely to correctly recognize an unprevoiced /d/ than an unprevoiced /b/.

The CART analyses showed that the presence or absence of prevoicing was by far the strongest cue for listeners to identify Dutch initial plosives as voiced or voiceless. This was true for both places of articulation. Both labial and alveolar plosives were perceived as being voiced when produced with prevoicing. The perception of plosives without prevoicing was different for the two places of articulation. Voicing perception in labials produced without prevoicing was influenced by the F_0 movement, while voicing perception of alveolar plosives without prevoicing was influenced by the SCG. These cues could only influence a small subset of the responses and the strengths of these cues were small in comparison to the strength of the main cue prevoicing (indicated by the length of the vertical branches of each split). Interestingly, however, the majority of the voiced plosives produced without prevoicing were still perceived as being voiced on the basis of other cues. These secondary cues were different for the two places of articulation.

In Table 5, performance levels of the listeners and of the regression tree analyses are summarized. Recall that the regression trees do not classify tokens as voiced or voiceless, rather they assign probabilities that the tokens are voiced. If the CART model would be forced to label tokens as voiced or voiceless they would simply choose the most probable category. For the labials, this means that all tokens without prevoicing would be labeled voiceless, because the

Table 5
Experiments 2 and 3: mean percentage of voiced decisions and correct (that is, in agreement with the intention of the speaker) decisions made by the listeners and by the CART analysis (Experiment 3), if it were forced to label tokens as voiced or voiceless on the basis of the probabilities at the end nodes. The numbers in the last row represent the mean percentage of such forced CART decisions in agreement with the decisions made by the majority of the listeners

		All tokens		All voiced tokens		Voiced tokens without prevoicing	
		Labial (%)	Alveolar (%)	Labial (%)	Alveolar (%)	Labial (%)	Alveolar (%)
Listeners	% voiced decisions	47	51	88	92	46	74
	% correct decisions	91	92	88	92	46	74
CART	% voiced decisions	40	51	80	86	0	51
	% correct decisions	90	86	80	86	0	51
% decisions of CART in agreement with the majority of the listeners		94	88	90	89	51	63

probability of voicedness is below 0.5 for the left and middle end nodes. As a result, the agreement between the CART model and the decisions made by the majority of the listeners is very low for the labial voiced tokens without prevoicing. For the alveolars, however, the probability voicedness of the middle node is higher than 0.5. Therefore, a forced choice would lead to the label voiced for these tokens. As a result, the agreement between listeners and the CART model is higher for the alveolars.

5. Summary and general discussion

This study investigated the production and perception of voicing in Dutch initial plosives. Experiment 1 focused on variation in prevoicing, which has been described as the primary cue for initial voiced plosives in Dutch. The productions of 10 different subjects indicated that there was a lot of variation among speakers in terms of number of prevoiced tokens and duration of prevoicing. Five out of 10 subjects prevoiced very consistently, with more than 90% of all their voiced tokens produced with prevoicing. The other five subjects produced prevoicing less frequently, but the proportion of prevoiced tokens varied considerably between those five less frequent prevoicers. Overall, 25% of all tokens produced by all 10 speakers were produced without prevoicing. Several factors appeared to have an effect on prevoicing production. First, the proportion of prevoiced tokens was higher for male speakers than for female speakers. Second, labial plosives were more often produced with prevoicing than alveolar plosives. Third, when the initial plosive was followed by a vowel, prevoicing was produced significantly more frequently and its duration was significantly longer than when the plosive was followed by a consonant.

Experiments 2 and 3 examined which acoustic properties are produced by speakers to signal the distinction between voiced and voiceless plosives, and which of these are used by listeners when they have to decide whether the plosive is voiced or voiceless. Several durational, spectral, and energy cues were measured. All acoustic properties but one had significantly different means for the two voicing categories. A CART analysis showed that of all these acoustic correlates, the presence or absence of prevoicing would be by far the most reliable cue to predict voicing. The tokens that were produced with prevoicing were all assigned to the voiced category. The tokens without prevoicing were further subdivided on the basis of another acoustic property. This property was different for the two places of articulation: the labials without prevoicing were split based on the F_0 difference, while the alveolars without prevoicing were split based on the SCG.

The perception study (Experiment 3) showed that the voicing feature of most tokens was perceived as intended by the speaker. Inspection of the mismatching responses showed that most mismatches appeared when the plosive was intended to be voiced but was produced without prevoicing. This suggests that prevoicing plays an important role in perception, which was confirmed by the outcomes of the CART analyses. The analyses showed that prevoicing was by far the strongest cue for the perception of the voicing distinction of both labial and alveolar plosives. Almost all tokens produced with prevoicing were identified as voiced. The majority of the tokens without prevoicing were perceived as voiceless, but a number of unprevoiced tokens were still perceived as voiced. The acoustic cue which most strongly influenced listeners' responses to tokens without prevoicing was different for the two places of articulation. The perception of voicing in labial plosives was influenced most strongly by the F_0 difference, while the perception of voicing in alveolar plosives was influenced most strongly by the SCG. The strength of these cues was fairly low in comparison to the strength of the presence of prevoicing. The correspondence between the analyses of the acoustic and perceptual data was very close.

The results show that the perception of voicing in Dutch plosives is asymmetric: the presence of prevoicing alone provides enough evidence for a listener to be sure that the plosive is voiced, while the absence of prevoicing alone does not provide enough evidence for a listener to categorise the plosive as voiceless. This asymmetry resembles findings in English by Port (1979). He reported that when audible glottal pulsing was maintained through the closure interval, an intervocalic plosive was heard as being voiced, no matter what the values of the other cues (duration of preceding vowel and closure duration) were. It was only when the closure interval was voiceless that the other cues could be effective. Removing all traces of glottal pulsing from the closure interval of a intervocalic /b/, however, did not change the phonemic percept. Only when the duration of the (silent) closure was increased was the plosive perceived as being voiceless.

It is important to note that the CART analysis was used as a statistical model to analyze the data and not as an explicit model for the listener's behavior. We do not claim that the listener's perceptual system works in the way the CART analysis does, namely by taking into account different cues in serial order. This would require a very complex model involving an extremely rapid succession of low-level decisions, in which listeners would first evaluate prevoicing and would only take other cues into account when this particular cue was absent. Instead, we support models of phonetic categorization in which listeners identify each token by considering in parallel all relevant cues that are available in the speech signal (e.g., Nearey, 1990; Smits, Ten Bosch, & Collier, 1996). These models claim that listeners first extract a number of perceptually relevant acoustic cues from the speech signal, which together constitute a point in a multidimensional feature space. Associated with this point is a set of probabilities of choosing each of the possible responses, on the basis of which the listener then makes a decision. Thus, the probability that a particular token in Experiment 3 belonged to the voiced category would be determined by all relevant cues. The CART analysis showed that the weight of the prevoicing cue was very high. The presence of prevoicing alone brought the probability of a voiced response so close to unity that variation in the other cues had no discernible effect. The absence of prevoicing, on the other hand, did not bring the voiced probability equally close to zero. When there was no prevoicing, other (weaker) cues, such as the F_0 difference for the labial plosives and the SCG for the alveolar plosives largely determined the class probabilities and therefore the decisions of listeners.

The present study shows that the voicing distinction is acoustically realized differently for labials and alveolars. First, the importance of prevoicing, which clearly plays the most important role in both labial and alveolar plosives, seems to differ between the two places of articulation. Experiments 1 and 2 showed that labials were produced more often with prevoicing than alveolars. Experiment 3 showed that listeners rely more strongly on prevoicing for labials than for alveolars, because the proportion of mismatching identification responses was larger for labials produced without prevoicing than for alveolars without prevoicing. Alveolar plosives, which have longer and stronger bursts than labial plosives, seem to carry more of the voicing distinction in the burst than labial plosives do. The finding that the difference in the SCG between voiced and voiceless plosives is larger for alveolars than for labials strengthened our impression that the place of articulation for /d/ and /t/ is slightly different. Experiment 3 showed that the SCG is an important cue in the perception of alveolar plosives produced without prevoicing but not for labial plosives.

The present study leaves us with an intriguing paradox. Prevoicing is the most reliable cue to the voicing distinction in Dutch initial plosives, yet in a quarter of all voiced plosives prevoicing is absent. Due to the presence of other cues not all voiced plosives without prevoicing were misperceived as voiceless. Nevertheless, although both the production and perception experiments were carried out under optimal conditions, almost 10% of the voiced plosives were mistakenly perceived as voiceless. This proportion is rather high in comparison to identification scores of English voiced plosives (e.g., Smits, 2000), for which the proportion of correct responses was close to 100%. This raises the question as to whether voicing in plosives is not communicated very accurately in Dutch. The data show that it is not the case that prevoicing is simply difficult to perceive, since all tokens produced with prevoicing were correctly identified as voiced. The voiced tokens which were mistakenly perceived as voiceless, however, were all produced without prevoicing. So the puzzling question is this: given the importance of prevoicing, why do speakers not produce prevoicing more reliably?

A possible explanation is that Dutch is undergoing sound change. One of the reviewers (Jessen) noted that along with Afrikaans and Yiddish, Dutch is the only Germanic language that makes a phonemic distinction between voiced and voiceless unaspirated plosives. The potential diminishing of prevoicing may be caused or boosted by the large influence of English on Dutch. All Dutch students (from which our subject group was drawn) speak English fluently, having received 6 years of formal training in schools form the each of 12, and being exposed to native English speech on television and radio on a daily basis, as all foreign TV shows are not dubbed but subtitled. Future research on other prevoicing languages and the influence of other language in which prevoicing is not important, should give more insight into this paradox between production and perception.

The present study attempted to give a detailed analysis of the production and perception of Dutch initial plosives in natural speech. Among all different acoustic properties that signal the voicing distinction in these plosives, the presence of prevoicing is by far the strongest cue to the perception of the plosive as belonging to the voiced category. Prevoicing is, however, not produced consistently by all speakers. Although the presence of prevoicing signals that the plosive is unmistakably voiced, prevoicing is not a prerequisite for a plosive to be perceived as being voiced. Other acoustic properties can provide sufficient evidence for the plosive to be voiced when prevoicing is absent. These cues are, however, weak in comparison to prevoicing. Interestingly, the voicing distinction in alveolar plosives seems to be realized with a small difference in the place of articulation, which makes the voicing distinction in alveolar plosives more robust than in labial plosives when prevoicing is absent.

Acknowledgements

Preliminary report of this research was given at the 143rd Meeting of the Acoustical Society of America, Pittsburgh (Pennsylvania), June 3–7, 2002. The authors are grateful to James McQueen for discussions throughout the project, to Harald Baayen for his help with the CART analyses, and to James McQueen, Mirjam Ernestus, and Taehong Cho for comments on an earlier version of this paper. Furthermore, we thank Jose Benki, Michael Jessen, and Pat Keating for very helpful reviews.

Appendix A

For Materials Experiment 1 see Table 6.

Table 6
Materials Experiment 1

	Vowel context		Consonant context	
	Labial	Alveolar	Labial	Alveolar
NW−comp	baag	daaf	bleep	dreek
	/baːχ/	/daːf/	/bleːp/	/dreːk/
	beugt	darf	blog	drens
	/bøːχt/	/dɑrf/	/blɔχ/	/drɛns/
	bimp	deust	brelt	dweum
	/bɪmp/	/døːst/	/brɛlt/	/dʋøːm/
	borf	dorg	brim	dwomp
	/bɔrf/	/dɔrχ/	/brɪm/	/dʋɔmp/
NW + comp	bark	daart	bluim	draan
	/bɑrk/	/daːrt/	/blœym/	/draːn/
	bech	dest	bluk	droost
	/bɛχ/	/dɛst/	/blʉk/	/droːst/
	biek	dint	brijs	dwaalf
	/biːk/	/dɪnt/	/brɛɪs/	/dʋaːlf/
	bijn	doon	broef	dwijg
	/bɛɪn/	/doːn/	/bruːf/	/dʋɛɪχ/
W + comp	baars (*perch*)	dag (*day*)	bloem (*flower*)	draad (*thread*)
	/baːrs/	/deur/	/bluːm/	/draːt/
	beest (*beast*)	duer (*door*)	blos (*blush*)	draf (*trot*)
	/beːst/	/døːr/	/blɔs/	/drɑf/
	berg (*mountain*)	dons (*down*)	bries (*breeze*)	dwaas (*foolish*)
	/bɛrχ/	/dɔns/	/briːs/	/dʋaːs/
	biels (*sleeper*)	duik (*dive*)	brood (*bread*)	dwars (*diagonal*)
	/biːls/	/dœyk/	/broːt/	/dʋars/
W + comp	baars (*perch*)	dak (*roof*)	blaag (*brat*)	drab (*dregs*)
	/baːrs/	/dɑk/	/blaːχ/	/drap/
	bed (*bed*)	dolk (*dagger*)	blad (*leaf*)	drek (*muck*)
	/bɛt/	/dɔlk/	/blɑt/	/drɛk/
	bink (*hunk*)	doorn (*thorn*)	bril (*glasses*)	drol (*turd*)
	/bɪnk/	/do:rn/	/brɪl/	/drɔl/
	boot (*boat*)	duin (*dune*)	brul (*roar*)	druk (*pressure*)
	/boːt/	/dœyn/	/brʉl/	/drʉk/

Appendix B

For Materials Experiments 2 and 3 see Table 7.

Table 7
Materials Experiments 2 and 3

	Vowel context		Consonant context	
	Labial	Alveolar	Labial	Alveolar
NW+comp	bark-park /bɑrk/-/pɑrk/ -(*park*)	daart-taart /daːrt/-/taːrt/ -(*pie*)	bluim-pluim /blœym/-/plœym/ -(*feather*)	draan-traːn /draːn/-/traːn/ -(*tear*)
	bech-pech /bɛχ/-/pɛχ/ -(*bad luck*)	dest-test /dɛst/-/tɛst/ -(*test*)	bluk-pluk /blʉk/-/plʉk/ -(*tuft*)	droost-troost /drosːt/-/trosːt/ -(*comfort*)
	biek-piek /biːk/-/piːk/ -(*peak*)	dint-tint /dɪnt/-/tɪnt/ -(*hue*)	breis-preis /brɛɪs/-/prɛɪs/ -(*pice*)	dwaalf-twaalf /dʋaːlf/-/tʋaːlf/ -(*twelve*)
	bijn-pijn /bɛɪn/-/pɛɪn/ -(*pain*)	doon-toon /doːn/-/toːn/ -(*tone*)	broef-proef /bruːf/-/pruːf/ -(*trail*)	dwijg-twijg /dʋɛɪχ/-/tʋɛɪχ/ -(*twig*)
W−comp	balk-palk /bɑlk/-/pɑlk/ (*beam*)-	damp-tamp /dɑmp/-/tɑmp/ (*vapour*)-	blok-plok /blɔk/-/plɔk/ (*block*)-	draad-traad /draːt/-/traːt/ (*thread*)-
	beek-peek /beːk/-/peːk/ (*brook*)-	deugd-teugd /døχt/-/tøχt/ (*virtue*)-	brood-proot /broːt/-/proːt/ (*bread*)-	drop-trop /drɔp/-/trɔp/ (*liqourice*)-
	beurs-peurs /børs/-/pørs/ (*grant*)-	dons-tons /dɔns/-/tɔns/ (*down*)-	braam-praam /braːm/-/praːm/ (*blackberry*)-	dwaas-twaas /dʋaːs/-/tʋaːs/ (*foolish*)-
	borg-porg /bɔrχ/-/pɔrχ/ (*bail*)-	duim-tuim /dœym/-/tœym/ (*thumb*)-	breed-preed /breːt/-/preːt/ (*broad*)-	dwerg-twerg /dʋɛrχ/-/tʋɛrχ/ (*dwarf*)-
W+comp	baars-paars /baːrs/-/paːrs/ (*perch*)-(*purple*)	dak-tak /dɑk/-/tɑk/ (*roof*)-(*branche*)	blaag-plaag /blaːχ/-/plaːχ/ (*brat*)-(*plague*)	drab-trap /drɑp/-/trɑp/ (*dregs*)-(*stairs*)
	bek-pek /bɛk/-/pɛk/ (*mouth*)-(*pitch*)	dolk-tolk /dɔlk/-/tɔlk/ (*dagger*)-(*interpreter*)	blad-plat /blɑt/-/plɑt/ (*leaf*)-(*flat*)	drek-trek /drɛk/-/trɛk/ (*muck*)-(*pull*)
	beul-peul /bøl/-/pøl/ (*brute*)-(*pod*)	doorn-toorn /doːrn/-/toːrn/ (*thorn*)-(*anger*)	blind-plint /blɪnt/-/plɪnt/ (*blind*)-(*skirting*)	drol-trol /drɔl/-/trɔl/ (*turd*)-(*troll*)

Table 7 (*continued*)

	Vowel context		Consonant context	
	Labial	Alveolar	Labial	Alveolar
	bink-pink /bɪnk/-/pɪnk/ (*hunk*)-(*little finger*)	duin-tuin /dœyn/-/tœyn/ (*dune*)-(*garden*)	bril-pril /brɪl/-/prɪl/ (*glasses*)-(*young*)	druk-truck /drʉk/-/trʉk/ (*pressure*)-(*truck*)

Note: W stands for word, NW stands for nonword, +comp stands for with voiceless word competitor and −comp stands for without voiceless word competitor.

References

Allen, G. D. (1985). How the young French child avoids the pre-voicing problems for word-initial voiced stops. *Journal of Child Language*, *12*, 37–46.

van den Berg, J. (1958). Myoelastic theory of voice production. *Journal of Speech and Hearing Research*, *1*, 244–277.

Breiman, L., Friedman, J. H., Olshen, R. H., & Stone, C. J. (1984). *Classification and regression trees*. New York: Chapman & Hall.

Caramazza, A., & Yeni-Komshian, G. H. (1974). Voice onset time in two French dialects. *Journal of Phonetics*, *2*, 239–245.

Cho, T., Jun, S., & Ladefoged, P. (2002). Acoustic and aerodynamic correlates of Korean stops and fricatives. *Journal of Phonetics*, *30*, 193–228.

Cho, T., & Ladefoged, P. (1999). Variation and universals in VOT: Evidence from 18 languages. *Journal of Phonetics*, *27*, 207–229.

Docherty, G. J. (1992). *The timing of voicing in English obstruents*. Berlin: Foris Publication.

van Dommelen, W. A. (1983). Some observations on assimilation of voicing in German and Dutch. In: M. van den Broecke, V. van Heuven & W. Zonneveld (Eds.), *Sound structures: studies for Anthonie Cohen*. Dordrecht: Foris.

Ernestus, M. (2000). *Voice assimilation and segment reduction in casual Dutch*. Utrecht: Holland Institute of Generative Linguistics.

Flege, J. E., & Eefting, W. (1987). Cross-language switching in stop consonant perception and production by Dutch speakers of English. *Speech Communication*, *6*, 185–202.

Forrest, K., Weismer, G., Milenkovic, P., & Dougall, R. N. (1988). Statistical-analysis of word-initial voiceless obstruents—preliminary data. *Journal of the Acoustical Society of America*, *84*, 115–123.

Haggard, M. P., Ambler, S., & Callow, M. (1970). Pitch as a voicing cue. *Journal of the Acoustical Society of America*, *47*(2), 613–617.

Haggard, M., Summerfield, Q., & Roberts, M. (1981). Psycho-acoustical and cultural determinants of phoneme boundaries—evidence from trading F_0 cues in the voiced–voiceless distinction. *Journal of Phonetics*, *9*, 49–62.

Honda, K., Hirai, H., & Kusakawa, N. (1993). Modeling vocal tract organs based on MRI and EMG observations and its implication on brain function. *Annual Bulletin*, No. 27 (pp. 37–49). Research Institute of Logopedics and Phoniatrics, University of Tokyo.

Houde, R. A. (1968). *A study of tongue body motion during selected speech sounds*. Santa Barbara: Speech Communication Research Laboratory, Monograph No. 2.

House, A. S., & Fairbanks, G. (1953). The influence of consonant environment upon the secondary acoustical characteristics of vowels. *Journal of the Acoustical Society of America*, *25*, 105–113.

Keating, P. A. (1984). Phonetic and phonological representation of stop consonant voicing. *Language*, *60*, 286–319.

Keating, P. A., Linker, W., & Huffman, M. (1983). Patterns in allophone distribution for voiced and voiceless stops. *Journal of Phonetics*, *11*, 277–290.

Keating, P. A., Mikos, M. J., & Ganong, W. F. (1981). A cross-language study of range of voice onset time in the perception of initial stop voicing. *Journal of the Acoustical Society of America*, *70*, 1261–1271.

Kewley-Port, D., & Preston, M. S. (1974). Early apical stop production: A voice onset time analysis. *Journal of Phonetics*, *2*, 195–210.

Kingston, J., & Diehl, R. L. (1994). Phonetic knowledge. *Language*, *70*, 419–454.

Kohler, K. J. (1985). F_0 in the perception of lenis and fortis plosives. *Journal of the Acoustical Society of America*, *78*, 21–32.

Konefal, J. A., & Fokes, J. (1981). Voice onset time: The development of Spanish–English distinction in normal and language disordered children. *Journal of Phonetics*, *9*, 437–444.

Lehiste, I., & Peterson, G. E. (1961). Some basic considerations on the analysis of intonation. *Journal of the Acoustical Society of America*, *33*, 419–425.

Lisker, L., & Abramson, A. S. (1964). A cross-language study of voicing in initial stops—acoustical measurements. *Word*, *20*, 384–422.

Löfqvist, A. (1978). Intrinsic and extrinsic F_0 variation in Swedish tonal accents. *Phonetica*, *31*, 228–247.

Macken, M. A., & Barton, D. (1980). The acquisition of the voicing contrast in English: A study of voice onset time in word-initial stop consonants. *Journal of Child Language*, *7*, 41–74.

Mohr, B. (1971). Intrinsic variations in speech signal. *Phonetica*, *23*, 65–93.

Nearey, T. M. (1990). The segment as a unit of speech perception. *Journal of Phonetics*, *18*, 347–373.

Ohala, J. J. (1983). The origin of sound patterns in vocal tract constraints. In P. F. MacNeilage (Ed.), *The production of speech* (pp. 189–216). New York: Springer.

Ohde, R. N. (1984). Fundamental-frequency as an acoustic correlate of stop consonant voicing. *Journal of the Acoustical Society of America*, *75*, 224–230.

Port, R. F. (1979). Influence of tempo on stop closure duration as a cue for voicing and place. *Journal of Phonetics*, *7*, 45–56.

Rothenberg, M. (1968). The breath-stream dynamics of simple-released-plosive production. *Bibliotheca Phonetica*, 6.

Rothman, G. B., Koenig, L. L., & Lucero, J. C. (2002). Intraoral pressure trajectories during voiced and voiceless stops in women and children. *Journal of the Acoustical Society of America*, *112*, 2416.

Slis, I. H., & Cohen, A. (1969). On complex regulating voiced–voiceless distinction. I. *Language and Speech*, *12*, 80–102.

Smith, B. (1978). Effects of place of articulation and vowel environment on voiced stop consonant production. *Glossa*, *12*, 163–175.

Smits, R. (1995). *Detailed versus gross spectro-temporal cues for the perception of stop consonants*. Unpublished doctoral dissertation, Institute for Perception Research, Eindhoven.

Smits, R. (2000). Temporal distribution of information for human consonant recognition in VCV utterances. *Journal of Phonetics*, *28*, 111–135.

Smits, R. L., Ten Bosch, L., & Collier, R. (1996). Evaluation of various sets of acoustic cues for the perception of prevocalic stop consonants. II. Modeling and evaluation. *Journal of the Acoustical Society of America*, *100*, 3865–3881.

Stevens, K. N. (1993). Models for the production and acoustics of stop consonants. *Speech Communication*, *13*, 367–375.

Stevens, K. N. (1998). *Acoustic phonetics*. Cambridge, MA: MIT Press.

Svirsky, M. A., Stevens, K. N., Matthies, M. L., Manzella, J., Perkell, J. S., & Wilhelms-Tricarico, R. (1997). Tongue surface displacement during obstruent stop consonants. *Journal of the Acoustical Society of America*, *102*, 562–571.

Talkin, D. (1995). A robust algorithm for pitch tracking (RAPT). In W. B. Kleijn, & K. K. Paliwal (Eds.), *Speech coding and synthesis*. New York: Elsevier.

Umeda, N. (1981). Influence of segmental factors on fundamental-frequency in fluent speech. *Journal of the Acoustical Society of America*, *70*, 350–355.

Westbury, J. R. (1983). Enlargement of the supraglottal cavity and its relation to stop consonant voicing. *Journal of the Acoustical Society of America*, *73*, 1322–1336.

Westbury, J. R., & Keating, P. A. (1986). On the naturalness of stop consonant voicing. *Journal of Linguistics*, *22*, 145–166.

Whalen, D. H., Abramson, A. S., Lisker, L., & Mody, M. (1993). F_0 gives voicing information even with unambiguous voice onset times. *Journal of the Acoustical Society of America*, *47*, 36–49.

Yeni-Komshian, G. H., Caramazza, A., & Preston, M. S. (1977). A study of voicing in Lebanese Arabic. *Journal of Phonetics*, *5*, 35–48.

Yoshioka, H., Murase, S., & Uematsu, M. (1996). Palato-lingual contact patterns during voiced and unvoiced consonant production. *Journal of the Acoustical Society of America*, *100*, 2661.