

# Lexical frequency and voice assimilation

Mirjam Ernestus,<sup>a)</sup> Mybeth Lahey, and Femke Verhees

Max Planck Institute for Psycholinguistics & Radboud University Nijmegen, Wundtlaan 1,  
6525 XD Nijmegen, The Netherlands

R. Harald Baayen

Radboud University Nijmegen & Max Planck Institute for Psycholinguistics, Wundtlaan 1,  
6525 XD Nijmegen, The Netherlands

(Received 14 July 2005; revised 11 February 2006; accepted 12 May 2006)

Acoustic duration and degree of vowel reduction are known to correlate with a word's frequency of occurrence. The present study broadens the research on the role of frequency in speech production to voice assimilation. The test case was regressive voice assimilation in Dutch. Clusters from a corpus of read speech were more often perceived as unassimilated in lower-frequency words and as either completely voiced (regressive assimilation) or, unexpectedly, as completely voiceless (progressive assimilation) in higher-frequency words. Frequency did not predict the voice classifications over and above important acoustic cues to voicing, suggesting that the frequency effects on the classifications were carried exclusively by the acoustic signal. The duration of the cluster and the period of glottal vibration during the cluster decreased while the duration of the release noises increased with frequency. This indicates that speakers reduce articulatory effort for higher-frequency words, with some acoustic cues signaling more voicing and others less voicing. A higher frequency leads not only to acoustic reduction but also to more assimilation. © 2006 Acoustical Society of America. [DOI: 10.1121/1.2211548]

PACS number(s): 43.70.Fq, 43.70.Bk, 43.71.An [AL]

Pages: 1040–1051

## I. INTRODUCTION

In everyday speech, words are often acoustically reduced compared to their citation form. In a study of conversational American English, Johnson (2004) found that at least one segment was missing in 25% of the words, while a complete syllable was absent in no less than 6% of the words. Similarly, Ernestus (2000) showed that words and fixed expressions may lose all their unstressed syllables in conversational Dutch. Thus, *eigenlijk* [ˈeɪxələk] “actually” may be realized as [ˈeɪk] and *in ieder geval* [ɪn ˈiðər xəˈvɑl] “in any case” as [ˈɪvɑl].

Several studies have shown that segments are more likely to be shorter or absent in words with a high frequency of occurrence (see, e.g., Zipf, 1935; Fidelholz, 1975; Bybee, 2001). Thus, Jurafsky *et al.* (2001), studying a large corpus of English telephone conversations, reported that word final /t/s and /d/s tend to be shorter and have a greater probability to be completely absent in words of a higher frequency of occurrence. On the basis of the same corpus, Bell *et al.* (2003) showed that a higher frequency of occurrence also correlates with shorter acoustic durations for function words. Similar results have been obtained for spontaneous Dutch: Pluymaekers, Ernestus, and Baayen (2005) found that Dutch affixes tend to be shorter and their segments to be more often absent in words of a higher frequency.

The present study broadens the scope of quantitative research on the role of lexical frequency in speech production by investigating the correlation between frequency and de-

gree of assimilation. The test case is regressive voice assimilation (RVA) in Dutch affecting obstruent clusters spanning morpheme boundaries.

In Dutch, all syllable-final obstruents are voiceless (final devoicing). Before the voiced plosives /b/ and /d/, however, they may be realized as voiced, a phenomenon referred to as regressive voice assimilation (e.g., Booij, 1995; Wetzels and Mascaró, 2001). Thus, the compound *wetboek* “law book,” consisting of the parts *wet* [vɛt] “law” and *boek* [buk] “book,” is generally pronounced as [vɛdbuk], and *twaalfduizend* “twelve thousand,” consisting of *twaalf* [tʌlf] “twelve” and *duizend* [duyzənt] “thousand,” as [tʌlvduyzənt]. RVA is described in the literature as obligatory within derived words and more frequent in compounds than across words (Loots, 1983; Booij, 1995). Furthermore, RVA is less frequent in obstruent clusters preceded by another voiceless obstruent (Demeulemeester, 1962), in women's speech (Kaiser, 1958; Slis, 1982, 1986), and at lower speech rates (Kaiser, 1958; Slis, 1982; Menert, 1994).

The voicing of obstruents in clusters as perceived by the listener is cued by several characteristics of the acoustic signal. Van den Berg (1986) showed that the most important cue for Dutch clusters is the presence of glottal vibration. The first obstruent of a cluster is perceived as voiced if glottal vibration is present during its final part, or during the initial part of the second obstruent. This second obstruent is perceived as voiced if glottal vibration is present during its final part. Other cues to the perception of voicing are the duration of the obstruents and the duration of the preceding vowel (van den Berg, 1987). Longer cluster-final obstruents tend to be perceived as voiceless. Longer preceding vowels favor the perception of the initial obstruents as voiced. The voicing of

<sup>a)</sup>Electronic mail: mirjam.ernestus@mpi.nl

single obstruents is cued by additional characteristics of the acoustic signal (for an overview, see van Alphen and Smits, 2004), including the duration of the plosive's release noise (the burst and the following period of aspiration), which tends to be shorter for voiced than for voiceless plosives (e.g., Slis and Cohen, 1969). The relevance of most of these additional subtle cues has not been investigated for obstruents in clusters.

Several competing hypotheses may be formulated concerning the correlation between frequency and degree of RVA. Two hypotheses depart from the idea that highly frequent morphologically complex words which are losing their internal morphological structure become more similar to monomorphemic words (cf. Ernestus, 2000: 34; see also Booij, 1995). This change might affect RVA in two ways. First, since RVA is more frequent across weaker (prosodic) boundaries (Booij, 1995), it might also be more frequent and stronger in high-frequency complex words in which boundaries are weakened. This would lead to higher percentages of completely voiced obstruent clusters and to more voiced clusters. Second, since completely voiceless clusters are typical for Dutch monomorphemic words (Zonneveld, 1983), one might also predict that RVA should occur less often or be weaker in high frequency complex words: A cluster with only one voiced obstruent would be closer to the ideal for monomorphemic words than a completely voiced cluster. The assumption underlying both lines of reasoning is that frequency of occurrence affects all acoustic cues for voicing in such a way that they all signal more voicing (first hypothesis) or less voicing (second hypothesis) in high frequency words. We will therefore refer to these hypotheses as Phonemic hypotheses.

Alternatively, we can formulate a hypothesis, which we will call the Subphonemic hypothesis, based on the observation that speakers tend to realize words of a higher frequency with less articulatory effort. Reduction in articulatory effort may affect a wide range of acoustic characteristics signaling voicing. Some of these characteristics may signal more voicing and others less voicing in words realized with less articulatory effort. For the listener, the net result of a higher frequency therefore will depend on the precise details of how articulatory reduction affects the different acoustic characteristics of the signal and on the extent to which these characteristics cue perceived voicing.

We pitted these hypotheses against each other by means of the subcorpus Library for the blind, part of the recently developed Spoken Dutch Corpus (Oostdijk, 2000; Oostdijk *et al.*, 2002). This subcorpus consists of parts of stories read aloud for the blind. It contains over 900 000 words, recorded with hardly any background noise, and it is therefore ideal for acoustic measurements. There are both male and female speakers, of very different ages (born between 1917 and 1974), and originating from all parts of the Netherlands. All recordings are transcribed orthographically.

We first classified obstruents in clusters from this subcorpus as voiced, voiceless, or absent. We investigated the predictive value of lexical frequency for these classifications (Sec. II). Whereas the Phonemic hypotheses predict a correlation between the classifications and frequency, this correla-

tion is not compelling for the Subphonemic hypothesis as the effects of frequency on the different characteristics of the acoustic signal may cancel each other out for perceived voicing. We then proceeded with detailed analyses of four important cues to voicing (Sec. III). In contrast to the Subphonemic hypothesis, the Phonemic hypotheses predict that these cues all signal either more voicing or less voicing at higher frequencies. Finally, by combining the voice classifications from Sec. II and the acoustic measurements from Sec. III, we tested whether frequency has a predictive role for the classifications in addition to the acoustic measurements (Sec. IV).

## II. PHONETIC TRANSCRIPTIONS

### A. Method

We studied obstruent clusters resulting from the concatenation of morphemes (e.g., the compound *voetbal* "football" resulting from the concatenation of *voet* and *bal*). The first obstruent in these clusters is realized as voiceless in word final position. We restricted ourselves to bilabial (/p/), labiodental (/f/), and alveolar (/s/, /t/) obstruents as cluster-initial obstruents. The second obstruent in the cluster was either /b/ or /d/, which may induce RVA of the preceding obstruent. We excluded clusters of only alveolar or only bilabial plosives, since they tend to be realized as single plosives (e.g., Booij, 1995). Furthermore, we only studied obstruent clusters preceded by a sonorant (consonant or vowel), as especially these clusters show RVA (Demeulemeester, 1962). The speakers were educated in the Western and Southern parts of the Netherlands, but we also considered a small number of speakers born in the Western or Southern parts of the Netherlands but educated in other regions.

The resulting data set contained 908 word tokens, realized by 84 men and 79 women. The majority (107) of speakers were educated in the West, and a minority in the South (29) or in other regions (27). We decided not to include the speakers' regional background in the analyses, a decision that hardly affected the results of the statistical analyses reported in this paper, as was shown by explorative analyses. The data set contained 321 compound tokens, 138 prefixed word tokens, and 449 suffixed word tokens. Of the suffixed word tokens, 14 are prosodic compounds according to the phonological literature (e.g., Booij, 1995), as they end in suffixes with full vowels.

Three phoneticians, native speakers of Dutch, classified the obstruents in the clusters as voiced, voiceless, or absent. One phonetician came from the center of the Western part, henceforth the West, another from the West of the Southern part, henceforth the South-West, and the third phonetician came from the very South of the Southern part, henceforth the South, of the Netherlands.

Transcribers tend to base their classifications not only on the acoustic signal but also on their expectations (e.g., Vierge, 1987). We attempted to minimize the role of the phoneticians' expectations based on their knowledge of words (and word frequencies) by presenting them only with a small part of each word, the part consisting of the obstruent cluster, the preceding vowel, and the following vowel. Some obstruent clusters were separated from their surrounding vow-

TABLE I. Absolute and relative numbers of obstruent clusters unanimously classified as completely voiced (+voice +voice), unassimilated (–voice +voice), completely voiceless (–voice –voice), as containing just the second obstruent (absent present), or other.

Categorization	Assimilation	<i>N</i>	Percentage
+voice +voice	Regressive	261	42.9%
–voice +voice	None	121	19.9%
–voice –voice	Progressive	151	24.8%
Absent present		57	9.4%
Other		19	3.1%

els by intervening sonorant consonants, and these consonants were then also included in the stretches of speech presented to the phoneticians. Thus, the phoneticians listened to *oetba* from *voetbal* “football” and to *endbla* from *ochtendblad* “morning paper.”

## B. Results

Table I shows the classifications of the 609 obstruent clusters on which the three phoneticians were in full agreement (67.1% of the classified clusters). In this table, the clusters are classified as completely voiced, as unassimilated (i.e., the initial obstruent as voiceless and the final obstruent as voiced), as completely voiceless, as realized without the initial obstruent, or as other (the final obstruent is absent or the initial obstruent is voiced while the final obstruent is voiceless).

As expected, the majority (43%) of clusters were classified as completely voiced, reflecting RVA. Unexpected is the high percentage (25%) of completely voiceless clusters, reflecting progressive voice assimilation. This type of assimilation was at least as frequent as no assimilation (20%).

The frequency measure that we considered for predicting the realization of the obstruent clusters is the frequency with which the two morphemes forming the obstruent cluster co-occur in derived words and compounds. Thus, the frequency for the realization of the *tb* cluster in the prefixed word *ont+bijt* “breakfast” was the frequency of *ontbijt* itself plus the frequency of *ontbijtje* “small breakfast,” *ontbijten* “to have breakfast,” *ontbijtservies* “breakfast set,” etc. We determined these cumulative frequencies on the basis of the CELEX lexical database (Baayen, Piepenbrock, and Gulikers, 1995). The distribution of these frequencies was heavily skewed, and we therefore applied a logarithmic transformation (with base *e*). Henceforth, we will refer to these transformed cumulative frequencies simply as frequency. Furthermore, we will refer with Word to that part of the word token under consideration that consists of the two morphemes contributing to the obstruent cluster and on which the frequency count is based (e.g., *ontbijt* both for *ontbijt* itself and for *ontbijtservies*).

We analyzed all classifications contributed by each of the three phoneticians, irrespective of whether the phoneticians were in full agreement. We first investigated whether frequency predicts assimilation by comparing clusters realized without assimilation with completely voiced clusters on the one hand and with completely voiceless clusters on the

other hand. In a next step, we studied the role of frequency in whether an assimilated cluster is completely voiced versus voiceless. Finally, we investigated whether frequency is a predictor for the absence of the initial obstruent.

The boxplots in Fig. 1 show the frequency distributions for these four types of classifications. The boxes show the interquartile ranges, the dots in the boxes denote the medians, and the “whiskers” extend to the observations within 1.5 times the interquartile range. Outliers beyond this range are represented by individual circles.

All statistical analyses in this study made use of stepwise multilevel models of covariance with speaker and Word as crossed grouping factors (Pinheiro and Bates, 2000; Baayen, Tweedie, and Schreuder, 2002; Baayen, 2004; Quené and van der Bergh, 2004; Bates and Sarkar, 2005). Such models obviate the necessity of separate F1 and F2 analyses. Note that the by-item covariate frequency is confounded with the random effect of Word. Hence, our multilevel models are conservative with respect to the contribution of this covariate.

For the analyses of the phonetic classifications, we used generalized linear mixed models with a binomial link function, using penalized quaslikelihood (e.g., Venables and Ripley, 2002). Such models predict the probability of a (binary) classification. We entered as predictors the speakers’ Gender and Year of birth (minus 1900), the Type of the first obstruent in the cluster (fricative versus plosive), the Second obstruent (*/b/* versus */d/*), the Morphological class of the word (prefixed, suffixed, or compound), and Phonetician (three levels). We entered Frequency as the last predictor, again keeping the models as conservative as possible with respect to this covariate.

We first investigated the probability that a cluster was perceived as completely voiced versus unassimilated. A higher frequency increased the probability that the cluster was perceived as completely voiced ( $\hat{\beta}=0.14$ , with  $\hat{\beta}$  denoting the estimated (unstandardized) regression coefficient, following the notation of, e.g., Chatterjee, Hadi, and Price, 2000,  $F(1, 1664)=4.23, p<0.05$ ), though the effect was small (see Fig. 1). Furthermore, men’s clusters were more often classified as completely voiced (68.9% completely voiced classifications and 31.1% unassimilated classifications) than women’s clusters [56.2% completely voiced classifications and 43.8% unassimilated classifications,  $F(1, 1664)=10.43, p=0.001$ ], and the phonetician from the West classified fewer clusters (57.9%) as completely voiced than the phoneticians from the South-West (68.2%) and the South (63.3%,  $F(2, 1664)=17.81, p<0.0001$ ), who did not differ from each other in their classifications ( $p>0.05$ ).

In the second analysis, we investigated the likelihood that a cluster was realized as completely voiceless versus unassimilated. A higher frequency increased the probability that the cluster was perceived as completely voiceless [ $\hat{\beta}=0.19, F(1, 1252)=4.38, p<0.05$ ]. In addition, the phonetician from the West classified clusters as completely voiceless (50.0% completely voiceless classifications and 50% unassimilated classifications) more often than the phonetician from the South (48.8% completely voiceless classifications)

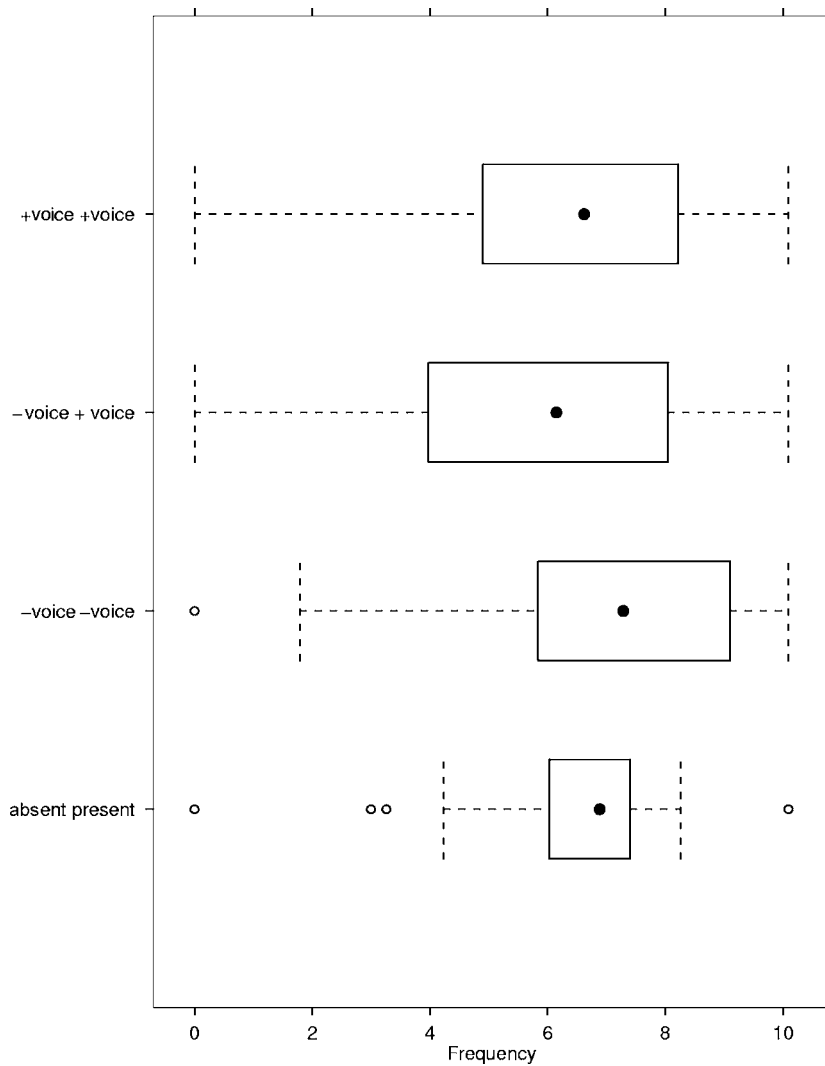


FIG. 1. Boxplot for the frequency distributions of the clusters that were classified as completely voiced (+voice +voice), unassimilated (-voice +voice), completely voiceless (-voice -voice), or as containing just the second obstruent (absent present).

but less often than the phonetician from the South-West [54.7%,  $F(2, 1252)=3.39$ ,  $p < 0.05$ ]. These two phoneticians from the South also differed significantly from each other in their classifications [ $F(1, 1117)=4.50$ ,  $p < 0.05$ ].

Third, we compared the completely voiced and the completely voiceless classifications. Plosive-initial clusters were more often classified as completely voiced (78.5% completely voiced classifications and 21.5% completely voiceless clusters) than fricative-initial clusters [58.4% completely voiced clusters,  $F(1, 1691)=5.02$ ,  $p < 0.05$ ] and the phoneticians from the South (64.4%) and South-West (64.0%) classified more clusters as completely voiced than the phonetician from the West [57.9%,  $F(2, 1691)=5.03$ ,  $p < 0.01$ ]. Frequency was not predictive.

Finally, we modeled the presence versus absence of the initial obstruent. The model did not converge when Word was included as a random effect. Hence, we report the model with only speaker as random effect. The initial obstruent was more likely to be absent in higher-frequency words [ $\hat{\beta} = 0.0002$ ,  $F(1, 2630)=111.64$ ,  $p < 0.0001$ ]. Furthermore, especially plosives were often absent [in 30.8% of cases, versus fricatives in 2.5% of cases,  $F(1, 2630)=115.57$ ,  $p < 0.0001$ ] and obstruents were more often classified as absent before /b/ (19.3% of the obstruents before /b/) than be-

fore /d/ [4.8%,  $F(1, 2630)=19.72$ ,  $p < 0.0001$ ]. These observations point to the generalization that especially alveolar plosives preceding /b/ are frequently absent.

### C. Discussion

Our data document a correlation between a word's frequency of occurrence and voice assimilation. We observed that clusters are more often perceived as completely voiced or completely voiceless in higher-frequency words. Especially the words with completely voiceless clusters and those with unassimilated clusters differed significantly with respect to their average frequency.

The high percentage of clusters perceived as completely voiceless is unexpected given the literature on Dutch phonology. The received wisdom is that in Dutch progressive voice assimilation is restricted to clusters ending in fricatives or in the initial plosives of function words (e.g., Demeulemeester, 1962; Zonneveld, 1983; Booij, 1995; Rietveld and van Heuven, 2001; but see Slis, 1986). Such clusters were not included in our data set, and, hence, our data show that this generalization is incorrect. Progressive voice assimilation also affects clusters ending in plosives that do not belong to function words, but to stems, to suffixes, or to content words

in compounds (cf. Slis, 1986). This falsifies phonological theories in which the initial plosives of function words receive special status because they would be the only plosives showing progressive voice assimilation (e.g., Zonneveld, 1983).

The phonetician from the South classified obstruents as voiced more often than the phonetician from the West, while the phonetician from the South-West patterned in general with the phonetician from the South. Since obstruent devoicing is more common in the West than in the South of the Netherlands (e.g., Collins and Mees, 1981: 159; Gussenhoven and Bremmer, 1983: 57), this suggests that the phoneticians' classifications reflected the likelihoods of voiced and voiceless obstruents in their own regiolects. This would be in line with the study by Coussé *et al.* (2004), who showed that trained phoneticians classify Dutch vowels as short or long in accordance with their own regional variety of Dutch.

Especially alveolar plosives before /b/ were classified as absent (cf. Mitterer and Ernestus, 2006). In languages related to Dutch, such as English (Marslen-Wilson, Nix, and Gaskell, 1995) and German (Wiese, 1996), alveolar plosives have been claimed to assimilate to the place of articulation of the following obstruent. Hence, a /tb/ cluster may be realized as a /pb/ or /bb/ cluster, which is difficult to distinguish from a single /b/. Also the high deletion rate in Dutch /tb/ clusters might be argued to be due to place assimilation. If so, alveolar plosives should be frequently absent also before velar plosives: A /tk/ cluster would be realized as a /kk/ cluster, and hence it should be difficult to distinguish from a single /k/. In order to test this prediction, the three phoneticians transcribed all 163 /tk/ clusters from the subcorpus Library for the blind consisting of an alveolar plosive and a velar plosive, in the context of surrounding sonorants and realized by speakers educated in the Western or Southern parts of the Netherlands. The /t/ was classified as absent in only one cluster. We conclude that the perceived high deletion rate of /t/ in /tb/-clusters is unlikely to result from place assimilation.

The explanation offered by Browman and Goldstein (1990: 360) for the absence of /t/ before /b/ in English appears to be more promising for our Dutch data. Browman and Goldstein argue that the absence of alveolar plosives before bilabial plosives may be the consequence of coarticulation. Bilabial plosives are realized with a constriction at the lips, and they thus can mask all sounds that are simultaneously realized within the vocal tract. If they are realized simultaneously with an alveolar plosive, they can mask at least the release of this alveolar plosive, which is the most important cue to its recognition. In such cases, the alveolar plosive is acoustically absent, although its articulatory gestures are present. Alveolar plosives cannot be masked by velar plosives, as velar plosives are realized with a constriction in the back of the mouth. The account by Browman and Goldstein thus not only explains the high deletion rate of /t/ before /b/, but also the low deletion rate of /t/ before /k/.

We now turn to detailed analyses of the acoustic characteristics of the speech signal that cue perceived voicing. We investigated whether frequency correlates with the duration of several components of the cluster and with the dura-

tion of the preceding vowel. Recall that according to the Phonemic hypotheses, these durations should all signal either more or less voicing at higher frequencies. The Subphonemic hypothesis predicts that in higher-frequency words the acoustic characteristics are realized with less articulatory effort.

### III. ACOUSTIC MEASUREMENTS

#### A. Method

Our data set contained 732 obstruent clusters, of which 711 had also been investigated in the classification study above. The obstruents in the additional 21 clusters could not be reliably classified as voiced, voiceless, or absent, due to background noise, whereas they could be measured accurately. Conversely, the data set lacked the 197 clusters incorporated in the classification data set that could not be reliably measured because of background noise or that were classified as containing only one obstruent.

Recall that voiced obstruents typically follow longer vowels than voiceless obstruents, that they are typically shorter than voiceless obstruents, and also contain shorter release noises. In addition, voiced obstruents in clusters tend to be produced with glottal vibration during their final parts (see Sec. I). Since clusters typically contain at best one interruption in glottal vibration, which starts during the first obstruent, the duration of glottal vibration provides information on its temporal location (i.e., the longer the period, the greater the chance that the final parts of both obstruents are realized with glottal vibration). We decided to measure the duration of the vowel preceding the obstruent cluster, the duration of the obstruent cluster itself, the period of glottal vibration during the cluster, and the duration of the release noise in the cluster.

In case the preceding vowel was separated from the obstruent cluster by a sonorant consonant (as in *ochtendblad* "morning paper" in which the cluster /tb/ is preceded by /n/), we included the duration of this sonorant in our measurement of the vowel, since it is in general impossible to determine exactly where the vowel ends in the acoustic signal and where the following sonorant consonant starts. We defined the beginning of the vowel as the beginning of the regular wave form with the characteristics of the vowel, and the end of the vowel (plus sonorant consonant) as the (sudden) end of this regular wave form. The end of the obstruent cluster was defined as the beginning of the regular wave form with the characteristics of the following sonorant. We assumed glottal vibration to be present in that part of the obstruent cluster where the wave form was periodic, the spectrogram contained a voice bar, and where we could hear the vocal fold vibration. Finally, we defined release noise to start at the sudden increase in amplitude after a closure and to end where the acoustic signal was regular or flat again. All measurements were in seconds. Figure 2 shows the segmentation for a token of the Word *afbeeld* "depict" containing the cluster /fb/.

The two phoneticians from the South and South-West of the Netherlands who previously had classified the clusters, made the acoustic measurements. Each phonetician mea-

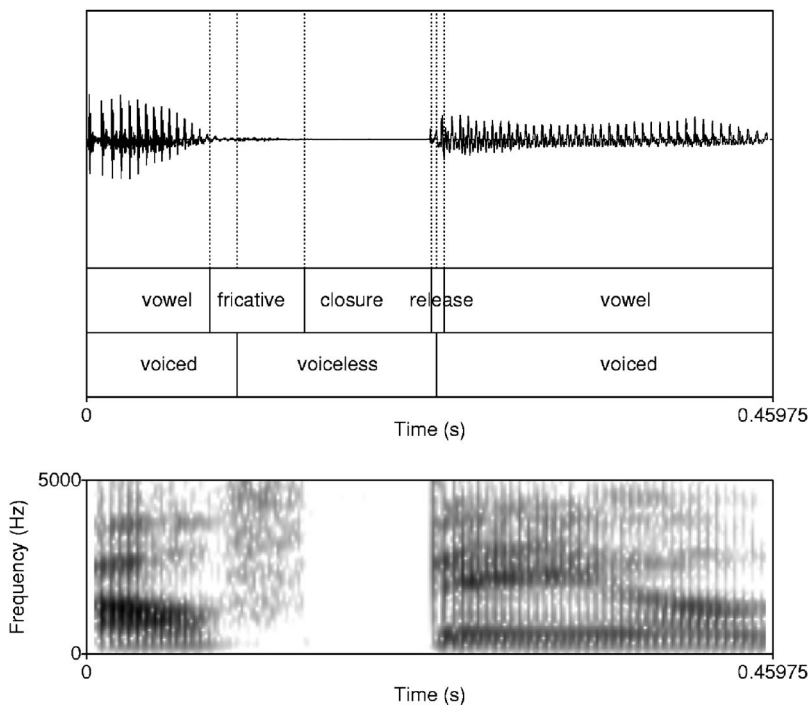


FIG. 2. Segmentation of [afbe] from *afbeeld* “depict.”

sured half of the clusters. One year later, each phonetician measured the durations for 100 tokens a second time. The first and second measurements differed on average 5 ms, with a standard deviation (after removal of one outlier) ranging from 4 ms (for cluster and burst duration) to 8 ms (for the period of glottal vibration).

## B. Results

We considered as predictors for the acoustic measurements the speakers’ Gender and Year of birth (minus 1900), the Type of the first obstruent (fricative versus plosive), the Second obstruent (/b/ versus /d/), and the Morphological class of the word (prefixed, suffixed, or compound). Since listeners interpret segment durations relative to the durations of surrounding segments (Miller, 1981), we entered as predictor for the duration of the preceding vowel also the duration of the cluster. For the other durations, we entered the acoustic duration of the preceding vowel, its phonological length (long versus short, see, e.g., Booij, 1995), and whether this vowel was stressed. Together, these additional predictors reflect speech rate. Finally, we entered Frequency as predictor.

### 1. Cluster duration

We first studied the duration of the obstruent cluster. Frequency correlated negatively with cluster duration [ $\hat{\beta} = -0.002$ ,  $F(1, 729) = 18.63$ ,  $p < 0.0001$ ]. In addition, clusters ending in /b/ (mean: 0.136 s) were on average 16 ms longer than those ending in /d/ [mean: 0.120 s,  $F(1, 729) = 10.78$ ,  $p = 0.001$ ].

Cluster duration itself may be predictive for the other acoustic durations signaling voicing. For instance, it is an obvious predictor for the duration of the period of glottal vibration during the cluster, since, by definition, this period cannot be longer than the cluster itself. We therefore in-

cluded cluster duration as a covariate in the statistical models for the durations of the period of glottal vibration, the release noise, and the preceding vowel. Since cluster duration correlates with Frequency, we decided not to include raw cluster duration. In order to avoid collinearity between our predictors (cluster duration and Frequency), we first fitted cluster duration against Frequency and entered the resulting residuals as covariates in the models. These residuals (henceforth Cluster residuals) show a correlation of  $r = 0.79$  ( $p < 0.0001$ ) with the cluster durations. In other words, greater cluster residuals imply longer clusters.

### 2. Period of glottal vibration

Of the 732 clusters, 253 were realized without any glottal vibration. We analyzed the complete absence versus presence of at least some glottal vibration by means of a generalized linear mixed model with a binomial link function. Women realized obstruent clusters more often without glottal vibration (in 43.8% of cases) than men [24.6%,  $F(1, 727) = 9.08$ ,  $p < 0.01$ ]. Obstruent clusters ending in /d/ were more often realized without glottal vibration (in 47.2% of cases) than obstruent clusters ending in /b/ [6.6%,  $F(1, 727) = 180.29$ ,  $p < 0.0001$ ]. Furthermore, we found that for clusters preceded by vowels longer than 120 ms, a longer vowel implied a higher probability of complete absence of glottal vibration [linear effect of vowel duration:  $\hat{\beta} = -3.84$ ,  $F(1, 727) = 646$ ,  $p = 0.01$ ; quadratic effect:  $\hat{\beta} = 15.34$ ,  $F(1, 727) = 10.87$ ,  $p = 0.001$ ].

The attested effects of gender and the Second obstruent are as expected. The larynx of men and women differ in their properties (e.g., shape, size, mass, stiffness) such that the oscillation conditions of the vocal folds are more restricted for women than for men (Lucero and Koenig, 2005). Furthermore, the oral volume can be expanded to a greater extent, both passively and actively, during the production of a

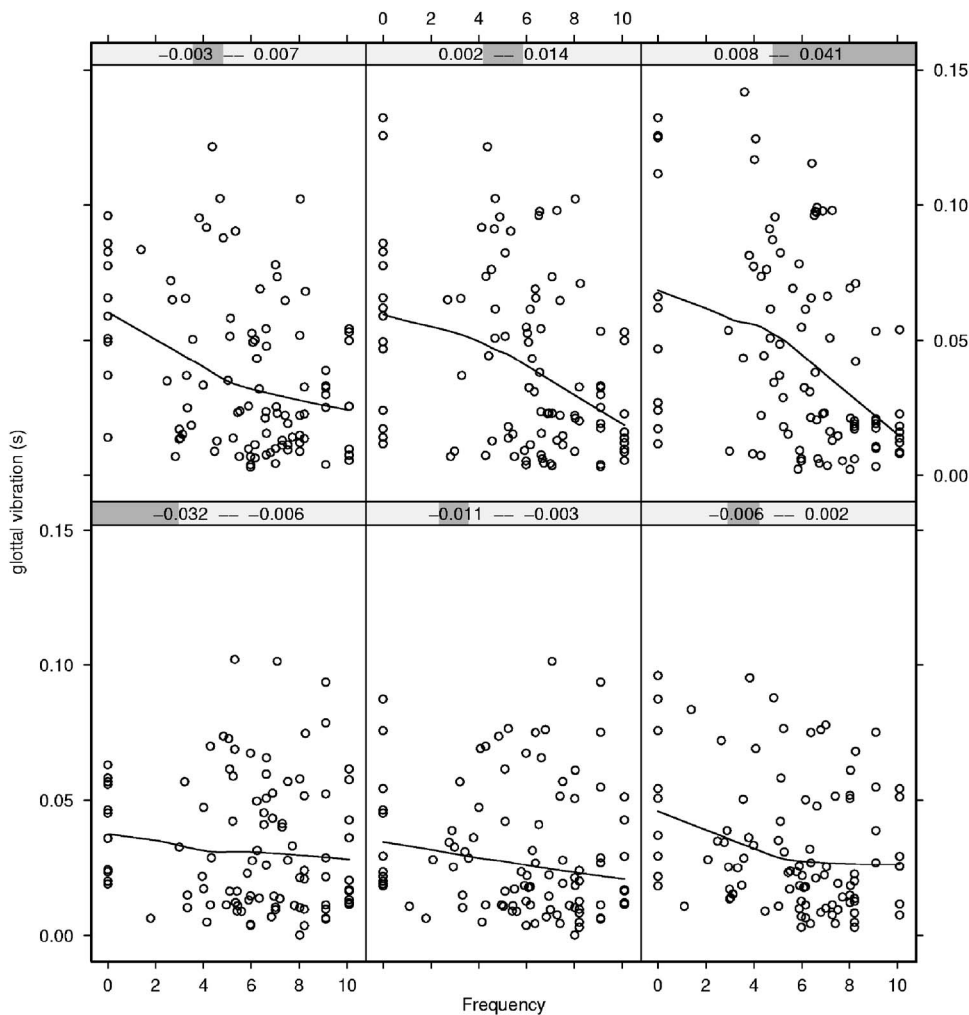


FIG. 3. The interaction of frequency and Cluster residuals on the duration of the period of glottal vibration. Every panel represents 97 clusters ordered by Cluster residuals: The bottom left panel represents the 97 clusters with the smallest Cluster residuals and the top right panel the 97 clusters with the greatest Cluster residuals (see the bar at the top of each panel). The solid line in each panel represents a non-parametric scatterplot smoother.

bilabial than an alveolar plosive (Ohala, 1983). As a consequence, the difference between the sub- and supraglottal air pressure, which is necessary for glottal vibration, is more easily maintained during [b] than [d].

In 137 out of the 479 clusters realized with glottal vibration, the vocal folds vibrated continuously from beginning to end. Men realized clusters with continuous glottal vibration more often (in 36.0% of cases) than women [in 19.2% of cases,  $F(1, 729)=12.55, p<0.001$ ]. Furthermore, clusters ending in /b/ were more often (in 33.3% of cases) produced with continuous glottal vibration than those ending in /d/ [24.8%,  $F(1, 729)=17.80, p<0.0001$ ]. These patterns are in line with those observed for the complete absence versus presence of some glottal vibration.

Finally, we modeled the length of the period of glottal vibration by means of linear regression analysis. Our data set only included the 342 clusters that were partly produced with glottal vibration. We did not include the clusters that were realized completely with or without glottal vibration since this would have led to an extreme bimodal distribution of the dependent variable, with a first peak at 0 s and a second peak at approximately 0.12 s, the average duration of the clusters.

We found a main effect of Frequency [ $F(1, 337)=6.25, p<0.05$ ] and of cluster residuals [ $\hat{\beta}=0.85, F(1, 337)=4.11, p<0.05$ ], in addition to an interaction of Frequency with Cluster residuals [ $F(1, 337)=12.08, p<0.001$ ]. Figure

3 illustrates the effect of Frequency and its interaction with Cluster residuals. Each panel in this trellis graph shows the correlation between Frequency (X axis) and the duration of the period of glottal vibration (Y axis) for clusters with the cluster residuals specified at the top of the panel. Cluster residuals increase from the bottom left panel to the bottom right panel and then from the top left panel to the top right panel. In general, frequency correlates negatively with the period of glottal vibration, but this effect is attenuated for clusters with smaller Cluster residuals.

In addition, the period of glottal vibration was, unsurprisingly, longer in clusters with greater Cluster residuals [ $\hat{\beta}=0.85, F(1, 337)=4.11, p<0.05$ ] and in clusters ending in /b/ [mean duration in clusters ending in /b/ was 0.051 s. and in clusters ending in /d/ 0.029 s,  $F(1, 337)=31.40, p<0.0001$ ]. The effects of these two variables varied with speaker (all  $p<0.001$ ).

### 3. Duration of the release noise

We investigated the total duration of the release noise(s) in the cluster. In fricative-initial clusters, only the final plosive is realized with a release noise, while in plosive-initial clusters, both the initial and the final plosive may be realized with release noises.

Frequency was a significant nonlinear predictor in the

model [linear:  $\hat{\beta} = -0.00045$ ,  $F(1, 725) = 0.67$ ,  $p > 0.1$ , quadratic:  $\hat{\beta} = 0.00013$ ,  $F(1, 725) = 11.04$ ,  $p < 0.001$ ]. From a log frequency of 4 onwards, the total duration of the release noises increased approximately linearly with increasing frequency. From 0 to 4, the duration increased only slightly with frequency. Furthermore, the total duration of the release noises was greater in clusters with greater cluster residuals [ $\hat{\beta} = 0.071$ ,  $F(1, 725) = 11.23$ ,  $p < 0.001$ ]. It was also greater if the initial obstruent was a plosive [mean duration of 0.033 s for plosive-initial clusters versus 0.021 s for fricative-initial clusters:  $F(1, 725) = 72.36$ ,  $p < 0.0001$ ], and if the final plosive was /d/ [mean duration for /d/-final clusters: 0.0232 s; for /b/-final clusters: 0.0230 s,  $F(1, 726) = 29.06$ ,  $p < 0.0001$ ]. The effect of the Type of the first obstruent varied with both speaker and Word while the effect of the Second obstruent varied with speaker (all  $p < 0.001$ ). Finally, the total duration of the release noises increased with longer durations of the preceding vowel [ $\hat{\beta} = 0.019$ ,  $F(1, 725) = 4.87$ ,  $p < 0.05$ ].

#### 4. Duration of the preceding vowel

Finally, we studied the duration of the vowel preceding the obstruent cluster, which is also a cue to perceived voicing in Dutch, though less strong than in English (Slis and Cohen, 1969). As expected, phonologically long vowels were on average longer (mean: 0.150 s) than phonologically short vowels [mean: 0.101 s,  $F(1, 728) = 210.13$ ,  $p < 0.0001$ ] and stressed vowels were on average longer (mean: 0.120 s) than unstressed vowels [mean: 0.077 s,  $F(1, 728) = 47.53$ ,  $p < 0.0001$ ]. Furthermore, vowels preceding fricative-initial clusters were longer (mean: 0.119 s) than those preceding plosive-initial clusters [mean: 0.099 s,  $F(1, 728) = 7.69$ ,  $p < 0.01$ ]. Frequency, however, was not predictive for the duration of the preceding vowel.

### C. Discussion

Summing up, as frequency increases, both the cluster and the period of glottal vibration become shorter, while the total release noise becomes longer. We now consider what these results imply with respect to the relation between frequency and articulatory effort.

Shorter periods of glottal vibration, as observed in words of higher frequencies, imply less articulatory effort. One necessary condition for the realization of glottal vibration is that the supraglottal air pressure is lower than the subglottal pressure. During the realization of fricatives, the air pressure above the glottis also has to be higher than the pressure beyond the constriction. These two requirements can be met simultaneously with undue effort only for a short time. In order to reduce articulatory effort, speakers may relax the adduction of the glottis, terminating glottal vibration but facilitating the realization of frication. During the realization of plosives, the air passing the vocal folds cannot escape from the vocal tract, and pressure builds up above the glottis. As a consequence, glottal vibration may simply stop in long plosives or in clusters of plosives, on average after 65 ms (Ohala, 1983). Since only little articulatory effort is necessary for glottal vibration in short clusters, reduction in articu-

latory effort hardly affects these clusters. This explains the observed interaction between Cluster residuals and Frequency for the period of glottal vibration (Fig. 3).

Reduction in articulatory effort is also achieved by reduction in the magnitudes of the constricting and release gestures, necessary for the production of the cluster, and by temporal overlap of these articulatory gestures. This results in shorter obstruent clusters. Moreover, it may result in longer release noises: If a plosive is realized with reduced constricting gestures, overlapping with the release gestures, its constriction is (partly) taken over by the release noise, which is consequently lengthened. As a consequence, the plosive may even sound as a fricative (cf. Ernestus, 2000: 203).

In conclusion, the attested correlations between the acoustic durations and frequency reflect reduction in articulatory effort. Our data are thus in line with the Subphonemic hypothesis, which states that a higher lexical frequency results in reduction of articulatory effort affecting a wide range of characteristics of the acoustic signal.

Shorter periods of glottal vibration and longer release noises result in less perceived voicing, whereas shorter obstruent clusters result in more perceived voicing (e.g., Slis and Cohen, 1969; van den Berg, 1986, 1987). Thus, in words of higher frequency, some cues signal more while others signal less perceived voicing. This finding is also in line with the Subphonemic hypothesis, while falsifying the two Phonemic hypotheses.

In Sec. II, we observed that frequency correlates with the phonetic classification of clusters as completely voiced, unassimilated, or completely voiceless. The three phoneticians listened only to the obstruent clusters and the surrounding vowels (plus intervening sonorant consonants, if present). In the majority of cases, they thus could not recognize the words, and their classifications could not be affected by their knowledge of the words, including the words' frequencies. This leads to the prediction that frequency is not a predictor for the phoneticians' classifications over and above the acoustic durations measured in Sec. III. We tested this prediction in Sec. IV.

## IV. NATURE OF THE FREQUENCY EFFECTS IN THE CLASSIFICATIONS

### A. Method

We combined our acoustic measurements with the voice classifications obtained in Sec. II. The resulting data set contained 711 clusters. Since every cluster was classified by three phoneticians, it consisted of in all 2133 classifications. We investigated whether frequency is a predictor for the phoneticians' judgments over and above the acoustic measurements, by means of stepwise generalized linear multilevel models with a binomial link function.

### B. Results

Figure 4 shows the distributions of the durations of the complete cluster, of the period of glottal vibration during the cluster, of the release noise(s), and of the preceding vowel,



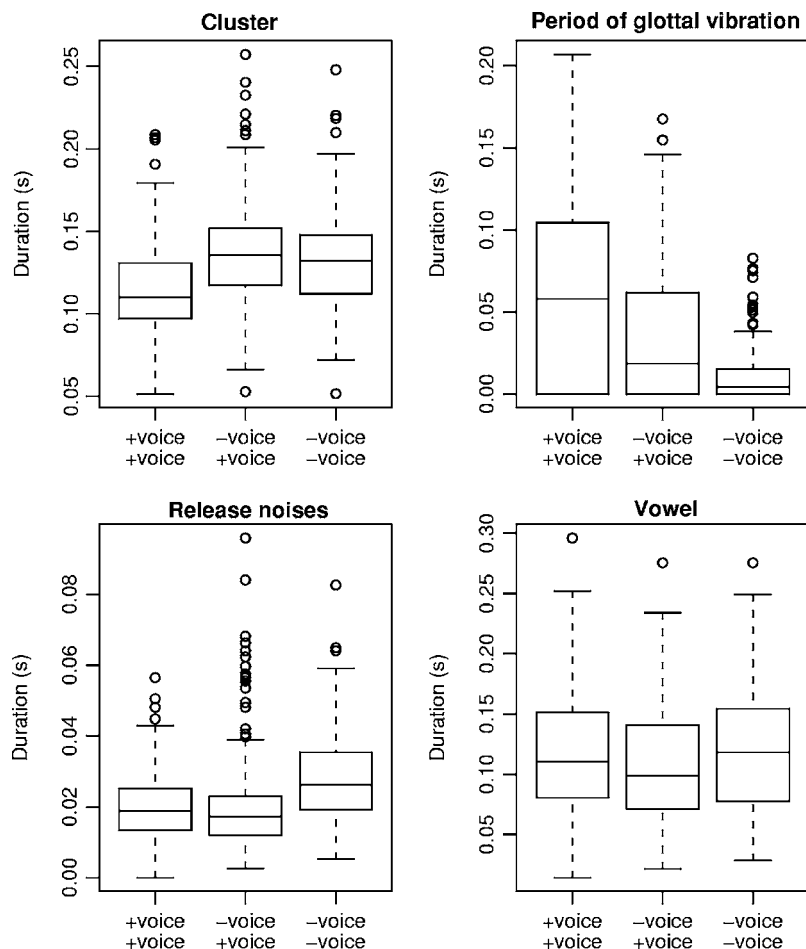


FIG. 4. Distribution of the durations of the obstruent cluster, of the period of glottal vibration, of the total release noise(s), and of the preceding vowel for completely voiced (+voice +voice), unassimilated (-voice +voice), and completely voiceless (-voice -voice) clusters.

for obstruent clusters classified as completely voiced, as unassimilated (-voice +voice), and as completely voiceless.

We investigated the relevance of Frequency, by contrasting unassimilated clusters first with completely voiced and then with completely voiceless clusters. Our independent variables were Frequency, entered as the first predictor, the durations of the preceding vowel, of the period of glottal vibration, of the release noise, and of the cluster. In addition, we included (interactions with) the speakers' Gender and Year of birth (minus 1900), the Type of the first obstruent in the cluster (fricative, plosive), the Second obstruent (/b/, /d/), the Morphological class of the word (prefixed, suffixed, compound), and Phonetician.

The probability ( $\hat{\beta}$ 's for log-odds) that a cluster was classified as completely voiced versus unassimilated decreased with increasing cluster duration [ $\hat{\beta} = -68.68, F(1, 1431) = 153.27, p < 0.0001$ ], increased nonlinearly with the period of glottal vibration [linear:  $\hat{\beta} = -13.38, F(1, 1431) < 1, p > 0.1$ ; quadratic:  $\hat{\beta} = 421.62, F(1, 1431) = 88.71, p < 0.0001$ ], and also increased with the duration of the preceding vowel [ $\hat{\beta} = 11.52, F(1, 1431) = 10.21, p < 0.01$ ]. Furthermore, the phoneticians tended to classify plosive-initial clusters as completely voiced more often (69.84%) than fricative-initial clusters [57.9%,  $F(1, 1431) = 16.40, p < 0.0001$ ]. Finally, Phonetician emerged as significant [ $F(1, 1431) = 27.03, p < 0.0001$ ]. Frequency was not predictive.

The probability that a cluster was classified as completely voiceless versus unassimilated increased with longer release noises [ $\hat{\beta} = 133.38, F(1, 1157) = 86.47, p < 0.0001$ ] while it decreased with longer periods of glottal vibration [ $\hat{\beta} = -40.50, F(1, 1157) = 42.99, p < 0.0001$ ], and with the duration of the preceding vowel [ $\hat{\beta} = -12.56, F(1, 1157) = 12.28, p < 0.0005$ ; note that the sign of the coefficient of this partial effect is negative even though Fig. 4 suggests that it should be positive. This difference in sign is due to by-Word differences that are taken into account in the multilevel analysis but that are ignored in the overall means depicted in Fig. 4]. In addition, the phoneticians tended to classify more fricative-initial clusters (51.7%) than plosive-initial clusters (42.3%) as completely voiceless [ $F(1, 1157) = 8.62, p < 0.005$ ]. Again, Frequency was no longer predictive.

Finally, the probability that a cluster was classified as completely voiceless versus completely voiced decreased with longer periods of glottal vibration [ $\hat{\beta} = -114.02, F(1, 1448) = 100.31, p < 0.0001$ ] and with longer preceding vowels [ $\hat{\beta} = -32.49, F(1, 1448) = 39.49, p < 0.01$ ], but increased with longer release noises [ $\hat{\beta} = 176.04, F(1, 1448) = 229.12, p < 0.0001$ ] and longer obstruent clusters [ $\hat{\beta} = 71.54, F(1, 1448) = 39.49, p < 0.0001$ ]. These results are as expected given the literature on acoustic cues for the perceptual voicing of single obstruents and obstruent clusters (e.g., Slis, 1982; van den Berg 1986, 1987).

### C. Discussion

The durations of the complete cluster, of the period of vocal fold vibration, of the release noises, and of the preceding vowel cued perceived voicing, which is consistent with the observations in the literature (e.g., Slis and Cohen, 1969; van den Berg, 1986, 1987). The distinction between completely voiced and unassimilated clusters was mainly cued by the period of vocal fold vibration, followed by the duration of the complete cluster, as can be seen in Fig. 4. The duration of the release noises was not predictive, probably because plosive-initial clusters tend to be realized without their initial plosives in connected speech. The perceived voicing of the final obstruent was signaled mainly by the period of glottal vibration and by the duration of the release noise(s) (see Fig. 4).

Frequency only has a predictive value for the unassimilated versus completely voiced and completely voiceless classifications if the acoustic characteristics of the speech signal are left out of consideration (compare the models developed in this section with those in Sec. II). This suggests that the correlations between the classifications and frequency observed in Sec. II were carried exclusively by the acoustic cues to voicing. Frequency affected the classifications only indirectly through the acoustic cues, which we predicted given that the phoneticians did not listen to complete words, but to fragments just containing the clusters with the surrounding vowels.

Our finding that frequency is not predictive for the voice classifications over and above the studied acoustic characteristics is also supported by the combination of the models developed in the different sections of this paper. The models showing frequency effects on the acoustic characteristics discussed in Sec. III combined with the acoustic models for perceived voicing (this section) show exactly the same correlations between frequency and the phoneticians' classifications as revealed in Sec. II.

First, we discuss the probability of a completely voiceless cluster versus an unassimilated cluster. This probability is especially high for plosive-initial clusters. It increases with the duration of the release noise(s) and decreases with the period of glottal vibration and with the duration of the preceding vowel. For higher-frequency words, the duration of the release noise(s) is longer and the period of glottal vibration is shorter. The frequency effects on these two cues thus work in the same direction and lead to increased probabilities for completely voiceless clusters in higher-frequency words. Hence, the combined acoustic models lead to exactly the same conclusion as the direct analysis of the role of frequency in the classifications.

Second, we consider the probability of a completely voiced cluster versus an unassimilated cluster. This probability is especially high for fricative-initial clusters. It decreases with cluster duration while it increases with the period of glottal vibration and with vowel duration. As frequency increases, both the cluster and the period of glottal vibration become shorter. The frequency effects on the two cues to voicing therefore work in opposite directions. Hence, the net effect of frequency on the classifications depends on the

strength of the frequency effects on the durations of the cluster and the period of glottal vibration as well as on the contributions of these two durations to perceived voicing.

$$lo = -68.682 * cd - 13.383 * gv + 421.626 * gv^2 \quad (1)$$

$$cd = 0.15 - 0.002 * Freq \quad (2)$$

$$gv = -0.018 + 0.541 * cd + 0.007 * Freq - 0.064 * Freq * cd \quad (3)$$

Equation (1) shows part of the model developed above for the log odds ( $lo$ ) of a completely voiced versus an unassimilated cluster. It shows the coefficients for the durations of the cluster ( $cd$ ) and the period of glottal vibration ( $gv$ ), ignoring the variables that are insensitive to Frequency. Equation (2) is the acoustic model for the duration of /b/-final clusters (Clusters ending in /d/ are predicted to be 4 ms shorter), with  $Freq$  standing for frequency. The model for the period of glottal vibration built in Sec. III contained cluster residuals as predictor. Equation (3) shows the model with raw cluster duration instead of Cluster residual as predictor, for /b/-initial words (in /d/-initial words the period of glottal vibration is predicted to be 7 ms shorter). Combining the three equations and varying Frequency from 0 to 10 reveals an almost linear positive correlation between Frequency and the log-odds: The higher the frequency, the higher the probability of a completely voiced cluster, also exactly as concluded in Sec. II.

### V. GENERAL DISCUSSION

Several studies have documented that segments tend to be shorter in higher-frequency words (e.g., Jurafsky *et al.*, 2001; Bell *et al.*, 2003; Pluymaekers *et al.*, 2005). The present study is the first to report a correlation between lexical frequency and voice assimilation. It is based on a Dutch corpus of read speech. Three phoneticians classified obstruents as voiced, voiceless, or absent in clusters consisting of a morpheme-final obstruent and the voiced initial plosive of the following morpheme. Completely voiced clusters (RVA) as well as completely voiceless clusters (progressive assimilation) occurred more often in higher-frequency words. Unassimilated clusters typically occurred in words of a lower frequency.

The high percentage of words with completely voiceless clusters (25%) shows that progressive voice assimilation also occurs in Dutch word-internal clusters ending in plosives. This is in contrast with the received wisdom according to which progressive voice assimilation in Dutch would be restricted to fricative-final clusters and to clusters ending in the initial voiced plosives of function words (e.g., Demeulemeester, 1962; Zonneveld, 1983; Booij, 1995; Rietveld and van Heuven, 2001; but see Slis, 1986). Thanks to larger corpora, we now have the means to go beyond introspection and laboratory speech, and discover the, sometimes, surprising characteristics of connected speech.

Acoustic analyses showed that as frequency increased, clusters tended to be shorter, to be realized with shorter periods of glottal vibration, and to have longer release noises.

This indicates that speakers tend to realize higher-frequency words with reduced articulatory effort. The probability of a completely voiceless versus unassimilated cluster is correlated with the durations of the period of glottal vibration and the release noises, which both signal less voicing at higher frequencies. The probability of a completely voiced versus unassimilated cluster, in contrast, is co-determined by the period of glottal vibration, which signals less voicing at higher frequencies, and by cluster duration, which signals more voicing in higher-frequency words. The opposite effects of frequency on these acoustic cues may explain why the effect of frequency on the perception of clusters as completely voiced versus unassimilated is small.

The frequency that we studied for each cluster was the sum of the frequencies of all words containing the morphemes that form the cluster. Thus, the frequency studied for the *tb* cluster in *ontbijt* “breakfast” was based on the frequencies of, among others, *ontbijt* itself, *ontbijtje*, and *ontbijtbord*. It might be argued that another frequency measure is in fact more important, namely, the frequencies of the morphemes in the complex words. If these frequencies are low, one might expect less assimilation due to a greater likelihood of compositional production (see also Bien, Levelt, and Baayen, 2005; Hay and Baayen, 2005). We therefore added the frequencies of the individual morphemes (e.g., *ont* and *bijt*), cumulated over all words in which they occur in the CELEX morphological parsings, as predictors to our models for the acoustic measurements. These morpheme frequencies turned out never to be predictive.

The present findings challenge full decomposition models of speech production. In the model by Levelt, Roelofs, and Meyer (1999) for instance, the production of a morphologically complex word such as /vutbal/ “football”, implies the activation of the word’s lemma (“football”), followed by the activation of its morphological constituents (the lexemes /vut/ and /bal/). In turn, lexemes activate their associated phonemes and syllables. Since the model locates frequency effects at the level of the lexeme, it cannot straightforwardly explain a frequency effect of the lemma on clusters consisting of the last phoneme of the first lexeme and the first phoneme of the second lexeme.

The present frequency effects also have implications for our understanding of speech comprehension. Previous research has documented that phoneme transitions that are infrequent within morphemes may be more frequent at morpheme boundaries, and that listeners use low-probability junctural transitions as cues for detecting morphological structure (e.g., Seidenberg, 1987; Hay, 2003). The predictivity of lexical frequency for subtle acoustic details of obstruent clusters documented in the present study shows that subphonemic cues may also contribute to the identification of morphological structure. Morphologically complex words produced with little assimilation at the subphonemic level, that is, with well-articulated obstruent clusters, tend to be less frequent and therefore are more easily recognized via their parts. In contrast, words showing more assimilation between their parts, that is, with hypo-articulated clusters, tend

to be of a higher-frequency and are more likely to be accessed on the basis of their full forms (e.g., Hay and Baayen, 2005).

Finally, the frequency effects documented in this study are completely unexpected given phonological theory, in which assimilation is described in terms of a phoneme sharing its phonological feature with a neighboring phoneme (e.g., Booij, 1995). Our findings show that voice assimilation is graded. Moreover, the different cues to voicing may be affected independently from each other. The literature on acoustic cues to voicing (e.g., Slis and Cohen, 1969; van den Berg, 1986, 1987) in combination with the literature on the relation between lexical frequency and articulatory effort (e.g., Jurafsky *et al.*, 2001; Bell *et al.*, 2003; Pluymaekers *et al.*, 2005) allows us to understand the observed independent and partly opposite effects of frequency on the fine-grained acoustic cues to voicing and their consequences for the perception of voicing in obstruent clusters.

- Baayen, R. H. (2004). “Statistics in psycholinguistics: A critique of some current gold standards,” *Mental Lexicon Working Papers*, pp. 1–45.
- Baayen, R. H., Piepenbrock, R., and Gulikers, L. (1995). *The CELEX Lexical Database (CD-ROM)* (Linguistic Data Consortium, University of Pennsylvania, Philadelphia, PA).
- Baayen, R. H., Tweedie, F. J., and Schreuder, R. (2002). “The subjects as a simple random effect fallacy: Subject variability and morphological family effects in the mental lexicon,” *Brain Lang* **81**, 55–65.
- Bates, D. M., and Sarkar, D. (2005). *The lme4 library*, on-line available: <http://lib.stat.cmu.edu/R/CRAN/>.
- Bell, A., Jurafsky, D., Fosler-Lussier, E., Girand, C., Gregory, M., and Gildea, D. (2003). “Effects of disfluencies, predictability, and utterance position on word form variation in English conversation,” *J. Acoust. Soc. Am.* **113**, 1001–1024.
- Bien, H., Levelt, W. M. J., and Baayen, R. H. (2005). “Frequency effects in compound production,” *PNAS* **102**, 17876–17881.
- Booij, G. E. (1995). *The Phonology of Dutch* (Clarendon, Oxford).
- Browman, C. P., and Goldstein, L. (1990). “Tiers in articulatory phonology, with some implications for casual speech,” in *Between the Grammar and Physics of Speech*, Papers in Laboratory Phonology Vol. 1, edited by J. Kingston and M. E. Beckman (Cambridge University Press, Cambridge), pp. 341–376.
- Bybee, J. L. (2001). *Phonology and Language Use* (Cambridge University Press, Cambridge).
- Chatterjee, S., Hadi, A., and Price, B. (2000). *Regression Analysis by Example* (Wiley, New York).
- Collins, B., and Mees, I. (1981). *The Sounds of English and Dutch* (Leiden University Press, Leiden).
- Coussé, E., Gillis, S., Kloots, H., and Swerts, M. (2004). “The influence of the labeller’s regional background on phonetic transcriptions: Implications for the evaluation of spoken language resources,” in *Proceedings of the Fourth International Conference on Language Resources and Evaluations*, edited by M. T. Lino, M. F. Xavier, F. Fereirra, R. Costa, and R. Silva, pp. 1447–1450.
- Demeulemeester, F. (1962). “Assimilatie van stemloze explosieven en fricatieven voor b en d” (“Assimilation of voiceless plosives and fricatives before b and d”), *Taal en Tongval* **14**, 20–36.
- Ernestus, M. (2000). *Voice Assimilation and Segment Reduction in Casual Dutch: A Corpus-Based Study of the Phonology-Phonetics Interface* (LOT, Utrecht).
- Fidelholz, J. (1975). “Word frequency and vowel reduction in English,” in *Papers from the 11th Regional Meeting of the Chicago Linguistic Society*, edited by R. E. Grossman, L. J. San, and T. J. Vance, Vol. **11**, pp. 200–213.
- Gussenhoven, C., and Bremmer, R. H. J. (1983). “Voiced fricatives in Dutch: Sources and present-day usage,” *North-Western Europ. Lang. Evol.* **2**, 55–71.
- Hay, J. (2003). *Causes and Consequences of Word Structure* (Routledge, New York).
- Hay, J. B., and Baayen, R. H. (2005). “Shifting paradigms: Gradient structure in morphology,” *Trends in Cognitive Science* **9**, 342–348.

- Johnson, K. (2004). "Massive reduction in conversational American English" in *Spontaneous Speech: Data and Analysis, Proceedings of the First Session of the Tenth International Symposium* (The National International Institute for Japanese Language, Tokyo, Japan), pp. 29–54.
- Jurafsky, D., Bell, A., Gregory, M., and Raymond, W. D. (2001). "Probabilistic relations between words: Evidence from reduction in lexical production," in *Frequency and the Emergence of Linguistic Structure*, edited by J. L. Bybee and P. Hopper (Benjamins, Amsterdam), pp. 229–254.
- Kaiser, L. (1958). "Onderzoek naar assimilatieverschijnselen" ("Research in assimilation phenomena"), in *Album Edgard Blancquaert: De Gehuldigde Aangeboden ter Gelegenheid Van Zijn Emeritaat Door Kollega's Vakgenoten en Oud-Leerlingen* (George Michels, Tongeren), pp. 31–41.
- Levelt, W. J. M., Roelofs, A., and Meyer, A. S. (1999). "A theory of lexical access in speech production," *Behav. Brain Sci.* **22**, 1–38.
- Loots, W. (1983). "Syntax and assimilation of voice in Dutch," in *Sound Structures: Studies for Antonie Cohen*, edited by M. van den Broecke, V. van Heuven, and W. Zonneveld (Foris, Dordrecht), pp. 173–180.
- Lucero, J. C., and Koenig, L. L. (2005). "Phonation thresholds as a function of laryngeal size in a two-mass model of the vocal folds (L)," *J. Acoust. Soc. Am.* **118**, 2798–2801.
- Marslen-Wilson, W., Nix, A., and Gaskell, G. (1995). "Phonological variation in lexical access: Abstractness, inference and English place assimilation," *Lang. Cognit. Processes* **10**, 285–308.
- Menert, L. (1994). "Experiments on voice assimilation in Dutch: Prosodic structures and tempo," Ph.D. dissertation, University of Utrecht, Utrecht.
- Miller, J. L. (1981). "Some effects of speaking rate on phonetic perception," *Phonetica* **38**, 159–180.
- Mitterer, H., and Ernestus, M. (2006). "Listeners recover /t/ that speakers reduce: Evidence from /t/-lenition in Dutch," *J. Phonetics*, **34**, 73–103.
- Ohala, J. J. (1983). "The origin of sound patterns in vocal tract constraints," in *The Production of Speech*, edited by P. F. MacNeilage (Springer, Berlin), pp. 189–216.
- Oostdijk, N. (2000). "The Spoken Dutch Corpus Project," *The ELRA Newsletter* **5**, 4–8.
- Oostdijk, N., Goedertier, W., Eynde, F. van, Boves, L., Martens, J., Moortgat, M., and Baayen, R. H. (2002). "Experiences from the Spoken Dutch Corpus Project," in *Proceedings of the Third International Conference on Language Resources and Evaluation*, edited by M. González Rodríguez and C. Paz Suarez Araujo, pp. 340–347.
- Pinheiro, J. C., and Bates, D. M. (2000). *Mixed-effects Models in S and S-PLUS* (Statistics and Computing, Springer, New York).
- Pluymaekers, M., Ernestus, M., and Baayen, R. H. (2005). "Lexical frequency and acoustic reduction in spoken Dutch," *J. Acoust. Soc. Am.* **118**, 2561–2569.
- Quené, H., and van der Bergh, H. (2004). "On multi-level modeling of data from repeated measures designs: A tutorial," *Speech Commun.* **43** 103–121.
- Rietveld, A., and van Heuven, V. (2001). *Algemene Fonetiek (General Phonetics)* (Coutinho, Bussum).
- Seidenberg, M. (1987). "Sublexical structures in visual word recognition: Access units or orthographic redundancy," in *Attention and Performance XII*, edited by M. Coltheart (Erlbaum, Hove), pp. 245–264.
- Slis, I. H. (1982). "Assimilatie van stem in het Nederlands" ("Assimilation of voice in Dutch"), *Glott* **5**, 235–261.
- Slis, I. H. (1986). "Assimilation of voice in Dutch as a function of stress, word boundaries, and sex of speaker and listener," *J. Phonetics* **14**, 311–326.
- Slis, I. H., and Cohen, A. (1969). "On the complex regulating the voiced-voiceless distinction, II," *Lang Speech* **12**, 80–102; (1969) **12**, 137–155.
- van Alphen, P. M., and Smits, R. (2004). "Acoustical and perceptual analysis of the voicing distinction in Dutch initial plosives: The role of prevoicing," *J. Phonetics* **32**, 455–491.
- van den Berg, R. (1986). "The effect of varying voice and noise parameters on the perception of voicing in Dutch two-obstruent sequences," *Speech Commun.* **5**, 355–367.
- van den Berg, R. (1987). "Effects of durational factors on the perception of voicing in Dutch two-obstruent sequences," *J. Phonetics* **15**, 259–271.
- Venables, W. N., and Ripley, B. D. (2002). *Modern Applied Statistics with S-Plus* (Springer, New York).
- Vierregge, W. H. (1987). "Basic aspects of phonetic segmental transcription," in *Probleme der Phonetischen Transkription (Problems with Phonetic Transcriptions)*, edited by A. Almeida and A. Braun (Franz Steiner Verlag Wiesbaden GMBH, Stuttgart), pp. 5–55.
- Wetzels, W., and Mascaró, J. (2001). "The typology of voicing and devoicing," *Lang.* **77**, 207–244.
- Wiese, R. (1996). *The Phonology of German* (Clarendon, Oxford).
- Zipf, G. K. (1935). *The Psycho-Biology of Language* (Houghton Mifflin, Boston).
- Zonneveld, W. (1983). "Lexical and phonological properties of Dutch devoicing assimilation," in *Sound structures: Studies for Antonie Cohen*, edited by M. van den Broecke, V. van Heuven, and W. Zonneveld (Foris, Dordrecht), pp. 297–312.