

Lexical frequency and acoustic reduction in spoken Dutch

Mark Pluymaekers

Radboud University Nijmegen

Wundtlaan 1, 6525 XD Nijmegen, The Netherlands

Mirjam Ernestus

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

(25th July 2005)

Running title: Lexical frequency and acoustic reduction

ABSTRACT

This study investigates the effects of lexical frequency on the durational reduction of morphologically complex words in spoken Dutch. The hypothesis that high-frequency words are more reduced than low-frequency words was tested by comparing the durations of affixes occurring in different carrier words. Four Dutch affixes were investigated, each occurring in a large number of words with different frequencies. The materials came from a large database of face-to-face conversations. For each word containing a target affix, one token was randomly selected for acoustic analysis. Measurements were made of the duration of the affix as a whole and the durations of the individual segments in the affix. For three of the four affixes, a higher frequency of the carrier word led to shorter realizations of the affix as a whole, individual segments in the affix, or both. Other relevant factors were the sex and age of the speaker, segmental context, and speech rate. To accommodate for these findings, models of speech production should allow word frequency to affect the acoustic realizations of lower-level units, such as individual speech sounds occurring in affixes.

PACS numbers: 43.70.Bk, 43.70.Fq

I. INTRODUCTION

In everyday speech, words are often pronounced shorter than their citation forms would suggest. This is not a marginal phenomenon: In Johnson's (2004) study on conversational American English, 25% of the words had one or more segments deleted. Deletion of complete syllables occurred in 6% of the words. Similar observations were made by Ernestus (2000) for Dutch. For example, the word *natuurlijk* ('of course') was sometimes reduced to [tyk]. Despite the frequent nature of these reductions, their presence has not yet been accommodated in any of the main psycholinguistic theories (e.g., Garrett, 1975; Dell, 1986; Levelt, 1989; Bock, 1995).

Reductions have often been linked to word frequency (e.g., Jespersen, 1922; Zipf, 1929), the hypothesis being that high-frequency words are more reduced than low-frequency words. Several explanations have been offered for this relationship, such as the compression of motor routines as a result of practice (Bybee, 2001), or the fact that high-frequency words are more predictable for the listener (e.g., Jurafsky, Bell, Gregory, & Raymond, 2001). Many studies have confirmed the pivotal role of word frequency in predicting diachronic phonetic abbreviations (Zipf, 1929; Bybee, 2001). It has proven more difficult, however, to demonstrate synchronic effects of frequency on acoustic realizations.

The main problem lies in the lack of suitable reference material. Since words generally differ not only in frequency, but also in at least one of their speech sounds, they are bound to differ in duration as well. Therefore, most authors have restricted themselves to comparing instances of the same phoneme occurring in different words. Umeda (1977) found that in American English, word-initial [s]-es were shorter if the frequency of their carrier word was high. Likewise, Cooper and Paccia-Cooper (1980) showed that palatalization of [d] before [j] was more likely in high-frequency than in low-frequency words. Van Coile (1987) used word frequency as a criterion to distinguish between function words and content words in Dutch, and found that vowels occurring in function words were shorter than the same vowels occurring in

content words. Finally, Jurafsky *et al.* (2001) compared tokens of word-final [d] and [t] occurring in English words with different frequencies. In words with a high frequency, the plosive had a greater chance of being deleted, and if it was present in the signal, its duration was significantly shorter.

Over the years, attempts have been made to demonstrate frequency effects on units larger than the phoneme as well. Wright (1979) used pairs of rare and common words matched on length in letters, but, as most speech researchers will agree, this type of matching does not offer enough experimental control for comparing durations. Gregory, Raymond, Bell, Fosler-Lussier, & Jurafsky (1999) measured the durations of a large number of words ending in *-t* or *-d*, and found an effect of word frequency on these durations. However, their target words probably differed on other dimensions as well, such as the number of phonemes and their complexity (Landauer & Streeter, 1973). Therefore, the evidence for effects of word frequency on the durations of larger linguistic units remains inconclusive.

To overcome the difficulties sketched above, we decided to focus on morphemes that can occur in a large number of words with different frequencies: affixes. This approach is similar to that of Aylett & Turk (2004), who compared syllables occurring in different words. The main difference lies in the fact that affixes by definition carry meaning, while for syllables this is not necessarily the case. An additional advantage of affixes is that most of them never bear stress, providing us with valuable control over factors like word stress and sentence accent.

Most studies on word frequency and reduction used English materials. Recently, a new database of spoken Dutch has become available, containing a large section of spontaneous, face-to-face conversations. This provided us with an excellent opportunity to investigate the effects of frequency on acoustic reduction in a language other than English.

In summary, this study investigates the effects of word frequency on the durations of Dutch affixes. Durational shortening is of course not the only acoustic correlate

of reduction, but the nature of the materials (spontaneous, overlapping speech) precluded us from studying variables such as mean amplitude (e.g., Shields & Balota, 1991) or Center of Gravity (e.g., Van Son & Pols, 2003). Section II describes our materials and method of measurement. Section III is concerned with the control variables we incorporated. Sections IV to VII contain the results for the different affixes. Section VIII, finally, discusses the results and their implications for future research.

II. METHODOLOGY

A. Materials

All materials were drawn from the Corpus of Spoken Dutch (Oostdijk, 2000). This corpus contains approximately 800 hours of speech recordings, of which only the 225 hours of spontaneous, face-to-face conversations were considered for the present study. Orthographic transcriptions are available for the entire corpus. We restricted ourselves to Dutch speakers, since they have been shown to use reduced forms more than speakers from Flanders (Keune, Ernestus, Van Hout, & Baayen, *in press*; see Adank, Van Hout, & Smits, 2004, for other acoustic differences between the two varieties of Dutch).

The affixes under investigation were the prefixes *ge-*, *ver-*, and *ont-*, and the suffix *-lijk*. *Ge-* is used mainly to create the perfect participle in Dutch, although it can also function as a nominal or verbal prefix. In this study, we restricted ourselves to the participial use of *ge-*. *Ver-* and *ont-* are verbalizing prefixes expressing states of change (*ver-*) and reversal or inchoation (*ont-*). For example, *ver-* + *plaats* ‘place’ gives *verplaatsen* ‘to move’, and *ont-* + *eigen* ‘own’ gives *onteigenen* ‘to disown’. The suffix *-lijk* can be found in adverbs and adjectives (e.g., *natuurlijk*, ‘natural(ly)’, and *eigenlijk*, ‘actual(ly)’). The citation forms of these four affixes are [xə], [vər], [ɔnt], and [lək].

For each of the affixes, a randomized list was made of all occurrences in the corpus. For each word type containing a target affix, the first token on the list was selected for further analysis. If the quality of the recording was too poor for acoustic analysis, it was replaced with the next token on the list. We considered as word types not only words belonging to different lemmas, but also different word forms of the same lemma. Thus, the sample for *ont-* included both *ontwikkelt* (‘develops’) and *ontwikkelde* (‘developed’).

B. Measurements

Acoustic measurements of the target words were made using the software package PRAAT (Boersma, 2001). For all words, we measured the duration of the affix and the durations of the individual segments in the affix (both in milliseconds). Since the amount of background noise differed considerably between tokens, it was hard to establish a general segmentation strategy (see also Vorstermans, Martens, & Van Coile, 1996). Figure 1 shows the manual segmentations for the prefix *ont-* in the tokens *ontwaken* ‘to wake’ (top), and *ontwijken* ‘to avoid’ (bottom), including the previous word and the first syllable of the stem. *Ontwaken* was relatively easy to segment, since there was hardly any background noise and no overlapping speech around the prefix. The sample for *ontwijken* contained more background noise, resulting in a waveform in which the different segments could not be easily distinguished. In all cases, we placed the segment boundaries where we found clear formant transitions in the spectrogram supported by visible changes in the waveform pattern.

INSERT FIGURE 1 APPROXIMATELY HERE

III. CONTROL VARIABLES

A. Probabilistic measures

Besides frequency, other measures of word probability are known to affect acoustic realizations as well. Fowler & Housum (1987) found that the second realization of a word in a monologue was shorter than the first one. Bard, Anderson, Sotillo, Aylett, Doherty-Sneddon, & Newlands (2000) replicated this effect for dialogues, showing that it was present irrespective of whether the speaker or the listener uttered the first token of the word. To check whether our target words might be subject to repetition effects, we counted how often the target word (or a word from the same inflectional paradigm) had been used in the conversation prior to the occurrence of the selected token. Since most tokens turned out to be first occurrences, this factor was not included in the final analyses. In addition, we counted how often the affix under investigation had already occurred. This variable, which varied in value between 0 and 72, turned out to have no effect.

The probability of occurrence of a word also depends on neighbouring words. In recent years, numerous studies have addressed the relationship between predictability from neighbouring words and acoustic reduction (e.g., Hunnicutt, 1985; Fosler-Lussier & Morgan, 1999; Gregory *et al.*, 1999; Bush, 2001; Jurafsky *et al.*, 2001; Bell, Jurafsky, Fosler-Lussier, Girand, Gregory, & Gildea, 2003). To determine the predictability of their target words, most authors have used measures like *conditional probability* or *mutual information*, which are computed using frequency estimates from large speech corpora. Both measures capture the likelihood of a certain word occurring given one or more of its neighbouring words. Mutual Information is arguably the most elegant of the two, as it combines conditional probability with the frequency of the word itself. The corresponding equation is as follows (X and Y denote either the previous word and the target word or the target word and the following word):

$$MI(X; Y) = \log \frac{(\text{Frequency}(XY))}{(\text{Frequency}(X)) * (\text{Frequency}(Y))}$$

From this equation, it is easy to see why Mutual Information could also be relevant for our purposes. Since word frequency is incorporated in the denominator, any effect observed for frequency might in fact be an effect of Mutual Information in disguise. Our sampling method prevented us from computing Mutual Information values for all of our target words, as some of them were at the beginning or end of utterances. For those words for which Mutual Information could be computed, we checked whether it was a better predictor of duration than word frequency alone. This was never the case.

As we are dealing with morphologically complex words, probabilistic variables affecting morphological processing should also be taken into account. Hay (2003) found that derived words that are more frequent than their stems are judged less morphologically complex by language users. This suggests that speakers may only consider an affix a separate morpheme if the stem is at least as frequent as the combination of stem and affix. If on the other hand the affix-stem combination is more frequent, the word is more likely to be accessed as a whole, implying less psychological reality for the affix. Since Hay showed that this perceived morphological complexity can also affect acoustic realizations, we included the ratio between the frequency of the target word and that of its stem (Word-Stem Ratio) in the analyses.

All frequency estimates, including the ones used to compute Mutual Information and Word-Stem Ratio, were taken from the Corpus of Spoken Dutch and logarithmically transformed.

B. Other control factors

This section discusses the non-probabilistic variables that were incorporated in this study. When rate of speech is high, words have a higher probability of deviating from the standard (Fosler-Lussier & Morgan, 1999). We estimated speech rate

by computing the number of syllables per second in the longest stretch of speech containing the target word that did not contain an audible pause. For all words in the utterance except the target word, the number of syllables was determined on the basis of the orthographic transcription. For the target word, we used the information from the manual transcription instead.

Sociolinguistic variables such as sex, age, and regional origin of the speaker also have a considerable impact on pronunciation (e.g., Byrd, 1994; Keune *et al.*, *in press*). Information about these three factors was gathered for all speakers. Age was operationalized by subtracting 1900 from the Year of Birth of the speaker.

Words that are positioned at the beginning of an utterance are often acoustically strengthened, while words at the end of an intonational phrase can show durational lengthening (e.g., Fougeron & Keating, 1997; Cambier-Langeveld, 2000; Bell *et al.*, 2003). This was controlled for by coding all target words for position in their carrier utterance (beginning, end, or middle).

Fox Tree & Clark (1997) and Bell *et al.* (2003) showed that words that occur near disfluencies are lengthened compared to words occurring in fluent contexts. Therefore, the presence of a false start or filled pause directly before or after the target word was also coded.

Not all phonetic environments are equally suitable for reduction. Zsiga (1994) found that word-final consonants are more likely to be reduced if they are followed by another consonant. We determined for each token whether the segment following the affix was a consonant or a vowel. For the prefixed words, we also counted the number of consonants in the onset of the stem (henceforth referred to as Onset Complexity).

Finally, the absence of certain segments in the affix was sometimes included as an extra factor in the analyses. If, for example, the final segment in the affix is absent, this may have implications for the durations of the other segments, as well as for the duration of the affix as a whole. Segments were considered as absent if they could not be isolated in the acoustic signal.

To evaluate the effects of word frequency on duration while controlling for all other possibly relevant factors, we used least squares regression. The application of this method on speech data is described in detail by Bell *et al.* (2003). The signs of the reported beta coefficients indicate whether there was a positive or a negative correlation between two variables. Since we do not want to report effects that depend crucially on a single data point, we excluded observations that were outliers with regard to leverage or Cook’s distance values (Chatterjee, Hadi, & Price, 2000).

IV. RESULTS FOR GE-

A. Duration of the prefix as a whole

For *ge-*, the sample consisted of 428 words uttered by 132 different speakers. No speaker contributed more than 12 tokens to the sample. Broad phonetic transcriptions of the encountered realizations are [xə], [x], [χə], and [χ].

We fitted a stepwise multiple regression model to the data with the duration of *ge-* as the response variable. There were four outliers, which were removed. We found main effects of Frequency ($\hat{\beta} = -4.1, t(420) = -3.05, p < 0.005$), Speech Rate ($\hat{\beta} = -8.6, t(420) = -5.62, p < 0.0001$), and Onset Complexity ($\hat{\beta} = -7.3, t(420) = -2.03, p < 0.05$). Words with a higher frequency had shorter realizations of *ge-*. When Speech Rate was high, the prefix was also shorter, as was the case if it was followed by a large number of consonants. The amount of variance explained by this model (also referred to as R^2) was 10%.

B. Durations of the individual segments

To gain more insight into the articulatory dynamics underlying the above-mentioned effects, separate models were fitted for the two segments in *ge-*.

For the fricative, a model was fitted to the entire data set, including the data points that were outliers in the model for the entire prefix. After the removal of three new outliers, there were main effects of Frequency ($\hat{\beta} = -3.6, t(421) = -3.71, p < 0.0005$), Speech Rate ($\hat{\beta} = -3.5, t(421) = -3.79, p < 0.0005$), and Word-Stem Ratio ($\hat{\beta} = 34.7, t(421) = 2.46, p < 0.05$). This model explained 7% of the variance.

The vowel was present in 414 tokens (97%). Four outliers were removed. Vowel duration was predicted by Frequency ($\hat{\beta} = -2.1, t(406) = -2.31, p < 0.05$), Speech Rate ($\hat{\beta} = -3.0, t(406) = -2.93, p < 0.005$), and Initial Position ($\hat{\beta} = 14.5, t(406) = 2.45, p < 0.05$). The vowel was longer in words that were in Initial Position. The R^2 of this model was 6%.

C. Discussion of the results for *ge-*

INSERT TABLE I APPROXIMATELY HERE

The results for *ge-* are summarized in Table I. The observed effects of Frequency, Speech Rate, Initial Position, and Onset Complexity all went in the expected direction. The Frequency effect was present for both the fricative and the vowel, as was the effect of Speech Rate.

It might seem counter-intuitive that a higher ratio between the frequency of the word and the frequency of its stem should lead to longer fricatives. After all, a higher value of this ratio is supposed to be associated with less morphological complexity, and hence, less psychological reality for the prefix. We return to this issue in our general discussion in Section VIII.

V. RESULTS FOR *ONT-*

A. Duration of the prefix as a whole

There were 102 word types starting with *ont-* in the corpus. The tokens in the sample were uttered by 63 different speakers, who contributed no more than four

tokens each to the data set. The realizations we encountered ranged from canonical [ɔnt] to highly reduced [ət].

A model was fitted to the data with duration of *ont*- as the response variable. Three outliers were removed. Prefix duration was predicted by Year of Birth ($\hat{\beta} = -1.4, t(95) = -4.81, p < 0.0001$). Younger speakers produced shorter prefixes. Frequency was not significant as a main effect, but it was in interaction with Speech Rate ($\hat{\beta} = -2.9, t(95) = -3.31, p < 0.005$) and Year of Birth ($\hat{\beta} = 0.2, t(95) = 2.84, p < 0.01$). The interaction between Frequency and Speech Rate is shown in Figure 2. Frequency had either a lengthening or no effect when Speech Rate was low (the bottom left and middle panels), a shortening effect when Speech Rate was neither low nor high, and no effect when Speech Rate was extremely high (the top right panel). In Figure 3, the interaction between Frequency and Year of Birth is illustrated. For the youngest speakers (the top middle and right panels) the effect of Frequency was absent, whereas for the other age groups a higher Frequency correlated with shorter realizations. All in all, this model accounted for 24% of the variance.

INSERT FIGURES 2 AND 3 APPROXIMATELY HERE

B. Durations of the individual segments

Since the vowel was present in all 102 tokens, we fitted a model for vowel duration to the entire data set. Three observations were identified as outliers and removed. Younger speakers produced shorter vowels ($\hat{\beta} = -0.3, t(96) = -2.19, p < 0.05$), while women’s vowels were longer ($\hat{\beta} = 8.7, t(96) = 2.33, p < 0.05$). The R^2 of this model was 10%.

The nasal was produced in 97 tokens, three of which were outliers. The duration of the nasal was affected by the Presence of the Plosive ($\hat{\beta} = -34.2, t(89) = -5.58, p < 0.0001$) and Year of Birth ($\hat{\beta} = -0.2, t(89) = -2.74, p < 0.01$). Younger speakers

produced shorter nasals, and if the plosive was absent the nasal was longer. We also found a significant interaction between the Presence of the Plosive and Frequency ($\hat{\beta} = 5.9, t(89) = 3.67, p < 0.0005$). Frequency was only significant if there was no plosive. Together, these three predictors explained 48% of the variance.

Finally, we fitted a model for the duration of the plosive. Three outliers were removed. There were main effects of Frequency ($\hat{\beta} = -17.8, t(66) = -3.50, p < 0.001$), Year of Birth ($\hat{\beta} = -1.0, t(66) = -3.87, p < 0.0005$), and Speech Rate ($\hat{\beta} = -3.6, t(66) = -2.39, p < 0.05$). All effects went in the expected direction. Furthermore, there was an interaction between Frequency and Year of Birth ($\hat{\beta} = 0.3, t(66) = 3.30, p < 0.005$), which was similar to the one observed for the entire prefix (see Figure 3). In total, 28% of the variance in the duration of the plosive was explained by this model.

C. Discussion of the results for *ont-*

INSERT TABLE II APPROXIMATELY HERE

Table II provides an outline of the results for *ont-*. There were Frequency effects in the expected direction for the duration of the plosive if it was present and for the duration of the nasal if the plosive was not present. This did not lead to a main effect of Frequency for the prefix as a whole, but there were two significant interactions.

The interaction with Speech Rate suggested that the effect of Frequency was limited to situations in which Speech Rate was not extremely low or high. The absence of a Frequency effect when Speech Rate is high can be explained by assuming that speakers try to avoid complete deletion of the prefix, since this may hamper communication. This is confirmed by the fact that no tokens were encountered in which *ont-* was completely absent. When Speech Rate is low, on the other hand, there is less need to reduce articulatory effort, which also diminishes the probability of finding a Frequency effect.

The interaction with Year of Birth showed that the Frequency effect was not present for the youngest speakers. This can be accounted for by the finding that younger speakers already produced shorter realizations (as evidenced by the negative main effects of Year of Birth observed in all models fitted for *ont-*).

VI. RESULTS FOR VER-

A. Duration of the prefix as a whole

The prefix *ver-* occurred in 140 different word types produced by 82 different speakers. The maximum number of tokens uttered by a single speaker was eight. Observed pronunciations included [vər], [və], and [f].

We fitted a model to predict the duration of *ver-*. After removing three outliers, there were significant main effects of Year of Birth ($\hat{\beta} = -0.5, t(134) = -2.58, p < 0.05$) and Onset Complexity ($\hat{\beta} = -14.7, t(134) = -2.64, p < 0.01$). Younger speakers produced shorter prefixes. If the number of consonants in the onset of the stem was high, the prefix was shorter as well. The R^2 of this model was 11%.

B. Durations of the individual segments

We fitted separate models only for the fricative and the rime (i.e., the combination of the vowel and [r]), since the vowel and [r] (if present) could not be reliably distinguished.

The fricative was present in all cases. Four outliers were removed. We found main effects of Onset Complexity ($\hat{\beta} = -7.4, t(133) = -2.37, p < 0.05$) and Sex of the speaker ($\hat{\beta} = -12.1, t(133) = -3.43, p < 0.001$). Women produced shorter fricatives. These variables explained 12% of the variance.

For the rime, a model was fitted to the 117 data points for which it was present. Three outliers were removed. There were main effects of Onset Complexity ($\hat{\beta} =$

$-12.8, t(111) = -3.49, p < 0.001$) and Year of Birth ($\hat{\beta} = -0.3, t(111) = -2.52, p < 0.05$), together explaining 17% of the variance.

C. Discussion of the results for *ver-*

Beta coefficients and significance values of the effects for *ver-* are given in Table III. For this prefix, there were no effects of Frequency. The effect of Onset Complexity (the higher the number of consonants in the onset of the stem, the shorter the prefix) can be traced back to both the fricative and the rime. As was the case for the other prefixes, younger speakers produced shorter realizations of *ver-*. This effect was mainly due to durational shortening of the rime.

INSERT TABLE III APPROXIMATELY HERE

VII. RESULTS FOR -LIJK

A. Duration of the suffix as a whole

The data set for the suffix *-lijk* consisted of 158 tokens, uttered by 88 different speakers. No speaker contributed more than six tokens to the data set. The realizations we observed ranged from the citation form [lək] to [ə] or [k]. Because of the semantic opacity of many of the words containing *-lijk*, the Word-Stem Ratio we discussed in Section III was not included in the analyses.

Again, we first fitted a model to predict the duration of the affix as a whole. Visual inspection of the fitted model revealed that the variance in the residuals was much larger for words in Final Position than for words in Non-Final Position ($F(42, 114) = 4.33, p < 0.0001$). Therefore, separate models were fitted for words in Final and in Non-Final Position. For words in Final Position (43 observations), four outliers were removed. The data set for Non-Final words contained 115 tokens, four of which were outliers.

For words in Non-Final Position, there were main effects of Frequency ($\hat{\beta} = -7.7, t(108) = -3.73, p < 0.0005$) and Year of Birth ($\hat{\beta} = -0.7, t(108) = -3.12, p < 0.005$), explaining 18% of the variance.

If the word was in Final Position, the duration of *-lijk* was affected by Speech Rate ($\hat{\beta} = -36.4, t(36) = -4.24, p < 0.0005$) and the Presence of the Plosive ($\hat{\beta} = 147.5, t(36) = 3.23, p < 0.005$). If the plosive was absent, the suffix was shorter. The R^2 of this model was 47%.

B. Durations of the individual segments

The [l] was produced in 140 tokens. This time, a split on the basis of Position was not necessary, as visual inspection of the residuals did not reveal any abnormalities. We removed four outliers and found main effects of Frequency ($\hat{\beta} = -2.1, t(130) = -2.95, p < 0.005$), Speech Rate ($\hat{\beta} = -4.2, t(130) = -4.14, p < 0.0001$), Year of Birth ($\hat{\beta} = -0.2, t(130) = -2.96, p < 0.005$), and Final Position ($\hat{\beta} = 10.0, t(130) = 3.66, p < 0.0005$). All effects went in the expected direction. Together, they accounted for 32% of the variance.

The vowel was realized in all but eight tokens. Three outliers were removed. Vowel duration was affected by Speech Rate ($\hat{\beta} = -4.2, t(143) = -2.77, p < 0.01$), Final Position ($\hat{\beta} = 12.2, t(143) = 3.03, p < 0.005$), and the Presence of the Plosive ($\hat{\beta} = -30.6, t(143) = -4.77, p < 0.0001$). If the plosive was absent, the vowel was longer. The R^2 of this model was 23%.

For the plosive, the situation was similar to that of the entire suffix. The variances of the residuals of the initially fitted model differed significantly between tokens that were in Final and Non-Final Position ($F(38, 105) = 7.27, p < 0.0001$). Therefore, we fitted separate models for the 39 Final and 106 Non-Final plosives.

For the Non-Final plosives, three outliers were removed. The only significant effect we found was one of Following Disfluency: if there was no disfluency following the target word, the plosive was shorter ($\hat{\beta} = -66.2, t(101) = -5.73, p < 0.0001$).

By itself, this factor accounted for 25% of the variance. For the Final plosives, we removed three outliers and found an effect of Speech Rate ($\hat{\beta} = -22.9, t(34) = -3.28, p < 0.005$) in the expected direction, explaining 24% of the variance.

C. Discussion of the results for -lijk

INSERT TABLE IV APPROXIMATELY HERE

An outline of the results for *-lijk* is given in Table IV. The duration of *-lijk* was most strongly affected by whether the word was in Final Position or not. For the suffix as a whole and the plosive, this effect was so pervasive that separate models had to be fitted to the Final and the Non-Final data points. For the other two segments, such drastic measures were not necessary, although Final Position remained a significant predictor. The effect always went in the same direction: Final words had longer realizations of *-lijk*. This could be explained by referring to the well-documented phonetic effect of phrase-final lengthening (e.g., Fougeron & Keating, 1997), but we feel that such an explanation would be too restricted. The majority (70%) of the words in Final Position were predicates in an utterance of the type ‘that is [Adjective]’. When saying something like *dat is belachelijk* ‘that is ridiculous’, it would make no sense to reduce articulatory effort on the word *belachelijk*, as it is the only information carrier in the utterance. This might also be reflected in the presence of sentence accent, leading to durational lengthening (e.g., Nooteboom, 1972).

Despite the dominance of Final Position as a predictor, we still found effects of Frequency, although these were restricted to the Non-Final suffixes and realizations of [l]. Speech Rate was significant for all durations except that of the Non-Final plosive. The effect of Following Disfluency we found was in line with the earlier findings of Fox Tree & Clark (1997) and Bell *et al.* (2003).

Interestingly, the two significant effects for Presence of the Plosive went in opposite directions. If the plosive was absent, Final suffixes were shorter, but vowels were

longer. It might be the case that speakers more or less compensate for the absence of the plosive by lengthening the vowel. For the subset of Final suffixes, this effect might not have surfaced since the entire suffix was lengthened.

VIII. GENERAL DISCUSSION

This study provides strong evidence for the relationship between lexical frequency and acoustic reduction. For the Dutch affixes *ge-*, *ont-*, and *-lijk*, we found effects of frequency on the durations of individual segments, the affix as a whole, or both. Apparently, the effect of word frequency in speech production is not restricted to the speed of lexical retrieval; it manifests itself in the subtle acoustic details of the word as well. This lends further support to the probabilistic framework developed by Bybee (2001), Pierrehumbert (2003), and others. They view probabilistic information as an integral part of our linguistic knowledge, exerting its influence at every level of language processing, including articulation.

How can these findings be incorporated in models of speech production? Most current theories are based on either speech errors or reaction time data, and have not been concerned with fine-grained differences in articulation. Nevertheless, models like the one proposed by Levelt, Roelofs, & Meyer (1999) can be modified in such a way that they can account for the results of the current study. One possible modification is the inclusion of reduced word forms in the lexicon, which are selected if the conceptual structure of the message specifies a word as redundant. Although this might work for obviously reduced forms such as [tyk] for *natuurlijk* ‘of course’, it seems less appropriate to include all possible durational variants of words in the lexicon. A more efficient solution is to pass information about redundancy on to the Articulator, where it could influence the amount of effort put into articulation.

Another possibility is that effects of frequency arise during phonological processing. All words in a speaker’s mental lexicon are believed to have a certain resting activation level. When this activation level is high (as is the case for highly fre-

quent words), activation may spread to the constituent speech sounds more quickly than when it is low, resulting in quicker preparations of the speech sounds and thus, shorter articulatory durations (e.g., Balota, Boland, & Shields, 1989). Our current data do not allow us to distinguish between these two hypotheses, as the words under investigation were produced after both conceptual and articulatory preparation had taken place. If, however, frequency effects were to be found in situations where no conceptual preparation was required (e.g., in shadowing or speeded naming tasks), this would suggest that at least part of the effect of frequency on durations arises during the later stages of the speech production process.

One assumption made by many theories of speech production remains problematic, however. They regard the syllable as the principal unit of articulation, as well as the primary locus of frequency effects below the word form level (e.g., Levelt, Roelofs, & Meyer, 1999; Sevald, Dell, & Cole, 1995). Given the results of the present study, this view appears to be too restricted. First of all, we find evidence for word-specific frequency effects that operate below the level of the syllable. More importantly, the different segments in a syllable were shown to be subject to different, sometimes even contradictory, forces. This suggests that the motor program that is executed during articulation is very likely not the syllable, since one would expect similar processes to apply to all subcomponents of such a program.

Related to this issue is the question why some segments were affected by frequency while others were not. Although it is very difficult to abstract a general pattern from the data, it is at least clear that durational reduction is not restricted to vowels. Furthermore, we find some evidence that speakers sometimes compensate for the absence of one segment by lengthening another. This presents interesting challenges for frameworks such as Articulatory Phonology (Browman & Goldstein, 1992), since it suggests that reduction of articulatory effort does not necessarily involve either increased overlap or reduced magnitude of speech gestures.

Of course, the morphological status of affixes is relevant as well. Hay (2003)

showed that the psychological reality of affixes depends on the ratio between the frequency of the morphologically complex word and the frequency of its stem. Moreover, she found that *t*'s preceding the English suffix *-ly* were more likely to be deleted if this ratio was high. We included the ratio in our analyses and found that a higher value led to longer fricatives in the prefix *ge-*. Apparently, less psychological reality for the affix does not necessarily imply shorter realizations. A possible explanation for this finding is that in most Dutch monomorphemic words, the first syllable receives stress (Booij, 1995). If speakers no longer regard the prefix as a separate morpheme, it will eventually become a 'normal' word-initial syllable and therefore more likely to receive stress. The longer duration we observed for the fricative in *ge-* might be the precursor of such a change.

One of the four affixes we investigated, *ver-*, failed to show an effect of frequency. This could be related to the fact that the initial fricative can be devoiced, leading to more variation in the sample. In addition, verbal prefixes like *ver-* and *ont-* contribute much more to the meaning of their carrier words than *ge-* and *-lijk*, which merely signal grammatical function. Since the main objective of speakers is to get the meaning of their utterances across, it is not inconceivable that relatively meaningful units are less affected by frequency than more or less meaningless ones. Redundancy has many different dimensions, and the role of semantics in defining it should not be underestimated. This is confirmed by other studies, which also report differences between words in their sensitivity to reduction (e.g., Jurafsky *et al.*, 2001; Bell *et al.*, 2003).

A useful step in combining different dimensions of redundancy was taken by Gregory *et al.* (1999), who incorporated Latent Semantic Analysis in their study. Latent Semantic Analysis (Landauer & Dumais, 1997) is a method to compute semantic relatedness scores from large-scale co-occurrence statistics. Since relatedness scores can be computed between a word and the whole discourse preceding it, this measure effectively combines word repetition, predictability from neighbouring words,

and semantic association with other words used in the conversation. Importantly, Gregory *et al.*'s (1999) results also show that other predictors remain relevant as well. For example, they report effects of mutual information and repetition over and above semantic relatedness. This suggests that this particular measure of semantic relatedness is not the only variable that should be taken into account.

One might ask whether our methodology could also be used to investigate acoustic reduction in stems. Indeed, it would be possible to compare the durations of identical stems that are combined with different affixes. However, in most languages the number of stems that lend themselves to such an approach will be rather restricted, as stems are generally less productive than affixes.

If the aim is to compare reduction in stems to reduction in affixes, word frequency is probably not the best variable to focus on. In non-agglutinative languages, the number of words containing an identical combination of stem and affix will be too small to make a comparison of words with different frequencies possible. For example, the English stem-affix combination *disable* only occurs in *disable*, *disables*, *disabled*, *disabling*, and *disablement*. Therefore, it might be a better idea to study measures like conditional probability or mutual information, and see whether they affect stems and affixes differently (Pluymaekers, Ernestus, & Baayen, *submitted*).

With regard to future research, several issues need to be addressed. First of all, we need to know more about the way the different measures of probability interact. This requires large databases of carefully segmented speech, so that multiple tokens of the same words can be examined in their respective contexts. Second, more attention needs to be paid to semantic variables. Taken together, these lines of research could result in a model in which probabilistic and semantic relationships are exploited to the fullest and play a role at the finest level of acoustic detail.

ACKNOWLEDGMENTS

This research was supported by Netherlands Organization for Scientific Research (NWO) grant number 360-70-130 to the third author. We thank Sarah Hawkins, Willem Levelt, Anders Löfqvist, and an anonymous reviewer for their helpful comments on earlier versions of this paper.

REFERENCES

- Adank, P., Van Hout, R. and Smits, R. (2004), An acoustic description of the vowels of Northern and Southern Standard Dutch, *J. Acoust. Soc. Am.* **116**, 1729–1738.
- Aylett, M. and Turk, A. (2004), The smooth signal redundancy hypothesis: a functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech, *Lang Speech* **47**, 31–56.
- Balota, D., Boland, J. and Shields, L. (1989), Priming in pronunciation: beyond pattern recognition and onset latency, *J. Memory Lang.* **28**, 14–36.
- Bard, E., Anderson, A., Sotillo, C., Aylett, M., Doherty-Sneddon, G. and Newlands, A. (2000), Controlling the intelligibility of referring expressions in dialogue, *J. Memory Lang.* **42**, 1–22.
- 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.
- Bock, J. (1995), Sentence production: From mind to mouth, in J. Miller and P. Eimas (eds), *Handbook of perception and cognition, Vol. 11: Speech, language, and communication*, Academic Press, pp. 181–216.
- Boersma, P. (2001), Praat, a system for doing phonetics by computer, *Glott International* **5**, 341–355.
- Booij, G. E. (1995), *The phonology of Dutch*, Clarendon Press, Oxford.
- Browman, C. and Goldstein, L. (1992), Articulatory phonology: An overview, *Phonetica* **49**, 155–180.
- Bush, N. (2001), Frequency effects and word-boundary palatalization in English, in J. Bybee and P. Hopper (eds), *Frequency and the emergence of linguistic structure*, Benjamins, Amsterdam, pp. 255–280.

- Bybee, J. L. (2001), *Phonology and language use*, Cambridge University Press, Cambridge.
- Byrd, D. (1994), Relations of sex and dialect to reduction, *Speech Comm.* **15**, 39–54.
- Cambier-Langeveld, T. (2000), *Temporal marking of accents and boundaries*, LOT, Utrecht.
- Chatterjee, S., Hadi, A. and Price, B. (2000), *Regression analysis by example*, John Wiley & Sons, New York.
- Cooper, W. and Paccia-Cooper, J. (1980), *Syntax & Speech*, Harvard University Press, Cambridge, Mass.
- Dell, G. (1986), A spreading-activation theory of retrieval in sentence production, *Psychol. Rev.* **93**, 283–321.
- Ernestus, M. (2000), *Voice assimilation and segment reduction in casual Dutch, a corpus-based study of the phonology-phonetics interface*, LOT, Utrecht.
- Fosler-Lussier, E. and Morgan, N. (1999), Effects of speaking rate and word frequency on pronunciations in conversational speech, *Speech Comm.* **29**, 137–158.
- Fougeron, C. and Keating, P. (1997), Articulatory strengthening at the edges of prosodic domains, *J. Acoust. Soc. Am.* **101**(6), 3728–3740.
- Fowler, C. and Housum, J. (1987), Talkers' Signalling of "New" and "Old" Words in Speech and Listeners' Perception and Use of the Distinction, *J. Memory Lang.* **26**, 489–504.
- Fox Tree, J. and Clark, H. (1997), Pronouncing 'the' as 'thee' to signal problems in speaking, *Cognition* **62**, 151–167.
- Garrett, M. (1975), The analysis of sentence production, in G. Bower (ed.), *The psychology of learning and motivation*, Academic Press, New York, pp. 133–178.

- Gregory, M., Raymond, W., Bell, A., Fosler-Lussier, E. and Jurafsky, D. (1999), The effects of collocational strength and contextual predictability in lexical production, *Chicago Ling. Soc.* **35**, 151–166.
- Hay, J. (2003), *Causes and consequences of word structure*, Routledge, New York and London.
- Hunnicuttt, S. (1985), Intelligibility versus redundancy - conditions of dependency, *Lang Speech* **28**, 47–56.
- Jescheniak, J. D. and Levelt, W. J. M. (1994), Word frequency effects in speech production: Retrieval of syntactic information and of phonological form, *J. Exp. Psychol. Learn Mem. Cogn* **20**(4), 824–843.
- Jespersen, O. (1922), *Language: its nature, development and origin*, George Allen & Unwin Ltd, London.
- Johnson, K. (2004), Massive reduction in conversational American English, *Spontaneous speech: data and analysis. Proceedings of the 1st session of the 10th international symposium*, The National International Institute for Japanese Language, Tokyo, Japan, pp. 29–54.
- Jurafsky, D., Bell, A., Gregory, M. and Raymond, W. (2001), Probabilistic relations between words: Evidence from reduction in lexical production, in J. Bybee and P. Hopper (eds), *Frequency and the emergence of linguistic structure*, John Benjamins, Amsterdam, pp. 229–254.
- Keune, K., Ernestus, M., Van Hout, R. and Baayen, R.H. (in press), Social, geographical, and register variation in Dutch: From written ‘mogelijk’ to spoken ‘mok’, to appear in *Corpus Linguistics and Linguistic Theory*.
- Landauer, T. K. and Dumais, S. (1997), A solution to Plato’s problem: The Latent Semantic Analysis theory of acquisition, induction, and representation of

- knowledge, *Psychol. Rev.* **104**, 211–240.
- Landauer, T. K. and Streeter, L. A. (1973), Structural differences between common and rare words: Failure of equivalence assumptions for theories of word recognition, *J. Verb. Learn. Verb. Behav.* **12**, 119–131.
- Levelt, W. J. M. (1989), *Speaking. From intention to articulation*, The MIT Press, Cambridge, Mass.
- 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.
- Nooteboom, S. G. (1972), *Production and perception of vowel duration: A study of the durational properties of vowels in Dutch*, University of Utrecht, Utrecht.
- Oostdijk, N. (2000), The Spoken Dutch Corpus project, *The ELRA Newsletter* **5**, 4–8.
- Pierrehumbert, J. B. (2003), Probabilistic phonology: Discrimination and robustness, in R. Bod, J. Hay and S. Jannedy (eds), *Probability theory in linguistics*, The MIT Press, pp. 177–228.
- Pluymaekers, M., Ernestus, M. and Baayen, R. (submitted), Articulatory planning is continuous and sensitive to informational redundancy, submitted to *Phonetica*.
- Sevold, C., Dell, G. and Cole, J. (1995), Syllable structure in speech production: Are syllables chunks or schemas?, *J. Memory Lang.* **34**, 807–820.
- Shields, L. and Balota, D. (1991), Repetition and associative context effects in speech production, *Lang Speech* **34**, 47–55.
- Umeda, N. (1977), Consonant duration in American English, *J. Acoust. Soc. Am.* **61(3)**, 846–858.

- Van Coile, B. (1987), A model of phoneme durations based on the analysis of a read Dutch text, *in* J. Laver and M. Jack (eds), *Proceed. Europ. Conf. Speech Techn.*, vol. 2, Edinburgh, United Kingdom, pp. 233–236.
- Van Son, R. and Pols, L. (2003), Information structure and efficiency in speech production, *Proceed. Eurospeech-2003*, Geneva, Switzerland.
- Vorstermans, A., Martens, J. and Van Coile, B. (1996), Automatic segmentation and labelling of multi-lingual speech data, *Speech Comm.* **19**, 271–293.
- Wright, C. (1979), Duration differences between rare and common words and their implications for the interpretation of word frequency effects, *Memory Cogn.* **7(6)**, 411–419.
- Zipf, G. (1929), Relative frequency as a determinant of phonetic change, *Harv. Studies Classi Philol.* **15**, 1–95.
- Zsiga, E. (1994), Acoustic evidence for gestural overlap in consonant sequences, *J. Phonetics* **22**, 121–140.

TABLE I. Beta coefficients and significance values of the effects for *ge-*. The beta coefficients indicate the magnitude of the effect in milliseconds. "–" means there was no significant effect. The bottom row shows the amount of variance explained (R^2) by each model.

| Predictor | Prefix | Fricative | Vowel |
|------------------------------|----------|-----------|--------|
| Frequency | –4.1** | –3.6*** | –2.1* |
| Initial Position | – | – | 14.5* |
| Onset Complexity | –7.3* | – | – |
| Speech Rate | –8.6**** | –3.5*** | –3.0** |
| Word-Stem Ratio | – | 34.7* | – |
| Explained variance (R^2) | .10 | .07 | .06 |

* = $p < 0.05$ ** = $p < 0.01$ *** = $p < 0.001$ **** = $p < 0.0001$

TABLE II. Beta coefficients and significance values of the effects for *ont-*. The beta coefficients indicate the magnitude of the effect in milliseconds. "–" means there was no significant effect. The bottom row shows the amount of variance explained (R^2) by each model.

| Predictor | Prefix | Vowel | Nasal | Plosive |
|------------------------------|----------|-------|-----------|----------|
| Frequency | – | – | – | –17.8*** |
| Frequency * Speech Rate | –2.9** | – | – | – |
| Frequency * Year of Birth | 0.2** | – | – | 0.3** |
| Frequency * Plosive Present | – | – | 5.9*** | – |
| Plosive Present | – | – | –34.2**** | – |
| Sex | – | 8.7* | – | – |
| Speech Rate | – | – | – | –3.6* |
| Year of Birth | –1.4**** | –0.3* | –0.2** | –1.0** |
| Explained variance (R^2) | .24 | .10 | .48 | .28 |

* = $p < 0.05$ ** = $p < 0.01$ *** = $p < 0.001$ **** = $p < 0.0001$

TABLE III. Beta coefficients and significance values of the effects for *ver-*. The beta coefficients indicate the magnitude of the effect in milliseconds. "–" means there was no significant effect. The bottom row shows the amount of variance explained (R^2) by each model.

| Predictor | Prefix | Fricative | Rime |
|------------------------------|---------|-----------|----------|
| Onset Complexity | –14.7** | –7.4* | –12.8*** |
| Sex | – | –12.1*** | – |
| Year of Birth | –0.5* | – | –0.3* |
| Explained variance (R^2) | .09 | .12 | .17 |

* = $p < 0.05$ ** = $p < 0.01$ *** = $p < 0.001$ **** = $p < 0.0001$

TABLE IV. Beta coefficients and significance values of the effects for *-lijk*. The beta coefficients indicate the magnitude of the effect in milliseconds. "–" means there was no significant effect. The bottom row shows the amount of variance explained (R^2) by each model.

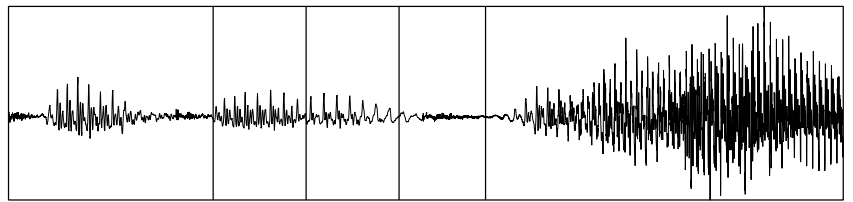
| Predictors | Suffix | Suffix | [l] | Vowel | Plosive | Plosive |
|------------------------------|-------------|----------|----------|-----------|-------------|---------|
| | (Non-Final) | (Final) | | | (Non-Final) | (Final) |
| Frequency | –7.7*** | – | –2.1** | – | – | – |
| Final Position | – | – | 10.0*** | 12.2** | – | – |
| Following Disfluency | – | – | – | – | –66.2**** | – |
| Plosive Present | – | 147.5** | – | –30.6**** | – | – |
| Speech Rate | – | –36.4*** | –4.3**** | –4.2** | – | –22.9** |
| Year of Birth | –0.7** | – | –0.2** | – | – | – |
| Explained variance (R^2) | .18 | .47 | .32 | .23 | .25 | .24 |

* = $p < 0.05$ ** = $p < 0.01$ *** = $p < 0.001$ **** = $p < 0.0001$

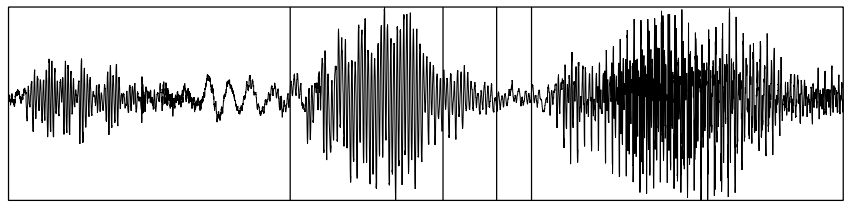
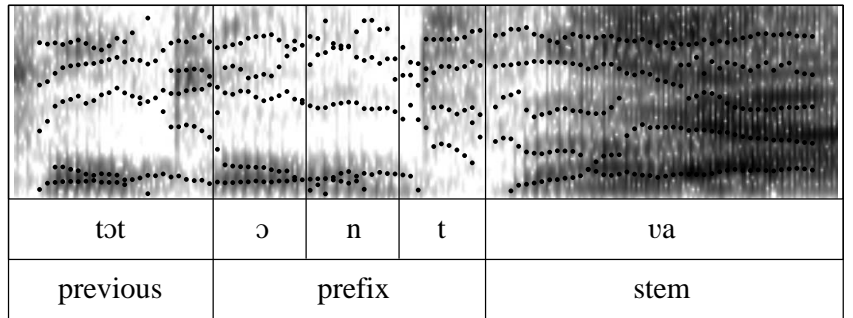
FIG. 1. Manual segmentations for the tokens *ontwaken* (top) and *ontwijken* (bottom). *Ontwaken* was a more or less ideal case, in which there was no background noise or overlapping speech. For *ontwijken*, the amount of background noise was much greater. In both cases, we placed boundaries where we could see both visible changes in the waveform pattern supported by abrupt formant transitions in the spectrogram.

FIG. 2. Interaction between Frequency and Speech Rate as observed for the duration of *ont-*. The panels should be read from left to right and from bottom to top: Speech Rate is lowest in the bottom left panel and highest in the top right panel. There is no Frequency effect when Speech Rate is low (the bottom left and middle panels) or extremely high (the top right panel).

FIG. 3. Interaction between Frequency and Year of Birth as observed for the duration of *ont-*. The panels should be read from left to right and from bottom to top: the oldest speakers are in the bottom left panel and the youngest speakers in the top right panel. There is no effect of Frequency for the youngest age groups (top middle and right panels).



0 0.676712



0 0.667130

