Morphological predictability and acoustic duration

of interfixes in Dutch compounds

Victor Kuperman and Mark Pluymaekers

Radboud University Nijmegen,

Wundtlaan 1,

6525 XD Nijmegen,

The Netherlands

Mirjam Ernestus and Harald Baayen

Radboud University Nijmegen and Max Planck Institute for Psycholinguistics,

Wundtlaan 1,

6525 XD Nijmegen,

The Netherlands

(Dated: January 9, 2007)

Abstract

This study explores the effects of informational redundancy, as carried by a word's morphological paradigmatic structure, on acoustic duration in read aloud speech. The hypothesis that the more predictable a linguistic unit is, the less salient its realization, was tested on the basis of the acoustic duration of interfixes in Dutch compounds in two datasets: One for the interfix -*s*- (1155 tokens) and one for the interfix -*e(n)*- (742 tokens). Both datasets show that the more probable the interfix is, given the compound and its constituents, the *longer* it is realized. These findings run counter to the predictions of information-theoretical approaches and can be resolved by the Paradigmatic Signal Enhancement Hypothesis. This hypothesis argues that whenever selection of an element from alternatives is probabilistic, the element's duration is predicted by the amount of paradigmatic support for the element: The most likely alternative in the paradigm of selection is realized longer.

I. INTRODUCTION

One of the organizing principles of speech production is the trade-off between economy of articulatory effort and discriminability of the speech signal (Lindblom, 1990). Speech communication often takes place in noisy conditions. In order to ensure robust recognition of their acoustic output, speakers need to invest effort in articulation. Yet clear and careful articulation is costly and hence tends to be dispensed efficiently (cf., Aylett and Turk, 2004; Hunnicutt, 1985). As a consequence, elements with low information load (or high predictability) have shorter or otherwise less salient realizations than relatively more informative elements of an utterance.

The informational redundancy of speech elements is often operationalized in terms of the probability (relative frequency of occurrence) of a linguistic unit (e.g., phoneme, syllable, word, or phrase) in its context. High probability has been observed to correlate with acoustic reduction in a large variety of language domains: Syntactic, discourse-related, phonological and prosodic, and lexical (e.g., Aylett and Turk, 2004; Bard *et al.*. 2000; Fowler and Housum, 1987; Jurafsky *et al.*, 2001; Lieberman, 1963; McAllister *et al.*, 1994; Pluymaekers, Ernestus and Baayen, 2005a; Pluymaekers, Ernestus and Baayen, 2005b; Samuel and Troicki, 1998; Scarborough, 2004; Van Son and Pols, 2003; Van Son and Van Santen, 2005). The attested types of reduction include — apart from widely reported durational shortening of syllables and individual phonemes — deletion of phonemes and complete syllables (e.g., Ernestus, 2000; Johnson, 2004), decrease in spectral center of gravity (Van Son and Pols, 2003), decrease in mean amplitude (Shields and Balota, 1991), higher degree of centralization of vowels (Munson and Solomon, 2004), and lower degree of coarticulation (Scarborough, 2004). The informational redundancy associated with a particular unit is a juxtaposition of the unit's probabilities given all relevant contexts. For instance, a word can be predictable because it has a high frequency, but also because it is frequently used with the word that precedes it. Both factors diminish the word's informativeness and both are expected to correlate with durational shortening.

The information-theoretical framework developed by Shannon (1948) has been used to explain the association between acoustic salience and informational redundancy. The efficiency of information transmission is optimal if the information in the signal is distributed equally, or smoothly, per time unit (e.g., Aylett and Turk, 2004; Aylett and Turk, 2006). When an important element is transmitted for a longer time, the probability of losing this element to noise decreases and the probability of the element being recognized correctly increases. This theoretical paradigm views acoustic duration as a means of smoothing the amount of information in the signal over time.

The present paper shows how the information carried by morphological paradigmatic structure modulates acoustic duration. Previous research (cf., Hay, 2003; Losiewicz, 1992) reported morphological effects on the acoustic duration of affixes in complex words. A related line of research demonstrated the influence of lexical neighborhood density on durational characteristics and coarticulation in speech production (e.g., Munson and Solomon, 2004, Scarborough, 2004, Vitevitch, 2002). The morphological objects that are central in the present study are interfixes in Dutch noun-noun compounds. We will show that the acoustic duration of these interfixes creates an apparent paradox for the proposed information-theoretical principle of "less information, more reduction", which underlies the Smooth Signal Redundancy Hypothesis (Aylett and Turk, 2004), the Probabilistic Redundancy Hypothesis (Jurafsky *et al.*, 2001), and research on speech efficiency (e.g., Van Son and Pols, 2003). In our data, the more predictable the interfix is, the *longer* its articulation.

The distributional characteristics of the interfixes in Dutch compounds provide a clear-cut example of probabilistic, non-categorical morphological structure. Compounding is very productive in Dutch and is defined as the combination of two or more lexemes (or constituents) into a new lexeme (cf. Booij, 2002). In this paper we based our decisions of whether a given word is a compound and what its constituents are on the morphological parsing provided in the CELEX lexical database (Baayen, Piepenbrock and Gulikers, 1995). Compounds in Dutch can be realized with the interfix -*s*- (e.g., *oorlog-s-verklaring*, "announcement of war"), or with the interfix -*en*- (or its variant -*e*-) (e.g., *dier-en-arts*

"veterinary"). Most compounds in Dutch, however, have no interfix (e.g., *oog-arts* "oph-thalmologist"): For ease of exposition, we will henceforth refer to these latter words as compounds with the zero-interfix, or -∅-. In the frameworks that adopt deterministic rules, the distribution of interfixes in Dutch is enigmatic and inexplicable. Krott, Baayen and Schreuder (2001), however, have shown that the distribution of interfixes follows probabilistic principles defined over constituent families. The left (or right) constituent family of a compound is the set of all compounds which share the left (or right) constituent with this compound. For instance, the left constituent family of the compound *banknote* includes *bankbill, bankbook, bank-draft, bank-rate,* and *bankroll.* Krott, Baayen and Schreuder (2001), Krott *et al.* (2002) and Krott, Schreuder and Baayen (2002) show that the selection of the interfix is biased towards the interfix that is most commonly used with the given left constituent and, to a lesser extent, with the right constituent. Thus, besides having their own probability of occurrence, interfixes exhibit dependencies on larger morphological units both to the left and to the right. For this reason, interfixes serve as an appealing testing ground for studying the consequences of morphological predictability for acoustic realization.

The primary focus of the present study is the relationship between the predictability of the interfix given the morphological constituents of the compound, and its duration. We study the information-theoretical approach for two datasets with interfixed compounds and against the backdrop of multiple sources of redundancy, ranging from morphological to phonological and lexical information. Along the way, we replicate findings of laboratory studies of durational reduction for lively read-aloud speech.

II. METHODOLOGY

A. Materials

Acoustic materials were obtained from the Read Speech (or the "Library for the Blind") component of the Spoken Dutch Corpus (Oostdijk, 2000). Within this corpus of approximately 800 hours of recorded speech, the Read Speech component comprises 100 hours of

recordings of written texts read aloud by speakers of Northern Dutch from the Netherlands and Southern Dutch from the Flanders area of Belgium. In the preparation of the recordings, speakers were pre-screened for the quality of their voice and clarity of pronunciation, and texts were made available to the speakers beforehand for preparatory reading. We chose to concentrate on read speech primarily because of the low level of background noise of the recordings. Quality was essential, since Automatic Speech Recognition (henceforth, ASR) was used for obtaining the segmental durations (see below). It should be noted that since these texts of fiction were read for the collection of the Library for the Blind, the reading style was a lively, rather than monotonous recitation, especially in the dialogs, where readers often mimicked casual speech.

Two datasets of Dutch noun-noun compounds were compiled: One with tokens containing the interfix -*s*- and one with compounds containing the interfix -*e(n)*-. Tokens in which the interfix -*s*- was either preceded or followed by the phonemes [s], [z] or [ʃ] were excluded from the dataset, since such an environment makes it difficult to reliably segment the interfix from its neighboring segments. The final dataset for the interfix -*s*- consisted of 1155 tokens. Similarly, tokens in which the second constituent begins with the segments [n] or [m] were taken out off the dataset of -*e(n)*- interfixes, resulting in a dataset of 742 tokens.

B. Measurements

Acoustic analysis of the selected tokens was performed using ASR technology. This was done for several reasons. First of all, the ASR technology allows to process a large volume of data in a relatively short time, which was important given the size of datasets used in this study. Moreover, it is possible to train an ASR device that bases its decisions purely on the characteristics of the acoustic signal, without reference to general linguistic knowledge. This is very difficult for human transcribers, who are bound to be influenced by expectations based on their knowledge of spelling, phonotactics, and so on (Cucchiarini, 1993). Second, ASR devices are perfectly consistent: Multiple analyses of the same acoustic signal always

6

yield exactly the same result. Finally, the reliability of segmentations generated by an ASR system is equal to that of segmentations made by human transcribers (Vorstermans, Martens and Van Coile, 1996), provided that a phonemic transcription of the signal is available to the ASR algorithm.

For the present analysis, we utilized a Hidden Markov Model (HMM) speech recognizer. This recognizer was trained using the software package HTK (Young *et al.*, 2002), comprises 37 phone models representing the 36 phonemes of Dutch and silence, and uses for each model 3-state HMMs with 32 gaussians per state (Kessens and Strik, 2004). The HTK recognizer operates in two modes: If it is provided with the transcription of the speech recording, it determines segmental temporal boundaries; if no such transcription is provided, it identifies both the phonemes and the positions of their temporal boundaries. The accuracy of segmentation is higher in the transcription-based mode. The sample rate of the HTK is 10ms. The reliability of the ASR's segmentation with predefined transcriptions was established in a test in which the positions of phoneme boundaries placed by the ASR were compared to the positions of the same boundaries placed by a trained phonetician. The materials used for this test consisted of 189 words spoken in isolation. Comparison between the ASR-generated and manual segmentations revealed that, after post-processing, 81% of the automatic boundaries were placed within 20 milliseconds of the corresponding hand-coded boundaries. This level of accuracy is in accordance with international standards (Vorstermans *et al.*, 1996), and we considered it sufficient for present purposes.

Acoustic analysis proceeded as follows. First, the speech signal corresponding to the target compound was manually excised from its utterance context and parameterized using Mel Frequency Cepstral Coefficients. The parameterized signal was then supplied to a Viterbi segmentation algorithm, along with a phonemic transcription of the word. This transcription was taken from the CELEX lexical database. However, for words with the interfix -*e(n)*-, a cursory inspection of sound files established that many instances of this interfix were not realized as [ə] (the canonical pronunciation in CELEX), but rather as [ən]. An inspection of the sound files from the dataset with the interfix -*s*- revealed cases where the

interfix was realized as [s] instead of the CELEX transcription [z] due to the regressive voice assimilation. Therefore, two trained phoneticians independently transcribed the realization of interfixes in both datasets. Initially, they disagreed on 10% of tokens from the *en*-dataset and 13% of tokens from the *s*-dataset. In both cases, they subsequently carried out a joint examination of the problematic tokens and came up with consensus transcriptions. The resulting transcriptions were provided to the segmentation algorithm, which estimated the boundaries of the phonemes in the acoustic signal. In this way, we obtained information about the durations of all segments for all words.

The acoustic duration of the whole interfix (henceforth, *InterfixDuration*) was taken as the main dependent variable in this study.

III. MORPHOLOGICAL VARIABLES

As shown in Krott *et al.* (2001), the more frequent an interfix is for the left constituent family of a compound, the more biased speakers are to use this interfix in that compound. The measures for this morphologically based bias will be at the center of our interest. They are defined as the ratio of the number of compounds where the left constituent is followed by -*s*-, -*e(n)*-, or -∅- respectively, and the total number of compounds with the given left constituent (henceforth, the left family size). To give an example, the Dutch noun *kandidaat* "candidate" appears as the left constituent in one compound with the interfix -*s*-, *kandidaat-s-examen* "bachelor's examination", in one compound with the interfix -*en*-, *kandidat-en-lijst* "list of candidates", and in one compound without an interfix *kandidaat-stelling* "nomination". The type-based bias of this left constituent family towards the interfix -*s*- is $1/(1 + 2) = 0.33$. The bias of the interfix -*e(n)*- has the value of $1/(1 + 2) = 0.33$ as well, and so does the bias of the zero-interfix. The measures of bias are labeled *TypeSBias*, *TypeEnBias* and *TypeZeroBias*.

Alternative, token-based, estimates of the bias are defined in terms of the frequencies of occurrence, rather than the type count of the compounds. The performance of token-based

8

measures is consistently worse in our models than that of the type-based ones. Therefore, the token-based measures are not reported here. Furthermore, we only consider left constituent families, since the effect of the right bias is reported as either weak or absent (Krott, Schreuder and Baayen, 2002; Krott *et al.*, 2004).

The predictivity of constituent families for the duration of the interfix may extend beyond the bias measures, which only estimate the ratio of variants in the constituent family, without taking the magnitude (size, frequency, or information load) of the constituent family into account. However, these magnitudes are expected to exhibit effects in our analysis, since they repeatedly emerged as significant predictors in both the comprehension and production of Dutch compounds (e.g., Bien, Levelt and Baayen, 2005; De Jong *et al.*, 2002; Krott *et al.*, 2004). To estimate the magnitude of constituent families, we incorporate in our study position-specific measures of entropy proposed by Moscoso del Prado Martín, Kostić and Baayen (2004). These measures employ the concept of Shannon's entropy (Shannon 1948), which estimates the average amount of information in a system on the basis of the probability distribution of the members of that system. The probability of each member (p_{sys}) is approximated as the frequency of that member divided by the sum of the frequencies of all members. The entropy of a system with n members is then the negative weighted sum of log-transformed (base 2) probabilities of individual members:

$$H = -\sum_{i=1}^{n} p_{sys} * \log_2 p_{sys}$$

Note that the entropy increases when the number of paradigm members is high (i.e. family size is large) and/or when the members are equiprobable.

Let us consider the positional entropy measure of the left constituent family of the Dutch noun *kandidaatstelling*. This family consists of three members: *kandidaatsexamen* has a lemma frequency of 22, *kandidaatstelling* has a lemma frequency of 15, and *kandidatenlijst* has a lemma frequency of 19 in the CELEX lexical database, which is based on a corpus of 42 million word forms. The cumulative frequency of this family is $22 + 15 + 19 = 56$, and the relative frequencies of these three family members are $22/56 = 0.39$ for *kandidaatsexamen*,

15/56 = 0.27 for *kandidaatstelling* and 19/56 = 0.34 for *kandidatenlijst.* The left positional entropy of this constituent family therefore equals $-(0.39 * \log_2 0.39 + 0.27 * \log_2 0.27 + 0.34 * \log_2 0.34) = 1.57$ bit.

We consider the positional entropy measures for both the left and the right constituent families, henceforth *LeftPositionalEntropy* and *RightPositionalEntropy* as potential predictors of the acoustic duration of the interfix. The informativeness of the right constituent family is meaningful as a measure of the cost of planning the right constituent: Planning upcoming elements with a low information load has been shown to predict reduction in the fine phonetic detail of the currently produced elements (Pluymaekers *et al.*, 2005a).

IV. OTHER VARIABLES

Since acoustic duration is known to depend on a wide range of factors, we used stepwise multiple regression to bring these factors under statistical control. Two sets of factors were considered: Lexical frequency-based probabilities, and phonetic, phonological and sociolinguistic variables.

A. Probabilistic factors

Phrasal level: A higher likelihood of a word given its neighboring words has been shown to correlate with vowel reduction, segmental deletion, and durational shortening (Bell *et al.*, 2003; Jurafsky *et al.*, 2001; Pluymaekers *et al.*, 2005a). To quantify this likelihood, for each compound token in our data we calculated its mutual information with the preceding and the following word (*BackMutualInfo, FwdMutualInfo*) by using the following equation (X and Y denote either the previous word and the compound, or they denote the compound and the following word; XY denotes the combination of the two words):

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

The measures were computed on the basis of the Spoken Dutch Corpus, which contains 9 million word tokens. All frequency measures were (natural) log-transformed. Obviously, the

values could not be computed for the instances where the target word was utterance-initial or utterance-final, respectively.

For those words for which mutual information with the preceding or the following word could be computed, we checked whether it was a significant predictor of the duration of the interfix over and beyond other factors. Neither *BackMutualInfo* nor *FwdMutualInfo* reached significance in our datasets. This result may originate in the properties of the datasets which comprise relatively low-frequency compounds. Obviously, these low-frequency compounds have even lower frequencies of cooccurrence with their neighboring words. For instance, for the s-dataset the average frequency of cooccurrence of the compounds with the preceding word is a mere 1.63 ($SD = 0.77$), and with the following word a mere 1.20 ($SD = 0.30$). Another explanation may be that effects of contextual predictability do not extend to phonemes in the middle of long compounds. They may only emerge for segments at word boundaries (e.g., Jurafsky *et al.*, 2001; Pluymaekers *et al.*, 2005a).

Word level: The lexical frequency of a word is known to codetermine articulation and comprehension (e.g., Jurafsky *et al.*, 2001; Pluymaekers *et al.*, 2005a; Scarborough, Cortese and Scarborough, 1977; Zipf, 1929). Moreover, previous research has shown that whole word frequency robustly affects production and comprehension of compounds even in the low-frequency range (cf. e.g., Bertram and Hyönä, 2003, Bien *et al.*, 2005). Therefore we include the natural log-transformed compound frequency (*WordFrequency*) as a control variable in the analyses. Together with the measure of the bias and the left positional entropy, this variable forms a cluster of predictors that capture different aspects of the same phenomenon. The measure of the bias estimates the *proportion* of the positional family of compounds that supports the interfix. The corresponding entropy estimates the number and average information load of the members in this family, i.e., it gauges the reliability of the knowledge base for the bias. Finally, a high compound frequency quantifies the evidence for the cooccurrence of the left and right constituents with the interfix. We expect these variables to behave similarly in predicting the durational characteristics of the interfix.

11

Segmental level: Another dimension of predictability for segmental duration is the amount of lexical information in the individual segment given the preceding fragment of the word (i.e., given the "word onset"). Following Van Son and Pols (2003), we define an information-theoretic measure that quantifies segmental lexical information (*TokenSegmentalInfo*):

$$I_L = -\log_2 \frac{\text{Frequency([word\_onset] + target\_segment)}}{\text{Frequency([word\_onset] + any\_segment)}}$$

Van Son and Pols (2003) interpret this measure as estimating the segment's incremental contribution to word recognition. The occurrence of a segment that is improbable given the preceding fragment of the word limits the cohort of matching words substantially and thus facilitates recognition. To give an example, the amount of lexical information of the segment [s] given the preceding English word fragment [kaʊ] is calculated as the negative log-transformed ratio of the cumulative frequency of words that begin with the string [kaʊs] (e.g., *cows, cowskin, cowslip, cowslips*) and the cumulative frequency of the words that begin with the string [kaʊ] plus any segment (e.g., *cows, cowpat, cowshed, cowskin, cowslip, cowslips, etc.*). In the present study, segmental lexical information measures are based on the frequencies of single words, such as made available in CELEX, and do not account for combinations of words, even if those may acoustically be valid matches for the phonetic string. For instance, the combination *cow stopped* is not included in the calculation of the lexical information for the segment [s] in the string [kaʊs].

A positive correlation of this token-based segmental lexical information and segmental duration was reported in Van Son and Pols (2003) for different classes of phonemes grouped by manner of articulation: For read speech, the r-values of correlations that reached significance ranged between 0.11 and 0.18 (55811 df). If segmental lexical information indeed modulates fine phonetic detail, it is a potential predictor of the duration of the interfix.

To this token-based measure of segmental lexical information (*TokenSegmentalInfo*), we add a type-based measure, *TypeSegmentalInfo*, which is based on the *number* of words matching the relevant strings, rather than their cumulated frequencies:

$$S_L = -\log_2 \frac{\text{Number}([\text{word onset}] + \text{target segment})}{\text{Number}([\text{word onset}] + \text{any segment})}$$

We validated both the token-based and the type-based measures of segmental lexical information against our own dataset to establish how the performance of the type-based estimate S_L compares with that of the token-based measure I_L. Our approach differs from that of Van Son and Pols (2003) in that it considers the divergence of phonemes from their mean durations, rather than the raw durations of these phonemes. Different phonemes, even those that share manner of articulation, intrinsically differ in their durations. Therefore, pooling the durations of large classes of phonemes introduces unnecessary noise in the correlation analyses. We gauged the divergence of each instantiation of every phoneme from the mean duration of this phoneme and tested whether this divergence can be explained by the amount of lexical information carried by the phoneme. Our survey is based on *all* segments in the *s*-dataset and in the compounds of the *en*-dataset in which the interfix is realized as [ə].

We collected the data on mean durations from the Read Text component of the IFA corpus, a hand-aligned phonemically segmented speech database of Dutch (Van Son, Binnenpoorte, Van den Heuvel, Pols, 2001). We log-transformed the individual durations and computed the means and standard deviations of all tokens of each phoneme. Then, moving phoneme by phoneme through our compound dataset we calculated the z-score for each phoneme, that is, the difference between its actual log-transformed duration and its mean log duration, in units of standard deviation from the mean. The correlation between the observed durational difference and the corresponding amount of type-based segmental lexical information yields an r-value of 0.06 ($t(17694) = 7.41, p < 0.0001$). This order of magnitude is comparable with the results that Van Son and Pols (2003) obtained for the token-based measure of lexical information. The observed correlation is a rough estimate of the baseline effect that segmental lexical information may have on acoustic duration. The correlation is highly significant but the correlation coefficient is quite small. This is expected, given the multitude of phonetic, phonological, sociolinguistic and probabilistic factors that deter-

13

mine acoustic duration in speech production that are not taken into account here. As the type-based measure is predictive for durations of segments across the dataset, we decided to include it in our analyses of the interfix durations. Thus, we take as control variable the value of *TypeSegmentalInfo* for the (first) segment of the interfix.

Importantly, the durations show a weaker correlation with the token-based segmental lexical information, proposed by Van Son and Pols (2003) ($r = 0.03, t(17694) = 4.25, p < 0.0001$), than for its type-based counterpart ($r = 0.06$). This measure also performs worse in the models reported below. Since the token- and type-based measures are highly correlated, we incorporated only *TypeSegmentalInfo* in our analysis.

B. Phonetic, phonological and sociolinguistic variables

Speech rate is an obvious predictor of acoustic duration (e.g., Crystal and House, 1990; Fosler-Lussier and Morgan, 1999; Pluymaekers *et al.*, 2005a). Two different measures estimating speech rate were included as control variables. First, we defined an utterance-based rate of speech, *SpeechRate*, as the number of syllables in the utterance divided by the acoustic duration of the utterance. Utterance is defined here as the longest stretch of speech containing the compound and not containing an audible pause.

Second, we defined a more local speech rate for the interfix -*s*-. In the *s*-dataset, the interfix -*s*- always belongs to the coda of the preceding syllable. We measured the average segmental duration in the interfix-carrying syllable minus the -*s*- interfix, and considered it as an estimate of the local speed of articulation in the part of the syllable that precedes the interfix -*s*-, henceforth *SyllableSpeed*. The syllable from which the final segment [s] was subtracted is structurally complete, with an onset, a vowel and (in 83% of tokens) a coda of one or more consonants. Note that for words with the interfix -*e(n)*- this measure of local speech rate is not meaningful. It would subtract the complete rhyme of the relevant syllable, leaving only the onset, the duration of which is above all determined by the number and types of its consonants.

14

Nooteboom (1972) observed that segments are shorter the greater the number of syllables or segments in the word. We therefore considered the total number of segments in the word, *NumberSegments*, and the number of segments following the interfix, *AfterSegments*.

We also took into account the sex, age and language variety of the speaker (cf., Keune, Ernestus, Van Hout and Baayen, 2005). The binary variable *SpeakerLanguage* encodes the speaker's variant as Southern Dutch or Northern Dutch. If the information about age was missing, we filled in the average age of our speakers' population.

Prosody may affect the duration of segments as well. For instance, words at the beginning and the end of utterances show articulatory strengthening (e.g., Bell *et al.*, 2003; Cambier-Langeveld, 2000; Fougeron and Keating, 1997). To control for the word's position in the utterance, we coded each token with two binary variables *UtteranceInitial* and *UtteranceFinal*.

Furthermore, stressed syllables are pronounced longer than unstressed ones (e.g., Ladefoged, 1982). We coded each compound with the interfix *-s-* for whether its interfix-containing syllable carries a (primary or secondary) stress (the binary variable *Stressed*).

The interfix *-e(n)-* is never stressed. The common stress pattern for compounds with the interfix *-e(n)-* is for the primary stress to fall on the syllable immediately preceding the interfix-containing syllable, and the secondary stress on the syllable immediately following the interfix-containing syllable: The insertion of *-e(n)-* prevents a stress clash between the two constituents. The rhythmic structure of compounds has been proposed as a factor codetermining the selection of the interfix, in addition to lexical constituent families and several other factors (Neijt *et al.*, 2002). To test the acoustic consequences of the rhythmic pattern, we coded each compound in the *en*-dataset as to whether the interfix syllable intervenes between two immediately adjacent stressed syllables (the binary variable *Clash*).

Compounds with the interfix *-e(n)-* were coded for the presence or absence of [n] in the acoustic realization of the interfix (*NPresent*), as established by two phoneticians (see section II). Similarly, compounds with the interfix *-s-* were coded for whether the interfix was realized as [z], variable *PhonemeZ*.

15

Finally, the immediate phonetic environment can make a segment more or less prone to reduction. Unstressed vowels in Dutch tend to lengthen before oral stops (cf., Waals, 1999). Therefore, each compound in the dataset with the *-e(n)-* interfix was coded for the manner of articulation of the following segment (binary variable *FollowedbyStop*).

V. RESULTS

A. The interfix *-s-*

The dataset for the interfix *-s-* included 1155 tokens. The number of different word types was 680, and their token frequencies followed a Zipfian distribution ranging from 1 to 19. We fitted a stepwise multiple regression model with the acoustic duration of the interfix as the dependent variable. The values of this variable were (natural) log-transformed to remove skewness of the distribution. The resulting variable *InterfixDuration* has a mean of 4.37 of log units of duration ($SD = 0.35$). The log-transformation in this model and the models reported below was applied purely for statistical reasons, such as reducing the likelihood that the estimates of the coefficients are distorted by atypically influential outliers. The coefficients of the regression models that are presented here in log units of duration can easily be converted back into milliseconds by applying the exponential function e^F to the fitted values (F) of the model.

We identified 21 data points that fell outside the range of -2.5 to 2.5 units of SD of the residual error, or had Cook's distances exceeding 0.2. These outliers were removed from the dataset and the model was refitted. Below we only report variables that reached significance in the final model.

The strength of the bias for the *-s-* interfix, *TypeSBias*, emerged as a main effect with a positive slope: Surprisingly, the duration of *-s-* was longer for compounds with a greater bias for this interfix [$\hat{\beta} = 0.35, t(1125) = 5.20, p < 0.0001$]. A positive correlation with duration was present for the predictor *RightPositionalEntropy* as well [$\hat{\beta} = 0.07, t(1125) = 4.10, p < 0.0001$], indicating that the duration of the interfix increases with the informational

16

complexity of the right constituent. These main effects were modulated by an interaction between *TypeSBias* and *RightPositionalEntropy* $[\hat{\beta} = -0.07, t(1125) = -3.67, p = 0.0003]$. Inspection of conditioning plots revealed that the influence of the bias measure was greater when the value of the right positional entropy was low. In addition, *WordFrequency* had an unexpected positive slope that just failed to reach significance: $[\hat{\beta} = 0.01, t(1125) = 1.95, p = 0.0510]$. We found no effect of the *LeftPositionalEntropy*.

Importantly, the lexical segmental information of the interfix was predictive in the expected direction: Segments conveying more information tended to be longer [*TypeSegmentalInfo*: $\hat{\beta} = 0.12, t(1125) = 3.86, p < 0.0001$].

Among the phonological and phonetic variables, the measure of the speech rate also demonstrated the expected behavior. The greater the local speed of articulation, the shorter the realization of this interfix [*SyllableSpeed*: $\hat{\beta} = -0.51, t(1125) = -5.27, p < 0.0001$]. Whether the interfix-carrying syllable was stressed was a significant predictor as well, with stress predicting durational shortening of the interfix [*Stressed*: $\hat{\beta} = -0.09, t(1125) = -3.96, p < 0.0001$]. Finally, interfixes realized as [z] were shorter than those realized as [s], as expected given the findings by, for instance, Slis and Cohen (1969) [*PhonemeZ*: $\hat{\beta} = -0.16, t(1125) = -3.17, p = 0.0016$].

All significant predictors were tested for possible non-linearities; none reached significance. The bootstrap validated R^2 of the model was 0.104. The unique contribution of the morpholexical factors *TypeSBias*, *PositionalEntropyRight*, and *WordFrequency* to the explained variance over and above the other predictors was 2.0%, as indicated by the drop in R^2 when these variables were removed from the model.

B. Discussion

Three related morpholexical variables emerge as significant predictors of the duration of the interfix: *TypeSBias*, *RightPositionalEntropy* and (marginally) *WordFrequency*. The positive correlations of *TypeSBias* and *WordFrequency* with the duration of the interfix lead

to the paradoxical conclusion that a greater likelihood for a linguistic unit may lead to a longer acoustic realization of that unit, contradicting the information-theoretical approach to the distribution of acoustic duration. We will address this issue in the General Discussion.

The interaction of the right positional entropy with the bias hints at planning processes at work. According to Pluymaekers *et al.* (2005b), the planning of upcoming linguistic elements may interfere with the planning and production of preceding elements. We interpret the right positional entropy measure as tapping into the costs of planning the right constituent. The observed interaction indicates that the bias allows greater durational lengthening of the interfix when planning the next constituent is easy.

In accordance with previous reports (e.g., Van Son and Pols, 2003), a high amount of lexical information carried by an individual segment (*TypeSegmentalInfo*) predicts the acoustic lengthening of this segment. In other words, segments with a larger contribution to the word's discriminability are produced with increased articulatory effort, and hence prolonged duration. This highlights the paradox with which we are confronted: Conventional measures, such as the segmental lexical information, behave as expected, while measures for the likelihood of the interfix exhibit exceptional behavior.

The effects of *TypeSegmentalInfo* and of *TypeSBias* may appear to contradict each other: For the same segment [s], the former variable predicts acoustic reduction, while the higher bias correlates with acoustic lengthening. Yet the two variables operate independently on different levels: The level of morphological word structure for the bias, and the segmental level for the lexical information. In the model, their (opposite) effects are simply additive.

The position of the compound in the utterance did not affect the durational characteristics of the interfix significantly, which is in line with observations by Cambier-Langeveld (2000). Cambier-Langeveld argues that final lengthening in Dutch only applies to the last syllable in the word or, if the vowel in this last syllable is [ə], to the penultimate syllable. Thus, the interfix lies beyond the scope of this effect. Similarly, the interfix emerges as outside the domain of influence of initial lengthening.

Segments are typically longer in a stressed syllable. This may have gone hand in hand

with compensatory shortening of the duration of the following -s-. Compensatory reduction of the -s- in the coda of a stressed syllable may therefore provide an explanation for the observed effect of *Stressed.* Alternatively, acoustic reduction of the interfix may have arisen from the fact that stress on the syllable preceding the interfix -s- correlates with a higher local speech rate, which we calculated as the number of segments in the syllable (minus -s-) divided by the total duration of the syllable (minus -s-). This finding may appear counterintuitive, but it derives from the following observation. It is true that stressed syllables in our dataset have longer realizations than unstressed ones [two-tailed t-test: $t(1097) = 30.0$, $p < 0.0001$], but more importantly, they consist of more segments [two-tailed t-test: $t(1146) = 22$, $p < 0.0001$]. The net effect is the greater speech rate at stressed syllables. To test whether the latter finding is idiosyncratic to our dataset, we computed the number of segments for each syllable in Dutch monomorphemic words using CELEX phonological transcriptions. Again, we found that stressed syllables contained more segments than unstressed ones (2.76 vs. 2.17 segments per syllable, two-tailed t-test: $t(192546) = 208.8$, $p < 0.0001$). This difference retained significance when the counts were corrected for ambisyllabicity. We conclude that a higher local speech rate may have contributed to the shortening of -s-interfixes that follow stressed syllables.

C. The interfix *-e(n)-*

The *en*-dataset contained 742 tokens of compounds. The number of different word types equalled 305, and the Zipfian distribution of tokens per type ranged from 1 to 74. We log-transformed the acoustic durations of the interfixes, which then had a mean of 4.065 log units of duration ($SD = 0.420$). We fitted a stepwise multiple regression model to these durations. This time, 19 data points fell outside the range of -2.5 to 2.5 units of SD of the residual error or had Cook's distances exceeding 0.2. These outliers were removed from the dataset, and the model was refitted. Only predictors that reached significance are reported.

The morpholexical predictors performed as follows: A higher bias for the interfix *-e(n)-,*

TypeEnBias, correlated with longer interfixes: $[\hat{\beta} = 0.14, t(716) = 5.39, p < 0.0001]$. The positional entropy of the right constituent family also had a positive main effect $[\hat{\beta} = 0.08, t(716) = 4.56, p < 0.0001]$. The interaction of these two variables was not significant $(p > 0.4)$. *LeftPositionalEntropy* and *WordFrequency* did not reach significance either $(p > 0.1)$.

As in the model for the interfix *-s-*, a higher amount of lexical information, as attested by *TypeSegmentalInfo* for the first segment of the interfix, correlated with longer articulation $[\hat{\beta} = 0.07, t(716) = 3.09, p = 0.002]$. This effect is again in line with predictions of the information-theoretical approach.

The interfixes of 226 tokens (29%) in the dataset were realized as [ən], while 561 tokens were pronounced with [ə]. As expected, the presence of [n] in the interfix implied a substantial increase in the total duration of the interfix. The factor *NPresent* was the most influential predictor $[\hat{\beta} = 0.71, t(716) = 37.80, p < 0.0001]$, and its unique contribution to the explained variance of this duration was 55%.

Two phonetic factors contributed to the duration of the interfix. Unsurprisingly, the interfix was shorter when the utterance-based speech rate was higher [*SpeechRate*: $\hat{\beta} = -0.04, t(716) = -4.17, p < 0.0001$]. Factor *FollowedbyStop* also had an effect $[\hat{\beta} = 0.23, t(716) = 13.10, p < 0.0001]$, which supports the observation by Waals (1999) that an unstressed vowel is pronounced longer before oral stops. It is noteworthy that Waals' observation, which was made under thoroughly controlled laboratory conditions, is replicated here in more natural read aloud speech.

All significant predictors in the model were checked for non-linearities, none of which reached significance. The bootstrap validated R^2 value for the model was 0.72. The unique contribution of the morphological predictors *TypeEnBias* and *RightPositionalEntropy* to the variance explained by the model was 2.3%, as indicated by the drop in R^2 after the removal of these variables from the model. This contribution is close to that provided by the morpholexical predictors in the *s*-dataset (2.0%).

D. Discussion

The analysis of the *en*-dataset replicates the unexpected direction of the influence of the morphologically-determined redundancy that we reported for the dataset with the interfix *-s-*: We found again that higher values for the bias estimates correlate with a longer duration of the interfix. We will return to this role of the bias in the General Discussion.

The positive simple main effect of the right positional entropy supports the hypothesis of continuous planning of articulation, according to which the planning complexity of upcoming elements may modulate acoustic characteristics of preceding elements.

Given the dominant contribution of the variable *NPresent* to the explained variance, we set out to establish what factors affected the selection of the variant [ən] versus [ə]. The interfix *-e(n)-* is spelled as either *-e-* or *-en-*, depending on orthographic rules. Compounds spelled just with *-e-* are unlikely to be pronounced with [ən]. The subset of compounds spelled with *-en-* contains 653 tokens. We fitted a logistic regression model that predicted the log odds of the selection of [ən] versus [ə] in this subset. The model uses the binomial link function and considers the presence of [n] in the realization of the interfix as a success, and its absence as a failure. The results demonstrate no effect of *TypeEnBias* on the selection of the phonetic variant ($p > 0.5$). Apparently the realization of an extra phoneme in the interfix is independent of the morphological likelihood of the interfix. The presence of [n] was more likely when *WordFrequency* was high [$\hat{\beta} = 0.63, p < 0.0001$], *RightPositionalEntropy* was high [$\hat{\beta} = 2.11, p < 0.0001$], the speaker's language was Southern Dutch [$\hat{\beta} = 1.37, p < 0.0001$], the number of segments after the interfix, *AfterSegments*, was high [$\hat{\beta} = 2.06, p < 0.0001$], and a stress clash was attenuated [$\hat{\beta} = 4.19, p < 0.001$]. The likelihood of [n] was lower when *LeftPositionalEntropy* was high [$\hat{\beta} = -0.60, p < 0.0001$].

In a second supplementary analysis, we investigated whether morpholexical factors are better predictors for acoustic duration if we consider the duration of [ə] as the dependent variable, rather than the duration of the whole interfix. In such a model, we expect the presence of [n] to exercise less influence and the morpholexical predictors to have greater

explanatory value than in the model for the duration of the interfix as a whole. We fitted a stepwise multiple regression model to the data with the (natural) log-transformed acoustic duration of the phoneme [ə] in the interfix as the dependent variable. After removal of 25 outliers, the model was refitted against the remaining 717 datapoints.

In line with our expectations, we observe a decrease in the predictive power of *NPresent* to only 15% of the explained variance, while the share of morphological variables *TypeEnBias* and *RightPositionalEntropy*, which retain significance as predictors of acoustic lenghtening, increases to 4.3% of the explained variance. We conclude that morphological structure code-termines the acoustic characteristics of the interfix *-e(n)-* over and beyond major phonological and phonetic predictors. [1]

VI. GENERAL DISCUSSION

According to the information-theoretical approach to acoustic salience developed in the last decade, a higher likelihood of a linguistic unit is correlated with more acoustic reduction. The main finding of the present study is that the effect of morphologically-determined probability on the duration of interfixes in Dutch compounds runs counter to this prediction. This pattern of results is especially puzzling, since our data also provide evidence *in favor of* the information-theoretical approach in the form of an effect of segmental lexical information. Thus, we do find that a higher probability of a segment given the preceding word fragment leads to more acoustic reduction.

The speakers in the Spoken Dutch Corpus read the compounds and thus received un-ambiguous visual information about the correct interfix. It is therefore remarkable that we nevertheless observed effects of morpholexical factors on the planning and implementation of speech production. We note, however, that the bias of the interfix as determined by the left constituent family is known to predict the speed of reading comprehension of novel and existing compounds (Krott, Hagoort and Baayen, 2004). We therefore expect the acoustic consequences of the bias to have a larger scope when visual cues to the appropriate

morphemes are absent, as in spontaneous speech genres.

What may be the solution for the problem that the present data appear to pose for the information-theoretical framework? One explanation might be that morphological information has a fundamentally different status from other types of linguistic information, and is typically associated with careful articulation. However, this line of reasoning is refuted by research on prefixes and suffixes in English (e.g., Hay, 2003) and Dutch (e.g., Pluymaekers *et al.*, 2005a, Pluymaekers *et al.*, 2005b).

Another solution might refer to the fact that interfixes are homophonous with plural markers in Dutch (cf., *boek-en* "books" and the compound *boek-en-kast* "bookshelf"). The frequency of the plural word forms might codetermine the duration of the interfix and be confounded with the bias. This explanation, however, can be discarded on the following grounds. First, there was no consistency in the correlation between the frequency of plural nouns and the bias of the interfix across datasets. For the -*s*-dataset the correlation was positive [$r = 0.12, t(1154) = 4.24, p < 0.0001$], while for the -*en*-dataset it was negative [$r = -0.28, t(740) = -8.15, p < 0.0001$]. Second, the frequency of the plural homophonous forms did not reach significance when included as a covariate in the regression models for both datasets. Finally, previous work on German compounds by Koester, Gunter, Wagner and Friederici (2004) has shown that plural suffixes and interfixes may not be perfectly homophonous in terms of systematic fine phonetic detail: Compound constituents followed by an interfix are shorter and have a higher pitch than their stand-alone plural counterparts.

The hypothesis that we would like to offer as a solution for the present paradox is that fine phonetic detail in speech is governed by two orthogonal dimensions, a syntagmatic dimension and a paradigmatic dimension. The information-theoretical approach that underlies the Smooth Signal Redundancy Hypothesis (Aylett and Turk, 2004) and the Probabilistic Reduction Hypothesis (Jurafsky *et al.*, 2001), as well as research on speech efficiency (Van Son and Pols, 2003; Van Son and Van Santen, 2005), views information from the syntagmatic perspective by considering the probability of a linguistic unit in its phonetic, lexical, or syntactic context. These syntagmatic relationships are inherently sequential and govern

the temporal distribution of information in the speech stream. For instance, the extent to which a segment contributes to the identification of the word *given the preceding word fragment* (Van Son and Pols, 2003) is a syntagmatic measure that is positively correlated with duration: The greater the contribution of the segment, the longer its acoustic implementation.

The syntagmatic measures proceed upon the premise that there is no (probabilistic) variation in the elements forming the word or the syntactic clause to be realized by the speaker. When the speaker wants to express the concept *book*, there is no doubt that the element following [bʊ] is [k].

However, the identity of the elements is not always known with such certainty: The interfix in Dutch compounds is one such example. We label such elements "pockets of indeterminacy". Paradigmatic relations, here defined over constituent families, provide the probabilistic basis for resolving this indeterminacy. The bias measures quantify the extent of support provided by paradigmatics for the different interfixes available for selection: A greater support increases the likelihood of a given interfix. Our experimental results indicate that such a greater likelihood is paired with a longer acoustic realization. Moreover, we have shown that a higher frequency of a compound correlates with an increased chance of a more salient realization of the interfix -*e(n)*- as [ən], rather than [ə].

Whereas the syntagmatic dynamics of lexical disambiguation are intrinsically temporal, paradigmatic inference is a-temporal in nature. In the a-temporal domain of paradigmatic inference for positions of choice, a greater probability implies a broader empirical basis for selection of a given alternative, and comes with increased acoustic duration.

Importantly, paradigms as a source of support for alternatives for selection are not restricted to morphological structure: We consider paradigms in a general Saussurean sense, as sets of linguistic elements over which the operation of selection is defined (de Saussure, 1966).

The amount of evidence for the alternatives apparently determines the confidence with which an interfix is selected. That a lack of confidence may lead to a decrease in acoustic

duration may be illustrated by an analogy: When producing case endings of German nouns, non-native speakers of German may hush up their realizations if they have doubts about the appropriate morpheme, but articulate the endings carefully and clearly if they are certain about which ending to choose. This example serves as an analogy only, and there is no implication that speakers make deliberate, conscious choices based on the morphological bias. The support measured as the bias is rather an estimate of the "naturalness" of the association between the available interfixes and the constituents of the compound.

Our hypothesis that paradigmatic inference for pockets of indeterminacy leads to longer (or otherwise more salient) realizations, henceforth the Paradigmatic Signal Enhancement Hypothesis, offers straightforward, testable predictions at various levels of linguistic structure. First consider the level of morphology. It is well known that English irregular verbs cluster into sets according to the kind of vocalic alternation that they exhibit in the past tense form (*keep/kept, run/ran*). The Paradigmatic Signal Enhancement Hypothesis predicts that a past-tense vowel — a pocket of indeterminacy — is realized with increased acoustic salience when the vocalic alternation is supported by a larger set of irregular verbs. Effects of paradigmatic gangs might even be found for the vowels of regular verbs (Albright and Hayes, 2003).

At the interface of morphology and phonology, we call attention to the phenomenon of final devoicing. In German and Dutch, a stem-final obstruent may alternate between voiced and voiceless, compare Dutch [hɔnt] *hond* ('dog') with [hɔndə] *honden* ('dogs'). Ernestus and Baayen (2003, 2004) have shown that this alternation, traditionally regarded as idiosyncratic, is affected by paradigmatic structures driven by the rhyme of the final syllable. In addition, they have shown that devoiced obstruents (e.g., the [t] of [hɔnt]) may carry residual traces of voicing, and that listeners are sensitive to these residual traces (Ernestus and Baayen, 2006). The Paradigmatic Signal Enhancement Hypothesis builds on these findings by predicting that greater paradigmatic support for voicing will correlate with enhanced acoustic salience of residual voicing in the devoiced obstruent.

Additional evidence for the Paradigmatic Signal Enhancement Hypothesis emerges from

25

research on intrusive /r/ in New Zealand English (Hay and Maclagan, in press): The more likely speakers are to produce intrusive /r/ given a range of linguistic and sociolinguistic factors, the more salient its realization (as reflected in the degree of constriction).

Finally, the probabilistic dependencies between morphemes, such as exist between the interfix, the compound's left and right constituents, and the whole compound, challenge the fully decompositional theory of morphological encoding in speech production, developed by Levelt, Roelofs and Meyer (1999). According to this model, an abstract lemma representation provides access to a word's individual constituents. The planning for articulation of these individual constituents is fully encapsulated from all other morphemes and their paradigmatic relations. This model is challenged not only by the present findings, but also by those of Van Son and Pols (2003), Pluymaekers *et al.* (2005a), Pluymaekers *et al.* (2005b), Hay (2003), and Ernestus *et al.* (2006). What the present paper adds to this literature is the surprising observation that fine phonetic detail is not only determined by the properties of the word itself and its nearest phonological neighbors, but also by its morphological paradigmatic structure.

ACKNOWLEDGMENTS

Endnotes

1. If a compound is spelled with -*e(n)*-, it can be realized as [ən] or [ə] in speech. We have shown that a higher word frequency favors the presence of [n] in the realization of the interfix. Might it be the case that the realization of the interfix as [ə] is longer in a compound that is more often realized with [ən]? To check this possibility, we

computed the percentage of tokens realized as [ən] for each -e(n)-compound. This percentage was not a significant predictor of acoustic duration of [ə] (p > 0.05). Thus we rule out an impact of the relative frequency of [ən]-realization (more probable in read speech) on [ə]-realization (more probable in spontaneous speech).

References

Albright, A. and Hayes, B. (**2003**). "Rules vs. analogy in English past tenses: A computational/experimental study", Cognit. **90**, 119–161.

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.

Aylett, M. and Turk, A. (**2006**). "Language redundancy predicts syllabic duration and the spectral characteristics of vocalic syllabic nuclei", J. Acoust. Soc. Am. **119**, 3048–3058.

Baayen, R. H., Piepenbrock, R., and Gulikers, L. (**1995**). *The CELEX lexical database (CD-ROM)* (Linguistic Data Consortium, University of Pennsylvania, Philadelphia, PA).

Bard, E., Anderson, A., Sotillo, C., Aylett, M., Doherty-Sneddon, G., and Newlands, A. (**2000**). "Controlling the Intelligibility of Referring Expressions in Dialogue", J. Mem. Lang. **42**, 1–22.

Bell, A., Jurafsky, D., Fosler-Lussier, E., Girand, C., 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.

Bertram, R. and Hyönä, Y. (**2003**). "The length of a complex word modifies the role of morphological structure: Evidence from eye movements when reading short and long finnish compounds", J. Mem. Lang. **48**, 615–634.

Bien, H., Levelt, W., and Baayen, R. (**2005**). "Frequency effects in compound production", PNAS **102**, 17876–17881.

Booij, G. (**2002**). *The Morphology of Dutch* (Oxford University Press, Oxford).

Cambier-Langeveld, T. (**2000**). *Temporal marking of accents and boundaries* (LOT, Amsterdam).

Crystal, T. and House, A. (**1990**). "Articulation rate and the duration of syllables and stress groups in connected speech", J. Acoust. Soc. Am. **88**, 101–112.

Cucchiarini, C. (**1993**). *Phonetic transcription: A methodological and empirical study* (University of Nijmegen, Nijmegen).

De Jong, N. H., Feldman, L. B., Schreuder, R., Pastizzo, M., and Baayen, R. H. (**2002**). "The processing and representation of Dutch and English compounds: Peripheral morphological, and central orthographic effects", Brain Lang. **81**, 555–567.

De Saussure, F. (**1966**). *Course in General Linguistics* (McGraw, New York).

Ernestus, M. (**2000**). *Voice assimilation and segment reduction in casual Dutch: A Corpus-Based Study of the Phonology-Phonetics Interface* (LOT, Utrecht).

Ernestus, M. and Baayen, R. H. (**2003**). "Predicting the unpredictable: Interpreting neutralized segments in Dutch", Lang **79(1)**, 5–38.

Ernestus, M. and Baayen, R. H. (**2004**). "Analogical effects in regular past tense production in Dutch", Ling **42**, 873–903.

Ernestus, M. and Baayen, R. (**2006**). "The functionality of incomplete neutralization in Dutch: The case of past-tense formation", in *Lab Phon 8*, edited by L. Goldstein, D. Whalen, and C. Best, 27–49 (Mouton de Gruyter, Berlin).

Ernestus, M., Lahey, M., Verhees, F., and Baayen, R. H. (**2006**). "Lexical frequency and voice assimilation", J. Acoust. Soc. Am. **120**, 1040–1051.

Fosler-Lussier, E. and Morgan, N. (**1999**). "Effects of speaking rate and word frequency on pronunciations in conversational speech", Speech Commun. **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. Mem. Lang. **26**, 489–504.

Hay, J. (**2003**). *Causes and Consequences of Word Structure* (Routledge, New York and

London).

Hay, J. and Maclagan, M. (**in press**). "Social and phonetic conditioners on the frequency and degree of "intrusive /r/" in New Zealand English", in *Methods in Sociophonetics*, edited by D. Preston and N. Niedzielski.

Hunnicutt, S. (**1985**). "Intelligibility versus redundancy - conditions of dependency", Lang. Speech **28**, 47–56.

Johnson, K. (**2004**). "Massive reduction in conversational American English", in *Spontaneous speech: data and analysis. Proceedings of the 1st session of the 10th international symposium*, 29–54 (The National International Institute for Japanese Language, Tokyo, Japan).

Jurafsky, D., Bell, A., Gregory, M., and Raymond, W. (**2001**). "Probabilistic relations between words: Evidence from reduction in lexical production", in *Frequency and the emergence of linguistic structure*, edited by J. Bybee and P. Hopper, 229–254 (Benjamins, Amsterdam).

Kessens, J. and Strik, H. (**2001**). "Lower WERs do not guarantee better transcriptions", in *Proc. Eurospeech-2001*, 1721–1724 (Aalborg, Denmark).

Keune, K., Ernestus, M., Van Hout, R., and Baayen, R. (**2005**). "Social, geographical, and register variation in Dutch: From written 'mogelijk' to spoken 'mok"', Corpus Ling. Ling. Theory **1**, 183–223.

Koester, D., Gunter, T. C., Wagner, S., and Friederici, A. D. (**2004**). "Morphosyntax, prosody, and linking elements: the auditory processing of German nominal compounds", J. Cognit. Neurosci. **16**, 1647–1668.

Krott, A., Baayen, R. H., and Schreuder, R. (**2001**). "Analogy in morphology: modeling the choice of linking morphemes in Dutch", Ling. **39**, 51–93.

Krott, A., Hagoort, P., and Baayen, R. H. (**2004**). "Sublexical units and supralexical combinatorics in the processing of interfixed Dutch compounds", Lang. Cognit. Processes **19**, 453–471.

Krott, A., Krebbers, L., Schreuder, R., and Baayen, R. H. (**2002a**). "Semantic influence on

linkers in Dutch noun-noun compounds", Folia Linguis. **36**, 7–22.

Krott, A., Schreuder, R., and Baayen, H. (**2002**b). "Linking elements in Dutch noun-noun compounds: Constituent families as predictors for response latencies", Brain Lang. **81**, 708–722.

Ladefoged, P. (**1982**). *A course in phonetics*, 2 edition (Hartcourt, Brace, Jovanovich, New York).

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.

Lieberman, P. (**1963**). "Some effects of semantic and grammatical context on the production and perception of speech", Lang. Speech **6**, 172–187.

Lindblom, B. (**1990**). "Explaining phonetic variation: A sketch of the H&H theory", in *Speech production and speech modeling*, edited by W. Hardcastle and A. Marchal, 403–440 (Kluwer, Dordrecht).

Losiewicz, B. (**1992**). "The effect of frequency on linguistic morphology", Ph.D. thesis, University of Texas.

McAllister, J., Potts, A., Mason, K., and Marchant, G. (**1994**). "Word duration in monologue and dialog speech", Lang. Speech **37**, 393–405.

Moscoso del Prado Martín, F., Kostić, A., and Baayen, R. H. (**2004**). "Putting the bits together: An information theoretical perspective on morphological processing", Cognit. **94**, 1–18.

Munson, B. and Solomon, N. (**2004**). "The effect of phonological neighborhood density on vowel articulation", J. Speech, Lang. Hearing Research **47**, 1048–1058.

Neijt, A., Krebbers, R., and Fikkert, P. (**2002**). "Rhythm and semantics in the selection of linking elements", in *Linguistics in the Netherlands 2002*, edited by H. Broekhuis and P. Fikkert, 117–127 (Benjamins, Amsterdam).

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.

Pluymaekers, M., Ernestus, M., and Baayen, R. (**2005**a). "Articulatory planning is continuous and sensitive to informational redundancy", Phonetica **62**, 146–159.

Pluymaekers, M., Ernestus, M., and Baayen, R. (**2005**b). "Lexical frequency and acoustic reduction in spoken Dutch", J. Acoust. Soc. Am. **118(4)**, 2561–2569.

Samuel, A. and Troicki, M. (**1998**). "Articulation quality is inversely related to redundancy when children or adults have verbal control", J. Mem. Lang. **39**, 175–194.

Scarborough, D. L., Cortese, C., and Scarborough, H. S. (**1977**). "Frequency and repetition effects in lexical memory", J. Exp. Psychol.: Human Percept. Perform. **3**, 1–17.

Scarborough, R. (**2004**). "Degree of Coarticulation and Lexical Confusability", in *Proceedings of the 29th Meeting of the Berkeley Linguis. Soc., February 14-17, 2003.*

Shannon, C. E. (**1948**). "A mathematical theory of communication", Bell System Technical J. **27**, 379–423.

Shields, L. and Balota, D. (**1991**). "Repetition and associative context effects in speech production", Lang. Speech **34**, 47–55.

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

Van Son, R., Binnenpoorte, D., van den Heuvel, H., and Pols, L. (**2001**). "The IFA Corpus: a phonemically segmented Dutch Open Source speech database", in *Proc. Eurospeech-2001* (Aalborg, Denmark).

Van Son, R. and Pols, L. (**2003**). "Information Structure and Efficiency in Speech Production", in *Proceedings of Eurospeech-2003* (Geneva, Switzerland).

Van Son, R. and Van Santen, J. (**2005**). "Duration and spectral balance of intervocalic consonants: A case for efficient communication", Speech Commun. **47**, 100–123.

Vitevitch, M. S. (**2002**). "The influence of phonological similarity neighborhoods on speech production", J. Experiment. Psychol.: Learning, Mem. and Cognit. **28**, 735–747.

Vorstermans, A., Martens, J., and Van Coile, B. (**1996**). "Automatic segmentation and labelling of multi-lingual speech data", Speech Commun. **19**, 271–293.

Waals, J. (**1999**). *An experimental view of the Dutch syllable* (Holland Academic Graphics,

The Hague).

Young, S., Evermann, G., Hain, T., Kershaw, D., Moore, G., Odell, J., Ollason, D., Povey, D., Valtchev, V., and Woodland, P. (**2002**). *The HTK Book 3.2* (Entropic, Cambridge).

Zipf, G. K. (**1929**). "Relative frequency as a determinant of phonetic change", Harv. Studies Classical Philol. **15**, 1–95.

List of Figures