Effects of Word Frequency on the Acoustic Durations of Affixes

Mark Pluymaekers, Mirjam Ernestus, R. Harald Baayen

Interfaculty Research Unit for Language and Speech
Radboud University Nijmegen & Max Planck Institute for Psycholinguistics,
Nijmegen, The Netherlands

{Mark.Pluymaekers, Mirjam.Ernestus, Harald.Baayen}@mpi.nl

Abstract

This study investigates whether the acoustic durations of derivational affixes in Dutch are affected by the frequency of the word they occur in. In a word naming experiment, subjects were presented with a large number of words containing one of the affixes *ge-*, *ver-*, *ont*, or *-lijk*. Their responses were recorded on DAT tapes, and the durations of the affixes were measured using Automatic Speech Recognition technology. To investigate whether frequency also affected durations when speech rate was high, the presentation rate of the stimuli was varied. The results show that a higher frequency of the word as a whole led to shorter acoustic realizations for all affixes. Furthermore, affixes became shorter as the presentation rate of the stimuli increased. There was no interaction between word frequency and presentation rate, suggesting that the frequency effect also applies in situations in which the speed of articulation is very high.

Index terms: speech production, word frequency, acoustic duration

1. Introduction

Lexical frequency has long been recognized as a possible predictor of acoustic duration. High-frequency words are hypothesized to be shorter than low-frequency words, because speakers would save articulatory effort when producing them. Thus far, most evidence for this relationship has come from studies investigating phoneme durations. In [1], for example, it was shown that word-initial [s]-es in American English are shorter if the frequency of their carrier word is high. In [2], vowels occurring in high-frequency words were found to be shorter than the same vowels occurring in low-frequency words. In [3], the durations of word-final [d] and [t] in English were investigated. Again, a higher frequency of the carrier word was correlated with shorter durations, even after partialling out the effects of speech rate, segmental context, and number of syllables.

Recently, effects of word frequency have been demonstrated on units larger than the phoneme. The study described in [4] investigated four derivational affixes in Dutch, all occurring in a large number of words with different frequencies. Materials were gathered from a large database of face-to-face conversations. For each word type in the database containing one of the affixes, one token was randomly selected for acoustic analysis. Measurements were made of the durations of the affixes and their individual segments. For three of the four affixes under investigation, a higher frequency of occurrence of the carrier word led to shorter realizations of the affix as a whole, individual segments in the affix, or both. This showed that word frequency can also affect the durations of units that carry meaning, such as affixes.

Only a few studies have investigated frequency effects on durations experimentally. [5] showed that lists containing high-frequency words had shorter total reading times than lists containing low-frequency words, but, as argued in [6], this total reading time included pauses between words and could therefore not provide conclusive evidence about acoustic durations. A more serious methodological problem in both [5] and [6] was that high- and low-frequency words were matched on length in letters only. As most speech researchers will agree, this does not provide enough experimental control for comparing durations. More sophisticated matching procedures were used in [7], but these came with the disadvantage that they precluded the authors from drawing conclusions about the durations of complete words. Instead, effects of word frequency could only be demonstrated on the durations of onsets.

In the current study, we extend the methodology used in [4] to the more controlled environment of a word naming experiment. More specifically, we investigate whether the durations of four Dutch derivational affixes differ as a function of the frequency of the word they occur in. This provides perfect phonetic control, as well as the possibility to study a large number of words spanning the whole frequency range. Furthermore, we investigate whether there are limits to the circumstances in which the word frequency effect applies. In [4], the frequency effect was sometimes found to be absent if speech rate was high. This could point to the existence of an upper threshold of articulation speed, above which frequency no longer plays a role. In the current experiment, we tested this hypothesis by varying the presentation rate of the stimuli.

2. Method

2.1. Participants

Twenty-one subjects participated in the experiment in exchange for pay. Most of them were undergraduate students from the participant pool of the Max Planck Institute for Psycholinguistics. There were 11 male and 10 female participants, all of whom were native speakers of Dutch. All had normal or corrected-to-normal vision.

2.2. Materials

The affixes under investigation were the prefixes *ge-*, *ver-*, and *ont-*, and the suffix *-lijk*. All of these affixes are unstressed. For each affix, we selected 60 words containing that affix. The selected target words spanned the entire range of word frequencies and always consisted of three syllables. To draw attention away from the affixes, 210 filler words were included. These fillers were mono-, bi-, or trisyllabic, and word stress alternated between the first, second, and third syllable. All words in the experiment had regular spelling.

The 240 targets and 210 fillers were divided into 15 lists of 30 words each. Each list contained 16 randomly selected targets (four for each affix) and 14 randomly selected fillers. The first five words on a list were always fillers, and the remaining words were ordered so that two targets containing the same affix were never directly adjacent. All words occurred just once in the experiment.

2.3. Design

The main experimental manipulation was the presentation rate of the stimuli. Three different rates were employed. In the slow condition, 1500 milliseconds separated the onsets of two subsequent stimuli. In the medium condition, this interval was 1100 milliseconds, while the fast condition allowed participants only 700 milliseconds to give their response.

Participants produced 150 different words in each condition. This means that presentation rate was varied not within, but between subjects. The words were divided over the conditions so that over the whole experiment (21 participants), each word occurred in each condition exactly seven times. Participants always started with the slow condition, followed by the medium condition and the fast condition. The main reason for keeping the order of conditions constant was our concern that participants who had seen a number of lists in the fast condition might not be able to adopt a slower pace for subsequent lists.

2.4. Procedure

Participants were seated in a sound-attenuated booth, behind a Sennheiser microphone. The words were presented on a computer screen, and participants were required to name them. Participants' responses were recorded on a Tascam DA-20 DAT tape recorder and digitized with a sampling frequency of 16 kHz. In addition, naming latencies were registered by means of a voice key device.

The procedure was the same for all participants. First, five lists were presented in the slow condition, followed by five lists in the medium condition. Finally, the remaining five lists were presented in the fast condition. Between two lists in the same condition, participants were allowed to pause for a few seconds. Between two conditions, they were allowed to leave the booth. During these longer breaks, participants were reminded by the experimenter that the presentation rate of the next set of lists would be considerably faster than that of the previous set, but that they should be able to keep up with it if they reacted as quickly as possible. To give participants an idea about the tempo they should adopt, each list was preceded by five fixation points presented at the same rate as the words in that list.

2.5. Acoustic measurements

The current experiment generated an enormous amount of speech data. Since hand-measuring the durations of all affixes would be extremely time-costly, we decided to use Automatic Speech Recognition (ASR) technology. Recent research has shown that the reliability of segmentations generated by an ASR system is equal to that of segmentations made by human transcribers [8,9], provided that a phonemic transcription of the signal is available to the segmentation algorithm.

We trained a Hidden Markov Model (HMM) speech recognizer using the software package HTK [10]. To optimize the ASR's performance on segmentation, we used context-independent, continuous density HMMs with 32 gaussians per state [11]. In total, 37 phone models were trained, representing the 36 phonemes of Dutch and silence. The training material was taken from the phonemically transcribed portion of the subcorpus 'Library for the blind' of the Corpus of Spoken Dutch [12]. In total, the training sample consisted of 13328 read utterances produced by 134 different speakers. The combined duration of these utterances was 6 hours and 39 minutes.

Acoustic analysis proceeded in two steps. First, the sound file containing a particular target word was parameterized using Mel Frequency Cepstral Coefficients. Then, the parameterized signal was provided to a Viterbi segmentation algorithm, together with a phonemic transcription of the word. This transcription was determined on the basis of two separate reference transcriptions, made by two different transcribers. If the transcribers agreed, the consensus transcription was used. If there was disagreement, a third transcriber picked the best transcription of the two.

The reliability of the ASR was examined in a pre-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 randomly selected words from the current experiment. Comparison between the ASR-generated and hand-made segmentations revealed that 76% of the automatic boundaries were placed within 20 milliseconds of the corresponding hand-coded boundary. The main discrepancies involved the beginnings of plosives and liquids, which were generally placed earlier by the ASR than by the phonetician. If these automatic boundaries were shifted 10 and 7 milliseconds to the right, respectively, the percentage of boundaries placed within 20 milliseconds of each other increased to 81%. This level of accuracy is in accordance with international standards [8,9], and was considered sufficient for the present purposes.

3. Results

3.1. Statistical analysis

Words that were preceded by a hesitation or mispronunciation were excluded from the data set, as were words for which the naming latency had not been registered. To the remaining 3981 data points, we fitted a linear mixed effect model with participant and word as random effects. This analysis was performed using the lme4 library [13] and the statistical software package R. The two main predictor variables were the log frequency of the stimulus, obtained from the word frequency list of the Corpus of Spoken Dutch, and the presentation rate (Slow, Medium, or Fast). The dependent variable was the log duration of the affix. In addition to the two predictor variables, the following control variables were entered into the model:

- Affix (*ge-*, *ver-*, *ont-*, or *-lijk*)
- Length of the stimulus in letters
- Number of realized segments in the stem
- Position of the word in the list
- Sex of the speaker
- Frequency of the previous word in the list

Variables and interaction terms between variables were retained only if they showed a significant effect. Data points for

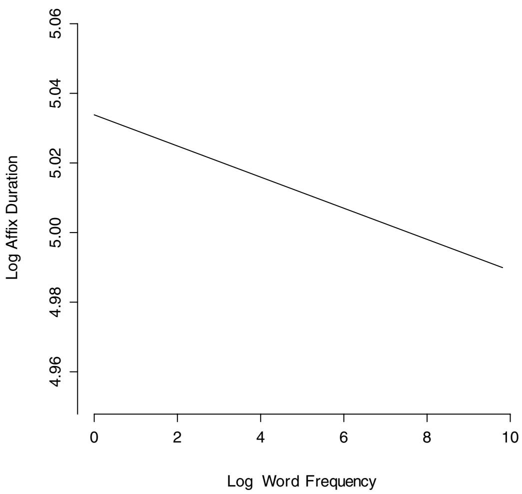

Figure 1: *The effect of word frequency on affix duration when the other significant variables are held constant.*

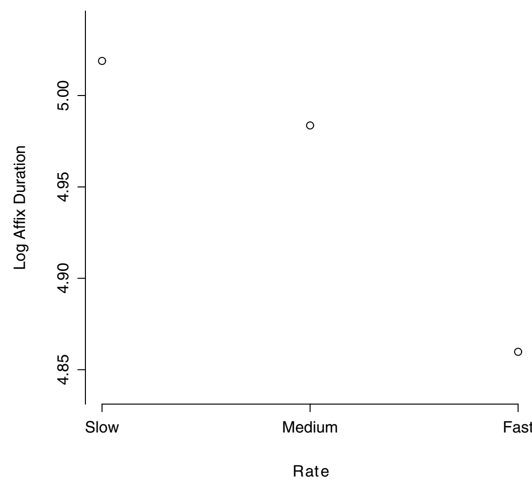

Figure 2: *The effect of presentation rate on affix duration when the other significant variables are held constant.*

which the residual value was more than 2.5 times the standard deviation of residuals were considered outliers and removed from the data set (89 data points, 2% of the complete data set). Then, the model was refitted to the remaining data.

3.2. Regression results

In the final model, the following significant effects were observed:

- The higher the lexical frequency of the word, the shorter the duration of the affix ($\hat{\beta} = -0.04, t(3882) = -2.57, p < 0.05$). This effect is illustrated in Figure 1, which plots the log frequency of the word against the log duration of the affix when the other significant variables are held constant.

- The faster the presentation rate, the shorter the duration of the affix (Slow vs. Medium: $\hat{\beta} = 0.04, t(3882) = -4.79, p < 0.0001$; Slow vs. Fast: $\hat{\beta} = 0.14, t(3882) = -17.65, p < 0.0001$; Medium vs. Fast: $\hat{\beta} = 0.11, t(3882) = -13.39, p < 0.0001$). This effect is shown in Figure 2.

- The more letters in the stimulus, the shorter the duration of the affix ($\hat{\beta} = -0.03, t(3882) = -3.41, p < 0.001$)
- The higher the number of realized segments in the stem, the shorter the duration of the affix ($\hat{\beta} = -0.04, t(3882) = -3.44, p < 0.001$). This is a well-known phonetic effect, previously reported in e.g. [14].
- Affixes differed with respect to their (intrinsic) durations ($F(3, 3882) = 400.1, p < 0.0001$).
- Word frequency interacted with the number of realized segments in the stem ($\hat{\beta} = 0.006, t(3882) = 2.28, p < 0.05$). The effect of frequency became smaller as the number of segments in the stem increased.

There were no effects of position in the list, sex of the speaker, or frequency of occurrence of the previous word. Furthermore, word frequency did not interact with affix or presentation rate.

4. Discussion

The results presented above show that the acoustic durations of the Dutch affixes *ge-*, *ver-*, *ont-*, and *-lijk* are indeed affected by word frequency. The higher the frequency of the word the affix occurs in, the shorter the affix. This finding replicates for laboratory speech the findings reported in [4] for spontaneous speech. However, the results of the two studies are not completely identical. For example, in the current study no interaction was observed between word frequency and affix, suggesting that the effect of frequency was equally strong for all four affixes. In [4], however, no frequency effects were found for the prefix *ver-*. There are a number of possible explanations for this difference. First of all, the design of the current experiment yielded considerably more data points per affix than that of the corpus study. The current data set contained 964 tokens beginning with *ver-*, compared to 140 tokens in the data set analyzed in [4]. Furthermore, the target words were more evenly distributed over the frequency range, mainly because we had now selected the words ourselves. Finally, the nature of the materials was completely different in the two studies. Presumably, the clearly recorded laboratory speech generated in the current experiment offered better acoustic conditions for finding effects of word frequency than the noisy, partly overlapping speech studied in the corpus survey. Moreover, many more factors need to be controlled for in spontaneous speech than in laboratory speech.

Presentation rate also showed a significant effect on affix durations. Durations were shorter in the Medium condition than in the Slow condition, and shorter in the Fast condition than in the Medium condition. This shows that our manipulation of presentation rate succeeded in speeding up articulation. However, our main interest was in whether presentation rate would interact with word frequency. This turned out not to be the case: The frequency effect was equally strong for each of the three presentation rates. Surprisingly, there is no upper threshold of articulation speed above which the frequency effect does no longer apply. On the other hand, the interaction we observed between frequency and the number of realized segments in the stem shows that the frequency effect becomes smaller as the number of other segments that need to be pronounced increases. This suggests that the magnitude of the frequency effect is not modulated by the time that is available to articulate the word, but rather by the total number of segments that need to be produced, even if there is plenty of time available.

Finally, the results of the current study indicate that ASR technology can successfully be used to segment speech data obtained in production experiments. Even though the minimum intrinsic durations specified by the HMM's topology lead to a certain loss of sensitivity, there is still sufficient variation present in the data set to reliably demonstrate effects of word frequency and speech rate on acoustic durations. On the other hand, existing ASR systems could benefit from the knowledge that word frequency is a robust source of pronunciation variation in meaningful units such as affixes.

5. Conclusions and future directions

The following conclusions can be drawn from this study:

- Derivational affixes in Dutch are shorter the higher the lexical frequency of the word they occur in.
- This is true for laboratory speech as well as spontaneous speech.
- In laboratory speech, the frequency effect also applies at very high speech rates.
- ASR technology can be used to investigate the effects of word frequency and other relevant variables on acoustic durations.

With regard to future research, it would be interesting to investigate whether a similar effect of word frequency on durations can be observed for affixes or other meaningful units in languages other than Dutch. From a psycholinguistic perspective, the most pressing issue at this point is to determine the processing stage at which this effect, which has not yet been accounted for in speech production theories, arises. Further experimental research is needed to shed more light on this issue.

6. Acknowledgement

This research was supported by Netherlands Organization for Scientific Research (NWO) grant number 360-70-130 to the third author.

7. References

[1] Umeda, N., "Consonant duration in American English", *J. Acoust. Soc. Am.*, 61:846–858, 1977.

[2] Van Coile, B., "A model of phoneme durations based on the analysis of a read Dutch text", *Proceed. Europ. Conf. Speech Techn.*, 2:233–236, 1987.

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

[4] Pluymaekers, M., Ernestus, M., and Baayen, R.H., "Lexical frequency and acoustic reduction in Dutch", *J. Acoust. Soc. Am.*, 118:2561–2569, 2005.

[5] Wright,C.E., "Duration differences between rare and common words and their implications for the interpretation of word frequency effects", *Mem. & Cogn.*, 7:411–419, 1979.

[6] Geffen, G. and Luszcz, M.A., "Are the spoken durations of rare words longer than those of common words?", *Mem. & Cogn.*, 11:13–15, 1983.

[7] Kawamoto, A.H., Kello, C.T., Higareda, I., and Vu, J.V.Q., "Parallel Processing and Initial Phoneme Criterion in Naming Words: Evidence from Frequency Effects on Onset and Rime Duration", *J. Exp. Psych.: Learn., Mem., & Cogn.*, 25:362–381, 1999.

[8] Vorstermans, A., Martens, J.P., and Van Coile, B., "Automatic segmentation and labelling of multi-lingual speech data", *Speech Communication*, 19:271–293, 1996.

[9] Sjölander, K., "Automatic alignment of phonetic segments", *Lund University, Dept. of Linguistics Working Papers*, 49:140–143, 2001.

[10] Young, S., Evermann, G., Hain, T., Kershaw, D., Moore, G., Odell, J., Ollason, D., Povey, D., Valtchev, V., Woodland, P., *The HTK Book*, Entropic, Cambridge, 2002.

[11] Kessens, J.M. and Strik, H., "Lower WERs do not guarantee better transcriptions", *Proc. Eurospeech*, pp. 1721–1724, 2001.

[12] Oostdijk, N. "The Spoken Dutch Corpus Project", *ELRA Newsletter*, 5:4–8, 2000.

[13] Bates, D.M. and Sarkar, D., "The lme4 library", available on-line at http://lib.stat.cmu.edu/R/CRAN, 2005.

[14] Nooteboom, S., "Production and perception of vowel duration: a study of durational properties of vowels in Dutch", Ph.D. thesis, Utrecht University, The Netherlands, 1972.