The Domain of Final Lengthening in the Production of Dutch

Tina Cambier-Langeveld

0. Introduction

Speakers have numerous ways at their disposal to mark the prosodic structure of their speech. The cues signalling this structure are potentially useful to the listener. One way for speakers to signal prosodic structure is by marking the depth of a boundary between two constituents, using a variety of prosodic features. For example, the end of an intonational phrase is typically marked by a pause, a boundary-marking pitch movement and/or lengthening of the preboundary segments (Downing 1970). Which prosodic features are used to mark a certain boundary may depend on the language and on the type of boundary involved, but is also to some degree variable, e.g. across speakers (Sanderman 1996). We will refer to the lengthening of preboundary segments as final lengthening.

In the past, final lengthening was thought to indicate syntactic boundaries (Klatt 1975). With the development of prosodic phonology, it turned out that the application of final lengthening is triggered by prosodic boundaries (just like other phonetic processes) rather than by syntactic ones. A final lengthening effect at utterance boundaries has been firmly established by numerous phonetic studies (Crystal and House 1988). Other prosodic boundaries are marked by final lengthening as well (Beckman and Edwards 1990). Furthermore, the amount of final lengthening has been shown to be related to the depth of the following boundary, both as produced by the speaker and as required by the listener (Nooteboom and Doodeman 1980, Gussenhoven and Rietveld 1992). Thus, a deeper prosodic boundary is, and should be, marked by more final lengthening. However, very little attention has been paid to the question which segments are affected; both production and perception experiments have generally been concerned with lengthening within a fixed domain (such as the vowel in the final syllable, or the final rhyme). So, previous research has shown that at least in the word-final rhyme the amount of lengthening is related to boundary depth, but we do not know whether boundary depth has any effect on the size of the domain that is lengthened. This leads us to the present research question:

---

1 This research was supported by a scholarship from the Holland Institute of Generative Linguistics/University of Amsterdam. The work was done under the supervision of Marina Nespor and Vincent van Heuven.
Is there a relationship between boundary depth and the domain of final lengthening?

In this paper, we will be looking at the domain of final lengthening in the production of Dutch only; future research will be concerned with perception. Work on the production of final lengthening has often been part of more general studies (Klatt 1975, Crystal and House 1988), which often limited their attention to the vowel of the word-final syllable, and only compared utterance final against non-utterance final durations. However, studies of final lengthening which focus on the duration of the vowel may be overlooking a significant aspect of the lengthening effect, since the effect is progressively distributed from left to right across a number of segments (Berkovits 1994).

One study which is concerned with the domain of final lengthening in production is an American English data-base study by Wightman et al. (1992). Their results seem to indicate that in speech production final lengthening is confined to the rhyme of the final syllable. However, their analysis considers only four pre-boundary units: the final coda, consonants (if any), the final vowel, any segments between the last stressed vowel and the final vowel, and the last stressed vowel. The third increment (from 'final vowel' to 'any segments between the last stressed vowel and the final vowel'), however, may be too crude to reveal small effects in more than the final rhyme, e.g. in the final onset or the pre-final rhyme. Another problem with this study, as often with data-base studies, is that factors which may influence the effect under observation are overlooked, such as word structure, (final-)syllable weight, word stress, etc. The usage of controlled material will allow us to answer the following question:

Do stress and the structure of the final syllable/word have any effect on the domain of final lengthening?

The results of a production experiment by Hofhuis (in prep.) indicate that final syllable structure indeed influences the size of the lengthened unit in Dutch. Her data consist of monosyllabic and disyllabic words, where the latter differ from the former only in the addition of a word-final schwa, occurring at five different boundaries. In the bisyllabic words, an effect of boundary depth on segment duration was found not only in the final syllable, but also in the penultimate rhyme (a full vowel), and sometimes even in the penultimate onset (i.e., in the whole word). This may be interpreted as evidence for the final foot (van der Hulst 1984) as the domain of lengthening, since in Dutch closed syllables or syllables with a final diphthong form a mono-syllabic foot, whereas syllables ending in a vowel may form one bisyllabic foot with a preceding syllable (Booij 1995). However, having only schwa-final bisyllabic words makes it impossible to differentiate between foot structure or vowel quality as the relevant factor determining the domain of final lengthening. Furthermore, the lengthening of the
penultimate syllable in Hofhuis' bisyllabic words can also be expressed as lengthening of the whole word, or as the last stressed syllable being the onset of final lengthening.

1. Approach

To answer the research questions in (1,2), material is needed in which a certain sequence of segments is followed by prosodic boundaries of different depths. Any other factors which may influence the duration of segments must be held constant. By placing the same string of segments at the crucial boundary every time, it is possible to abstract away from inherent differences in duration between segments and from the influence of adjacent segments. Prominence relations and overall speech rate must also be kept constant. Since prosodic features such as these cannot be represented in spelling, it was decided to present the material both visually (i.e. printed on paper) and auditorily. The auditory stimuli to be repeated by the subjects were meant to convey the desired (i.e. constant) intonation contour, prominence relations, speech rate etc. to the subjects. All temporal markers were eliminated from this input speech so that any temporal boundary markers in the subjects' reproduction must have been 'implemented' by the speakers themselves, thus reflecting natural lengthening effects in speech.

The above approach assumes that the correct prosodic features in the input speech will be copied, whereas the incorrect temporal structure will be automatically corrected by the speakers. To check whether this is indeed what subjects do, a pilot study was carried out prior to the main experiment. The pilot study also serves to give us some ideas for the set-up of the main experiment. The latter will include more words, which have segmentally varying final syllables in order to answer the research question in (2).

In section 2, a report is given of the preliminary experiment forming the pilot study; in section 3 the main experiment will be described.

2. Pilot Study

Introduction. The preliminary experiment described in this section not only serves as a try-out for the main experiment described in the following section, but is also necessary in order to answer some specific questions concerning the required form of the input for the repetition task.

In order to see if final lengthening interacts in any way with lengthening due to prominence (Sluijter and van Heuven 1995, and references given therein), an extra variable in the experiment will be focus distribution. The results will be used to determine if in the main experiment the target words should be placed in or out of focus (or both).
Material. The target word which is used for this experiment has four syllables: *rododendron* ‘rhododendron’. Such a long word allows us to distinguish between several possible domains of final lengthening such as the whole word, the final syllable *dron* (which is also the final foot), and *dendron*, i.e., the domain starting with the stressed syllable. The results should thus give us some idea of the domain(s) that should be looked for, and how the target words for the main experiment should be varied in order to answer the main research question in detail.

Four sentences containing the target word were composed in such a way that they contained the same number of syllables in total, as well as the same segmental material. Each sentence consists of an intonational phrase, followed by a parenthetical, i.e., by an additional remark. The parenthetical was realized without a pitch accent, so that every utterance contained only one pitch accent. When the target word is utterance-final, the parenthetical precedes the intonational phrase that contains the target word.

The four sentences, given in (3), differ in the type of boundary that follows the target word. The boundaries will henceforth be referred to as the PW (prosodic word), PHP (phonological phrase), IP (intonational phrase) and U (utterance) boundaries (Nespor and Vogel 1986):

(3) PW Piet wil die rare *rododendron*planten, gek als hij is.
‘Piet wants those strange rhododendron plants, crazy as he is’

PhP Piet wil die rare *rododendron* planten, gek als hij is.
‘Piet wants to plant that strange rhododendron, crazy as he is’

IP Piet wil die rare *rododendron*, plantengek als hij is.
‘Piet wants that strange rhododendron, plant crazy as he is’

U Plantengek als hij is wil Piet die rare *rododendron*.
‘Plant crazy as he is, Piet wants that strange rhododendron’

To answer the question whether final lengthening interacts in any way with lengthening due to prominence, each of the four sentences above occurs with two different focus distributions. Using precursor questions, either the target word or the proper name ‘Piet’ is put in focus. These introductory questions were spoken by the author and recorded directly onto a computer disk. Three different intonation contours were used in the input speech, but the position of the pitch accent was always in accordance with the focus distribution imposed on the sentence by the preceding question.

Diphone synthesis was used for the input speech, because it allows us to make speech without any temporal effects. A diphone consists of a transition between two phonemes, with roughly half a phoneme on either side (Dixon and Maxey, 1968). Thus, a word like ‘prosody’ will be built up from the following diphones, where # indicates silence: #p-pr-ro-os-so-od-di:-i:#. The idea behind diphone synthesis is that different realisations of one phoneme differ least in the middle of
that phoneme. In a sequence of diphones, the first half of one phoneme is joined to the second half of that same phoneme. Any discrepancies are then smaller and easier to adjust than those between two different phonemes. As each diphone is taken from the same (accented) position out of the same type of (nonsense) word, there are no temporal effects in the original set. Rules for lengthening have been added to the synthesis program (DS, Rijnsoever 1988), but were turned off for the present purposes. The diphone set used is bloem30_nostand.ind, made by Drulman and Collier, and spoken by the professional Dutch speaker Bloemendal. This diphone set affords the highest intelligibility for Dutch at this time (van Bezooijen and Pols 1993).

The complete material for the pilot study consists of 24 utterances (1 target word x 4 boundary depths x 2 focus distributions x 3 intonation contours).

Subjects. Three male and two female speakers of standard Dutch (ages 36 to 55) participated in the experiment. Two male and one female subjects were phonetically trained, while the other two had no phonetic background. Both trained and untrained subjects are included because for the main experiment it is important to know whether subjects should have some (phonetic) sophistication or not.

Procedure. The experiment was run in individual sessions. The material was presented to the subjects both auditorily using a Sony SS-E34 loudspeaker and visually on paper. The 24 question-and-answer pairs were randomized and presented to three subjects first in the order 1-24, and after a 10 minute break they were presented again, but in the reversed order 24-1. The other two subjects started with the sentences in the order 24-1 and continued with the order 1-24 after the break.

Subjects were seated in a sound insulated booth and instructed to repeat the answers only, with the same accentuation as they had just heard. Both the questions (as recorded by the author) and the following answers (produced by diphone synthesis) were played back by a Silicon Graphics computer (placed outside the recording booth), and the speech produced by the subjects was recorded directly onto computer disk through a Sennheiser MKH-416 directional condenser microphone. Warning tones indicated the beginning and end of the recording time (5 sec).

Before the experiment began, four practice examples were given which were taken from the actual material. This was done so that the subjects could get used to the type of speech and contrasts involved in the experiment and to the pace at which the material was presented.

Results. Every stimulus sentence was repeated twice by every speaker. Only the first recordings were segmented, while the second recording served as back-up in case a mistake or other problem had been overheard during the experiment. In these cases the second recording replaced the first. The results are thus based on
24 sentences x 5 speakers = 120 measurements. The target words were segmented according to the guidelines given by Rietveld and van Heuven (1997).

An analysis of variance (ANOVA) was performed on the data, with boundary type, focus distribution and intonation contour as fixed factors, with speaker as repeated measure and total word duration as the dependent variable.

Neither intonation contour nor focus distribution had any significant effect on the results (F[2,117]<1, and F[1,118]=2.1, p=.149, respectively). Only the type of boundary at which the target word occurs has a significant effect on the duration of the target word (F[3,116]=22.0; p<.001). On the basis of these findings, a series of oneway ANOVA's was run with segment duration as the dependent variable, and only boundary type as a fixed factor, to find out where the interaction lies between duration and boundary type, i.e., which segments of the word are lengthened in relation to the boundary at which the word occurs. The results of this analysis form the preliminary answer to our research question in (1) concerning the influence of boundary depth on the domain of final lengthening.

In Figure 1, the duration of each segment in the word *rododendron* for each of the four prosodic boundaries is given. On the x-axis, the number next to the segment indicates the syllable in which this segment occurs; o2 thus stands for the /o:/ in the second syllable. The silent interval between the target word and the following word is indicated by ‘si’; this interval is undefined for the U boundary, since no word follows the target word in that condition.

It is clear from Figure 1 that there is no effect of final lengthening until the final syllable of the word is reached; only in the final syllable a difference between the boundary types is found. In o4 and n4, there is a significant effect of boundary type on word duration (F[3,116]=12.2, p<.001 for o4; F[3,116]=68.0, p<.001 for n4). Two homogeneous subsets are formed (Newman-Keuls procedure with α=5%): PW and PhP do not differ from each other, nor does IP differ from U. The silent interval after the IP-boundary is also significantly longer than after the shallower boundaries (F[2,87] =23.1, p<.001). The effect of boundary type in the onset of the last syllable is not significant (F[3,116]=1.8, p=.15 for d4; F[3,116]= 2.1, p=.10 for r4). From Figure 1 it is clear that the gap between the boundaries of different depths gets larger as the target phoneme occurs closer to the word edge. This indicates that indeed final lengthening is progressive, thus agreeing with the results of Berkovits (1994). There is also a small tendency for the PhP-boundary to be marked more by final lengthening than the PW-boundary,
and for the U-boundary to be marked more than the IP-boundary, but this tendency does not reach significance. Still, the results agree with the observation by others that there is more final lengthening at deeper boundaries.

**Discussion.** As for the general questions addressing the appropriateness of the present approach for the type of experiments discussed here, one may conclude that the subjects were quite capable of doing the task. Generally, no difference could be detected between phonetically trained and phonetically naive subjects. The subjects had no problem interpreting the structure of the sentences and adding the temporal organisation appropriate for such a structure. This is exactly what is required for the set-up of this experiment.

The effect of final lengthening was restricted to the final syllable, and only significant in the final rhyme. This implies that neither the whole word nor the last stressed syllable are affected by final lengthening. Since in the case of the target word *rododendron* the final syllable is also the final foot, additional material is needed that can differentiate between the final syllable and the final foot as the domain of final lengthening, i.e., target words which have a final bisyllabic foot.

A consistent difference was found between an intonational phrase boundary and a phonological phrase boundary. Other levels are not consistently marked; in fact, the PhP-boundary never differed significantly from the prosodic word boundary, and only one speaker made a significant distinction between the intonational phrase and the utterance. If the following experiment yields similar results, the conclusion may be drawn that only intonational phrase boundaries are obligatorily marked by final lengthening, and the choice for marking finer distinctions in boundary depth is up to the speaker.

Focus did not have a significant effect on the absolute duration of the segments. This is somewhat surprising, since accented words in Dutch are normally lengthened by 10% with respect to unaccented words (Sluijter and van Heuven 1995, and references given therein). One explanation could be that subjects were aware of the syntactic and prosodic contrasts involved in the sentences they had to repeat, which probably urged the speakers to take special care in articulating the utterances with the right structural properties, and to pay less attention to the focus distribution. Another possibility is that the temporal effect of focus is neutralized in postnuclear position. Possibly, the speaking rate after the nuclear accent is lower than before it. Thus, the non-focused target word is perhaps not as short as it would have been if it had been used in prenuclear position. In any case, we cannot answer the question whether lengthening due to prominence interacts with final lengthening, since no lengthening due to prominence was found.
3. Main Experiment

Introduction. The pilot study described in the previous section has shown that a repetition task using diphone synthesis without any temporal effects elicits speech that contains the desired temporal effects. This means that the speakers did not copy the temporal structure of the input speech, but implemented their own temporal structure in their speech. This temporal structure can thus be said to reflect natural temporal effects, which speakers always display in their speech.

Neither focus distribution nor intonation contour had any effect on segment length. Consequently, only one focus condition (+ focus) and one intonation contour will be used in the main experiment. Since not all four boundaries in the pilot study (PW, PhP, IP, U) were consistently distinguished by either a different domain or a different amount of lengthening, there is no reason to add more subtle distinctions in boundary depth to the experiment. Therefore the same four types of boundaries will be used in the main experiment.

The results of the pilot study indicate that the domain of final lengthening is restricted to either the final foot or the final syllable (which coincide in the case of the used target word *rododendron*). In the following experiment, words will be used which differ in the structure of their final two syllables. Furthermore, the final syllable will have varying vowel quality of the nucleus (full vs. schwa) and varying degrees of stress (secondary stress vs. non-stress).

Material. Five words are used in this experiment, given in (4), with capitals indicating main word stress (V=full vowel, C=consonant):

(4) final VV rhyme : harMOnika ('concertina')
    final V + C rhyme : MArathon (id.)
    final schwa rhyme : MOde ('fashion')
    final schwa + C rhyme : TANdem (id.)

These words can be used to answer the research question in (2) concerning the role of syllable structure and/or stress on the domain of final lengthening.

The choice in the pilot study for an equal number of syllables in each utterance together with the other requirements on the (differences between the) utterances implies that the number of syllables in the intonational phrase in which the target word occurs is smaller when the target word is followed by an IP-boundary or an U-boundary than when it is followed by a PW-boundary or a PhP-boundary (cf. (3)). To check whether the length of the intonational phrase rather than the length of the utterance should be kept constant, two versions of the IP- and U-boundary sentences were included in the experiment, one construed similarly to those in the pilot study (i.e., with the same number of syllables in the utterance as in the PW- and PhP-boundaries; henceforth 'non-compensated') and
one in which the utterance is lengthened so as to obtain the same number of syllables in the intonational phrase as in the PW- and PhP-boundary conditions (henceforth ‘compensated’). An example is given in (5):

(5)  
PW  Jan was een goede marathonloper, in zijn jonge jaren.
    ‘Jan was a good marathon runner, in his young years’
IP_{non-c}  Jan liep een goede marathon, lopend in zijn jonge jaren.
    ‘Jan ran a good marathon, running in his young years’
IP_{comp}  Jan Scholten liep een goede marathon, lopend in zijn jonge jaren.
    ‘Jan Scholten ran a good marathon, running in his young years’

The sentences for each word differ minimally in order to form grammatical sentences with the four boundary types immediately following the target word. Each sentence consists of an intonational phrase and a parenthetical, as in the pilot study.

All sentences were synthesized as in the pilot study. Each utterance was realized with a single accent-lending (rise-fall) pitch movement on the primary stressed syllable of the target word. The material for the present experiment thus consists of 30 utterances (5 target words x 6 sentences).

Subjects. Three male and three female subjects (ages 34 to 55) participated in the experiment. Four of these are phonetically trained. Two of the subjects had also taken part in the pilot study. All of the subjects were native speakers of standard Dutch, as judged by a two-member panel of native Dutch phoneticians.

Procedure. All utterances were preceded by a precursor question, which put focus on the target word. Recording procedures and instructions were exactly the same as those in the pilot study.

Results. Both repetitions (of every sentence by each speaker) were segmented. First of all, the two versions (non-compensated and compensated) of the U-boundary and of the IP-boundary were compared. They never differed significantly from one another (F[1,118]<1 for both U and IP). However, the means for the compensated versions are a little shorter than the non-compensated versions, i.e., they come closer to the PW-and PhP-durations. This is to be expected, since a shorter intonational phrase will be spoken more slowly than a long one, so that the non-compensated versions have slightly longer durations. In order not to bias the results towards a large distinction between PW and PhP on the one hand and IP and U on the other, only the compensated versions will be used in the analyses. The results are thus based on 5 words x 4 boundaries x 6 speakers x 2 repetitions = 240 measurements.
One way ANOVA's were done for each target word with segment duration as the dependent variable, boundary type as a fixed factor and with speaker as repeated measure. The durations of each segment per word are given in Figure 2.

In all five words, there is a significant effect of boundary type on the duration of the final rhyme (F[3,236]= 15.6, p<.001 for the vowel; F[3,92]=21.2, p<.001 for the coda consonant). The onset is significantly affected in the words marathon (F[3,44]=4.3, p=.010), tandem (F[3,44]=3.9, p=.015) and yucca (F[3,44]=4.6, p=.007), but not in the words mode (F[3,44]<1) and harmonika (F[3,44]=2.0, p=.129). When the final vowel is full (in marathon, harmonika and yucca), there is no lengthening effect before the final syllable. However, in the word mode, a large effect is found in the penultimate nucleus (F[3,44]=9.6, p<.001). The penultimate nucleus of the word tandem is not significantly lengthened ((F3,44)=2.5, p=.073), but does show a tendency which is not found in the words having a final full vowel. The data for the two final syllables of the target words are presented in a different way in Figure 3; the lengthening of the penultimate and of the final syllable are shown for each target word, with the duration at the shallowest boundary (PW) taken as 'zero duration'. In the diagram showing the lengthening of the penultimate syllable, it can be seen that this syllable is lengthened by some 30 ms in mode, and that tandem lies somewhere in between mode and the other words.

**Figure 2. Segment durations for each boundary type; one diagram for each word**
4. Conclusions

The results of the main experiment confirm the results of the pilot study. Both experiments show that:

- the domain of final lengthening is insensitive to stress position,
- the amount of lengthening is largest in the final segment and decreases in the preceding segments,
- a Dutch PhP-boundary is not marked by significantly more lengthening than a prosodic word boundary,
- the end of an intonational phrase and of an utterance are clearly lengthened, and
- a deeper boundary does not lead to a larger domain of lengthening.

The results of the main experiment also give additional information. Words ending in a long vowel do not behave any differently than words ending in a closed syllable. This implies that the domain of final lengthening is not the final foot, since final VV-words have a final bisyllabic foot, while closed syllables form a mono-syllabic foot. However, words with a final schwa have a larger domain of lengthening than words with a full vowel in the final syllable; the effect in mode reaches back into the penultimate syllable. The word tandem, having a final schwa + coda consonant, seems to lie in between mode and the other words. This leads to the conclusion that when the rhyme of the final syllable contains only a schwa (i.e., with final super-light syllables), final lengthening begins in the penultimate syllable. This effect can also be found, but does not reach significance, in words ending with a schwa and a coda consonant, i.e., in the next lightest type of rhyme.

The effect of syllable weight on the domain of final lengthening may be explained by the fact that in general, light syllables are phonetically shorter than heavy syllables. A final schwa, being shorter than a final long vowel, may not be able to be lengthened to the degree that is required by the boundary depth. Thus,
segments preceding a short final segment will have to participate in the final lengthening, resulting in a larger domain which is lengthened. The domain of final lengthening is then determined by purely phonetic considerations (such as inherent duration of segments and their expandability; Allen et al. 1987, chapter 9). This may also clarify the non-systematic behaviour of the final onset; the variation in the absolute duration of the final rhyme (due to factors such as segment identity, stress, speaking rate, etc.), together with (inherent) differences in expandability, result in a non-systematic effect in the final onset.

References

Hofhuis, E. (in prep.) Segment Duration and Prosodic Structure, PhD dissertation, Katholieke Universiteit Nijmegen.