ON THE NON-LINEAR LENGTHENING OF FOCAL ACCENTED SWEDISH WORDS

Mattias Heldner

1. Introduction

This study deals with the perceptual relevance of a specific non-linear lengthening pattern in focally accented Swedish words, namely that found in stressed syllables where a phonologically short vowel is followed by a long consonant.

Swedish is a quantity language and there is a distinction between phonologically long and short vowels in the stressed syllable. In addition, there is a complementary relation between the vowel and the immediately following consonant – the consonant being short whenever the vowel is long (V:C) and long (VC:) or part of a consonant cluster (e.g. VCC) when the vowel is short (Elert, 1964). In other words, a quantity contrast may also be said to appear within the stressed syllable.

Lengthening is a well-established acoustic correlate of focal accents in Swedish, as well as in many other languages. The amount of lengthening reported in the literature varies, but, typically, words with a focal accent are about 25% longer than the same words when non-focal. However, previous studies of Swedish (e.g. Bannert, 1979; Fant, Kruckenber & Nord, 1991; Heldner & Strangert, forthcoming) have shown that the segments within the focally accented words are sometimes lengthened in a non-linear fashion. That is, some segments are lengthened relatively more than others.

Figure 1 shows typical focal accent lengthening patterns for stressed syllables with a short vowel followed by a long consonant (CVC:) as well as for syllables with a long vowel and a short postvocalic consonant (CV:C) (data from Heldner & Strangert, forthcoming). The non-linear lengthening patterns occurs within the CVC:-syllable. Here, the short vowel remains practically unaffected by the presence of a focal accent, while the postvocalic long consonant is lengthened considerably. This pattern is qualitatively different from that observed in the stressed CV:C:-syllable, where all segments are lengthened, including the short postvocalic consonant.
Figure 1. Examples of lengthening patterns in CVC:- (left panel) and CV:C-syllables (right panel). Segment durations (in ms) within the stressed syllable in a focal accented (filled circles) and non-focal word (empty squares). The empty triangles show a linear lengthening pattern where the amount of lengthening across the syllable is equal to that in the focally accented syllable, but where the lengthening is evenly distributed.

Figure 1 also shows hypothetical linear lengthening patterns for both syllable types. The stressed syllable durations are the same for the focally accented (and non-linearly lengthened) syllables and the linearly lengthened ones, but the lengthening is differently distributed. A comparison between the linear lengthening patterns and the non-linear ones shows that the short vowel in the CVC:-syllable is lengthened relatively less and the long consonant relatively more than the syllable as a whole. As a result, the temporal contrast between the short vowel and the long consonant is strengthened when the word carries a focal accent. In the CV:C-syllable, however, all segments, including the short postvocalic consonant, are lengthened approximately linearly, which means that the temporal contrast is left unaffected.

Thus, the non-linear lengthening of CVC:-syllables has two distinguishing features as compared to the approximately linear lengthening of CV:C-syllables. Firstly, the phonologically short segment in the CVC:-syllable (i.e. the vowel) resists lengthening, while everything else is lengthened. Secondly, there is a strengthening of the temporal contrast between the short and long segments within the CVC:-syllable. It is not altogether obvious why the stressed CVC:-syllable should be lengthened in this non-linear fashion but there seems to be something special about CVC:-syllables and phonologically short vowels.

The first experiment in the present study has been designed to investigate experimentally whether this non-linear lengthening pattern in CVC:-syllables is
important for the perceived naturalness of focal accented words. What it attempts
to establish, in other words, is whether it is important to model focal accent
lengthening in CVC:-syllables in this non-linear fashion, or whether one might
just as well lengthen all the segments linearly, as is done in CV:C-syllables. An
obvious motivation for this question, of course, is that if non-linear lengthening
has an impact on perceived naturalness, CVC:- and CV:C-syllables have to be
treated differently in speech synthesis.

In addition, the relative importance of the two features distinguishing the non-
linear from the linear lengthening pattern, that is, (i) the resistance to lengthening
in the short vowel and (ii) the strengthening of temporal contrast will be examined
in a second experiment.

2. Method

2.1. Collection of production data

As the main question in this paper pertains to the perceived naturalness of non-
linear lengthening patterns, the principal part of the investigation will be the
perceptual experiments. However, prior to this, some production data were
collected to serve as a model for the stimuli to be used in the perceptual
experiments. To this end, six three-word sentences were constructed, in which the
middle words (the test words) were disyllabic verbs with stress on the first
syllable. These sentences are shown in Table 1. All the test words contained a
stressed syllable with a phonologically short vowel followed by a long consonant.

<table>
<thead>
<tr>
<th>Sentence</th>
<th>Transcription</th>
<th>Translation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mannen dricker vinet.</td>
<td>/drikːər/</td>
<td>‘The man is drinking wine.’</td>
</tr>
<tr>
<td>Mannen klipper persilja.</td>
<td>/klipːər/</td>
<td>‘The man is cutting parsley.’</td>
</tr>
<tr>
<td>Mannen knäcker äggen.</td>
<td>/kneːkːər/</td>
<td>‘The man is breaking eggs.’</td>
</tr>
<tr>
<td>Mannen grillar köttet.</td>
<td>/grilːər/</td>
<td>‘The man is grilling meat.’</td>
</tr>
<tr>
<td>Mannen gräddar brödet.</td>
<td>/greːdːər/</td>
<td>‘The man is baking bread.’</td>
</tr>
<tr>
<td>Mannen pressar citronen.</td>
<td>/preːsːər/</td>
<td>‘The man is squeezing lemon.’</td>
</tr>
</tbody>
</table>

There were two versions of each sentence, one of which had a focal accent on the
middle word and the other on the final word. The test words thus occurred in focal
as well as in non-focal (i.e. pre-focal) position. One speaker (the author) read each
of these sentences six times. Subsequently, the durations of all segments in all
sentences were measured, and the amount of focal accent lengthening for each segment of the test word, as well as for the whole test word, was determined. In addition, a set of f₀ turning points associated with word accents, focal accents and boundary tones were measured. The measured segment durations are shown in Figure 2, while the corresponding word durations and amounts of lengthening are shown in Table 2. Figure 2 also presents a linear lengthening relative to the segment durations in non-focal position using the amounts of focal accent lengthening recorded in Table 2.

Table 2. Mean word durations (in ms) for the test words in focal and non-focal position and the amount of focal accent lengthening.

<table>
<thead>
<tr>
<th></th>
<th>dricker</th>
<th>klipper</th>
<th>knäcker</th>
<th>grillar</th>
<th>gräddar</th>
<th>pressar</th>
</tr>
</thead>
<tbody>
<tr>
<td>focal</td>
<td>516</td>
<td>514</td>
<td>571</td>
<td>524</td>
<td>523</td>
<td>506</td>
</tr>
<tr>
<td>non-focal</td>
<td>340</td>
<td>325</td>
<td>400</td>
<td>383</td>
<td>360</td>
<td>360</td>
</tr>
<tr>
<td>% lengthening</td>
<td>52%</td>
<td>58%</td>
<td>43%</td>
<td>37%</td>
<td>45%</td>
<td>41%</td>
</tr>
</tbody>
</table>

2.2. Preparation of stimuli

As mentioned above, the production data was collected only to serve as a model for stimuli in perceptual experiments. Thus, the stimuli in the two experiments in this study were different synthetic versions of the sentences in Table 1. All of these sentences were spoken by a male mbrola voice and prepared using the software WaveSurfer (Sjölander & Beskow, 2000) with a text-to-speech plug-in.

The stimuli in the two experiments were similar in many respects. The general aspects of the stimuli will be described here, whereas the particularities will be dealt with in the following sections. In general, the segmental durations and f₀-contours in all sentences were modeled after sentences with focal accents on words in medial position in the sentence. Moreover, the experimental variation occurred in the sentence-medial words only. Everything was kept constant in the initial and final words.

Furthermore, all experimental variations concerned segmental durations. However, as a consequence of these manipulations the slopes of the f₀-movements within the test words were also affected. This was because the values of the f₀-turning points were kept constant while these points were anchored relative to the segments (e.g. to the boundary between the stressed vowel and the postvocalic consonant). Therefore, lengthening of a segment also resulted in a less steep slope of f₀-movements in that segment.
Figure 2. Mean segment durations (in ms) for each of the test words in focal and non-focal position and a linear lengthening relative to the segment durations in non-focal position. C1=first consonant, C2=second consonant, V1=first vowel etc.
2.3. The first perceptual experiment

The first perceptual experiment was designed to investigate whether a non-linear lengthening pattern in CVC:-syllables is important for the perceived naturalness of focal accented words. Accordingly, listeners were asked in this experiment to compare pairs of synthesized sentences, where the amount of lengthening of the focally accented word was the same, while the distribution of this lengthening differed. One test word in each pair reflected the non-linear lengthening patterns in focally accented words (as in the filled circles of Figure 2). The other test word in the pair contained a linear lengthening of all segments relative to the segment durations for the non-focal words, that is, all segments within the word were lengthened by the same percentage (as in the empty triangles of Figure 2). Six pairs of sentences were prepared: one pair for each sentence in Table 1.

Eleven native speakers of Swedish participated in the first experiment. They were given the forced choice task to judge whether the first or second sentence in each pair sounded more natural. The order of presentation within each pair, that is whether the linear or non-linear version occurred first, was random. The presentation order of the sentence pairs was also random. Each listener judged each pair ten times. Thus, 110 judgements were obtained of each pair.

2.4. The second perceptual experiment

The second perceptual experiment was conducted to compare the perceptual importance of the two features distinguishing the non-linear from the linear lengthening patterns. Thus, the purpose was to find out whether one of the features (i) strengthening of temporal contrast and (ii) short vowel remaining short was more important than the other.

Again, listeners were asked to compare the naturalness of pairs of synthesized sentences where the segmental durations in the sentence-medial word differed. However, in this experiment both sentences in each pair deviated from the patterns observed in production. Thus, the listeners had to judge which deviation affected the naturalness least.

The experimental variations were restricted to the stressed vowel and the immediately following consonant and started out from the non-linear lengthening patterns (as shown in Figure 2). In one of the sentences in each pair, the vowel remained short while the consonant was shortened to the same duration as in the linear pattern (see Figure 2). This means that the temporal contrast was weakened compared to what was the case in the non-linear lengthening pattern, while the vowel stayed short. In the other sentence in the pair, both the vowel and the consonant were lengthened in such a way as not to affect the temporal contrast.
compared to the contrast displayed by the non-linear version. Here, the vowel had the same duration as in the linear pattern (see Figure 2) while the consonant was longer than in the non-linear reference. These lengthening patterns are illustrated in Figure 3.

![Figure 3](image.png)

Figure 3. An illustration of the lengthening patterns weakened contrast (empty circles) and lengthened vowel (empty squares) together with the non-linear lengthening reference (filled circles) in the second perceptual experiment.

Contrary to the manipulations carried out in the first experiment, those in the second experiment affected the total duration of the stressed syllable in the test words and consequently also the rhythm of the sentence.

Ten native speakers of Swedish participated in the second experiment. Their task and the number of presentations of each sentence pair was the same as in the first experiment. Every sentence pair was judged 100 times (10 times per speaker).

3. Results of the first perceptual experiment

Turning now to the results of the first perception experiment, Table 3 shows the observed frequencies for linearly and non-linearly lengthened stimuli judged to be more natural than the other sentence in a pair. It is quite obvious that stimuli with non-linear lengthening were judged to be more natural. The Chi-square tests across all speakers are significant for all test words except grillar: $\chi^2(1, N=110)=44.54, p<.01$; klipper: $\chi^2(1, N=110)=44.54, p<.01$; knäcker: $\chi^2(1, N=110)=37.24, p<.01$; grillar: $\chi^2(1, N=110)=2.33, p=.13$; gräddar: $\chi^2(1, N=110)=47.13, p<.001$; pressar: $\chi^2(1, N=110)=26.51, p<.001$. Moreover, the non-significant Chi-square test for grillar is probably due to the small durational differences between the linear and non-linear versions (cf. Figure 2).
Table 3. Frequencies across all listeners of linearly and non-linearly lengthened stimuli judged to be more natural than the other sentence in a pair.

<table>
<thead>
<tr>
<th></th>
<th>dricker</th>
<th>kclipper</th>
<th>knäcker</th>
<th>grillar</th>
<th>gräddar</th>
<th>pressar</th>
<th>Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear</td>
<td>20</td>
<td>20</td>
<td>23</td>
<td>47</td>
<td>19</td>
<td>28</td>
<td>157</td>
</tr>
<tr>
<td>Non-linear</td>
<td>90</td>
<td>90</td>
<td>87</td>
<td>63</td>
<td>91</td>
<td>82</td>
<td>503</td>
</tr>
<tr>
<td>Totals</td>
<td>110</td>
<td>110</td>
<td>110</td>
<td>110</td>
<td>110</td>
<td>110</td>
<td>660</td>
</tr>
</tbody>
</table>

Although a majority of the listeners were sensitive to the durational differences they were exposed to during the experiment, two of the listeners reported that they did not notice any difference between the two versions in each sentence pair. An examination of their judgments showed no preference for either non-linear or linear lengthening. In other words, not all listeners were sensitive to the durational variations.

4. Results of the second perceptual experiment

The results of the second perception experiment are shown in Table 4. The table shows observed frequencies for stimuli judged to be more natural than the other sentence in a pair. Clearly, words with a lengthened vowel in combination with the maintaining of temporal contrast between the short vowel and the long consonant were judged to be less natural than words where the vowel remained short while the temporal contrast was weakened.

Table 4. Frequencies across all listeners of stimuli with lengthened vowels and weakened contrasts, respectively, judged to be more natural than the other sentence in a pair.

<table>
<thead>
<tr>
<th></th>
<th>dricker</th>
<th>kclipper</th>
<th>knäcker</th>
<th>grillar</th>
<th>gräddar</th>
<th>pressar</th>
<th>Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lengthened vowels</td>
<td>6</td>
<td>16</td>
<td>5</td>
<td>33</td>
<td>9</td>
<td>12</td>
<td>81</td>
</tr>
<tr>
<td>Weakened contrasts</td>
<td>94</td>
<td>84</td>
<td>95</td>
<td>67</td>
<td>91</td>
<td>88</td>
<td>519</td>
</tr>
<tr>
<td>Totals</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>600</td>
</tr>
</tbody>
</table>

The preference for the weakened contrast (or short vowel remaining short) stimuli was statistically significant in all words as the Chi-square tests across all speakers were significant for all test words: [dricker: $\chi^2(1, N=100)=77.44, p<.01$; kclipper: $\chi^2(1, N=100)=46.24, p<.01$; knäcker: $\chi^2(1, N=100)=81.00, p<.01$; grillar: $\chi^2(1,$
Moreover, a comparison of the results of the two experiments shows that the listeners were more certain of their judgments in the second experiment. The listeners preferred the non-linear lengthening pattern in 76% of the judgments in the first experiment while the stimuli where the short vowel remained short (weakened contrast) in the second experiment were preferred in 87% of the judgments (cf. Tables 3 and 4).

5. Discussion

The first perceptual experiment in this study showed that a majority of the listeners were sensitive to the durational differences they were exposed to during the experiment. These listeners preferred a non-linear lengthening of focally accented CVC:-syllables to a linear expansion. In other words, the listeners preferred the kind of lengthening patterns previously observed in production studies (e.g. Bannert, 1979; Fant et al., 1991; Heldner & Strangert, forthcoming). Taken together, these findings show that a linear time expansion is neither an appropriate description of focal accent lengthening in Swedish from a production perspective, nor from a perception one. Moreover, Swedish seems to be different in this respect from other languages. For example, as shown by Sluijter (1995), the temporal contribution of an accent in Dutch and English is an almost linear time expansion of the entire word.

Previous production studies have shown that the non-linear lengthening of CVC:-syllables has two distinguishing features as compared to the lengthening of CV:C-syllables: firstly, the phonologically short segment (i.e. the vowel) restricts lengthening, and, secondly, there is a strengthening of the temporal contrast between short and long segments within the syllable (e.g. Bannert, 1979; Heldner & Strangert, forthcoming).

The second perceptual experiment has shown that it is more important for the perceived naturalness that the short vowel remains short than that the temporal contrast be strengthened. This is an indication that the linearly lengthened stimuli in the first experiment were judged to be less natural than non-linearly primarily because the short vowel was lengthened. Perhaps more importantly, it indicates the existence of an expandability constraint at the phonemic level analogous to the compressibility constraint proposed by Klatt (e.g. Klatt, 1976). The precise explanation of an expandability constraint in short vowels in Swedish remains to be discovered, but it is certainly not physiologically determined, as compressibility constraints will often be. A phonological explanation in terms of
maintaining sufficient dissimilarity between long and short vowels (or CVC:- and CV:C-syllables) when lengthened (Bannert, 1979) would seem to be more plausible.

Moreover, the second perceptual experiment indicates that the strengthening of temporal contrast observed in production data is a consequence of the expandability constraint in the short vowel. If the vowel remains short while the syllable is lengthened, the following consonant has to be lengthened more than the syllable as a whole.

To summarize, perceived naturalness increases as a result of modeling focal accent lengthening of a CVC:-syllable with a non-linear pattern as compared to using a linear lengthening pattern. Moreover, the most important feature of this non-linear pattern is for the vowel to be maintained short. For text-to-speech applications, this clearly implies that the lengthening of CVC:-syllables should be modeled differently from that of CV:C-syllables.

Acknowledgements
The author would like to thank Jonas Beskow for technical assistance.

References


