Distal rate effect for Finnish epenthetic vowels

Michael O’Dell¹, Tommi Nieminen²

¹University of Tampere
²University of Eastern Finland

michael.odell@uta.fi, tommi.nieminen@uef.fi

Index Terms: speech rate, syllable perception, epenthesis, distal prosody, Finnish

1. Introduction

1.1. Speech rate and syllable perception

In recent years Dilley and colleagues [1, 2, 3] have conducted a series of experiments involving the effect of context tempo on the perception of syllables. They found that “entire words can seem to disappear or appear as a function of their context speech rate” [1]. Using appropriate pairs of sentences, e.g. Mom said these are (our) gray gloves, slowing the context made the extra one syllable word (our in this example) disappear from perception, while speeding up the context increased its probability of perception.

These results have been interpreted as indicating that syllable perception is partially dependent on the rhythm of a wider speech context. A vowel duration long enough to be heard as a syllable in a faster context may not be sufficient to trigger syllable perception in a slower context.

So far this distal speech rate effect has been observed in English, Russian and Mandarin Chinese. We wanted to see whether a corresponding effect could be demonstrated in Finnish as well. Because Finnish is a full fledged quantity language, it follows that entire words can seem to disappear or appear. The main factors affecting word perception in a slower context are therefore the rhythm of the wider speech context. A vowel duration long enough to be heard as a syllable in a faster context may not be sufficient to trigger syllable perception in a slower context.

1.2. Epenthetic vowels in Finnish

In Finnic studies, an epenthetic vowel (called švaa or svarabhakti) occurring within certain heterorganic consonant clusters has been reported for most Finnish dialects. Crucial contexts almost always described as attesting this intrusive or transitional vowel in the traditional “epenthetic dialects” are [l]—initial two-or three-consonant clusters {[lp],[lk],[lv],[lvt],[lh],[lm]}. The quality of the epenthetic vowel is generally described as identical to the vowel preceding the cluster (although in a few dialects it is held to be intermediate between preceding and following vowel), except in the case of [lj], where the epenthetic vowel is invariably [l]. Other possibilities include [r]—initial, [h]—initial, and [h]—final clusters, but these clusters are described as containing an epenthetic vowel only in a few individual studies, often ones conducted with experimental methods.

Traditionally dialectologists in Finland have reported only three separate variants for these consonant clusters, one for each dialect. Either a given dialect has no epenthetic vowel (e.g. [help.po] helppo ’easy’), or it has a very short epenthetic nonsyllabic vowel ( [hel.po] ), or it has a full syllabic vowel ([hel.po]). In other words, existence of an epenthetic vowel has only been acknowledged if there is a vocalic segment where the standard (written) language has a consonant cluster, and only if that vowel is almost long enough to be counted as syllabic.

However, perceptually as well as articulatorily, the phenomenon is quite variable and forms a continuum. As already evidenced in research the acoustic epenthetic vowel is much more common in Finnish than one might surmise based on existing dialect descriptions alone. In an earlier study [4] we found an acoustic epenthetic vowel in 83 percent of spontaneous productions of the word pilvi (‘cloud’). The words were taken from the Variation of Finnish Prosody Database (Prosovar) [5], and the word pilvi was chosen for study only because it was represented in the database quite often due to a particular picture description task used to elicit speech.

In other languages similar epenthetic vowels have been interpreted as arising from the timing of gestures in a consonant cluster [6, 7, 8], a view compatible with Articulatory Phonology [9]. An example of the relevant timing differences from this point of view is shown schematically in Figure 1.

A short acoustic vowel segment in speech might thus be interpreted either as a separate syllable (kuluma ‘abrasion’) or as a normal transition inside a consonant cluster (kulma ‘angle’), at least when both interpretations correspond to actually occurring words. This means there may be more factors in the perception of the number of syllables in a word than just whether or not there is an actual acoustic vowel between consonants.

1.3. Research questions

1. Can a syllable produced by a speaker be interpreted by the listener as a (nonsyllabic) transition between consonants in a cluster in Finnish? For instance, can the word kuluma be heard as the word kulma?

2. Does contextual tempo or rhythm have an effect on the perception of such a syllable?

3. Can a distal speech rate effect be demonstrated in Finnish, or is it less noticeable for instance because of stronger effects of local rhythm (proximal prosody), given that Finnish is a quantity language?
Table 1: Original test sentences, which contain test words with “extra” vowel (meaning of words without vowel in parentheses)

<table>
<thead>
<tr>
<th>Sentence</th>
<th>Original Test</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nokia’s longtime spending discipline (tramp) doesn’t cause concern. Nokian pitkääikainen kulukari ei aiheuta haolia.</td>
<td>kulukari</td>
<td>vaunujen suurin sallittu kuluma on luultavasti liian pieni. The maximum abrasion (angle) allowed for cars is probably too small.</td>
</tr>
<tr>
<td>In some situations salami (a strait) is quite dangerous. Joissakin tilanteissa salamikuluma on aika vaarallinen.</td>
<td>salami</td>
<td>nokian pitkäaikainen kulukuri ei aiheuta huolia. on luultavasti liian pieni. The maximum abrasion (angle) allowed for cars is probably too small.</td>
</tr>
<tr>
<td>The tires have defects caused by the casting (ready) process. Renkaissa on valamisprosessin aiheuttamia vikoja.</td>
<td>salami</td>
<td>nokian pitkäaikainen kulukuri ei aiheuta huolia. on luultavasti liian pieni. The maximum abrasion (angle) allowed for cars is probably too small.</td>
</tr>
<tr>
<td>For example some tallow (ap pulley) can be found in the hall closet. Esimerkiksi talia löytyy etisen kaapista.</td>
<td>tallow</td>
<td>nokian pitkäaikainen kulukuri ei aiheuta huolia. on luultavasti liian pieni. The maximum abrasion (angle) allowed for cars is probably too small.</td>
</tr>
<tr>
<td>He knew that only a rake (a few) could perform such a task. Hän tiesi, että vain harava pystyy poistamaan suoritukseen.</td>
<td>rake</td>
<td>nokian pitkäaikainen kulukuri ei aiheuta huolia. on luultavasti liian pieni. The maximum abrasion (angle) allowed for cars is probably too small.</td>
</tr>
</tbody>
</table>

2. Methods

2.1. Stimuli

A group of 6 native speakers of standard Finnish was first recorded producing 12 sentences, 6 test sentences each containing a test word of interest (see Table 1) as well as 6 distraction sentences. Sentences were presented to the speakers visually in random order and speakers were instructed to produce the sentence from memory after a small pause. Test sentences and distraction sentences were not distinguished in any way, and test words within test sentences were not marked. One instance of each sentence was then selected for use in the perception experiment, each from a different speaker, so that duration of the “extra” syllable was as constant as possible.

Each of the selected sentences was manipulated in Praat [10] to form two versions: In the Base-version (context tempo same as test word) the entire sentence was manipulated evenly so the resulting duration of the second syllable of the test word was 90 ms (see Figure 2). In the Slow-version (context tempo slower than test word) the entire test word was the same as in the Base-version, but the rest of the sentence was manipulated so that it was 1.6 times longer in duration (see example in Figure 3). Similar manipulations were carried out for the distraction sentences, using an arbitrarily chosen centrally located word as the “test word”.

2.2. Perception test

In the perception test 84 subjects (69 women and 15 men aged 19–64 years) listened to the sentences in 21 different random orders (but always with test sentences and distraction sentences alternating). For each sentence subjects were presented visually with a written version of the sentence with the test word replaced with dots. In each presentation of the test each sentence was presented once. Half of the sentences were Base-versions and half were Slow-versions, so each listener heard only one version of each sentence. The listener’s task was to write the missing words on an answer form based on what was heard (no alternatives were provided).

2.3. Statistical analysis

The empirical response data was analyzed using a hierarchical logistic regression model (cf. equations 1–3 below) in which probability of perceiving a syllable (\( \pi_{ijk} \)) may depend on RATE (i.e. Base-context vs. Slow-context), WORD (\( n = 6 \) different word/sentence pairs) and SUBJ (\( n = 84 \) subjects). Because not all subjects were assigned to different random orders (test runs), it was also possible to obtain a separate estimate of variation due to test run (RUN, \( n = 21 \)). Also included in the model were the interactions WORD \( \times \) RATE and SUBJ \( \times \) RATE (cf. equation 1). Other interactions could not be estimated because of the experimental setup (each subject heard only one version of each sentence).

---

Note that orthographic talia corresponds to [talija], ie. talja with an epenthetic [i].
Coefficients for all effects are normally distributed on the logit scale with mean zero (equation 2; \(X = \text{RATE}, \text{WORD}, \text{SUBJ}, \text{RUN}, \text{WORD} \times \text{RATE}, \text{WORD} \times \text{SUBJ}\)) and standard deviation parameters with half Cauchy prior distributions (equation 3, see [11]) with unit scale parameter. Effects in the model were evaluated using Bayesian inference [11], and posterior distributions for model parameters were calculated using JAGS [12] in conjunction with R [13].

\[
\logit(\pi_{ijk}) = \mu + \beta^\text{RATE}_{k} + \beta^\text{WORD}_{i} + \beta^\text{SUBJ}_{j} + \beta^\text{RUN}_{h[j]} + \beta^\text{WORD} \times \text{RATE}_{ik} + \beta^\text{SUBJ} \times \text{RATE}_{jk}
\]

(1)

\[
\beta^X \sim \text{Normal}(0,\sigma^X)
\]

(2)

\[
\sigma^X \sim \text{Half-Cauchy}(1)
\]

(3)

3. Results

Figure 4 shows the estimated standard deviations of coefficients for effects in the model. Probability of syllable perception was clearly affected by \(\text{WORD}\) (the standard deviation is clearly positive). This result is expected, since the different sentences and word pairs are expected to differ in probability even before the stimuli are heard.

Traditional studies of Finnish dialects have claimed that dialects in some areas have an epenthetic vowel whereas others do not. However, although we collected background information as to where our subjects lived in childhood and as adults, no clear correlation was found with subjects’ propensity to interpret the acoustic second vowel as a syllable. Test \(\text{RUN}\) probably also had an effect, although its size remains unclear.

The most important effect for the present research was the \(\text{RATE}\) of the stimulus carrier sentence. As can be seen in Figure 4, \(\text{RATE}\) had a very clear effect on perception, possibly the largest effect in the model. Interactions \(\text{WORD} \times \text{RATE}\) and \(\text{SUBJ} \times \text{RATE}\) had relatively small effect, which makes it easier to interpret the results.

Figure 5 shows the estimated probabilities of perceiving an extra syllable for both \(\text{RATE}\)s (\(\text{Base}\) vs. \(\text{Slow}\)), overall (with \(\text{WORD}\) effect removed, below) as well as for each word pair separately (above). It is clear that the observed strong effect of \(\text{RATE}\) on perception means that syllable perception is more likely in the \(\text{Base}\) condition, that is, when the carrier sentence tempo is the same as the test word tempo.
In the Slow condition, when the carrier sentence has been stretched, syllable perception is less likely. Posterior probabilities \( \Pr(\pi_{\text{Slow}} < \pi_{\text{Base}}) \) that syllable perception is less likely in the Slow condition than in the Base condition are as follows: kulukari: 0.970, kuluma: 0.992, salami: > 0.999, valamis: > 0.999, talia: 0.991, harava: 0.905. The overall posterior probability is \( \Pr(\pi_{\text{Slow}} < \pi_{\text{Base}}) > 0.999 \).

4. Discussion

Perhaps the simplest interpretation of the results is that by slowing the context by a factor of 1.6 we have correspondingly shifted any durationally based perception of the test word by the same amount. For instance, probability of syllable perception in the Base-context as a function of the duration of the acoustic epenthetic vowel might look something like the solid curve in Figure 6. Presumably this function is roughly the ratio of the probability density curves for the two categories weighted by prior expectations of the two words in the given sentence. If we assume this probability depends solely on durations in the context, then slowing the context by 1.6 would be expected to multiply the time axis by 1.6, producing the dashed curve in Figure 6. For any given vowel duration, such as that indicated by the arrow in the figure, a syllable perception would be less likely in the Slow-context (open circle) compared to the Base-context (filled circle).

![Figure 6: Schematic example of perceptual shift from Base-context to Slow-context](image)

Of course this account is too simplistic, because the timing of the word itself must also have an impact, which would be expected to move the Slow-curve in Figure 6 closer to the Base-curve. It does, however, illustrate why syllable perception should be less likely in the Slow-context, and this effect might well be stronger the more regular the rhythm of the carrier sentence (cf. [14]).

Why weren’t segments other than the second vowel of the test word affected? For instance, in the token of salami used as test word the duration of the first [a] was slightly less than the second [a] (see Figure 2), and yet only the second [a] was sometimes not perceived as a separate syllable. Perhaps both vowels were affected, but the a priori probabilities of possible Finnish words dominated. In Finnish both salmi (‘strait’) and salami (‘salami’) are actual words, and both represent normal phonotactic patterns, which is not the case e.g. with slami.

It is also interesting to note that the phonological opposition itself between a consonant cluster allowing a transitional epenthetic vowel and a separate syllable is not very reliable. In spite of the fact that all test words were originally produced as words with an “extra” syllable, they were quite often perceived as words without that syllable even in the Base-context. Syllable perception was most consistent in the case of salami in the Base-context and consonant cluster perception was most consistent in the case of harava (‘rake’) perceived as harva (‘a few’) in the Slow-context. Overall syllable perception amounted to about 64% in the Base-context and about 28% in the Slow-context.

In the future we plan to investigate the pronunciation in Finnish of consonant clusters and epenthetic vowels from an articulatory point of view. It will be interesting to see what features beside the duration of acoustic vowel ultimately affect perception. Especially interesting would be to have informants from both traditional epenthetic dialects (e.g. the Eastern dialects) as well as other dialects (e.g. the Southwestern dialects).

In the present study stimuli were formed from original words containing a whole syllable instead of a consonant cluster (e.g. kuluma and salami, etc.), instead of kulma and salmi, paired with expanded contexts in addition to the base versions. The research by Dilley et al. [1, 2, 3] utilized both original recordings containing an extra syllable and the corresponding recordings without the extra syllable, as well as contexts which were both expanded and contracted in addition to the base versions. In the future we would also like to extend this study analogously to include test words formed from recordings of words with (phonological) consonant clusters (kulma ['kulma ~ kulimus'], salmi ['salmi ~ salmi'], etc.), with contracted contexts as well as expanded contexts to see whether an original consonant cluster could be heard as an extra syllable, and whether a faster context would increase the probability of perceiving a syllable.

5. Conclusions

The main result of the present study was that it is indeed possible for a syllable to sound like a consonant cluster in Finnish under specific circumstances. In addition, we found that contextual tempo has a clear effect on this perception. We found clear evidence of a distal speech rate effect, and its direction follows the predictions of Dilley et al. [1, 2, 3], although the magnitude of the effect is slightly smaller than in their findings. This result was not fully expected as many features in this research differed from the previous studies (cf. Table 2).

<table>
<thead>
<tr>
<th>Dimension</th>
<th>this study</th>
<th>Dilley et al.</th>
</tr>
</thead>
<tbody>
<tr>
<td>language</td>
<td>Finnish</td>
<td>English, Russian, Mandarin</td>
</tr>
<tr>
<td>variation source</td>
<td>epenthesis</td>
<td>reduction</td>
</tr>
<tr>
<td>test word</td>
<td>3–4 syllables</td>
<td>one syllable</td>
</tr>
<tr>
<td>sentence position</td>
<td>stressed</td>
<td>unstressed</td>
</tr>
<tr>
<td>tempo difference</td>
<td>1.6</td>
<td>1.9</td>
</tr>
</tbody>
</table>

Previous observations have been made for non-quantity languages and for one syllable function words in unstressed positions. We have extended the range of cases in which a distal speech rate effect can be observed to include Finnish, a quantity language. It also appears that the effect of context tempo persists to some extent even to the middle of multisyllable content words in stressed position within the sentence.
6. References


