

Local speaking rate and perceived quantity

Diana Krull¹, Hartmut Traunmüller¹ and Wim A. van Dommelen²

¹*Department of Linguistics, Stockholm University;* ²*Department of Language and Communication Studies, NTNU, Trondheim*

In an earlier experiment, we have shown that the local speaking rate affects the perception of quantity of Estonian listeners. In order to see if this effect is language dependent, we presented the same stimuli to Finns and a subset to Norwegians, whose languages have a different and smaller functional load of quantity distinctions. The results obtained with Estonian and Finnish listeners are compatible with a model of speech perception in which variations in speaking rate are reflected in the pace of an “inner clock” by which listeners measure segment durations. More ‘absolute’ and narrow scoped results with Norwegians are compatible with such a model only if the inner clocks of listeners with a less demanding linguistic background are assumed to resist such an influence to a higher degree.

1. Introduction

The present investigation addresses the question of how quantity is communicated with respect to global or local variations in speaking rate. If the perception of quantity were based on the absolute duration of segments, it would be difficult to maintain the opposition at different speaking rates. If, instead, quantity perception were based on durational ratios between segments (or larger units), there would be no such problem. However, this would not apply to the same extent to all languages. Speakers of Finnish and Estonian, for example, in whose languages quantity carries a heavy functional load, have been shown to maintain the relevant durational ratios more consistently than speakers of Swedish (Engstrand & Krull, 1994). A model of speech perception in which variations in speaking rate are handled by an “inner clock” whose pace is determined by the speaking rate of the speech listened to (Traunmüller, 1994) appears to be best suited to describe the behavior of listeners. However, also here there may be differences between languages in the way listeners estimate the local speaking rate. In Estonian, the duration of the vowel in a following syllable is known to affect the perceived quantity of a preceding V or C strongly (Lehiste, 1970–75), and there is a similar effect in Finnish (Lehtonen, 1970). An effect of the same kind, albeit smaller, has recently been observed in Norwegian (Dommelen, 1999).

In an earlier experiment, the effect of the speech rate of a short preceding or following context was studied in addition to the duration of a V, C or VC-sequence that carried a quantity distinction in Estonian words (Krull & Traunmüller, 2000; Traunmüller & Krull, *subm.*) In order to see if the observed perceptual effects are language dependent, the same stimuli were presented to Finns and a subset to Norwegians. Finnish is closely related to Estonian but there are differences in the quantity systems. While Estonian distinguishes between three degrees of quantity in both vowels and consonants of accented syllables,

Finnish distinguishes only two degrees but does so even in unaccented syllables. Norwegian belongs to a different language family. It has two distinctive degrees of quantity for stressed vowels, while the duration of a following consonant is inversely related to that of the vowel

2. Method

The stimuli were obtained by manipulating the durations of selected sections of recordings of the Estonian word *saate* [sa:t:e] 'get' (2nd p. pl) produced by a female speaker, preceded by *ja* [ja] 'and', in isolation, and followed by *ka* [ka] 'also', with list reading intonation. (The segment durations of the utterances are listed in Table 1.) This resulted in different original durations of the segments [s], [a], [t] and [e] in the three different contexts. All the response alternatives, the Estonian ([sate], [sa:te], [sa::te], [sat:e], [sat::e], [sa:t:e], [sa::t:e]), 4 Finnish ([sate], [sa:te], [sat:e], [sa:t:e]), and 2 Norwegian ([sate], [sa:te]) are common or possible words in the respective language.

The durations of the [a:] and [t:] in [sa:t:e] (occlusion + burst) were modified in steps of a factor $2^{n/8}$ with $-8 < n < 8$, allowing for a deviation within \pm _ pitch period for phase-clean joints within the [a:]. Figure 1 illustrates the arrangement of stimuli in series.

Table 1. Segment durations in ms (upper half) and in % of the duration of [a:t:] in the original utterance (lower half).

[ja]	[s]	[a:]	[t:]	[e]	[ka]	[a:t:]
*	96	137	152	101	378	289
*	121	175	188	201	*	363
160	105	176	171	198	*	347
*	33.2	47.4	52.6	34.9	130.8	100
*	33.3	48.2	51.8	55.4	*	100
46.1	30.3	50.7	49.3	57.1	*	100

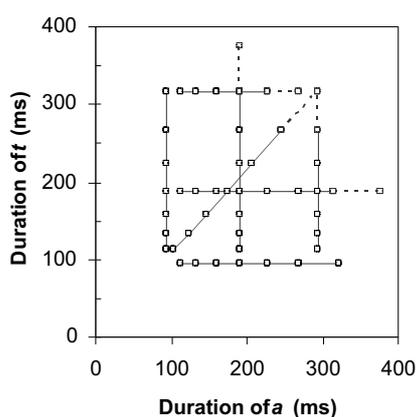


Figure 1. Stimuli, represented by circles, arranged in 7 series. Duration values represent word spoken in isolation. Longest durations (dashed lines) only used when final [e] or [e ka] prolonged. Vertical series: [a] constant, [t] lengthened step by step, horizontal: [t] constant, [a] varied, diagonal: [a] and [t] varied together.

Norwegian listeners were presented three series with a constant duration of [t] (133, 188 and 266 ms) and one with a constant duration of [a] (191 ms). The durations of those parts of the utterances that preceded or followed the [at]-sequence were varied as follows: 84% ($2^{-2/8}$) of the original, unchanged, and 119% ($2^{+2/8}$) of the original.

There were 25 Estonian, 23 Finnish and 25 Norwegian subjects. Finnish and Estonian subjects listened to the same 1013 test stimuli arranged in 161 series, Norwegian subjects listened to 484 test stimuli arranged in 92 series. Each stimulus was presented once in a series with successively increasing duration of the segment in focus.

3. Results and discussion

Percent responses was calculated for [V:] (all three languages), [V::] (Estonian), [C:] (Finnish and Estonian) and [C::] (Estonian). Figures 2a–2d show these responses for the three listener groups as a function of the durational change of the first vowel, [a], with that of the second, [e], as a parameter. Analogous figures drawn for consonant quantity as perceived by Estonian and Finnish listeners look similar but are not shown here. However, Figure 2e shows the vowel quantity results of the Norwegian listeners with [t] duration as a parameter.

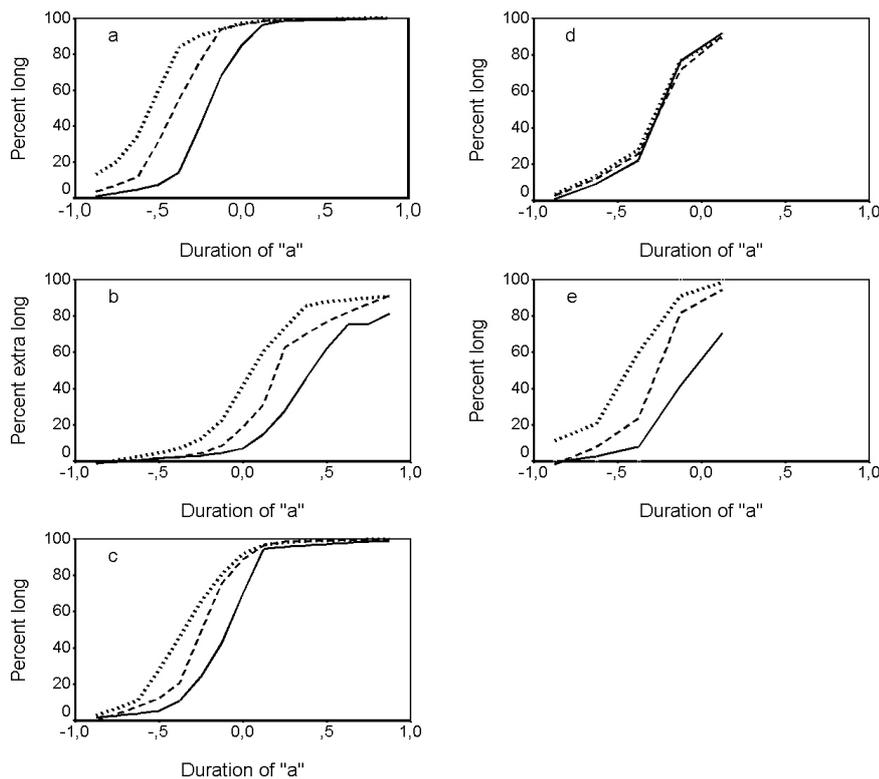


Figure 2a–2e. Percent [long] or [extra long] responses for [a] in [(ja)sate(ka)] as perceived by listeners of different linguistic provenience: Estonian long (a) and extra long (b), Finnish (c) and Norwegian (d, e). X-axis: power of 2 by which duration of [a] was modified (-1 = 50%, +1 = 200% of original duration). Dotted lines: shortened [e] (84%); broken lines unchanged [e]; unbroken lines: lengthened [e] (119%). In Figure 2e duration of [t] was modified instead.

Table 2. Sum of effects of [(ja)s] and [e(ka)], for Norwegian also of [t], as a percentage of effect of segment in focus. Within parentheses: Norwegian effect without that of [t].

	[a:]	[t:]	[a::]	[t::]
Estonian	-70	-52	-80	-81
Finnish	-61	-46		
Norwegian	-42 (-9)			

The results show that the duration of the [e] in the second syllable had a large effect on the identification of [a] of the first syllable by Estonian listeners. (A similar effect was found on the identification of [t]). For Finnish listeners, the effect of [e] was slightly smaller, and very small for Norwegians. Also the combined effect of [(ja)s] and [e(ka)] was much smaller in Norwegian (see Table 2), which can be interpreted as a narrower reference frame or a shorter window of analysis. In Norwegian, the [t] that follows the vowel appears to carry the function that the [e] in the following syllable has in Finnish and Estonian. For Estonian and Finnish listeners, the [t] could hardly take this role since it is itself subject to a quantity opposition. For Finnish, a similar argument could be raised concerning the vowel of the second syllable, although variation in quantity is less common in this position. On the basis of the present experiments, we cannot exclude the possibility that Finnish listeners use a window of analysis that is still wider than that of Estonians.

The results obtained with Estonian and, less clearly, Finnish listeners are compatible with a model of speech perception in which variations in speaking rate affect the pace of an “inner clock” by which listeners measure segment durations. However, the results obtained with Norwegian listeners are compatible with such a model only if the inner clocks of listeners with a less demanding linguistic background are assumed to resist such influence to a higher degree. Apparently, listeners tend to assume that the speaking rate is more normal than it actually is, unless they are forced by their language to be more careful in taking the speaking rate into account. However, the most substantial difference between the Norwegian and the Fennic listeners consisted in the narrower reference frame of the Norwegians.

4. References

- Dommelen, W. A. van (1999) Auditory accounts of temporal factors in the perception of Norwegian disyllables and speech analogs. *Journal of Phonetics* 27: 107–123.
- Engstrand, O. & Krull, D. (1994) Durational correlates of quantity in Swedish, Finnish and Estonian: Cross-language evidence for a theory of adaptive dispersion. *Phonetica* 51: 80–91.
- Krull, D. & Traunmüller, H. (2000) Perception of quantity in Estonian. *Proceedings, Fonetik 2000*: 85–88. (Dept. of Languages, University of Skövde).
- Lehiste, I. (1970–75) Experiments with synthetic speech concerning quantity in Estonian *Congressus Tertius Internationalis Fenno-Ugristarum, Tallinnae habitus 17.–23. VIII 1970. Pars I: Acta Linguistica*. Ed. by V. Hallap, 254–69. Tallinn: Valgus
- Lehtonen, J. (1970) *Aspects of quantity in standard Finnish*. Studia Philologica Jyväskyläensia. University of Jyväskylä.
- Traunmüller, H. (1994) Conventional, biological, and environmental factors in speech communication: A modulation theory, *Phonetica*, 51, 170–183.
- Traunmüller, H. & Krull, D. (submitted) The effect of local speaking rate on the perception of quantity shown by Estonian subjects.