MEASURING RELATIVE ARTICULATION RATE IN FINNISH UTTERANCES

Jussi Hakokari[1,3], Tuomo Saarni[2], Tapio Salakoski[2], Jouni Isoaho[3], and Olli Aaltonen[1]

[1]Department of Phonetics, University of Turku
[2]Turku Centre for Computer Science
[3]Department of Information Technology, University of Turku
jussi.hakokari@utu.fi, tuomo.saarni@utu.fi

ABSTRACT

This paper presents two investigations into articulation rate, or the distribution of segmental duration, in a Finnish language speech corpus. The first study, rank ordering of short utterances according to their component words' articulation rate, reveals that 75 % or more of Finnish utterances can be expected show some level of final lengthening. Also initial shortening, or accelerated speaking rate in the beginning of utterances, is present in amounts clearly above chance level. The second study, an investigation into how relative duration progresses in utterances, confirms the observations mentioned before. Furthermore, the second study shows the initial and final effects are statistically significant. Importantly, the results are near-identical to those obtained independently from Southern Swedish, even though the languages and corpora in question are entirely different.

Keywords: final lengthening, initial shortening, speaking rate, corpus, segmental duration

1. INTRODUCTION

Final lengthening (FL), the tendency to slow down articulation at the ends of utterances, has been observed in almost any language in which the matter has been adequately investigated. The only exceptions mentioned in the literature are quantity languages such as Finnish, Hungarian, Estonian, and Japanese (cf. [8]). Even in these the tendency has been confirmed later on [1][4][5][7]. Our own recent studies into Finnish FL [9] reveal how much lengthening different phases of finality induce at phone level, but do not address the question of pervasiveness. Subjective examination of the speech signal clearly gives the impression that FL, although common, does not affect every utterance. This study will answer the question how often we can expect FL in formal Standard Finnish. Furthermore, the study at hand will follow the progress of articulation rate from the beginning to the end in utterances of varying length.

We have adopted the methodology of Hansson [2][3] with a few adjustments. As one of many experiments in her dissertation, she studied utterance-level effects on speaking rate in Southern Swedish (Scanian). Unlike most studies, that have relied on laboratory speech and even nonsense words, her materials consisted of dialect recordings made in natural settings (10 speakers, 518 phrases) in addition to elicited speech. Her methodology was to measure syllable duration within the domain of word. The average syllable duration in final and initial words was thus contrasted by that in medial words. Since the number of syllables in any word was not of concern, the approach could be called a "syllable duration on word level" study. The approach presented in this study could be equally labeled a "phone duration on word level", as the level of observation is word but the level of measurement is the phone. Additionally, the method could be described as semi-phonemic, as phones are measured according to their membership in broad phonemic categories as explained in methods and materials.

2. METHODS AND MATERIAL

2.1. Speech material

The speech material consists of two speech corpora previously studied separately. Since they both have been deemed very similar in aspects of segmental duration (FL included) [9][10], they were considered fit to be combined and studied as a single corpus in the study at hand. The first corpus consists of individual sentences of various lengths, picked out from a Finnish-language periodical Suomen Kuvalehti and read aloud by an adult male speaker. The second consists of television news, weather broadcasts and radio presentations. Being

of roughly the same size, together the corpora add up to 2115 utterances, 72720 phones, or ~93 min of continuous speech flow with any acoustic pauses other than voiceless plosives removed.

The corpora were manually annotated at phone and word levels. The annotation was done with duration studies in mind; great emphasis was put on temporal accuracy (e.g. segment boundaries were determined at the level of a single waveform).

2.2. Methods

The first experiment was designed to give an estimate on how often in fact FL occurs. While our previous study [9] confirmed FL exists in Finnish, subjective experience tells us it is not unavoidable and it does not occur in every utterance. To rank order words of a single utterance according to their speaking rate is a methodological dilemma. We first divided all the phones in the corpus into the following categories: non-plosive consonants, plosives, and vowels. These were further divided into phonologically short and long consonants, as Finnish is a quantity language with short/long distinction in consonants and vowels alike. Monophthong long vowels were also separated from diphthongs. The term 'utterance' refers to any continuous speech flow with no internal pauses. An utterance may thus be either a terminal or non-terminal intonation unit, but the criterion for the end of an 'utterance' was acoustic pause. Internal syntactic boundaries alone did not delimit utterances in this study.

Second, each phone's duration was compared against the mean duration of its respective category. For instance, a 80 ms short vowel was compared against mean duration of all the short vowels in the material (64.9 ms), and given a factor of 80 ms/64.9 ms ≈ 1.23. Third, the entire word was given a mean factor based on all of its component phones. Finally, each sentence was reordered (rank ordered) according to the factors of their component words without losing the information of the original word order. The operation was performed on all 2, 3, 4, and 5-word utterances found in the speech material (n=1020).

The second experiment was designed to show how word-level speaking rate progresses in utterances. The experiment used the same information as the first one, but this time the factors assigned to words were brought together. For instance, all 5-word utterances were merged so that the factor for its component initial word was a mean of every first word in every 5-word utterance in the material. Utterances from 2 to 9 words in length were investigated for the second experiment (n=1708).

Durations were calculated and processed for statistical analysis using specially designed scripts which extract information from Praat TextGrid files. TextGrid files contain the word and phone level annotation.

3. RESULTS AND DISCUSSION

3.1. Rank Ordering

The results from the rank ordering are summarized in table 1. The columns represent utterances of various lengths. The ordinals on the left hand side indicate the words with the slowest articulation rate (greater segmental duration). For instance, of the 240 two-word utterances only 36 were ones in which the 1st word is in fact longer in segmental duration than the second. FL % stands for the percentage of utterances in which the final word is the longest, IS % (e.g. initial shortening) for the proportion of utterances in which the initial word is the shortest. The final row ("à Hansson") shows the frequency of FL as with Hansson's [3] criteria; any utterance in which the final word is longer than the penultimate is considered to exhibit FL.

Table 1: Summary of rank ordering results.

	2-word (n=240)	3-word (n=255)	4-word (n=281)	5-word (n=244)
1st word	36	17	0	2
2nd word	204	26	23	5
3rd word		212	42	18
4th word			216	36
5th word				183
FL %	85.00	83.14	76.87	75.00
IS %	85.00	63.13	55.87	48.77
FL % à Hansson	85.00	89.01	86.97	87.29

The results show that in as much as 75-85 % of the utterances there is evidence supporting FL and in ~49-63 % evidence supporting initial shortening. It must be noted that at chance level the percentages would be in the order of 50%, 33 %, 25 %, and 20 % for the utterance lengths examined. Hansson's methodology put the Swedish figures at 71-82 %; in our material FL was somewhat more pervasive (85-89 %). For reference, Mihkla [6] reported Estonian (a language closely related to Finnish) newsreaders

producing FL 60 % of the time depending on the context, but the details of their methodology remain obscure.

3.2. Relative Duration

The figures below (1-4) show the mean relative durations and 95 % confidence intervals ($p<0.05$) of 2 to 9-word utterances. Average articulation rate is represented by 1.0 on the left hand side. Above 1.0 their segmental durations exceed that average, consequently slowing down articulation rate. Below 1.0 speaking rate is accelerated and shorter phones are produced. In the legend below, 1st always refers to the first (i.e. initial) word in the utterances; the last (i.e. final) word is always the rightmost one.

Figure 1: Relative articulation rates in 2 (n=240) and 3-word utterances (n=255)

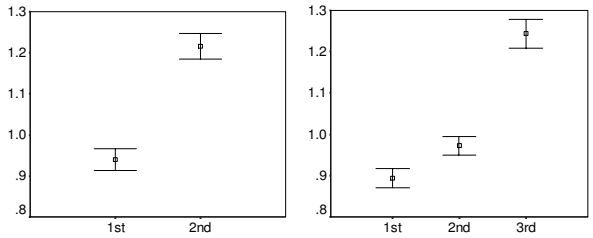


Figure 2: Relative articulation rates in 4 (n=284) and 5-word (n=244) utterances.

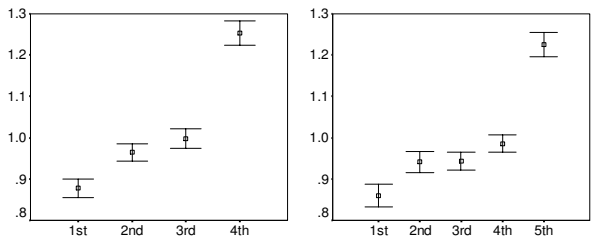


Figure 3: Relative articulation rates in 6 (n=239) and 7-word (n=205) utterances.

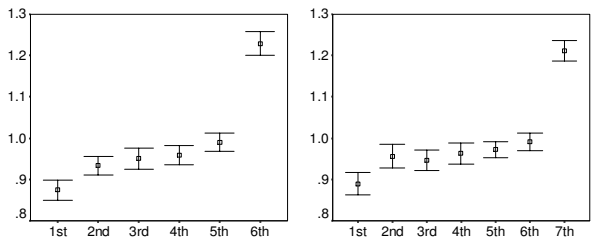
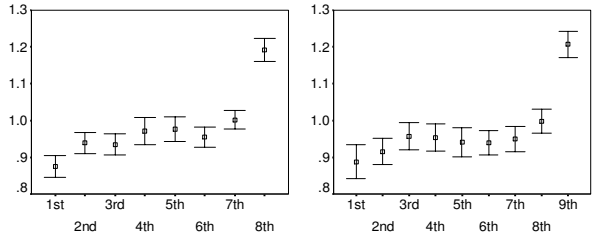


Figure 4: Relative articulation rates in 8 (n=142) and 9-word (n=99) utterances.



The articulation rate, as calculated in relative durations of their constituent phones, show results uniform across the various utterance lengths and are near-identical to Hansson's results. The results follow a simple pattern. The initial word is characterized by modest yet statistically significant shortening. The medial words are of roughly the same length, yet there is a non-significant tendency for phones to grow in duration towards the end of the utterance. Penultimate words expectedly stand out as longer than the rest of the medial words (cf. [1]), but final words are significantly lengthened. Our unpublished results further suggest that FL is not limited to the last one or two syllables, but in fact may commence earlier and gradually grow towards the boundary.

Utterance-initial shortening or compression, as it may be called, is present in each sentence type (2 to 9). In 9-word sentences it is no longer significant, though. There is a similar tendency in all phrase lengths in Hansson's [2][3] material, although the effect is not significant at 0.05 level. Nevertheless, since initial shortening is present in all Swedish utterance lengths (2 to 5 words) as well as in Finnish ones (2 to 9 words), it is fairly safe to assume the difference between the first and the second word is not a product of chance but some real articulatory or linguistic phenomenon. The factor is remarkably uniform with means between 0.85 – 0.90. For Hansson [3], the factor can be calculated to vary from 0.82 to 0.86, the 5-word phrases with a figure as low as 0.70 excluded.

Medial words are clearly discernible from the initial and the final. They are mostly level and between the initial and final words, but closer to initial than final. Hansson's Swedish data compares well. Medials in penultimate environment, having longer durations than the other medials, show a similar but weaker tendency in the Swedish data. Since Hansson has only studied phrases with 5 or fewer words, penultimate words can be compared against other medials only

in two phrase types, the 4 and 5-word ones. It is necessary to examine longer Swedish utterances before any reliable comparison can be made.

FL makes the final word considerably longer than the others ($p<0.001$). Across utterance types, the duration factor is in the excess of 1.2. From Hansson's [3] results we can deduce the corresponding factor vary from ~1.24 to ~1.38.

4. CONCLUSIVE REMARKS

We have been able to identify two phenomena related to articulation rate in our corpora. There is a considerable amount of utterance-final lengthening in the domain of utterance-final word. Second, there is a minor yet statistically significant and systematic shortening of the utterance-initial word. No conclusive effects were observed in medial words; articulation rate remains fairly stable mid-utterance regardless of the utterance's length. In other words, individuals studied here sharply began speaking with an accelerated speaking rate, and then slowed down to 'normal' until they finally slowed down considerably before pausing.

While the corpus was not annotated for stress or accent, there is no reason to attribute the findings to stress or prominence. The formal Finnish news speech is typically not accentuated to a great extent. The entire speech material used in the study is practically devoid of strong contrastive accent. The first two authors examined 21.4 % of the speech corpus and found noticeable prominence in only 13.9 % (first author) or 11.9 % (second author) of final words. That amount cannot contribute much to FL.

Perhaps the most important observation is that the results are strikingly similar between this study and the reference study on Southern Swedish dialects. The differences between the two are limited to certain details. In our material, both FL and initial shortening are more frequent but lesser in magnitude. That may have to do with individual variation or speaking style; ours is strictly formal literary style while the Swedish dialect recordings are a combination of natural and elicited speech. At this point there is no reason to suggest a language-specific background for the phenomenon. It ought to be noted that, in despite of geographical vicinity, there is no particular direct contact between Finnish and Southern Swedish (Scanian) speakers to speak of. The two groups are rarely exposed to each others speech, the languages are not genetically related, and they are very different both in terms of segmental and suprasegmental phonology. Final lengthening, as found in practically every language investigated so far, is likely to be a product of muscular and pulmonary mechanics involved in articulation, not a learned linguistic feature with a self-purposeful communicative function.

5. REFERENCES

[1] Hakokari, J., Saarni, T., Salakoski, T., Isoaho, J., Aaltonen, O. 2005. Determining prepausal lengthening for Finnish rule-based speech synthesis. *Proceedings of Speech Analysis, Synthesis and Recognition: Applications of Phonetics (SASR 2005)*, Kraków.

[2] Hansson, P. 2002. Articulation rate variation in South Swedish phrases. In: Bel, b. Marlien, I. (eds), *Proceedings of Speech Prosody 2002,* Aix-en-Provence, 371-374.

[3] Hansson, P. 2003. Prosodic phrasing in spontaneous Swedish. An academic dissertation. *Travaux de l'institut de linguistique de Lund 43*. Lund University.

[4] Hockey, B.A., Fagyal, Zs. 1999. Phonemic length and pre-boundary lengthening: an experimental investigation on the use of durational cues in Hungarian. *Proceedings of the XIVth International Congress of Phonetics Sciences*, San Francisco, 313-316.

[5] Krull, D. 1997. Prepausal lengthening in Estonian: evidence from conversational speech. In: Lehiste, I. & Ross, J. (eds.), *Estonian Prosody: Papers from a Symposium, Proceedings of the International Symposium on Estonian Prosody,* Tallinn, 136-148.

[6] Mihkla, M. 2005. Modelling pauses and boundary lengthenings in synthetic speech. *Proceedings of the Second Baltic Conference on Human Language Technologies (HLT'2005)*, Tallinn, 305-310.

[7] Vaissière, J. 1983. Language independent prosodic features. In: Cutler, A. & Ladd, R. (eds.) *Prosody: Models and Measurements*, 53-65.

[8] Venditti, J., van Santen, J. 1998. Modeling segmental durations for Japanese text-to-speech synthesis. Third ESCA Workshop on Speech Synthesis (SSW3-1998), 31-36.

[9] Hakokari, J., Saarni, T., Isoaho, J., Salakoski, T., Aaltonen, O. Prepausal lengthening in Finnish: further evidence for a phonetic universal. Submitted.

[10] Saarni, T., Hakokari, J., Aaltonen, O., Isoaho, J., Salakoski, T. 2007. Utterance-initial duration Finnish non-plosive consonants. *Proceedings of the 16th Nordic Conference of Computational Linguistics (NODALIDA 2007)*, Tartu, 160-166.