Determining Prepausal Lengthening for Finnish Rule-Based Speech Synthesis

Jussi Hakokari[1] & Tuomo Saarni[2]

Tapio Salakoski[2], Jouni Isoaho[2] & Olli Aaltonen[1]

[1]Phonetics Laboratory

[2]Department of Information Technology

University of Turku, Turku, Finland

jussi.hakokari@utu.fi & tuomo.saarni@utu.fi

ABSTRACT

We are developing a Finnish rule-based TTS system. Our primary concern is to enhance naturalness in the synthesis by observing tendencies in natural language and implementing the findings into the synthesis. We have concentrated on modeling duration, which is essential to the Finnish language due to contrasting phonemic length and the fact that the durations of individual phones are highly sensitive to their position within a word. One of the many durational considerations is prepausal lengthening. Most languages exhibit lengthening of speech sounds at the ends of phrases and sentences, but there is no applicable data available on the phenomenon with regard to Finnish. Data mining a speech corpus, we have set out to investigate what is the best justified way to implement prepausal lengthening to a Finnish rule-based TTS system. The results show that there is considerable lengthening of the entire prepausal word. Moreover, some of the effects appear to extend to the penultimate word. The data acquired in the study provides us adequate information for modeling prepausal lengthening in TTS.

1. Introduction

We are in the process of developing a Finnish rule-based TTS system. Our approach is to tackle the problem of naturalness associated with the rule-based methods by data mining natural speech corpora and trying to model the set of rules accordingly to best correspond to natural speech. For now, we are using a single-speaker corpus and aim to model that speaker instead of making compromises between a host of voices and speaking styles. Furthermore, we are developing a signal generator software that would perform better than the previous synthesizers such as Klatt. While a fully synthetic rule-based synthesis has its advantages, tolerable naturalness is rightly associated with the concatenative methods that rely on samplings

of recorded human speech. Much of our research has so far concentrated on quantity, which is a central issue in Finnish, a language with contrasting phonemic length and long, multisyllabic words [4].

We set out to model prepausal (final, preboundary) lengthening, a common phenomenon in speech, in our synthesis and came to an unexpected discovery. While there is virtually no literature on the subject specifically with regard to Finnish, several published and unpublished sources claim that namely the Finnish, Estonian (a language closely related to Finnish), and Japanese languages exhibit either little or no prepausal lengthening. Vaissière [11] presents the claim, referring to Lehiste [6]. Lehiste's work, however, deals with phonological length (the phonemic contrast between short and long speech sounds) and the syllabic structures of Finnish and Estonian. She does not discuss any acoustic phonetic duration, prepausal lengthening or other. Curiously, D'Odorico & Carubbi [3] report final syllable lengthening (FSL), a more specific instance of prepausal lengthening, to be actively suppressed in languages such as Finnish, Estonian, and Japanese. They cite Oller [9] as the source; Oller's study again mentions Finnish in no way. Conversely, Bishop [1] refers to a study by Johnson & Martin [5] claiming that final lengthening effects have been reported in Finnish among other European languages. Johnson and Martin, on the other hand, refer to Lehtonen [7], whose study deals with segmental durations in a number of syllable and phoneme environments. Lehtonen does not, however, present any conclusive data namely about prepausal lengthening. As a conclusion, it has to be admitted there is little explicit evidence that Finnish speakers would lack prepausal lengthening, a phenomenon thought by many to be a universal tendency and physiological in nature. A prepausal lengthening effect has been observed in Estonian, but it appears weaker than in English and not as pervasive [8].

Vainio's [10] investigations on the same corpus used in this study revealed a prepausal lengthening effect. Vainio's study, however, involved prosody in general and he did not describe the effect in much detail. We set out to investigate the effect more closely in order to obtain sufficient data to implement in Finnish TTS. We used data mining to compare mean durations of segments in prepausal words to those in other positions. While shedding some light on the issue of prepausal lengthening in Finnish and thus contributing to anyone interested in its potential status as a universal feature is a worthy goal, our primary concern is to find a phonetically justified and computationally inexpensive manner of implementing the effect to a TTS system.

2. Methods

2.1. Speech Material

The speech material was a single-speaker corpus of 692 Finnish declarative sentences read aloud by a 39-year-old male from Helsinki. The corpus contained ~6500 words. The sentence durations ranged from ~2 to ~20 seconds adding up to 69 minutes of recorded speech. The sentences are randomly picked from a Finnish periodical (Suomen Kuvalehti), with the exception that foreign words and foreign proper names have been avoided. The material is clearly articulated and represents Standard Finnish. The speaker has a relatively low pitch voice, and has a distinctive creak (final devoicing) in the ends of sentences, a common trait among Finnish speakers.

The corpus was segmented and annotated at word and phone levels. Pauses (silence) were separately marked in the phone level annotation. The annotation appears accurate and consistent. All the data mined information presented in the paper ultimately relies on the annotator's judgement on phone boundaries. A thorough description of the corpus and the annotation criteria are found in Vainio [10].

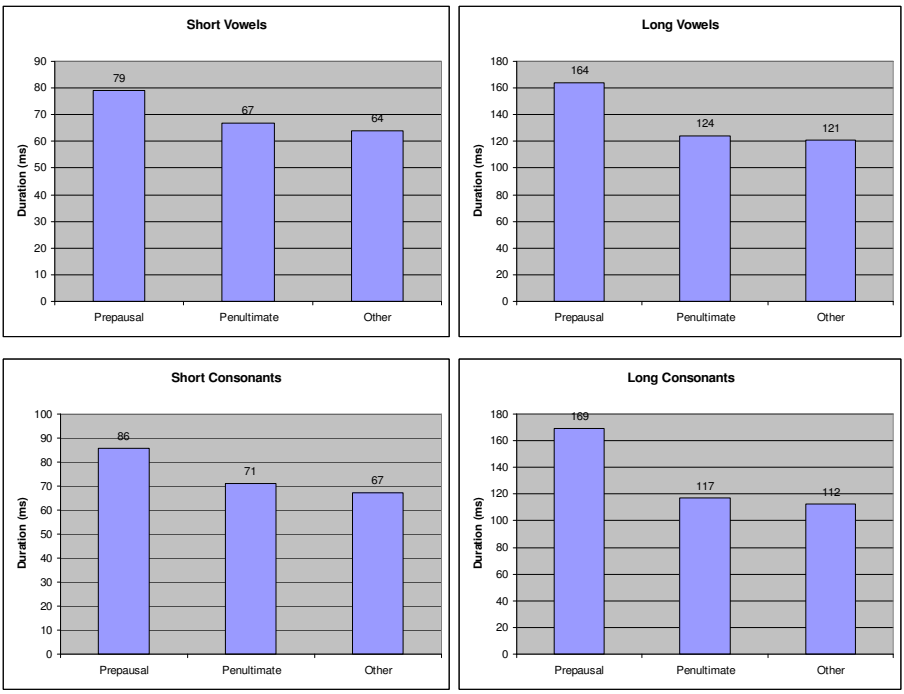
2.2. Data Mining and Procedure

Calculating mean segmental durations was done automatically using the word and the phone level annotation in the corpus. The mean segmental durations of the entire phrase-final (prepausal) words are contrasted with the rest of the speech material. Thus, syllables, final or penultimate, are not examined alone. Multisyllabic wordforms are more the rule than the exception in Finnish; the words in a prepausal position may consist of a number of syllables. Furthermore, the ends of phrases and the ends of sentences were not separated; the method treats all the words that precede a silence or the end of an utterance in the recording equally (with the exception of silences during plosive consonants, which, of course, do not trigger a pause). The penultimate words, the ones that precede the prepausal word, are also examined to determine whether the effect is extended beyond the last word before a pause.

Both the prepausal and the penultimate words were compared against the rest that consisted of phrase-initial and medial words exclusively. The rest, in other words, did not include the prepausal or the penultimate words themselves. There were two levels of examination. First, we examined the data by grouping individual phones into four categories: phonemically short consonants (C), long consonants (C:), short vowels (V), and long vowels (V:). The idea was to find out whether there is a lengthening effect and whether it applies to all the four categories. Second, we examined the lengthening effect phoneme-by-phoneme to find out if there were differences in the behaviour of different speech sounds.

3. Results

The present data suggests there is a clear prepausal lengthening effect even when the entire prepausal word is examined. That applies to all the speech sounds. The phonemically short consonants were 28.4 % longer in a prepausal position than in other positions (penultimate positions excluded). The corresponding figures for the long consonants are 50.9 %, 23.4 % for the short vowels, and 35.5 % for the long vowels. Contrary to what was expected, the lengthening selectively affected even the penultimate word. All the four categories were longer in the penultimate words compared to the overall duration of words in the medial and the phrase-initial positions (C:= 4.5 % longer, V= 4.7 %, V:= 2.5 %, and C= 6.0 %).



Picture 1. *The amount of lengthening in prepausal and penultimate positions compared to other positions.*

Table 1. *Sample sizes of all the prepausal, penultimate, and initial/ medial phones in the study.*

	Sample Size N
Prepausal Short Vowels	3422
Prepausal Long Vowels	397
Prepausal Short Consonants	3700
Prepausal Long Consonants	405
Penultimate Short Vowels	2915
Penultimate Long Vowels	276
Penultimate Short Consonants	3152
Penultimate Long Consonants	293
Other Short Vowels	12836
Other Long Vowels	1289
Other Short Consonants	13694
Other Long Consonants	1407

The prepausal lengthening effect was also examined on the phoneme level. All the vowels and the consonants were affected, including the phonemically long and short sounds, but to varying degrees. The short consonants with a sample size less than 10 are all non-native and only occur in loan words. Of these the voiced velar plosive /g/ (N=7) was an exception; it was actually shorter in prepausal positions. The lengthening of the penultimate words was not examined at the phoneme level.

Table 2. *Phoneme-to-phoneme results. The sample size refers to occurrences in a prepausal position.*

Phoneme	Percentage of prepausal lengthening	Sample size N
ø	25.4	52
æ	18.1	273
ɑ	17.3	861
e	24.6	569
i	28.4	854
o	27.8	366
u	27.1	309
y	34.8	138
b	15.1	8
d	32.9	78
f	36.4	6
g	-7.3	7
h	34.6	187
j	24.2	93
k	27.2	400
l	22.0	317
m	21.6	224
n	14.1	499
p	23.2	141
r	22.0	202
s	32.9	618
t	27.0	669
ʉ	17.4	223
ŋ	49.9	24
ʃ	132.5	4

Phoneme	Percentage of prepausal lengthening	Sample size N
øː	52.7	2
æː	42.0	64
ɑː	35.5	133
eː	44.5	56
iː	32.7	52
oː	41.8	19
uː	25.6	55
yː	17.8	16
kː	42.2	36
lː	53.5	106
mː	37.3	19
nː	48.3	31
pː	25.2	14
rː	91.9	9
sː	60.3	78
tː	52.0	98
ŋː	39.5	13

4. Discussion

Our data clearly suggests a prepausal lengthening effect of the phrase-final word. The current data is based on a single speaker. While the sample size of individual phones is adequate, we cannot generalize the effect to the entire Finnish-speaking community; a future study must include several speakers, preferably representatives of different dialect backgrounds. Our data contains declarative sentences exclusively; a future investigation should establish whether the phenomenon is present in questions and commands as well. The current data represents newspaper speech. Our study does not answer the question whether prepausal lengthening takes place in (spontaneous) conversational speech. However, a recent study of prepausal lengthening in Russian by Volskaya & Stepanova [12] suggests that the effect is not style specific. This study has only addressed segmental duration on the word level. The role of syllabic structure is yet to be determined. Furthermore, it is necessary to produce a more detailed description of how the effect is distributed towards the phrase boundary.

The phoneme-to-phoneme examination revealed that there are differences between individual phonemes, but we failed to find any systematics in the variation other than that the short voiceless consonants were lengthened somewhat more than the voiced ones (~29.1 % vs. ~20.0 %). That is in line with the tendency first observed by Cooper & Danly [2] The intrinsic mean durations of phonemes are easy to implement as such in a TTS system, and more naturalness may be achieved by using multipliers for phones in penultimate and prepausal positions. That presents a very simple and straightforward solution for introducing prepausal lengthening into a Finnish rule-based TTS system and warrants experimentation.

5. Conclusion

The categorical notion that the Finnish language lacks prepausal lengthening does not seem to hold in the light of our current data. Also the claim that Finnish speakers somehow resist prepausal lengthening effects lacks evidence. We have shown a significant lengthening effect that is observable even when the entire prepausal word is under examination. We have also extended the lengthening effect to the penultimate word; the penultimate words in their entirety show a weak effect especially in the phonemically long consonants. Its significance is dubious, however, and we are inclined to conduct a more detailed study on the distribution of the lengthening effect. Even though the data is based on a single-speaker corpus and should not be overly generalized, it serves as a basis for designing a preliminary module that introduces prepausal lengthening into our Finnish rule-based TTS system.

References

[1] Bishop, J. B. 2003. *Aspects of intonation and prosody in Bininj Gun-wok: an autosegmental-metrical analysis*. PhD thesis, School of Graduate Studies, University of Melbourne.

[2] Cooper, W. E., Danly, M. 1981. Segmental and Temporal Aspects of Utterance-Final Lengthening. *Phonetica*, 38, pp. 106-115.

[3] D'Odorico, L., Carubbi, S. Prosodic characteristics of early multi-word utterances in Italian children. *First Language*, 23(1), pp. 97-116.

[4] Hakokari, J., Saarni, T. Jalonen, M., Aaltonen, O., Isoaho, J., Salakoski, T. 2005. Word model-determined segmental duration in Finnish speech synthesis and its effect on naturalness. *Proceedings of The Second Baltic Conference on Human Language Technologies*, Tallinn.

[5] Johnson, K., Martin, J. 2001. Acoustic vowel reduction in Creek: effects of distinctive length and position in the word. *Phonetica*, 58(1-2), pp. 81-102.

[6] Lehiste, I. 1965. The function of quantity in Finnish and Estonian. *Language 41*, 447-456.

[7] Lehtonen, J. 1970. *Aspects of quantity in Standard Finnish*. Jyväskylä: K.J. Gummerus Oy.

[8] Mihkla, M. 2005. Modelling pauses and boundary lengthenings in synthetic speech. *Proceedings of The Second Baltic Conference on Human Language Technologies*, Tallinn.

[9] Oller, K. 1973. The effect of position in utterance on speech segment duration in English. *The Journal of the Acoustical Society of America,* 51, pp. 1235-1247.

[10] Vainio, M. 2001. Artificial Neural Network Based Prosody Models for Finnish Text-to-Speech Synthesis. Academic dissertation, University of Helsinki.

[11] Vaissière, J. 1983. Language independent prosodic features. A. Cutler & R. Ladd (Eds.) *Prosody: Models and Measurements*, (53-65). Springer Verlag.

[12] Volskaya, N., Stepanova, S. 2004. On the temporal component of intonational phrasing. *Proceedings of the 9th Conference "Speech and Computer"*.