Working Papers 49 (2001), xx–xx

Gender differences in vowel duration in read Swedish: Preliminary results

Christine Ericsdotter & Anna M. Ericsson
Dept. of Linguistics, Stockholm University
ericsdotter@ling.su.se, annaer@ling.su.se

Abstract
Swedish data from 9 subjects suggest that women use greater vowel duration contrasts than men do. The results are discussed with reference to research by Simpson (Simpson, 1998; Simpson, in press; Simpson, submitted) and his linguistic / biomechanical approach to the problem.

1. Introduction
Recent research by Simpson (Simpson, 1998; Simpson, in press; Simpson, submitted) show that female speakers of German and American English produce vowels with longer durations than male speakers do. Simpson's studies were based on German acoustic data of both read and spontaneous speech from *The Kiel Corpus* (Simpson, Kohler & Rettstadt, 1997), and on American English acoustic and articulatory data of read speech from the *University of Wisconsin X-ray Microbeam Speech Production Database* (Westbury, 1994). Analyzing the articulatory data of the American English speech samples, and considering the similarity between the German and the English results, Simpson suggests that the male–female difference can be explained using a biomechanical account, rather than having its origin in e.g. sociophonetics, which have been proposed in other studies (Whiteside, 1986; Byrd, 1992). Simpson argues that differences in vocal tract cross-section require males and females to traverse different articulatory distances to achieve analogous phonetic targets. In an attempt to attain similar perceptual shape in a diphthong such as [aɪ], males and females must compensate for differences in distance by moving their tongues at different speeds. One consequence of the slower female movement might be a movement with greater temporal extent. Amplitude of gesture and temporal extent are related through stiffness in the simple mass-spring model (Saltzman, 1986).

The cross-linguistic similarities in the results give weight to the biomechanical explanations, and have a number of implications for gender dynamics in speech synthesis, typology, etc. The purpose of the present study was to collect and analyze a small set of Swedish data to see if similar durational patterns are to be found in Swedish.

2. Method
2.1 Subjects
The subjects were 5 women and 5 men from the basic course of linguistics at Stockholm University. They were all in their twenties and speakers of a Central Standard Swedish.

2.2 Speech material

The speech material consisted of a single sentence where three target vowels [u] [i] [ɒ] occurred in real Swedish one-syllable words with /tVk/ structure, in bold: *Nej, här bor en **tok** med en **tik** på sitt **tak*** (≈No, a **crazy fellow** with a **she-dog** on his **roof** lives here). Three different contrastive stress patterns were elicited by questions from the experimenters, who pretended to have mistaken one of the target words for a word rhyming with the target word.

2.3 Recordings

The recordings took place in the anechoic room at the Stockholm University Phonetics Lab, using a Brüel & Kjær-microphone type 4145 connected to a Panasonic SV-3800 DAT-recorder with a 48 kHz sampling frequency. The subjects were familiar with the environment and were given time to rehearse the test phrase with the experimenter before the recording. Six repetitions of each stress pattern in the speech materials were recorded, which amounts to about 5 minutes per subject. The total time per subject, including rehearsal, questions, calibration of the sound level and recording, was about 10 minutes.

2.4. Duration measurements

The data were transferred from DAT tapes to Cool Edit (http://www.syntrillium.com) using a Sound Blaster Live sound card, and downsampled to 16 kHz mono files. The data from one female subject were excluded from the study due to errors in the stress pattern and long hesitation pauses in the middle of sentences. The phrase segments *tok med en tik på sitt tak* were labeled to vowel and phrase segment boundaries on parallel spectrograms and oscillograms using the xs tool from the Snack module (Gustafson & Sjölander, 1998). The following segmentation criteria were used: the start of the vowel was set to the first glottis pulse after the first stop's postaspiration; the end of the vowel was set to the start of the second stop's occlusion; the start of the phrase segment was set to the release of the first stop in /tok/; the end of the phrase segment was set to the release of the last stop in /tak/. Vowel and phrase segment durations were calculated, compared and statistically tested with two-tailed t-tests using MS Excel.

3. Results

3.1. Raw data

The results are shown in Table 1 and Figure 1a-1b. The male phrase segment durations were on average 100 ms longer than those of the women. In words not carrying contrastive stress, the male vowels tended to be longer than the female counterparts, the differences significant for vowels [u] and [i]. The opposite relation was true in words carrying contrastive stress, but the difference was only significant for the vowel [i].

3.2. Normalized data

To examine the role of the phrase segment duration, the data were normalized by setting the duration of all phrase segments to 1.6 s by taking *(1.6 / phrase segment duration)* phrase segment duration*, and changing the vowel durations linearly by taking *(1.6 / phrase segment duration)* vowel duration.*

The results are shown in Table 1 and Figure 1c-1d. In words not carrying contrastive stress, the male vowels still seemed to be longer than the female counterparts, but the difference was only statistically significant in the vowel [u]. In words carrying contrastive stress, female vowels were always longer than male vowels. This difference difference was significant for all vowels.

Table 1. Raw data and normalized data for phrase segment and vowel durations, men (M) and women (F). TTEST-symbols: *** $\alpha < 0.001$, ** $0.001 < \alpha < 0.01$, * $0.01 < \alpha < 0.05$, ns $\alpha > 0.05$.

PHRASE SEGMENTS	M	F	TTEST						
Tokens	90	72							
Mean (ms)	1647	1554	***						
Standard dev. (ms)	162	89							
Mean (norm ms)	1600	1600							
V IN UNSTRESSED WORDS	u			i			ɒ		
	M	F	TTEST	M	F	TTEST	M	F	TTEST
Tokens	60	48		60	48		60	48	
Mean (ms)	140	117	***	120	105	***	119	114	*ns*
Standard dev. (ms)	29	14		24	20		28	28	
Mean (norm ms)	135	120	**	115	108	*ns*	121	119	*ns*
Standard dev. (norm ms)	27	16		21	20		29	30	
V IN STRESSED WORDS	u			i			ɒ		
	M	F	TTEST	M	F	TTEST	M	F	TTEST
Tokens	30	24		30	24		30	24	
Mean (ms)	153	162	*ns*	136	154	**	180	151	*ns*
Standard dev. (ms)	24	33		17	25		69	37	
Mean (norm ms)	138	180	**	139	158	**	153	170	*
Standard dev. (norm ms)	36	63		22	21		25	34	

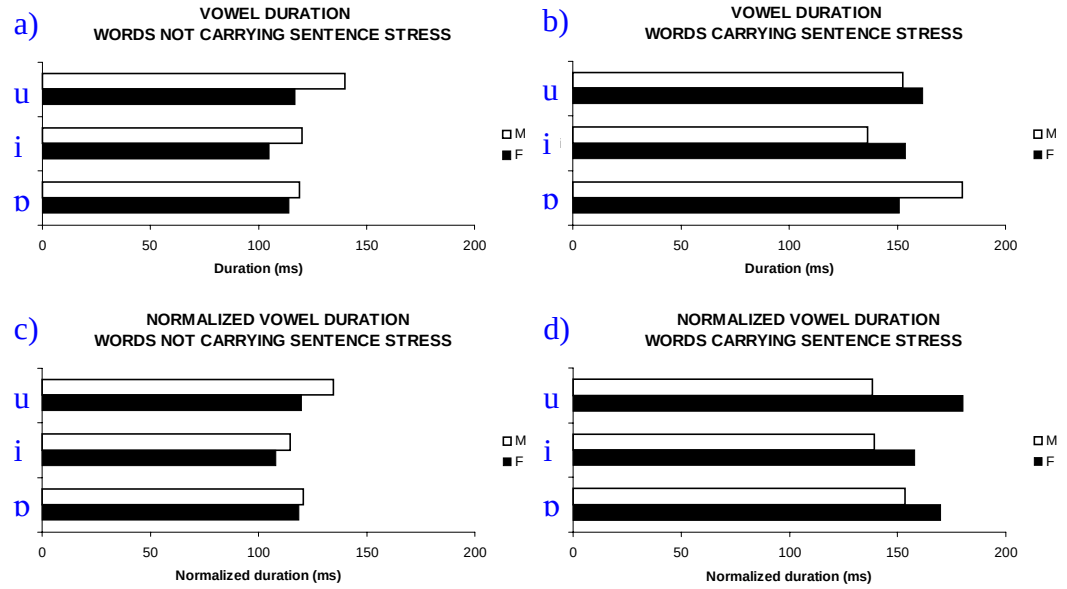

Figure 1. Comparisons of male (M) and female (F) vowel durations. a-b: raw data. c-d: normalized data.

4. Discussion

The Swedish data indicate that women use greater vowel duration *contrasts* than men do, the women producing shorter or similar vowel durations in non-stressed positions, and longer vowel durations in stressed positions. At first sight, these results are not in line with Simpson's results, where the women *always* produced longer vowels. However, Simpson's data were not analyzed with respect to stress, the American English data were vocalic stretches involving sentence stress, while in the German data, both stressed and unstressed vowels were combined in the calculation of mean values. Initial findings for three Germanic languages (Simpson & Ericsdotter, in prep), suggest an equally complex set of gender-specific durational differences in the German and American English data as those we have reported here for Swedish.

A possible interpretation of the pattern can be made with reference to Simpson's biomechanical account. Unstressed vowels tend to create a smaller articulatory and acoustical vowel space than do stressed vowels (Lindblom, 1963). If greater vocal tract dimensions require higher articulatory speed and shorter temporal extent of the movement, this would have a greater effect on stressed vowels than unstressed, the articulatory distances being longer in a wider vowel space.

Another interpretation of the pattern is based on the spontaneous impression when listening to the data that men and women appeared to use different strategies for bringing out the phrase stress pattern. The men tended to use duration, loudness and a small pitch change focused on the word to be stressed, while the women used duration, loudness and a great pitch change, starting earlier in the phrase and having its maximum on the word to be stressed. If wider pitch excursions also require longer vowel durations, this can be a factor in the female stressed vowels. The different strategies may however not be correlated to gender in general but be an effect of the present experimental situation: the experimenters were women using a female stress pattern in the "mis-hearings" which may have made the interviewed women more co-operative and comfortable in "joining the game".

Acknowledgements

Elilta Arefaine, Anna-Karin Luisti & Viveka Lundgren helped preparing the data. The authors are grateful to Adrian Simpson for his contribution to this study.

References

Byrd, D. (1992). 'Preliminary results on speaker-dependent variation in the TIMIT database.' *Journal of the Acoustical Society of America* **92**: 593-596.

Gustafson, J. & K. Sjölander (1998). 'Educational tools for speech technology.' In the proceedings of Fonetik -98, May 27-29 1998, Stockholm University: 176-179.

Lindblom, B. (1963). 'Spectrographic study of vowel reduction.' *Journal of the Acoustical Society of America* **35**: 1773-1781.

Saltzman, E. (1986). Task dynamic coordination of the speech articulators: a preliminary model. In *Generation and Modulation of Action Patterns*. **15**. Edited by H. Heuer & C. Fromm. New York, Springer.

Simpson, A. P., K. J. Kohler & T. Rettstadt (1997). *The Kiel Corpus of Read/Spontaneous Speech Acoustic data base, processing tools and analysis results.* Arbeitsberichte des Instituts für Phonetik und digitale Sprachverarbeitung der Universität Kiel AIPUK **32**.

Simpson, A. P. (1998). *Phonetische Datenbanken des Deutschen in der empirischen Sprachforschung und der phonologischen Theoriebildung*. Arbeitsberichte des Instituts für Phonetik und digitale Sprachverarbeitung der Universität Kiel AIPUK **33**.

Simpson, A. P. & C. Ericsdotter (in prep). 'Gender-specific duration differences in English, German and Swedish read speech.'

Simpson, A. P. (in press). 'Dynamic consequences of differences in male and female vocal tract dimensions.' *Journal of the Acoustical Society of America*.

Simpson, A. P. (submitted). 'Gender-specific articulatory-acoustic relations in vowel sequences.'

Westbury, J. R. (1994). *Microbeam Speech Production Database User's Handbook, Version 1.0*. Madison, WI.

Whiteside, S. P. (1986). 'Temporal-based acoustic-phonetic patterns in read speech: Some evidence for speaker sex differences.' *Journal of the International Phonetic Association* **26**: 23-40.