https://doi.org/10.3176/lu.1998.3.02

DIANA KRULL (Stockholm)

PERCEPTION OF ESTONIAN WORD PROSODY. A STUDY OF WORDS EXTRACTED FROM CONVERSATIONAL SPEECH

Background

The Estonian quantity system is unusually intricate and involves different levels of the prosodic hierarchy: segment, syllable, foot and word. The domain of syllabic quantity is a disyllabic unit where the first, stressed syllable increases in duration with increasing degree of quantity while the second, unstressed syllable correspondingly decreases. Each quantity degree has its characteristic duration ratio between the two initial syllables of a word: 2:3 for Q1 (short), 3:2 for Q2 (long) and 2:1 for Q3 (overlong) (Lehiste 1960). The relatively small temporal difference between Q2 and Q3 is complemented by different F0 patterns in the initial syllable: an early F0 peak followed by a fall in Q3, a late peak and no fall in Q2. The F0 pattern has been shown to be important for the recognition of Q3 (Lehiste 1989).

Both temporal and tonal characteristics are normally present in "laboratory speech" (word lists, prepared sentences, etc.). Studies of conversational speech, however, have shown that only the differences between duration ratios remain stable, while the characteristic F0 fall in Q3 is often absent, except in sentence final position when followed by a pause (Krull 1997). However, I. Lehiste (1989) has shown using synthetic stimuli that listeners need F0 information in addition to duration ratios in order to perceive Q3. Therefore, the question arises: can quantities in conversational speech be distinguished by their acoustic characteristics alone, or do listeners have to make use of the semantic context?

Method

To address this question, a perception test was carried out with words from the conversational speech recorded for an earlier experiment (Krull 1997) and presented to listeners without context. The test material consisted of disyllabic words excised from the conversational speech of three Estonians, one male (AE), and two females (AT, MT). The speech consisted of near monologues of 1—1.5 hours in duration. The selected words had the form $(C)V_1CV_2$ where C was a short consonant, V_1 a short, long or overlong vowel, and V_2 a short vowel. Only such words were chosen where a different quantity degree would change the meaning of the word, for example Q1 *veri* nom.sg. 'blood', Q2 *veeri*! 2p.imp. 'read slowly!', Q3 *veeri* part.pl. 'fringes, edges'(Q2 and Q3 have the same spelling), or Q1 *tuba* nom.sg. 'room', Q2 *tuuba* nom.sg. 'tuba'. In previous studies of Estonian conversational speech (Krull 1993; Engstrand, Krull 1994; Krull 1997) only content words were used.

In the present study, disyllabic pronouns, pre- and postpositions were included. The number of stimuli varied with speakers: AE had 18 Q1 words, 22 Q2 and 32 Q3. The corresponding numbers for AT were 6, 6 and 17, and for MT 2, 10 and 16.

The excised words were transferred to a digital tape, for each speaker separately. The words occurred in random order, each word twice in different surroundings. The word series of each speaker was preceded by a short practice session. The listeners were 24 native speakers of Standard Estonian, all but four originating from the Northwestern parts of the country. The words were presented through a loudspeaker, each stimulus twice with 1 s. in between and with 3 s. between stimulus pairs. The stimuli were arranged in blocks of ten pairs with 8 s. between blocks. The listeners' task was to identify the quantity degree of each word, if necessary by guessing, and mark their choice on an answer sheet.

Results

Figure 1 shows percent Q1, Q2 and Q3 answers to stimuli of each of the three degrees of quantity. The general pattern is similar for the three speakers, although the number of correct answers varies. It can be seen that recognizing Q1 did not cause difficulties and that Q2 was a little more difficult. More serious difficulties were encountered only in connection with Q3, particularly with the stimuli of speaker AT. The recognition rate of Q3 words was very variable: for example, one word could be recognized by all 24 listeners while another, similar word spoken by the same person was not recognized at all. There could be several reasons for this difference, which will be discussed further on. The most obvious reason could lie in the acoustic properties of the stimuli.

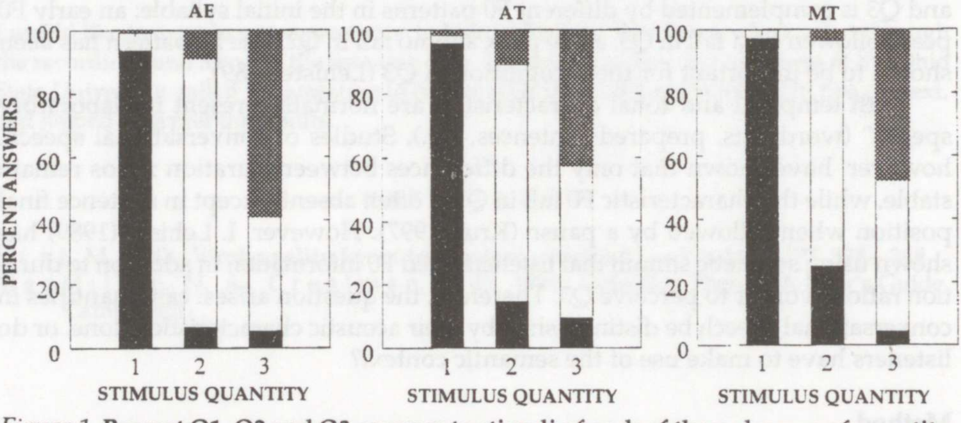

Figure 1. **Percent Q1, Q2 and Q3 answers to stimuli of each of these degrees of quantity. Black = Q1, white = Q2, and grey = Q3.**

A multiple correlation analysis on four acoustic properties of the Q3 stimuli showed that V_1 duration had the strongest influence on the listeners decisions, followed by F0 change within V_1 (see Table 1). V_1/V_2 duration ratio had a weaker influence and was statistically significant only for stimuli of speaker AE. F0 movement across the intervocalic consonant was not significant in any of the cases. Similar results were obtained for the recognition of Q3 in combined Q2 and Q3 stimuli, although with slightly higher correlation coefficients: $R2 = 0.724$ (AE), $R2 = 0.604$ (AT), and $R2 = 0.713$ (MT). For combined Q1 and Q2 stimuli, only V_1 duration and, for speaker AE V_1/V_2 duration ratio had a statistically significant effect on listeners' answers.

Table 1

Multiple correlation analysis of four acoustic properties as predictors of the recognition of Q3 stimuli

	AE		AT		MT	
Variable	**Std coeff.**	**p**	**Std coeff.**	**p**	**Std coeff.**	**p**
V_1 duration	0.627	0.000	0.630	0.001	0.627	0.000
V_1/V_2 duration ratio	0.107	0.172	0.080	0.618	0.013	0.923
% F0 change in V_1	–0.403	0.000	–0.149	0.410	–0.380	0.011
% F0 change across C	–0.042	0.585	–0.059	0.674	0.036	0.954
Multiple *R*2	0.688		0.537		0.654	

F0 peak position — typically early for Q3 and late for Q2 — had no statistically significant effect in the cases where it was present (AE in 8 Q2 and 22 Q3 stimuli, AT in 2 Q2 and 10 Q3; MT 6 Q2 and 4 Q3). However, adding peak postition raised the squared multiple correlation to *R*2 = 0.882 for the prediction of Q3 answers in AE's 20 Q2 and Q3 stimuli, *R*2 = 0.834 for Q3 stimuli alone.

To illustrate the relation between acoustic properties and listeners' perception of each degree of quantity, the stimuli were arranged into three groups according to their recognition rate: low (0—20%), mid (21—80%), and high (81—100%). (The mid group covers a larger percentage of answers because only small differences were found within it). The four acoustic characteristics best known to influence the perception of quantity were plotted against these groups. The results for speaker AE can be seen in Figure 2. The upper left graph shows the difference in V_1 duration with different recognition rates: for a good recognition of the degree of quantity V_1 duration was clearly important for all three degrees of quantity. The graph to the upper right shows large differences in the V_1/V_2 duration ratios between quantities. On the lower left graph a large difference between Q1 and Q2 on the one hand and Q3 on the other is seen for the F0 change within V1. On the lower right, finally, the placement of the F0 peak within V_1 is shown. In this case only 22 stimuli (i.e. 11 of the 54 Q2 and Q3 words) were involved for speaker AE.

Table 2

Mean temporal and tonal values in stimuli recognized by more than 80% of the listeners

	Spk	Q1	SD	n	Q2	SD	n	Q3	SD	n
V_1 duration	AE	84	20.1	35	145	30.5	29	245	56.3	26
	AT	52	1.2	4	114	3.9	5	214	20.4	5
	MT	63	29.0	12	162	0	2	213	21.5	9
V_1/V_2 duration ratio	AE	0.66	0.16	35	1.95	0.57	29	3.44	0.96	26
	AT	0.82	0.01	4	1.75	0.73	5	2.27	0.84	5
	MT	0.64	0.20	12	1.25	0	2	2.95	1.20	9
% F0 change in V_1	AE	2.39	9.74	35	–1.82	7.13	29	–27.39	14.77	26
	AT	0.81	1.52	4	–4.83	3.76	5	–17.68	7.09	5
	MT	–0.19	8.61	12	17.2	0	2	–26.37	19.16	9
% F0 change across C	AE	–10.6	16.23	35	–2.99	9.01	26	–0.94	9.11	26
	AT	–2.59	1.22	4	6.71	0.81	5	–15.58	47.46	5
	MT	–2.33	2.01	12	–7.5	0	2	16.46	34.49	9
Peak placement (in % of V_1 dur.)	AE	–	–	–	0.64	0.04	8	0.31	0.10	12
	AT	–	–	–	–	–	0	–	–	0
	MT	–	–	–	0.68	0	2	0.40	0.05	3

Percent stimuli that were recognized by more than 80% of the listeners was for speaker AE 97% for Q1, 66% for Q2 and 41% for Q3. The corresponding values for AT were 100%, 25% and 16%, and for MT 100%, 17% and 26%. The varying recognition rate of Q3 corresponds well with the results of the multiple correlation analyses: AE had three variables with a highly significant influence on Q3 the recognition, AT one (V_1 duration), and MT two (V_1 duration and F0 change in V_1). The large differences in F0 change across the intervocalic consonant may be due to the influence of the surroundings in the conversational speech.

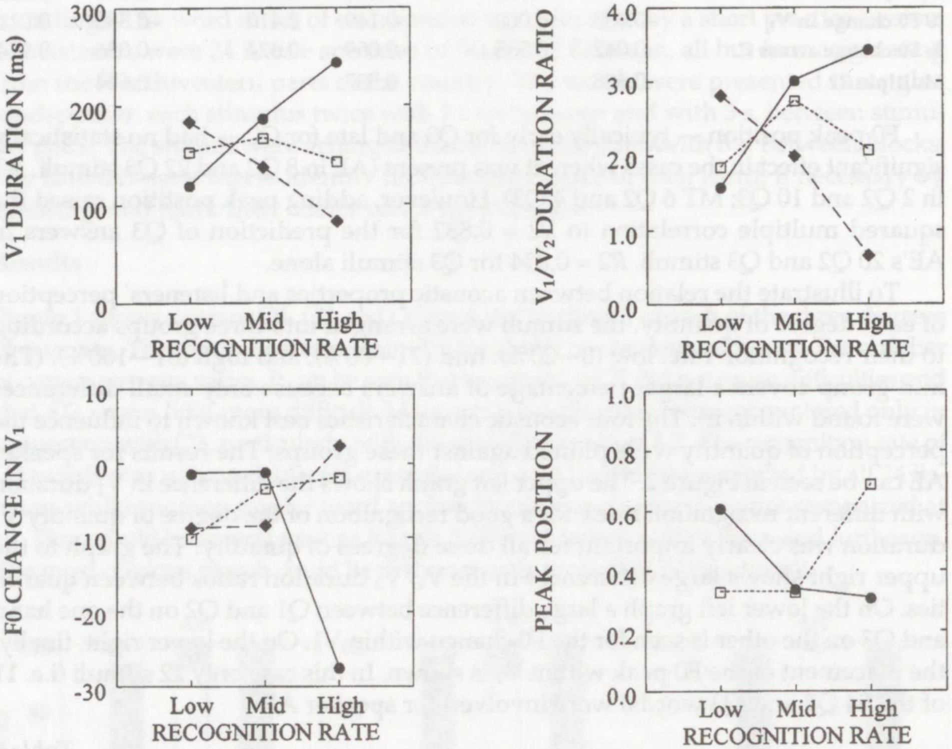

Figure 2. **V_1 duration, V_1/V_2 duration ratio, percent F0 change within V_1 and F0 peak position (as ratio of V_1 duration) in relation to different recognition rates of the quantity degree of the stimulus. Filled squares = Q1, unfilled squares = Q2, filled circles = Q3. See text for details.**

Discussion

An earlier study using partly the same material (Krull 1997) showed that the typical duration ratios between the first and second syllable constituted the largest and most stable difference between Q2 and Q3 in conversational speech. Even differences in V_1 duration were robust, in spite of a changing speaking rate. The characteristic F0 fall in Q3, however, tended to be neutralized. Several studies have shown that differences in duration, even intersyllabic duration ratios alone are not sufficient for the recognition of Q3 (Lehiste 1989; Eek 1980a; 1980b; Eek, Meister 1997) If this is the case, listeners to conversational speech probably have to use the semantic context in order to understand the words.

The present study seems to confirm this: without their context the majority of Q3 stimuli were not recognized. According to a statistic analysis, listeners used V_1 duration as the main indicator of quantity. At least in part, this influence may have been enhanced by the absence of the context and in speech tempo. A certain minimum V_1 duration seems to be necessary for a high recognition rate of Q3, in the present material the minimum was 180 ms. However, it is not sufficient without additional characteristics: for example, stimuli considerably longer than 180 could have a high rate of Q2 identifications if F0 was not falling.

The importance of F0 is further exemplified by the following comparison. The material in Krull 1997 contained two kinds of words: those that could change their meaning with the degree of quantity, and those that could not. In the present study, the second kind of words were not included. In the earlier study quite often a F0 rise — instead of the typical fall — could be found in V_1 of Q3 words, with the exception of sentence final prepausal words. This was particularly pronounced in the case of speaker AE: in the earlier study he had a mean F0 rise of about 8% in V_1 of sentence internal Q3 words, in the present material, he had, instead, a fall of over 12%. It thus seems that a speaker may — albeit unconciously — be more careful when pronouncing words whose meaning could be changed by changing the degree of quantity. This phenomenon should be studied more in detail in the future.

There may be additional factors influencing the listeners' decisions. For example, the relative frequency of occurrence of a word in the listeners' vocabulary could influence their choice of answer. It is not unusual that a CV_1CV_2 combination appears much more frequently in one degree of quantity than in another. In view of this and other possible uncontrolled influences, the prediction of the recognition rate for Q3 stimuli could be surprisingly high: e.g. $R2 = 0.834$ for AE's Q3 words with a typical F0 peak. A study of addidional factors such as the distribution of energy within the disyllabic unit as proposed by A. Eek (1986) could improve it still more.

Acknowledgements

Many thanks to those who helped me to carry out this test: dr Arvo Eek, Einar Meister, dr M. Ehala, and the students of The Pedagogical University in Tallinn who served as listeners.

REFERENCES

Eek, A. 1980a, Estonian Quantity: Notes on the Perception of Duration. — Estonian Papers in Phonetics, 5—30.

—— 1980b, Further Information on the Perception of Estonian Quantity. — Estonian Papers in Phonetics, 31—57.

Eek, A., Meister, E. 1997, Simple Perception Experiments on Estonian Word Prosody: Foot Structure vs. Segmental Quantity. — Estonian Prosody: Papers from a Symposium. Proceedings of the International Symposium on Estonian Prosody, Tallinn, Estonia, October 29—30, 1996, Tallinn, 71—99.

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. 1993, Word-Prosodic Features in Estonian Conversational Speech: Some Preliminary Results. — Department of Linguistics, Stockholm University. PERILUS XVII, Stockholm, 45—54.

—— 1997, Prepausal Lengthening in Estonian: Evidence from Conversational Speech. — Estonian Prosody: Papers from a Symposium. Proceedings of the International Symposium on Estonian Prosody, Tallinn, Estonia, October 29—30, 1996, Tallinn, 136—148.

Lehiste, I. Segmental and Syllabic Quantity in Estonian. — Indiana University, American Studies in Uralic Linguistics 1, Bloomington, 21—82.

—— 1989, Current Debates Concerning Estonian Quantity. — FUSAC'88. Proceedings of the Sixth Annual Meeting of the Fenno-Ugric Studies Association of Canada...., Lanham—New York—London, 77—86.

—— 1997, Search for Phonetic Correlates in Estonian Prosody. — Estonian Prosody: Papers from a Symposium. Proceedings of the International Symposium on Estonian Prosody, Tallinn, Estonia, October 29—30, 1996, Tallinn, 11—35.