Journal of Phonetics (1989) **17**, 167–174

Discrimination of duration ratios in bisyllabic tokens by native English and Estonian listeners

Robert Allen Fox

Division of Speech & Hearing Science, The Ohio State University, 324 Derby Hall, 154 N Oval Mall, Columbus, OH 43210, U.S.A.

and

Ilse Lehiste

Department of Linguistics, The Ohio State University, 204 Cunz Hall, 1841 Millikin Avenue, Columbus, OH 43210, U.S.A.

Received 6th July 1988, and in revised form 8th November 1988

In an earlier study [*Journal of Phonetics* (1987), **15**, 349–363], we examined the ability of Estonian and English speakers to discriminate among pairs of noise bursts whose durations were in the ratios of 1:2, 2:3, 3:2, and 2:1. The results obtained showed that both groups of listeners clearly recognized only two contrastive patterns: 1:2 and 2:3 *vs.* 3:2 and 2:1. Since there may be a difference in the perception of duration ratios when listeners are in a speech mode of perception rather than a non-speech mode, the present study is a replication of the earlier study using bisyllabic tokens as stimuli instead of noise-burst sequences. Again, the obtained results show the same contrastive patterns. The data support a reanalysis of the three-way quantity contrast in Estonian into two-way contrast based on duration (short–long *vs.* long–short) and a contrast based upon F_0 contours.

1. Introduction

Estonian has been commonly described as having a three-way quantity (duration) distinction among disyllabic word structures (Lehiste, 1960; Eek, 1983). In particular, speakers produce words in quantity 1 with a ratio of the durations of the first syllable and the second syllable of approximately 2:3. Disyllabic words in quantity 2 are produced with a duration ratio of approximately 3:2 and words in quantity 3 with a duration ratio of approximately 2:1 (Lehiste, 1960).

Fox & Lehiste (1987) recently obtained data which supported a reanalysis of this three-way quantity distinction into two separate binary decisions: one based on quantity (short–long *vs.* long–short disyllabic structures) and one possibly based on fundamental frequency differences. Their study required both Estonian and American English listeners to discriminate among pairs of noise bursts whose durations were in the ratio of 1:2, 2:3, 3:2 or 2:1. Their results showed that both groups of listeners clearly

0095-4470/89/030167 + 08 $03.00/0

recognized only two contrastive durational patterns: 1:2 and 2:3 *vs.* 3:2 and 2:1. Of special importance for understanding quantity contrasts in Estonian was the discovery that Estonian listeners (as well as English listeners) consistently failed to discriminate between the ratios 3:2 and 2:1.

However, as Fox & Lehiste (1987, p. 361) pointed out, "the stimuli used in this study were distinctly non-speech like in nature and the possibility does exist, however unlikely, that the failure to obtain a 3:2 *vs.* 2:1 duration ratio distinction with the Estonian listeners stems from a difference between the non-speech and speech modes of perception." The present study was designed to determine the extent to which listeners were better able to make duration ratio distinctions when the stimuli were distinctly speech-like. The basic experimental design was the same as that used in Fox & Lehiste (1987).

2. Method

2.1. *Subjects*

A total of 46 listeners, naive to the purposes of the study, participated in the experiment. Of these, 30 listeners were native speakers of Estonian with no known speech or hearing impairment who lived in or around Tallinn, Estonia. The other 16 listeners were native speakers of American English who were students at The Ohio State University. None of the American listeners was familiar with Estonian or any other language in which duration plays a phonologically contrastive role. The listening tests for the Estonian speakers were run in Tallinn, Estonia, while the listening tests for the English speakers were run in Columbus, OH, U.S.A.

2.2. *Stimuli*

The stimuli were synthetic bisyllabic tokens created using a cascade/parallel synthesizer (Klatt, 1980) and the resulting waveforms edited with the ILS system (Signal Technology, Inc.) to ensure exact temporal measurements. The stimuli consisted of sequences that would correspond to the conventional spellings of *babab* in English and *pabab* in Estonian. The bilabial plosives were voiceless and unaspirated in each case, which is one possible realization of the phoneme /b/ in English and the standard realization of the phoneme /p/ in Estonian, spelled either *p* or *b* according to the rules of the language. The temporal structure of these bisyllabic tokens was based upon the noise-burst sequences utilized in Fox & Lehiste (1987). Each bisyllabic token was composed of two steady-state vowels and their associated CV and VC transitions. The two vowels were separated by 80 ms of silence which corresponded to the intervocalic consonant closure. A schematic of the formant structure of the stimulus tokens is shown in Fig. 1.

Although the presence of transitions toward a final consonant created the perceptual impression of a closed syllable in final position, which, for Estonian, would make the two syllables dissimilar (the first being open, the second being closed), it has been shown (Lehiste, 1968) that in Estonian, the presence or absence of a short consonant has practically no effect on the duration of the preceding vowel. The same was assumed for English. The final consonant was provided in order to make the two syllables of each bisyllabic token identical except for the duration of the steady-state portion of the vowel.

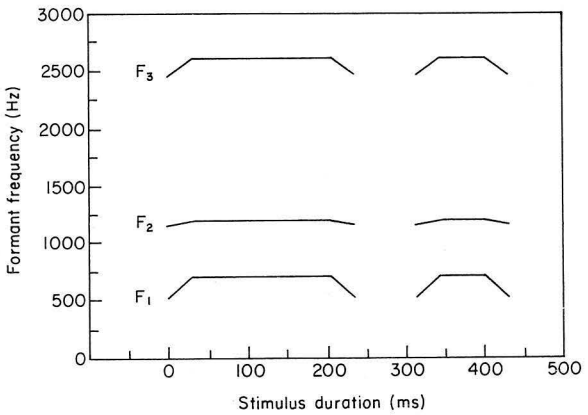

Figure 1. Schematic diagram of the formant structure of the bisyllabic tokens. This schematic represents a token with a total vowel duration of 350 ms with a duration ratio of 2 : 1.

In Fox & Lehiste (1987) there were two sets of noise-burst sequences. In one set the total duration of noise 1 + noise 2 was 350 ms, while in the second set the total duration of the noise bursts was 450 ms. These durations did not include the 80 ms of silence between the noise bursts. There were four different noise burst sequences within each of these stimulus sets corresponding to noise 1/noise 2 duration ratios of 1 : 2, 2 : 3, 3 : 2, and 2 : 1. The same basic pattern was used in creating the stimuli for the present experiment.

A total of eight bisyllabic tokens was created. In four of the tokens the total duration of the two vowels (including formant transitions) was 350 ms; in the other four tokens the total vowel duration was 450 ms. The actual duration of the entire stimulus token, including the 80 ms of intervocalic silence, was thus either 430 or 530 ms. In each of these two sets of tokens there were four different vowel 1/vowel 2 duration ratios. These ratios included 1 : 2, 2 : 3, 3 : 2, and 2 : 1; the latter three ratios are similar to those duration ratios found in Estonian (Lehiste, 1960). Changes in vowel durations were made by increasing or decreasing the length of the steady-state portion of the vowels only.

An experimental trial consisted of two bisyllabic tokens separated by a 500 ms inter-stimulus interval. There were 384 experimental trials in the experiment. Of these, 192 trials were "same" trials (in which both bisyllabic tokens had the same duration ratio, regardless of total duration) and 192 were different trials (in which the two bisyllabic tokens had different duration ratios). Please refer to Fox & Lehiste (1987) for a more detailed description.

These 384 experimental trials were randomized and divided into four blocks of 96 trials each with a short time interval between blocks (approximately 10 s). The stimuli were output under computer control at a sampling rate of 10 kHz and low-pass filtered at 4.5 kHz. These stimuli were recorded using a high quality cassette deck (Harmon/Kardon 301). A set of 50 practice items (randomly selected) was also recorded together with a small set of example trials for use while instructing the listeners about the experiment. These were 40 example trials: 10 trials with same duration ratio/same total duration token pairs; 10 trials with same ratio/different total duration; 10 trials with different duration ratio/same total duration token pairs; and 10 trials with different duration ratio/different total duration token pairs.

2.3. *Procedure*

The experimental instructions given to the listeners described the nature of the bisyllabic tokens and the vowel 1/vowel 2 duration ratio variations. The instructions used were only minimally different from those used in Fox & Lehiste (1987)—changes were only made in order to be consistent with the stimulus differences between that study and the present study. Listeners heard each of the four sets of example trials twice. The text of the instruction pamphlet described in detail (including appropriate illustrations) the nature of duration ratio differences among the bisyllabic tokens as well as the total duration differences. Two sets of instructions were constructed, one for each language group.

All subjects were run individually. Listeners took approximately 15 m to read the instruction pamphlet and hear the sample pairs. After the instructions had been completed, listeners completed the 50 practice trials followed by the four blocks of experimental trials. There was a short (2–3 m) pause between the second and third experimental blocks. The experimental task involved listening to an experimental trial and indicating (by circling the appropriate item in the test booklet) whether the vowel 1/vowel 2 duration ratios of the two bisyllabic tokens were the "same" or "different". The entire experiment took approximately 60 m.

3. Results and discussion

Mean discrimination scores in this study (transformed into percent "same" scores) are shown in Tables I and II. For easy comparison with the results obtained in our previous study (Fox & Lehiste, 1987) we have included the discrimination scores obtained using noise burst sequences in parentheses in the appropriate table columns. In general, the data show the same overall pattern for both the speech and non-speech stimuli and there is little difference between the responses from the Estonian and English listeners. With both the speech and non-speech stimuli, the percent "same" scores for identical ratio comparisons (e.g. 1 : 2/1 : 2, 3 : 2/3 : 2) are very high when the two stimuli being compared agree in terms of total duration, but these percentages decrease markedly when the two

TABLE I. Percent "same" responses for Estonian and English listeners when both stimuli in a trial have the same total duration. The percentages in parentheses are from Fox & Lehiste (1987) and correspond to discrimination of duration ratios using noise-burst sequences

Duration ratio of first stimulus	Duration ratio of second token			
	1 : 2	2 : 3	3 : 2	2 : 1
Estonian listeners				
1 : 2	95.7 (96.7)	86.3 (89.7)	10.4 (10.7)	3.8 (9.8)
2 : 3	87.9 (85.7)	94.0 (95.5)	11.7 (18.8)	4.2 (8.9)
3 : 2	4.2 (6.7)	10.0 (12.9)	96.9 (97.6)	92.5 (93.3)
2 : 1	4.5 (5.8)	6.3 (6.3)	85.4 (93.3	96.9 (98.7)
English listeners				
1 : 2	94.5 (93.9)	88.3 (77.7)	12.5 (18.3)	7.0 (11.6)
2 : 3	91.4 (78.6)	94.5 (92.3)	20.3 (33.0)	8.5 (17.9)
3 : 2	16.4 (18.3)	34.4 (14.7)	96.1 (93.6)	87.5 (91.1)
2 : 1	7.8 (14.2)	15.6 (15.2)	94.5 (89.2)	95.1 (93.3)

TABLE II. Percent "same" responses for Estonian and English listeners when the total durations of the two stimuli in a pair are different. The percentages in parentheses are from Fox & Lehiste (1987) and correspond to discrimination of duration ratios using noise-burst sequences

Duration ratio of first stimulus	Duration ratio of second token			
	1:2	2:3	3:2	2:1
Estonian listeners				
1:2	72.1 (61.6)	53.3 (52.2)	6.7 (9.8)	6.7 (10.3)
2:3	58.8 (45.5)	50.8 (53.7)	13.3 (12.9)	7.5 (8.9)
3:2	5.8 (8.9)	10.0 (14.2)	61.5 (68.1)	62.9 (53.5)
2:1	3.3 (7.6)	6.3 (6.3)	50.0 (52.2)	82.2 (73.8)
English listeners				
1:2	69.5 (63.8)	52.3 (50.9)	8.6 (15.2)	6.3 (17.4)
2:3	53.9 (51.3)	63.3 (60.7)	12.5 (22.3)	7.0 (13.8)
3:2	6.3 (12.1)	17.2 (16.5)	71.1 (69.0)	63.3 (62.1)
2:1	3.1 (11.6)	9.4 (10.7)	60.1 (52.2)	77.1 (75.4)

stimuli have different total durations. With both stimulus types listeners consistently identified (incorrectly) the 1:2/2:3 and the 3:2/2:1 ratio comparisons as "same". These responses were also affected by differences in total duration. All listeners were much more accurate in discriminating among the other remaining duration ratios (i.e., 2:3/3:2, 1:2/3:2, 2:3/2:1, and 1:2/2:1) under both stimulus type conditions. As in Fox & Lehiste (1987) we have partitioned the discrimination data into three separate groups to facilitate analysis and discussion of the results. These three groups include: (1) the identical ratio comparisons (1:2/1:2, 2:3/2:3, 3:2/3:2, and 2:1/2:1); (2) the non-identical but similar ratio comparisons (1:2/2:3 and 3:2/2:1); and (3) the non-identical and dissimilar ratio comparisons (2:3/3:2, 1:2/3:2, 2:3/2:1, and 1:2/2:1).

3.1. *Identical ratio comparisons*

The identical ratio comparisons involved the presentation of two stimuli having the same duration ratios but not necessarily the same total duration. The correct response in each case was "same". The obtained scores were remarkably similar for both language groups (Estonian, 81.0%; English, 80.3%) and both stimulus types (bisyllabic tokens, 81.8%; noise bursts, 80.5%). The discrimination responses for both language groups and both stimulus groups tended to be more accurate when the two stimuli being compared had the same total duration than when they had different total durations.

To determine the overall effect of duration ratio, total duration, language group, and stimulus type upon listener responses, a four-way repeated-measures mixed-design analysis of variance with the factors Ratio (1:2, 2:3, 3:2, 2:1), Duration (same or different total duration), Language (Estonian or English), and Stimulus type (bisyllabic token or noise-burst sequence) was done on the discrimination scores.[1] Since there was an unequal number of listeners in the Language × Stimulus type groups, the analyses

[1]One factor that was not considered in the analyses of variance was that of ratio order within a trial (e.g., 1:2/2:3 *vs.* 2:3/1:2). As in Fox & Lehiste (1987), this factor was not included because of the small number of data points (4) that would be included in each Duration × Ratio × Language × Stimulus type cell for each listener.

were done using the SAS General Linear Model program for unbalanced designs (Ray, 1982). As in Fox & Lehiste (1987), the data were converted to "rationalized" arcsine values (Studebaker, 1985) to avoid the statistical problems often associated with proportional (percentage) data. Note that although the numbers referred to in the text represent the raw percentage scores all relevant analyses were completed using the arcsine-transformed data.[2]

There was a significant main effect of Duration [$F(1,784) = 856.6, p < 0.0001$] which demonstrated that listeners were significantly more accurate when the stimuli being compared had the same total duration (95.4%) than different total durations (66.7%). There was also a significant effect of Ratio [$F(3,784) = 16.5, p < 0.0001$]. Duncan's multiple range comparison showed that listeners' responses were significantly more accurate (at the 0.05 level) for the 2:1 duration ratio (86.7%) than for the duration ratios 1:2 (80.9%), 2:3 (75.2%), and 3:2 (81.6%). Responses to the 2:3 duration were also significantly less accurate (at the 0.05 level) than those for the 1:2 and 3:2 duration ratios. The lowest scores were found with the 2:3 ratios. There were no significant main effects due to either Language [$F(1,784) = 0.28, p > 0.59$] or Stimulus type [$F(1,784) = 1.66, p > 0.19$].

Only the Duration × Ratio interaction was significant [$F(3,784) = 9.10, p < 0.0001$] which was due to the fact that the responses to the individual ratios were relatively uniform in the same total duration condition, but varied significantly in the different total duration condition. The Duration × Language interaction, which was significant in the noise burst data alone (Fox & Lehiste, 1987), was only near significance [$F(1,784) = 7.68$, $p = 0.057$]. This showed that the Estonian listeners tended to be more accurate than the English listeners in the same duration condition (Estonian, 96.5%; English 94.0%) but less accurate in the different duration condition (Estonian, 65.4%; English, 68.3%). None of the other two-, three- or four-way interactions approached significance.

3.2. *Non-identical but similar ratio comparisons*

The non-identical but similar ratio comparisons included the 1:2/2:3 (both short–long ratios) and 3:2/2:1 comparisons (both long–short ratios). The latter comparisons are representative of the quantity 2 *vs.* quantity 3 distinction in Estonian. Note that the "same" responses are actually incorrect responses as each of these experimental trials consisted of *different* duration ratios. The responses were very similar between the two language groups (Estonian, 71.5%; English, 70.8%), but there was some difference between the stimulus type groups (bisyllabic tokens, 72.7%; noise sequences, 69.9%). The total duration factor again played a significant role. In general, the percentage "same" responses in the same duration condition (Estonian, 89.3%; English, 86.5%) was 30–40% higher than in the different duration conditions where responses were near chance level (Estonian, 53.7%; English, 55.2%).

As in the previous section, the data were arcsine transformed and analyzed using a four-way repeated-measures mixed-design analysis of variance with the factors Ratio (1:2/2:3 and 3:2/2:1), Duration (same or different total duration), Language (Estonian or English), and Stimulus type (bisyllabic token or noise-burst sequence). As before, this represents an unbalanced design and the SAS General Linear Model program was used (Ray, 1982). There were significant main effects of Duration [$F(1,392) = 344.9$,

[2]The means and standard deviations of the arcsine transformed bisyllabic token data for each of the three groups of data are available upon request.

$p < 0.0001$], Ratio [$F(1,392) = 6.89, p < 0.009$], and Stimulus type [$F(1,392) = 6.30$, $p < 0.013$]. As noted above, the significant Duration effect stems from the fact that listeners are much less likely to say "same" if the total durations of the two stimuli being compared are different. The significant Ratio effect was obtained because listeners were more likely to say "same" (incorrectly) to the 3 : 2/2 : 1 comparison (73.7%) than to the 1 : 2/2 : 3 comparison (68.7%). The significant Stimulus type effect was obtained because listeners made more errors (i.e., incorrect "same" responses) when the stimuli were bisyllabic tokens (72.7%) than noise-burst sequences (69.9%). This gives strong evidence that the results obtained in Fox & Lehiste (1987) were *not* a function of having listeners operate in a non-speech as opposed to speech mode of perception. None of the two-, three-, and four-way interactions approached significance.

3.3. *Non-identical and dissimilar ratio comparisons*

The next group of ratio comparisons (2 : 3/3 : 2, 1 : 2/2 : 3, 3 : 2/2 : 1, and 1 : 2/2 : 1) represented experimental trials in which one stimulus exhibited a short–long structure, while the other stimulus exhibited a long–short structure. These data include ratio comparisons which represent the quantity 1 *vs.* quantity 2 (2 : 3/3 : 2) and quantity 1 *vs.* quantity 3 (2 : 3/2 : 1) distinctions in Estonian. These scores also represent percentage incorrect discrimination responses.

In general, the listeners are much more accurate in making these discriminations than in the non-identical but similar ratio comparisons. Here for the first time we find a language group difference, that is, Estonian listeners are somewhat more accurate than are English listeners.

As above, these data were arcsine transformed and analyzed with a four-way repeated-measures mixed-design analysis of variance with the factors Ratio (2 : 3/3 : 2, 2 : 1/3 : 2, 2 : 3/2 : 1, and 1 : 2/2 : 1), Duration (same or different total duration), Language (Estonian or English), and Stimulus type (bisyllabic token or noise-burst sequence) using the SAS General Linear Model program for unbalanced designs. For these data there was no significant main effect of Duration [$F(1,784) = 3.32, p < 0.069$] but there were significant main effects of Ratio [$F(1,784) = 11.09$, $p < 0.0009$], Language [$F(1,784) = 39.5$, $p < 0.0001$], and Stimulus type [$F(1,784) = 17.63$, $p < 0.0001$]. The Ratio effect was further analyzed using Duncan's multiple range comparison which showed that the percent "same" responses to the 2 : 3/3 : 2 ratio (16.6%) were significantly greater than those for any of the other three responses (10.8%, 9.4%, and 4.7% for the comparisons 1 : 2/3 : 2, 2 : 3/2 : 1, and 1 : 2/2 : 1, respectively). The same trend can be observed in both the Estonian and English data which is interesting since the 2 : 3/3 : 2 ratio comparison corresponds to a quantity 1 *vs.* quantity 2 distinction in Estonian. The significant Language effect stems from the fact that Estonian listeners are, in general, more accurate than are English listeners (Estonian, 8.7%; English 14.9%). The significant Stimulus type effect was obtained because listeners are somewhat more accurate when the stimuli are bisyllabic tokens (9.0%) rather than noise bursts (13.2%).

Only the Duration × Language interaction was significant [$F(1,784) = 4.61$, $p < 0.03$]. This was obtained because the English responses were significantly less accurate in the same duration condition (17.0%) than in the different duration condition (12.7%)—the Estonian responses were approximately the same in both conditions (same duration, 8.6%; different duration, 8.7%). None of the other two-, three-, or four-way interactions approached significance.

4. General discussion

The pattern of discrimination results obtained using both the bisyllabic tokens and the noise-burst sequences (Fox & Lehiste, 1987) demonstrates that listeners seem to be able to distinguish between short–long and long–short duration ratios on the basis of durational information alone. They seem unable to make more precise discriminations which involve comparing two duration ratios that are both either short–long (e.g., 1 : 2/2 : 3) or long-short (e.g., 3 : 2/2 : 1). Neither the linguistic background of the listener (i.e., whether or not the listener spoke a quantity language) nor the linguistic status of the stimulus tokens (i.e., noise-bursts or bisyllables) seem to have a significant effect upon the ability to make these more precise discriminations. As suggested in Fox & Lehiste (1987), these data may reflect a psychophysical limitation based on the number of possible internal duration representations (see Fraisse, 1946, 1956; Povel, 1981). The question that remains to be answered is what kind of information is present in the acoustic signal that listeners of Estonian use to distinguish among words exhibiting all three duration ratios (2 : 3, 3 : 2, 2 : 1)? One of the authors (IL) has treated this question in a separate article (Lehiste, 1988) and has argued that fundamental frequency differences play an important role. Clearly, as evidenced by the data described here and in Fox & Lehiste (1987), duration ratios alone are insufficient to convey the linguistic distinction.

This research was supported by Grant 1 RO1 NS21121-01 from NINCDS, National Institutes of Health. The authors gratefully acknowledge the help of their colleagues Arvo Eek, Mati Hint, and Kullo Vende in conducting the perceptual tests with the Estonian subjects in Tallinn.

References

Eek, A. (1983) Kvantiteet ja rõhk eesti keeles. *Keel ja Kirjandus*, **26**, 481–559.

Fox, R. A. & Lehiste, I. (1987) Discrimination of duration ratios by native English and Estonian listeners, *Journal of Phonetics*, **15**, 349–363.

Fraisse, P. (1946) Contribution a l'étude du rythme en tant que forme temporelle, *Journal de Psychologie Normale et Pathologique*, **39**, 293–304.

Fraisse, P. (1956) *Les structures rythmiques*. Louvain: Publications Universitaires de Louvain.

Klatt, D. H. (1980) Software for a cascade/parallel formant synthesizer, *Journal of the Acoustical Society of America*, **67**, 971–995.

Lehiste, I. (1960) Segmental and syllabic quantity in Estonian. In *American Studies in Uralic linguistics*, pp. 21–82. Bloomington: Indiana University.

Lehiste, I. (1968) Vowel quantity in word and utterance in Estonian, *Congressus Secundus Internationalis Fenno-ugristarum*, Helsinki, pp. 293–303.

Lehiste, I. (1988) Current debates concerning Estonian quantity. In *FUSAC '88: Proceedings of the sixth annual meeting of the Fenno-Ugric Studies Association of Canada* (J. Nevis, editor).

Povel, D.-J. (1981) Internal representation of simple temporal patterns, *Journal of Experimental Psychology*, **7**, 3–18.

Ray, A. A. (1982) *SAS User's Guide: Statistics*, SAS Institute, Cary NC.

Studebaker, G. A. (1985) A "rationalized" arcsine transformation, *Journal of Speech and Hearing Research*, **28**, 455–462.