Journal of Phonetics (1987) **15**, 349–363

Discrimination of duration ratios by native English and Estonian listeners

Robert Allen Fox

Speech & Hearing Science Section, Department of Communication, The Ohio State University, 324 Derby Hall, 154 North 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 April 1987, *and in revised form 17th June* 1987

It has been suggested that Estonian has a three-way quantity distinction among disyllabic word structures in terms of the ratio of the duration of first syllable compared to the duration of the second syllable. The present study examines the ability of Estonian speakers and American English speakers to discriminate among a set of duration ratios independent of other phonetic factors such as fundamental frequency or segmental variations. Subjects were required to discriminate among pairs of noise bursts whose durations were in the ratios of 1:2, 2:3, 3:2 and 2:1. In half the noise sequences the combined duration of noise 1 + noise 2 was 350 ms, in the other half the combined duration was 450 ms. The results obtained show that both groups of listeners clearly recognized only two contrastive patterns: 1:2 and 2:3 vs. 3:2 and 2:1. In addition, overall sequence duration had a significant effect upon duration ratio discrimination. In terms of a language group difference, the Estonian and English groups differed only in terms of responses to very different ratios (e.g. 1:2 vs. 2:1; 2:3 vs. 2:1), in particular, the Estonian group produced fewer errors. The data support a reanalysis of the three-way quantity contrast into two distinct binary decisions: one based on quantity (short–long vs. long–short disyllabic structures) and one possibly based on fundamental frequency differences.

1. Introduction

Estonian has been described by Lehiste (1960) as having a three-way quantity (duration) distinction among disyllabic word structures. In particular, she found that Estonian speakers produced 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 were produced with a duration ratio of approximately 3:2, and disyllabic words in quantity 3, with a ratio of approximately 2:1. The average durations of the syllables analyzed in the study were 106 and 151 ms for words in quantity 1; 295 and 187 ms for words in

0095-4470/87/040349 + 15 $03.00/0

quantity 2; and 435 and 195 ms for words in quantity 3. Duration of the first syllable is, of course, contrastive at the syllabic level and the average durations of the syllable can be classified into the three categories of short, long, and overlong, as has been traditional in Estonian phonetics and phonology.

These durational differences are accompanied by different fundamental frequency (F_0) patterns. All three factors (i.e. duration ratio, first syllable length, F_0 variations) are phonetically present in these disyllabic words. Several linguists have been interested in determining the hierarchy of relative importance among these three factors and much work has been done in describing the three-way quantity opposition in Estonian. Among the most important papers is the recent study by Eek (1983) which is directly relevant to the present study.

Eek (1983) related the ratios between syllable durations to speech tempo and to F_0 patterns with the disyllabic sequence. Basically, he found that words were heard as being in quantity 1 when the ratio of the second vowel to the first vowel was equal to or larger than 1.2. The word was assigned to quantity 2 when the V_2/V_1 ratio was between 0.57 and 0.81, and to quantity 3 when the ratio was equal to or smaller than 0.43. Differences in tempo and F_0 also played important roles. According to Eek's study, quantities 1 and 2 differ primarily in duration, as quantity 1 could be turned into quantity 2 (and vice versa) by manipulation of duration alone. Additional phonetic features are required for the perception of quantity 3.

However, one problem in interpreting Eek's data is that the range of ratio values described for each quantity seems too precise and too complex. Intuitively, one might expect the ratios to represent relatively simple numbers such as 2:3, 3:2 and 2:1 (as suggested by Lehiste, 1960). There seems to be some experimental support for this intuitive notion in studies by Fraisse (1946, 1956) and Povel (1981).

Fraisse (1946, 1956) examined the ability of subjects to produce and perceive durations in the context of temporal patterns. Fraisse (1946) found that subjects who were asked to produce temporal patterns by tapping (consisting of 2–6 taps) basically used only two different durations (which he termed *long* and *short*). The longer duration was typically at least twice as long as the shorter one. Stimuli in which the longer duration was less than double the duration of the shorter interval tended to be reproduced with a long: short ratio of 2:1. Fraisse thus suggested that the subjects seem to be able to conceptualize only two distinct durations that roughly relate as 2:1.

Povel (1981) agreed that Fraisse's data could be understood in terms of a two-duration limitation on internal representations, but felt that this solution seemed too simple. He re-examined how well subjects could imitate simple temporal patterns using a wider variety of duration ratios than did Fraisse (from 1:4 to 4:5). Povel required subjects to imitate (by tapping) sequences of 150-ms beeps the onset intervals of which were varied systematically. The duration ratios of the intervals between beeps were relationships which can be expressed numerically as 1:4, 1:3, 2:5, 1:2, 3:5, 2:3, 3:4, and 4:5 (in Povel's study the shorter duration interval was always first, resulting in ratio values smaller than 1). The results of his first two experiments (consisting of two-, or three-interval stimuli) demonstrated that the only correctly reproduced duration ratio was 1:2 (i.e. 0.50). The errors in production were systematic. There was a tendency for the subjects' responses to drift toward the 1:2 ratio so that ratios below 0.50 were increased (e.g. a ratio of 0.40 was reproduced as 0.45) and ratios above 0.50 were made smaller (e.g. a ratio of 0.66 was reproduced as 0.49). Under special circumstances set up for a third experiment, subjects were able to imitate interval relations of 1:3 and 1:4 accurately,

but this required the first interval duration to be repeated at least twice. The two-interval stimuli used in the first experiment correspond most closely to the Estonian disyllabic word structures and here Povel's results support the hypothesis that subjects are limited to only two different internal duration representations.

Povel conducted his experiments at Indiana University and his subjects were presumably native speakers of American English. It is legitimate to ask whether similar results would be obtained if the subjects had been speakers of a language in which duration plays a contrastive role. This would be especially interesting if the subjects were native speakers of a language like Estonian which has the three-way quantity contrast described above.

The theoretical interest in the problem is at least two-fold. There are linguistic schools of thought that claim that all linguistic oppositions are binary. In terms of duration, this would mean that sounds could be long or short, but there would be no three-way durational contrasts of the type 2 : 3 vs. 3 : 2 vs. 2 : 1. The data from Fraisse (1946, 1956) and Povel (1981) seem to support this point of view. If speakers of a language with this type of three-way quantity opposition likewise can only identify duration ratios of short–long (or long–short), the three-way opposition must be manifested by other means. If, however, speakers of such a language can identify additional durational ratios, especially those that occur in their native language, then it would be the case that the native language of a subject could influence his or her performance in a psychoacoustic test.

As the perception of a duration ratio involves some comparison of the duration of one stimulus to the duration of a second stimulus, a short discussion of the considerable work which has been done in determining the just-noticeable-differences (JNDs) in duration perception may be relevant at this point. Lehiste (1970) described a number of different studies which examined JNDs for duration discrimination using simple tone stimuli. Two studies (Stott, 1936; Henry, 1948) found JNDs of approximately 10–40 ms in the stimulus duration range of 30–300 ms (which is about the range of duration of interest in the present study). A third study (Ruhm, Mencke, Milburn, Cooper, and Rose, 1966) found much smaller JNDs to the order of 2.6 ms with a stimulus duration of 100 ms. Abel (1972) found that the JNDs necessary for 75% correct duration discrimination of noise bursts low-pass filtered at a variety of frequencies was between 5 and 40 ms for stimulus durations between 50 and 400 ms.

On the basis of these JND results, particularly those from Ruhm *et al.* (1966), one might expect that listeners could make very fine duration ratio discriminations. However, an experimental task requiring listeners to determine whether or not one stimulus has the same duration as a second stimulus is very different from an experimental task requiring listeners to make some type of response based on an internally computed duration ratio. For example, in the task used by Abel (1972), listeners heard two non-speech stimuli in sequential order and were asked which of the two stimuli was longer. Listeners did not have to compute a duration ratio value, nor were they required to compare one paired set of stimuli to a second paired set of stimuli. However, in the experimental task used in the present study, listeners had to compare duration ratios between two difference noise burst sequences. This required listeners to compare the results of two duration ratio computations. In terms of speech perception processes, if a language has a three-way quantity distinction it is insufficient for a listener to make a simple discrimination between the duration of two successive syllables (or vowels); rather, some type of ratio computation must be done. This being the case, then, it does

not seem that the JND data for duration discrimination are directly relevant to the present study and thus may not easily predict listeners' success at making duration ratio discriminations.[1]

The present study was designed to examine (i) the ability of listeners to discriminate among such duration ratios independent of other phonetic factors such as fundamental frequency or segmental variations, and (ii) whether speakers of a quantity language (e.g. Estonian) are better able to make such discriminations than speakers of a non-quantity language (e.g. English). In particular, the experiment required listeners whose native language was Estonian or English to discriminate between pairs of noise burst sequences which differed in terms of the ratios of the duration of first noise burst to the second noise burst.

2. Method

2.1. *Subjects*

A total of 56 listeners, naive to the purposes of the study participated in the experiment. Of these, 28 listeners, all students at the Ohio State University, were native speakers of English with no known speech or hearing impairment. None of these speakers were familiar with Estonian or any other language in which duration plays a phonologically contrastive role. The other 28 listeners were native speakers of Estonian with no known speech or hearing impairment who lived in or around Tallinn, Estonia. The listening tests for the English speakers were run in Columbus, OH, while the listening tests for the Estonian speakers were run in Tallinn, Estonia. In addition, five phonetically trained listeners (one native English speaker and four native Estonian speakers) participated in the experiment. Two of the trained listeners (RF and IL) knew the structure of the stimuli (though not the order of presentation). Their responses should represent the limiting case for whether any listener can accurately make all these ratio discriminations.

2.2. *Stimuli*

The basic stimulus consisted of a sequence of two noise bursts (speech noise) separated by a pause of 80 ms. It was felt that these stimuli were more similar to the disyllabic word patterns in Estonian (e.g. two vowels separated by a consonant) than were Povel's (1981) beep stimuli. The noise bursts were generated by a set of Coulborne modules (including a noise generator, electronic gates and timers) with a 10 ms rise/fall time to prevent audible clicks. Eight different noise-burst sequences were created. In four of the sequences the total duration of noise burst 1 (NOISE1) and noise burst 2 (NOISE2) was 350 ms; in the other four sequences the total duration of NOISE1 and NOISE2 was 450 ms. This total duration difference was included because we wanted to ensure that listeners were comparing duration ratios and not, for example, the durations of the first noise burst in

[1] We have collected some pilot data using a ratio comparison task to determine duration ratio JNDs for a selected number of duration ratios with the same type of noise-burst sequences used in the present study. For the 2:1 ratio (NOISE1 = 233 ms, NOISE2 = 117 ms) a level of 75% correct discriminations was obtained when it was compared to a sequence with a 6:5 ratio (NOISE1 = 191 ms, NOISE2 = 159 ms). These data suggest that duration changes much greater than used by Ruhm *et al.* (1966) are necessary to produce discriminable duration ratio differences. Given that these pilot data involved only a single overall NOISE1 + NOISE2 duration they almost certainly underestimate actual duration ratio JNDs.

each sequence. This stimulus variation should prevent the confounding of duration ratio differences with noise burst duration differences.

In each of these two sets of sequences, there were four different NOISE1/NOISE2 duration ratios. In particular, for each of the two duration groups, there were four stimuli in which the ratio of the duration of NOISE1 to NOISE2 was either 1:2, 2:3, 3:2, or 2:1. Three of these duration ratios (i.e. 2:3, 3:2, and 2:1) are similar to the syllable 1/syllable 2 duration ratios found in Estonian (Lehiste, 1960). The ratio 1:2 was included partly for the sake of symmetry, to offer an equal number of cases in which the first member of the pairs of noise bursts was longer than the second as compared to cases in which the second member was long, and partly to increase the comparability with the stimuli used by Povel (who had used the 1:2 ratio among his set of stimuli).

An experimental trial consisted of two noise-burst sequences separated by a 500 ms interstimulus interval. There were 384 experimental trials in the experiment. Of these, 192 trials were "same" trials and 192 trials were "different" trials.

The "same" trials consisted of two noise burst sequences whose NOISE1/NOISE2 duration ratios were the same, but whose total durations could be the same or different. In particular, each noise burst sequence was paired with a sequence having the same duration ratio and either the same total duration (12 times) or a different total duration (12 times). Each of the eight different noise burst sequences thus appeared in 24 "same" trials.

The "different" trials were composed of two noise burst sequences whose NOISE1/NOISE2 duration ratios were different and whose total durations could be the same or different. In particular, each noise burst sequence was paired (eight times) with each of the six other possible duration ratios (three different duration ratios having the same or different total durations). Each of the eight noise burst sequences thus occurred in 24 "different" trials as the first sequence in the trial and in 24 "different" trials as the second sequence. Each combination of "different" duration ratio stimuli (e.g. 1:2 vs. 2:3) occurred 32 times (ignoring order and overall duration variations).

These 384 experimental trials were randomized and divided into four blocks of 96 trials each with a short time interval between blocks. The trials were tape recorded using a high quality cassette deck (Harmon/Kardon 301). A set of 50 practice trials (randomly selected) were also recorded together with a small set of example trials for use while instructing the listeners about the experiment. There were 40 example trials: 10 trials with same ratio/same total duration stimulus pairs; 10 with same ratio/different total duration stimulus pairs; 10 with different ratio/same total duration stimulus pairs; and 10 with different ratio/different total duration stimulus pairs.

The experimental instructions given to the listeners described the nature of the noise burst sequences (including schematic illustrations) and the NOISE1/NOISE2 duration ratio. Listeners were explicitly told that they were to decide whether the duration ratios of the two noise burst sequences were the same or different and were warned to ignore any total duration differences between the two stimuli. Listeners heard each of the four sets of example trials twice. The accompanying text in the instruction pamphlet described whether or not the duration ratios of the example sequences were the same or different and whether or not their total durations were the same or different. Two sets of instructions and response sheets were constructed, one for each language group. Both groups heard the same stimulus tape.

2.3. *Procedure*

All subjects were run individually in the experiment. The instructions for the experiment (complete with example trials) took approximately 15 min. After receiving these instructions, listeners completed the 50 practice tokens followed by the four blocks of experimental trials. There was a short (2–3 min) 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 NOISE1/NOISE2 duration ratios for the two noise burst stimuli in the trial were the "same" or "different". The entire experiment took approximately 60 min.

3. Results and discussion

Mean discrimination scores obtained in this study (transformed into percent "same" scores) are shown in Tables I–IV. Each table shows the responses to a particular ratio comparison when the noise burst sequences have either the same total duration or different total durations. In general, the data show the same overall pattern of responses

TABLE I. Per cent "same" responses for Estonian listeners when both noise burst sequences are of equal total duration

	Ratio of second sequence			
Ratio of first sequence	1:2	2:3	3:2	2:1
1:2	96.7	89.7	10.7	9.8
2:3	85.7	95.5	18.8	8.9
3:2	6.7	12.9	97.6	93.3
2:1	5.8	6.3	93.3	98.7

TABLE II. Per cent "same" responses for Estonian listeners when the total durations for the two noise burst sequences are different

	Ratio of second sequence			
Ratio of first sequence	1:2	2:3	3:2	2:1
1:2	61.6	52.2	9.8	10.3
2:3	45.5	53.7	12.9	8.9
3:2	8.9	14.2	68.1	53.5
2:1	7.6	6.3	52.2	73.8

TABLE III. Per cent "same" responses for English listeners when both noise burst sequences are of equal total duration

	Ratio of second sequence			
Ratio of first sequence	1:2	2:3	3:2	2:1
1:2	93.9	77.7	18.3	11.6
2:3	78.6	92.3	33.0	17.9
3:2	18.3	14.7	93.6	91.1
2:1	14.2	15.2	89.2	93.3

TABLE IV. Per cent "same" responses for English listeners when the total durations for the two noise burst sequences are different

Ratio of first sequence	Ratio of second sequence			
	1:2	2:3	3:2	2:1
1:2	63.8	50.9	15.2	17.4
2:3	51.3	60.7	22.3	13.8
3:2	12.1	16.5	69.0	62.1
2:1	11.6	10.7	52.2	75.4

for both the Estonian and English listeners. The mean percent scores for identical ratio comparisons (e.g. 1:2 vs. 1:2) are very high when both noise burst sequences are of the same total duration, but decrease markedly when they have different total durations. Both listener groups consistently identified (incorrectly) the 1:2 vs. 2:3 and the 3:2 vs. 2:1 ratio comparisons as the "same". These responses were also affected by the total duration difference. Listeners were much more accurate in discriminating among the other "different" duration ratios. To facilitate analysis and discussion of these results, we have partitioned the discrimination data into three separate groups: (1) identical ratio comparisons, (2) non-identical but similar ratio comparisons (1:2 vs. 2:3 and 3:2 vs. 2:1), and (3) non-identical and dissimilar ratio comparisons (1:2 vs. 3:2, 1:2 vs. 2:1, 2:3 vs. 3:2, and 2:3 vs. 2:1).

3.1. *Identical ratio comparisons*

The identical ratio comparisons required listeners to determine whether or not two noise burst stimuli had the same NOISE1/NOISE2 duration ratios, independent of whether the two stimuli were of the same total duration. The obtained responses are shown in Fig. 1. The mean per cent scores, collapsed across ratios and durations were remarkably

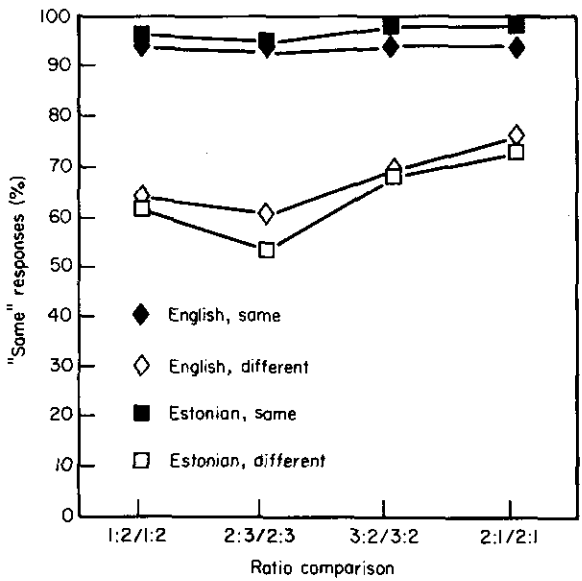

Figure 1. Responses to the identical ratio comparisons collapsed across listeners within language groups.

similar for the two language groups (Estonian, 80.6%; English, 80.3%)[2] and listeners in both language groups show the tendency to be more accurate in duration ratio discriminations when the two noise burst sequences have the same total duration (Estonian, 97.0%; English, 93.3%) than when they have different total durations (Estonian, 64.2%; English, 67.3%).

To determine the overall effect of duration ratio, total duration, and language group variations upon listener responses a three-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 durations in the two noise burst sequences being compared), and LANGUAGE (Estonian or English) was done on the discrimination scores.[3] RATIO and DURATION were within-subject factors and LANGUAGE was a between-subjects factor. The basic data consisted of the per cent "same" responses for each subject to a given ratio comparison. However, given that the variance for such per cent data may not be homogeneous (especially when scores near 100% or 0% are included in the data), these per cent scores were converted to "rationalized" arcsine transform values (Studebaker, 1985). Note that, although we will discuss the effects in terms of per cent "same" scores, all relevant analyses were done using the arcsine transformed data. The means and standard deviations of the transformed scores for each of the three groups of data appear in the Appendix.

There was a significant main effect of DURATION [$F(1,378) = 817.8$, $p < 0.0001$] which shows that listeners were much more accurate when the noise burst sequences being compared had the same total duration (95.2%) rather than different total durations (65.8%). There was also a significant effect of RATIO [$F(3,378) = 13.2$, $p < 0.001$]. Duncan's multiple range comparison showed that ratios 1:2 (78.9%) and 2:3 (75.1%) were both significantly different (at the 0.05 level) from ratios 3:2 (82.0%) and 2:1 (85.1%). Thus listeners were slightly more accurate in making judgments on the long/short duration ratio sequences. The responses for the 2:1 ratios were significantly different from all other ratios. There was no significant difference for the third main effect of LANGUAGE, [$F(1,54) = 171.2$, $p > 0.69$].

There was a significant DURATION × RATIO interaction [$F(3,378) = 6.04$, $p < 0.001$] which was mainly due to the fact that although the responses for all ratios were similar in the same duration condition (95.3%, 93.9%, 95.6%, and 96.0% for 1:2, 2:3, 3:2, and 2:1, respectively), the responses for 1:2 (62.7%) and 2:3 (57.2%), were significantly lower than those for 3:2 (68.6%) and 2:1 (74.6%) in the different duration condition. There was also a significant LANGUAGE × DURATION interaction [$F(1,378) = 9.94$, $p < 0.001$] indicating that, although the Estonian listeners had a greater number of "same" responses in the same duration condition, they had fewer "same" responses in the different duration condition. The LANGUAGE × RATIO [$F(3,378) = 0.77$] and LANGUAGE × DURATION × RATIO [$F(3,378) = 0.20$] interactions were non-significant.

Although these results demonstrate the strong effect which total duration difference may have upon listener responses, there was no clear suggestion that Estonian listeners do better at the task, in general, than do English listeners. However, this set of data does

[2]All discrimination scores cited in the text were significantly different from chance level (50% same responses) unless otherwise noted.

[3]One factor that was not considered in the analyses of variance was that of order (e.g. 1:2/2:3 vs. 2:3/1:2). We did not include this factor because of the relatively small number of data points (4) that would be included in each DURATION × LANGUAGE × RATIO × ORDER cell for each listener.

not examine whether or not the two language groups respond differentially to comparisons involving noise burst sequences having different duration ratios. The second set of data involve such discriminations and include the 3:2 vs. 2:1 comparison, which is representative of the quantity 2 vs. quantity 3 distinction in Estonian.

3.2. *Non-identical but similar ratio comparisons*

The two non-identical but similar ratio comparisons represented experimental trials in which the pair of noise burst sequences were minimally different. These comparisons included 1:2 vs. 2:3 (both short–long sequences) and 3:2 vs. 2:1 (both long–short sequences) and the mean per cent responses are shown in Fig. 2. Note that these "same" responses are actually incorrect responses, as each of these experimental trials are composed of two *different* duration ratios. The responses were again remarkably similar across the two language groups (Estonian, 70.7%; English, 69.1%). Duration variations also played a significant role, as the per cent "same" responses for the same duration condition (Estonian, 90.5%; English, 84.2%) were 30–40% higher than the responses for the different duration condition (Estonian, 50.8%; English, 54.1%) which were near chance level (the Estonian responses in the different duration condition were not significantly different from chance).

As above, these data (after arcsine transformation) were analyzed with a three-way repeated-measures mixed-design analysis of variance with the factors RATIO (1:2 vs. 2:3 and 3:2 vs. 2:1), DURATION, and LANGUAGE. There were significant main effects of DURATION [$F(1,162) = 269.0$, $p < 0.001$] and RATIO [$F(1,162) = 7.34$, $p < 0.008$]. The significant RATIO effect was obtained because listeners were more likely to say "same" incorrectly to the 3:2 vs. 2:1 comparison than to the 1:2 vs. 2:3 comparison (a difference of about 5%). There was no significant main effect of LANGUAGE [$F[1,54] = 0.03$] and all of the interaction effects including DURATION × RATIO [$F(1,162) = 3.26$], LANGUAGE × DURATION [$F(1,162) = 3.06$], LANGUAGE × RATIO [$F(1,162) = 1.38$], and LANGUAGE × DURATION × RATIO [$F(1,162) = 0.15$] were non-significant.

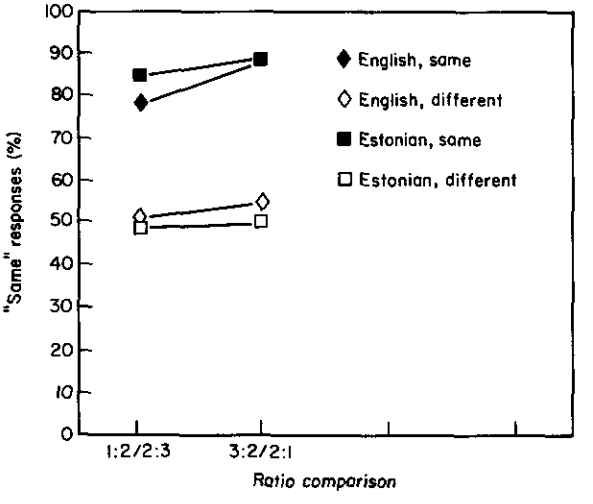

Figure 2. Responses to the non-identical but similar ratio comparisons collapsed across listeners within language groups.

There was, again, no evidence to suggest that the listener groups differed in their ability to discriminate between the duration ratios. For example, there was less than 2% difference between the mean responses for the Estonian listeners and the English listeners (Estonian, 69.4%; English, 71.5%) for the 3:2 vs. 2:1 duration ratio comparison (reflecting a quantity 2 vs. quantity 3 comparison in Estonian).

3.3. *Non-identical and dissimilar ratio comparisons*

The four non-identical and dissimilar ratio comparisons (1:2 vs. 3:2, 1:2 vs. 2:1, 2:3 vs. 3:2, and 2:3 vs. 2:1) represented experimental trials in which the two noise burst sequences exhibited a relatively greater degree of variation in terms of duration ratio differences than in the previous set of data. In particular, this set of data included all those comparisons in which the NOISE1:NOISE2 ratio was long–short (e.g. 3:2 and 2:1) in one sequence but short–long (e.g. 1:2 and 2:3) in the other. These data include comparisons which are similar to the quantity 1 vs. quantity 2 (2:3 vs. 3:2) and quantity 1 vs. quantity 3 (2:3 vs. 2:1) distinctions in Estonian. The mean percentage responses are shown in Fig. 3.

In general, listeners are much more accurate in making these four distinctions than in the non-identical but similar ratio comparisons, and the Estonian listeners made about half the number of incorrect responses than did the English listeners (Estonian, 9.9%; English, 16.4%). There was a small overall difference between the same (Estonian, 10.0%; English, 17.9%) and different (Estonian, 9.9%; English, 15.0%) total duration conditions, but this difference was not significant, unlike the significant differences found in the identical and non-identical but similar ratio comparisons. Of greater interest with this set of data is whether or not the language group difference is significant.

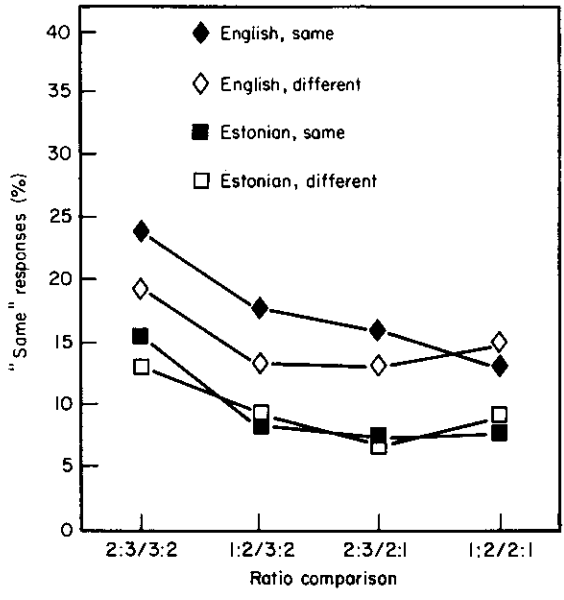

Figure 3. Responses to the non-identical and dissimilar ratio comparisons collapsed across listeners within language groups.

These data were arcsine transformed and analyzed with a three-way repeated-measures mixed-design analysis of variance with the factors RATIO, DURATION, and LANGUAGE. There was no significant main effect due to DURATION [$F(1,378) = 1.14$], but there were significant main effects due to RATIO [$F(1,378) = 14.3, p < 0.001$] and LANGUAGE [$F(3,54) = 4.04, p < 0.05$]. The significant LANGUAGE effect shows that Estonian listeners are significantly less likely to make incorrect "same" responses than are the English listeners. The contribution of individual duration ratio difference to the overall RATIO effect was tested using Duncan's multiple range comparison which showed that the per cent "same" responses to the 2:3 vs. 3:2 ratio comparison was significantly different from all other comparisons (at the 0.05 level). In particular, listeners from both language groups were more likely to incorrectly respond "same" to this comparison than in any other comparison. This result is all the more interesting in the case of the Estonian listeners for whom the 2:3 vs. 3:2 duration ratio comparison corresponds to a quantity 1 vs. quantity 2 difference.

None of the interaction effects including DURATION × RATIO [$F(3,378) = 1.07$], LANGUAGE × DURATION [$F(1,378) = 0.79$], LANGUAGE × RATIO [$F(3,378) = 0.21$], and LANGUAGE × DURATION × RATIO [$F(3,378) = 0.49$] were significant.

3.4. *Phonetically trained listener responses*

Mean discrimination scores obtained for the phonetically trained listeners are shown in Tables V and VI. The small sample size for the phonetically trained group ($N = 5$) prevents easy statistical comparison with the data from the two untrained groups. However, general tendencies in discrimination responses among the three groups can be discussed.

The mean per cent scores for the identical ratio comparisons were only 3% higher for the phonetically trained listeners (83.3%) than for either of the two other groups, which

TABLE V. Per cent "same" responses for phonetically trained listeners when both noise burst sequences are of equal total duration

	Ratio of second sequence			
Ratio of first sequence	1:2	2:3	3:2	2:1
1:2	96.7	87.5	20.0	20.0
2:3	72.5	96.7	32.5	12.5
3:2	10.0	12.5	95.0	75.0
2:1	5.0	15.0	87.5	90.8

TABLE VI. Per cent "same" responses for phonetically trained listeners when the total durations for the two noise burst sequences are different

	Ratio of second sequence			
Ratio of first sequence	1:2	2:3	3:2	2:1
1:2	80.8	67.5	20.0	10.0
2:3	55.0	64.2	22.5	12.5
3:2	5.0	10.0	67.5	62.5
2:1	2.5	5.0	47.5	75.0

could be a function of sampling error. As with the untrained listeners, the trained listeners were also more accurate in duration ratio discriminations when the two noise bursts had the same total duration (94.8%) than when they had different total durations (71.9%).

In the non-identical but similar ratio comparisons, mean "same" responses from the trained listeners collapsed across duration conditions (69.4%) were within 1.0% of the mean responses from the other two language groups. For this group, the per cent of "same" responses in the same duration condition (80.7%) was slightly more than 20% greater than in the different condition (58%), which was near chance level. This pattern is very similar to that found for the other two listener groups.

In the non-identical and dissimilar ratio comparisons, the trained listeners again showed a pattern of responses similar to that of the untrained listeners. They were much more accurate in making correct "different" responses (13.5% "same" responses overall) than in the non-identical but similar ratio comparisons. This per cent value was midway between the corresponding responses for the Estonian and English group means. There was again a small overall difference between the same (16.0%) and different (11.0%) total duration conditions.

In general, the pattern of responses for the phonetically trained listeners was extremely similar to those from the other two listener groups. In addition, the data from the two informed listeners who were familiar with the nature of the stimulus set as well as the goals of the experiment performed no better and no worse than did the other phonetically trained listeners. In particular, listeners were well able to discriminate between duration ratios 1 : 2 and 2 : 3 on the one hand, and ratios 3 : 2 and 2 : 1 on the other hand, but were unable to make accurate 1 : 2 vs. 2 : 3 and 3 : 2 vs. 2 : 1 discriminations. As before, total duration differences affected listener responses (particularly in the identical comparisons and nonidentical but similar comparisons), suggesting that subjects sometimes found it difficult to make discriminations based purely upon duration ratios.

4. General discussion

Let us return to the theoretical questions posed earlier. Listeners seem to be able to distinguish between shorter and longer signals and to decide whether the first or second member of a given sequence was longer. However, under the conditions of this experiment, the listeners did not distinguish between the ratios 1 : 2 and 2 : 3 on the one hand, and 3 : 2 and 2 : 1 on the other hand. The linguistic background of the listeners did not have any important effect upon this outcome although Estonian listeners were less likely to make incorrect "same" responses when the two stimuli were quite different. The amount of phonetic training also did not have any important effect upon listener responses. Trained listeners produced the same pattern of responses as did untrained listeners. These results may reflect a psychophysical limitation which prevents listeners from being able to make these distinctions. The hypothesized limitation upon the number of internally represented durations suggested by Fraisse (1946, 1956) and supported by experiments 1 and 2 in Povel (1981) would explain these data reasonably well.

From the point of view of Estonian prosody, the following conclusions may be drawn. The results clearly imply that words in quantity 1, with a duration ratio of 2 : 3, are perceived as distinct from words in quantities 2 and 3, with duration ratios of 3 : 2 and 2 : 1, respectively. These two long quantities, however, are not distinguished on the basis

of duration ratio. As under normal conditions listeners do indeed recognize the difference between words in quantities 2 and 3, other phonetic factors must provide the decisive information. Fundamental frequency contours are the most likely candidate, but further research may bring new information and new ideas. The present experiment suggests that Estonian should rightfully be considered an accent language, in which phonetic factors other than durational ones play a significant role.[4] The experiment also demonstrates that phonetics may, indeed, provide crucial information that must be taken into account when questions of linguistic theory are to receive satisfactory solutions.

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. This study was improved by following suggestions made by Prof. Mari Jones, Prof. van den Broecke, and two anonymous reviewers. The authors also thank the statistical consultants at the Instructional and Research Computer Center for their help in analyzing the data. Reprints may be obtained from the first author.

References

Abel, S. (1972) Duration discrimination of noise and tone bursts, *Journal of the Acoustical Society of America*, **52**, 1219–1223.

Eek, A. (1983) Kvantiteet ja rohk eesti keeles (I). *Keel ja Kirjandus*, **26**(9), 481–489; **26**(10), 549–559.

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.

Henry, F. (1948) Discrimination of the duration of a sound, *Journal of Experimental Psychology*, **38**, 734–743.

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

Lehiste, I. (1970) *Suprasegmentals*, Cambridge, MA: MIT Press.

Povel, D.-J. (1981) Internal representation of simple temporal patterns, *Journal of Experimental Psychology*, **7**(1), 3–18.

Ruhm, H. B., Mencke, E. O., Milburn, B., Cooper, W. A., and Rose, D. E. (1966) Differential sensitivity to duration of acoustic stimuli, *Journal of Speech and Hearing Research*, **9**, 371–384.

Shearer, W. M. (1982) *Research Procedures in Speech, Language, and Hearing*. Baltimore: Williams & Wilkens.

Stott, L. H. (1936) Time-order errors in the discrimination of short tonal durations, *Journal of Experimental Psychology*, **18**(7), 41–766.

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

Winer, B. J. (1962) *Statistical Principles in Experimental Design*. New York: McGraw-Hill.

Appendix

"Rationalized" arcsine transformation of the proportional data

It is well known that with proportional scales (such as per cent correct scores) the variances are correlated with the means and the data are not normally distributed about the mean (Studebaker, 1985). In order to reduce these problems in performing statistics upon proportional data (such as have been obtained in this study) many researchers suggest the use of arcsine scale transformation (e.g. Shearer, 1982; Winer, 1962). The

[4]We note that 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 non-speech and speech modes of perception. To address this possibility, we are developing a stimulus set with duration ratio differences similar to the noise burst stimuli used in the present study but much more speech-like. These perceptual tests will also use both Estonian and English listeners.

TABLE AI. Mean "rationalized" arcsine transformed values corresponding to proportion correct scores for the identical ratio comparisons grouped by language, ratio, and duration (same vs. different) (standard deviations are given in parentheses)

	Duration ratio comparison				
	1:2/1:2	2:3/2:3	3:2/3:2	2:1/2:1	Mean
English					
Same duration	101.4(14.6)	99.0(16.7)	100.5(14.4)	100.3(15.5)	100.3(15.2)
Different duration	63.7(18.8)	60.5(21.5)	69.5(19.9)	76.4(17.4)	67.5(20.1)
Mean	82.5(25.3)	79.7(27.2)	85.0(23.3)	88.3(20.3)	83.9(24.2)
Estonian					
Same duration	105.1(12.0)	102.0(12.9)	107.2(10.1)	108.1 (8.7)	105.6(11.1)
Different duration	62.2(22.3)	53.8(22.9)	68.9(20.2)	74.0(22.2)	64.7(22.9)
Mean	83.6(28.0)	77.9(30.5)	88.0(25.0)	91.0(24.0)	85.1(27.3)
Overall mean	83.1(26.6)	78.8(28.8)	86.5(24.1)	89.7(22.2)	84.5(25.8)

TABLE AII. Mean "rationalized" arcsine transformed values corresponding to proportion correct scores for the non-identical but similar ratio comparisons grouped by language, ratio, and duration (same vs. different) (standard deviations are given in parentheses)

	Duration ratio comparison		
	1:2/2:3	3:2/2:1	Mean
English			
Same duration	78.9(19.0)	92.4(19.5)	85.7(20.3)
Different duration	51.2(18.6)	54.9(19.4)	53.1(18.9)
Mean	65.0(23.3)	73.7(27.0)	69.4(25.5)
Estonian			
Same duration	86.9(19.8)	93.4(19.4)	90.2(19.7)
Different duration	49.7(19.1)	50.0(21.8)	49.8(20.3)
Mean	68.3(26.9)	71.7(30.0)	70.0(28.4)
Overall mean	66.7(25.1)	72.7(28.4)	69.7(26.9)

TABLE AIII. Mean "rationalized" arcsine transformed values corresponding to proportion correct scores for the non-identical and dissimilar ratio comparisons grouped by language, ratio, and duration (same vs. different) (standard deviations are given in parentheses)

	Duration ratio comparison				
	2:3/3:2	1:2/3:2	2:3/2:1	1:2/2:1	Mean
English					
Same duration	21.9(24.4)	15.3(20.5)	12.2(22.5)	8.5(21.5)	14.5(22.5)
Different duration	17.7(21.1)	9.9(19.6)	10.3(20.8)	10.8(22.4)	12.2(21.0)
Mean	19.8(22.7)	12.6(20.1)	11.3(21.5)	9.7(21.8)	13.3(21.7)
Estonian					
Same duration	13.0(19.8)	2.8(17.6)	2.4(14.5)	1.6(17.7)	4.9(17.6)
Different duration	9.9(18.8)	4.3(17.6)	1.2(14.1)	3.6(18.1)	4.7(17.3)
Mean	11.5(17.6)	3.6(17.4)	1.8(14.2)	2.6(17.8)	4.9(17.6)
Overall mean	15.6(21.3)	8.1(19.3)	6.5(18.8)	6.1(20.0)	9.1(20.2)

standard arcsine transformation has the disadvantage, however, that the transformed values have little intuitive relationship to the proportional data as normally described (e.g. in terms of per cent values). To address this problem, Studebaker (1982) developed what he termed a "rationalized" arcsine transform. This transform is a simple linear transformation of the more common arcsine transform which produces transformed values very close to the corresponding per cent values of the data (see Studebaker, 1982, for details about the computation of these values).

Tables AI–AIII provide the means and standard deviations for the transformed proportional data used in the three ANOVAs described in the text. Note that the values go slightly higher than 100 for per cent values which correspond to 100% and slightly lower for 0 for per cent values which correspond to 0%. Individual entries in the tables for a particular duration ratio comparison, for a particular language group, with either the same or different overall duration, are based upon $N = 28$ (the number of listeners in a particular group).