PERCEPTION OF PROMINENCE BY ESTONIAN AND ENGLISH LISTENERS*

ILSE LEHISTE
and
ROBERT ALLEN FOX
The Ohio State University

This study deals with the perception of prominence in speech and nonspeech signals by listeners who are native speakers of Estonian or English. The stimuli consisted of repetitions of the synthetic syllable [ɓɑ] at a constant fundamental frequency (F_0) or signal-correlated noise tokens whose amplitude envelope matched that of the [ɓɑ]. The basic token was 400 msec in duration. Listeners heard sequences of four stimulus tokens separated by 120 msec of silence. One token in the sequence could be lengthened to 425, 450, 475, or 500 msec and/or increased in amplitude by 3 or 6 dB; changes in duration and amplitude were independent. The speech ([ɓɑ]) and nonspeech (noise) listening tests were run separately. The listeners' task was to indicate which of the four tokens was "most prominent" The responses showed that, for English-speaking listeners, amplitude cues overrode duration cues, while Estonian listeners were more responsive to duration cues. For both groups of listeners, smaller increments in duration and/or amplitude sufficed to produce perceived prominence in the noise condition than in the speech condition. This is interpreted as being due to the absence of F_0 variation. In spoken language, F_0 is a significant component of stress both in production and perception. Thus the absence of F_0 variation should have greater effect on the perception of speech-like stimuli than noise stimuli. The results of the experiments show this to be the case.

Key words: stress perception, Estonian

INTRODUCTION

Many studies have shown that a listener's native language may affect perceptual judgment of phonetic segments in a way that reflects language-specific phonetic or systematic properties. For example, Bladon and Lindblom (1981) and Flege, Munro, and

---

* A partial report of Experiment 1 was presented to the Spring 1991 meeting of the Acoustical Society of America in Baltimore, MD. A partial report of Experiment 2 was presented to the Spring 1990 meeting of the Acoustical Society of America in State College, PA. These experiments are described in the reverse order in this report for purposes of clarifying the relationship between the two stimulus sets used and the conclusions that can be drawn from each study. This work was partially supported by a grant from NSF to the first author (BNS–8719891).

Address correspondence to Ilse Lehiste, Department of Linguistics, The Ohio State University, Columbus, OH 43210.

Fox (1991) have argued that a listener's previous linguistic experience will affect vowel dissimilarity judgments. Identification of consonants may, similarly, be affected by language background. Good examples are provided by cross-language studies of the production and perception of initial plosive consonants (Caramazza, Yeni-Komshian, Zurif, and Carbone, 1973; Williams, 1977).

The present study is one in a series in which we have explored the possible influence of listeners' linguistic background on their perception of suprasegmental cues in the speech signal. Languages are frequently classified according to their suprasegmental structure into stress languages, tone languages, and quantity languages. Intensity, fundamental frequency, and duration play different roles in these languages in the manifestation of their suprasegmental structure. It may be expected that listeners speaking a language of one of these types may perceive suprasegmental cues differently from listeners speaking a language of a different suprasegmental type.

Estonian belongs among the languages in which duration is significantly employed in signaling phonological oppositions. Disyllabic Estonian words are characterized by a quantity pattern distributed over two syllables in such a way that words in the short quantity (Q1) have a duration ratio between the syllables amounting to 2:3, words in the long quantity (Q2) have a ratio of 3:2, and words in the overlong quantity (Q3) have a ratio of 2:1 (*cf.* Lehiste, 1968). English disyllabic words, on the other hand, have no such systematic duration patterns. In earlier experiments, we tested both Estonian and English listeners' ability to discriminate among pairs of tokens varying only in terms of their duration ratios. The tokens in the first study (Fox and Lehiste, 1987) represented sequences of white noise bursts bearing no resemblance to speech sounds. Synthesized speech sounds (versions of [babab]) were used in the second experiment (Fox and Lehiste, 1989). We expected that listeners speaking the two languages would react differently to the noise signals used in the first experiment; however, we found no significant difference between the two groups. We repeated the experiment with speech-like signals to test whether a difference would appear in a situation where the listeners may be expected to perform in a "speech mode" Contrary to expectations, we found no evidence for a difference in listener responses that could be attributed to the difference in language background. We also found no significant difference between results for speech and noise stimuli. Since Estonian speakers and listeners distinguish between the three disyllabic word types under normal speech conditions, Lehiste (1989) argued that a characteristic fundamental frequency (F_0) pattern associated with durational patterns contributes to the identification of the Estonian contrast. Such contrastive F_0 patterns are normally present in spoken Estonian, but they were absent from the synthetic speech stimuli used in these experiments.

The absence of F_0 cues for the distinction between disyllabic word types could reasonably confuse the listeners in a speech-like situation, but it does not explain the lack of positive three-way contrast perception by Estonian listeners when listening to nonspeech stimuli, since no expectation of a contrastive F_0 contour should exist there in the first place.

Eek (1987) carried out a comparison of word stress perception in Estonian and Russian. Estonian words are normally stressed on the first syllable of the word. The

second syllable may be stressed only in interjections, in sequences containing a proclitic, in compounds, and in foreign words with the structure of a compound. Both stressed and unstressed initial syllables and also an unstressed second syllable can be overlong. Russian, a stress language like English, has free stress that may fall on any syllable of a word. Eek modified the F_0 pattern and the syllable durations of a disyllabic nonsense word. The intensities of the two syllables were kept equal in some of the tests; in others, the intensity of the second syllable was lowered by 4 dB. Listening tests showed a significant difference in the perception of these stimuli by Estonian and Russian subjects. As Eek had hypothesized, the stress judgments of Estonians were mostly associated with F_0 patterns. A considerable F_0 difference failed to be the determining factor only when it was applied to a duration pattern that does not occur in Estonian words. For Russians, however, duration served as the leading parameter associated with the perception of stress; the effect of F_0 became noticeable only when the two syllables were of equal duration. Intensity played a negligible role for listeners with either language background.

Fry (1972) reports results of experiments in which native French listeners were asked to make stress judgments in response to the same set of stimuli that had been used earlier to explore stress perception by (British) English listeners. The stimuli consisted of a set of synthesized versions of five English word pairs, the noun and verb forms of *object, subject, digest, contract,* and *permit.* Duration and intensity of the two syllables were systematically varied (F_0 data are not reported). The listening tests showed that either additional duration or additional intensity were effective in influencing French listeners to perceive stress on one or the other syllable. For English listeners, the acoustic cue having the greatest weight in stress judgments was the relative duration of the vocalic portions, with relative intensity next in order of importance and the formant structure of the vowels third. Neither French nor English uses contrastive duration as a distinctive suprasegmental feature.

No clear pattern emerges from these results. For the perception of stress in English test words by English listeners, duration ranks above intensity in Fry's study; for the perception of stress in the same test words by French listeners, either one is sufficient; for the identification of the stressed syllable in disyllabic nonsense words by Russian listeners, duration outranks other suprasegmental cues. None of these languages uses duration for phonemic purposes, i.e., for distinguishing between two lexical items when all other factors are kept constant. Eek's study showed that speakers of a quantity language, Estonian, use F_0 as primary cue provided the durational pattern of the disyllabic test word corresponds to that of a possible Estonian word. Our studies of the perception of durational ratios by English and Estonian listeners (Fox and Lehiste, 1987, 1989) showed that there was no significant difference between these two groups in the perception of durational patterns that characterize Estonian words.

In a quantity language like Estonian, durational differences are phonemic in the basic sense of distinguishing between the lexical meanings of words. It appears reasonable to us to expect that the linguistic background of speakers has an influence on their perception of acoustic signals, and that speakers of languages in which duration plays a distinctive role are more sensitive to durational differences than speakers of a stress language like Russian or English, in which duration is not independently contrastive,

but serves as one of the phonetic characteristics of stressed syllables. As already mentioned, the results of both Eek's experiment and our experiments with durational ratios are not in accordance with this expectation. To explore further the question of possible influences of language background (here Estonian and English) on the perception of suprasegmental distinctions, we conducted two experiments that focused on the role of duration and intensity (excluding the role of pitch) in the perception of prominence of auditory stimuli.

We chose the term "prominence" in order to be able to apply it to both speech and non-speech stimuli. "Prominence" is the more inclusive term; "stress" is a subset of "prominence". The purpose of the experiments was, first of all, to establish the differences (if any) in the relative contribution of the two acoustic dimensions of duration and amplitude in perception of acoustic signals by listeners with differing language backgrounds. At this level, we wanted to avoid direct influence of knowledge of word-level stress patterns in either language; it had been shown by Eek that it makes a difference to Estonian listeners whether the suprasegmental structure of a stimulus resembles that of a possible Estonian word (F_0 cues were ineffective in words with non-Estonian durational patterns). The problem was solved by making the stimuli comparable in duration to monosyllabic words, and by using the term "prominence" rather than "stress" (and their Estonian equivalents) in instructions provided to the listeners. Both Estonian and English listeners heard a sequence of four tokens – the equivalent of four monosyllabic words – and judged which of these was most prominent, reacting to the duration and intensity cues without any necessary reference to word structure. In the first experiment we utilized sequences of four nonsense CV tokens which differed only in terms of overall length and amplitude; the test tokens used in the second experiment represented signal correlated noise versions of these CV tokens (perceived as noise signals bearing little resemblance to the original tokens).

The hypotheses tested were the following: (1) There exists a significant difference between the two groups of listeners in the perception of these stimuli; (2) speakers of a quantity language (here Estonian) are more sensitive to durational differences than speakers of a stress language (here English) in the perception of prominence; and (3) the influence of the linguistic background of the listeners carries over to their perception of nonspeech stimuli.

Experiment 1: Perception of Prominence in Speech Tokens

The main goal of the first experiment was to determine whether language background affects the extent to which listeners utilize duration or amplitude cues in making prominence judgments about CV tokens. In each experimental trial, listeners heard a sequence of four syllables. Within a sequence, all syllables could be of the same amplitude with one of the tokens longer in duration; one of the syllables could have increased length and amplitude (by 3 or 6 dB); or one of the syllables could be higher in amplitude while a different token was longer. Since both duration and amplitude contribute to perceived prominence, it was expected that simultaneous enhancement of duration and amplitude

TABLE 1

Outline of the different combinations of amplitude and duration variations present in the stimulus sequences across the four token positions. The same pattern of manipulations is present in the stimulus sets for both Experiments 1 and 2

Amplitude Variations		Equal Duration All Positions	Duration Increases (425, 450, 475, or 500 msec)			
			Pos. 1	Pos. 2	Pos. 3	Pos. 4
	Equal Amplitude All Pos.	1	4	4	4	4
Amplitude Increases (3 dB or 6 dB)	Pos. 1	2	8	8	8	8
	Pos. 2	2	8	8	8	8
	Pos. 3	2	8	8	8	8
	Pos. 4	2	8	8	8	8

Total different stimulus sequences = 153

on the same token would dramatically increase the number of times that token would be identified as most prominent.

Crucially important for determining the relative importance of duration and amplitude are trials in which these two cues provide conflicting information – that is, those cases in which one token is longer than the three other tokens, but a different token is the loudest of the four. The longer token in a sequence will be called hereafter the *target token.*

Methods

Stimuli. The stimulus sequences were generated from a basic CV monosyllable [ɓɑ] created with the Klatt cascade software synthesizer (Klatt, 1980). The basic token was 400 msec in duration with 40 msec formant transitions and a 360 msec steady state vowel, synthesized on a monotone F_0 of 115 Hz. The frequencies of F_1, F_2, and F_3 began at 500, 1020, and 2250 Hz, respectively, and changed linearly to 700, 1220, and 2600 Hz, respectively, at the end of the transition, where they remained for the rest of the duration of the token. The frequencies of F_4 and F_5 remained at 3300 and 4500 Hz, respectively, throughout the token. No attempt was made to synthesize a prevoiced closure prior to the formant transitions so phonetically the initial consonant could be characterized as an unvoiced stop with 0 msec voice onset time ([ɓ]).

During the experimental trials listeners heard four tokens in sequence with a constant silent interstimulus interval (ISI) of 150 msec. One token of each sequence could be

lengthened from the basic duration of 400 msec to 425, 450, 475, or 500 msec, and/or one token (not necessarily the same) could be increased in overall amplitude by 3 dB or 6 dB. These independent manipulations resulted in 153 different trials (as outlined in Table 1). Each of these trials occurred twice in random order for a total of 306 experimental trials.

Increases in duration were accomplished by lengthening only the steady state portion of the vowel. The duration of all [ɓɑ] tokens was within the range of possible durations of Estonian monosyllabic words, all of which are of overlong quantity. Thus the Estonian listeners were making prominence judgments rather than making phonemic decisions between long and overlong quantity. Between trial sequences there was an interval of 3.5 seconds.

Subjects. The subjects were 33 native speakers of English in Columbus, Ohio, and 40 native speakers of Estonian in Tallinn, Estonia. The English listeners participated to fulfil, in part, a requirement for an undergraduate course in Speech and Hearing Science. The Estonian listeners participated as paid volunteers. All were naive as to the experimental variables being manipulated and as to the goal of the experiment.

Procedure. On each trial a listener heard a sequence of four [ɓɑ] tokens and was required to indicate which of the four tokens was most "prominent" by circling one of four numbers on a prepared response form. Listeners were not instructed as to what criteria (e.g., loudness, duration) to use in their prominence judgments. Oral and written instructions were given in English in Columbus, and in Estonian in Tallinn.[1]

Results and discussion

As noted above, since both duration and amplitude contribute to perceived prominence, it was expected that simultaneous enhancement of duration and amplitude on the same token would dramatically increase the number of prominent responses to the target token. This expectation was confirmed. Shown in Figure 1 are the percentage of "prominent" responses to the longer syllable (the target token) when a different syllable in the sequence was increased in amplitude only (diff-3 dB and diff-6 dB) or when the longer syllable was simultaneously increased in amplitude (same-3 dB and same-6 dB). For both language groups, when a syllable had both longer duration and greater amplitude it was usually perceived as the most prominent token. However, the most crucial data for our purposes involve prominence judgment of those sequences in which the target token did not receive any amplitude increase (and may provide competition between duration and amplitude cues). Thus the following analyses do not include those data in which the target token was simultaneously increased in amplitude and duration.

In the initial analyses the data were not broken down by token position because each listener would then be contributing only two data points in each cell. Figure 2 shows the mean percentages of "prominent" responses to the target token in a trial as a function of its duration and of the amplitude of the other three tokens (all at 0 dB or one of the

[1] The terms used in Estonian-language instructions were "kõige silmapaistvam, kõige väljapaistvam" – words treated as synonyms in the Estonian–English Dictionary of Paul Saagpakk (Saagpakk, 1982) and glossed as "distinguished, eminent, outstanding, prominent, remarkable, noteworthy, notable, respectable, striking, conspicuous, salient, in evidence".

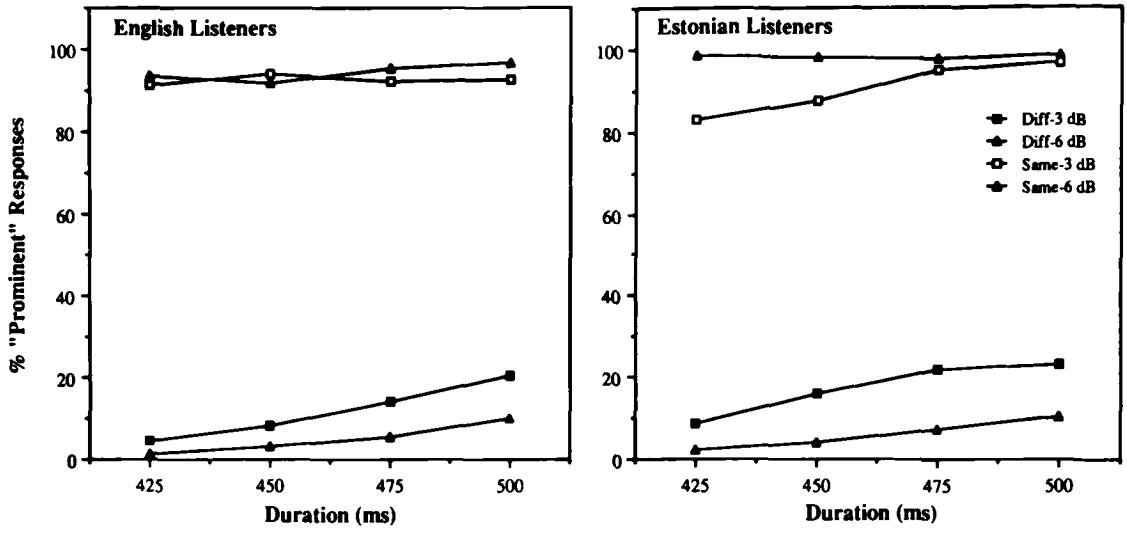

Fig. 1. Mean percentage of "prominent" responses to the longer tokens in Experiment 1. In each figure the open symbols show the responses to tokens increased in both amplitude and duration. Closed symbols show the responses to tokens increased in duration only when a different syllable in the sequence has been increased in amplitude.

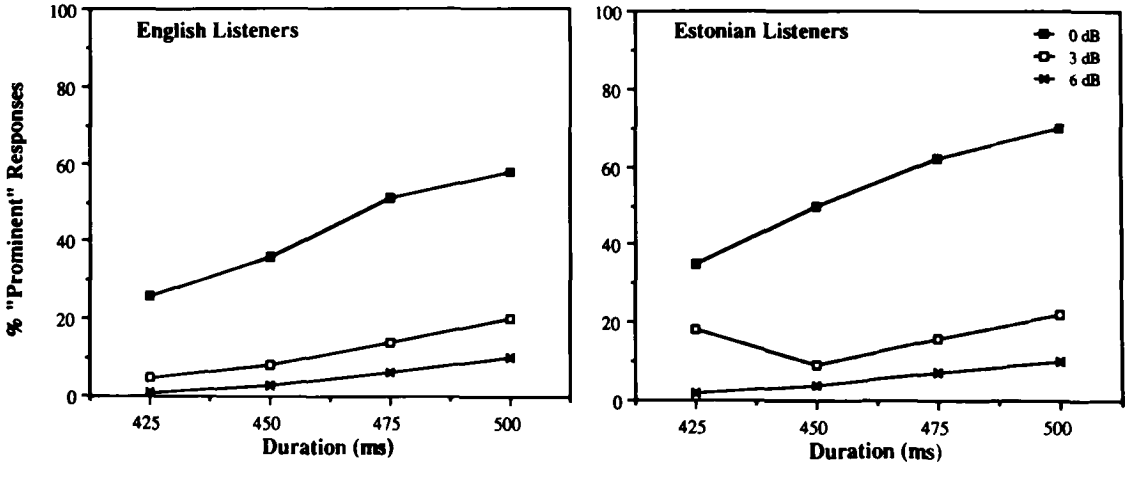

Fig. 2. Mean percentage of "prominent" responses to the target tokens in Experiment 1. Closed squares indicate responses when all tokens in a trial are presented with baseline (0 dB) amplitude. Open squares and crosses indicate responses when one of the non-target tokens in a trial sequence is presented 3 dB or 6 dB greater in amplitude, respectively, than the other three tokens in the sequence.

three at 3 or 6 dB), separately for the two listener groups. These data were analyzed using a three-way, repeated-measures analysis of variance with the between-subject factor language group and the within-subject factors target duration and non-target amplitude.[2]

As can be seen from the figure, when amplitude was kept constant, increases in duration resulted in gradually increasing numbers of prominence judgments for both groups of listeners. Presence of increased amplitude on one of the nontarget tokens reduced identification of the longer token as "prominent" by a significant amount [$F(2, 142) = 431.2$, $p < 0.001$]. When all tokens in a trial were presented at the baseline amplitude of 0 dB, the longer token was identified as "prominent" 48.9% of the time, on the average. (Unless separately identified, the averages here and in the rest of the paper refer to combined results for all listeners.) However, when a nontarget token had greater amplitude – either 3 dB or 6 dB – this average percentage dropped to 14.8 and 5.5%, respectively. (In these sequences, the higher amplitude nontarget tokens received the majority of the "prominent" responses.)

The pattern of prominence responses was significantly affected by the native language of the listeners [$F(1, 71) = 7.03$, $p < 0.01$], Estonian listeners identifying the longer token as "prominent" 25.9% of the time, *vs.* 19.9% for English listeners. This result could be interpreted as showing that duration is a somewhat more salient feature of "prominence" for Estonian listeners than for English listeners.

Perceived prominence was also significantly increased with the duration of the target token [$F(3, 213) = 65.1$, $p < 0.001$], for both groups of listeners.

A significant language by amplitude interaction [$F(2, 142) = 8.97$, $p < 0.001$] was obtained because the Estonian listeners were different from the English listeners at 0 dB and 3 dB (but not in the 6 dB condition; this may be attributed to a floor effect).

An amplitude by duration interaction [$F(6, 426) = 27.04$, $p < 0.001$] was obtained because the effect of target token duration was greatest when all tokens in the sequence had the same amplitude (0 dB). The duration effect was reduced in both the 3 dB and 6 dB conditions. This demonstrates that neither group of listeners was depending exclusively upon the duration cue (ignoring token amplitude) in making the prominence decision. Conversely, the residual effect of duration in the 3 dB and 6 dB condition shows that the amplitude cue was not completely dominant either.

Neither the language by duration interaction [$F(3, 213) = 0.53$] nor the amplitude by duration by language interaction [$F(6, 426) = 0.73$] were significant. As shown in Figure 2, the basic pattern of responses across amplitudes and durations was very similar for both groups. However, the higher sensitivity of Estonian listeners to duration increments is evident even in the 425 msec/0 dB condition, where they (but not the English listeners) performed above chance level (25%).

---

[2] All percentage data were arcsine transformed prior to statistical analysis to avoid well-known problems related to proportional data (see Studebaker, 1985).

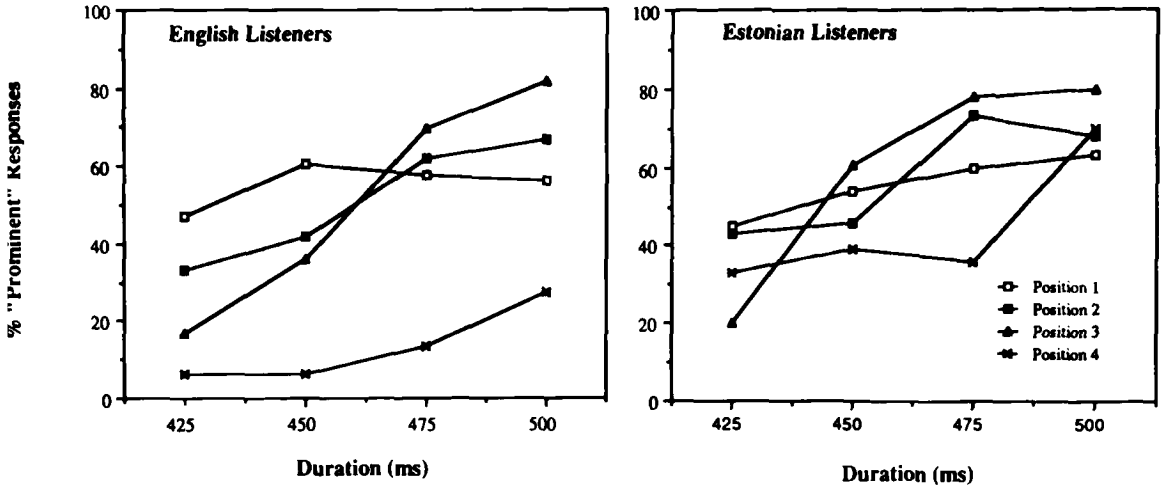

Fig. 3. Mean percentage of "prominent" responses to the longer target tokens (CVs) in Experiment 1 when all tokens in a trial are presented with baseline (0 dB) amplitude, broken down by position of the target token in the trial sequence.

Figure 3 shows results from the 0 dB condition for both language groups broken down by the position of the target token in the trial sequence. To examine response differences due to target position, these data were analyzed using a three-way, repeated-measures analysis of variance with the between-subject factor language group and the within-subject factors target position and target duration.

There were significantly fewer "prominent" responses [$F(3, 213) = 26.12$, $p < 0.001$] to target tokens appearing in final position. Target tokens in position 1, 2, and 3 were identified as "prominent" 55.3%, 54.5%, and 55.8% of the time, respectively, but target tokens in final position were only identified as "prominent" 30.3% of the time. However, a significant language group by target position interaction [$F(3, 213) = 6.34$, $p < 0.001$] demonstrated that English listeners showed significantly fewer "prominent" responses in final position (55.3%, 51.1%, 51.1%, and 13.3% for positions 1–4, respectively) than Estonian listeners (55.3%, 57.2%, 59.7%, and 44.4%, for positions 1–4, respectively).

There was also a significant target position by target duration effect [$F(9, 639) = 9.38$, $p < 0.001$] showing that the effects of position were not the same across all different target durations. For example, although the percentage of "prominent" responses increased steadily as target duration increased (to 425, 450, 475, and 500 msec) for targets in position 2 (38.4%, 44.5%, 67.8% and 67.2%, respectively) and position 3 (18.5%, 50.0%, 74.0%, and 80.8%, respectively), there was very little increase when the target occurred in initial position (45.9%, 56.9%, 58.9%, and 59.6%, respectively). The precise interaction between target position and target duration was somewhat different between the two language groups, which led to a marginally significant language group by target position by target duration interaction [$F(9, 639) = 2.27$, $p < 0.05$].

One should note that this study was not primarily directed at positional differences, and that each listener contributed only two responses for each combination of target position, target duration, and amplitude. However, it seems clear that the responses of English listeners, in particular, were affected by positional variations. A possible explanation for these results might be the relatively greater importance of preboundary lengthening in English, as compared to Estonian (for English, *cf.* Lehiste, 1979, and the discussion below).

EXPERIMENT 2: PERCEPTION OF PROMINENCE IN NONSPEECH TOKENS

The pattern of responses obtained in Experiment 1 is consistent with the hypothesis that a listener's native language can affect his/her perception of relative prominence in a set of synthetic speech tokens. Experiment 2 was designed to determine whether the effect of native language is limited to the perception of speech tokens or whether it extends to nonspeech as well. If it does not, the results of the experiment confirm the existence of a difference between listening in a "speech mode" *vs.* listening to nonspeech psychoacoustic stimuli; if it does indeed carry over from listening to speechlike stimuli to listening to nonspeech stimuli, reconsideration of some earlier generalizations concerning perception is called for.

Methods

Stimuli. The [ɓɑ] tokens used in Experiment 1 were converted to nonspeech noise tokens by randomly (50% probability) changing the polarization of individual digital samples. This created a set of signal-correlated noise tokens having the same amplitude envelope and duration as the speech tokens, but without identifiable pitch or formants. Except for this change in stimuli, the design and procedure were exactly the same as in Experiment 1.

Subjects. The subjects were 24 native speakers of English in Columbus, Ohio, and 40 native speakers of Estonian in Tallinn, Estonia. None of them had participated in Experiment 1.

Results and discussion

As in Experiment 1, there were many more "prominent" responses to the target token when it was increased in both duration and amplitude simultaneously than when it was increased in duration only (and a different syllable had an amplitude increase of 3 or 6 dB). Although not shown, the obtained pattern of responses was practically identical to Figure 1. As before, however, the crucial data do not involve judgment of sequences in which both duration and amplitude increases occur on the target token, and those data were eliminated from the remaining analyses.

Figure 4 shows the results from Experiment 2 pooled across positions. The format of the figure is analogous to that of Figure 2.

As expected, there was a significant main effect of amplitude [$F(2, 124) = 254.5$,

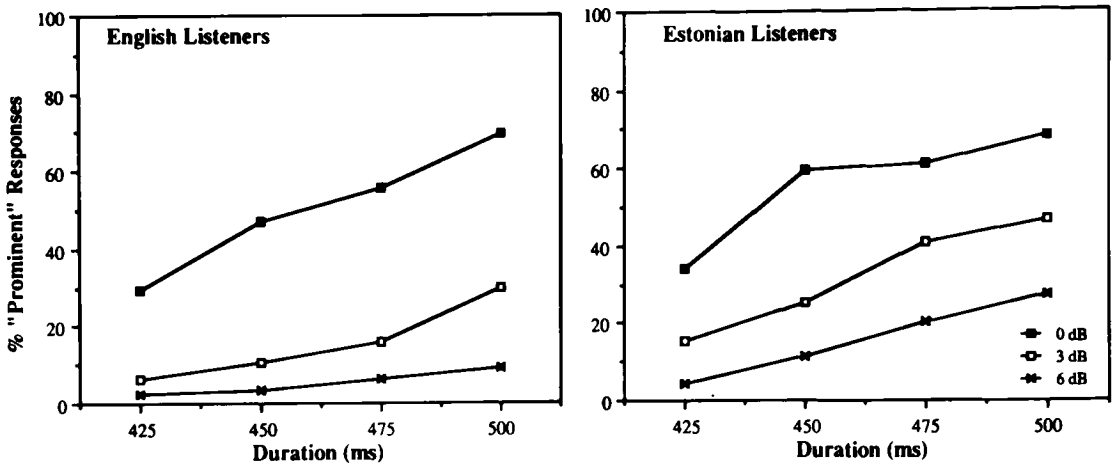

Fig. 4. Mean percentage of "prominent" responses to the target tokens in Experiment 2. Closed squares indicate responses when all tokens in a trial are presented with baseline (0 dB) amplitude. Open squares and crosses indicate responses when one of the non-target tokens in a trial sequence is presented 3 dB or 6 dB greater in amplitude, respectively, than the other three tokens in the sequence.

$p < 0.001$]. When all tokens in a trial were 0 dB, the longer target token was identified as "prominent" 53.7% of the time. However, when one token had a greater amplitude – either 3 dB or 6 dB – this percentage dropped to 25.9% and 11.8%, respectively.

Native language of the listener again produced a significant effect [$F(1, 652) = 11.9$, $p < 0.001$]: Estonian listeners identified the longer token as "prominent" 34.5% of the time, English listeners only 23.7% of the time. Unlike Experiment 1, the difference between the two groups was evident in all three amplitude conditions, and particularly in the 3 and 6 dB conditions. The language by amplitude interaction was nonsignificant [$F(2, 124) = 2.2$, $p > 0.10$]. As the duration of the target token increased, it was more likely to be identified as "prominent" [$F(3, 186) = 65.1$, $p < 0.001$]. The language by duration interaction was nonsignificant [$F(3, 186) = 2.1$, $p > 0.09$]. However, there was a significant amplitude by duration interaction [$F(6, 372) = 13.6$, $p < 0.001$] because the increase in the number of "prominent" responses with duration was affected by whether or not a louder token occurred at some other position in the trial sequence. Thus it is again clear that the listeners were not using the duration cue exclusively to make their prominence decision. There was also a significant amplitude by duration by language interaction [$F(6, 372) = 3.43$, $p < 0.003$], because the amplitude by duration interaction was more pronounced for the English group.

Figure 5 shows the results from the 0 dB condition for both language groups broken down by the position of the target token in the trial sequence. As above, in order to examine effects of target position these data were analyzed using a three-way, repeated-measures analysis of variance with the between subject factor language group and the within-subject factors target position and target duration.

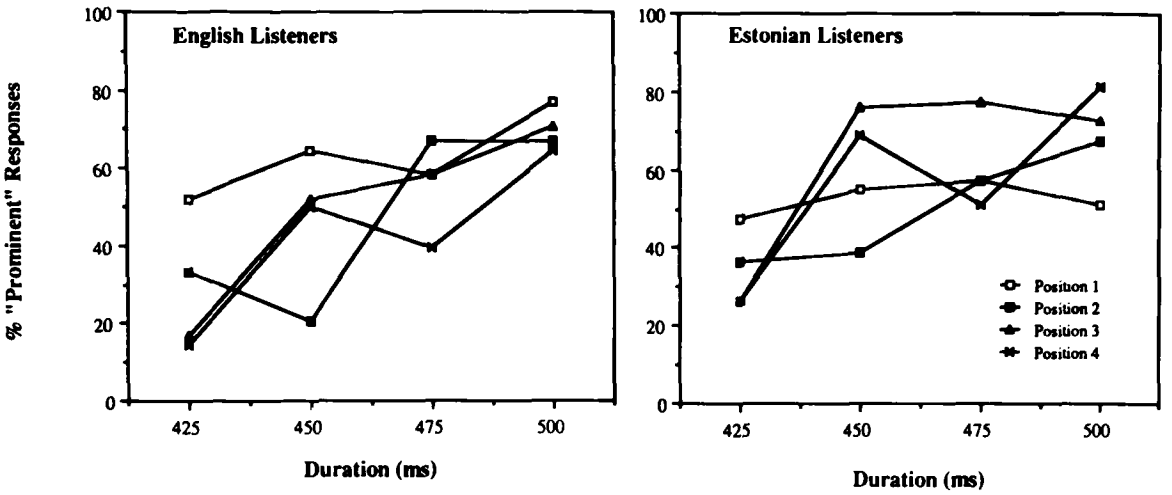

Fig. 5. Mean percentage of "prominent" responses to the target tokens in Experiment 2 when all tokens in a trial are presented with baseline (0 dB) amplitude, broken down by position of the target token in the trial sequence.

There was a significant effect of target position [$F(3, 186) = 3.00$, $p < 0.05$], but in these data the final position did not produce the fewest number of "prominent" responses. In particular, the percentages of "prominent" responses for positions 1–4 were 56.6%, 48.8%, 58.0%, and 51.4%, respectively. However, the effect of target position on the responses from the two language groups was different. In particular, there was a significant language group by target position interaction. For the English listeners, final position did produce the fewest "prominent" responses (63.0%, 46.9%, 49.5%, and 42.2% for positions 1–4, respectively) but it did not for the Estonian listeners (52.8%, 50.0%, 63.1%, and 56.9%, respectively).

As in Experiment 1, there was a significant target position by target duration interaction [$F(9, 558) = 8.59$, $p < 0.001$]. Again, when target tokens occurred in initial position, increases in target duration (to 425, 450, 475, or 500 msec) produced little increase in the percentage of "prominent" responses (49.2%, 58.6%, 57.8%, and 60.9%, respectively). Target tokens appearing in position 3 also produced little increase in the percentage of "prominent" responses in the three longest target durations (21.9%, 67.2%, 70.3%, and 71.9%). Targets occurring in position 2 showed increased "prominent" responses only in the two longest target durations (35.2%, 32.0%, 61.0%, and 67.2%). Targets in position 4 showed a gradual increase in "prominent" responses with increased duration (21.9%, 61.9%, 46.9%, and 75.0%) except that more "prominent" responses were obtained with a duration of 475 msec than at 450 msec.

In general, there were fewer "prominent" responses to target tokens appearing in position 1. A tentative explanation for this result may be that the first element in the sequence establishes a base against which other elements are evaluated. In order to hear

the first token as prominent, the listeners must listen to the entire sequence to make sure that these later items can be eliminated from the competition.

Unlike Experiment 1, the language group by target position by target duration interaction was not significant [$F(9, 558) = 1.33, p > 0.2$).

This result suggests also that instructions given to listeners may have an effect on their performance. In an earlier study (Lehiste, 1979), listeners had been asked to identify the "longest" token; in the current study, they were asked to identify the token that is "most prominent". In the 1979 study, the fourth (noise) token required greater lengthening than tokens occurring in the other three positions in order to be perceived as "longest" by English-speaking listeners.

Comparison of Speech and Nonspeech Data

To determine whether listener performance differed when the tokens were speech-like rather than noise, the data from Experiments 1 and 2 were combined and subjected to a mixed analysis of variance with the additional between-subject factor of stimulus type (noise and [ɓɑ] tokens).

Not unexpectedly, there were significant main effects of language, amplitude, and duration. However, of greater interest to the present study was the fact that there was a significant stimulus type by language by amplitude interaction [$F(2, 266) = 6.85$, $p < 0.001$] and a significant stimulus type by language by amplitude by duration interaction [$F(6, 798) = 3.40$, $p < 0.003$]. The Estonian listeners produced about the same number of "prominent" responses in the 0 dB condition for both the noise and [ɓɑ] tokens (55.7 and 54.4%, respectively). However, the number of "prominent" responses to the 3 dB and 6 dB conditions dropped more with the [ɓɑ] tokens (17.4 and 5.9%, respectively) than with the noise tokens (32.1 and 15.8%, respectively). For English listeners the decrease in the number of "prominent" responses across amplitude conditions was similar for the [ɓɑ] and noise tokens.

General Discussion

The study started with three hypotheses. The first concerned the possible difference between listeners who are native speakers of either a stress language or a quantity language in making prominence judgments on the basis of duration and amplitude cues (in the absence of F_0 cues). Our findings show significant differences between English and Estonian listeners in this respect. Specifically, Estonian listeners identified a longer token as "prominent" more frequently than did English listeners. This was particularly striking when stimuli occurred in final position in a speech sequence: English listeners required a much larger increase in duration than Estonian listeners in order to perceive the final token as being more prominent. It may be that different degrees of preboundary lengthening are used in the two languages to signal completion of an utterance (for Estonian, *cf.* Lehiste, 1981; for English, *cf.* Klatt, 1976; for recent cross-linguistic

comparisons, *cf.* Berkovits, 1991; Fletcher, 1991). Alternatively or additionally, it may be that speakers of a quantity language are simply more sensitive to differences in duration.

The second hypothesis related to the relative importance of duration and amplitude in prominence judgments, claiming more specifically that speakers of a quantity language (Estonian) are more sensitive to durational differences than speakers of a stress language (English) in the perception of prominence. Here it appears that listeners of both groups used both cues, but the two groups differed in the relative importance of the cues under certain conditions. When speech tokens all had the same amplitude, Estonian listeners were more sensitive to duration, particularly in final position.

A further difference between the two groups of listeners emerged in the interaction of duration and amplitude cues. Estonian listeners reacted to the difference in amplitude cues by responding differently to duration increases in the zero (baseline amplitude) condition *vs.* amplitude-increased conditions. In the 0 dB condition, they produced the same number of "prominent" responses for the noise and the [ɓɑ] tokens, but when another token had greater amplitude, the "prominent" responses dropped considerably more for the [ɓɑ] tokens than for the noise tokens. English listeners had the same decrease across amplitude conditions for [ɓɑ] and noise tokens. (We have no immediate suggestion as to the possible cause for this difference.)

The third hypothesis concerned the influence of the linguistic background of the listeners on their perception of the two kinds of signals. In setting up the hypothesis, we started with the preliminary assumption that if such an influence is present, it is more likely to be found in listener responses to speech signals; if the two groups perform differently in responding to noise signals as well, a stronger form of the hypothesis would be supported. Furthermore, different performance on the two types of stimuli might provide additional confirmation of the special status of speech.

As has been described above, the results do indeed support the stronger form of the hypothesis. Significant differences were also observed between the two types of stimuli.

Both listener groups gave more "prominent" responses to the longer noise tokens than to the longer [ɓɑ] tokens. We interpret this to mean that absence of F_0 cues has a greater influence on the perception of speechlike signals (when a constant F_0 is present) than nonspeech signals (where F_0 is absent). We suggest that in order to be perceived as prominent, speechlike signals raise an expectation in the listener that F_0, too, will change. As the expectation was not met, the [ɓɑ] stimuli received lower scores than otherwise similar noise stimuli (similar ideas are discussed in Bentin and Mann, 1991).

The results of these experiments thus support the contention that the linguistic background of listeners influences their perception of prominence – particularly in terms of the utilization of the cues of amplitude and duration. This effect was observed in both speechlike stimuli and noise stimuli. In addition, speechlike stimuli and noise stimuli were reacted to by listeners in significantly different ways depending on their linguistic background. Our results suggest strongly that speech experience has an effect on general auditory perception. Now, most discussions of perception appear to take it for granted that the human auditory mechanism can be studied without taking into account the previous linguistic experience of the subjects; we suggest caution in drawing

generalizations from listening tests in the design of which the linguistic background of the subjects has not been taken into account.

This study was designed as an experiment in the perception of prominence cued by two of the suprasegmental features that also serve as cues to the presence of linguistic stress. To the extent possible, we avoided prompting listeners to make linguistic judgments about stress. Word-level stress is not contrastive in Estonian in the manner in which it is in English; duration does not play the same phonological role in English as it does in Estonian. The nature of the stimuli was such that what we termed "prominence" might be considered the equivalent of sentence-level stress or emphasis – the type of stress that does not have a phonological function at the word level. The results of the experiments nevertheless show that elements of word-level phonological structure enter into perception of utterance-level prosody. Results such as emerged from this study may find practical application in teaching the prosody of a second language to speakers of languages with differing suprasegmental structures (*cf.* Mochizuki-Sudo and Kiritani, 1991, for an example of problems that arise in teaching English stress to Japanese learners of English as a second language).

(Received January 3, 1992; accepted August 13, 1992)

REFERENCES

BENTIN, S., and MANN, V.A. (1991). Masking and stimulus intensity effects on duplex perception: A confirmation of the dissociation between speech and nonspeech modes. *Journal of the Acoustical Society of America*, **88**, 64–74.

BERKOVITS, R. (1991). The effect of speaking rate on evidence for utterance-final lengthening. *Phonetica*, **48**, 57–66.

BLADON, R.A.W., and LINDBLOM, B. (1981). Modeling the judgment of vowel quality differences. *Journal of the Acoustical Society of America*, **69**, 1414–1422.

CARAMAZZA, A., YENI-KOMSHIAN, G., ZURIF, E., and CARBONE, E. (1973). The acquisition of a new phonological contrast: The case of stop consonants in French-English bilinguals. *Journal of the Acoustical Society of America*, **54**, 421–428.

EEK, A. (1987). The perception of word stress: A comparison of Estonian and Russian. In R. Channon and L. Shockey (eds.), *In Honor of Ilse Lehiste* (pp. 19–32). Dordrecht: Foris Publications.

FLEGE, J.E., MUNRO, M.J., and FOX, R.A. (1991). The interlingual identification of Spanish and English vowels. *Journal of the Acoustical Society of America*, **90**, Pt. 2, 2252 (Abstract).

FLETCHER, J. (1991). Rhythm and final lengthening in French. *Journal of Phonetics*, **19**, 193–212.

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

FOX, R.A., and LEHISTE, I. (1989). Discrimination of duration ratios in bisyllabic tokens by native English and Estonian listeners. *Journal of Phonetics*, **17**, 167–174.

FRY, D.B. (1972). French and the tonic accent. In A. Valdman (ed.), *Papers in Linguistics and Phonetics to the Memory of Pierre Delattre* (pp. 201–208). s'Gravenhage: Mouton.

KLATT, D.H. (1976). Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *Journal of the Acoustical Society of America*, **59**, 1208–1221.

KLATT, D.H. (1980). Software for a cascade/parallel formant synthesizer. *Journal of the Acoustical Society of America*, **67**, 979–995.

LEHISTE, I. (1968). Vowel quantity in word and utterance in Estonian. *Congressus Secundus Internationalis Fenno-Ugristarum* (pp. 293–303). Helsinki: Societas Finno-Ugrica.

LEHISTE, I. (1979). Perception of duration in sequences of four intervals. *Journal of Phonetics*, **7**, 313–316.

LEHISTE, I. (1981). Sentence and paragraph boundaries in Estonian. *Congressus Quintus Internationalis Fenno-Ugristarum* (pp. 164–169). Turku: Suomen Kielen Seura.

LEHISTE, I. (1989). Current debates concerning Estonian quantity. In J. Nevis (ed.), *FUSAC '88: Proceedings of the Sixth Annual Meeting of the Finno-Ugric Studies Association of Canada* (pp. 77–86). Lanham, MD: University Press of America.

MOCHIZUKI-SUDO, M., and KIRITANI, S. (1991). Production and perception of stress-related durational patterns in Japanese learners of English. *Journal of Phonetics*, **19**, 231–248.

SAAGPAKK, P. (1982). *Estonian-English Dictionary*. New Haven: Yale University Press.

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

WILLIAMS, L. (1977). The perception of stop consonant voicing by Spanish-English bilinguals. *Perception & Psychophysics*, **21**, 289–297.