Spectral balance as a cue in the perception of linguistic stress

Agaath M. C. Sluijter,[a] Vincent J. van Heuven,[b] and Jos J. A. Pacilly[c]
Holland Institute of Generative Linguistics, Phonetics Laboratory, Leiden University, Cleveringaplaats 1, P.O. Box 9515, 2300 RA Leiden, The Netherlands

(Received 28 March 1995; revised 1 August 1996; accepted 2 August 1996)

In this study, the claim that intensity, as an acoustic operationalization of loudness, is a weak cue in the perception of linguistic stress is reconsidered. This claim is based on perception experiments in which loudness was varied in a naive way: All parts of the spectrum were amplified uniformly, i.e., loudness was implemented as intensity or gain. In an earlier study it was found that if a speaker produces stressed syllables in natural speech, higher frequencies increase more than lower frequencies. Varying loudness in this way would therefore be more realistic, and should bring its true cue value to the surface. Results of a perception experiment bear out that realistic intensity level manipulations (i.e., concentrated in the higher frequency bands) provide stronger stress cues than uniformly distributed intensity differences, and are close in strength to duration differences. [S0001-4966(97)00412-8]

PACS numbers: 43.71.Es, 43.70.Fq [RAF]

INTRODUCTION

Dutch and English are languages with word stress: one of the syllables of a word, especially when pronounced in citation form, is perceived as the most prominent one, the so-called lexical stress position of the word. The phonetic correlates of lexical stress in these languages are pitch, duration, loudness, and vowel quality (Lehiste, 1970; Beckman 1986, and references mentioned there). Of these, pitch and duration have been found the most important perceptual cues; intensity, as an acoustical operationalization of loudness, is generally claimed to be of lesser importance (among others: Fry, 1955, 1958; van Katwijk, 1974), while vowel quality is the least important cue (Fry, 1965; Rietveld and Koopmans-van Beinum, 1987). When words are spoken outside focus, i.e., without a pitch accent on the stressed syllable, the position of the stress has to be inferred from the remaining cues such as duration and intensity.

In the older linguistic and phonetic literature it was generally held that languages such as English and Dutch are characterized by so-called dynamic (rather than melodic) stress. That is to say, stressed syllables are produced with greater pulmonary and glottal effort, with greater loudness as the primary perceptual correlate (Sweet, 1906; Bloomfield, 1933). With the advent of speech synthesis techniques in the fifties this view was quickly discredited, when manipulating intensity (i.e., gain), as an operationalization of loudness variation, proved virtually inconsequential for stress perception (Fry, 1955, 1958 for English; Mol and Uhlenbeck, 1956 for Dutch; Issatchenko and Schädlich, 1966 for German).

In the present study, the claim that loudness is a weak cue in the perception of linguistic stress is reconsidered. Recently, Sluijter and van Heuven (1996) showed that intensity level differences between stressed and unstressed Dutch syllables are concentrated in the higher parts of the spectrum, whereas intensity differences in the lower part of the spectrum, i.e., below 500 Hz, were negligible. We assume that these differences in the higher parts of the spectrum are caused by a difference in the shape of the glottal waveform, due to an increase in vocal effort when producing stressed syllables, and are therefore a reflection of effort, and are perceived in terms of greater loudness.

The assumption that vocal effort is related to the perception of loudness was explored by Brandt *et al.* (1969). They independently varied vocal effort and intensity of continuous speech stimuli. In their experiments speech samples that were produced with greater effort, were estimated as louder than the same samples spoken with less effort, even when the mean intensity was adjusted so as to be constant. They considered the acoustic spectrum to be a special cue for the perception of vocal effort. Glave and Rietveld (1975) also examined the role of effort in speech loudness; their results confirmed that greater vocal effort is related to greater perceived loudness. Furthermore, they showed that the spectra of vowels spoken with greater effort have more intensity in the higher-frequency region, which they assumed to be caused by the changes in the source spectrum due to a more pulse-like shape of the glottal waveform.

This operationalization of loudness variation, i.e., increasing intensity in the higher frequency bands only, differs substantially from implementing loudness in terms of changing the gain factor uniformly across the spectrum as was done in the perceptual experiments above. Therefore, varying the acoustical correlate of loudness in a more realistic way, i.e., by varying the spectral balance,[1] should bring out the true cue value of loudness for stress perception.

If, indeed, varying intensity level in the higher frequency bands only is a perceptually more effective stress cue than applying uniform intensity level increments, a second question arises: What is the importance of the loudness cue relative to other stress cues? In order to keep this second question within manageable proportions, we will examine

[a]Now at KPN Research, P.O. Box 421, 2260 AD Leidschendam, The Netherlands. Electronic mail: a.m.c.sluijter@research.kpn.com
[b]Electronic mail: heuven@rullet.leidenuniv.nl
[c]Electronic mail: pacilly@rullet.leidenuniv.nl

the importance of intensity level manipulations relative to that of duration manipulation, i.e., the cue that has been advanced as the most reliable stress cue so far.

It is not the intention of the present study to question the primacy of the $F0$ cue in stress perception, since we regard $F0$ movement as a cue for sentence accent rather than for linguistic word stress. There is ample evidence, e.g., in Dutch, that an $F0$ movement with the appropriate excursion size ($\geqslant$4 semitones) and time alignment (cf. 't Hart *et al.*, 1990; Hermes and Rump, 1993) is a sufficient cue for accent, and *a fortiori* for stress, since accents are normally associated with the lexically stressed syllable of a word. In fact, when the accent is shifted to a nonstressed syllable so as to signal a metalinguistic contrast as in *I said SUGgest not DIgest,*[2] the original stress cues in the second syllable of *suggest* are almost completely obliterated and transferred to the initial syllable, cf. Sluijter and van Heuven (1995). However, the $F0$ cues are not invariant stress cues, since they disappear at the sentence level when the word is deaccented through focus manipulation (cf. van Heuven, 1987; Sluijter and van Heuven, 1996). Formant changes, finally, have consistently been reported as the least important cue for word stress (and sentence accent).

We will therefore examine the relative strength of the two implementations of loudness and duration in unaccented, i.e., nonfocused, targets.

In the experiment described below we studied the perception of stress position in the disyllabic Dutch nonsense word *nana* by manipulating vowel duration, spectral balance (intensity level increments in the higher frequency bands only) and intensity (uniformly distributed gain increments) in accordance with our production data (Sluijter and van Heuven, 1996). The hypothesis to be tested is that spectral balance is a stronger stress cue than overall intensity, and that the importance of spectral balance as a stress cue will approximate (or even surpass) that of duration. The possible finding that more realistic loudness manipulations provide a stronger stress cue than the traditional operationalization of loudness as gain/intensity should then, at least in part, rehabilitate the claim of the above mentioned older literature by Sweet (1906) and Bloomfield (1933).

I. PERCEPTION EXPERIMENT I

A. Methods

1. Material

We used the reiterant nonsense word pair /ˈnaːnaː/-/naːˈnaː/. This type of speech allows us to vary duration, spectral balance and intensity without taking into account segmental differences between both syllables, e.g., differences in intrinsic duration (Peterson and Lehiste, 1960) and intrinsic intensity (Lehiste and Peterson, 1959) of vowels, and possible perceptual compensation for these features. Reiterant speech was also used by Morton and Jassem (1965), van Katwijk (1974), Berinstein (1979) and many others in similar experiments and is assumed to be like nonreiterant speech in all aspects which are important in the study of prosody (Larkey, 1982).

TABLE I. Overview of the duration manipulations yielding seven duration steps. Durations are given (in ms) for first ($\sigma1$) and second syllable ($\sigma2$) separately, as well as total word duration ($\sigma1+\sigma2$).

Duration	$\sigma1$	$\sigma2$	$\sigma1+\sigma2$
1 *NAna*	250	185	435
2	230	200	430
3	210	215	425
4 *neutral stimulus*	190	230	420
5	170	245	415
6	150	260	410
7 *naNA*	130	275	405

We used the unstressed syllable *na* of the sentence *Wil je naˈna zeggen* /vɪ| jə naːnaː zɛɣə/ 'Will you [naˈna] say,' uttered by a male speaker with a pitch movement on *zeggen*, taken from the production study. This speaker was chosen out of a set of ten because the quality of his voice was preserved best in LPC resynthesis in comparison with the other male and female speakers.

We concatenated two syllables *na* to form the disyllabic nonsense word *nana*. The duration of the syllables was varied in seven steps from *ˈnana* to *naˈna* in accordance with our production data (Sluijter and van Heuven, 1996). We took a representative duration range for reiterant speech averaged over the speakers. This led to the following experimental values: the initial syllable was varied in seven steps of 20 ms from 250 to 130 ms, the second syllable was varied in seven steps of 15 ms from 185 to 275 ms. Note that an increase of the duration of the first syllable covaries with a decrease of the duration of the second syllable. The stimulus with an initial syllable of 190 ms and a final syllable of 230 ms (number 4) was meant to be temporally ambiguous for stress perception. The longer average duration of the second syllable was copied from actual speech production so as to reflect the influence of word-final lengthening (Wightman *et al.*, 1992; Sluijter and van Heuven, 1996). Table I gives an overview of the resulting stimuli.

In order to reduce the dimensionality of the stimulus space, we implemented spectral balance in terms of variable intensity levels below and above 0.5 kHz. It appeared from our production data (Sluijter and van Heuven, 1996) that the intensity levels in the three octave bands (B2–B4) were correlated (r^2 between 0.45 and 0.57), whereas there was no correlation between the base band B1, and any of the higher octaves (r^2 between 0.04 and 0.23). The spectral balance of the syllables was therefore varied by increasing the levels of the frequency components above 500 Hz by 3, 6, or 9 dB, in either the initial or the final syllable. We used the digital filtering facilities of the speech and signal processing package XAudlab (Lagendijk, 1992) implemented on a Silicon Graphics Indigo/Irix computer. The filtering and filter design algorithms implemented in this package use the standard FIR structure and DFT approach. The spectral balance steps were a straightforward quantization of the differences between the stressed and unstressed realizations of the syllables *na* in our production study. We applied uniform intensity level increments to all the frequencies above 500 Hz, although strictly speaking the intensity differences in the third filter band

TABLE II. In the left-hand part of the table the intensity level manipulations per step are presented. Levels were increased for components above 500 Hz. These manipulations caused overall intensity level increases of the syllables, which are presented in the right part of the table. These values were used to vary intensity level uniformly in all bands.

	Increased levels above 500 Hz		Increase in overall intensity level (incl. baseband)	
Step	$\sigma 1$	$\sigma 2$	$\sigma 1$	$\sigma 2$
1	+9 dB	⋯	+3 dB	⋯
2	+6 dB	⋯	+2 dB	⋯
3	+3 dB	⋯	+1 dB	⋯
4	⋯	⋯	⋯	⋯
5	⋯	+3 dB	⋯	+1 dB
6	⋯	+6 dB	⋯	+2 dB
7	⋯	+9 dB	⋯	+3 dB

(1.0–2.0 kHz) should be a little larger than those in the second (0.5–1.0 kHz) and fourth (2.0–4.0 kHz) filter bands. Crucially, however, we did not add any intensity to the base band.

Larger differences than the 9-dB increase in the higher bands occur occasionally in our production data, but this value was chosen as the maximum increment as stimuli with larger intensity level differences in the higher bands sounded less than acceptable.

These vocal effort/spectral balance manipulations yielded overall intensity level changes of approximately 1, 2, or 3 dB, respectively. Consequently, these steps were used to vary overall intensity level. Overall intensity level was varied by simply multiplying the sample values of either the initial or the final syllable by 1.12, 1.26, and 1.41, respectively. Table II gives an overview of the manipulations.

As can be seen in Table II, the overall intensity level differences in both stimulus sets are identical. There are seven duration levels, seven intensity levels, and two implementation methods. This nominally yields 98 stimuli but there were only 91 in practice since stimuli with the neutral intensity level (i.e., step 4) are identical for the two methods. The first part *Wil je, nana*, and the last part of the sentence *zeggen* were concatenated and resynthesized using straightforward LPC synthesis. As a consequence spectral discontinuities were smoothed over a window length of 25 ms. A sample frequency of 10-kHz, 4.5-kHz low-pass filter and 12-bit amplitude resolution were used for both analysis and resynthesis (18 reflection coefficients, Hamming window length 25.6 ms, window shift 10 ms).

Stimuli were presented without a pitch movement on the target in a fixed carrier phrase *Wil je [target] zeggen* (Will you [target] say). The carrier sentence was synthesized with a declining pitch contour, modeled after the pitch contour of the original sentence, such that the target was part of a falling declination line. An accent-lending pitch movement was realized on the first syllable of *zeggen*. The targets were presented in their original context since presenting stimuli out of their original context induces strong perceptual bias to perceive the stress on the first syllable (van Heuven and Menert, 1996). The prefinal position in the sentence was originally chosen to avoid preboundary lengthening in the targets; in the present experiment it is therefore necessary to avoid perceptual compensation for preboundary lengthening by maintaining this position.

2. Subjects and procedure

One stimulus tape was prepared containing the 91 stimuli in two different random orders. The 182 stimuli were presented in blocks of 13 utterances with 2-s intervals between utterances, offset to onset, and a larger interval and a 500-ms tone of 1000 Hz separating the blocks. This was done to prevent subjects from losing their way on the answer sheet, and to give them time to turn the pages of their answering booklet. The tape started with five practice utterances to familiarize the subjects with their task. Forty-six listeners participated in the test. Twenty-four subjects (phonetically trained staff and students of the Faculty of Arts) were tested in two groups in a language laboratory at Leiden University. They listened to the tapes over headphones. Twenty-two (phonetically naive) subjects participated in the test as part of a phonetics class taught by the second author, and were tested in a classroom at Leiden University. They listened to the tape over loudspeakers. Subjects were instructed to determine the stress position of *nana* in each utterance (with binary forced choice) and to note their responses on the response sheets provided. The experiment lasted approximately 30 min.

B. Statistical analysis

We determined the number of judgments favoring initial stress for each stimulus and expressed this as a percentage, henceforth $p(init)$.

There were three goals for the statistical analysis. The primary goals were to establish the relative strengths of duration and intensity level manipulations as stress cues, and to determine to what extent the way of varying intensity (overall versus above 500 Hz only) interacts with the effects of duration and intensity level. An additional goal was to determine to what extent the way of presentation interacts with the above effects. A four-way analysis of variance was performed, with $p(init)$ as the dependent variable, and with *presentation* (headphones versus loudspeakers), *method* of varying intensity (intensity level increments in all bands versus spectral balance, i.e., increasing intensity above 0.5 kHz only), *duration* (seven steps) and *intensity level* (seven steps) as fixed effects and with repetition as repeated measure.[3] The effects of *duration* and *intensity level* variations will show up as main effects in the ANOVA. The importance of *method* and *presentation* will be visible in their interactions with *duration* and *intensity* level. The main effects of *presentation* and *method* are irrelevant in this research, since they will merely reflect a difference in overall bias favoring one stress position over the other.

C. Results

1. Global presentation

We computed the consistency of each subject by comparing their answers on the first and the second presentation of the stimuli. Subjects who were not consistent in more than

TABLE III. Main effects and interactions of duration, intensity level, presentation (headphones versus loudspeakers), and method (of varying intensity: overall versus high frequency bands only) on *p(init)*. *F* ratio, significance of *F* and percentage of explained variance (η^2) are given.

Effects	*F*	sign.	η^2
Main effects			
Duration	761.3	<0.001	68
Intensity level	129.4	<0.001	12
Presentation	2.8	NS	0
Method of variation	3.8	NS	0
Two-way interactions			
Duration * intensity level	7.8	<0.001	4
Duration * presentation	53.5	<0.001	5
Duration * method	5.4	<0.001	0
Intensity level * presentation	7.8	<0.001	1
Intensity level * method	46.1	<0.001	4
Presentation * method	<1	NS	0
Three-way interactions			
Duration * int. level * presentation	1.9	0.003	1
Duration * int. level * method	2.4	<0.001	1
Duration * presentation * method	1.6	NS	0
Int. level * presentation * method	4.1	0.001	0

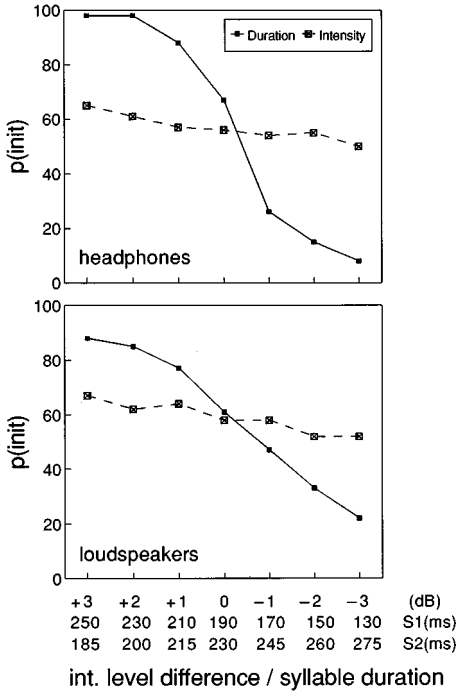

FIG. 1. Percentage of listeners ''initial stress'' judgments, $p(init)$, for the 91 stimuli *nana* as a function of syllable duration (solid lines) and overall intensity (dashed lines). The differences in *intensity level* ($IL_{\sigma 1}-IL_{\sigma 2}$ in dB), obtained by spectrally uniform amplification, are given along the *x* axis, top line. *Duration* values (in ms) are given on the middle and bottom lines for the first and second syllable, respectively. The results are presented for each presentation condition separately: headphones (upper panel) and loudspeakers (lower panel).

60% of the cases were omitted from further analysis. The 60% consistency cutoff point was chosen as there was a clear discontinuity between the six poorest subjects and the 40 individuals who remained in the analysis. Twenty-one subjects who listened to the tape over headphones and 19 subjects who listened to the tape over loudspeakers were used for further analysis.

The listening test yielded a total of 7280 responses (91 stimuli * two repetitions * 40 subjects). Overall, 57% of the responses favored initial stress, which indicates that there is a slight bias for initial stress. This bias is above chance, as determined by a binomial test ($p<0.001$).

In Table III the main effects and interactions of *duration*, *intensity level*, *presentation*, and *method* are given.

There is a large effect of both *duration* and *intensity level* on $p(init)$. In answer to our question if varying intensity level in a more realistic way, i.e., by varying the spectral balance, has an effect on stress perception, we can provisionally conclude from the highly significant interaction of *intensity level* with *method*, that the method of variation has at least a considerable influence on the effect of *intensity level* on $p(init)$. Furthermore, the significance of the two- and three-way interactions with *presentation* means that the way of presentation has an influence on both the effect of *duration* and *intensity level* on $p(init)$. Given the significant two- and three-way interactions we decided to study the main effects of duration and intensity level separately for each presentation condition (headphones versus loudspeakers) and for each method of varying intensity (overall level versus manipulating spectral balance). Therefore, we ran two separate two-way analyses of variance with *duration* and *intensity* (uniformly distributed gain increments, henceforth *intensity*) as fixed effects and with repetition as repeated measure and two more analyses with *duration* and *spectral balance* as fixed effects. The results are described below in separate subsections for each way of varying intensity level.

2. Intensity (uniformly distributed gain increments)

In this subsection, the effect of *duration* and *intensity*, the latter varied by spectrally uniform amplification, on $p(init)$ is examined. Figure 1 shows the decrease of the percentage perceived initial stress as a function of duration and intensity level difference. The duration of the first syllable decreases from left to right, while at the same time the duration of the second syllable increases. The intensity scale gives the difference in overall intensity level (IL) between the initial syllable and the final syllable ($IL_{\sigma 1}-IL_{\sigma 2}$). The upper panel displays the results for the stimuli presented over headphones, the lower panel those for the stimuli presented over loudspeakers. This way of presenting the data does in no way mean that we assume the duration and the intensity range to be absolutely identical. However, the similarity of both ranges is that they are both a representative reflection of ranges found in our production data (see Sec. I A 1).

When stimuli are presented over headphones, the whole range of intensity change produces only a slight decrease of *p(init)*: from 65% to 50%. The range of duration change produces a much larger decrease of *p(init)*: from 98% to 8%. *Duration*, *intensity*, and their interaction together explain 97% of the variance. Although the contribution of *intensity* is statistically significant [$F(6,91)=4.9$, $p<0.001$], it is only small compared to that of *duration* [$F(6,91)=315.5$, $p<0.001$]. *Intensity* alone explains a mere 2% of the variance. *Duration* on the other hand, explains as much as 93% of the variance. There is a significant interaction between *duration* and *intensity* [$F(36,49)=1.8$, $p=0.26$], which ex-

plains 2% of the variance. This interaction is due to the fact that overall intensity level variations have little or no influence at the extremes of the duration scale, where judgments are mainly guided by duration differences, whereas they have a larger influence on *p(init)* in the temporally more ambiguous stimuli.

As can be seen in the lower panel of Fig. 1, presenting the stimuli over loudspeakers mainly affects the effectiveness of duration as a stress cue and hardly influences the perceptual contribution of intensity level differences. In this case duration produces a less steeply sloping decrease, from 88% to 22%, whereas intensity again produces a decrease of 15%. Again, the effects of both *duration* and *intensity* are significant [$F(6,91)=90.5$, $p<0.001$ and $F(6,91)=5.1$, $p=0.001$, respectively]. *Duration* explains 80% of the variance and *intensity* 5%. Together with their interaction, they explain 93% of the variance, although the interaction was not significant in this condition [$F(36,49)=1.4$, NS].

Our intermediate conclusion is that intensity level variation, as used in this experiment, implemented by spectrally uniform amplification, is only a minor stress cue, whether stimuli are presented over headphones or over loudspeakers.

3. Spectral balance (intensity level variation by increments in the higher frequency bands only)

Figure 2 shows the decrease of *p(init)* as a function of duration ratio and difference in spectral balance. The duration range is the same as in Fig. 1, but now the intensity level differences are obtained by increasing the levels in the higher frequency bands only. The intensity level scale gives the difference in spectral balance between the initial syllable and the final syllable ($B_{\sigma 1}-B_{\sigma 2}$). Again, the upper panel presents the data of the stimuli presented over headphones, the lower panel of the stimuli presented over loudspeakers.

The whole range of spectral balance produces a decrease of 41%: from 77% to 36% when stimuli are presented over headphones. The duration range produces a decrease of 86%: from 95% to 9%. *Duration, spectral balance* and their interaction together explain 99% of the variance. Both *duration* and *spectral balance* have a significant effect on *p(init)* [*duration*: $F(6,91)=420.5$, $p<0.001$; *spectral balance*: $F(6,91)=73.7$, $p<0.001$]. *Duration* alone explains 76% of the variance, whereas *spectral balance* explains 13% of the variance. The significant interaction between *duration* and *spectral balance* [$F(36,49)=9.0$, $p<0.001$] is again due to the fact that variations in spectral balance have less influence on stress judgments at the extremes of the duration range.

When stimuli are presented over loudspeakers, the effect of *duration* on *p(init)* decreases. However, while intensity (Sec. I C 2) proves equally ineffective through headphones as over loudspeakers, *presentation* strongly influences the relative strength of effort and duration as stress cues. *Duration* and *spectral balance* produce an almost equal decrease of *p(init)*: 80% to 24% for *duration* versus 86% to 20% for *spectral balance*. This, in fact, means that subjects rely more heavily on differences in spectral balance than on duration differences when stimuli are presented over loudspeakers. Both *duration* and *spectral balance* have a highly significant effect on *p(init)* [*duration*: $F(6,91)=77.2$, $p<0.001$; *spectral balance*: $F(6,91)=115.5$, $p<0.001$]. Together with their interaction they explain 96% of the variance. *Duration* alone explains ''only'' 35%, whereas *spectral balance* explains as much as 53%. The significant interaction of *duration* and *spectral balance* [$F(36,49)=3.1$, $p<0.001$] is due to the fact that the more extreme values of one parameter add disproportionally more weight as the other parameter is more ambiguous.

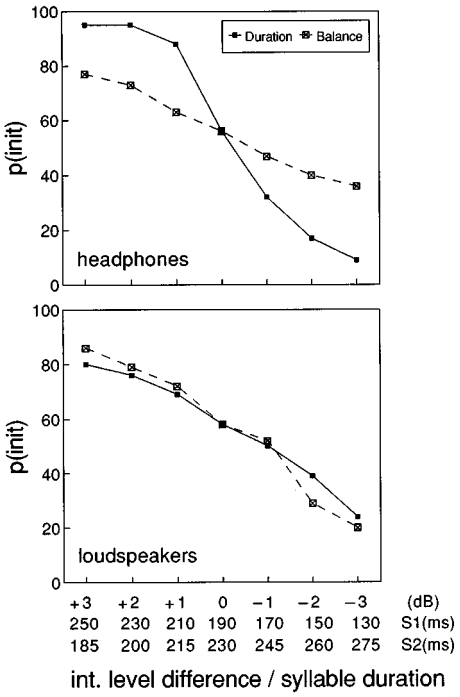

FIG. 2. Percentage of listeners ''initial stress'' judgments, *p*(*init*), for the 91 stimuli *nana* as a function of syllable duration (solid lines) and overall intensity (dashed lines). The differences in *spectral balance* ($B_{\sigma 1}-B_{\sigma 2}$ in dB), obtained by amplification of frequency components above 500 Hz only, are given along the *x* axis, top line. *Duration* values (in ms) are given on the middle and bottom lines for the first and second syllable, respectively. The results are presented for each presentation condition separately: headphones (upper panel) and loudspeakers (lower panel).

We conclude from these results that realistic intensity level manipulations (i.e., mimicking speech production effort by incrementing intensity level in the higher frequency bands only) provide a relatively strong stress cue, and in fact approximate the cue value of duration differences, whereas overall intensity level differences do not provide a substantial stress cue.

Since the reliability of duration as a cue is degraded when the stimuli are presented over loudspeakers, the relative cue value of spectral balance in this situation becomes more important. One explanation could be that subject differences (phonetically trained versus phonetically naive) were responsible for the difference in effectiveness of the duration cue. Of course, an alternative explanation of this interaction is that accurate perception of duration differences suffers from reverberation of the acoustic signal in the room in which the subjects were tested. Locating syllable boundaries in reverberant speech is more difficult since their exact locations are obscured by energy reflections of preceding segments. As a result, the variation in vocal effort became

relatively more important as a stress cue since its acoustical correlate (spectral balance) is not easily affected by reverberation. The experiment reported on in the next section was specifically set up to allow us to choose between the two alternative explanations suggested above.

II. PERCEPTION EXPERIMENT II

A. Effect of ''reverberation'' on the perception of differences in duration and spectral balance

In a room, the acoustic signal produced by either a talker or a loudspeaker may reach a listener by many individual soundpaths. The original speech at the talker's (or loudspeaker's) position and the resulting sound at the listener's position are not identical. Comparing the specific distribution of sound intensity over frequency and time of the original speech with that of the transmitted speech, a certain degree of smearing of the finer details is found: the temporal intensity distribution will be blurred by the combined effects of the many individual soundpaths with various time delays (Houtgast and Steeneken, 1973, 1985; Duquesnoy and Plomp, 1980).

We assume that reverberation, which is a result of myriad reflected sound waves, and is mainly a distortion in the temporal domain, is responsible for the fact that the relative importance of duration as a cue in stress perception decreased when the stimuli were presented over loudspeakers. It has been amply demonstrated that reverberation has a considerable effect on speech intelligibility. These effects appear to be due to the reflections that arrive at the subjects' ear(s) later than about 30 ms after the direct signal, while earlier reflections are integrated with the direct sound (Gelfand and Silman, 1979 and references mentioned there).

In order to rule out alternative explanations for the reverberation effect based on subject differences (see above), we ran a control experiment. We presented both nonreverberant and reverberant stimuli over headphones with the same duration and intensity level manipulations as in the previous experiment and asked subjects *in a within-subjects design* to determine the stress position of each stimulus.

B. Methods

1. Stimulus material

The reverberant stimuli were produced by processing the master test recordings through a Yamaha SPX 90II digital multi-effect processor. The SPX 90II creates a highly natural sounding reverberation. Reverberation time for this particular processor is defined as the length of the time it takes for the level of reverberation at 1 kHz to decrease by 60 dB. Usually natural reverberation varies according to the frequency of the sound: the higher the frequency the more the sound tends to be absorbed by walls, furnishings and even air. We decided not to alter the reverberation time of the high frequencies in proportion to the mid-frequency reverberation time.

We decided to use a reverberation time of 0.6 s for our stimuli. This value was chosen so that an impulse recorded in a sound insulated booth but processed through the SPX 90II sounded and looked more or less identical to an impulse recorded in the reverberant room in which the stimuli were presented in the previous experiment. Figure 3 presents an example of a test item (*Wil je* ˈ*nana zeggen*) with and without artificial reverberation.

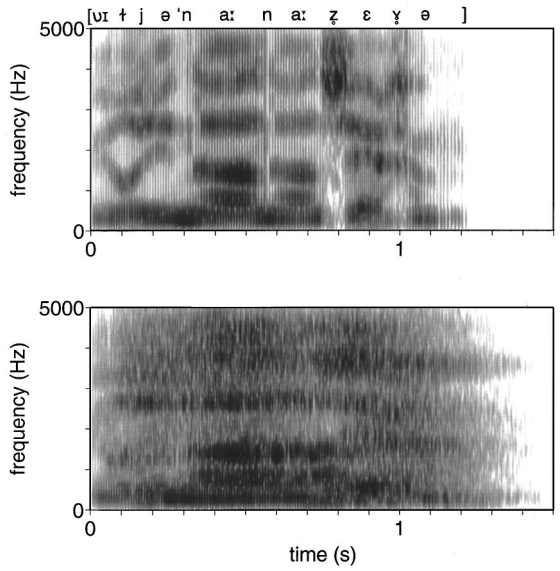

FIG. 3. Example of a test item (*Wil je* ˈ*nana zeggen*) without (upper panel) and with (lower panel) artificial reverberation.

3. Subjects and procedures

A stimulus set was prepared containing the 182 stimuli (91 with and 91 without reverberation) in four different random orders. The third and fourth orders were identical to the first and second, the only difference being that they were recorded in reverse sequence. The 182 stimuli were presented on-line in blocks of 13 utterances with 2-s intervals between utterances and a larger interval between blocks. The procedure was similar to that in the first experiment. Forty-four subjects (staff and students of the Faculty of Arts) participated in the experiment. Seven subjects were phonetically trained and 37 were phonetically naive. The latter subjects were paid for their service. They were tested in four groups in a language laboratory at Leiden University. Each group listened to one of the four different orders. They listened to the stimuli over good quality stereo headphones.

C. Results

1. Global presentation

The reliability of the subjects was determined by relating their individual scores to the composite group score. In order to know how each of them affected the reliability of the group, Cronbach's α was calculated when each of the subjects was removed from the group in turn. We wanted to use the same number of subjects as in the first experiment. We therefore eliminated the four subjects whose exclusions yielded the largest increase of α. Consequently, 40 subjects were used for further analysis.

We determined the number of judgments favoring initial stress for each stimulus and calculated the percentage, *p(init)*. The listening test yielded a total of 7280 responses (182 stimuli * 40 subjects). Overall 56% of the responses

TABLE IV. Main effects and interactions of duration, intensity level, presentation (nonreverberant versus reverberant stimuli), and method (of varying intensity: overall versus high-frequency bands only) on *p(init)*. *F* ratio, significance of *F* and percentage of explained variance (η^2) are given.

Effects	*F*	sign.	η^2
Main effects			
Duration	338.0	<0.001	60
Intensity level	78.8	<0.001	14
Presentation	28.8	<0.001	1
Method of variation	24.5	<0.001	1
Two-way interactions			
Duration * intensity level	2.5	0.004	3
Duration * presentation	53.5	<0.001	9
Duration * method	3.4	0.009	1
Intensity level * presentation	5.4	<0.001	1
Intensity level * method	27.0	<0.001	5
Presentation * method	<1	NS	0
Three-way interactions			
Duration * int. level * presentation	2.9	0.001	3
Duration * intensity level * method	1.4	NS	1
Duration * presentation * method	4.3	0.002	1
Int. level * presentation * method	1.7	NS	0

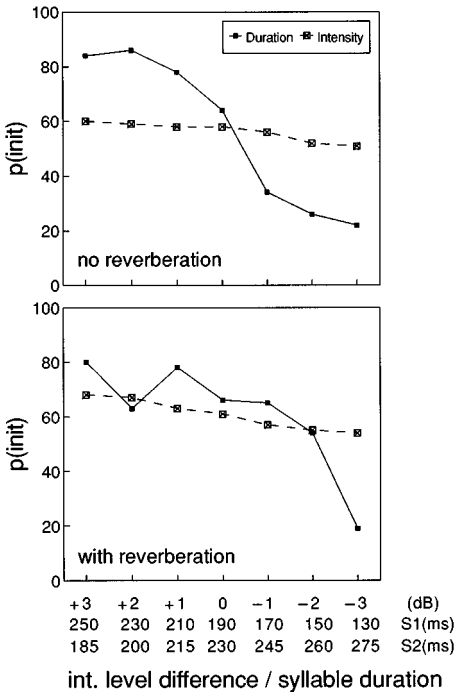

FIG. 4. Percentage of listeners ''initial stress'' judgments, *p*(*init*), for the 91 stimuli *nana* as a function of syllable duration (solid lines) and overall intensity (dashed lines). The differences in *intensity level* ($IL_{\sigma 1}$–$IL_{\sigma 2}$ in dB), obtained by spectrally uniform amplification, are given along the *x* axis, top line. *Duration* values (in ms) are given on the middle and bottom lines for the first and second syllable, respectively. The results are presented for each reverberation condition separately: no reverberation (upper panel) and with artificial reverberation (lower panel).

favored initial stress, which indicates that there is a slight bias for initial stress. This bias is above chance, as determined by a binomial test ($p<0.001$).

As in the previous experiment, we ran a four-way analysis of variance, with *p(init)* as the dependent variable, and with *presentation* (reverberant versus nonreverberant), *method* (adding intensity in all bands versus adding intensity in higher bands only), *duration* (seven steps) and *intensity level* (seven steps) as fixed effects. There were no repeated measures. Since there is no residual variance, the variance caused by the fourth-order interaction was used as the error term. In Table IV the main and interaction effects are given. As can be seen in Table IV, the crucial main effects and interactions are quite similar to those in the previous experiment. There are large effects of both *duration* and *intensity level* on *p(init)*, although the effect of *duration* on *p(init)* is smaller than in the first experiment. The significant main effect of *presentation* indicates that there was a difference in stress bias between reverberant stimuli and nonreverberant stimuli: 59% versus 54%, respectively, which we attribute to the fact that the end of the second syllable of *nana* is more strongly demarcated by the unvoiced fricative [z], than the initial syllable, which is succeeded by an identical syllable. Therefore, the perceived length of the initial syllable is possibly more strongly influenced by reverberation than the second syllable.

The significance of the two- and three-way interactions with *presentation* means that reverberation has an influence on both the effect of *duration* and *intensity level* on *p(init)*. Crucially, significant two- and three-way interactions with *presentation* are found similar to the interactions in the first experiment. This indicates that the effect of reverberation is highly comparable to the effect of the way of presentation in the first experiment. This is an indication that reverberation was indeed (at least for the greater part) responsible for the difference in relative importance of *duration* and *spectral balance* between the two presentation conditions. As in Secs. I C 2 and 3, we will now study the main effects of *duration* and *intensity level* in more detail separately for *presentation* (reverberant versus nonreverberant) and *method* (uniform intensity level versus spectral balance). Results are presented in the next subsection.

2. Reverberant versus nonreverberant speech

We ran two separate two-way analyses of variance with *duration* and *intensity* as fixed effects and two more analyses with *duration* and *spectral balance* as fixed effects. There were no repeated measures: only percentages of explained variance but no *F* ratios could be computed.[4] Figure 4 shows the decrease of the percentage perceived initial stress, *p(init)*, as a function of duration ratio and intensity presented as in Fig. 1 with uniform intensity level differences. The upper panel shows the data for the nonreverberant stimuli, the lower panel shows the data for the reverberant stimuli. Figure 5 shows similar data, but now with differences in spectral balance as in Fig. 2.

Figures 4 and 5 show that the effectiveness of duration deteriorates considerably for the reverberant stimuli.[5] As can be seen in Fig. 4, *intensity* does not serve as a stress cue at all for the nonreverberant stimuli. The effectiveness of this cue slightly increases for the reverberant stimuli. This tendency was also observed in the previous experiment.

The results for *spectral balance* (Fig. 5) are comparable to those in the previous experiment: again a considerable

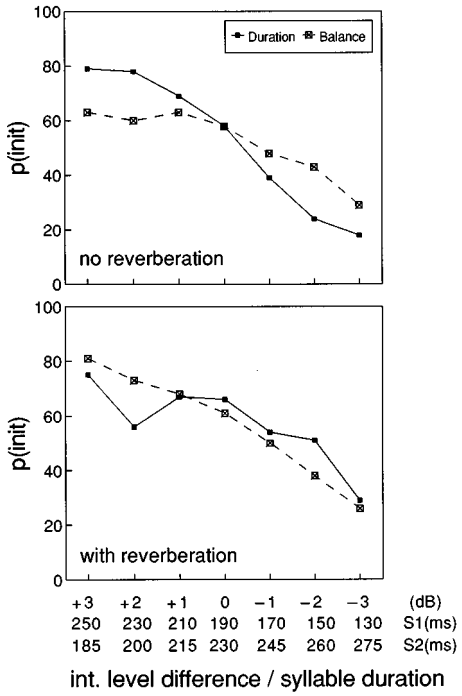

FIG. 5. Percentage of listeners ''initial stress'' judgments, $p(init)$, for the 91 stimuli *nana* as a function of syllable duration (solid lines) and overall intensity (dashed lines). The differences in *spectral balance* ($B_{\sigma 1}-B_{\sigma 2}$ in dB), obtained by amplification of frequency components above 500 Hz only, are given along the *x* axis, top line. *Duration* values (in ms) are given on the middle and bottom lines for the first and second syllable, respectively. The results are presented for each reverberation condition separately: no reverberation (upper panel) and with artificial reverberation (lower panel).

increase in effectiveness of *spectral balance* is found for the reverberant stimuli.

In the next section we will compare the results of both experiments in more detail.

III. COMPARISON OF EXPERIMENTS 1 AND 2

In Table V, we present an overview of the percentages explained variance for duration, intensity and spectral balance in both experiments to compare the relative strength of the stress cues in both experiments. The left-hand part of the table presents the data for experiment 1, in which stimuli were presented to half of the subjects over headphones and to half of the subjects over loudspeakers (*separate conditions*). The right-hand part of the table presents the data of the present experiment (2), in which both reverberant and nonreverberant stimuli were presented in a *within-subjects design* over headphones (*mixed condition*).

TABLE V. Relative strength of stress cues (in % explained variance η^2) in reverberant and nonreverberant stimuli, presented in separate and mixed conditions (experiment 1 and experiment 2, respectively).

	Experiment 1 separate conditions		Experiment 2 mixed condition	
	Overall int. (Fig. 1)	Spectral balance (Fig. 2)	Overall int. (Fig. 4)	Spectral balance (Fig. 5)
No reverb				
Duration	93	76	94	73
Intensity level	2	13	0	18
Dur. * int.	2	10	6	9
Residue	3	1	…	…
Reverb				
Duration	80	35	84	35
Intensity level	5	53	6	57
Dur. * int.	8	8	10	8
Residue	7	4	…	…

As can be seen in Table V the percentages explained variance in both experiments are almost identical. We conclude on the basis of these results that duration indeed suffered from reverberation and that reverberation was therefore responsible for the relative increase in effectiveness of spectral balance when stimuli were presented over loudspeakers.

In the present experiment, variations in duration did not lead to an equally large change in *p(init)* as in the previous experiment. In the nonreverberant speech condition, *p(init)* decreased with roughly 60% from about 80% to 20%, whereas in the previous experiment in this condition a range was covered between 98% and 8%. This could possibly be due to the fact that reverberant and nonreverberant stimuli were presented in random succession, which might have prevented our listeners from tuning in to one specific speech type.[6]

In summary, the importance of duration as a cue to stress perception decreased under reverberation ($T=0.6$ s), whereas the relative contribution of spectral balance manipulations increased strongly. The magnitude of the effects in both experiments were in the same range. The effectiveness of overall intensity, however, was hardly affected by reverberation and was equally poor in both experiments. On the basis of these results we conclude that the use of duration as a cue for stress suffers from reverberation. As a result, loudness (as a reflection of vocal effort) becomes relatively more important as a stress cue showing that its acoustical correlate (spectral balance) is not easily affected by reverberation.

IV. GENERAL DISCUSSION AND CONCLUSIONS

In this study we reconsidered the general claim that loudness is a weak cue in the perception of stress. This traditional claim was based on perception experiments in which loudness was varied in a naive way: All parts of the spectrum were amplified uniformly. We hypothesized that varying loudness more realistically will make it a stronger stress cue, and that we could possibly rehabilitate the traditional claim that languages such as Dutch and English have dynamic (rather than melodic or temporal) stress.

From the results of both experiments, we conclude that loudness implemented as a difference in overall intensity level (i.e., manipulating gain without changing spectral balance) provides only a marginal stress cue. Of course, we need not be surprised that intensity level variations turn out to provide only a marginal stress cue. In fact, it seems to us that intensity level variation will never have communicative significance, for the simple reason that intensity level is too susceptible to noise. If the speaker accidentally turns his head, or passes a hand across his mouth, intensity level drops of greater magnitude than those caused by the difference between stressed and unstressed syllables will easily occur. For this reason, manipulating intensity in stress perception

experiments seemed ill-advised. The reason why it was used in the classical studies by Fry (1955) and Mol and Uhlenbeck (1956) must have been that there were simply no alternatives available for investigating the role of loudness in stress perception.

In contrast, loudness realistically implemented as the acoustical reflection of greater vocal effort, is a reliable stress cue, close in strength to duration. Moreover, the differences in spectral balance provide an even stronger stress cue than duration when accurate perception of syllable and segment boundaries is hampered, for instance in a reverberant environment. Examples of such reverberant listening conditions in daily life abound. In fact, studying speech communication in rooms, halls etc. is probably more realistic than in sound-insulated booths and free-field situations. Therefore, it seems that listeners have different cooperating cues at their disposal to determine linguistic stress position. The effectiveness of the different cues depends on environmental circumstances in which speech is perceived.

Results of a perception experiment carried out by Beckman (1986) for English and Japanese showed that these two languages differed greatly as to the relative importance of $F0$, duration and loudness as perceptual cues to stress. Both Japanese and English listeners were presented with disyllabic words in which all these parameters were varied according to production data. Japanese is an archetypal nonstress-accent language, a so-called pitch-accent language, with $F0$ as the most consistent acoustical correlate of stress/accent. English is an archetypal stress-accent language with the same acoustical correlates of stress and accent as Dutch. The comparison between English and Japanese listeners showed that Japanese listeners seemed to rely heavily on differences in $F0$ and they hardly used any of the other cues. English listeners also relied heavily on $F0$, although to a much lesser extent. Loudness, however, was also found to be a very effective cue for English listeners in stress perception. Loudness in this experiment was operationalized as ''total amplitude,'' a measure of power integrated over the entire duration of the vocalic nucleus (i.e., energy), rather than as peak intensity. Beckman assumes this measure to be closely related to loudness and she attributes the success of this cue to this relation:

> Thus the total amplitude may be a better correlate of stress than is either duration or intensity alone and it may be a more consistent perceptual cue simply because it is a better measure of loudness,... (Beckman, 1986, p. 197).

In our view this measure of loudness is equally unrealistic as overall intensity level manipulations are. Beckman in fact measured the combined effect of peak intensity and duration. It is therefore no surprise that this measure yields considerably better results than either duration or peak intensity alone. It has only been established for pure tones of a relatively short duration that differences in duration are responsible for differences in the perception of loudness. Although the literature agrees about the fact that there is a certain threshold value above which duration changes no longer influence loudness, the literature largely disagrees as to determining the exact value of this threshold. However, despite the great variability of results regarding the threshold value among the various studies, they largely agree on the fact that temporal integration of energy occurs at very short durations (Beckman, 1986 and references mentioned there). Therefore, although this measure may have some relevance for plosives (i.e., the longer a noise burst, the louder it is perceived), it has no relevance for vowels and sonorants, since these sounds are no short acoustic events. Therefore, in our view, this operationalization of loudness has no relevance in vocalic nuclei.[7]

In our view, the ultimate test to investigate whether English listeners are more sensitive to loudness than Japanese listeners, would be to synthesize similar stimuli as used in Beckman (1986) while separating focused and nonfocused material and varying loudness in the way described in the present article. If it is indeed true that languages such as Dutch and English have dynamic accent as opposed to pitch accent in languages such as Japanese, Japanese listeners will be insensitive to these more realistic loudness manipulations as well, whereas English listeners would make considerable use of these differences.

In addition to the above mentioned, more linguistically oriented implications, the findings of the present study have some more practical, application-based implications as well. The results can probably be used to improve the quality of speech synthesis. In future research, experiments should be executed investigating if stress and focus domains could be more optimally synthesized if we take the present results into account. There are elaborate rule-sets in Dutch text-to-speech systems to predict whether or not a word should be accented (Quené and Kager, 1993; Dirksen and Quené, 1993). If a word is accented, all its syllables, stressed as well as unstressed, should be lengthened, at least in Dutch, relative to syllables of a word that remains unaccented (Eefting, 1991; van Heuven, 1993). The stressed syllables of both accented and unaccented words should be marked by a combination of (extra) longer duration[8] and greater loudness. The present experiments showed that the relative importance of these cues depends on the listening circumstances; it is therefore necessary to represent both cues optimally in synthetic speech to guarantee adequate stress perception independent of listening circumstances especially because for unaccented words these cues are the only remaining cues to stress. Furthermore, in our experiment stress was varied so as to reflect production data. However, intensity level, spectral balance and duration could be combined in a more extreme way, for instance by both adding and shifting intensity levels. Listeners could probably prefer more strongly marked stress positions when listening to synthetic speech, because of the fact that there is not always a one-to-one mapping of what speakers do and what listeners want. The intelligibility of synthesized speech in text-to-speech systems could possibly improve by a more accurate marking of stress and accent since the former facilitates the recognition of words in continuous speech (cf. van Heuven, 1988), while the latter prompts the listener to give priority to bottom-up processing exactly there where it matters (cf. Terken and Nooteboom, 1987; van Donselaar, 1995).

We assumed the differences in spectral balance to be caused by a more pulse-like shape of the glottal waveform while producing stressed syllables. Future manipulations could be made even more realistically by manipulating the glottal pulse separately instead of using digital filtering of the oral output.

To conclude this paper, the most important finding of this study is that listeners are more susceptible to intensity level variations when detecting stress position than hitherto has been assumed. This is due to the fact that intensity level differences in our experiments were implemented in a more realistic way, i.e., by amplification in the higher frequency bands only, as the acoustical reflection of an increase in vocal effort used to produce stressed syllables. The results can be viewed as a first step to rehabilitation of the claim that languages such as Dutch and English have dynamic stress, with perceived loudness as its most reliable cue.

ACKNOWLEDGMENTS

Portions of this research have been presented at the ESCA workshop on Prosody, Lund (September 1993) and at the 127th meeting of the Acoustical Society of America, Cambridge, MA (June 1994). The authors would like to thank H. Traunmüller, J. W. de Vries, S. G. Nooteboom, and one anonymous reviewer for comments on earlier versions of this paper, ideas, and discussion.

[1]It was pointed out to us by Hartmut Traunmüller that ''spectral emphasis'' might be a better term. We agree, but stick to the term ''spectral balance'' to insure terminological uniformity with our earlier publications.

[2]Note, however, that we used Dutch words. There is no guarantee that English will behave like Dutch. As a case in point, a Dutch word spoken without a pitch accent is pronounced some 15% faster than its accented counterpart (linear time compression, cf. Eefting, 1991; Sluijter and van Heuven, 1995). A similar experiment showed that only the accented foot, but not the entire word, is time-expanded in American English (Turk and Sawush, 1995).

[3]A similar analysis was performed on the arcsine transformed percentages (cf. Studebaker, 1985). There were no crucial differences, so we decided to use the nontransformed percentages in all the analyses performed on the data in this paper.

[4]In this type of situation it is not uncommon to adopt the highest interaction as the numerator term. The second-order interaction is the only interaction in this analysis and it is inherent to this type of experiment that this interaction plays a systematic role: When one cue is ambiguous the other one becomes more important; consequently, the interaction is not a suitable numerator term. Since the primary goal of this analysis is to quantify the relative magnitude of the effects (the significance of which has been shown in earlier experiments), rather than to determine the significance of the effects, we decided to refrain from any significance testing at all.

[5]Unexpectedly, in this condition subjects hardly used duration as a cue in duration step 2 (230–200), whereas they heavily relied on duration in step 3. We do not have an explanation for this effect and we assume that there is some unknown acoustic interference of reverberation and duration in some of the stimuli.

[6]Besides, the subjects were mainly students who had never participated in listening experiments before. The results of the 20 most reliable subjects, as determined with Cronbach's α, cover a much larger range, comparable to the range covered in the first experiment. The seven phoneticians who participated in the present experiment all belonged to this group. This means that subject differences could partly be held responsible for the distortion of the duration results.

[7]Beckman (1986) did not consistently separate focused and nonfocused material. The relative strength of $F0$ may therefore be overestimated, in any case in English and probably also in Japanese.

[8]The relative importance of stress cues may differ from language to language. Specifically stress cues such as duration (Berinstein, 1979) and pitch (Potisuk *et al.*, 1996) assume a lower position in the rank order of cues as these parameters are simultaneously exploited in other linguistic contrasts (vowel quantity and lexical tone, respectively).

Beckman, M. E. (**1986**). *Stress and Non-Stress Accent* (Foris, Dordrecht).
Berinstein, A. E. (**1979**). ''A Cross-linguistic study on the perception and production of stress,'' *Working Papers in Phonetics* (University of California, Los Angeles), No. 47.
Bloomfield, L. (**1933**). *Language* (Holt, Rinehart and Winston, New York).
Brandt, J. F., Ruder, K. P., and Shipp, Jr., I. (**1969**). ''Vocal Loudness and Effort in Continuous Speech,'' J. Acoust. Soc. Am. **46**, 1543–1548.
Dirksen, A., and Quené, H. (**1993**). ''Prosodic analysis: The next generation,'' in *Analysis and Synthesis of Speech: Strategic Research Towards High-Quality Text-To-Speech Generation*, edited by V. J. van Heuven and L. C. W. Pols (Mouton de Gruyter, Berlin), pp. 131–144.
Donselaar, W. van (**1995**). ''Effects of accentuation and given/new information on word processing,'' doctoral dissertation, Utrecht University, Utrecht.
Duquesnoy, A. J., and Plomp, R. (**1980**). ''Effect of reverberation and noise on the intelligibility of sentences in cases of presbyacusis,'' J. Acoust. Soc. Am. **68**, 537–544.
Eefting, W. Z. F. (**1991**). ''The effect of information value and accentuation on the duration of Dutch words, syllables, and segments,'' J. Acoust. Soc. Am. **89**, 412–424.
Fry, D. B. (**1955**). ''Duration and intensity as physical correlates of linguistic stress,'' J. Acoust. Soc. Am. **27**, 765–768.
Fry, D. B. (**1958**). ''Experiments in the perception of stress,'' Lang. Speech **1**, 126–152.
Fry, D. B. (**1965**). ''The dependence of stress judgments on vowel formant structure,'' in *Proceedings of the 5th International Congress on Phonon Science, Münster 1964* (Karger, Basel), pp. 306–311.
Gelfand, S. A., and Silman, S. (**1979**). ''Effects of small room reverberation upon the recognition of some consonant features,'' J. Acoust. Soc. Am. **66**, 22–29.
Glave, R. D., and Rietveld, A. C. M. (**1975**). ''Is the effort dependence of speech loudness explicable on the basis of acoustical cues?,'' J. Acoust. Soc. Am. **58**, 875–879.
Hart, J. 't, Collier, R., and Cohen, A. (**1990**). *A Perceptual Study of Intonation; An Experimental-Phonetic Approach to Speech Melody* (Cambridge U.P., Cambridge).
Hermes, D. J., and Rump, H. H. (**1993**). ''The role of pitch in lending prominence to syllables,'' in *Proceedings of an ESCA Workshop on Prosody*, Working papers **41**, Department of Linguistics, Lund University, edited by D. House and P. Touati, pp. 28–31.
Heuven, V. J. van (**1987**). ''Stress Patterns in Dutch (Compound) Adjectives: Acoustic Measurements and Perception Data,'' Phonetica **44**, 1–12.
Heuven, V. J. van (**1988**). ''Effects of stress and accent on the human recognition of work fragments in spoken context: gating and shadowing,'' in *Proceedings of the 7th FASE/Speech-88 Symposium*, edited by W. A. Ainsworth and J. N. Holmes (The Institute of Acoustics, Edinburgh), pp. 811–818.
Heuven, V. J. van (**1993**). ''On the temporal domain of focal accent,'' in *Proceedings of an ESCA Workshop on Prosody*, Working papers **41**, Department of Linguistics, Lund University, edited by D. House and P. Touati, pp. 132–135.
Heuven, V. J. van, and Menert, L. (**1996**). ''Why stress position bias?,'' J. Acoust. Soc. Am. **100**, 2439–2451.
Houtgast, T., and Steeneken, H. J. M. (**1973**). ''The modulation transfer function in room acoustics as a predictor of speech intelligibility,'' Acustica **28**, 66–73.
Houtgast, T., and Steeneken, H. J. M. (**1985**). ''A review of the MTF concept in room acoustics and its use for estimating speech intelligibility in auditoria,'' J. Acoust. Soc. Am. **77**, 1069–1077.
Issatchenko, A. V., and Schädlich, H. J. (**1966**). ''Untersuchungen über die deutsche Satzintonation,'' Stud. Grammatica **7**, 7–64.
Katwijk, A. van (**1974**). *Accentuation in Dutch; An Experimental Linguistic Study* (Van Gorcum, Amsterdam).
Lagendijk, M. (**1992**). *The XAudlab User Manual* (Speech Processing Expertise Centre, Leidschendam, The Netherlands).
Larkey, L. S. (**1982**). ''Reiterant speech: An acoustic and perceptual validation,'' J. Acoust. Soc. Am. **73**, 1337–1345.
Lehiste, I. (**1970**). *Suprasegmentals* (MIT, Cambridge, MA).

Lehiste, I., and Peterson, G. E. (**1959**). ''Vowel amplitude and phonemic stress in American English,'' J. Acoust. Soc. Am. **31**, 428–435.
Mol, H. G., and Uhlenbeck, G. M. (**1956**). ''The linguistic relevance of intensity in stress,'' Lingua **5**, 205–213.
Morton, J., and Jassem, W. (**1965**). ''Acoustic correlates of stress,'' Lang. Speech **8**, 148–158.
Peterson, G. E., and Lehiste, I. (**1960**). ''Duration of syllable nuclei in English,'' J. Acoust. Soc. Am. **32**, 693–703.
Potisuk, S., Gandour, J., and Harper, M. P. (**1996**). ''Acoustic correlates of stress in Thai,'' Phonetica **53**, 200–220.
Quené, H., and Kager, R. (**1993**). ''Prosodic sentence analysis without parsing,'' in *Analysis and Synthesis of Speech: Strategic Research Towards High-Quality Text-to-Speech Generation*, edited by V. J. van Heuven and L. C. W. Pols (Mouton de Gruyter, Berlin), pp. 115–130.
Rietveld, A. C. M., and Koopmans-van Beinum, F. J. (**1987**). ''Vowel reduction and stress,'' Speech Commun. **6**, 217–229.
Sluijter, A. M. C., and Heuven, V. J. van (**1995**). ''Effects of Focus Distribution, Pitch Accent and Lexical Stress on the Temporal Organization of Syllables in Dutch,'' Phonetica **52**, 71–89.
Sluijter, A. M. C., and Heuven, V. J. van (**1996**). ''Spectral balance as an acoustic correlate of linguistic stress,'' J. Acoust. Soc. Am. **100**, 2471–2485.
Studebaker, G. A. (**1985**). ''A ''rationalized'' arcsine transform,'' J. Speech Hear. Res. **28**, 455–462.
Sweet, H. (**1906**). *A Primer of Phonetics* (Clarendon, Oxford).
Terken, J. M. B., and Nooteboom, S. G. (**1987**). ''Opposite effects of accentuation and deaccentuation on verification latencies for 'given' and 'new' information,'' Language Cogn. Process. **2**, 145–163.
Turk, A. E., and Sawush J. (**1995**). ''The domain of the durational effects of accent,'' MIT Speech Comm. Group Working Papers **X**, also J. Phon (to appear).
Wightman, C. W., Shattuck-Hufnagel, S., Ostendorf, M., and Price, P. J. (**1992**). ''Segmental durations in the vicinity of prosodic phrase boundaries,'' J. Acoust. Soc. Am. **91**, 1707–1717.