Spectral balance as an acoustic correlate of linguistic stress

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

(Received 20 December 1994; revised 16 April 1996; accepted 29 April 1996)

Although intensity has been reported as a reliable acoustical correlate of stress, it is generally considered a weak cue in the perception of linguistic stress. In natural speech stressed syllables are produced with more vocal effort. It is known that, if a speaker produces more vocal effort, higher frequencies increase more than lower frequencies. In this study, the effects of lexical stress on intensity are examined in the abstraction from the confounding accent variation. A production study was carried out in which ten speakers produced Dutch lexical and reiterant disyllabic minimal stress pairs spoken with and without an accent in a fixed carrier sentence. Duration, overall intensity, formant frequencies, and spectral levels in four contiguous frequency bands were measured. Results revealed that intensity differences as a function of stress are mainly located above 0.5 kHz, i.e., a change in spectral balance emphasizing higher frequencies for stressed vowels. Furthermore, we showed that the intensity differences in the higher regions are caused by an increase in physiological effort rather than by shifting formant frequencies due to stress. The potential of each acoustic correlate of stress to differentiate between initial- and final-stressed words was examined by linear discriminant analysis. Duration proved the most reliable correlate of stress. Overall intensity and vowel quality are the poorest cues. Spectral balance, however, turned out to be a reliable cue, close in strength to duration.

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

INTRODUCTION

A fair number of the languages of the world employ a structural parameter called stress. Stress is a structural, linguistic property of a word that specifies which syllable in the word is, in some sense, stronger than any of the others. An important topic of phonetic research has always been the acoustical and perceptual characterization of the properties by which the stressed syllable distinguishes itself from the unstressed syllables surrounding it (syntagmatic comparison) or, in a more controlled approach, how a stressed realization of a syllable differs from an unstressed realization of the same syllable (paradigmatic comparison).

In this article we will be concerned with the characterization of linguistic stress in Dutch, a ''stress-accent'' language, as is English (Beckman, 1986). Stress-accent differ from nonstress-accent languages such as Japanese, in that pitch accents are not only characterized by a pitch movement but also by other phonetic correlates such as greater duration and loudness (Beckman, 1986).

In stress-accent languages, a speaker may present a word as communicatively important by realizing a pitch accent on the prosodic head of that word by executing a prominence-lending pitch movement (a rise, fall, or combination of the two). The prosodic head within the word is the stressed syllable. For this reason pitch movement has always been advanced as the most important phonetic correlate of linguistic stress. In line with other theoretical and empirical work, i.e., Vanderslice and Ladefoged (1972), Huss (1978), Pierrehumbert (1980), Beckman and Edwards (1994), and others, we take the view, however, that this is not necessarily the most insightful analysis of the phenomenon. Pitch movement is the correlate of accent, rather than of lexical stress:

''In short utterances, however, pitch excursions are more likely to be interpreted in terms of the sequence at a nuclear accent, as in Fry's 1958 experiment showing the salience of the $F0$ contour in cueing stress in pairs as *pérmit* versus *permít*. This is probably the major source of the common misunderstanding in the experimental literature that $F0$ excursion is a direct acoustic correlate of the feature ''stress,'' a misunderstanding that has been incorporated into several standard textbooks, (...)'' (Beckman and Edwards, 1994, p. 13).

Beckman and Edwards (1994) present English prominence as a unidimensional system with four qualitative levels: the highest stress occurs on a syllable with a full vowel bearing a nuclear pitch accent, the second highest stressed syllables contain a full vowel with a nonnuclear pitch movement, the next highest stressed syllables contain a full vowel with no pitch movement, and the lowest level (i.e., unstressed) syllables are reduced.

We, however, argue that stress and accent are distinct (though nonorthogonal) dimensions: syllables in a word are either stressed or unstressed. Accentuation is used to focus and is determined by the communicative intentions of the speaker, i.e., accentuation is dependent on language behavior. Stress is a structural, linguistic property of a word that specifies which syllable in the word is the strongest. In our view ''stressed'' refers to syllables which are the potential docking sites for accent placement. They have an accent-lending pitch movement associated with them when they occur within a single word in a narrow focus. In our view, stress is therefore determined by the language system, and accent by language behavior. The positions of stressed syllables in Dutch (and English) are to a certain extent predict-

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

able by quantity-sensitive rules, any remaining exceptions are marked in the lexicon (dictionary) as receiving lexical stress (Langeweg, 1988; Kager, 1989).

Stressed vowels always have full vowel quality. Unstressed vowels in Beckman and Edwards' system are always reduced. The amount of reduction, however, depends on the context in which the vowel is uttered (van Bergem, 1993) and probably also on the language: (American) English is claimed to be more sensitive to vowel reduction than Dutch.

Any syllable can be accented so as to express focus by placing a pitch movement on it (Sluijter and van Heuven, 1995). Moreover all the syllables in a word containing an accented syllable are linearly expanded in time (Nooteboom, 1972; Eefting, 1991); expansion of nonaccented syllables is even found when a pitch accent is executed on a lexically nonstressed syllable in narrow focus (Sluijter and van Heuven, 1995). Whether these effects are language specific or not is not a topic of this article. The crucial fact is that we agree with Beckman and Edwards that studies of phonetic correlates of stress in both English and Dutch may yield contradictory results if there is no systematic control for the levels of stress hierarchy involved.

If a word in a stress accent language remains unaccented, the stressed syllable can still be distinguished, both perceptually (van Heuven, 1988) and acoustically (van Heuven, 1987; Sluijter *et al.*, 1995), by a combination of longer duration, greater loudness, and full phonetic quality (i.e., absence of spectral reduction). In the older literature (Sweet, 1906; Bloomfield, 1993), stress in languages such as Dutch and English was often referred to as dynamic stress, as opposed to melodic stress, indicating that its primary phonetic correlate was greater loudness. Indeed, greater acoustical intensity has been consistently reported as a reliable correlate of stress (cf. Lea, 1977; Rietveld, 1984; Beckman, 1986; Slootweg, 1987). In all these studies, however, stressed syllables were also accented, so that the greater intensity is caused by the larger amplitude of voicing (cf. Sluijter *et al.*, 1995). When overall intensity was varied in artificial speech, it inevitably proved a weak stress cue, much weaker than duration (Fry, 1955; van Katwijk, 1974), and only marginally stronger than vowel quality (Fry, 1965).

We asked ourselves whether overall intensity would still provide a reliable acoustic correlate of stress if the target word/syllable were not pronounced with an accent-lending pitch movement. Of course, we need not be surprised if intensity variations should turn out to provide only a marginal stress cue. In fact, it would seem to us that intensity variation will never have communicative significance for the simple reason that intensity is too susceptible to noise. If the speaker accidentally turns his head, or passes a hand before his mouth, intensity 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 seems ill-advised. The reason why it was used in the classical studies by Fry (1958, 1965) must have been that there were simply no alternatives available for investigating the role of loudness in stress perception.

We would like to defend the view that the older literature was essentially correct when it suggested loudness as a correlate of stress. Loudness is a subjective property of a sound that allows a listener to rank sounds along a weak–strong scale running from faint (barely audible) to blaring (nearly deafening) (Green, 1976, p. 278). The subjective impression of loudness corresponds with greater acoustic intensity as well as with distribution of intensity over the spectrum.

Crucially, intensity in the mid-frequency range contributes more to perceived loudness than intensity above 5 kHz and, especially, below 0.5 kHz (Handel, 1989, pp. 66–67). We also know that perceived loudness of a speech sound corresponds with the amount of effort that a speaker spends in producing it (Brandt *et al.*, 1969; Glave and Rietveld, 1975). There is little debate, even today, that stress is produced by expending more effort in the production of a syllable, whether at the pulmonic, glottal, or articulatory stage (Ladefoged, 1967, 1971). Effort was suggested as a physiological correlate of linguistic stress almost a hundred years ago by Sweet (1906, pp. 47, 49). Essentially the same view was expressed later by Bloomfield (1933, pp. 110–111).

Although these views are largely correct, they were wrong in one important respect. When more effort is expended in speech production, the result is not just greater amplitude of the (glottal) waveform, although this is certainly part of it. As we know from more recent studies, increased vocal effort generates a more strongly asymmetrical glottal pulse: the closing phase is shortened, such that the trailing flank of the glottal pulse is steep. As a result of this, there is a shift of intensity over the spectrum so that low frequency components are hardly affected that the intensity increase is concentrated in the higher harmonics only. Such differential effects of effort were reported by Glave and Rietveld (1975) and Gauffin and Sundberg (1989), who all noticed that intensity below 500 Hz was not affected by effort (or even reduced), and that all extra intensity was located in the frequency region between 500 and 4000 Hz.

We also know, from the work by Zwicker and Feldtkeller (1967), that overall intensity is certainly not the only acoustic correlate of loudness. These authors show, quite elegantly, that perceived loudness can be predicted by integrating intensity within specific frequency bands (critical bands), and then calculating a weighted sum across the critical bands. Crucially, the energies in the low frequency bands add little to perceived loudness, while the contribution of the higher bands is much stronger.

Simplifying this to some extent, we suggest that the acoustical correlate of greater physiological effort is a decrease of negative spectral tilt, or even a positive tilt. A relatively rising spectrum, in turn, is associated with greater loudness, so that the traditional claim of loudness as a perceptual cue for stress seems justified. If this line of reasoning is accepted, it follows that measuring overall intensity is not the only valid operationalization of increased physiological effort; we should at least consider intensity distribution, or spectral tilt, as well. Spectral tilt, in contradistinction to overall intensity, is not easily obscured by environmental factors, so that this operationalization of greater vocal effort seems communicatively more robust than overall intensity.

In a production experiment we therefore examine the intensity distribution of stressed and unstressed vowels in four contiguous frequency bands. We expect that the intensity in the higher frequencies of the spectrum of a stressed syllable increases more than the intensity in the lower frequencies as the stressed syllable is produced with greater vocal effort than its unstressed counterpart.

However, before concluding that differences in intensity in the higher regions are caused by increased physiological effort due to stress, we have to take alternative explanations into account. It is often possible to attribute intensity shifts to the effect of stress on formant frequencies. In order to disentangle the possibly confounded effects of stress on vowel quality, i.e., formants shifting to more extreme positions along their respective continua (Rietveld and Koopmans, 1987), and that on spectral slope, we will measure both types of parameter, and proceed by showing that the intensity increase in the upper frequency bands cannot reasonably be the result of formant frequency shifts.

In the past decades a great deal of research has been directed towards the acoustical realization of stress (e.g., Fry, 1955, 1965; Lehiste and Peterson, 1959; Lehto, 1969; Adams and Munro, 1978; Berinstein, 1979) and the relative strength of these parameters in separating stressed from unstressed tokens (Rietveld, 1984; Beckman, 1986). However, at the moment it seems that much of this research suffers from covariation of accent and stress.[1] Moreover, no one seems to have compared *all* the acoustical correlates of linguistic stress including spectral tilt and vowel quality. In this article we will therefore study the already known acoustic correlates of linguistic stress as well as the proposed new correlate: spectral balance. We predict that a combination of the higher octave filter levels should yield a more successful separation of stressed and unstressed tokens than overall intensity and vowel quality. Whether spectral balance should be a better correlate of linguistic stress than duration will be answered on the basis of our results.

In this study, we ask the following concrete research questions: (1) Is overall intensity still a reliable acoustic correlate of linguistic stress when possible confounding with high $F0$ due to accent is undone? (2) Are intensity differences as a function of stress mainly located in the higher regions of the spectrum? (3) Are the intensity differences in the higher regions caused by an increase in physiological effort rather than by shifting formant frequencies due to stress? Finally, the last specific question is: (4) To what extent can each acoustic correlate of stress be used to differentiate between initial-stressed and final-stressed words?

In order to answer these questions, a production study was carried out in which we examined syllable duration, overall intensity, intensity distribution (as a measure of spectral balance), and formant frequencies (as an acoustic correlate of vowel quality) of stressed and unstressed vowels spoken by four males and six females with and without an accent, using a single Dutch minimal stress pair and its reiterant-speech copy.

We will not be concerned with the measurement of fundamental frequency since we take the view that pitch movements are the correlate of accent rather than of stress. It is possible that stress (in the sense of force of articulation) might induce some minor and unreliable $F0$ changes; however such $F0$ changes can and should be distinguished from deliberate (macrointonational) uses of pitch.

I. METHOD

A. Material

We selected the Dutch minimal stress pair *canon-kanon* /kaːnɔn/-/kaːˈnɔn/ ''cannon''—''canon'' differing in stress position only. We also used the reiterant version of this word pair (repetition of the same syllable) where each syllable was replaced by the syllable *na* yielding nonsense words: /ˈnaːnaː/-/naːˈnaː/. Reiterant speech allows us to study prosodic phenomena while abstracting from segmental influences (Liberman and Streeter, 1978; Nakatani and Schaffer, 1978). The vowel /aː/ was chosen because it is the most open, longest vowel in Dutch. This vowel has the highest $F1$ value of all Dutch vowels, resulting in the largest distance between $F0$ and $F1$ (Pols *et al.*, 1973).

The target words were embedded in prefinal position in a carrier sentence: *Wil je* [target] *zeggen* /ʋɪl jə [target] zɛχə(n)/ ''Will you [target] say.'' Targets were spoken with and without a pitch movement on the stressed syllable.

B. Subjects and procedure

The resulting four stimulus types (2 stress positions * 2 accent conditions) with their reiterant versions were read eight times each by six male and six female speakers. The speakers were individually recorded on audio tape in a sound insulated booth, using a Sennheiser MKH-416 directional condenser microphone and a Revox B77 MKII tape recorder. The subject's head was strapped to a headrest to ensure a constant distance between mouth and microphone.

Stimulus sentences were presented in Dutch orthography (i.e., not in phonetic symbols) in two different counterbalanced random orders on a computer monitor that was placed inside the booth in front of the subject. The condition with the target outside focus (henceforth [−F]) was realized by placing a single (contrastive) accent on the last word of the sentence: *zeggen*. In the other focus condition (henceforth [+F]) a single accent was placed on the stressed syllable of the target, placing the target in focus. The syllable to be accented appeared in capitals on the monitor. When without an accent on the target, the intended stress pattern was indicated in bold face. In the instructions it had been pointed out to the speakers that the word containing the capitalized (accented) syllable was to be interpreted as expressing a narrow focus contrast with another word within the same semantic domain, as follows:

condition with an accent on the target
Wil je KAnon zeggen (en niet liedje)
''Will you canon say (rather than song)''
Wil je kaNON zeggen (en niet geweer)
''Will you cannon say (rather than rifle)''

condition without an accent on the target
Wil je **ka**non ZEGgen (en niet opschrijven)
''Will you canon say (rather than write down)''
Wil je ka**non** ZEGgen (en niet opschrijven)
''Will you cannon say (rather than write down).''

Each lexical stimulus was followed by a reiterant stimulus with exactly the same accent and stress pattern. Subjects always produced lexical and reiterant versions of each stimulus in immediate succession before going on to the next stimulus.

Each stimulus type was presented four times (orders 1, 2, 3, and 4) in the first part of the reading session and four more times (orders 5, 6, 7, and 8) in the second part of the reading session. After each stimulus, whether lexical or reiterant, a 5-s pause was observed, during which interval the subject inhaled prior to initiating the next utterance.

Accents were realized as prominence-lending rise–fall pitch movements on the appropriate syllable (configuration 1 & A in 't Hart *et al.*, 1990). Two phonetically trained listeners (i.e., the present authors) verified the location and the realization of the accents. There was no disagreement on this point. One of the male speakers realized accents on all the target words in the [−F] condition. Another male speaker could not read aloud in a satisfactory way. These speakers were excluded from further analysis, leaving four male and six female speakers.

C. Data analysis

The 640 utterances (2 stress positions * 2 focus conditions * 2 versions [i.e., lexical versus reiterant] * 10 speakers * 8 repetitions) were digitized (10 kHz sampling frequency, 4.8 kHz low-pass filtering, 12-bit amplitude resolution) on a VAX/VMS computer. The maximum amplitude range was utilized by normalizing the output levels for each individual speaker.

We selected four repetitions (orders 2, 3, 6, and 7) yielding 320 utterances for further research. This was done to remove item initial and final effects, since the eight repetitions were presented in blocks of four stimuli. Only if one of these realizations were affected by hesitation, mispronunciation, or incorrect accentuation, was it replaced by one of the other realizations (orders 1, 4, 5, or 8).

1. Vowel quality

Formant frequencies were determined by analyzing the digital waveform of male speakers into 10 LPC coefficients (25.6-ms analysis window, 10-ms time shift). The filter was calculated in coefficients of a cascade of second-order filters. These coefficients were sorted and forced to be complex conjugate (resonating) pairs, yielding five spectral peaks (arguably formants). All vowel quality and intensity measurements for both stressed and unstressed vowels were determined at the point in the vowel where the $F1$ reached its maximum. It was sometimes difficult to determine this maximum adequately in the syllable *non*, in which case we used the temporal midpoint of the syllables. The same procedure was followed for female speakers, this time analyzing the waveform into eight LPC coefficients, yielding four spectral peaks. In addition, formant frequencies were estimated by locating the strongest harmonic of the formants in a fast Fourier transform (FFT) spectrum. Both values for each formant were compared, and if they were within ± 1 interharmonic distance, the value determined by the former method was used; if they did not agree, the value taken from the FFT spectrum was used. In some cases it was impossible to determine a reliable value for $F1$, mostly for female speakers because of interference of $F1$ with $F0$. Unreliable $F1$ measurements were excluded from further data processing.

2. Spectral level

Intensity was measured in four contiguous frequency bands B1–B4: 0–0.5, 0.5–1.0, 1.0–2.0, and 2.0–4.0 kHz. The spectrum level of a frequency band was defined as the base-10 logarithm of the summed power (squared amplitude) Fourier coefficients in that frequency band relative to the maximum output level of the VAX/VMS analog-digital (AD) converter (12 bits, 10 kHz) which we defined as 60 dB. Following Gauffin and Sundberg (1989), the lowest band was chosen such that it included the fundamental frequency. The second, third, and fourth bands were chosen such that these bands included $F1$, $F2$, and $F3$, respectively. The mean fundamental frequency of male and female speakers varied between 100 and 400 Hz. The frequency value of the first three formants of /aː/ are $F1$: 750 Hz, $F2$: 1300 Hz, and $F3$: 2500 Hz for male speakers (Pols *et al.*, 1973) and 986, 1443, and 2778 Hz, respectively, for female speakers (van Nierop *et al.*, 1973). However, we have to be aware of the fact that $F1$ and $F2$ of the vowel /ɔ/ both fall within B2 (400 and 900 Hz, respectively, for male speakers and 578 and 933 Hz for female speakers). When it was not possible to base our conclusions on the findings of the lexical data, we based them on the findings of the reiterant word pair.

3. Duration

Syllable durations of the target words were measured using the high resolution waveform editor SESAM (Broeder, 1990). Segmentation boundaries were determined in a straightforward fashion by the visual criteria described by Van Zanten *et al.* (1991).

4. Overall intensity

The overall intensity of the stressed and unstressed vowels of each word was defined as the base-10 logarithm of the summed power (squared amplitude) Fourier coefficients between 0 and 5 kHz relative to the maximum output level of the VAX/VMS AD converter (12 bits, 10 kHz).

D. Statistical analysis

To examine the significance of the effects of stress and focus condition (accent) on syllable duration, overall intensity and spectrum levels (reiterant speech only), we ran three-way analyses of variance for both lexical and reiterant speech data with focus condition, syllable position, and stress as fixed effects and with repetition and speaker as repeated measures. We ran two more analyses of variance on the spectrum levels of *ka* and *non* with focus condition and stress as fixed effects. We did not include the sex of the

TABLE I. Mean syllable duration (in ms) of the first (σ1) and second (σ2) syllables of initial and final stressed *kanon* (lexical) and *nana* (reiterant). Standard deviations are presented in parentheses. The differences in duration between the stressed and unstressed syllables are presented syntagmatically (ΔS) and paradigmatically (ΔP). The data are presented per focus condition (in focus: [+F], outside focus: [−F]; stressed syllables are in bold face).

		Lexical			Reiterant		
Focus	Stress	σ1	σ2	ΔS	σ1	σ2	ΔS
[+F]	Initial	**254** (33)	233 (40)	21	**261** (37)	209 (43)	52
	Final	151 (24)	**278** (37)	127	162 (26)	**289** (40)	127
	ΔP	103	45		99	80	
[−F]	Initial	**227** (30)	214 (41)	13	**235** (28)	190 (37)	45
	Final	142 (22)	**262** (41)	120	157 (25)	**260** (37)	103
	ΔP	85	48		78	70	

TABLE II. Mean overall intensity (in dB) of the first (σ1) and second (σ2) syllables of initial and final stressed *kanon* (lexical) and *nana* (reiterant). Standard deviations are presented in parentheses. The differences in dB between the stressed and unstressed syllables are presented syntagmatically (ΔS) and paradigmatically (ΔP). The data are presented per focus condition (in focus: [+F], outside focus: [−F]; stressed syllables are in bold face).

		Lexical			Reiterant		
Focus	Stress	σ1	σ2	ΔS	σ1	σ2	ΔS
[+F]	Initial	**47.8** (3.9)	43.9 (4.4)	3.9	**46.5** (3.6)	43.1 (3.4)	3.4
	Final	41.4 (4.9)	**47.5** (3.3)	6.1	44.3 (3.8)	**46.0** (3.1)	1.7
	ΔP	5.4	4.6		2.2	2.9	
[−F]	Initial	**44.9** (3.2)	43.1 (4.1)	1.8	**44.6** (3.2)	42.5 (3.9)	2.1
	Final	42.7 (3.5)	**44.6** (4.1)	1.9	44.4 (3.3)	**43.5** (3.7)	0.9
	ΔP	2.2	1.5		0.2	1.0	

speakers in the design of the analyses; although the spectral slopes of our male speakers were slightly more level than those of the female speakers, preliminary analyses revealed no interaction between the sex of the speaker and any of the linguistic factors (i.e., stress, focus, and lexicality). The effects of sex are only tangential to this study, but they will be dealt with briefly in Appendix A.

We ran three-way analyses of variance on formant frequencies for each syllable position and speech type separately with focus, stress, and sex as fixed effects and with repetition, syllable position, and speaker as repeated measures. Missing cases were excluded from the analyses. For all analyses included in this article we use an α of .05.

To determine how well these acoustic measures can be applied to determine the stress position of a word, we carried out linear discriminant analyses (LDA) for *ˈnana/naˈna* and for *ˈcanon/kaˈnon* for each focus condition separately. Discriminant analysis is primarily a data reduction method in which parameters are collapsed onto orthogonal discriminant functions so that the functions maximally separate the groups. Discriminant functions are linear combinations of weighted variables in which the standardized weights reflect the importance of the associated variables. In all analyses the stress positions functioned as groups: *ˈcanon* versus *kaˈnon*, with 40 data points (10 speakers $*$ 4 repetitions) per group.

The results are presented below in separate subsections for duration (Sec. II A), overall intensity (Sec. II B), formant frequencies (Sec. II C), and the intensity in the four separate filter bands (Sec. II E).

II. RESULTS

A. Duration

In Table I mean absolute syllable durations are broken down by speech type, focus condition, and stress position. The differences in duration between stressed and unstressed syllables were determined syntagmatically (differences within words) as well as paradigmatically (differences across words).

As can be seen in Table I, stressed syllables are longer than unstressed syllables [lexical: $F(1,318)=337.2$, $p<0.001$; reiterant: $F(1,318)=440.6$ $p<0.001$]. The presence or absence of an accent affects the duration of both stressed and unstressed syllables. Accented words [+F] have longer syllables than unaccented words [−F], in accordance with earlier findings of Eefting (1991), Sluijter (1992), and Sluijter and van Heuven (1995) [lexical: $F(1,318)=22.0$, $p<0.001$; reiterant: $F(1,318)=26.3$, $p<0.001$]; there were no significant interactions between focus and stress [lexical: $F(1,316)<1$; reiterant: $F(1,316)=3.7$, ns].

The differences between stressed and unstressed syllables in the initial stressed words are relatively small compared to the differences in final stressed words. Final syllables are longer than initial syllables due to preboundary lengthening (Klatt, 1976; Wightman *et al.*, 1992) [lexical: $F(1,318)=192.0$, $p<0.001$; reiterant: $F(1,318)=74.3$ $p<0.001$]. Due to the effect of stress and preboundary lengthening, the longest duration is found for a stressed final syllable, whereas the shortest duration is found for initial unstressed syllables. However, there is a (almost) significant interaction between syllable position and stress [lexical: $F(1,316)=39.0$, $p<0.001$; reiterant: $F(1,316)=3.3$, $p=0.072$], indicating that combined effects of stress and final lengthening are not completely additive. It has been suggested by others (e.g., Nooteboom, 1972; Klatt, 1976) that the effects of stress and preboundary lengthening are nonadditive, arguing that additive effects would lengthen a syllable beyond its ceiling duration.[2] There were no significant interactions between focus and syllable position [lexical: $F<1$; reiterant: $F(1,316)=1.2$, ns].

We examined the effectiveness of duration as an acoustic separator between initial and final stressed words for each focus condition separately. In a LDA in which the duration of syllables 1 and 2 were used as the predictors to separate *ˈkanon from kaˈnon*, 98% and 100% correct discrimination were reached for lexical and reiterant speech, respectively, in the [+F] condition. The results in the [−F] condition were almost identical, 99% correct grouping for both lexical and reiterant speech. This means that duration is a very robust acoustic correlate of stress, which remains stable despite the potential confounding influence of accent.

B. Overall intensity

In Table II means and standard deviations of the overall intensity data are summarized. The differences (in dB) be-

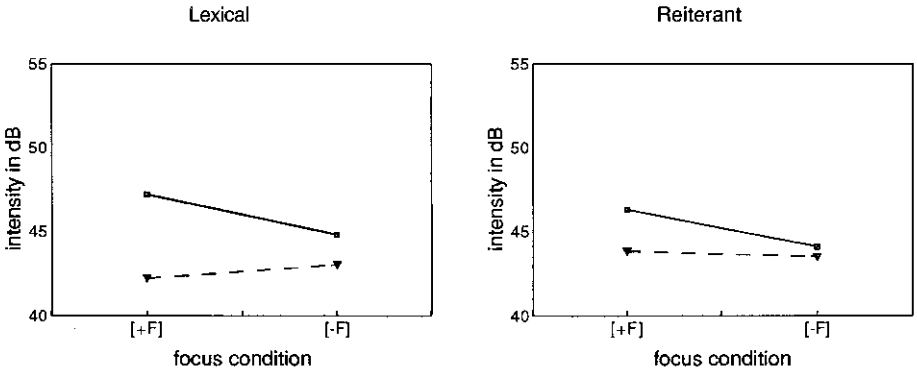

FIG. 1. Mean overall intensity (in dB) broken down by focus condition and stress (−stress is indicated by dashed lines, +stress by solid lines) for lexical speech data [left-hand panel] and reiterant speech data [right-hand panel].

tween stressed and unstressed vowels were determined syntagmatically and paradigmatically (cf. Sec. II A).

The initial syllables in the reiterant speech condition are somewhat louder than the final syllables in this condition [$F(1,318)=8.8$, $p=0.003$]. There were no other statistically significant main effects or interactions involving the factor syllable position [lexical $F(1,318)=1.7$, ns; all interactions $F<1$].

A difference of about 5 dB, determined both syntagmatically and paradigmatically, is found between stressed and unstressed vowels of *'canon* and *ka'non* in the [+F] condition. The differences between the stressed and unstressed vowels of *'nana* and *na'na* in this focus condition are about 3 dB. Outside the focus, there was only a slight difference of about 2 dB between stressed and unstressed vowels of *'canon* and *ka'non*, and an even smaller difference of about 1 dB between the stressed and unstressed vowels of *'nana* and *na'na*. Stress appeared to be significant for both lexical and reiterant speech [lexical: $F(1,318)=59.3$ $p<0.001$; reiterant: $F(1,318)=15.7$, $p<0.001$]. Focus only caused a significant effect for the reiterant speech data [lexical: $F(1,318)=3.1$, ns; reiterant: $F(1,318)=9.4$ $p=0.002$. Crucially, the interaction between focus and stress for both lexical and reiterant speech data is significant [lexical: $F(1,316)=12.9$, $p<0.001$; reiterant: $F(1,318)=5.9$, $p=0.015$]. The effects of stress on overall intensity in the [+F] condition are stronger than the effects in the [−F] condition. Figure 1 displays the relation between the factors stress and focus for overall intensity.

As can be seen in the right-hand part of Fig. 1, which shows reiterant speech data, only stressed syllables in the [+F] condition have a higher overall intensity. There is hardly any difference between stressed and unstressed vowels in the [−F] condition. Moreover, the intensity values for unstressed syllables are similar in the [+F] and [−F] conditions. The lexical speech data show a similar effect: there is only a slight difference between stressed and unstressed vowels in the [−F] condition, whereas there is a considerable difference between stressed and unstressed syllables in the [+F] condition. The overall intensity of unstressed vowels in the [−F] and [+F] conditions is virtually identical. We assume that this effect can be explained by the fact that in the [+F] condition a rise–fall configuration, marking the accent on the stressed syllable, is realized on the stressed vowel. This leads to a higher overall intensity of this syllable. One explanation can be given for this effect: each glottal pulse has a larger amplitude of voicing due to more speaker effort.

In a LDA in which the overall intensity of syllables 1 and 2 was used as the predictor to separate *'canon* from *ka'non*, 88% correct discrimination was reached in the [+F] condition, whereas only 69% correct discrimination was reached in the [−F] condition. The separation is even less clear for the reiterant speech data: 80% correct discrimination in the [+F] condition and only 63% correct discrimination in the [−F] condition. The effects are fully corroborated by the acoustical analysis, i.e., in the latter condition, where the effects of lexical stress were examined in the abstraction from the confounding accent variable, there is hardly any difference between the overall intensity of stressed and unstressed vowels. Overall intensity is therefore more likely to be an acoustic correlate of accent than of stress.

On the basis of these data, we should expect that there would be a high degree of uncertainty in listeners' judgments for the different stress positions in the [−F] condition if they have to infer the stress position of a word from overall intensity alone. In the [−F] condition, intensity is one of the remaining cues to determine the lexical stress position of the words since the accent marking pitch movement is absent from that syllable. Therefore, we expect other cues such as duration and possibly spectral balance to be more helpful in determining stress position.

C. Vowel quality

We performed three-way analyses of variance on each formant value for each syllable position and each speech type separately, with focus condition, stress, and sex as fixed factors, and with repetition and speaker as repeated measures.

Focus never had a significant main effect on any of the dependent variables $F1-F4$ [lexical: all cases $F<1$]. Moreover, there was no significant interaction involving the factor focus. We therefore decided to collapse the results over focus conditions.

In Table III the means (and standard deviations) of $F1-F4$ are summarized for each sex separately. The results are broken down for speech type and syllable position. As can be

TABLE III. Mean formant frequencies $F1-F4$ (in Hz) for stressed and unstressed vowels of the first ($\sigma1$) and second ($\sigma2$) syllables of lexical (*kanon*) and reiterant (*nana*) speech produced by four male (IIIa) and six female (IIIb) speakers.

a. **Male**			Vowel	−Stress		+Stress	
kanon	$\sigma1$	$F1$	/aː/	570	(49)	668	(44)
		$F2$		1276	(61)	1382	(82)
		$F3$		2238	(133)	2269	(110)
		$F4$		3366	(324)	3453	(253)
	$\sigma2$	$F1$	ɔ	326	(74)	361	(64)
		$F2$		829	(115)	811	(75)
		$F3$		2491	(315)	2496	(244)
		$F4$		3375	(309)	3322	(203)
nana	$\sigma1$	$F1$	/aː/	655	(59)	717	(60)
		$F2$		1390	(124)	1457	(128)
		$F3$		2617	(116)	2544	(125)
		$F4$		3711	(138)	3676	(162)
	$\sigma2$	$F1$	/aː/	665	(68)	702	(63)
		$F2$		1367	(82)	1440	(143)
		$F3$		2578	(155)	2556	(155)
		$F4$		3750	(202)	3708	(274)
b. **Female**			Vowel	−Stress		+Stress	
kanon	$\sigma1$	$F1$	/aː/	655	(68)	685	(67)
		$F2$		1632	(140)	1693	(120)
		$F3$		2657	(193)	2514	(253)
		$F4$		4028	(183)	3933	(216)
	$\sigma2$	$F1$	ɔ	476	(110)	488	(110)
		$F2$		1151	(107)	1117	(143)
		$F3$		2941	(304)	3016	(316)
		$F4$		4017	(217)	4086	(209)
nana	$\sigma1$	$F1$	/aː/	743	(90)	741	(90)
		$F2$		1664	(109)	1647	(102)
		$F3$		2768	(237)	2723	(208)
		$F4$		4088	(235)	3934	(194)
	$\sigma2$	$F1$	/aː/	767	(91)	765	(110)
		$F2$		1618	(99)	1636	(91)
		$F3$		2733	(259)	2722	(243)
		$F4$		3976	(201)	3974	(235)

seen in Table III, the formant frequencies of the female speakers are always higher than those of the male speakers. Sex causes a significant effect for all dependent variables in all analyses [all cases: $p \leqslant 0.001$]. $F1-F3$ of /aː/ in our data for male speakers roughly correspond to the values reported in the literature (Pols *et al.*, 1973). However, $F1$ of female speakers is somewhat lower than the value of 986 Hz reported by van Nierop *et al.* (1973), whereas $F2$ is somewhat higher than the reported value of 1443 Hz. These differences may well be caused by the fact that the consonantal context, /h (vowel) t/, used by van Nierop *et al.* (1973) differs from the consonantal context used in our experiment. $F3$ corresponds to the reported value.

Stress does not have a significant effect on the formant values of the /ɔ/ [all cases: $F(1,124)<1$]. There was also no significant interaction between the factors stress and sex for this particular vowel [all cases: $F(1,122)<1$].

Different results are obtained for the vowel of the initial syllable *ka* in that speakers tend to lower $F1$ and $F2$ in unstressed *ka* and to raise $F3$ [$F1$: $F(1,156)=33.3$, $p \leqslant 0.001$; $F2$: $F(1,156)=19.8$, $p \leqslant 0.001$; $F3$: $F(1,156)=5.6$ $p=0.020$; $F4$: $F(1,156)<1$]. This means that this vowel in Dutch, when unstressed, changes to an [ɑ]-like quality (van Bergem, 1993). We found significant interaction between stress and sex for $F1$ [$F(1,153)=12.1$, $p=0.001$], $F3$ [$F(1,153)=7.5$, $p=0.007$], and $F4$ [$F(1,153)=5.4$, $p=0.021$]. Male speakers lower $F1$ more when producing unstressed syllables than female speakers do, which indicates that male speakers tend to open their mouth less producing unstressed vowels than producing stressed vowels, whereas females do not. Male speakers lower $F3$ and $F4$ when a syllable is unstressed, whereas female speakers raise these formants.

TABLE IV. Percentage correct discrimination reached in a linear discriminant analysis, with each formant separately ($F1-F4$) used as a predictor variable, and with all formant values together used as predictor variables (all). The results are presented for each speech type (lexical and reiterant) and for each focus condition ([+F] and [−F]) separately.

Focus	Formant	Lexical (%)	Reiterant (%)
[+F]	$F1$	63	56
	$F2$	60	56
	$F3$	57	56
	$F4$	47	59
	$F1-F4$ (all)	84	68
[−F]	$F1$	65	58
	$F2$	65	56
	$F3$	56	55
	$F4$	61	58
	$F1-F4$ (all)	77	71

The effect of stress on the formant values of the vowels in *nana* in the initial syllable was only significant for $F4$ [$F1$: $F(1,156)=3.4$, ns; $F2$: $F<1$; $F3$: $F(1,156)=3.5$, ns; $F4$: $F(1,156)=11.5$, $p=0.001$] and for the second syllable only for $F2$ [$F(1,156)=5.7$, $p=0.018$; all other formants: $F<1$]. We found two significant cases of interaction between stress and sex for $F1$ and $F2$ of the initial syllables [$F1$: $F(1,154)=6.1$, $p=0.014$; $F2$: $F(1,154)=5.1$, $p=0.025$]. As can be seen in Table III male speakers tend to lower the $F1$ and $F2$ of an unstressed syllable, whereas female speakers realize virtually the same formant values for stressed and unstressed vowels.

The results of the LDA were used to determine how well each formant performed as a predictor of stress position. Table IV summarizes the results for both word pairs. The percentage correct discrimination is presented for each focus condition separately.

As can be seen in Table IV, single formant values are poor indicators of stress position in all conditions. Results improve if we use them in a multiple prediction; the lexical tokens in particular can be separated reasonably well (84% and 77%, respectively). This result can easily be explained by the fact that the vowel quality of the /aː/ in the initial stressed ˈ*canon* shifted towards [ɑ] in the final stressed *ka*ˈ*non* (cf. van Bergem, 1993).

D. Covariation of voice intensity and articulation

There is a possible covariation of voice intensity and properties of the filter. This is related to the finding that speakers, when talking louder, tend to use more open articu-

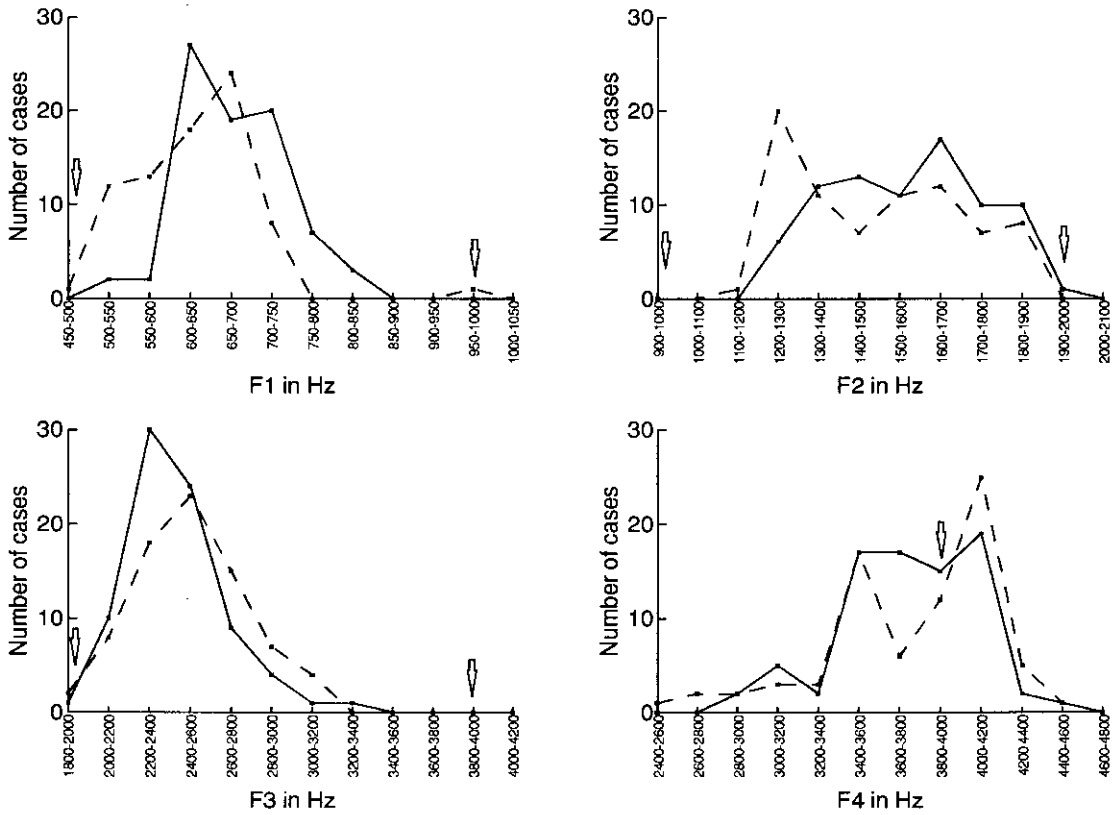

FIG. 2. An overview of the distribution of $F1$, $F2$, $F3$, and $F4$ of the 80 (2 focus conditions * 10 speakers* 4 repetitions) unstressed realizations (dashed line) and the stressed realizations (solid line) of the vowel /a:/ in the syllable /ka:/. The boundaries of frequency bands are indicated by arrows.

lation (van Son and Pols, 1990). These changes will also affect the spectral balance. We measured formant frequencies not only to determine the strength of vowel quality as an acoustic correlate of stress but also to determine their influence on the spectral balance. As described in Sec. II C, spectral levels were determined by measuring the intensity in four nonoverlapping contiguous frequency bands B1–B4: 0–0.5, 0.5–1.0, 1.0–2.0, and 2.0–4.0 kHz. Following Gauffin and Sundberg (1989), the lowest band was chosen so that it included the fundamental. The second, third, and fourth bands were chosen so that these bands included $F1$, $F2$, and $F3+F4$, respectively. We wanted to determine to what extent our speakers realized formant frequencies that fall within these four frequency bands. Figures 2, 3, and 4 present an overview of the distribution of the formant data for stressed and unstressed syllables collapsed over sex and focus condition. *Ka* and *non* are presented in Figs. 2 and 3, respectively. The reiterant speech data, collapsed over syllable positions, are presented in Fig. 4. The boundaries between the different frequency bands are marked in Figs. 2–4 by arrows.

As can be seen in Fig. 2, the peaks of the formant distributions remain well within the designated filters, but there is a considerable shift of the gravitational point of stressed tokens relative to unstressed tokens for both $F1$ and $F2$ and a slight spillover of $F1$ into the base band. In Fig. 3, showing *non* data, there is a considerable spillover of $F1$ into the base band and of $F2$ into B1. The distribution of $F1$ of stressed tokens shifts upwards, whereas the distribution of $F2$ shifts downwards. In Fig. 4, we only observe a very slight shift upwards for both $F1$ and $F2$. The only shift in distribution of $F3$ is found for the *ka* data. In all other cases, the distribution of both $F3$ and $F4$ does not shift. The $F1$, $F2$, and $F3$ data of both stressed and unstressed vowels in *ka* in *kanon* and *na* in *nana* do indeed largely fall within the designated frequency bands. However, as can be seen in Fig. 3, a part of the distribution of both $F1$ and $F2$ of the vowel /ɔ/ falls within B2. Due to the shift of $F1$ and $F2$ towards higher frequencies, possible differences in the spectral level are caused partly by differences in formant frequencies. We determined the influence of the shift of $F1$ on the amplitude of $F2$ (A2) using Eq. (1) and the influence of the shift of both $F1$ and $F2$ on the amplitude of $F3$ (A3) using Eq. (2) (Fant, 1960; Stevens, 1994):

$$\Delta \mathrm{A2(in\ dB)} = 40 \log \frac{F1_{-\mathrm{stress}}}{F1_{+\mathrm{stress}}} - 40 \log \frac{\sqrt{F2_{-\mathrm{stress}}{}^2 - F1_{-\mathrm{stress}}{}^2}}{\sqrt{F2_{+\mathrm{stress}}{}^2 - F1_{+\mathrm{stress}}{}^2}}, \quad (1)$$

$$\Delta \mathrm{A3(in\ dB)} = 40 \log \frac{F1_{-\mathrm{stress}} * F2_{-\mathrm{stress}}}{F1_{+\mathrm{stress}} * F2_{+\mathrm{stress}}}. \quad (2)$$

This allowed us to determine how much of the difference in spectral balance between stressed and unstressed syllables can be explained by the formant frequency shifts. Table V presents the mean differences in spectrum amplitude of $F2$ and $F3$ (A2 and A3, respectively) caused by the shift of

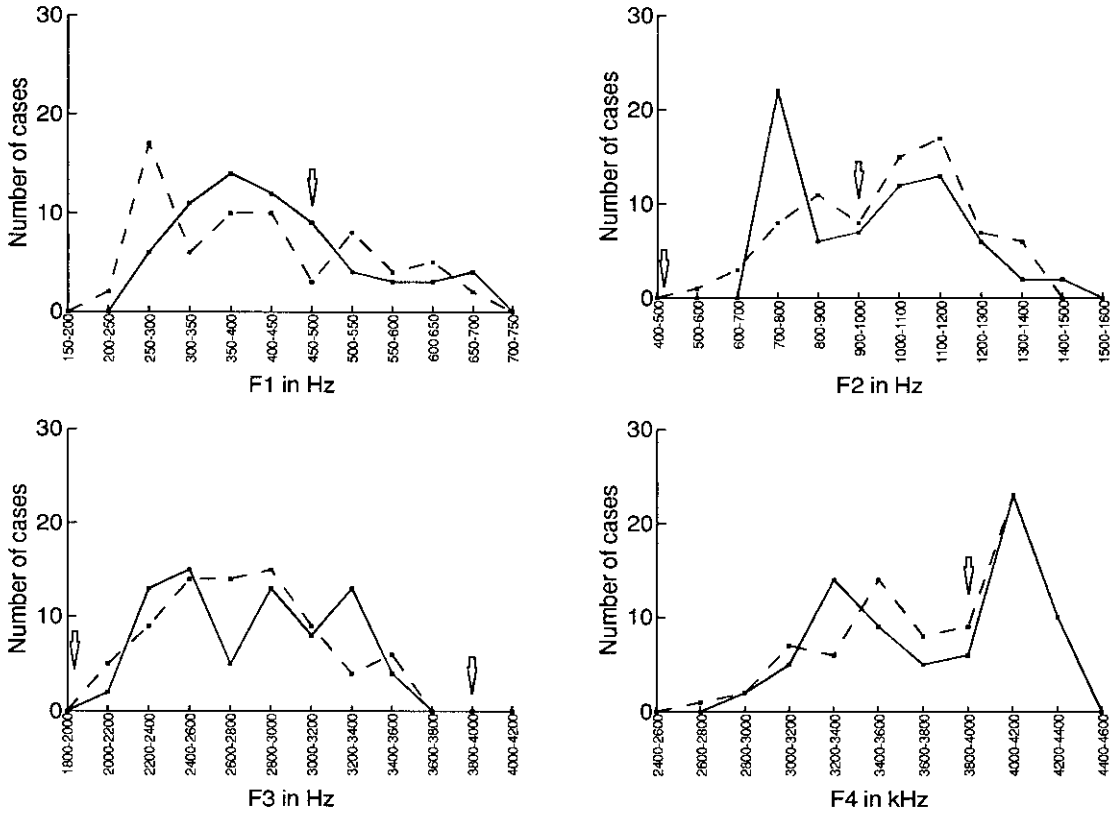

FIG. 3. An overview of the distribution of $F1$, $F2$, $F3$ and $F4$ of the 80 (2 focus conditions * 10 speakers * 4 repetitions) unstressed realizations (dashed line) and the stressed realizations (solid line) of the vowel /ɔ/ in the syllable /nɔn/. The boundaries of frequency bands are indicated by arrows.

either $F1$ or by both $F1$ and $F2$. The differences will be used to correct the raw filter levels in B1–B4 that influence formant shifts in $F1$ and $F2$.

As expected, the influence of the formant frequency shifts on the amplitude of $F2$ and $F3$ is negligible for the reiterant speech data. They are also negligible for the *non* data, as far as the influence of $F1$ and $F2$ on A3 is concerned. However, as mentioned above, the quality of the /aː/ in *ka* changed from /ɑ/ in the unstressed syllables to an /aː/ in the stressed syllables; this upward shift of F1 and F2 had considerable influence on the amplitudes of $F2$ and $F3$. In Sec. II E, we will correct the measured spectral level of stressed vowels for the influence of the vocal tract changes using Eqs. (1) and (2). In Eqs. (1) and (2) we used mean $F1$ and $F2$ values of the unstressed vowels in *ka*, *non*, and *na*, respectively, to correct the spectral levels of each individual stressed vowel. If formant values of a particular stressed syllable were missing, its spectral level was corrected by replacing the missing formant values by the mean value of the remaining three stressed realizations of that particular speaker and vowel in the same focus and speech condition.

E. Intensity differences in four contiguous filter bands

We hypothesized that the spectral level of a stressed syllable differs from its unstressed counterpart. We expected that the intensity in the higher part of the spectrum increases more than the intensity in the lower part when a syllable is stressed. We ran three-way analyses of variance on the (corrected) intensity levels in each filter band separately, henceforth B1–B4, with focus condition, stress, and syllable position as fixed effects, and with repetition and speaker as repeated measures for reiterant speech. For *ka* and *non*, we ran two-way analyses of variance for each syllable separately with focus and stress as fixed effects, and with repetition and speaker as repeated measures. Although we found a significant main effect of syllable position on the spectral levels of all the frequency bands [B1: $F(1,318)=7.9$, $p=0.005$; B2: $F(1,318)=6.8$, $p=0.009$; B3: $F(1,318)=4.9$, $p=0.027$; B4: $F(1,318)=23.5$, $p<0.001$], we did not find any significant interactions with the factor syllable position [focus * syllable position: all cases $F(1,316)<1$; stress * syllable position: B1 and B2: $F(1,316)<1$; B3: $F(1,316)=1.8$, ns; B4: $F(1,136)=3.1$, ns]. We therefore decided to collapse the reiterant speech data over syllable position in the following presentation of the data.

The spectral slopes of the stressed and unstressed vowels in our data are presented in Fig. 5, showing the spectra of the stressed and unstressed vowels in the reiterant and lexical speech data, on the basis of the mean intensity values (corrected and uncorrected for formant frequency shifts) in the four contiguous frequency bands: 0–0.5 kHz, 0.5–1 kHz, 1–2 kHz, and 2–4 kHz. The left-hand figures present the data in the [+F] condition; the right-hand figures present the [−F] data. In Appendix B the uncorrected means (and standard deviations) are summarized for both lexical and reiter-

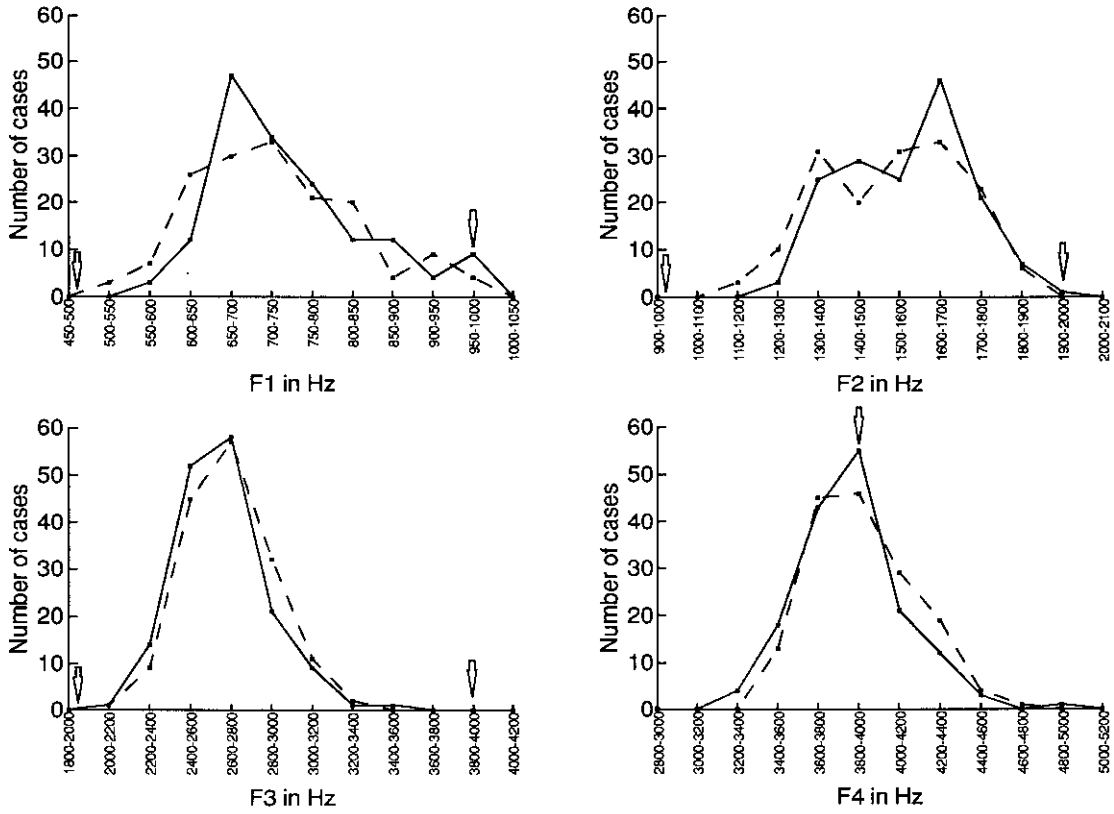

FIG. 4. An overview of the distribution of $F1$, $F2$, $F3$, and $F4$ of the 160 (2 focus conditions * 10 speakers * 4 repetitions * 2 syllable positions) unstressed realizations (dashed line) and the stressed realizations (solid line) of the vowel /a:/ in the syllable /na:/. The boundaries of frequency bands are indicated by arrows.

ant speech data for each filter band separately in Tables B I–B IV.

As can be seen in Fig. 5, the negative spectral tilt of unstressed vowels is steeper than that of stressed vowels. Accented, stressed vowels have a gentler negative spectral tilt than unaccented stressed vowels. The intensity in the lowest filter band is hardly affected by stress, whereas there are considerable intensity differences in the other three filter bands in both focus conditions.

Stress did not cause a significant effect on the intensity in the lowest frequency band of the reiterant speech data [$F(1,318)=2.9$, ns], but did exert a significant effect on the intensity in all the other frequency bands, with stressed syllables having more intensity in the higher frequency bands than unstressed syllables [all cases: $p\leqslant 0.001$]. For the lexical speech data, both *ka* and *non*, stressed syllables have more intensity in all the frequency bands, including the base band [*ka* B1: $F(1,158)=8.3$, $p=0.004$; B2: $F(1,158)=45.6$, $p<0.001$; B3: $F(1,158)=29.0$, $p<0.001$; B4: $F(1,158)=62.0$, $p<0.001$; *non*: B1: $F(1,158)=19.6$, $p<0.001$; B2: $F(1,158)=6.2$, $p=0.014$; B3: $F(1,158)=9.2$, $p=0.003$; B4: $F(1,158)=18.2$, $p<0.001$]. However, the *ka* data are comparable to the reiterant data, as can be seen in Fig. 5, by having the largest intensity differences in the highest three frequency bands. We explain the elevation of B1 of the *non* data by the fact that the $F1$ of the vowel in this syllable is located in the base band, whereas the $F1$ of the vowel in the other syllables with /a:/ is located in B2.

There is no difference in the intensity distribution over the four frequency bands between unstressed tokens in the [+F] and [−F] conditions (which makes sense, because focus affects only stressed syllables). It should be noted that the effects of stress on the filter levels (B2, B3, and B4) are clearly larger in [+F] tokens than in [−F] tokens. We found significant interaction between focus and stress in these bands for the *non* and the *na* data and in B3 for the *ka* data [*ka*: B2: $F(1,156)=3.4$, ns; B3: $F(1,156)=4.4$, $p=0.037$; B4: $F(1,156)=1.4$, $p=0.235$; *non*: B2: $F(1,156)=10.8$, $p=0.001$; B3: $F(1,156)=6.9$, $p=0.009$; B4: $F(1,156)=7.0$, $p=0.009$; *na*: B2: $F(1,316)=13.9$, $p\leqslant 0.001$; B3: $F(1,316)=20.4$, $p\leqslant 0.001$; B4: $F(1,316)=6.2$, $p=0.013$]. Therefore, these differences in spectral level due to stress are largely caused by the presence of a pitch movement. However, non-

TABLE V. Changes in dB of the spectrum amplitude of $F2$ and $F3$ (A2 and A3, respectively) caused by the shift of formant frequencies from unstressed to stressed vowels.

	Influence (dB) $F1$ on A2		Influence (dB) $F1$ and $F2$ on A3	
	[+F]	[−F]	[+F]	[−F]
ka	0.9	0.9	2.5	2.0
non	2.5	1.4	−0.4	0.2
na, $\sigma 1$	−0.3	0.2	0.8	0.4
na, $\sigma 2$	0.2	0.5	0.5	0.5

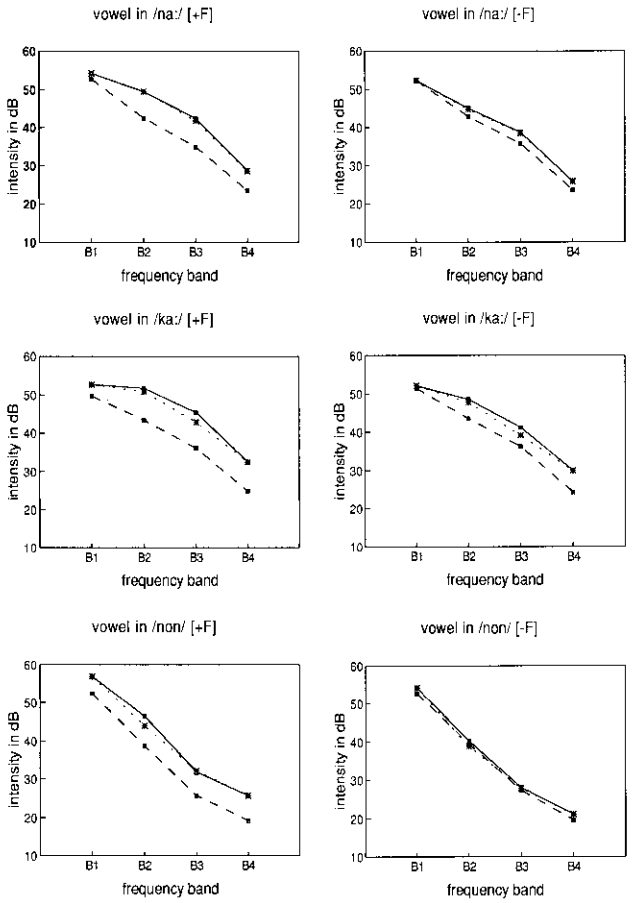

FIG. 5. Mean intensity (dB) of unstressed vowels (dashed lines) and stressed vowels (corrected values: dotted lines; uncorrected values: solid lines) in /na:/, /ka:/ and /nɔn/, respectively, for each focus condition separately: [+F] (left-hand side) and [−F] (right-hand side).

negligible effects of stress on spectral levels remain even in [−F] tokens for syllables containing /aː/. No effect of stress can be observed when the vowel is /ɔ/.

We conclude that, although there is an influence of the transfer function of the vocal tract on the spectral balance, voice source differences led to a difference in spectral balance.

To determine the capacity of intensity in different frequency bands as a predictor of stress position, we performed LDAs in which we used spectral levels in each frequency band as predictor variables, one by one, as well as simultaneously. We performed analyses on both the corrected and uncorrected values. We also ran analyses on the uncorrected values, because of the fact that these results could be of interest for applications in the field of speech recognition, whereas the results of corrected measures are of interest to those who are interested in the exact contribution of the voice source in the production of stress. Table VI summarizes the results for both word pairs. The percentage correct discriminations is presented for each focus condition separately.

As can be seen in Table VI the intensity in the lowest filter band, below 500 Hz, is the poorest indicator of stress position in all conditions. Results improve considerably if we use the intensity in the second, third, or fourth filter band.

TABLE VI. Percentage correct discrimination by linear discriminant analysis, with the intensity in each band separately used as predictor variables, and with the all intensity values together used as predictor variables (all). The results are presented for each speech type (lexical and reiterant), for each focus condition ([+F] and [−F]), and for corrected (C) and uncorrected (U) values separately.

		Lexical (%)		Reiterant (%)	
Focus	Frequency band	U	C	U	C
[+F]	B1	74	74	70	70
	B2	97	89	93	90
	B3	88	85	98	91
	B4	88	88	85	85
	all	99	94	100	96
[−F]	B1	61	61	58	58
	B2	73	61	79	66
	B3	74	58	83	66
	B4	78	78	73	73
	all	81	83	86	71

When we performed a LDA with four separate bands simultaneously as predictors, 100% correct grouping was reached in the [+F] condition, separating ˈ*nana* from *na*ˈ*na* (96% for corrected values). A 99% and a 94% correct grouping were reached by separating ˈ*kanon* from *ka*ˈ*non* in the same focus condition for uncorrected and corrected values, respectively. The same result is obtained if we omit the intensity in the base band as a predictor. This means that adding the base band does not lead to a significant improvement of the LDA.

The percentages of correct stress assignment for the tokens produced outside focus with uncorrected spectrum levels were 86% for *nana* and 81% for *kanon*, and 71% and 83% for the corrected values of *nana* and *kanon*, respectively. We conclude from these results that spectral balance is a clear acoustic correlate of stress and is even more reliable than overall intensity.

F. Comparing the strength of the four acoustic correlates of lexical stress

In Fig. 6 we compare the percentage correct discriminations by LDA for the four acoustic correlates of stress examined in the preceding sections.

It can be observed that in the [+F] condition vowel quality is the poorest correlate of stress. Spectral balance, operationalized as the intensity differences in different frequency bands after factoring out the effect of formant frequency shift, is a reliable correlate of stress, close in strength to duration. Overall intensity performs reasonably well in the [+F] condition. However, as was mentioned above, the higher overall intensity can be explained by the fact that in the [+F] condition a rise–fall configuration, marking the accent on the stressed syllable, is realized on the stressed vowel. Therefore, overall intensity is more likely to be an acoustic correlate of accent. Since this is in contrast to much earlier research on the acoustic realization of stress, we therefore examined the true correlates of stress without the confounding influence of accent by using speech data spoken without a pitch accent on the stressed syllable. Our results

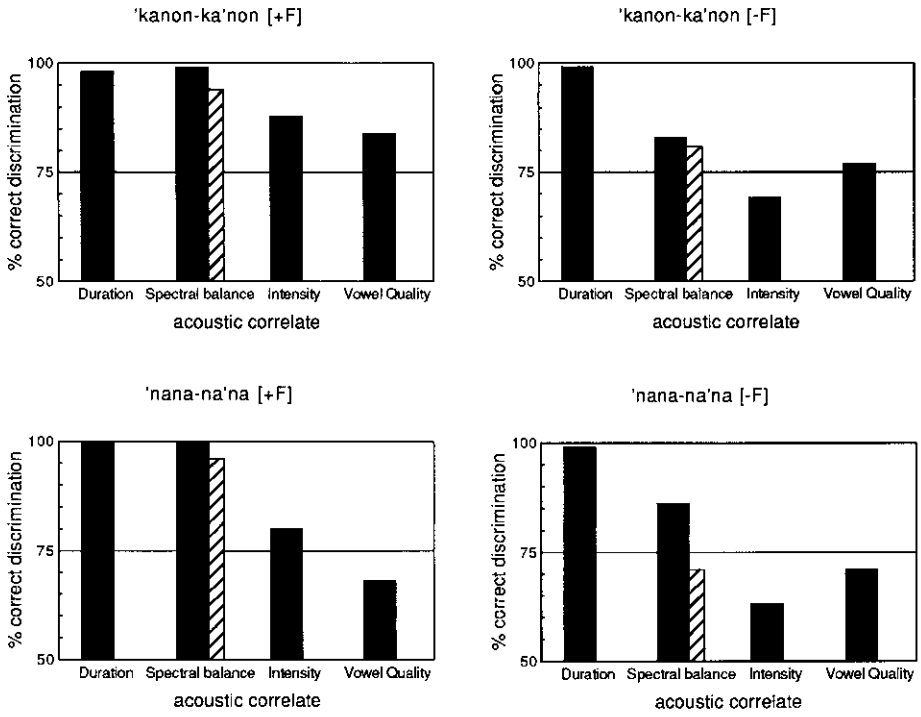

FIG. 6. An overview of the percentage correct discriminations for each acoustic correlate of stress. The upper parts present the results separating the lexical speech tokens; the lower parts present the results separating the reiterant speech tokens. Results are given for each focus condition separately: [+F] and [−F]. The percentages correct for the corrected spectral levels are presented by hatched bars; the uncorrected data by black bars.

show that the older literature was not correct in regarding overall intensity as a reliable acoustic correlate for stress. Overall intensity turned out to be the poorest correlate of stress position, even poorer than vowel quality. Duration remains the most stable acoustic correlate of stress position, but spectral balance also performs well in this condition and turned out to be the second best cue in stress assignment.

III. GENERAL DISCUSSION AND CONCLUSIONS

This study examined the acoustical correlates of stress and accent (other than pitch). Unlike earlier research on this topic, we measured the acoustical correlates of stress with and without the confounding effect of accent. We assumed that a pitch movement is a correlate of accent but not of stress. In this study, therefore, we investigated the acoustic correlates of stress in two conditions: with a pitch movement on the stressed syllable (condition [+F]) and without a pitch movement on the stressed syllable (condition [-F]).

The measurements of overall intensity supported our hypothesis that overall intensity is not a reliable correlate of stress. In the [-F] condition, in which no pitch accent was realized on the stressed syllable, there was hardly any difference between the overall intensity of stressed and unstressed vowels, whereas in the [+F] condition there was a considerable difference in overall intensity between stressed and unstressed vowels. A part of the rise of the rise–fall configuration marking the accent on the stressed syllable is realized on the stressed vowel, leading to a higher overall intensity of this syllable because of the fact that the pulses have a larger amplitude. Our finding limits the validity of earlier conclusions drawn by, e.g., Rietveld (1984) and Beckman (1986), who reported overall intensity as one of the most reliable acoustical means of stress to distinguish stressed from unstressed syllables. In these studies, however, stressed syllables were invariably accented so that the greater intensity is probably caused by the larger amplitude of the pulses. Our first research question (Is overall intensity still a reliable acoustic correlate of linguistic stress even without the possible confound of high *F*0?), therefore, has to be answered negatively.

Furthermore, we investigated spectral differences between stressed and unstressed vowels in order to answer our second research question (Are intensity differences due to stress mainly located in the higher regions of the spectrum?). As predicted, the results show that intensity differences between stressed and unstressed vowels are mainly concentrated in the three highest filter bands, above 0.5 kHz. Intensity in the higher bands (0.5–1, 1–2, and 2–4 kHz) was increased in stressed syllables by 5–10 dB, whereas the intensity in the lowest band was hardly affected at all.

These results are comparable to earlier findings by Glave and Rietveld (1975) on the effects of varying effort on spectral intensity distribution. They measured spectra of the vowel [ɛ] spoken with greater or lesser effort. The spectra of the vowel spoken with greater effort have more intensity in the higher-frequency region above 0.5 kHz and even show a decrease in intensity at the lower end of the frequency scale. With Glave and Rietveld (1975), we assume that the most important factor is probably the change of the source spectrum. We would argue that the increase in the higher part of the spectrum is caused by the more pulselike shape of the glottal source signal as the speaker expends more effort, nec-

essary to produce a stressed syllable. The glottal pulses of stressed and unstressed syllables may differ. These differences can arise because of the way the vocal folds and the glottis are configured during phonation. At a reduction of voice intensity, with a fixed location of all formants, the level of the harmonics situated at higher frequencies will decrease more than the level of harmonics at lower frequencies due to an increase in the negative slope of the source spectrum envelope. The relation between higher harmonics and the lowest ones strongly depends on the speed of glottal closure. The faster the glottis is closed, the more pulselike the excitation signal will be, resulting in a relatively flat harmonic spectrum. A more gradual pattern of glottal closure, as we assume to be the case for unstressed syllables, on the other hand, yields a steeper negative spectral slope, probably exceeding the 12-dB per octave rolloff that is often mentioned for the harmonic source spectrum (Fant, 1960; Childers and Lee, 1991).

However, the spectrum of a speech wave is not only influenced by the differences in voice source signal, since the intensity variations of a single harmonic or of a group of harmonics at a certain place along the frequency scale depends on both the source and the filter. There is a possible covariation of voice intensity and properties of the filter. This is related to the finding that speakers, when talking louder, tend to use more open articulations (van Son and Pols, 1990). These changes will also affect the spectral balance. The spectral peaks of a sound spectrum, i.e., the formants, reflect the resonances of the vocal tract. Formant frequencies and therefore the transfer function can change as a result of articulatory change, which affects the dimensions of the pharyngeal and the oral cavities (or as a result of nasal coupling). As a means of control for differences in the shape of the vocal tract between stressed and unstressed syllables and the influence of these differences on the spectrum, we compared formant frequencies of identical vowels in stressed and unstressed syllables. It is conceivable that speakers open their mouths more when producing stressed syllables than when producing unstressed syllables. The amount of mouth opening is reflected in the spectral tilt but counter to glottal sharpening it also directly influences the frequency of $F1$.

Our results show a difference in spectral balance between stressed and unstressed vowels, stressed vowels having more high-frequency emphasis than unstressed vowels. This difference is certainly not only due to differences in the shape of the vocal tract. The fact that open vowels tend to have higher formant frequencies when stressed can explain only part of the intensity increase in the higher-frequency bands. However, it was found that the effects of an upward shift of $F1$ and $F2$ on the spectral intensity levels are negligible for the reiterant speech data and quite small for the lexical speech data. We therefore conclude in answer to our third research question (Are the intensity differences in the higher regions caused by an increase in physiological effort in the laryngeal system rather than by shifting formant frequencies due to stress?), that the intensity differences in the higher-frequency bands between stressed and unstressed syllables are mainly caused by an increase in physiological effort rather than by differences in articulation.

Finally, we examined the potential of each acoustic correlate of stress to discriminate between initial-stressed words and final-stressed words. It turned out that duration is still the most effective correlate of stress, relatively unaffected by accent. Overall intensity and vowel quality are the poorest indicators of stress position. Spectral balance, however, seems to be a reliable cue even in the unaccented condition, close in strength to duration.

The following limitations should be considered in the interpretation of the results. First, the words that were investigated did not form a representative set of all words of Dutch. Only one disyllabic minimal stress pair was used. In further studies we will study multiple pairs of words and extend the scope of our study to other languages (Sluijter *et al.*, 1995). It is unclear at the moment to what extent vowel reduction plays a more important role to determine stress level in words with more than two syllables. Moreover, other languages, e.g., English, may be more sensitive to vowel reduction than Dutch.

In summary, the most important finding of this study is that spectral balance is an acoustic correlate of stress and that it can quite reliably distinguish stressed from unstressed tokens, irrespective of accent. Furthermore, as was mentioned in the Introduction, Zwicker and Feldtkeller (1967) showed that the energies in the low-frequency bands add little to perceived loudness, while the contribution of the higher bands is much stronger. Our results therefore suggest that the older literature, mentioned in the Introduction, was essentially correct when it referred to stress in languages such as Dutch and English as dynamic stress, as opposed to melodic stress, indicating that its primary phonetic correlate was greater loudness. A stressed syllable might be perceived as louder, and therefore more prominent, than an unstressed one due to the increased intensity levels in the higher part of the spectrum. Stress is not just a weaker degree of accent. One would expect to observe lower values along all measured correlates in stressed syllables of unaccented words. However, what we do observe is weakening along only those correlates that are related to the omission of the accent-lending pitch movement.

In subsequent research we have examined the perceptual relevance of the findings of the present study in an experiment in which we investigated the perception of stress position by manipulating vowel duration and intensity, the latter both in the classic way (i.e., uniform intensity differences) and in the more realistic way suggested by our production data (i.e., differences in higher bands only). These results will be presented in a separate article.

ACKNOWLEDGMENTS

Portions of this research were 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 K. N. Stevens, J. W. de Vries, S. G. Nooteboom, S. Shattuck-Hufnagel, G. Fant, D. R. Ladd, A. E. Turk, G. de Krom, R. Goedemans, J. Caspers, and one anonymous reviewer for ideas, discussion, and comments on earlier versions of this

article. Finally, thanks are due to J. Pacilly for the necessary programming and technical assistance.

APPENDIX A: EFFECT OF SEX ON SPECTRAL DISTRIBUTION OF INTENSITY

Speakers were normalized for overall intensity, since absolute differences in overall intensity across speakers were not controlled for in the recording procedures. Four-way analyses of variance were run on the intensity effects per filter band for lexical and reiterant speech data separately with focus (accent), stress, and sex as fixed factors. Speaker was nested as a random factor under sex, after randomly eliminating two female speakers from the data set in order to get the same number of speakers across sexes, as full orthogonality is required by this type of analysis. Female voices have a 1-dB greater intensity in the 0–0.5 kHz band, and a 2–3-dB weaker intensity in the higher frequency bands. These tendencies are in line with results reported for American English male versus female speakers (Holmberg *et al.* 1988; Sluijter *et al.*, 1995). However, in our data the main effects of sex are not significant for lexical nor for reiterant speech [lexical: B1: $F(1,6)=2.0$, $p=0.210$; B2: $F(1,6)<1$; B3: $F(1,6)=1.44$, $p=0.28$; B4: $F(1,6)=1.70$, $p=0.240$; reiterant: B1: $F(1,6)<1$; B2: $F(1,6)=1.38$, $p=0.285$; B3 and B4: $F(1,6)<1$]. Moreover, in the reiterant speech condition there were no significant interactions (second or higher order) involving sex. In the lexical speech condition out of all possible interactions involving the factor sex, only one (stress by sex) reached significance in one single frequency band [B3: $F(1,6)=21.3$, $p=0.004$].

On the basis of these results there was no need to incorporate sex as a factor in the final analysis of variance reported in Sec. II. By omitting sex there, we had the advantage that the data of all sex female speakers could be included in the analysis.

APPENDIX B

TABLE BI. Mean intensity (in dB) between 0 and 0.5 kHz of the first ($\sigma1$) and second ($\sigma2$) syllables of initial and final stressed *kanon* (lexical) and *nana* (reiterant). Standard deviations are in parentheses. The differences in dB between the stressed and unstressed syllables are presented syntagmatically (ΔS) and paradigmatically (ΔP). The data are presented per focus condition (in focus: [+F]; outside focus: [−F]).

Focus condition	Stress position	Lexical			Reiterant		
		$\sigma1$	$\sigma2$	ΔS	$\sigma1$	$\sigma2$	ΔS
[+F]	Initial	52.7 (3.5)	52.4 (4.5)	0.3	54.4 (3.5)	52.3 (3.7)	2.1
	Final	49.7 (5.3)	56.8 (3.7)	7.1	53.2 (4.2)	53.8 (3.1)	0.6
	ΔP	3.0	4.2		1.2	1.5	
[−F]	Initial	52.1 (3.1)	52.6 (4.2)	0.5	52.9 (3.4)	51.3 (4.3)	1.6
	Final	51.4 (3.5)	54.1 (4.2)	2.7	53.3 (3.7)	51.8 (4.2)	1.5
	ΔP	0.7	1.5		0.4	0.5	

TABLE BII. Mean intensity (in dB) between 0.5 and 1.0 kHz of the first ($\sigma1$) and second ($\sigma2$) syllables of initial and final stressed *kanon* (lexical) and *nana* (reiterant). Standard deviations are in parentheses. The differences in dB between the stressed and unstressed syllables are presented syntagmatically (ΔS) and paradigmatically (ΔP). The data are presented per focus condition (in focus: [+F]; outside focus: [−F]).

Focus condition	Stress position	Lexical			Reiterant		
		$\sigma1$	$\sigma2$	ΔS	$\sigma1$	$\sigma2$	ΔS
[+F]	Initial	51.7 (5.7)	38.6 (5.7)	13.1	49.7 (6.5)	41.5 (5.0)	8.2
	Final	43.3 (5.7)	46.4 (4.2)	3.1	43.5 (4.8)	49.2 (6.5)	5.7
	ΔP	8.4	7.8		6.2	7.7	
[−F]	Initial	48.6 (5.4)	39.5 (5.8)	9.1	45.6 (5.7)	41.5 (6.5)	4.1
	Final	43.5 (5.9)	40.2 (5.7)	3.3	44.2 (4.4)	44.7 (5.0)	0.5
	ΔP	5.1	0.7		1.4	3.2	

TABLE BIII. Mean intensity (in dB) between 1.0 and 2.0 kHz of the first ($\sigma1$) and second ($\sigma2$) syllables of initial and final stressed *kanon* (lexical) and *nana* (reiterant). Standard deviations are in parentheses. The differences in dB between the stressed and unstressed syllables are presented syntagmatically (ΔS) and paradigmatically (ΔP). The data are presented per focus condition (in focus: [+F]; outside focus: [−F]).

Focus condition	Stress position	Lexical			Reiterant		
		$\sigma1$	$\sigma2$	ΔS	$\sigma1$	$\sigma2$	ΔS
[+F]	Initial	45.3 (4.5)	25.6 (4.7)	19.7	42.6 (3.9)	33.6 (4.3)	9.0
	Final	36.1 (5.8)	31.7 (5.0)	−4.4	36.3 (4.1)	42.3 (4.6)	6.0
	ΔP	9.2	6.1		6.3	8.7	
[−F]	Initial	41.2 (4.5)	27.3 (3.9)	13.9	39.3 (4.1)	35.1 (5.2)	4.2
	Final	36.3 (5.2)	28.0 (5.4)	−8.3	36.5 (3.9)	38.5 (4.4)	2.0
	ΔP	4.9	0.3		2.8	3.4	

TABLE BIV. Mean intensity (in dB) between 2.0 and 4.0 kHz of the first ($\sigma1$) and second ($\sigma2$) syllables of initial and final stressed *kanon* (lexical) and *nana* (reiterant). Standard deviations are in parentheses. The differences in dB between the stressed and unstressed syllables are presented syntagmatically (ΔS) and paradigmatically (ΔP). The data are presented per focus condition (in focus: [+F]; outside focus: [−F]).

Focus condition	Stress position	Lexical			Reiterant		
		$\sigma1$	$\sigma2$	ΔS	$\sigma1$	$\sigma2$	ΔS
[+F]	Initial	32.5 (5.9)	19.2 (6.1)	16.3	29.1 (4.7)	21.6 (3.9)	7.5
	Final	24.8 (5.5)	25.6 (5.4)	0.8	25.4 (5.3)	28.0 (5.6)	2.6
	ΔP	7.7	6.4		3.7	6.4	
[−F]	Initial	30.0 (5.1)	19.6 (5.8)	11.4	27.0 (4.4)	21.9 (5.4)	5.1
	Final	24.3 (4.8)	21.1 (6.1)	−3.2	25.3 (3.7)	24.8 (5.2)	0.5
	ΔP	5.7	1.5		1.7	2.9	

[1]A notable exception is the research reported by Huss (1977).

[2]Beckman and Edwards (1988) show that preboundary lengthening and the presence versus absence of an accent have different influences on the syllable-internal organization. For preboundary lengthening the longer acoustic durations are associated with a disproportionate lengthening of the latter part of the vocalic gesture, whereas the presence of an accent is associated with a more even distribution of lengthening throughout the syllable. This means that the effects of preboundary lengthening and stress could be disentangled when the syllable-internal articulatory organization is studied in more detail.

Adams, C., and Munro, R. R. (**1978**). ''In search of the acoustic correlates of stress: Fundamental frequency, amplitude, and duration in the connected utterance of some native and non-native speakers of English,'' Phonetica **35**, 125–156.

Beckman, M. E. (**1986**). *Stress and Non-stress Accent* (Foris, Dordrecht).

Beckman, M. E., and Edwards, J. (**1988**). ''Articulatory timing and the prosodic interpretation of syllable duration,'' Phonetica **45**, 156–174.

Beckman, M. E., and Edwards, J. (**1994**). ''Articulatory evidence for differentiating stress categories,'' in *Phonological Structure and Phonetic Form. Papers in Laboratory Phonology III*, edited by P. A. Keating (Cambridge, U.P., Cambridge), pp. 1–33.

Bergem, D. van (**1993**). ''Acoustic vowel reduction as a function of sentence accent, word stress, and word class on the quality of vowels,'' Speech Commun. **12**, 1–23.

Berinstein, A. E. (**1979**). ''A cross-linguistic study on the perception and production of stress,'' University of California, Los Angeles Working Papers in Phonetics **47**.

Bloomfield, L. (**1993**). *Language* (Holt, Rinehart and Winston, New York).

Brandt, J. F., Ruder, K. P., and Shipp, I., Jr. (**1969**). ''Vocal loudness and effort in continuous speech,'' J. Acoust. Soc. Am. **46**, 1543–1548.

Broeder, D. G. (**1990**). ''Sesam handleiding; analyse en synthese van spraak,'' SPIN/ASSP report **24** (Speech Technology Foundation, Utrecht).

Childers, D. G., and Lee, C. K. (**1991**). ''Vocal quality factors: analysis, synthesis and perception,'' J. Acoust. Soc. Am. **90**, 2394–2410.

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.

Fant, G. (**1960**). *Acoustic Theory of Speech Production* (Mouton, The Hague).

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,'' Language and Speech **1**, 126–152.

Fry, D. B. (**1965**). ''The dependence of stress judgments on vowel formant structure,'' *Proceedings of the 5th International Congress on Phonetic Science, Münster, 1964* (Karger, Basel), pp. 306–311.

Gauffin, J., and Sundberg, J. (**1989**). ''Spectral correlates of glottal voice source waveform characteristics,'' J. Speech Hearing Res. **32**, 556–565.

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.

Green, D. M. (**1976**). *An Introduction to Hearing* (Erlbaum, Hillsdale, NJ).

Handel, S. (**1989**). *Listening: An Introduction to the Perception of Auditory Events* (MIT, Cambridge, MA).

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).

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 word fragments in spoken context: gating and shadowing,'' in *Proceedings of the 7th FASE Symposium, Speech '88* (Institute of Acoustics, Edinburgh), pp. 811–818.

Huss, V. (**1977**). ''English word stress in the post-nuclear position,'' Phonetica **35**, 86–105.

Kager, R. W. J. (**1989**). *A Metrical Theory of Stress and Destressing in English and Dutch* (Foris, Dordrecht).

Katwijk, A. van (**1974**). *Accentuation in Dutch: An Experimental Linguistic Study* (Van Gorcum, Amsterdam).

Klatt, D. H. (**1976**). ''Linguistic uses of segmental duration in English: Acoustic and perceptual evidence,'' J. Acoust. Soc. Am. **59**, 1208–1221.

Ladefoged, P. (**1967**). ''Stress and respiratory activity,'' in *Three Areas of Experimental Phonetics* (Oxford U.P., London), pp. 1–49.

Ladefoged, P. (**1971**). Preliminaries to Linguistic Phonetics (Univ. of Chicago, Chicago).

Langeweg, S. J. (**1988**). ''The stress system of Dutch,'' doctoral dissertation, Leiden University.

Lea, W. A. (**1977**). ''Acoustic correlates of stress and juncture,'' in Studies in Stress and Accent, edited by L. M. Hyman, Southern California Occasional Papers in Linguistics **4**, pp. 83–119.

Lehiste, I., and Peterson, G. E. (**1959**). ''Vowel amplitude and phonemic stress in American English,'' J. Acoust. Soc. Am. **31**, 428–435.

Lehto, L. (**1969**). English Stress and its Modification by Intonation: An Analytic and Synthetic Study of Acoustic Parameters (Suomalainen Tiedeakatemia, Helsinki).

Liberman, M. Y., and Streeter, L. A. (**1978**). ''Use of nonsense-syllable mimicry in the study of prosodic phenomena,'' J. Acoust. Soc. Am. **63**, 231–233.

Nakatani, L. H., and Schaffer, J. A. (**1978**). ''Hearing 'words' without words: Prosodic cues for word perception,'' J. Acoust. Soc. Am. **63**, 234–245.

Nierop, D. J. P. J van, Pols, L. C. W, and Plomp, R. (**1973**). ''Frequency analysis of 25 female speakers,'' Acustica **29**, 110–118.

Nooteboom, S. G. (**1972**). ''Production and perception of vowel duration,'' doctoral dissertation, Utrecht University.

Pierrehumbert, J. B. (**1980**). ''The phonology and phonetics of English intonation,'' Ph.D. dissertation, Massachusetts Institute of Technology.

Pols, L. C. W., Tromp, H. R. C., and Plomp, R. (**1973**). ''Frequency analysis of Dutch vowels from 50 male speakers,'' J. Acoust. Soc. Am. **53**, 1093–1101.

Rietveld, A. C. M. (**1984**). ''Syllaben, klemtonen en de automatische detectie van beklemtoonde syllaben in het Nederlands,'' doctoral dissertation, Catholic University of Nijmegen.

Rietveld, A. C. M., and Koopmans-van Beinum, F. J. (**1987**). ''Vowel reduction and stress,'' Speech Commun. **6**, 217–229.

Slootweg, A. (**1987**). ''Word stress and higher level prosodics,'' in *Linguistics in the Netherlands 1987*, edited by F. Beukema and P. Coopmans (Foris, Dordrecht), pp. 195–203.

Sluijter, A. M. C. (**1992**). ''Lexical stress and focus distribution as determinants of temporal structure,'' in *Linguistics in the Netherlands 1992*, edited by R. Bok-Bennema and R. van Hout (Foris, Dordrecht), pp. 247–259.

Sluijter, A. M. C. (**1995**). *Phonetic Correlates of Stress and Accent* (Holland Academic Graphics, The Hague).

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., Shattuck-Hufnagel, S., Stevens, K., and Heuven, V. J. van (**1995**). ''Supralaryngeal resonance and glottal pulse shape as correlates of prosodic stress and accent in American English,'' Proceedings of the 13th International Conference of Phonetic Sciences, Stockholm, **2**, 630–633.

Son, R. J. J. H. van, and Pols, L. C. W. (**1990**). ''Formant frequencies of Dutch vowels in a text, read at normal and fast rate,'' J. Acoust. Soc. Am. **88**, 1683–1693.

Stevens. (**1994**). Private communication.

Sweet, H. (**1906**). *A Primer of Phonetics* (Clarendon, Oxford).

Vanderslice, R., and Ladefoged, P. (**1972**). ''Binary suprasegmental features and transformational word accentuation rules,'' Language **48**, 819–838.

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.

Zanten, E. van, Damen, L., and Houten, E. van (**1991**), ''The ASSP Speech Database,'' SPIN/ASSP-report **41** (Speech Technology Foundation, Utrecht).

Zwicker, E., and Feldtkeller, R. (**1967**). *Das Ohr als Nachrichtenempfanger* (Hirzel, Stuttgart).