Why stress position bias?

Vincent J. van Heuven[a] and Ludmila Menert[b]

Department of Linguistics/Phonetics Laboratory, Holland Institute of Generative Linguistics (HIL),
Leiden University, Cleveringaplaats 1, P.O. Box 9515, 2300 RA Leiden, The Netherlands

There is ample evidence in the literature that English and Dutch listeners tend to perceive stress on the word-initial syllable. This bias is most easily seen in the perception of (nonsense) words containing repetitions of identical syllables. In four experiments the possible causes of this bias are investigated. The results show that the bias disappears when (i) words are preceded by a spoken context, when (ii) the voice source is replaced by noise (whisper), or when (iii) the fundamental frequency level of the utterance as a whole is lowered. The data are best explained by assuming that the listener interprets the onset of voicing of an isolated word as a (silent) pitch rise from the bottom of the speaker's pitch range. © *1996 Acoustical Society of America.*

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

INTRODUCTION

A. The problem of stress position bias

Stress is an abstract linguistic property of a word that defines the position of the strongest syllable in the word. In languages such as English and Dutch the stressed syllable is articulated with greater effort than its unstressed counterpart, so that its constituent sounds have greater intensity, longer duration, and approximate their ideal pronunciation (target values) more optimally. Moreover, if the word is accented, a perceptually salient pitch movement is placed on the stressed syllable. In this case, the syllable bearing the pitch movement is certain to be perceived as stressed: the pitch movement is a sufficient cue for stress. The relative perceptual strength of the other cues, intensity, duration, and spectral distribution, may vary across languages. For English and Dutch it is generally held that of these three features (vowel) duration is given the greatest weight, whereas spectral distribution and intensity are of lesser importance (Fry, 1955, 1965; van Katwijk, 1974; Lehiste, 1970).

Imagine that we synthesize a nonsense word consisting of an exact repetition of the same syllable (e.g., /tɪstɪs/), i.e., one with exactly the same sound segments, with identical intensity and duration, and with exactly the same pitch (contour). In such a stimulus there is no phonetic property by which one syllable would stand out more than the other, so that we would expect the listener to be unable to decide which syllable bears the stress. Counter to this naive expectation, there is ample evidence to show that, for this type of material, English and Dutch listeners typically perceive the stress on the first syllable (e.g., Morton and Jassem, 1965; van Katwijk, 1974; Berinstein, 1979). For listeners to perceive stress in any other position, the initial syllable has to be reduced in duration, intensity, or spectral distribution, or one of the noninitial syllables has to be given a pitch movement. This tendency to perceive stress on the first syllable will be called **stress bias for initial position** or, simply, stress bias.

In English, as well as in Dutch, the typical stress position is word initial (Delattre, 1965; Carlson *et al.*, 1985; van Heuven and Hagman, 1988). In contrast, for speakers of K'ekchi and Caqchiquel, Mayan languages with fixed stress in final position, the perceptual stress bias was found to favor the final syllable (Berinstein, 1979).[1] Therefore, the distribution of stress in the lexicon of the listener's language is obviously a major cause of the perceptual stress bias. It is possible, however, that there are additional factors that govern the direction and strength of the perceptual stress bias that reside in the speech signal itself. It is the primary purpose of this article is to examine several such factors underlying stress bias in Dutch, a language with a statistically dominant word-initial stress pattern (van Heuven and Hagman, 1988).

The phenomenon of stress position bias has typically been examined in perception experiments on the basis of isolated (nonsense) word stimuli only (see references above). The exclusive use of isolated word stimuli may have introduced an experimental artifact. Specifically, we believe that the task of judging stress position in utterance-initial words causes the listener to make assumptions as to the pitch of the preceding period of silence. The arguments leading up to this hypothesis are presented in the next section, together with a general indication of the type of experiment that is needed to test the consequences of this hypothesis.

B. The possible role of context in stress perception

As we have explained above, stress is cued by (a combination) of the acoustic parameters of intensity, spectral distribution, duration, and pitch. We believe that listeners evaluate the effect of the first two parameters on the basis of a strictly **within-word** comparison of differences between syllables. All else being equal, listeners will report the stress on the syllable that is least reduced (relative to some internalized norm for each sound or syllable) in terms of intensity and spectral distribution.

The effects of pitch and duration on stress perception involve different mechanisms. Let us first consider pitch. Counter to what is often stated, stress is not necessarily associated with the syllable containing the highest pitch within

[a] Electronic mail: heuven@rullet.LeidenUniv.nl
[b] Now at Research Institute for Language and Speech (OTS), Utrecht University, The Netherlands.

the word. The crucial element for stress perception is that a **change in pitch**, whether rise, fall, or both, can be associated with a particular syllable. For instance, for a pitch rise to be perceived as a cue to stress in Dutch, its peak value should lie within 50 ms after the vowel onset in the syllable ('t Hart *et al.*, 1990, p. 73; Hermes and Rump, 1994).[2] Even if the pitch does not change over the duration of the target word itself, the listener may still perceive a rise (or fall) on one of the marginal syllables if the pitch level of the preceding or following word is different. Given that the domain within which pitches are compared in stress perception extends beyond word boundaries, it is important to know what assumptions, if any, the listener makes with regard to the pitch of the context when a word is presented in isolation.

Although the perceptual evaluation of duration as a cue to stress is generally very much like that of intensity and spectral distribution, it also involves an element of context sensitivity. Duration cannot be evaluated on a strictly within-word basis because the duration of the word final syllable is increased considerably before a prosodic boundary (e.g., Klatt, 1976; Nooteboom and Doodeman, 1980). Therefore, the absence of temporal reduction in the final syllable does not necessarily imply stress on that syllable. However, the presence of temporal reduction in any other position is a sure sign of nonstress.

C. Predictions and experiments

If the listener, when asked to decide where the stress falls in an isolated word, makes no assumptions as to the pitch of the (silent) context, we should obtain the same results whether the target word is tonal or atonal. We reasoned that a whispered word, presented in isolation, would preclude any need on the part of the listener to make assumptions with regard to the pitch of the silent context. Therefore, in experiment I, we compared bias in stress perception between versions of the same words (both sense and nonsense) that crucially differed in the nature of the excitation signal of the synthesizer: periodic buzz versus aperiodic noise.

In experiment II we varied the presence versus absence of a spoken context. When present, the spoken context will provide the basis on which the pitch level of the onset syllable of the target word will be evaluated. Specifically, we will test the hypothesis that stress bias favoring the initial position will disappear, or will at least be attenuated, when a target is embedded in a context sentence with level pitch throughout, whereas substantial bias will be found when the same target is presented in isolation (silent context).

On the strength of the argument that the onset syllable of a word presented in isolation is typically perceived as stressed, we speculate that the listener assumes the pitch of the silent context to be lower than the pitch of the target word itself. We hypothesize that, as a result, the listener associates a pitch rise with the first syllable of the target, which is taken as a cue for stress. If this reasoning is correct, we predict that the stress bias favoring onset position will increase as the target is synthesized on a higher monotonous pitch. This hypothesis is tested in experiment III.

If we present a stimulus word containing a succession of two identical syllables, the listener may apply perceptual compensation for the final lengthening effect (see above). The final syllable, when followed by silence, should be longer than the other syllable(s). If not, it is interpreted as a reduced syllable, and the stress is judged to fall on the first syllable. Perceptual compensation for the effect of final lengthening provides one explanation for the initial stress bias. Therefore, experiment IV examines the applicability of perceptual compensation for final lengthening as an alternative or additional mechanism underlying stress bias. To this end target words will be presented in final position as well as embedded in a following context. Since final lengthening is blocked for embedded targets, the perceptual compensation hypothesis predicts no (or at least reduced) stress bias when the target is followed by a pause or prosodic boundary. In experiment IV we also tested one further alternative explanation of the context effect on stress bias, reasoning that an isolated target does not only stand out from its silent context by a (hypothetical) difference in pitch, but also by a greater difference in intensity.

I. EXPERIMENT I: EFFECTS OF VOCAL SOURCE

In this experiment we will investigate the effects of the vocal source on the strength of the stress bias favoring initial position in isolated words. Our first aim will be to replicate the finding reported by others (e.g., Morton and Jassem, 1965; van Katwijk, 1974; Berinstein, 1979) that the onset syllable is the preferred stress location when a nonsense word containing two identical syllables is presented.

To this end, we generated the same set of stimuli in four different versions, using a speech synthesizer based on the source-filter theory of speech production. The particular synthesizer used (Philips MEA 8000 chip) excites a complex filter (five formants; center frequencies F_1-F_3 variable, bandwidths B_1-B_4 variable, F_4, F_5, and B_5 fixed) by either a periodic source (F_0 variable between 50 and 512 Hz) or by white noise (for further details see Brück and Teuling, 1982). In the crucial condition the excitation signal was white noise, so that the stimulus is atonal, and sounds whispered throughout. If our reasoning is correct, there should be no bias favoring stress in the onset position when no meaningful pitch can be associated with the stimulus. The results of the noise-source condition will have to be compared with those obtained when the excitation signal is periodic (buzz).

It is not clear, *a priori*, what fundamental frequency course would constitute a truly neutral stimulus. It has been observed that there is a general tendency for the vocal pitch to gradually drift down over the course of an utterance. This phenomenon is known as declination (cf. Cohen and 't Hart, 1967; Cohen *et al.*, 1982; Pierrehumbert, 1979). It would therefore seem reasonable to assume that listeners expect utterances to exhibit some sort of gradually declining pitch. Hearing a word with a perfectly monotonous pitch is counter to this general tendency: the later syllables are higher pitched than they should be, so that perception of stress will be shifted towards the end of the word. We solved this dilemma by not using one but three buzz conditions for comparison, one with a perfectly monotonous pitch, one with gently falling pitch (declination), and one with gently rising pitch (in-

clination). We predict that inclining pitch will attenuate the stress bias for the initial position, whereas declining pitch will strengthen the bias.

In order to be able to further evaluate the strength of the effect of varying the excitation source, we included temporal organization as a second variable. In the absence of abrupt pitch movements, lengthening one syllable relative to the others is the most effective cue signaling stress in English and Dutch (Lehiste, 1970, Chap. IV; van Katwijk, 1974).

Finally, we did not want to rely on the exclusive use of nonsense words. It can be argued that naive listeners have no easy access to such theoretical notions as stress, and can judge stress position better if a difference in stress position corresponds to a difference in meaning. Fortunately, the Dutch lexicon contains a number of minimal stress pairs, so that in fact more than half of our stimuli were existing words.

A. Method

Eight-word pairs, differing in stress position only, were selected. In two pairs the stress difference corresponds with a lexical contrast:

kanon	/ˈkaːnɔn/	-	/kaˈnɔn/
	'canon		cannon'
servisch-	/ˈsɛrviˑs/	-	/sɛrˈviˑs/
servies	'Serbian		chinaware'

In another two pairs different stress position cues a difference in morphological structure (as well as in meaning) between the members of a pair. These complex verbs are separable when stressed on the prefix but are inseparable when stressed on the stem. In the latter case the verbs are intransitive and take a past participle without the participle marker *ge-*.

voorkomen	/ˈvoːrkoːmə/	-	/voːrˈkoːmə/
	'appear in court		prevent'
doorlopen	/ˈdoːrloːpə/	-	/doːrˈloːpə/
	'walk on		finish'

Next, two compound adjectives were included; these have variable stress depending on their rhythmic environment: when followed by a noun within the same phonological phrase, the first element bears the main stress; when used predicatively (i.e., at the end of a phonological phrase), main stress is on the second element (cf. English *un*ˈknown versus ˈ*unknown* ˈ*objects*, e.g., van Heuven, 1987; Liberman and Prince, 1977).

lichtgrijs	/ˈlɪχtχrɛɪs/	-	/lɪχtˈχrɛɪs/
	'light gray, attributive versus predicative use'		
matblauw	/ˈmatblɑʊ/	-	/matˈblɑʊ/
	'matt blue, attributive versus predicative use'		

Finally, two pairs contained reiterant speech (repetition of the same syllable, cf. Nakatani and Shaffer, 1978) yielding nonsense words:

saasaas	/ˈsaːsaːs/	-	/saːˈsaːs/
siesies	/ˈsiˑsiˑs/	-	/siˑˈsiˑs/

These nonsense words provide a necessary control condition: here the listener cannot apply any lexical knowledge (i.e., type or token frequency) to favor one syllable position over the other for stress. The stress rules for Dutch would predict initial stress if these sequences are considered to be compound nouns (compound stress rule as in English) but on the final syllable if these items are taken to be monomorphemic: superheavy syllables (i.e., ending in a branching rhyme VVC or VCC) attract stress (for a survey cf. Neijt and van Heuven, 1992).

These 4∗2 segment strings were synthesized from diphones, using a Philips MEA 8000 speech synthesizer controlled by a microcomputer. Our diphones were excised from the accented syllable spoken in nonsense words of the type /CəˈCVCə/, (where C was identical throughout the word) that had been recorded in a fixed carrier phrase (Elsendoorn and 't Hart, 1982, 1984), and then coded in terms of formant frequencies $(F_1 - F_5)$ and bandwidths $(B_1 - B_5)$, voiced/unvoiced and intensity, using linear predictive coding (LPC) analysis (for details cf. 't Hart *et al.*, 1982; Vogten, 1984). Since no temporal or spectral changes were made to the parameter values of the diphones, the resulting speech was constructed exclusively from segments with the spectral and temporal characteristics of primary stress. Theoretically, such speech provides no cues signaling stress position. Any explicit stress indication has to be introduced *post hoc*, e.g., by inserting a pitch movement, or by manipulating intensity, duration, and/or segmental quality.

Each word was then synthesized in three temporally different versions.

(1) Neutral, i.e., with durations copied from accented syllables as explained above (for specifics of the neutral vowel durations of all the target words used in the present series of experiments, see the Appendix).

(2) A relative lengthening of the first stress position, incrementing the first vowel by 32 ms while decrementing the vowel in the second stress position by 32 ms; this value was found to cause a reasonable shift in stress judgments, without leading to ceiling and/or bottom effects.

(3) A relative lengthening of the second stress position, decrementing the first vowel by 32 ms and incrementing the second vowel by 32 ms.

Next, each of the resulting 24 stimulus types were synthesized using four different source signals.

(1) Monotonous (MON) pitch fixed at 100 Hz.

(2) Gradually falling pitch (declination, DEC), starting at 100 Hz and dropping 1 Hz every 20 ms.

(3) Gradually rising pitch (inclination, INC), starting at 100 Hz and rising 1 Hz/20 ms.

(4) The voicing source replaced by white noise (NOI).

Sample parameter displays and oscillograms of the neutral versions of selected stimulus words are included in the Appendix. The entire set of 96 stimulus types was recorded on audio tape in quasirandom order, excluding the immediate repetition of the same lexical item or source signal, with 2 s interstimulus intervals (offset to onset). This tape was presented over good quality loudspeakers in an acoustically treated language laboratory to small groups of listeners (eleven in all), who were audiometrically normal, Dutch

speaking students at the Faculty of Arts, Leiden University, and who served voluntarily and without remuneration.

Subjects were instructed to decide which syllable for each stimulus was stressed (binary forced choice, no ties allowed), and to indicate their choice on answer sheets that contained orthographic listings of the material on the tape. They were given the opportunity to practice before the experiment. The entire procedure was repeated once after a short break to collect the answer sheets used, so that each listener produced two responses to each stimulus type.

B. Results

Figure 1 presents percent stress perceived on the first syllable and, by implication, on the second (i.e., the complementary score to 100%), broken down by temporal type (first syllable lengthened, neutral, second syllable lengthened) and by source signal (MONotone, DEClination, INClination, NOIse).

The results prove fairly straightforward. Changing the temporal makeup of the stimulus causes a 30% change from perceived stress on the first syllable to the second syllable, or vice versa. The overall effect, though less convincing than the duration effects reported in the literature (cf. Introduction), is significant by an analysis of variance (ANOVA) with lexical type, source signal, and temporal version as fixed factors, and with lexical instantiations as the error term $[F(2,93)=34.9, p<0.001]$;[3] the means for the neutral and lengthened first syllable versions do not differ significantly (Newman-Keuls test for contrasts, $p<0.05$). Crucially, a similar effect can be observed for the type of source signal. Stress is typically reported on the initial syllable when the source is either a monotonous or declining pitch; however, when the source is either an inclining pitch or noise, the distribution of stress judgments over the syllables is essentially random, indicating that no stress bias remains $[F(3,92)=17.6, p<0.001]$; all differences between means are signifi-
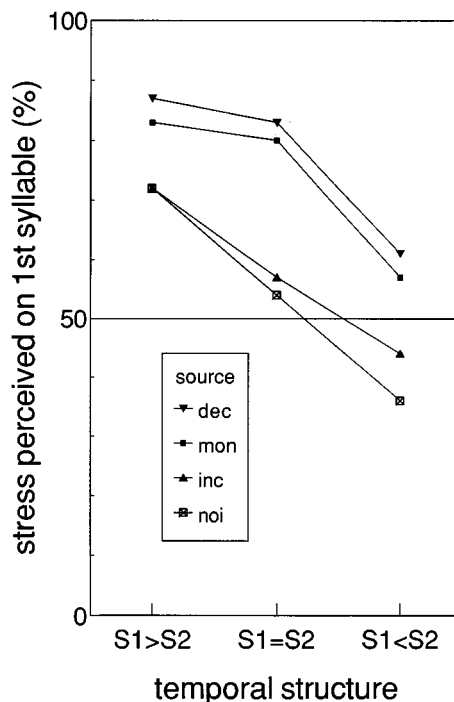

FIG. 1. Percent stress perceived on first syllable broken down by temporal structure of stimulus ($S1>S2$: first syllable lengthened, $S1=S2$: neutral, $S1<S2$: second syllable lengthened), and source signal (DEC: periodic with pitch declination, MON: periodic with monotonous pitch, INC: periodic with pitch inclination, NOI: with noise source).

cant, except for the pairs MON/DEC and INC/NOI. There is no interaction between the temporal version and signal source $[F(6,84)=1.6, p>0.1]$.

Figure 2 presents the results for the four categories of minimal stress pairs, broken down by temporal version but collapsed over source signals and lexical instantiations. As can easily be seen here, there are considerable differences in preferred stress location depending on the type of minimal
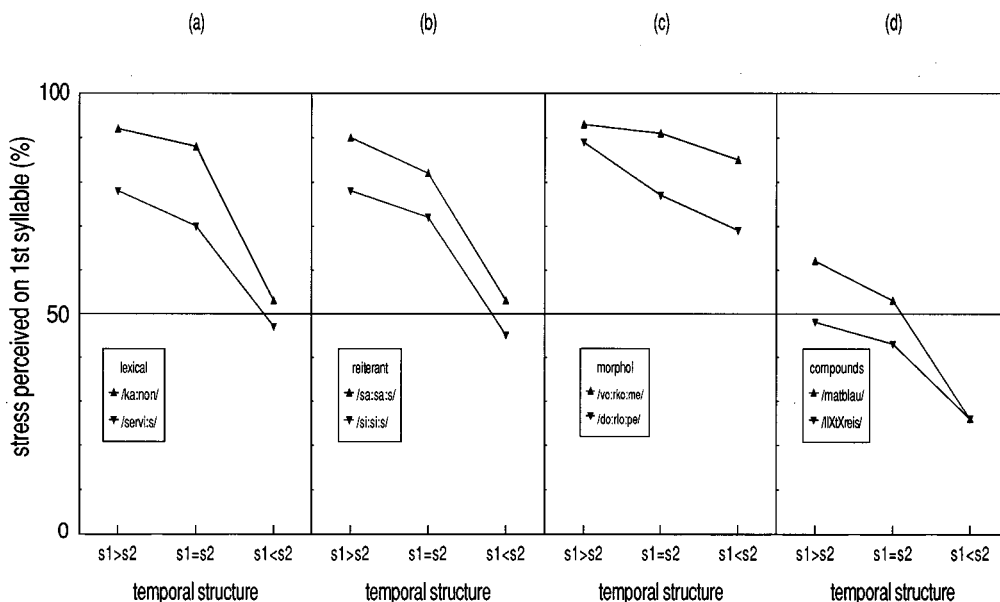

FIG. 2. Percent stress perceived on first syllable collapsed over source signals but broken down by temporal structure of stimulus (cf. Fig. 1) and by word type: (a) lexical stress, (b) reiterant item, (c) morphological stress, and (d) compound adjective.

stress pair. Most important, the adjectival compounds (*licht-grijs, matblauw*) are characterized by a modest trend for stress perceived in the second position, whereas all other items show a strong preference for stress in the initial position. The effect of minimal stress pair is significant [$F(3,92)$ $=10.6$, $p<0.001$]. There are no interactions with either temporal version [$F(6,84)<1$], or source signal [$F(9,80)=1.4$, $p>0.1$].

C. Conclusions

These results allow us to draw a number of preliminary conclusions. First, we have essentially replicated the results reported by Morton and Jassem (1965) for the effect of pitch slope. In the stimuli synthesized with a perfectly level pitch, there is considerable overall bias for stress in the first position (except for compound adjectives). When replacing the level pitch by a gently falling pitch (declination), the bias increases only marginally and by a statistically nonsignificant amount that may be due to a ceiling effect. Gently rising pitch, however, counteracts the initial bias, and leads to an essentially even distribution of stress judgments over the two syllables.

Second, the results reveal a strong general trend, averaged over all conditions, for stress to be perceived on the first syllable for all stimuli except the compound adjectives. For this type of word no preference for either first or second position was found.

Third, the data contain important evidence that the stress bias for the initial position is caused by the periodic character of the vocal source. When the stimuli are synthesized with a noise source, no stress bias remains in the data. The perceptual effect of replacing level pitch by noise is nearly identical (in fact, slightly stronger) to that of generating a rising pitch contour on the stimulus (inclination) such that the second vowel is between 1.7 semitones (for *siesies*) and 2.7 semitones (for *doorlopen*) higher (on average) than the first.

Finally, the statistical analysis shows that the lexical effect (word type) is independent of the vocal source effect: the effects of pitch direction and periodicity apply equally to all word types used. These data suggest that the effect of the source characteristic will also be found for other languages, and even for languages with fixed stress in some noninitial position.[4] Future cross-linguistic research will have to substantiate the claim made here that a rerun of our experiment with native listeners of a language with fixed final stress (e.g., French, K'ekchi, Caqchiquel) should show the same effects of source signal but with a larger overall bias favoring final stress, without interaction of language and source signal.

II. EXPERIMENT II: EFFECT OF PRECEDING CONTEXT

In this experiment we vary the presence versus absence of a spoken context as a factor potentially influencing the magnitude of the stress bias for initial syllables. If indeed the onset syllable of an utterance is perceived as having a higher pitch than the preceding silence (absence of context), bias for

stress perception in initial syllables should disappear if the target is embedded in a spoken context with level pitch throughout the utterance. It is this hypothesis that is tested here.

A. Method

The basic stimulus material was reduced to half its original size by selecting, at random, one exemplar from each lexical category, viz., *kanon* (lexical stress), *overkomen* (morphological stress), *lichtgrijs* (compound adjective), and *saasaas* (reiterant nonsense word). *Overkomen*/ˈoːvərkoːmə/ 'come across, fly over' versus /oːvərˈkoːmə/, 'happen to' replaces earlier words with the adverbial prefix *door-* or *voor-*, so as to obviate the possible criticism that vowel lengthening by /-r/ (cf. Nooteboom, 1972) may have caused the stress bias for the onset syllable.

These four words were synthesized as they were in Sec. I A, again with three temporal versions and four different source signals. However, each of the resulting 48 stimuli was generated once in isolation, and a second time in final position in a fixed carrier phrase *En toen zei ze ...* /ɛn tuːn zɛɪ zə/ 'And then she said ... .' The carrier was synthesized with declining pitch (1 Hz/20 ms) such that the pitch was 100 Hz at the onset of the target word. The onsets of isolated target words were characteristic of utterance-initial position, whereas in context the transition of carrier into target was fluently coarticulated. It is an inherent feature of diphone synthesis that first-order coarticulation effects are maintained: the diphone at the beginning of the target, containing the transition from silence to the consonant, was replaced by the appropriate diphone containing the transition from schwa to the following consonant. Parameter displays and oscillograms of selected stimuli are included in the Appendix.

The 96 stimuli were recorded on tape in quasirandom order, excluding the immediate repetition of the same context, excitation source, and lexical type.

The experiment was run with twelve listeners drafted from the same pool of subjects and participating under the same conditions described in Sec. I A.

B. Results

The results for this experiment are presented in Fig. 3, with items in isolation and in context presented in separate panels, that are analogous to Fig. 1. Clearly, the results of the previous experiment are replicated here. Interestingly, the distribution of stress judgments for the new item *overkomen* is again most strongly biased for initial stress, as was typical of verbs with adverbial prefixes.

The effect of context is to reduce the incidence of initial stress judgments from 65% on aggregate in isolation to 48% in context. This effect is significant by an ANOVA with context, temporal version, and source signal as fixed factors, and with lexical type as the error term[5] [$F(1,94)=11.7$, $p=0.001$]. Moreover, the presence or absence of context exerts a purely additive effect vis à vis the earlier variables, $F<1$ for second and third order interactions, so that a detailed description of the other effects can be omitted here: these were roughly the same as those described in Sec. I B.
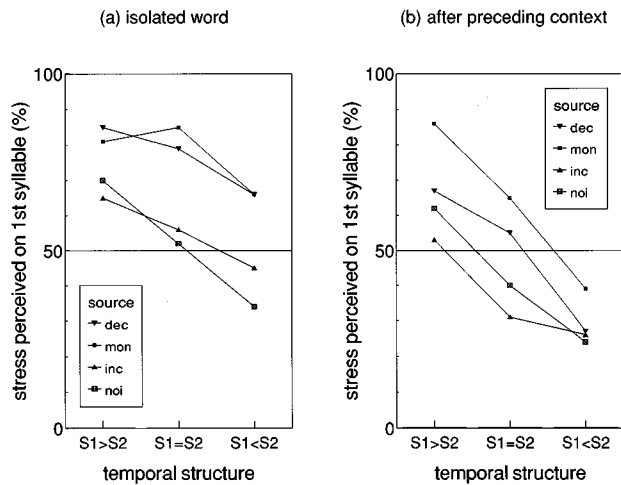
FIG. 3. Percent stress perceived on first syllable broken down by temporal structure of stimulus and source signal (cf. Fig. 1) for stimuli presented in isolation [panel (a)] and in initial position in a short carrier sentence [panel (b)].

C. Conclusion

In this experiment we have tested the prediction that the preference on the part of the listener to perceive the first syllable of a word as stressed will be reduced when the target is immediately preceded by a spoken context and fluently coarticulated with it. This prediction followed from our view that the listener assumes a low preceding pitch when there is no context. The difference between the actual onset pitch and the assumed (silent) context pitch is then evaluated as a rise in pitch cuing stress. If a preceding context is present and its pitch is equal to that of the target, there can be no perceived or inferred pitch rise, so that the stress bias for the target's onset syllable should disappear. This is precisely the result that we obtained.

III. EXPERIMENT III: EFFECTS OF PITCH LEVEL

As anticipated in the Introduction, we assume that the listener, on hearing an utterance-initial syllable, generates a reference pitch level that is equal to the lowest vocal pitch appropriate for the particular speaker's voice, i.e., his terminal frequency for a declarative sentence. For an average Dutch male this pitch would lie around 75 Hz. The actual pitch of the stimulus onset, which in our experiments was at 100 Hz, is then evaluated against this (lower) reference pitch. The difference between the actual pitch (100 Hz) and the reference pitch (75 Hz) is interpreted as a pitch jump (or 'silent rise'), and taken as a cue for stress on the first syllable.

When the stimulus is aperiodic, no actual pitch can be determined, and no pitch jump can be inferred. Hence, in experiment I, bias disappeared in 'whispered' targets. When the target is preceded by a spoken carrier phrase, the reference pitch is provided by the context. Since, in experiment II, the pitch was level throughout the relevant portion of the stimulus, no pitch jump was heard, and bias disappeared.

How can we show that this admittedly speculative account of stress bias is correct? If it is true that the difference between onset pitch and reference pitch is interpreted as a

pitch movement, it follows that a higher actual pitch onset, all else being equal, should yield a greater bias. Therefore we will vary the onset pitch between 70 Hz (roughly coincident with the assumed reference pitch) and 160 Hz. We hypothesize that a 70-Hz onset will generate little or no bias, but that higher onsets (100, 130, 160 Hz) will come out with ever larger bias.

Second, we will speculate on the mechanism by which the listener generates the reference pitch. Here we hypothesize, along with Laver and Trudgill (1979), that the listener assumes the presence of a large individual from a speech sample with relatively low formants (i.e., large resonance cavities), whom he associates with a low bottom pitch. Conversely, when the formants in the speech sample are relatively high, the speaker is apparently a small individual, with a correspondingly high-pitched voice.[6] To test this hypothesis we generated stimuli using three different formant settings: starting from a normal male with an average formant setting, we simulated a large individual with lowered formants, as well as a small individual with raised formants. If it is true that lowered formants are associated with a low-pitched voice, the listener will generate a lower reference pitch for this type of voice, so that a larger silent rise will be perceived relative to a fixed actual pitch onset. Similarly, if raised formants correspond to high-pitched voices, we predict stronger bias for stimuli with up-shifted formants.

A. Method

Seventy-two stimuli were synthesized from diphones by a DEC MicroVAX-II computer using the LVS speech analysis and (re-) synthesis software developed at IPO-Eindhoven ('t Hart *et al.*, 1982; Vogten, 1984). The diphones used here and in our earlier experiments were the same. The overall speech quality was improved, however, owing to the fact that hardware synthesis through the MEA 8000 chip allows severely quantized parameter specifications only; these limitations do not apply to our software synthesis.

The stimuli comprised 36 versions of the Dutch word *kanon* (see experiment I), and another 36 of the nonsense word *saasaas*. The 36 versions of each word were then obtained through orthogonal combination of three factors.

(1) F_0 was varied in four steps: 70, 100, 130, and 160 Hz; F_0 was level throughout the duration of the stimulus word.

(2) Formant range was varied in three steps. Starting from the formant frequencies $F_1 - F_5$ as calculated by the LPC analysis, a type of voice was synthesized that was typical of a large male (formants lowered to 0.85 of their original frequencies), and another type that suggested a small male (formants raised to 1.20 of their original values).

(3) Temporal type was varied in three steps, as in the previous experiments. However, instead of lengthening and shortening the vowels by 32 ms, now a 50-ms increment was used. The larger increment was chosen so as to insure a closer approximation of the duration effects reported in the traditional literature (cf. Introduction).

The 72 tokens were recorded onto audio tape in quasi-random order, preceded by eight practice items. This tape

(a)
formants raised

stress perceived on 1st syll (%)

timing
S1>S2
S1=S2
S1<S2

F0 level (Hz)

(c)
formants neutral

stress perceived on 1st syll (%)

timing
S1>S2
S1=S2
S1<S2

F0 level (Hz)

(b)
formants lowered

stress perceived on 1st syll (%)

timing
S1>S2
S1=S2
S1<S2

F0 level (Hz)

(d)
averaged over formant settings

stress perceived on 1st syll (%)
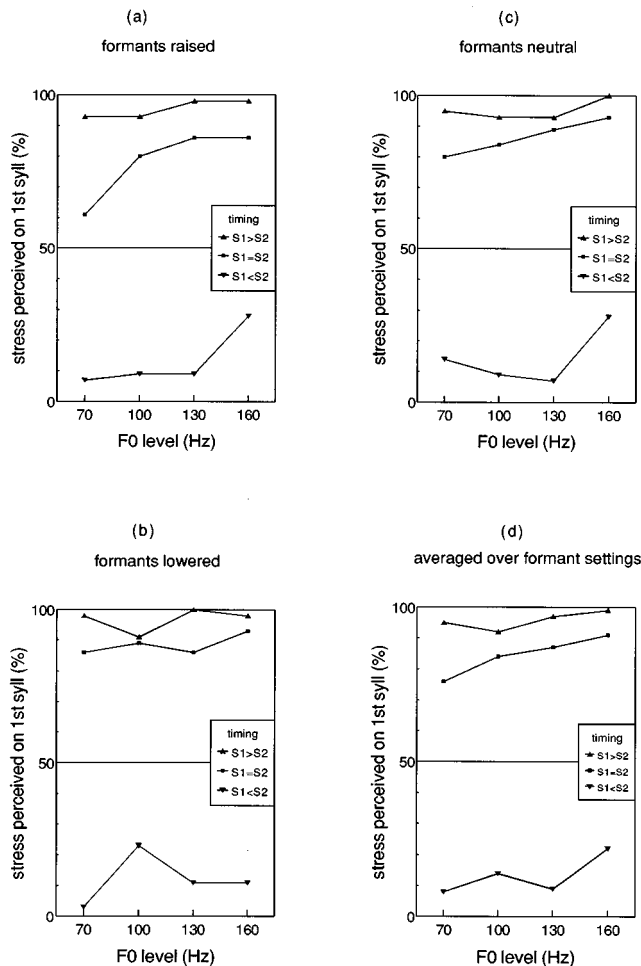
timing
S1>S2
S1=S2
S1<S2

F0 level (Hz)

FIG. 4. Percent stress perceived on first syllable as a function of F_0 level (Hz), broken down by temporal structure of the target word (cf. Fig. 1) and for formant setting: raised [panel (a)], neutral [panel (b)], and lowered [panel (c)]. Panel (d) presents the results averaged over the three formant settings.

was presented twice to eleven Dutch listeners over a good quality sound reproduction system (Quad ESL-63) in a small, well-insulated lecture room with some soft paneling attached to walls and ceiling to reduce reverberation. Other procedural details were as described in Secs. I A and II A.

B. Results

Figure 4 presents percent stress perceived on the first syllable broken down by pitch level, formant setting, and temporal version. Manipulating the relative duration of first and second syllable produces 96% stress perceived on the lengthened first syllable, 85% on a temporally neutral first syllable, and 13% on a shortened first syllable. This effect was significant by a four-way ANOVA on mean percent perceived initial stress with lexical type (/kaːnɔn/ versus /saːsaːs/), pitch level, formant setting, and temporal version as fixed factors, and with repetition as the error term [$F(2,141)=1120.1$, $p<0.001$]. This effect of duration manipulation was, in fact, larger than we had hoped for, and tends to crowd out the effects of the other factors.

Still, changing the F_0 level has a clear effect on stress perception. When averages are taken over the other variables

[panel (d) of Fig. 4], bias for the first syllable increases monotonically with the F_0 level: 60% stress for 70 Hz, 63% for 100 Hz, 64% for 130 Hz, and 71% for 160 Hz. Although this effect is smaller than that of temporal version, it is still substantial [$F(3,140)=8.5$, $p<0.001$]. The effect of F_0 is most pronounced in the temporally neutral versions with 76%, 84%, 87%, and 91% stress perceived on the first syllable [$F(3,45)=4.4$, $p=0.020$].

Counter to our prediction, bias does not disappear completely at 70 Hz. Whether a further reduction of bias can be obtained by lowering the F_0 level even more is doubtful: when constructing our stimuli we had to abandon pitches below 70 Hz, as these sounded highly unnatural (rough, creaky voice quality).

The predicted effect of formant setting is not borne out by our data. If anything, the results are in the wrong direction, but the effect of formant setting is insignificant [$F(2,141)=1.9$ $p>0.1$].

C. Conclusions

Again we have demonstrated that the perception of stress is not solely dependent on differences in fundamental frequency, intensity, duration, and timber within the word, as is generally maintained in the literature (see above).

We have shown here that the fundamental frequency of a perfectly level-pitched stimulus influences the perception of stress in isolated words: the higher the pitch level, the greater the bias favoring stress on the first syllable. The effect was especially clear for those stimuli where the duration cue was ambiguous.

Generally, then, the results of this experiment confirm our hypothesis that stress bias is caused by the listener perceiving a discrepancy between the actual pitch onset and some low reference pitch as a pitch rise cuing stress.

We have not been able to confirm our hunch that the reference pitch is derived from the average formant setting in the voice of the speaker. In fact, more recent publications have shown that, counter to what has been claimed in earlier studies, there is hardly any correlation between the physical size of the speaker (in terms of height and weight) and mean fundamental frequency (Künzel, 1989). Such correlations exist only between groups such as male versus female, or adult versus child; within a group of adult males any correlation disappears. We will therefore have to consider other possibilities. For instance, it may even be that the reference pitch is fixed and speaker-independent. This possibility will not be explored further in this article.

IV. EXPERIMENT IV: TESTING ALTERNATIVE EXPLANATIONS

We have already presented substantial experimental evidence in support of our view that word onset stress bias is the result of a perceptual mechanism whereby the listener interprets the difference between the onset pitch of the target word and the assumed low pitch of the silent context as a pitch rise. Our fourth and final experiment was designed to replicate the effect of pitch level (experiment III) and at the same time to rule out two alternative explanations for the

2445   J. Acoust. Soc. Am., Vol. 100, No. 4, Pt. 1, October 1996

V. J. van Heuven and L. Menert: Stress position bias   2445

bias phenomenon that have been entertained by others and by ourselves at some stage during the research.

A. Alternative hypotheses

Two alternative explanations will be identified in the following, and means of testing their consequences will be outlined.

1. Perceptual compensation for final lengthening

The final syllable of a word is generally pronounced longer than the earlier syllables in the word (e.g., Klatt, 1976; Nooteboom, 1972). Repp (1986) suggested that listeners know this, and consequently expect the final syllable to be longer than other syllables. Through perceptual compensation, the final syllable in a word containing two identical syllables will be evaluated as being shorter, and therefore less stressed, than the initial syllable.

Notice that this account does not explain the effects of noise source and preceding context, both of which reduce bias for the first position. Nevertheless, we decided to test this explanation, reasoning as follows. The effect of final lengthening is restricted to words spoken in isolation or in clause-final position (Klatt, 1976; Nooteboom and Doodeman, 1980). Therefore, no (perceptual compensation for) final lengthening should occur when the crucial word is fluently coarticulated with its immediately **following** spoken context. Hence we predict that the percentage of stress perceived on the first syllable of a word (with ambiguous stress position) will drop relative to its presentation in isolation if the word is followed by a coarticulated speech context.

2. Perceived intensity jump

As a second alternative, let us assume that the onset of an isolated word stands out from its background because of the sudden and large increase of auditory stimulation. This view is compatible with our finding that the initial syllable loses some of its prominence when it is preceded by a carrier phrase. This reasoning has its drawbacks, too: it is by no means clear that stress is cued by the energy increase in the stimulus onset. Although the claim has been made at least once before (Lehto, 1969), it was never perceptually tested. Still, the viability of this account is easily tested. Reduction of initial stress bias should be obtained when the leading syllable is closely preceded by any other type of intense auditory stimulation, e.g., by a noise burst not belonging to the spoken message.

B. Method

For this experiment we used one existing word with variable stress (*kanon*) and one reiterant item (*saasaas*). These items were synthesized in three temporal versions, as in experiments I and II (i.e., with 32-ms increments/decrements of vowel duration), either in isolation or fluently coarticulated with a preceding and/or following context phrase: *En toen zei ze... een beetje harder*, /ɛn tuˑn zɛɪ zə... əm beːcə hɑrdər/ 'And then she said... a little louder,' with the entire stimulus generated on a 100-Hz level pitch (see the

Appendix for the oscillogram and parameter display of the carrier). Three more conditions were created by replacing the speech contexts by noise bursts (ANSI or 'speech' noise, General Radio GR 1382 noise generator gated through a Grason Stadler 1284B electronic switch, 25-ms rise/fall time controlled by a Devices Digitimer D4030 microtimer). The noise was equal in duration (864 ms and 1136 ms for preceding and following context, respectively) to the speech material deleted, and matched in intensity with the loudest vowel of the crucial word, [aː]. The initial (and final) 15 ms of the noise overlapped with the offset (and onset, respectively) of the crucial word in order to ensure some degree of perceptual continuity. Isolated words synthesized with a level pitch of 160 Hz made up the last condition in this experiment, yielding a total of eight conditions.

The ensemble of 48-stimulus types was recorded on audio tape in quasirandom order, excluding immediate repetition of lexical item and context condition, and was preceded by six practice items. The tape was presented twice to a group of twelve listeners that had not been involved in any of the earlier experiments.

C. Results

The results of this experiment are summarized in Fig. 5, which shows percent stress perceived on the initial syllable accumulated over lexical types, subjects, and tokens, and broken down by the remaining independent variables. The scores for conditions with speech versus noise context are superposed [panels (b)–(d)], as is also done for the two pitch conditions with isolated words [panel (a)].

Examining first the results for the isolated words [panel (a)], we notice that raising the level pitch from 100 to 160 Hz has the predicted effect of increasing the bias favoring the initial position. The difference is not so much apparent in the temporally neutral and lengthened first syllable versions because of a ceiling effect, but it is **very strong** in the lengthened second syllable condition: here the perception of initial stress rises from 23% for the 100-Hz level pitch to 69% or 160 Hz. The overall effect of pitch level is significant by a two-way ANOVA on mean percent perceived initial stress with temporal makeup and pitch level as factors, assuming fixed effects, and with repetition as the error term [$F(1,11)$ $=31.7$, $p=0.001$]. Clearly, general effect of the pitch level found in experiment III is replicated here, lending further support to the silent pitch rise hypothesis.

Now turning to the effect of spoken context, abstracting from higher-order interactions we find a replication of the context effect of experiment II. Adding a preceding speech context reduces the bias for the first position: 69% initial stress on average in isolation (100-Hz condition) versus 57% when utterance-final in a carrier [lower line in panel (d)]. Adding a following speech context is inconsequential, reducing overall bias by a mere 4% [lower line in panel (b)]. Embedding the word in a preceding and following speech context [lower line in panel (c)] seems to combine the two effects: here mean bias is reduced to 53% (i.e., no bias re-
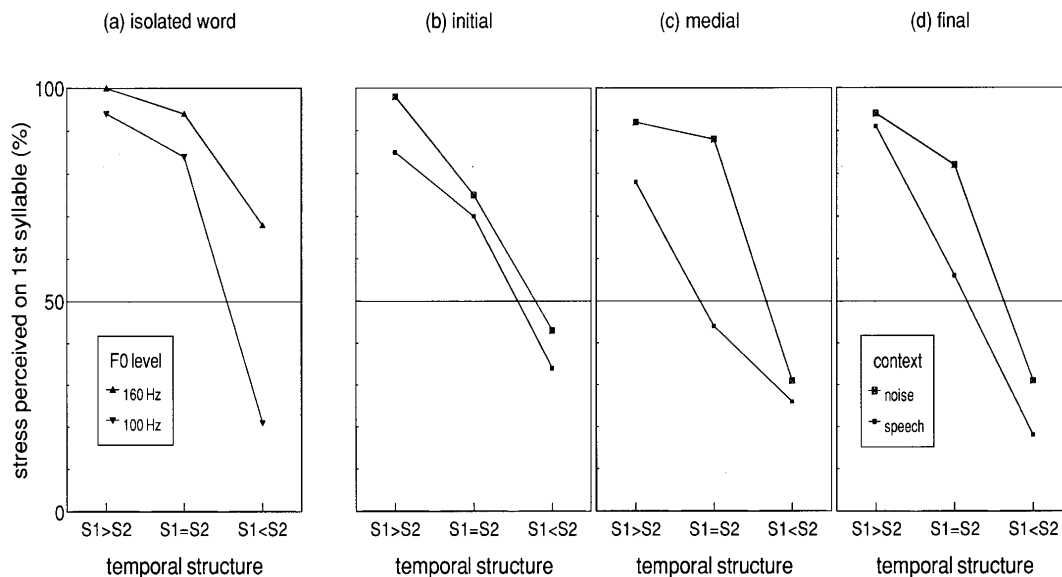
FIG. 5. Percent stress perceived on first syllable broken down by temporal structure of stimulus (cf. Fig. 1). In panel (a) results for words in isolation, with 160- and 100-Hz level pitch conditions are indicated separately. Panels (b)–(d) contain the data for words presented in context (initial, medial, final) with separate lines for speech and noise contexts. Here targets words and speech contexts were always synthesized on a 100-Hz monotone.

mains). Clearly it is the presence of a leading context that is mainly responsible for the reduction of bias.

Finally, the data bear out that the context effect is restricted to speech. A preceding noise burst does not reduce bias for first position at all, and in fact increases the bias by 1% on average. A following noise burst increases the bias by 3%, and the combined effect of preceding and following noise reduces bias by 1%. These values are just random fluctuations around the mean bias of 69% stresses reported for the first position established for isolated words.

D. Conclusions

First of all, the results of this experiment rule out the hypothesis that stress bias for initial position is the result of perceptual compensation for final lengthening. For this to be the case, the bias should disappear, or at least be severely reduced, when the crucial word is immediately **followed** by a spoken context. This result simply was not obtained.

Two opposing answers were suggested as to why a **preceding** spoken context reduces bias for stress perception in the initial position. Either a preceding auditory stimulation reduces the prominence of the leading syllable, or it prevents the listener from inferring a pitch jump. Our results unequivocally support the second hypothesis. On the one hand, the silent pitch-rise theory predicts an increase in stress bias favoring initial position when isolated words are given a higher (level) pitch. This effect manifested itself quite plainly, as it did in the previous experiment. The alternative view predicted a reduction of bias of equal magnitude for a preceding speech context as for a preceding stretch of noise of similar duration and intensity. This prediction was not borne out: no reduction was obtained when a noise burst was added as the auditory context. Bias was substantially reduced, however, when the context was speech. Conse-

quently, only the silent pitch-rise hypothesis remains compatible with all the experimental data found in our experiments.

V. GENERAL DISCUSSION

The starting point for this research was our hypothesis that the overall bias for initial position as reported in the literature could, at least partly, be ascribed to experimental artifacts such as the use of monotonous pitch and the presentation of stimulus words in isolation.

Four acoustic variables were pitted against one another. For reference purposes, one of these concerned a cue that has traditionally been mentioned as a reliable correlate of stress, viz., duration. The other three have never been considered as stimulus properties that might influence the perception of stress location: the presence versus absence of a (preceding) context, the characteristics of the excitation source (e.g., buzz versus noise), and the general pitch level of the isolated word stimulus.

For languages such as Dutch and English, duration generally ranks second in importance as a cue to stress position, after pitch movements, but before intensity (Fry, 1955, 1958; Morton and Jassem, 1965; van Katwijk, 1974) and vowel quality variation (Fry, 1965; Rietveld and Koopmans-van Beinum, 1987). Though we have no way of comparing the full set of cues, the traditional effect of duration seems to be replicated in our experiments: changing duration in our stimulus set could swing stress judgments by some 30% (experiments I, II, and IV) or even by 90% (experiment III), where the duration difference was larger.

With the majority of our stimulus words, stress was typically perceived in the onset position, even in nonsense items whose stress should fall on the final syllable according to the Dutch stress rules. Only compound adjectives seem to be exempt from this effect (experiments I and II). Notice, however, that this type of word has initial stress only if it is

immediately followed by a strong stress within the same phonological phrase (cf. van Heuven, 1987); this condition was not fulfilled in the present experiment.

We have also managed to demonstrate, however, that much of this bias is a function of the particular choice of stimulus presentation. Embedding the target in a preceding context fluently coarticulated with it effectively reduces stress bias for the initial position by about 20% (experiments II and IV). A similar reduction of initial position bias was found as an effect of the source signal (experiment I). When a monotonous or declining pitch was replaced by a gently inclining pitch or by a noise source, the distribution of stress judgments over the two positions changed from roughly 75%–25% to 50%–50%. Changing the overall pitch level of an isolated word between 70 and 160 Hz could swing stress judgments for the first syllable from 76% to 91% (experiment III, temporally neutral stimuli) or even counteract the lengthening of the second syllable in experiment IV (23% vs 69% stress on the first syllable for 100- and 160-Hz level pitches, respectively).

Generally, then, these nontraditional effects on stress perception are substantial and robust. They did not occur in just one experiment, but were replicated in several experiments in our series. On the whole we have been successful in showing that the well-known phenomenon of stress bias for the first position in Dutch, English, and many other languages is not merely a reflection of a statistical preference in the language to have words with stress on the first syllable. We now know that acoustic properties of the stimulus are at least partly responsible for the bias.

We have defended the view that the bias phenomenon is best explained by assuming that the listener interprets the difference between the actual pitch onset of the stimulus and the estimated bottom frequency of the speaker's pitch range as a silent pitch rise cuing stress. This explanation was the only one compatible with all the experimental data. Yet, it leaves many questions open, and in fact does not necessarily exclude other explanations that we have not yet considered.

As an example of the latter, it was suggested to us by Klatt (1987) that listeners may have learned through experience the average pitch of utterance-initial stressed and unstressed syllables. They might use this knowledge to decide, for each utterance onset, whether the first syllable is more likely to be stressed than unstressed. Although this suggestion is worth following up, there are at least two arguments that detract from its viability. First, listeners know that the onset pitch depends on the duration of the utterance: the longer the utterance, the higher the onset pitch (Cohen *et al.*, 1982; de Pijper, 1983; Grosjean, 1983). So it will often be the case that an unstressed syllable at the beginning of a long utterance will have an onset with a higher pitch than a stressed syllable at the beginning of a short utterance, i.e., there is potential overlap in the onset pitches of utterance-initial stressed and unstressed syllables (when both start on the low declination line). Second, and most important, Klatt's suggestion cannot explain why stress bias disappears when the excitation signal is noise.

It is obvious that the 'silent' or 'inferred' rise from the assumed bottom of the speaker's pitch range to the onset pitch of the stimulus is a less effective stress cue than a real rise or than a 'virtual' rise that takes place in the middle of an utterance during the silent interval of voiceless consonants. Recall the fact that raising the overall pitch level in experiment III by from 70 to 160 Hz never increased the number of judgments for stress in the initial position by more than ~15%. Had this been a real 90-Hz pitch rise, its effect on stress perception would have obliterated all other effects.

Finally, the silent rise hypothesis is clearly asymmetrical. We suggest that the listener makes specific assumptions with regard to the prosody, especially the pitch, of the period of silence preceding the stimulus **onset only**. We believe that the listener makes no assumptions as to the prosody of a following period of silence. This is necessary in order to exclude the possibility of silent falls generating stress bias in the final position. In making our mechanism asymmetrical we look to the fact that properties of the final part of the stimulus persist in auditory memory for a relatively long time, and decay only slowly during the following silence (cf. Plomp, 1964; van den Broecke and van Heuven, 1983). Therefore, it is important to find out after what period of silence the listener treats the next stimulus onset as a new utterance rather than as a continuation of the preceding utterance. Knowledge of this may help us to understand the mechanism by which the listener generates assumptions about the prosody of a silent context.

APPENDIX

Parameter displays and oscillograms for stimuli used in experiments I, II, and IV, are shown in Fig. A1. Each column of points represents a time frame of 12.8 ms. The upper panel displays intensity (dB), the middle panel contains an oscillogram, and the bottom panel shows the center frequencies (in kHz) of the formants $F_1 - F_4$. The length of the vertical lines drawn around the center frequencies represents the formants' perceptual prominence (Q factor, i.e., bandwidth divided by center frequency). A dot below the intensity panel indicates that the frame is voiceless. In order to improve legibility formant/bandwidth tracks were smoothed over an 8-frame time window. No smoothing was used during stimulus generation, however. Neutral durations of first and second vowels in target words are indicated in Table AI (rounded off to the nearest 10 ms).

2448   J. Acoust. Soc. Am., Vol. 100, No. 4, Pt. 1, October 1996

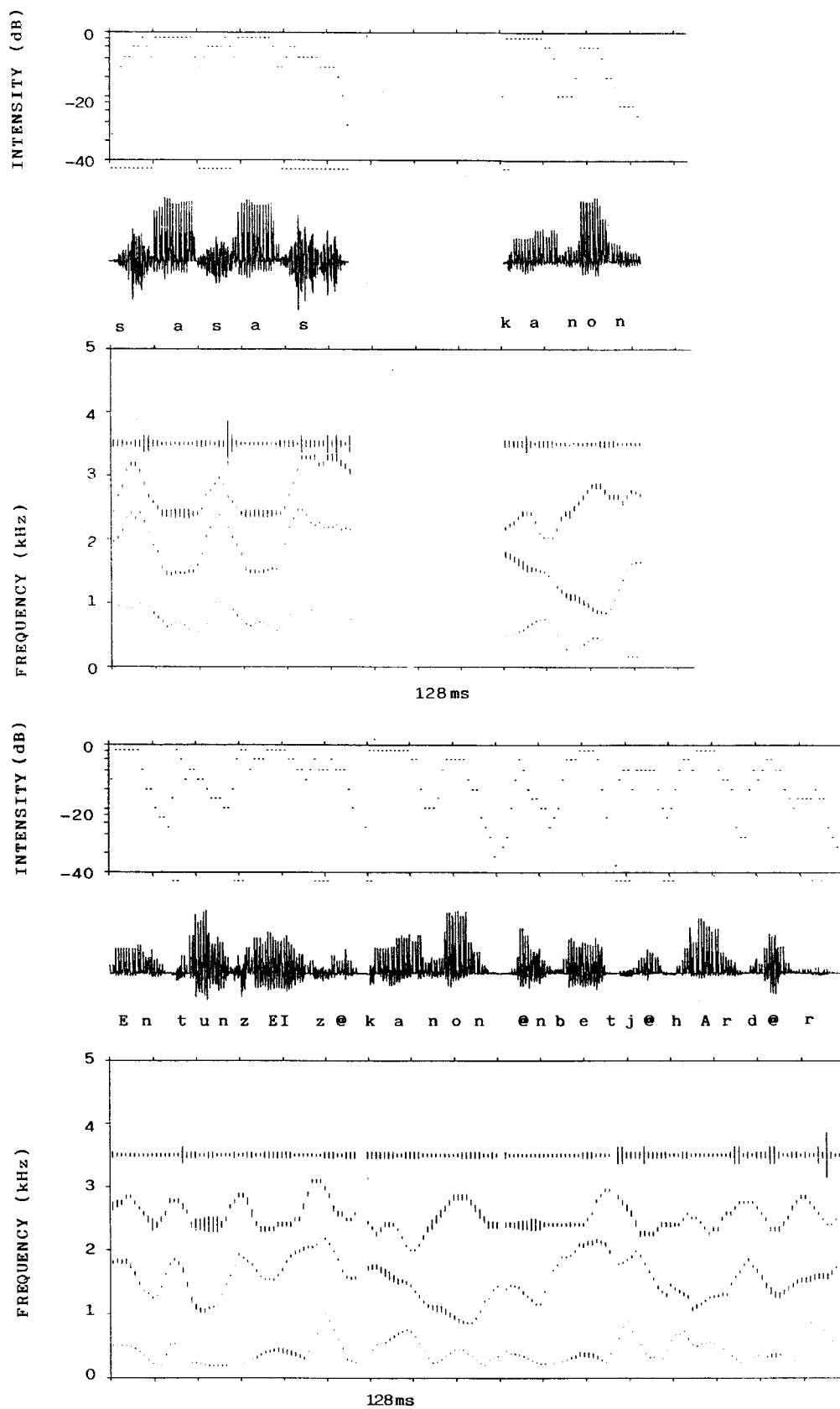V. J. van Heuven and L. Menert: Stress position bias   2448

FIG. A1. Parameter displays and oscillograms for stimuli used in experiments I, II, and IV.

TABLE AI. Neutral durations (rounded off to the nearest 10 ms) of first and second vowels in the target stimulus words used in experiments I–IV.

Lexical			Morphological			Compound adjective			Reiterant		
Word	V1	V2	word	V1	V2	word	V1	V2	word	V1	V2
kaːnɔn	170	100	voːrkoːmə	170	160	lɪχtχrɛɪs	80	170	saːsaːs	180	180
sɛrviˑs	140	100	doːrloːpə	170	150	mɑtblɑʊ	100	160	siˑsiˑs	100	100
			oːvərkoːmə	180	160						

[1] We are not aware of any other valid cross-linguistic study of bias in stress perception. In one apparent counterexample a comparison was made of English (Morton and Jassem, 1965) versus Polish (Jassem *et al.*, 1968) listeners' responses to the same set of disyllabic stimuli. Although the stress distributions of English (initial stress dominant) and Polish (fixed penultimate stress) differ, the predicted bias position coincides since the penultimate position is also the initial position in a disyllable.

[2] In recent studies, Caspers and van Heuven (1993a, b) presented evidence that the segmental alignment of the accent-lending Dutch pitch rise is more adequately expressed in terms of a coincidence of its onset with the beginning of the syllable.

[3] Preliminary analyses showed no systematic effects or interactions of lexical instantiations with any of the other factors in the design.

[4] It was pointed out to us by Ilse Lehiste that this claim may not hold for tone languages with (lexical) tone contrasts in the stressed syllable. Although it is possible to whisper in tone languages (see, e.g., Kloster-Jensen, 1958; Miller, 1961; Wise and Chong, 1957) no studies have been run on the perceptual location of stress in tone languages in whispered speech.

[5] Although lexical type was a significant effect in experiment I, we felt justified in making this decision since there were no interactions between lexical type and any of the other factors. Making lexical types the error term is conservative with respect to our predictions. Had it been a factor in the ANOVA, the effects of context, source signal, and temporal version would have been (even) more strongly significant.

[6] This does not necessarily imply that the relationship between physical size (low formants) and low pitch should be found for any given individual. Listeners may resort to vocal stereotypes when giving this type of judgment.

Berinstein, A. E. (**1979**). ''A cross-linguistic study on the perception and production of stress,'' UCLA Work. Papers Phonet. **47**, 1–59.

Broecke, M. P. R. van den and Heuven, V. J. van (**1983**). ''Effect and artifact in the auditory discrimination of rise and decay time: Speech versus nonspeech,'' Percept. Psych. **33**, 305–313.

Brück, H. D., and Teuling, D. J. A. (**1982**). ''Integrated voice synthesizer,'' Electron. Components Appl. **4**, 72–79.

Carlson, R., Elenius, K., Granström, B., and Hunnicut, S. (**1985**). ''Phonetic and orthographic properties of the basic vocabulary of five European languages,'' Speech Transmission Laboratory—Quarterly Progress and Status Report **1**, 63–94.

Caspers, J., and Heuven, V. J. van (**1993a**). ''Effects of time pressure on the phonetic realization of the Dutch accent lending pitch rise and fall,'' Phonetica **50**, 161–171.

Caspers, J., and Heuven, V. J. van (**1993b**). ''Perception of low-end vs. high-end anchoring of accent-lending pitch rises,'' in *Proceedings of an ESCA Prosody Workshop on Prosody*, edited by D. House and P. Touati (Working papers, Department of Linguistics, Lund University), Vol. 41, pp. 188–191.

Cohen, A., and Hart, J.'t (**1967**). ''On the anatomy of intonation,'' Lingua **19**, 177–192.

Cohen, A., Collier, R., and Hart, J. 't (**1982**). ''Declination: construct or intrinsic feature of speech pitch?'' Phonetica **39**, 254–273.

Delattre, P. (**1965**). *Comparing the Phonetic Features of English, German, Spanish, and French* (Julius Gross Verlag, Berlin).

Elsendoorn, B. A. G., and Hart, J. 't (**1982**). ''Exploring the possibilities of speech synthesis with Dutch diphones,'' IPO Annu. Prog. Rep. **17**, 63–65.

Elsendoorn, B. A. G., and Hart, J. 't (**1984**). ''Heading for a diphone synthesis system for Dutch,'' IPO Annu. Prog. Rep. **19**, 32–35.

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 6th International Congress of Phonetic Science*, edited by E. Zwirner and W. Bethge (Karger, Basel), pp. 306–311.

Grosjean, F. (**1983**). ''How long is the sentence? Prediction and prosody in the on-line processing of language,'' Linguistics **21**, 501–529.

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).

Hart, J. 't, Nooteboom, S. G., Vogten, L., and Willems, L. F. (**1982**). ''Manipulaties met spraakgebruik'' (Manipulations of speech sounds), Philips Techn. Tijdschrift **40**, 108–109.

Hermes, D. J. and Rump, H. H. (**1994**). ''Perception of prominence in speech intonation induced by rising and falling pitch movements,'' J. Acoust. Soc. Am. **96**, 83–92.

Heuven, V. J. van (**1987**). ''Stress patterns in Dutch (compound) adjectives: acoustic measurements and perception data,'' Phonetica **44**, 1–12.

Heuven, V. J. van and Hagman, P. (**1988**). ''Lexical statistics and spoken word recognition in Dutch,'' in *Linguistics in the Netherlands 1988*, edited by P. Coopmans and A. Hulk (Foris, Dordrecht), pp. 58–68.

Jassem, W., Morton, J., and Steffen-Batog, M. (**1968**). ''The perception of stress in synthetic speech-like stimuli by Polish listeners,'' Speech Anal. Synth. **1**, 289–308.

Kager, R. J. (**1989**). *Stress and Destressing in English and Dutch* (Foris, Dordrecht).

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

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

Klatt, D. H. (**1987**). Personal communication.

Kloster-Jensen, M. (**1958**). ''Recognition of tones in whispered speech,'' Word **14**, 187–196.

Künzel, H. J. (**1989**). ''How well does average fundamental frequency correlate with speaker height and weight?'' Phónetica **46**, 117–125.

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

Laver, J., and Trudgill, P. (**1979**). ''Phonetic and linguistic markers in speech,'' in *Social Markers in Speech*, edited by K. Scherer and P. Giles (Cambridge U.P., Cambridge), pp. 1–32.

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

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

Liberman, M., and Prince, A. (**1977**). ''On stress and linguistic rhythm,'' Linguistic Inquiry **8**, 249–336.

Miller, J. D. (**1961**). ''Word tone recognition in Vietnamese whispered speech,'' Word **17**, 11–15.

Morton, J., and Jassem, W. (**1965**). ''Acoustic correlates of stress,'' Lang. Speech **8**, 148–158.

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

Neijt, A. H., and Heuven, V. J. van (**1992**). ''Rules and exceptions in Dutch word stress,'' in *Linguistics in The Netherlands 1992*, edited by R. Bok-Bennema and R. van Hout (Benjamins, Amsterdam), pp. 185–196.

Nooteboom, S. G. (**1972**). ''Production and perception of vowel duration, a study of durational properties of vowels in Dutch,'' Doctoral dissertation, Utrecht University.

Nooteboom, S. G., and Doodeman, G. J. N. (**1980**). ''Production and perception of vowel length in spoken sentences,'' J. Acoust. Soc. Am. **67**, 276–287.

Pierrehumbert, J. (**1979**). ''The perception of fundamental frequency declination,'' J. Acoust. Soc. Am. **66**, 363–369.

Pijper, J.-R. de (**1983**). *Modelling British English Intonation* (Foris, Dordrecht).

Plomp, R. (**1964**). ''Rate of decay of auditory sensation,'' J. Acoust. Soc. Am. **36**, 277–282.

Repp, B. H. (**1986**). Personal communication.

Rietveld, A. C. M., and Koopmans-van Beinum, F. J. (**1987**). ''Vowel reduction and stress,'' Speech Commun. **6**, 217–230.

Vogten, L. (**1984**). ''Analyse, zuinige codering en resynthese van spraak'' (Analysis, economical coding, and resynthesis of speech), Doctoral dissertation, Technical University, Eindhoven.

Wise, C. M., and Chong, L. P.-H. (**1957**). ''Intelligibility of whispering in a tone language,'' J. Speech Hear. Dis. **22**, 335–338.