AN UNUSUAL EFFECT ON THE PERCEPTION OF STRESS

VINCENT J. VAN HEUVEN

Dept. of Linguistics/Phonetics Laboratory,
Leyden University, P.O. Box 9515,
2300 RA Leiden, The Netherlands

ABSTRACT

In our research we try to understand why (Dutch) listeners tend to perceive the first syllable of a phonated reiterant word as stressed when presented in isolation, but not when the word was whispered or generated in a preceding (but not following) spoken context. We suggested that the listener interpretes the periodicity onset of an isolated word as a(n accent-lending) pitch rise relative to the bottom of the speaker's pitch range. Some consequences of this view are further tested in the present study.

INTRODUCTION

Listeners tend to perceive lexical stress on the first syllable in isolated words. This stress bias is most conveniently demonstrated with so-called reiterant speech, i.e. words made up of repetitions of an identical syllable. The proportion of initial stress perceived in such stimuli was found to range between 59% (binary free choice in English [1]) and 80% (4-alternative forced choice in English [2]), with intermediate values reported for Dutch (65%, ternary free choice [3]), and Polish (79% binary free choice [4]).

The basic question that we set out to answer is: What causes this bias towards perceiving the initial syllable as stressed? Although it would seem obvious to relate this bias to the statistical distribution of lexical stress in the language, which in the above experiments favours onset position, we have reason to believe that the bias is at least partly caused by a perceptual mechanism of a more general nature: Van Heuven & Menert [5] established that:

(i) presenting a target word in a preceding speech context reduces initial position bias;
(ii) replacing the periodic (buzz) source signal in isolated targets by a noise source ("whisper") reduces bias likewise;
(iii) increasing the fundamental frequency (F0) of the periodic source in isolated words from a level 100 Hz to a level 160 Hz increases bias.

These findings were explained as follows. We assume that the listener, on hearing the 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: "virtal rise"), and taken as a cue for stress on the first syllable.

When the stimulus is a-periodic, no actual pitch can be determined, and no pitch jump can be inferred. Hence bias disappears in whispered targets. When the target is preceded by a spoken carrier phrase, the reference pitch is provided by the context. Since, in our experiment, the pitch was level throughout the entire 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 shall vary the onset pitch between 70 Hz (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. In our earlier experiment F0 varied in just two steps (100 versus 160 Hz). Unfortunately the 160 Hz condition comprised only a small proportion of the stimulus material. It was therefore an unusual condition within the experimental context, which by itself may have contributed to the increase in bias. In the present experiment each of the four pitch levels occurs equally often, so that the effect of pitch level can be examined more conclusively.

Secondly, we shall speculate on the mechanism by which the listener generates the reference pitch. Here we hypothesize that the listener assumes the presence of 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. 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 virtual 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.

METHOD

Seventy-two stimuli were generated by a DEC Micro-VAX-II computer using the LVS speech analysis and (re-) synthesis software developed at IPO-Eindhoven. The stimuli comprised 36 versions of the Dutch word 'kanon' and another 36 of the nonsense word 'saasaas'. Both words are ambiguous for (lexical) stress position: 'kanon' with stress on the first syllable means 'polyphonic hymn' but 'gun' with stress on the second syllable. The Dutch stress rules are compatible with stress on either syllable in the reiterant sequence 'saasaas' (see further [5]). The words were synthesised from diphones which had been excerpted from the accented syllable in nonsense words of the type [CiəCi'VCiə], and stored in parametrised form in computer memory using the AAP-LPC analysis program provided by the LVS-package, using 5 formants and 5 bandwidths in the frequency band up to 4.5 kHz, calculated over a 25-ms time-window that was shifted along the time axis in 10-ms steps. Given their origin, all the sound seg-ments in our diphone synthesis are equally suggestive of strong, primary stress. This is what makes this type of synthesis ideal for our purpose: in the absence of any experimenter-induced parameter changes the distribution of perceived stress should be neatly balanced over the two syllables in our words.

The 36 versions of each word were then obtained through orthogonal combination of three factors:

(i) F0 was varied in 4 steps: 70, 100, 130, and 160 Hz; F0 was level throughout the duration of the stimulus word.
(ii) Formant range was varied in 3 steps. Starting from the formant frequencies F1 through F5 as calculated by the LPC-analysis, a type of voice was synthesised that was typical of a large male (formants lowered to .85 of their original frequencies), and another type that suggested a small male (formants raised to 1.20 of their original values).
(iii) Temporal type was varied in three steps. Next to the original version consisting of temporally unadjusted diphones, a version with a noticeably longer first syllable was obtained by lengthening the steady state portion of the first vowel by 50 ms while shortening the second vowel by the same amount. A complementary version with a longer second syllable was generated by reversing this procedure.

The 72 tokens were recorded onto audio tape in quasi-random order, preceded by 8 practice items. This tape was presented twice to 11 Dutch listeners over a good quality sound reproduction system (Quad ESL-63) in a small, well insulated lecture soom with some soft paneling attached to walls and ceiling so as to limit reverberation. Listeners were instructed to decide for each stimulus whether the stress was on the first or on the second syllable, with binary forced choice. They were to indicate their choice by ticking the appropriate syllable on answer sheets that contained a listing of the 72 stimuli in the order in which they appeared on the tape, typed in ordinary Dutch spelling.

RESULTS

Examining table I, which presents % stress perceived on the first syllable broken down by pitch level, formant setting, and temporal version, we observe the following:

TABLE I: Percent stress perceived on first syllable broken down by F0-level, temporal version, and formant setting.

	F0-level (Hz)			
	70	100	130	160
1st syll long				
formants lowered	93	93	98	98
formants neutral	95	93	93	100
formants raised	98	91	100	98
1st/2nd syll equal				
formants lowered	61	80	86	86
formants neutral	80	84	89	93
formants raised	86	89	86	93
2nd syll longer				
formants lowered	7	9	9	28
formants neutral	14	9	7	28
formants raised	3	23	11	11

1. Manipulating the relative duration of first and second syllable produces 96% stress perceived on the lengthened first sylable, 85% on a temporally neutral first syllable and 13% on a shortened first syllable. This effect was significant by a classical three-way ANOVA with pitch level, formant setting and temporal version as fixed factors, $F(2,141)=1120.1$ ($p < .001$). This effect is well-known from the literature (cf. [6] and references given there) and can therefore serve as a baseline against which the strength of the remaining factors can be evaluated.
2. Changing F0 level has a clear effect on stress perception. Bias for the first syllable increases monotonically with F0 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, $F(3,140)=8.5$ ($p < .001$). The effect of F0 is most pronounced in the temporally neutral versions with 76, 84, 87, and 91% stress perceived on the first syllable, $F(3,45)=4.4$ ($p=.020$).

Counter to our prediction, bias does not disappear completely at 70 Hz. Whether a further reduction of bias can be obtained by lowering the F0 level still further remains doubtful: when constructing our stimuli we had to abandon pitches below 70 Hz, as these sounded highly unnatural (rough, creaky voice quality).

3. 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, $F(2,141)=1.9$ (ins).

CONCLUSION

Both in this and in our previous experiments we have demonstrated that the perception of stress is not solely dependent on differences between F0, intensity, duration and timbre within the word or utterance, as is generally maintained in the literature (cf. [6]).

We have presented convincing evidence here that the onset frequency of an otherwise perfectly level pitch influences the perception of lexical stress in isolated words: the higher the pitch level, the greater the bias favouring perceived stress on the first syllable. Generally, the results confirm our claim that stress bias is caused by the listener's perceiving the discrepancy between the actual pitch onset and some low reference pitch as a virtual pitch rise cueing stress, i.e. an auditory meachanism, rather than by the listener's knowledge of the statistical distribution of lexical stress in the language.

However, we have not been able to confirm our suspicion that the reference pitch is derived from the average formant setting in the voice of the speaker, which negative finding prompts at least two questions for further research. Firstly, is it really true that listeners associate a particular pitch range with a given formant setting, and secondly, could it be the case that the reference pitch is fixed and speaker independent? These questions will be taken up in our future research.

References:

[1] J. Morton, W. Jassem, "Acoustic correlates of stress", Language and Speech, 8, 1965, 148-158.

[2] A.E. Berinstein, "A cross-linguistic study on the perception and production of stress", UCLA Working Papers in Phonetics, 47, 1979, 1-59.

[3] A.F. van Katwijk, Accentuation in Dutch, an experimental linguistic study, Van Gorcum, 1974.

[4] W. Jassem, J. Morton, M. Steffen-batog, "The perception of stress in synthetic speech-like stimuli by Polish listeners", Speech Analysis and Synthesis, 1, 1968, 289-308.

[5] V.J. van Heuven, L. Menert, "Linguistic and perceptual causes of stress perception bias", ms. submitted to Journal of the Acoustical Society of America, 1968.

[6] I. Lehiste, Suprasegmentals, MIT-Press, 1970.