AN UNUSUAL EFFECT ON THE PERCEPTION OF STRESS

VINCENT J. VAN HEUVER

Dept. of Linguistics/Psychology, 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 retracted word as stressed when presented in isolation, and why this tendency is more marked when the word is generated in a preceding (but not following) spoken sentence. The hypothesis that is generated is that the listener interprets the periodicity onset of an isolated word as an accent landing pitch that is 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 retracted speech, i.e. words made up of repetitions of a syllable. The proportion of initial stress perceived in such stimuli was found to range from 60% for the primary free choice in English [1], and BIDI (4-alternative forced choice in English [2]), with intermediate values reported for Dutch [65], ternary free choice [3]), and Polish [72] binary choice ([47]).

The basic question that we set out to answer is: What causes this bias towards perceiving initial stress as stressed? Although it would seem obvious to base this bias on the statistical distribution of lexical stress in the language, which in the above experiments favored 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 Heuver & Honset [5] showed that

(1) presenting a target word in a preceding speech context reduces initial position bias;
(2) replacing the melodic (bend) noun with a, or generated in a preceding (but not following) spoken sentence.

Se 55.1.1

different formant settings: starting from a normal male with an average formant setting, we simulated a large individual variation of 20 formant settings, as well as small individual with raised formants. If it is true that listeners will perceive sentences with a low-pitched voice, the listener will generate a low-pitched voice, so that a larger vocal rise will be perceived relative to the voice, which in turn is perceived as a stronger bias for stimuli with unshifted formants.

METHODOLOGY

Twenty-two stimuli were generated by a DHC Micro

XIX-II computer using the USP speech analysis and synthesis software developed at EPO.

Furthermore, the stimuli comprised two vowels of the Dutch word 'kanon' and another 36 of the nonsense word 'nasanen'. Both words are ambiguous for (lexical) stress position: 'kanon' with stress on the first syllable, "nasanen" with stress on the second syllable. This Dutch stress pattern is in line with the vernacular, where the latter can be retracted in the relevant sequence 'nasaaanen' (see further [5]). The words were synthesized from digitized from phonies which had been excepted from the accepted syllable in nonsense words of the type /gaphi/ and stored in parameters for foreseen in computer memory using the AAP-UPC interface program by the USP-package, using 5 formants and 5 bandwidths in the frequency band up to 4.5 kHz, calculated over a 25-ns time-window that was shifted along the time axis in 10-ns steps. Given the original, all the sound segs in our diphthong analysis are equally suggestive of strong, primary stress. This makes this type of stress pattern is chosen for our purposes, but the experiments with unshifted parameter changes the distribution of perceived syllables of the two syllables in our list of stimuli.

The 36 versions of each word were then percribed through orthogonal combination of these factors:

(1) PD was varied in 4 steps: 70, 100, 130, and 160 Hz; PD was level throughout the duration of the stimulus.

(2) Formants can be varied in 3 steps. Starting from the formant frequencies F1 through F5 as calculated in the IPPC analysis, a type of formant synthesis that was typical of a large male sample (formants F1 through F5 of original frequencies), and another type that have been shifted upwards more than 1.20 of their original values.

(1) Temperal type can vary three steps: Next to the original version consisting of temporally unshifted stimuli, a version with pitch level, formant setting and temporal version as fixed factors, F1(100/33.3, 50,%)

The 72 recorded was played onto audio tape in random order, preceded by 8 practice items.

This tape was presented twice to 12 Dutch listeners over a good quality sound reproduction system (RLS-63) in a soundproof stimulus room of the University of Leiden. The listeners were instructed to judge whether the stress was on the first or on the second syllable, with binary forced choice experiment. They were asked to indicate their choice by ticking the appropriate syllable on an answer sheet containing a list of the 72 stimuli in the order in which they appeared on the tape, typed in ordinary Dutch spelling.

RESULTS

Examining table 1, which presents I stress perception on the first syllable broken down by pitch level, formant setting, and temporal version, we observe the following:

<table>
<thead>
<tr>
<th>RESULTS</th>
<th>1st syll long</th>
<th>1st syll equal</th>
<th>2nd syll long</th>
</tr>
</thead>
<tbody>
<tr>
<td>PD level</td>
<td>70 130 160</td>
<td>40 60 80</td>
<td>40 60 80</td>
</tr>
<tr>
<td>formant lowered</td>
<td>91 3 90 98</td>
<td>91 3 90 98</td>
<td>91 3 90 98</td>
</tr>
<tr>
<td>formants raised</td>
<td>91 3 90 98</td>
<td>91 3 90 98</td>
<td>91 3 90 98</td>
</tr>
</tbody>
</table>

1. Manipulating the relative duration of first and second syllable produces 96% stress perception on the long syllable, and 4% on a temporally neutral first syllable and 13% on a temporally neutral second syllable. This effect was significant by a classical three-way analysis of variance with pitch level, formant setting and temporal version as fixed factors, F1(100/33.3, 50,%).

2. Changing PD and formant settings has no clear effect on stress perception. Bias for the first syllable is consistent, and the formant settings remain unshifted (for PD level 70 Hz). This effect is smaller than that of temporal version. It is still substantial, F1(100/33.3, 50,%).

3. The effect of PD is most pronounced in the temporal neutral syllables, and the formants are shifted towards the first syllable, F1(100/33.3, 50,%).
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 mechanism, 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:


308 Se 55.1.3