# Perceptual cues of linguistic stress: intensity revisited

Agaath M.C. Sluijter & Vincent J. van Heuven
Holland Institute of Generative Linguistics /
Phonetics Laboratory
PO Box 9515, 2300 RA Leiden, The Netherlands

## ABSTRACT

*The general claim that intensity is a weak cue in the perception of stress is reconsidered. This claim is based on perception experiments in which intensity was varied in a naive way: all parts of the spectrum were increased with the same amount of energy. However, stressed syllables are produced with more vocal effort. If a speaker produces more vocal effort, higher frequencies increase more than lower frequencies. We show that the intensity differences as a function of stress are mainly located above 0.5 kHz. Varying intensity in this way would be much more realistic.*

## INTRODUCTION

The acoustical correlates of stress are pitch, duration, loudness and vowel quality. Of these, pitch and duration have been found the most important perceptual cues; intensity and vowel quality are generally claimed to be of lesser importance (Beckman 1986 and references mentioned there).

Indeed, an appropriate pitch movement provides an overriding cue, but is present only when the stress coincides with an accent (marking focus). When words are spoken outside focus, the position of the stress has to be inferred from the remaining cues. In our research we set out to determine the relative importance of vowel duration and intensity as cues for stress for materials spoken in and outside focus.

Traditionally, the cue value of intensity has been examined in a relatively naive and unrealistic way by simply manipulating the overall volume of syllables (or vowels). However, stressed syllables are produced with greater vocal effort than unstressed syllables. If a human speaker expends more vocal effort, as is required for the realization of a stress, intensity does not change uniformly across the spectrum, but higher frequencies are increased more than lower frequencies. Gauffin and Sundberg (1989) investigated the spectral consequences of changes in SPL during vocalization. Results show that with increasing loudness the levels of the higher bands (1.0-2.0 and 2.0-4.0 kHz) of a sustained vowel /ae/ increase more than the energy in the lower bands (0-0.5 and 0.5-1.0 kHz). Brandt, Ruder and Shipp (1969) independently varied vocal effort and intensity of continuous speech stimuli; speech produced with greater effort was estimated louder, even when intensity was held constant. Glave & Rietveld (1989) also showed that greater vocal effort is related to greater perceived loudness. Consequently, we expect that the spectral tilt of a stressed syllable differs from its unstressed counterpart: the energy in the higher frequencies of the spectrum increases more than the energy in the lower frequencies as this stressed syllable is produced with more vocal effort. In a production experiment we measured the energy in four contiguous frequency bands of stressed and unstressed vowels spoken in and outside focus, using both lexical and reiterant tokens.

## METHOD

We selected the minimal stress pair /ká:non/-/ka:nón/ (cannon - canon) differing in stress position only. The target words were embedded in a carrier sentence: *Wil je [target] zeggen* 'Will you [target] say'. Targets were spoken in and outside focus. The condition with the target outside focus was realized by placing an accent on the word *zeggen*. In the other focus condition an accent was placed on the stressed syllable of the target, placing the target in focus. To control for influences of syllable structure, 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: /ná:na:/-/na:ná:/. The vowel /a:/ was chosen because it is the most open, longest vowel in Dutch. The resulting four stimulus types (2 stress positions * 2 focus conditions) with their reiterant versions were read by four male and six female speakers of standard Dutch. The speakers were recorded individually in a sound insulated booth, using semi-professional equipment. The subject's head was strapped to the chair so as to insure a constant distance between mouth and microphone.

Stimulus sentences were presented in normal Dutch orthography on a computer monitor in 8 random counterbalanced orders. Subjects always produced lexical and reiterant versions of each stimulus in immediate succession before going on to the next stimulus. After each stimulus, whether lexical or reiterant a 5s. pause was observed, during which interval the subject was required to inhale prior to initiating the next utterance.

## RESULTS

The 640 utterances (2 stress positions * 2 focus conditions * 2 versions, i.e. lexical & reiterant * 10 speakers * 8 repetitions) were digitized. We used four repetitions (orders 2, 3, 7 and 8) yielding 320 sentences.

F1-maxima of the vowel in each syllable were determined using smoothed (30 ms integration) resograms. It was not possible to determine these maxima adequately in the syllable *non*, so these syllables were not used for further research. We measured the energy in dB in four contiguous filter bands: 0-0.5, 0.5-1.0, 1.0-2.0 and 2.0-4.0 kHz of stressed and unstressed vowels. In all cases we performed one-way analyses of variance for each filter band separately, with stress as a fixed factor.

In Figure 1 the energy differences in the four contiguous filter bands between the stressed initial syllable of *kánon* and the unstressed initial syllable of *kanón* are presented, i.e. in paradigmatic comparison. The results are broken down by gender and focus condition.

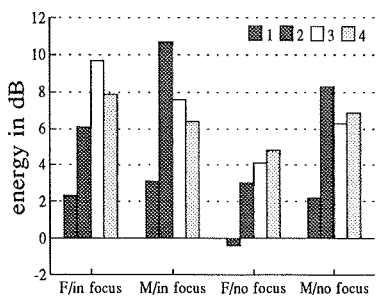


**Figure 1.** *Energy differences (in dB) in four contiguous filter bands (bar 1 = 0-0.5 kHz, bar 2: 0.5-1.0 kHz, bar 3: 1.0-2.0 kHz, bar 4: 2.0-4.0 kHz) between the stressed syllable 'ka' in 'kánon' and the unstressed syllable 'ka' in 'kanón'. The results are broken down by gender (M=male, F=female) and by focus condition (focus and no focus).*

As can be seen in Figure 1, there is an energy difference between stressed and unstressed vowels of about 3 to 11 dB in the highest three filter bands [all cases: $p<.01$], whereas there is only a slight difference in the base band [all cases: n.s., except the difference of 3.1 dB for male subjects in focus: $F_{1,31}= 9.4$, $p=.005$]. Male subjects have the largest energy difference between 0.5 and 1.0 kHz, whereas female subjects have the largest difference in the third filter band in words spoken in focus and in the fourth filter band in words spoken outside focus. This difference is probably due to the fact that female speakers usually have their formants at higher frequencies than males.

In Figure 2 the energy differences in the four contiguous filter bands between the stressed initial syllable of the reiterant version *nána* (replacing *kánon*) and the unstressed initial syllable of *naná* (replacing *kanón*) are presented in the left panel of the figure. The energy differences between the final syllables of the *nana* versions of the target words are presented in the right panel of the figure. As in figure 1, results are broken down by gender and focus condition.

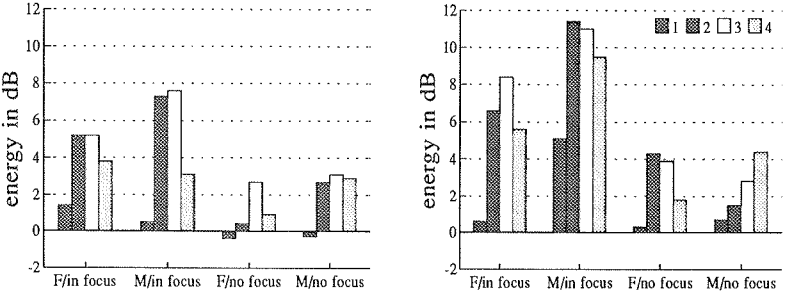


**Figure 2.** *Energy differences (in dB) in four filter bands between stressed and unstressed syllables 'na' (reiterant speech). In the left panel of the figure the initial syllables are presented, in the right panel the final syllables. Further see caption figure 1.*

The results for the initial syllables of the reiterant word pair (left-hand panel of figure 2), show that energy differences between stressed and unstressed vowels are mainly concentrated in the highest three filter bands with stronger effects in focus than outside focus. Outside focus female subjects only make a significant difference in energy between stressed and unstressed vowels between 1-2 kHz [$F_{1,47}=5.3$, $p=.025$, other cases $F<1$]. Male subjects only realize significant differences for targets spoken outside focus above 1 kHz [0-0.5 kHz: $F<1$; 0.5-1 kHz: $F_{1,31}=3.2$, n.s.; 1-2 kHz: $F_{1,31}=6.7$, $p=.014$; 2-4 kHz: $F_{1,31}=7.1$, $p=.013$]. The base band is hardly affected [all cases: n.s.].

The results for the final syllables of the reiterant word pair (right-hand panel of figure 2), show that for final syllables the effects are stronger, however comparable to the results for initial syllables (both reiterant and lexical).

In figure 3, the energy differences between the stressed and unstressed syllable of each member of the reiterant word pair are given separately. In the left panel of figure 3, the stressed and unstressed *na* of *kánon* are compared, in the right panel of *kanón* (syntagmatic comparison).

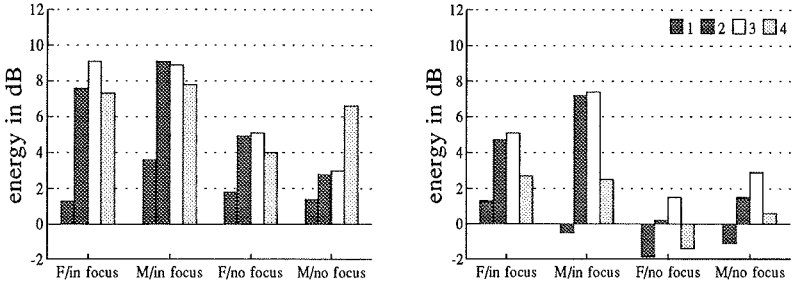


**Figure 3.** *Energy differences (in dB) of stressed and unstressed syllables within words. The left panel presents nána (for kánon), the right panel naná (for kanón). Further see caption figure 1.*

As can be seen in figure 3, the energy differences between the stressed and unstressed part of an initially stressed word are considerably larger than the differences in a finally stressed word, especially outside focus. Overall, the results are largely comparable to the results presented above.

## CONCLUSION

We conclude that the intensity of the base band (0-0,5 kHz) was hardly affected by stress; however, intensity in the higher bands (0,5-1, 1-2, and 2-4 kHz) increased in stressed syllables by 5-10 dB, with stronger effects for accented than for unaccented words.

We are about to perform a subsequent perception experiment in which we shall examine 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). Stimuli will be presented outside focus (without a pitch movement on the target) because of the fact that a pitch movement will always be the overriding cue in stress perception. We expect that realistic intensity manipulations (i.e., concentrated in the higher frequency bands) will provide stronger stress cues than can traditional intensity differences, and will be close in strength to duration differences. This finding then will rehabilitate the traditional claim that languages such as Dutch and English have dynamic (rather than melodic or temporal) stress.

## REFERENCES

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

J.F. Brandt, K.P. Ruder and I. Shipp jr. (1969), "Vocal Loudness and Effort in Continuous Speech", *J. Acoust. Soc. Am.*, Vol. 46, pp. 1543-1548.

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

R.D. Glave and A.C.M. Rietveld (1975), "Is the effort dependence of speech loudness explicable on the basis of acoustical cues", *J. Acoust. Soc. Am.*, Vol. 58, pp. 875-879.