

# A COMPARISON BETWEEN THE ACOUSTICS OF VOWEL AND CONSONANT REDUCTION\*

*R.J.J.H. van Son and Louis C.W. Pols*

## Abstract

Vowel reduction has been studied for years. It is a universal phenomenon that reduces the distinction of vowels in informal speech and in unstressed syllables. How consonants behave in situations where vowels are reduced is much less well known. In this paper we compare durational and spectral data for both intervocalic consonants and vowels segmented from read speech with otherwise identical segments from spontaneous speech. On a global level, it shows that consonants reduce like vowels when the speaking style becomes informal. On a more detailed level there are differences related to the type of the consonant.

## 1. Introduction

Vowel reduction is a well-established phenomenon that has found its place in phonetics textbooks (e.g., Clark and Yallop, 1990; O'Shaughnessy, 1987). Briefly summarized, vowels are pronounced more "sloppily" and with less distinction when speaking style is informal, or when the vowels are part of unstressed syllables (Koopmans-van Beinum, 1980). Essentially, vowels become more centralized and/or more like the phonemes that surround them. Although there is an ongoing debate about the details, vowel reduction is generally considered to be a universal phenomenon of speech (Van Bergem, 1995).

There have been studies that investigated acoustic and articulatory consonant reduction in relation to the corresponding vowel reduction, but these were generally limited to only a few classes of consonants, with only limited speech material (e.g., Byrd, 1994; Duez, 1995; Farnetani, 1995; Keating et al., 1994; Schmidt and Flege, 1995). From these studies it is difficult to discern the general effects of consonant reduction in "normal" speech situations.

To study how consonants reduce acoustically, we decided to contrast speech from reading aloud with that of "spontaneous" story telling. It is known that vowels spoken informally or spontaneously are severely reduced with respect to vowels that are read aloud from text. Consonant reduction too can be expected to show itself when informal speech is compared with read speech.

At the moment, any understanding of the way reduction affects the spectro-temporal structure of consonants and the way it influences consonant identification is seriously lacking. Therefore, it is difficult to point to specific features of articulation

---

\* This is an extended version of a paper presented at ICSLP '96 in Philadelphia (Van Son and Pols, 1996).

Table 1. Dutch consonants used in this paper. Columns: Place of articulation, Rows: Manner of articulation (Plosives, Fricatives, Nasals and Vowel-like).

	Velar	Palatal	Alveolar	Labial
Plosive	k g		t d	p b
Fricative	x ɣ	ʃ ʒ	s z	f v
Nasal	ŋ		n	m
Vowel-like	r	j	l ʎ	w

where reduction will affect the phonemic distinction of consonants. In this paper, we will limit ourselves to an inventory of consonant acoustics that parallel the vowel characteristics that are affected by vowel reduction. One important question that we want to answer is whether acoustic consonant reduction is indeed similar to vowel reduction.

Four aspects of vowels and consonants are studied to characterise consonant reduction:

1. Formant values
2. Duration
3. Center of Gravity of the spectrum (i.e., the “mean” frequency)
4. Sound energy difference between vowels and consonants

To be able to compare realizations across both speaking styles, we will ignore the ultimate form of consonant reduction, i.e., complete deletion.

## 2. Material and methods

For this study we used speech material of an experienced newscaster who first told some stories and anecdotes to an interviewer (who he knew quite well). This speech was transliterated and after some time he was asked to read aloud the transcription. This way, we obtained 2 times 20 minutes of speech (spontaneous and read). The whole orthographic script was transcribed to phonetic symbols by the Grapheme-to-Phoneme conversion module of an experimental speech synthesizer developed at the Department of Language and Speech at the University of Nijmegen. One of the authors checked the transcription and marked words for sentence accent by listening. All speech was sampled with 16 bit precision and 48 kHz sampling rate.

From the phonetic transcription, all Vowel-Consonant-Vowel (VCV) segments were located in the speech recordings (also those crossing word boundaries). In total, 4814 pairs of VCV realizations from corresponding positions in the read and spontaneous utterances were identified. For 1847 VCV of these pairs, both members were realized with identical syllable structure, syllable boundary type, and sentence and word stress. Of these 1847 VCV-pairs, a random subset of 791 pairs has been analyzed in detail for this paper (see table 1 and 2) and will be used here to study consonant reduction in more detail.

Table 2. Number of matched VCV pairs per consonant (ignoring voicing).

	Velar	Pal	Alve	Lab	Total
Plos	63		65	61	189
Fric	77	3	63	75	218
Nasal	14		72	63	149
V-like	60	21	94	60	235
Total	214	24	294	259	791

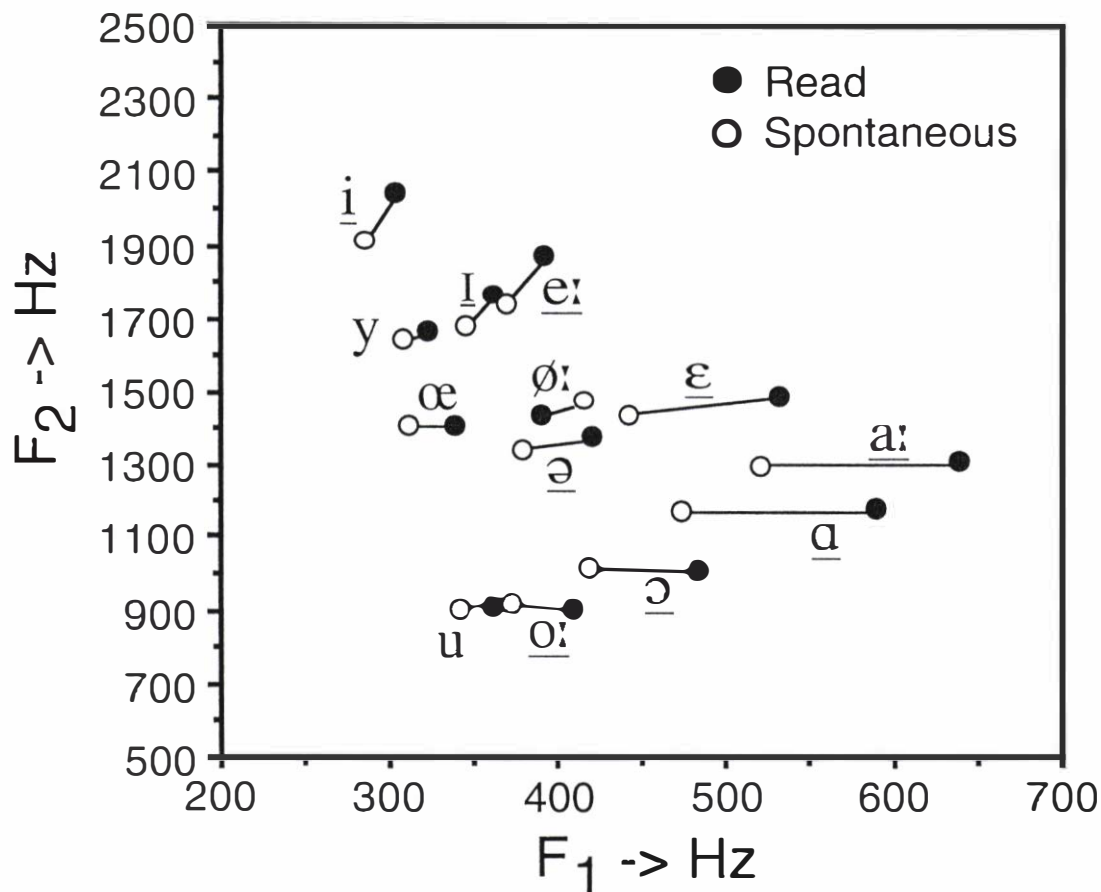


Figure 1: Spectral reduction in Dutch vowel space. Underlined symbols indicate statistical significant differences in either direction ( $p \leq 0.001$ , two tailed Sign test).

Phoneme boundaries were placed using a waveform display with audio feedback (Boersma) combined with synchronized displays of the Harmonicity-to-noise ratio, total energy, and the spectral balance, i.e., energy in the high- (above 3 kHz) versus low- (below 750 Hz), high- versus mid- (between 750 and 3000 Hz), and mid- versus low-frequencies. In cases where none of the displays suggested a boundary, audio cues were used exclusively. The boundaries between vowels and consonants were placed preferably on waveform zero-crossings that corresponded to “visible” changes in the spectral composition of the waveform. If present, priority was given to spectral changes that indicated the start or end of a constriction (e.g., abrupt changes in the spectral balance). LPC formant tracks were extracted using the Split-Levinson algorithm (after down sampling to 10 kHz, using 5 pole zero pairs).

### 3. Results

#### 3.1. Formant values

Vowel reduction is characterized by a centralization of the distribution of steady-state values in the  $F_1/F_2$  plane. The vowels from the spontaneous VCV segments used in this study show such a centralization with respect to those from read VCV segments (Figure 1, see also an independent analysis of the same speech, Koopmans-van Beinum, 1992).

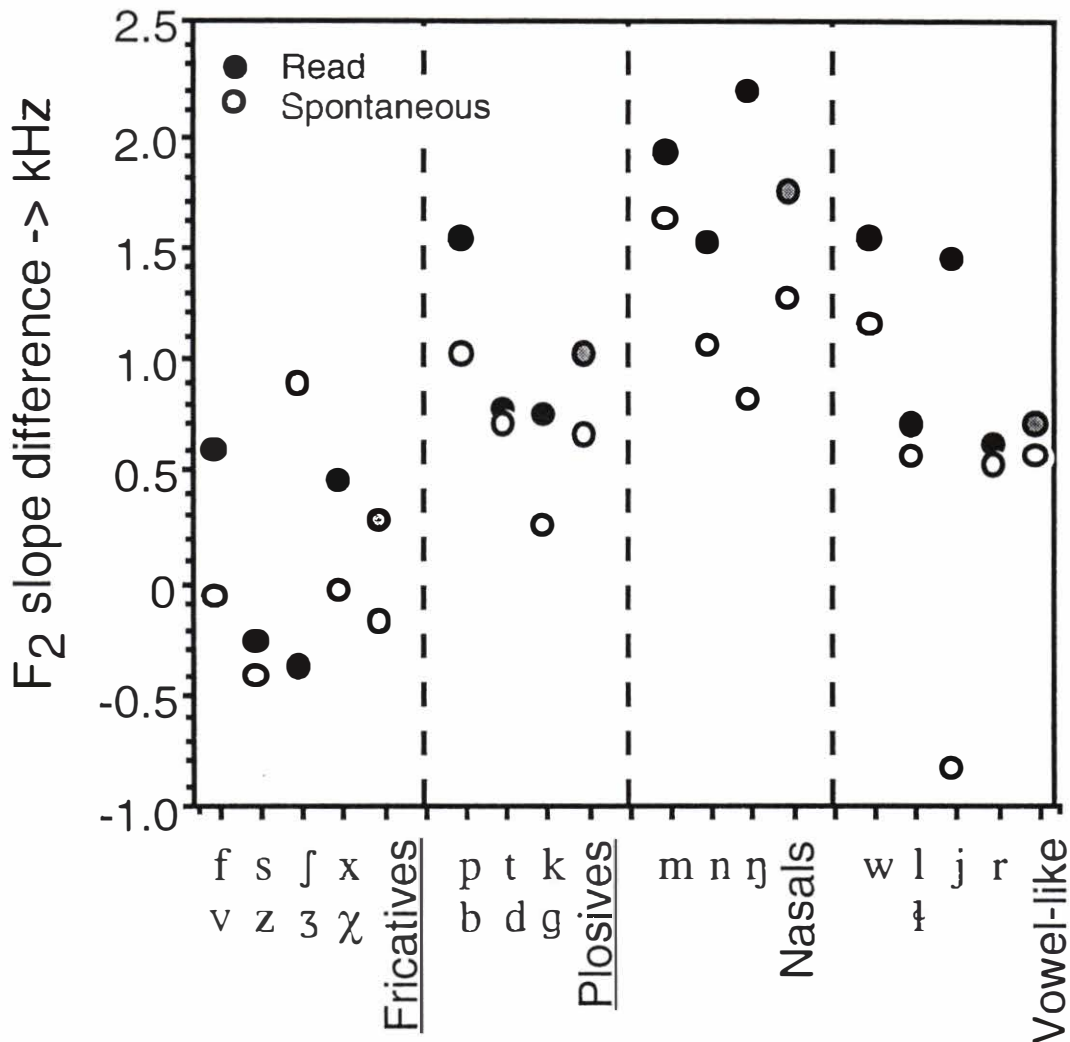


Figure 2. The differences between the slope of the F<sub>2</sub> formant at the consonant boundaries. Underlined symbols indicate statistical significance ( $p \leq 0.01$ , two tailed Sign test). Grey circles: pooled values.

The formant transitions in the vowel off- and onset bordering a consonant, especially of the F<sub>2</sub>, are both sensitive to coarticulation and are important cues for consonant identification (e.g., Clark and Yallop, 1990; O'Shaughnessy, 1987). To quantify the extent of acoustic coarticulation we determined the difference between the F<sub>2</sub> slopes at the CV- and the VC-boundaries (i.e., the F<sub>2</sub> slope difference). We used formant track slopes normalized for vowel duration because formant track shapes are largely invariant with speaking rate (Pols and Van Son, 1993) and because in perception one also normalizes for speaking rate (Miller and Bear, 1983). The slopes were calculated from the coefficients of a 4<sup>th</sup> order polynomial fit of the F<sub>2</sub> tracks of the vowels with the duration normalized to 1.

For the fricatives and plosives, as well as for all consonants pooled (not shown), there is a statistically significant difference between speaking styles with respect to the values of the F<sub>2</sub> slope difference ( $p \leq 0.001$ , two tailed Sign test). The F<sub>2</sub> slope difference between CV- and VC-boundaries is generally lower in spontaneous speech than in read speech. The behaviour of individual phonemes is very erratic (Figure 2, none reaches statistical significance).

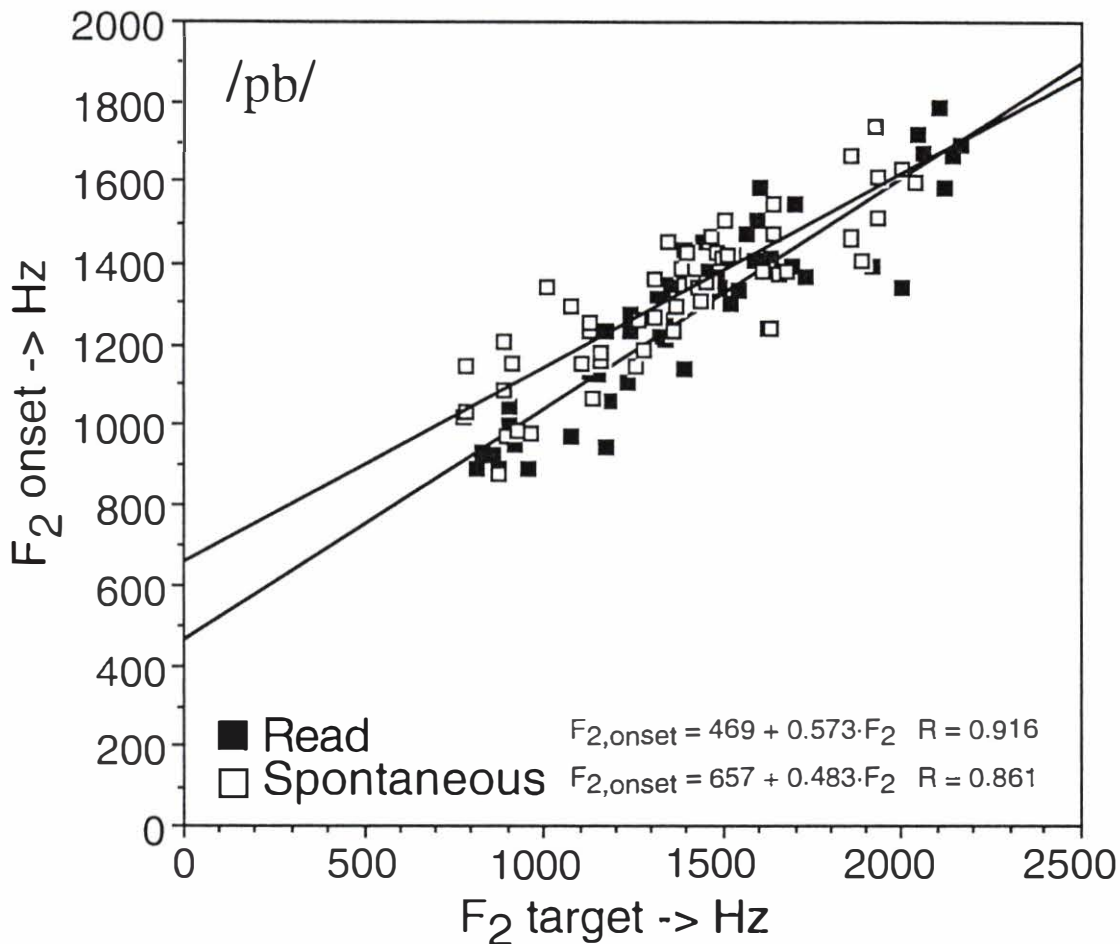


Figure 3a. F<sub>2</sub> onset and F<sub>2</sub> "target" frequencies of vowels preceded by /p/ or /b/. The equations of the regression lines through the data points are the F<sub>2</sub> Locus Equations. The equations and correlation coefficients (R) of the regression lines are given in the lower parts of the graphs ( $p \leq 0.001$ ).

### 3.2. F<sub>2</sub> Locus Equations

Another way to look at the "formant" values of consonants is to use the correlation between the F<sub>2</sub> frequency at the vowel onset and the vowel "target", the so called F<sub>2</sub> Locus Equations (Sussman et al., 1991, 1993, 1995). The vowel target is defined here as the position in the vowel realization where the formant frequencies reach their most extreme values. Examples of these Locus Equations and the way they are calculated are given for both the read and spontaneous realizations of the labial plosives (/p/ and /b/, Figure 3.a) and nasals (/m/, Figure 3.b).

The correlation between F<sub>2</sub>-onset and F<sub>2</sub>-target frequencies of the *following* vowel is generally quite strong and consonant specific. For the consonants /kgxytdszpbfvmw/ the correlation explains more than half the variance for both speaking rates and stressed and unstressed realizations alike (i.e.,  $R \geq 0.72$ , ignoring voicing contrasts, e.g., /p/ and /b/ are combined). For half the consonants examined, the correlation explains more than 75% of the variance (i.e.,  $R \geq 0.87$ ). Furthermore, there is a robust (cor-)relation between the Y-axis intercept value and the slope of the Locus equations which is linked to place of articulation (Sussman et al., 1995).



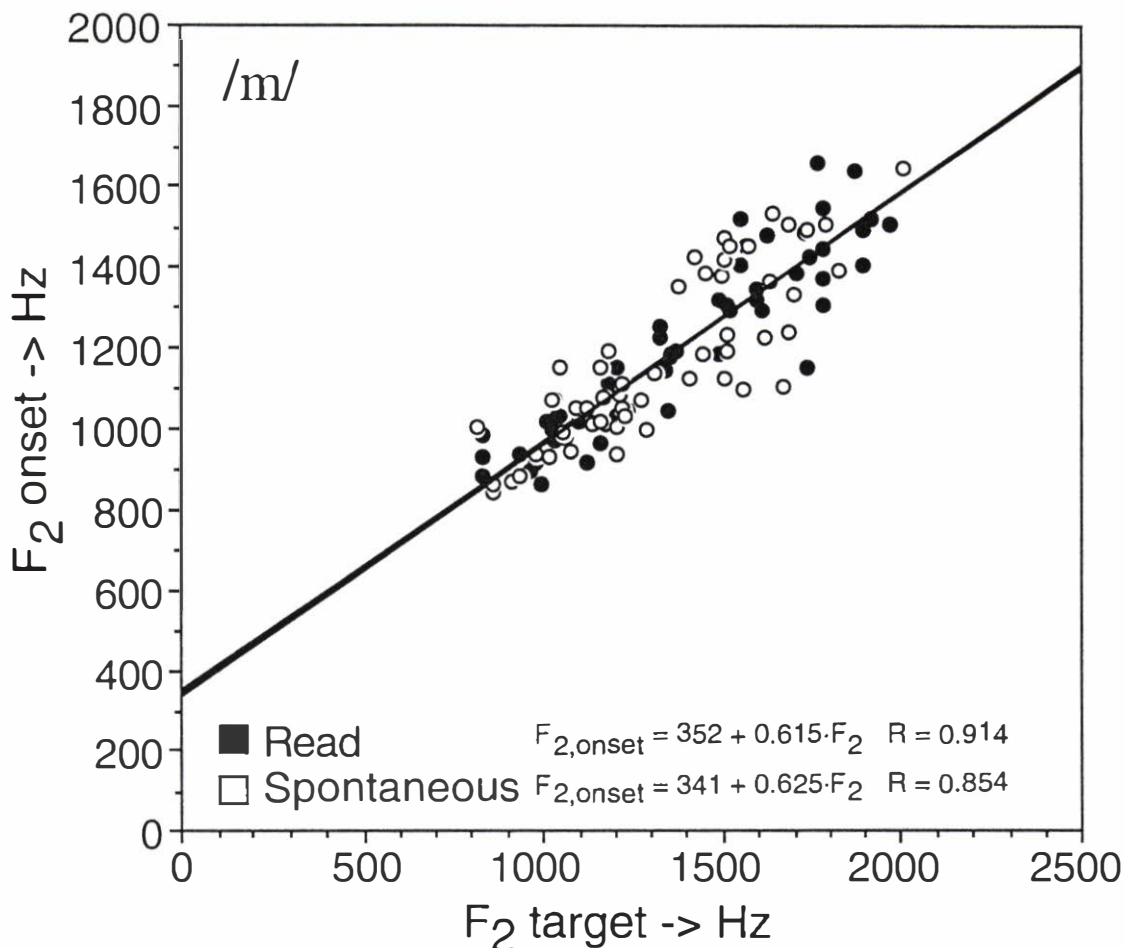


Figure 3b. F<sub>2</sub> onset and F<sub>2</sub> "target" frequencies of vowels preceded by /m/. The equations of the regression lines through the data points are the F<sub>2</sub> Locus Equations. The equations and correlations coefficients (R) of the regression lines are given in the lower parts of the graphs ( $p \leq 0.001$ ).

For our data, we plotted the values of the Y-intercept and the slopes of the Locus Equations in Figure 4, split on speaking style and vowel stress. We selected only realizations that were not word-final and, furthermore, only those consonants and conditions for which the correlation coefficient between F<sub>2</sub>-onset and -target, i.e., the Locus Equations, had a level of statistical significance of  $p \leq 0.05$  (two-tailed). This gives us values that are reasonable stable. We must stress that we do not consider this a convincing level of significance to decide whether these correlations are genuine. The total number of coefficients calculated is too large for that. We only consider correlations convincing when a level of significance of  $p \leq 0.001$  (two-tailed) is reached (the four points with  $p > 0.001$  are indicated in the legend of Figure 4). It can be seen in Figure 4 that the correlation between the Y-intercept and the slope of the Locus equations is very strong indeed,  $R = 0.909$ , i.e., it explains 82% of the variance. The consonants seem to cluster according to place of articulation as was found by Sussman et al. (1995).

It can be seen in Figure 4 that there is no obvious systematic difference with respect to the Locus equations between stressed and unstressed realizations or between read and spontaneous realizations. For the /td/, /pb/, /fv/, and /xɣ/ there were significant differences in correlation strength between stressed and unstressed or read and spontaneous realizations ( $p \leq 0.001$ ). However, although the correlation strength

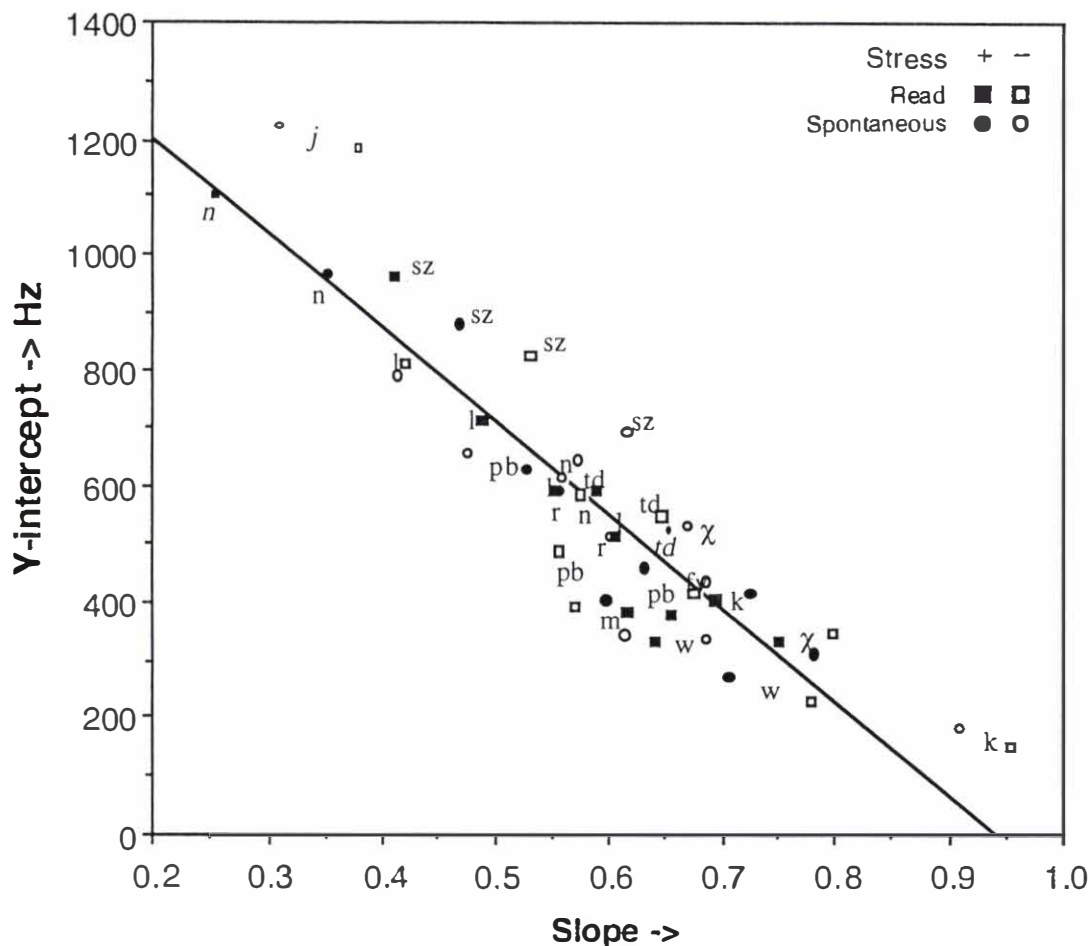


Figure 4. Y-intercept frequency versus slope of the F<sub>2</sub> Locus Equations. Data for read and spontaneous speech (squares and circles) and stressed and unstressed vowels (filled and open symbols) are displayed separately. The linear regression line between Y-intercept and the slope of the F<sub>2</sub> Locus Equations is drawn ( $R = 0.909$ ,  $p \leq 0.001$ , two-tailed). Only those consonants are shown for which the correlation between F<sub>2</sub>-onset and F<sub>2</sub>-target frequencies have a level of significance of  $p \leq 0.05$  (two-tailed, see text). The correlations between the F<sub>2</sub>-onset and F<sub>2</sub>-target frequencies are statistically significant ( $p \leq 0.001$ , two-tailed) for all but four points (both values of the /j/, the Read, Stressed realizations of the /n/, and the Spontaneous, Stressed realizations of the /td/, indicated by italic text).

was larger for stressed than for unstressed realization within the same speaking style, the correlation strength for read speech could be larger or smaller than that for spontaneous speech, depending on stress. All in all, consonant reduction seems to have no systematic effects on the F<sub>2</sub> Locus Equations.

We already showed that spontaneously uttered vowels were markedly reduced with respect to read vowels (Figure 1). Stress too had a strong effect (not shown). The fact that the strong correlations between the vowel F<sub>2</sub>-onset and -target frequencies is preserved irrespective of stress and speaking style indicates that consonant reduction is strongly linked to vowel reduction. As the F<sub>2</sub> "target" frequencies of the vowels change due to differences in stress or speaking style, the Locus Equations show that the F<sub>2</sub> onset frequencies change too. Formant frequencies are markers for articulatory movements, so we can assume that articulatory vowel reduction is accompanied by correlated changes in the (preceding) consonants.

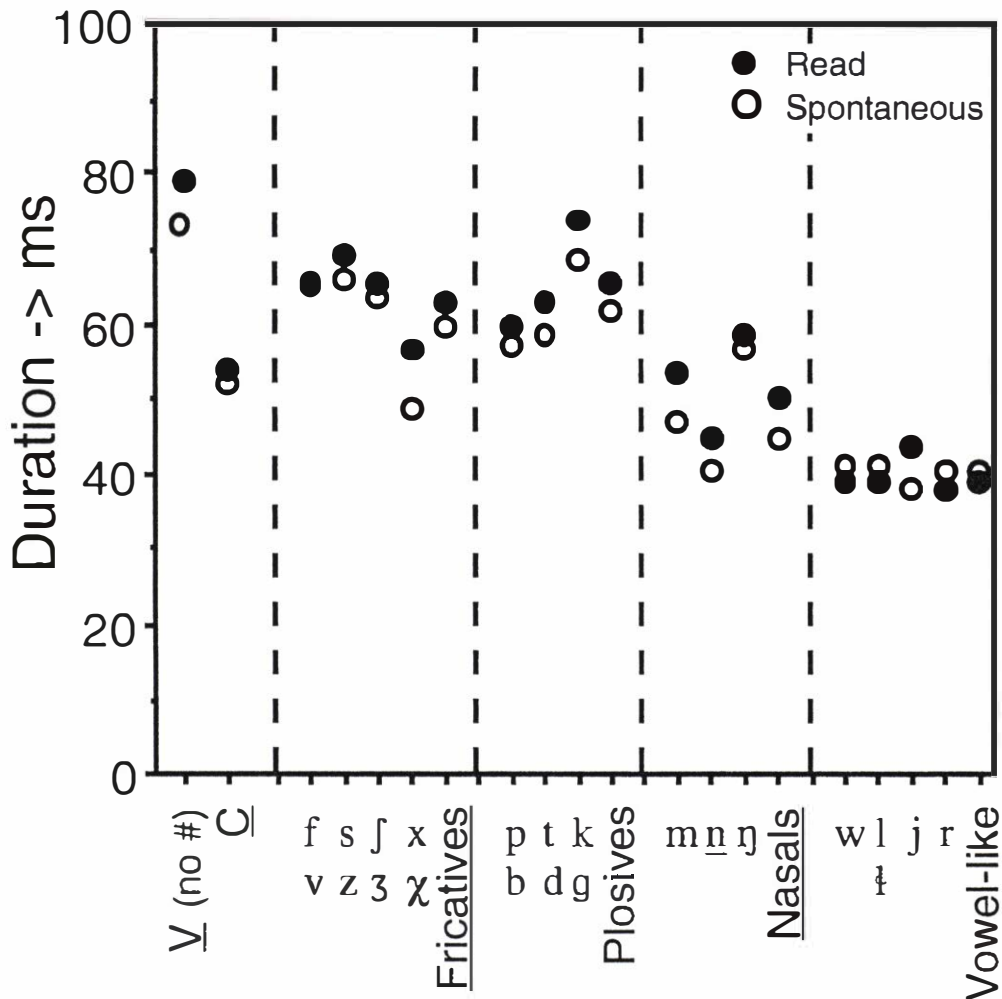


Figure 5. Durational reduction in Dutch vowels and consonants (V (no #): all vowels pooled excluding pauses; C: all consonants pooled). Underlined symbols indicate statistical significance ( $p \leq 0.001$ , two-tailed Sign test). Grey circles: pooled values.

### 3.3. Duration

Duration is one of the strongest correlates of *vowel* reduction (Van Bergem, 1995; Van Son, 1993). As is to be expected, there is a decrease in vowel duration in the spontaneous members of each pair (Figure 5, pooled values, see also Koopmans-van Beinum, 1992). The *consonant* realizations too are shorter in spontaneous speech (Figure 5, C, pooled values). This holds for all individual consonantal categories (not all differences are statistically significant, see Figure 5), except for the vowel-like consonants where duration seems to remain constant or to increase slightly (not significant).

Both vowels and consonants become shorter when spoken spontaneously. Furthermore, they become shorter by the same amount. The relative duration of consonants in the VCV segments, i.e., as a fraction of the total duration, does not change with speaking style (not shown).



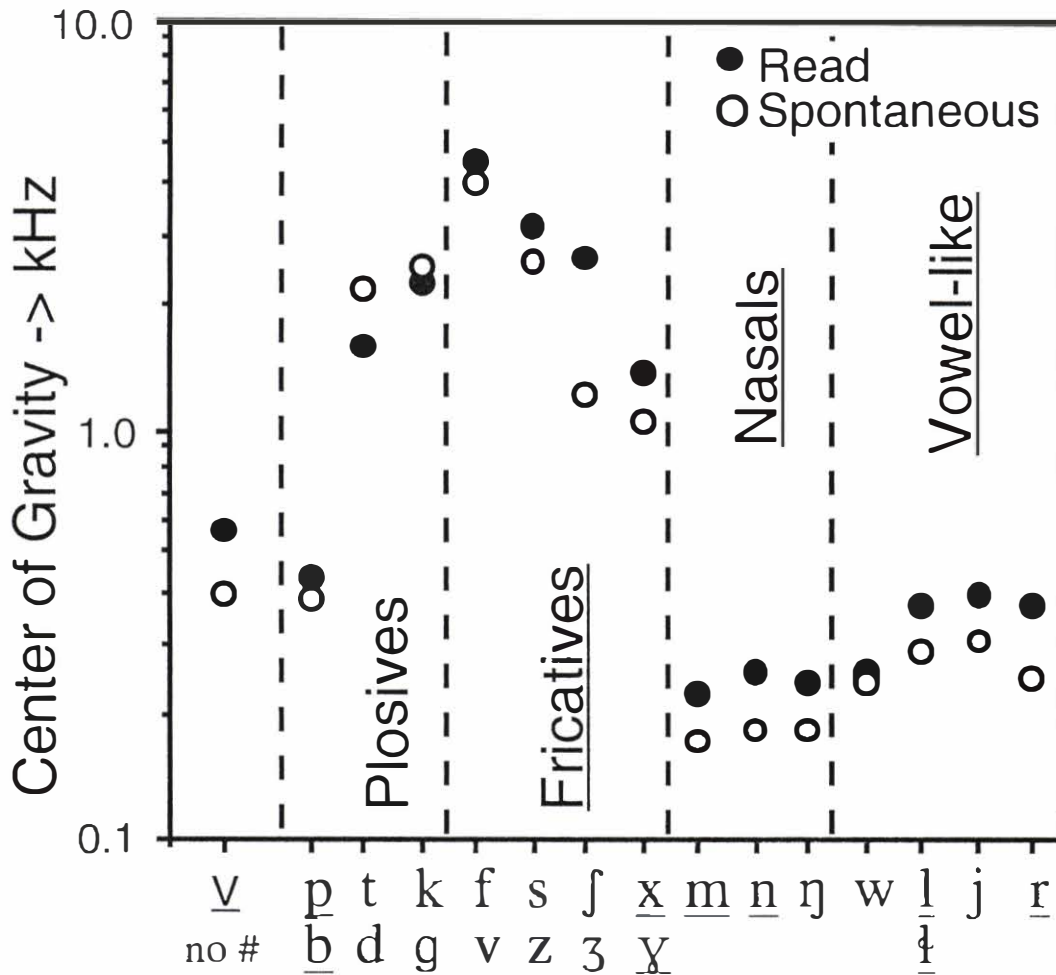


Figure 6. Reduction of the Center of Gravity for Dutch vowels and consonants. V(no #): all vowels pooled excluding pauses. See Figure 5 for details, underlined category names indicate statistically significant differences for pooled values (not shown).

### 3.4. Center of Gravity

The center of gravity of a spectrum (COG) is in a sense, the "mean" frequency. It is calculated by dividing  $\int f \cdot E(f) \cdot df$  by  $\int E(f) \cdot df$ . For sonorants, the COG is related to the spectral slope, the steeper the slope, the lower the COG. The steepness of the spectral slope, in its turn, is determined by the steepness of the glottal pulse which is a measure of speech effort. For turbulent noise, the COG is determined by the size of the quotient of (air flow speed) / (constriction area) which again is determined by speech effort.

For Dutch (and English), a more level spectral slope, i.e., a higher COG, strongly correlates with perceived sentence accent (Sluijter, 1995a, 1995b; Sluijter en Van Heuven, 1996). As the de-accentuation of vowels strongly correlates with vowel reduction, we can predict that reduction will show up as a lower COG. In Figure 6 this prediction bears out for the vowel realizations. For each vowel, spontaneous realizations have a lower COG than the read realizations (only shown for pooled data). For the sonorants and fricatives we see a similar picture (a lower COG for spontaneous realizations). For the release bursts of the plosives we see an erratic

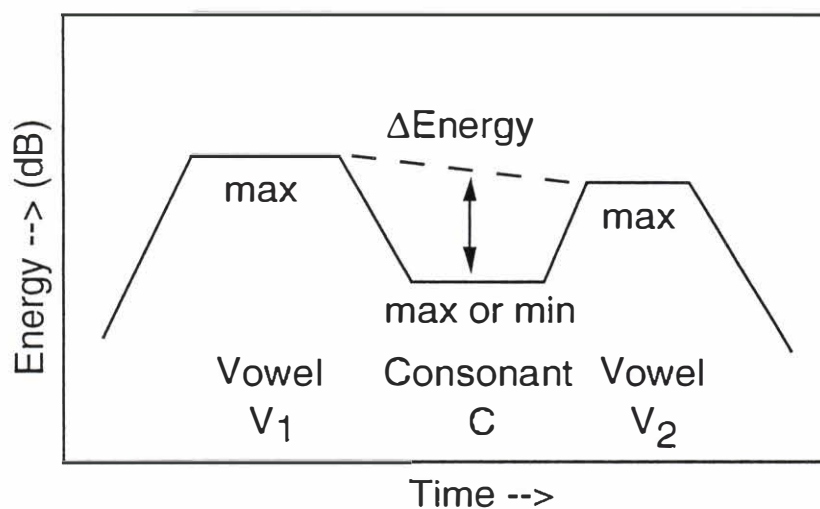


Figure 7. Definition of the intervocalic sound energy difference.  
 $V_{\max} = (V_{1,\max} + V_{2,\max})/2$ . For plosives and fricatives:  
 $\Delta E = V_{\max} - C_{\max}$ , and for nasals and vowel-like consonants:  
 $\Delta E = V_{\max} - C_{\min}$ .

behaviour that does not seem to indicate a definite difference in the COG with respect to speaking style.

A subdivision of the phonemes in categories can be seen in Figure 6. Very high absolute COG frequencies are found for most obstruents (plosives and fricatives). For fricatives, the COG frequency is inversely related to the size of the cavity in front of the noise source. For plosives the pattern is more intricate. The COG frequencies for /tdk/ from spontaneous speech are indistinguishable or higher than those from read speech (statistically not significant). The rather low COG frequencies for /pb/ (similar to that for vowels) show the influence of the open oral cavity behind the sound source. The overall distribution of COG values of obstruents is strongly bimodal due to the presence of approximants (not shown).

Quite low COG frequencies are found for sonorants (vowels and consonants) with vowels having higher values than nasals and vowel-like consonants. For the latter, the COG is dominated by the damping of the higher frequencies due to their closed articulation.

### 3.5. Intervocalic sound energy difference

One of the most salient differences between vowels and consonants is in their respective sound energy level. Vowels generally have a much higher sound energy level than consonants. Vowel reduction decreases the maximal sound energy level of vowels. Whether the energy level of consonants changes by the same amount can be determined by measuring the sound energy, or the relative energy, of consonants with respect to their flanking vowels. The sound energy difference is measured as indicated in Figure 7.

Figure 8 displays the sound energy differences for read and spontaneous speech. For all consonants, except for the nasals, the intervocalic sound energy difference is smaller in spontaneous speech. Altogether, the effects of speaking style on the intervocalic sound energy differences seem to be small, on the order of 1 dB. Therefore, changes in the sound level of the vowels seem to be largely matched by corresponding changes in the intervocalic consonants.

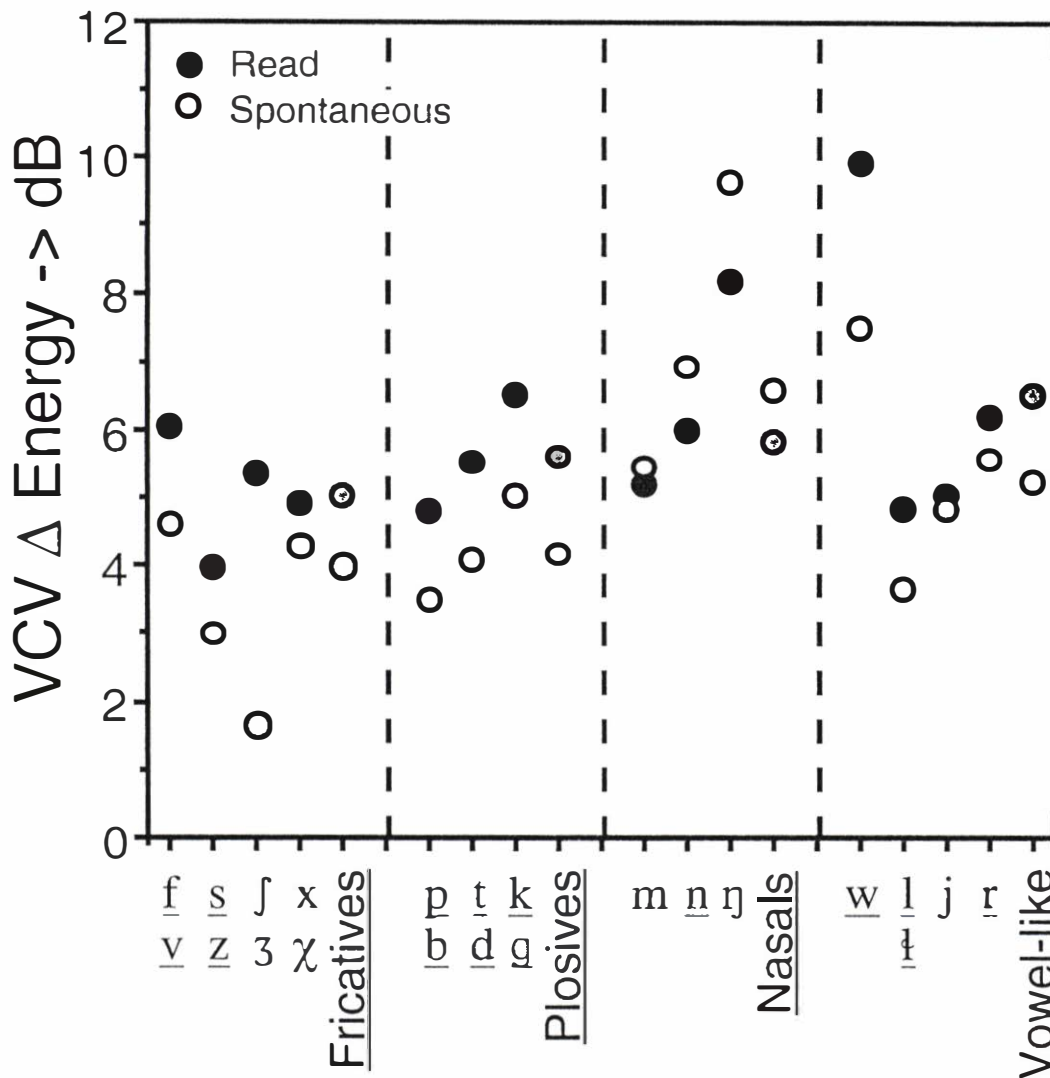


Figure 8. Reduction of intervocalic sound energy difference. See Figure 5 for details.

#### 4. Discussion

Four correlates of reduction have been studied for consonants with respect to speaking style: 1) F<sub>2</sub> slope differences and Locus Equations, 2) Duration, 3) Center of Gravity, and 4) Intervocalic sound energy difference.

In spontaneous speech, the nasal consonants “weaken” somewhat more than the neighbouring vowels whereas other consonants “weaken” somewhat less than the vowels (Figure 8).

The generally lower F<sub>2</sub> slope differences in spontaneous speech indicate a decrease of coarticulation strength. This is equivalent to the spectral effect of articulatory reduction found in vowel space. We also found strong, consonant specific, correlations between the F<sub>2</sub>-onset and -target frequencies of vowel realizations following these consonants. These robust correlations, called F<sub>2</sub> Locus Equations (Sussman et al., 1991, 1993, 1995), indicate that formant changes in consonants largely mirror those in vowels. The lack of any (systematic) differences between the F<sub>2</sub> Locus Equations with respect to differences in stress or speaking style strongly

suggests that the consonants also mirror any change in vowel F<sub>2</sub> frequencies that is related to vowel reduction.

In spontaneous speech, consonant realizations shorten like vowels. The decrease in duration of consonants is such that the relative duration, as a fraction of total VCV segment duration, remains unchanged (not shown). Therefore, the change in duration seems to be a "global" feature of a change in speaking style.

Except for the plosives, all consonants and vowels showed a decrease in COG in the spontaneous speech condition. This indicates that both the vowels and the non-plosive consonants show a diminishing source strength in spontaneous speech. This in return, implies a decrease in vocal and articulatory effort. As the COG is strongly linked to the spectral slope at high frequencies, this lowering might be expected to correlate with a decrease in the perceived stress of the vowels and, if consonants contribute to stress perception, the consonants (Sluijter, 1995a, 1995b; Sluijter en Van Heuven, 1996).

## 6. Conclusions

When spoken in a more informal style, consonant realizations show reduction in terms of diminishing articulatory precision and global effort. Furthermore, consonant reduction resembles vowel reduction in both type and extent of the changes in the produced sounds. Details of the changes in spectral and sound energy level of consonants due to speaking style differences depend on the type of phoneme.

## 7. Acknowledgements

The authors want to thank Florien Koopmans-van Beinum for supplying the speech recordings and Noortje Blauw for her transliteration of the spontaneous speech. This research was made possible by grant 300-173-029 of the Dutch Organization of Research (NWO).

## 8. References

- Byrd, D. (1994) "Relations of sex and dialect to reduction", *Speech Communication* 15, 39-54.
- Boersma, P. "Praat, a system for doing phonetics by computer" available at URL <http://fonsg3.let.uva.nl/>.
- Clark, J. and Yallop, C (1990). *An introduction to phonetics and phonology* (Basil Blackwell, Oxford, UK), 116-151, 1990, 1995<sup>2</sup>.
- Duez, D. (1995). "On spontaneous French speech: aspects of the reduction and contextual assimilation of voiced stops", *Journal of Phonetics* 23, 407-427.
- Farnetani, E. (1995). "The spatial and the temporal dimensions of consonant reduction in conversational Italian", *Proceedings of Eurospeech 95*, Madrid, 2255-2258.
- Keating, P.A., Lindblom, B., Lubker, J., and Kreiman, J. (1994). "Variability in jaw height for segments in English and Swedish VCVs", *Journal of Phonetics* 22, 407-422.
- Koopmans-Van Beinum, F.J. (1980). *Vowel contrast reduction, an acoustic and perceptual study of Dutch vowels in various speech conditions*, Ph.D. thesis of the University of Amsterdam, 163 p.
- Koopmans-Van Beinum, F.J. (1992). "The role of focus words in natural and in synthetic continuous speech: Acoustic aspects", *Speech Communication* 11, 439-452.
- Miller, J.L. and Baer, T. (1983). "Some effects of speaking rate on the production of /b/ and /w/", *Journal of the Acoustical Society of America* 73, 1751-1755.
- O'Shaughnessy, D. (1987). *Speech Communication* (Addison-Wesley, Reading, MA).
- Pols, L.C.W. and Van Son, R.J.J.H. (1993). "Acoustics and perception of dynamic vowel segments", *Speech Communication* 13, 135-147.

- Schmidt, A.M. and Flege, J.E. (1995). "Effects of speaking rate changes on native and non-native speech production", *Phonetica* 52, 41-54.
- Sluijter, A.M.C. (1995a). "Intensity and vocal effort as cues in the perception of stress", *Proceedings of Eurospeech 95*, Madrid, 941-944.
- Sluijter, A.M.C. (1995b). *Phonetic correlates of stress and accent* HIL dissertations 15, Ph.D. Thesis, University of Leiden.
- Sluijter, A.M.C. and Van Heuven, V.J. (1996). "Spectral balance as an acoustic correlate of linguistic stress", *Journal of the Acoustical Society of America* 100, 2471-2485.
- Sussman, H.M., McCaffrey, H.A.L., and Matthews, S.A. (1991). "An investigation of locus equations as a source of relational invariance for stop place categorization", *Journal of the Acoustical Society of America* 90, 1309-1325.
- Sussman, H.M., Hoemeke, K.A., and Ahmed, F. S. (1993). "A cross-linguistic investigation of locus equations as a phonetic descriptor of articulation", *Journal of the Acoustical Society of America* 94 1256-1268.
- Sussman, H.M., Fruchter, D., and Cable, A. (1995). "Locus equations derived from compensatory articulation", *Journal of the Acoustical Society of America* 97, 13112-3124.
- Van Bergem, D. (1995). *Acoustic and lexical vowel reduction*, in *Studies in Language and Language Use* 16. Ph.D. Thesis, University of Amsterdam.
- Van Son, R.J.J.H. (1993). *Spectro-temporal features of vowel segments*, in *Studies in Language and Language Use* 3. Ph.D. Thesis, University of Amsterdam.
- Van Son, R.J.J.H., and Pols, L.C.W. (1996). "An acoustic profile of consonant reduction", *Proceedings ICSLP 96*, Philadelphia, USA, 1529-1532.