

ACOUSTIC CONSONANT REDUCTION: A COMPARISON*

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 unstressed syllables. How consonants behave in situations where vowels are reduced is not known. In this paper we compare durational and spectral data for both intervocalic consonants and vowels, uttered in 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 manner of articulation of the consonants.

1. Introduction

Vowel reduction is a well established phenomenon that has found its place in phonetics textbooks (e.g., O'Shaughnessy, 1987; Clark and Yallop, 1990). 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. 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 (cf. Van Bergem, 1995).

Given such a universal phenomenon, it is quite natural to ask how *consonants* behave in situations where vowels become reduced. In other words, does something like "consonant reduction" really exist, and if it does, how can it be described? It must be stressed that there are reasons to expect that consonants do not behave exactly like vowels. Because of their lower acoustic energy and their often intricate structure, consonants are often more difficult to perceive and identify than vowels (however, see Steeneken, 1992). Therefore, the penalty for reduction, in terms of reduced intelligibility, could be much higher for consonants than for vowels. There have been studies that investigated acoustic and articulatory consonant reduction in relation to vowel reduction, but these were generally limited to only a few classes of consonants, with only limited speech material (some recent papers are, e.g., Byrd, 1994; Keating et al., 1994; Duez, 1995; Farnetani, 1995; Schmidt and Flege, 1995). From these studies it is difficult to discern the general effects of consonant reduction in "normal" speech situations.

To study whether and how consonants reduce, 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 were read aloud from text. When there exists a corresponding phenomenon that can be called

* A preliminary version of this paper was presented at Eurospeech 95, Madrid (Van Son and Pols, 1995).

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

	Back	Mid	Alve	LabD
Plos	k g		t d	p b
Fric	x ɣ	ʒ ʃ	s z	f v
Nasal	ŋ		n	m
V-like			l ʎ	

consonant reduction, this too can be expected to show itself when informal speech is compared with read speech.

Consonant reduction can happen in several ways. At the broadest level, it would manifest itself as a loss of distinction in the *manner* of articulation and in the *place* of articulation. The former would result in blurring the borders between, for example, vowel-like consonants and vowels, or fricatives and plosives. The latter could result in, for example, palatalization of fricatives and plosives (Byrd, 1994), or a lack of distinction between alveolar and labio-dental consonants due to incomplete or inappropriate closure. As the perception of place and manner of articulation for consonants also depends on the transition between vowels and consonants, consonants might additionally be perceived differently due to changes in the neighbouring vowel segments alone. 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 where reduction will affect the phonemic distinction of consonants. We will limit ourselves in this paper to an inventarization of aspects 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 characterize 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 limit of consonant reduction, i.e., complete deletion, where these aspects are undefined.

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 the transcription. This way,

Table 2. Number of matched VCV pairs, separated on consonant identity.

	Back	Mid	Alve	LabD	Total
Plos	62		65	61	188
Fric	77	3	62	75	217
Nasal	14		72	63	149
V-like			44		44
Total	153	3	243	199	598

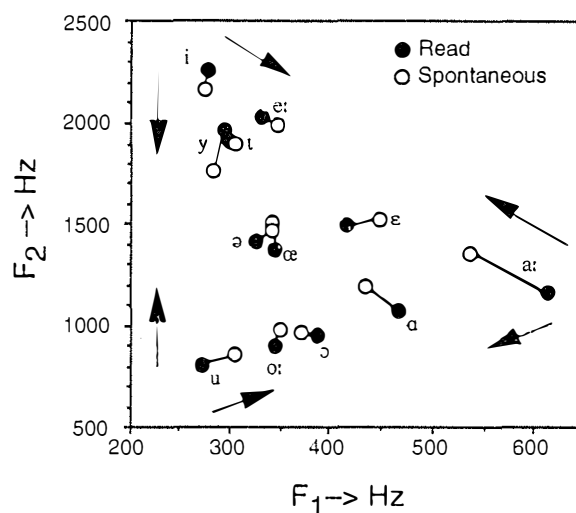


Figure 1. Spectral reduction in Dutch vowel space. In the vowel plane, reduced realizations occupy a more central position. Illustrated here by the difference between vowels from Spontaneous and Read speech. Arrows indicate direction of change under reduction.

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 Phonetics at the University of Nijmegen. One of the authors checked the transcription and marked words for sentence accent by listening.

From the phonetic transcription, all Vowel-Consonant-Vowel (VCV) segments were located in the speech recordings (also those crossing word boundaries). 1847 VCV realization pairs had both realizations originating from corresponding positions in the utterances with identical syllable structure, syllable boundary type, and sentence and word stress. Of these VCV-pairs, 598 have been analyzed in detail for this paper (see table 1 and 2) and will be used here to study consonant reduction in more detail.

Phoneme boundaries were placed using a waveform display with audio feedback (Boersma, 1996) combined with linked 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, as much as possible, 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). Formant analysis was done using LPC-based formant extraction algorithm.

3. Results

3.1. Formant values

Vowel reduction is characterized by a centralization of the distribution of vowel realizations 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 the independent analysis of the same speech by Koopmans-Van Beinum, 1992). Unfortunately, there is no corresponding compact consonant formant space where reduction could show up. However, the formant transitions in the vowel onset are important for consonant identification in CV sequences. There is a regularity

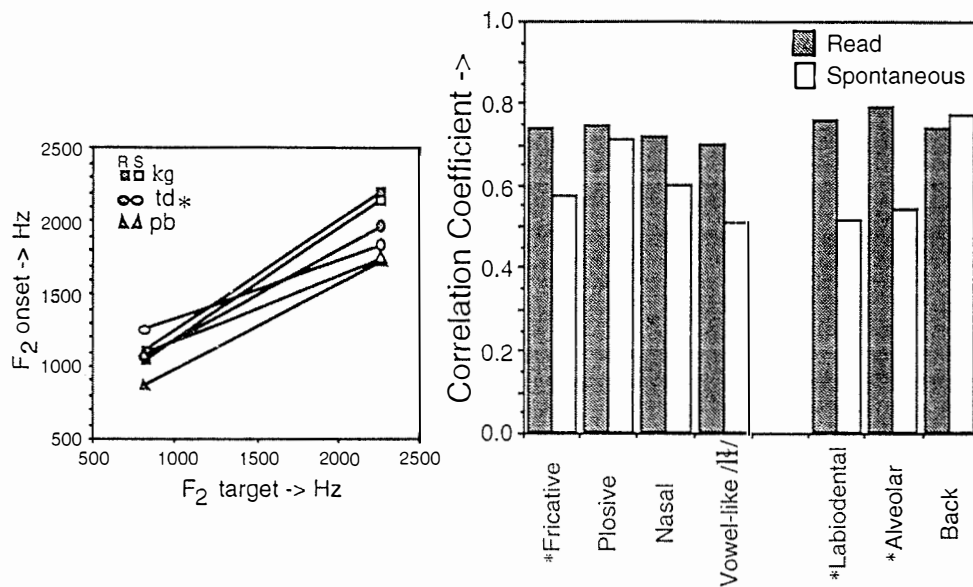


Figure 2. Left panel: examples of plots corresponding to F₂-locus equations for plosives. For each CV-transition, the F₂-onset frequency is plotted against the F₂-target frequency. For each individual consonant, the equation of the linear regression line through these points is the locus equation. Right panel: reduction in F₂-locus equations of groups of Dutch consonants. The correlation coefficients of F₂-locus equations indicate the "consistency" of CV (co-)articulation. The * symbol indicates a statistically significant difference between correlation coefficients in read versus spontaneous speech, $p \leq 0.01$, two tailed test of Fisher Z-transforms of the correlation coefficients

between the F₂ frequency at the start of the vowel and the frequency inside the vowel kernel. If the F₂ frequencies at the start of the vowel are plotted against the frequencies in the vowel center, then, for each consonant, the points cluster along a straight line.

The linear regression line through these points represents the F₂-locus equation for that consonant (e.g., figure 2, left panel; Sussman et al., 1991, 1993, 1995). The correlation coefficients of the linear regression lines indicate the strength of the relation

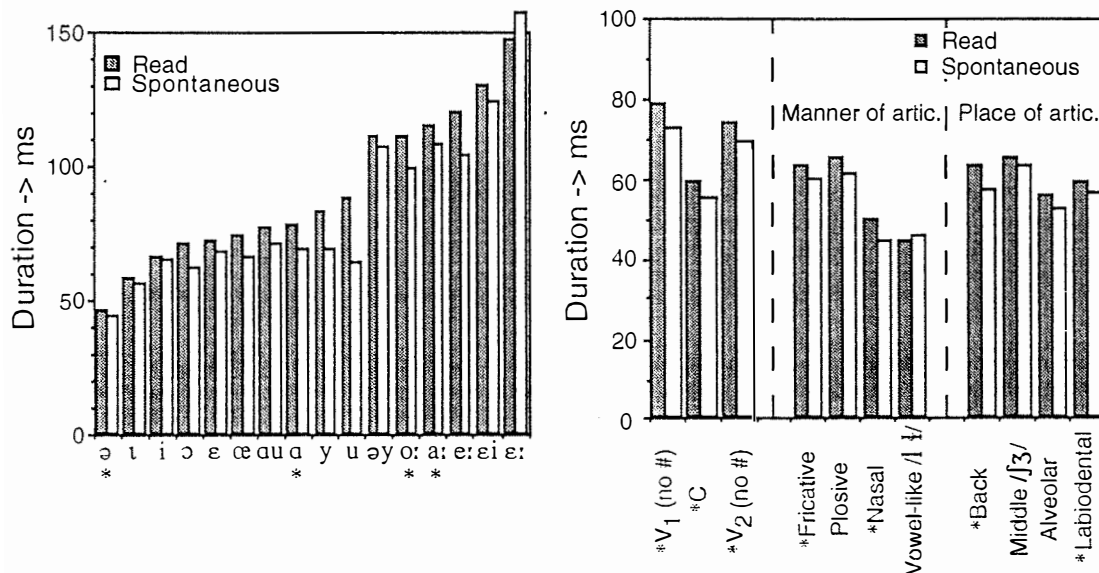


Figure 3. Durational reduction in Dutch vowels and consonants, left vowels (initial and final vowels in the VCV segments pooled), right consonants and mean values for vowels (V₁: initial; V₂: final; no #: excluding pauses). * $p \leq 0.01$, two tailed Sign test ($p \leq 0.001$ for all vowels pooled)

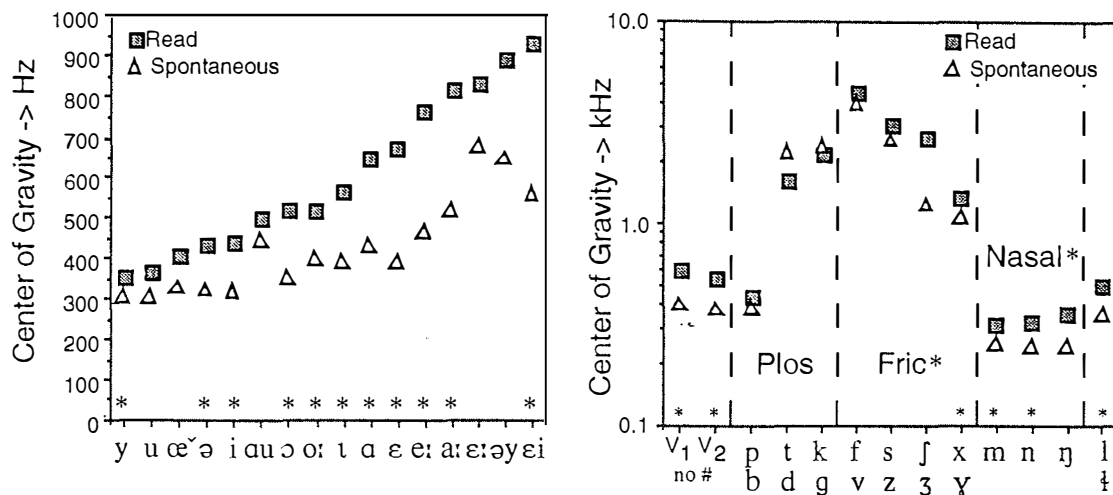


Figure 4. Reduction of the Center of Gravity for Dutch vowels and consonants, left vowels, right mean vowel values and consonants. Plotted are the averages of the maximal values within phoneme realizations. Note the logarithmic scale in the right hand plot. V₁: mean of all initial vowels, V₂: mean of all final vowels, no #: excluding pauses. * p ≤ 0.001, two tailed Sign test.

between onset and kernel values of the second formant and hence the consistency of articulation.

Usually, the F₂-locus equations are used only for plosives, but the regularities extend to other consonants as well (cf. the high correlation coefficients in the right panel of figure 2). No articulatory or perceptual correlates are known for these locus equations, but they do indicate a specific invariant distinctiveness in articulation. Therefore, the size of the correlation coefficient can be used to quantify the consistency of articulation of a certain consonant. The occurrence of consonant reduction should be visible as a less distinctive articulation resulting in lower correlation coefficients. In the right hand panel of figure 2 it can be seen that the correlation coefficients are indeed consistently lower in spontaneous speech for most consonants than in read speech, indicating reduction.

3.2. Duration

Duration is one of the strongest correlates of vowel reduction (e.g., Lindblom, 1963). As is to be expected, there is a very consistent decrease in vowel duration in the spontaneous members of each pair (figure 3, left hand panel for individual vowels, right hand panel for pooled values, see also Koopmans-Van Beinum, 1992). The odd one out is the /ε:/, which is extremely rare in Dutch. The consonant realizations too are shorter in spontaneous speech. This holds for all consonantal categories, except for the vowel-like laterals, /l ɫ/, where duration seems to remain constant.

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, does not change when speaking style changes (not shown).

3.3. Center of Gravity

The center of gravity of a spectrum 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 center of gravity is related to the

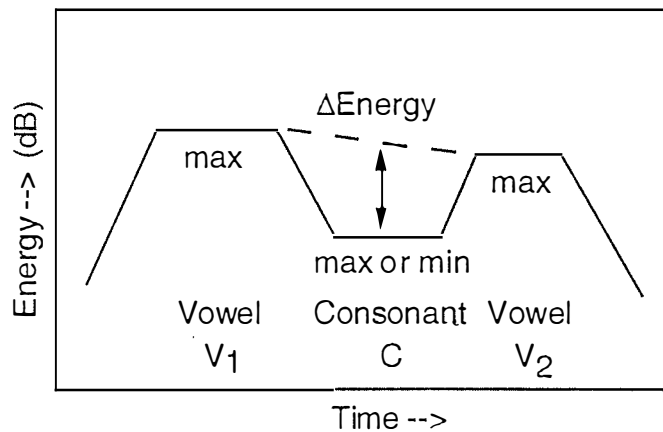


Figure 5. Definition of the intervocalic sound energy difference. $V_{max} = (V_{1,max} + V_{2,max})/2$; Plosives and fricatives: $V_{max} - C_{max}$, Nasals and vowel-like: $V_{max} - C_{min}$.

spectral slope, the steeper the slope, the lower the center of gravity. 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 center of gravity 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 center of gravity, strongly correlates with perceived sentence accent (Sluijter, 1995a,b). As the de-accentuation of vowels strongly correlates with vowel reduction, we can predict that reduction will show up as a lower center of gravity. In figure 4 this prediction bears out for the vowel realizations. For each and every vowel, spontaneous realizations have a lower center of gravity than the read realizations. For the sonorants and fricatives we see a similar picture (lower centers of gravity for spontaneous realizations). For the release bursts of the plosives we see an erratic behaviour that does not seem to indicate a definite difference in the center of gravity with respect to speaking style.

In figure 4, a subdivision of the phonemes can be seen. Very high absolute frequencies are found for the center of gravity of obstruents (plosives and fricatives) due to the strong high frequency component in the noise. For fricatives, the absolute height of the center of gravity is inversely related to the size of the cavity in front of the noise source. For plosives the pattern is more intricate. The frequencies for /tdkg/ from spontaneous speech are indistinguishable or higher than those from read speech (statistically not significant). The vowel-like frequencies for /pb/ show the mark of the open oral cavity behind the sound source.

Quite low frequencies are found for the sonorants (vowels and consonants) with vowels having higher values than consonants. For the consonant sonorants, the center of gravity is dominated by the damping of the higher frequencies due to their closed articulation.

3.4. Intervocalic sound energy difference

One of the most salient differences between vowels and consonants is in their respective sound energy. Vowels generally have much higher energy levels 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 5.

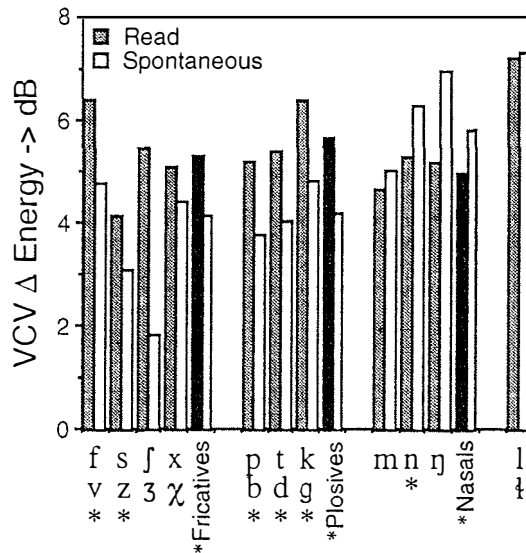


Figure 6. Reduction of intervocalic sound energy difference. * $p \leq 0.001$, two tailed Sign test.

Figure 6 displays the sound energy differences for read and spontaneous speech. For obstruents (plosives and fricatives) the difference is smaller in spontaneous speech than in read speech, for sonorant consonants (nasals and the /l ɫ/) it is larger. Altogether, the differences between the intervocalic sound energy differences in read and those in spontaneous speech seem to be small, of the order of 1 dB. Therefore, any change in the sound level of the vowels seems to be largely mirrored by the consonants. The differences between the behaviour of obstruents and sonorant consonants in this respect can, most likely, be traced back to the differences in the mechanism by which the speech sounds are generated. In sonorants, sound originates in the glottal pulses, in obstruents it originates in the air flow turbulence near an obstruction.

In section 3.3 (figure 4) it was shown that the spectral differences between read and spontaneous speech could be explained by a lower speech effort in spontaneous speech than in read speech. Such a reduced speech effort would not only result in a lower center of gravity, but also in a lower sound energy level. Figure 6 indicates that there is a relation between the center of gravity (i.e., the spectral slope) and the change in the sound energy: the lower the center of gravity (i.e., the steeper the spectral slope), the larger the difference in total sound energy between speaking styles. The size of the center of gravity of vowels lays in between that of nasals and that of fricatives (figure 4). The observed effect of a more informal speaking style on the intervocalic sound energy difference of nasals and fricatives implies that the absolute sound energy level of vowels changes by an amount in between of that of nasals and fricatives, i.e., less than nasals, more than fricatives (figure 6). In other words, a more informal speaking style results in an increase in the intervocalic sound energy difference for nasals (whose energy level reduces more than vowels) and a decrease for fricatives (that reduce less than vowels).

4. Discussion

Four correlates of reduction have been studied for consonants with respect to speaking style: 1) F_2 -locus equations, 2) Duration, 3) Center of Gravity, and 4) Intervocalic sound energy difference.

The generally lower correlation coefficients of the F_2 -locus equations in spontaneous speech indicate a loss of (articulatory) distinction. This is equivalent to the spectral effects of articulatory reduction found in vowel space.

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 in seems to be a "global" feature of the change in speaking style.

Except for the plosives, all consonants and vowels showed a decrease in center of gravity. 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 center of gravity is strongly linked to the spectral slope at high frequencies, this lowering can be expected to correlate with a decrease in the perceived stress (Sluijter, 1995a,b).

In spontaneous speech, the sonorant consonants (nasals and the /l ɫ/) "weaken" somewhat more than the neighbouring vowels whereas non-sonorants (fricatives and plosives) "weaken" somewhat less than the vowels (figure 6). Combined with the results for the spectral center of gravity this implies that the slope of the spectrum determines the size of the difference in sound energy that results from a difference in speaking style.

6. Conclusions

Uttered 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 spectral changes in consonants due to speaking style depend on the source of the speech sound: vocal folds or fricative noise.

In our future research we will extend coverage of the consonant used. Next to speaking style, we will also look at the effects of stress on reduction. The global measures of acoustic reduction presented here will be supplemented with more detailed types of analysis. Identification experiments to determine the effects of acoustic reduction on consonant identification are under way.

7. Acknowledgements

The authors want to thank Dr. Florien Koopmans-van Beinum for supplying the speech recordings and Dr. 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. (1996). '*Praat, a system for doing phonetics by computer*' available at our WWW site URL <http://fonsg3.let.uva.nl/>.
- Clark, J. & Yallop, C. (1990). *An introduction to phonetics and phonology* (Basil Blackwell, Oxford, UK), 1995², 116-151.
- 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. & 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. (1992). 'The role of focus words in natural and in synthetic continuous speech: Acoustic aspects'. *Speech Communication* 11, 439-452.
- O'Shaughnessy, D. (1987). *Speech Communication* (Addison-Wesley, Reading, MA).
- Pols, L.C.W. & Van Son, R.J.J.H. (1993). 'Acoustics and perception of dynamic vowel segments', *Speech Communication* 13, 135-147.
- Schmidt, A.M. & 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.
- Steeneken, H.J.M. (1992). '*On measuring and predicting speech intelligibility*', Ph.D. Thesis, University of Amsterdam.
- Sussman, H.M., McCaffrey, H.A.L. & 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. & 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. & Cable, A. (1995). 'Locus equations derived from compensatory articulation', *Journal of the Acoustical Society of America* 97, 3112-3124.
- 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. & Pols, L.C.W. (1995). 'What does consonant reduction look like, if it exists', *Proceedings of Eurospeech 95*, Madrid, 1909-1912.