AN ACOUSTIC PROFILE OF CONSONANT REDUCTION

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

Institute of Phonetic Sciences & IFOTT, University of Amsterdam
Herengracht 338, 1016 CG Amsterdam, The Netherlands
email: {rob, pols}@fon.let.uva.nl

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 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 [3, 9]. 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 [14].

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. [1, 4, 5, 6, 11]. 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 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.

	Velar	Pal	Alve	Lab
Plos	k ɡ		t d	p b
Fric	x ɣ	ʒ ʃ	s z	f v
Nasal	ŋ		n	m
V-like	r	j	l ɫ	w

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

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 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, 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. 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). 1847 VCV 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, 791 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.

	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

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

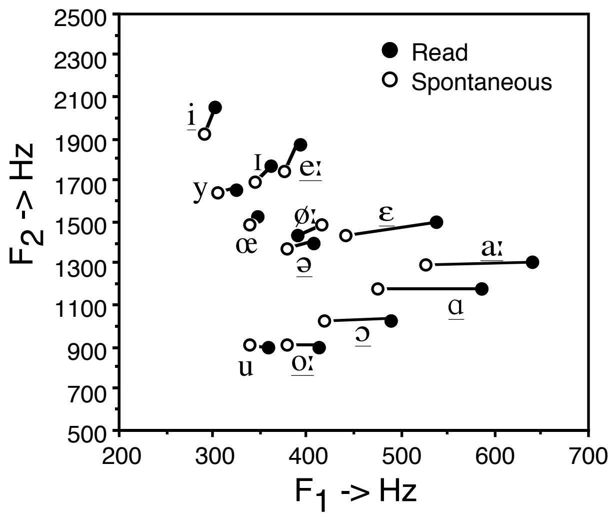

Figure 1: Spectral reduction in Dutch vowel space (pre-consonantal vowels). Underlined symbols:indicate statistical significance ($p \leq 0.001$, two tailed Sign test).

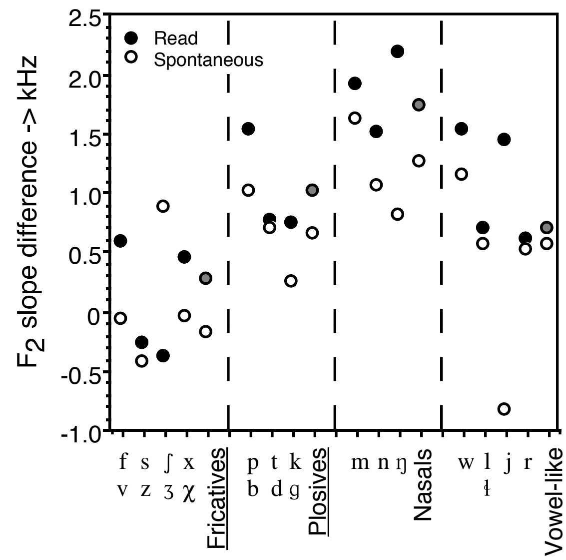

Figure 2. The differences between the slope of the F_2 formant at the consonant boundaries. Underlined symbols indicate statistical significance ($p \leq 0.01$, two tailed Sign test). Grey circles: pooled values.

Phoneme boundaries were placed using a waveform display with audio feedback [2] 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 were 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, [7]).

The formant transitions in the vowel off- and onset bordering a consonant, especially of the F_2, are both sensitive to coarticulation and are important cues for consonant identification [3, 9]. To quantify the extent of acoustic coarticulation we determined the difference between the F_2 slopes at the CV- and the VC-boundaries (i.e., the F_2 slope difference). We used formant track slopes normalized for vowel duration because formant track shapes are largely invariant with speaking rate [10] and because in perception one also normalizes for speaking rate [8]. The slopes were calculated from the coefficients of a 4^{th} order polynomial fit of the F_2 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 lower slope difference between speaking styles ($p \leq 0.001$, two tailed Sign test). The behaviour of individual phonemes is very erratic (figure 2, none reaches statistical significance).

3.2. Duration

Duration is one of the strongest correlates of *vowel* reduction [14, 15]. As is to be expected, there is a decrease in vowel duration in the spontaneous members of each pair (figure 3, pooled values, see also [7]). The *consonant* realizations too are shorter in spontaneous speech (figure 3, C, pooled values). This holds for all individual consonantal categories (not all statistically significant, see figure 3), 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, does not change when speaking style changes (not shown).

3.3. 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 determine by the size of the quotient of (air flow speed) / (constriction area) which again is determined by speech effort.

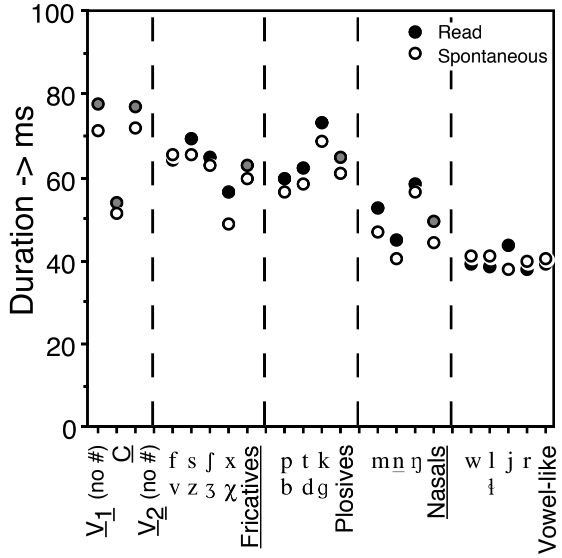

Figure 3. Durational reduction in Dutch vowels and consonants (V_1: initial; V_2: final; no #: excluding pauses). Underlined symbols indicate statistical significance ($p \leq 0.001$, two tailed Sign test). Grey circles: pooled values.

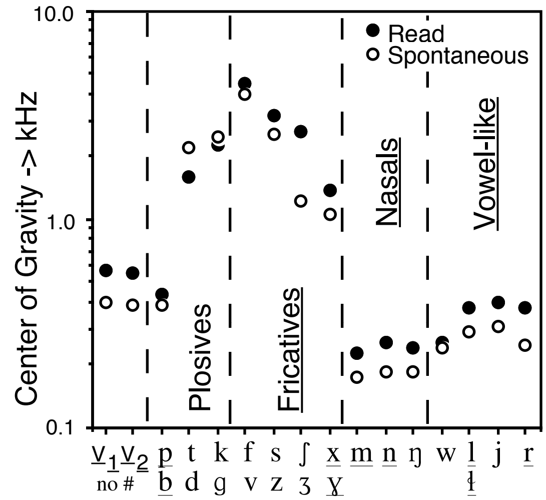

Figure 4. Reduction of the Center of Gravity for Dutch vowels and consonants. V_1 V_2: initial and final vowels, no #: excluding pauses. See figure 3 for details, underlined category names indicate pooled values (not shown).

For Dutch (and English), a more level spectral slope, i.e., a higher COG, strongly correlates with perceived sentence accent [12, 13]. 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 4 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 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 4. 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 /tdkg/ from spontaneous speech are indistinguishable or higher than those from read speech (statistically not significant). The vowel-like COG frequencies for /pb/ 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 aproximants (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.

3.4. 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 5.

Figure 6 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 changes 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.

4. DISCUSSION

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

The generally lower F_2 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.

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. 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 [12, 13].

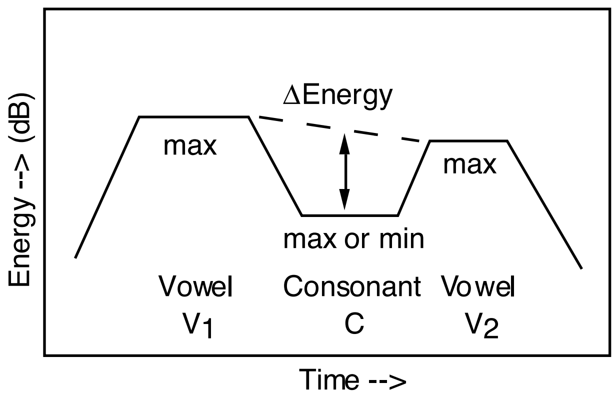

Figure 5. Definition of the intervocalic sound energy difference. Vmax = (V1,max+V2,max)/2.
For plosives and fricatives: $\Delta E = V_{max} - C_{max}$, and for nasals and vowel-like consonants: $\Delta E = V_{max} - C_{min}$.

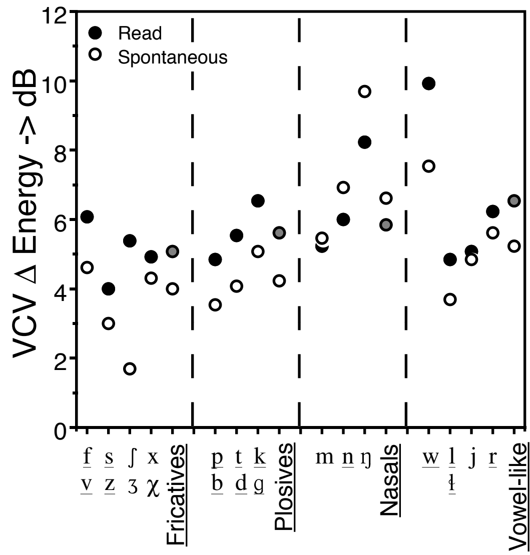

Figure 6. Reduction of intervocalic sound energy difference. See figure 3 for details

In spontaneous speech, the nasal consonants “weaken” somewhat more than the neighbouring vowels whereas other consonants “weaken” somewhat less than the vowels (figure 6).

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 spectral and sound energy level changes in consonants due to speaking style 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

1. Byrd, D. “Relations of sex and dialect to reduction”, *Speech Communication* **15**, 39-54, 1994.
2. Boersma, P. *“Praat, a system for doing phonetics by computer”* available at URL http://fonsg3.let.uva.nl/.
3. Clark, J. and Yallop, C. *An introduction to phonetics and phonology* (Basil Blackwell, Oxford, UK), 116-151, 1990, 1995[2].
4. Duez, D. “On spontaneous French speech: aspects of the reduction and contextual assimilation of voiced stops”, *Journal of Phonetics* **23**, 407-427, 1995.
5. Farnetani, E. “The spatial and the temporal dimensions of consonant reduction in conversational Italian”, *Proceedings of Eurospeech 95*, Madrid, 2255-2258, 1995.
6. Keating, P.A., Lindblom, B., Lubker, J. and Kreiman, J. “Variability in jaw height for segments in English and Swedish VCVs”, *Journal of Phonetics* **22**, 407-422, 1994.
7. Koopmans-Van Beinum, F.J. “The role of focus words in natural and in synthetic continuous speech: Acoustic aspects”, *Speech Communication* **11**, 439-452, 1992.
8. Miller, J.L. and Baer, T. “Some effects of speaking rate on the production of /b/ and /w/”, *Journal of the Acoustical Society of America* **73**, 1751-1755, 1983.
9. O’Shaughnessy, D. *Speech Communication* (Addison-Wesley, Reading, MA), 1987
10. Pols, L.C.W. and Van Son, R.J.J.H. “Acoustics and perception of dynamic vowel segments”, *Speech Communication* **13**, 135-147, 1993.
11. Schmidt, A.M. and Flege, J.E. “Effects of speaking rate changes on native and non-native speech production”, *Phonetica* **52**, 41-54, 1995.
12. Sluijter, A.M.C. “Intensity and vocal effort as cues in the perception of stress”, *Proceedings of Eurospeech 95*, Madrid, 941-944, 1995.
13. Sluijter, A.M.C. *“Phonetic correlates of stress and accent”*, *HIL dissertations **15***, Ph.D. Thesis, University of Leiden, 1995.
14. Van Bergem, D. *Acoustic and lexical vowel reduction*, in *Studies in Language and Language Use **16***. Ph.D. Thesis, University of Amsterdam, 1995.
15. Van Son, R.J.J.H. *Spectro-temporal features of vowel segments*, in *Studies in Language and Language Use **3***. Ph.D. Thesis, University of Amsterdam, 1993.