THE CORRELATION BETWEEN CONSONANT IDENTIFICATION AND THE AMOUNT OF ACOUSTIC CONSONANT REDUCTION

R.J.J.H. van Son & Louis C. W. Pols
Institute for Phonetic Sciences / IFOTT, University of Amsterdam, Herengracht 338,
NL-1016CG Amsterdam, The Netherlands, E-mail: {rob, pols}@fon.let.uva.nl

ABSTRACT

Reduction causes changes in the acoustics of consonant realizations that affect their identification. In this study we try to identify some of the acoustic parameters that are correlated with this change in identification. Speaking style is used to manipulate the degree of reduction. Pairs of otherwise identical intervocalic consonants from read and spontaneous utterances are presented to subjects in an identification experiment. The resulting identification scores are correlated to five different acoustical measures that are affected by the amount of consonant reduction: Segmental duration, spectral Center of Gravity, intervocalic sound energy difference, intervocalic F_2 slope difference, and the amount of vowel reduction in the syllable kernel. The identification differences between the read and spontaneous realizations are compared with the differences in each of the acoustic measures. It showed that only segmental duration and the spectral Center of Gravity are significantly correlated to identification scores.

1. INTRODUCTION

The relation between acoustic vowel reduction and vowel identification has been studied for years (for Dutch, e.g., [3],[5],[6],[7]). Much less is known about consonant reduction and its effect on consonant identification. From previous studies it is clear that reduction can change consonant realizations just as much, or even more, as it can change vowel realizations ([1],[2],[8],[10]). In this paper we will investigate which of these changes are important for consonant identification. Earlier studies have shown that speaking style, e.g., read versus spontaneous speech, has a profound effect on the amount of vowel reduction ([3],[4],[6]) and consonant reduction ([8],[10]). Speaking style is also known to affect phoneme identification (e.g., [3],[6]). Here we will try to link the effects of reduction on consonant acoustics and identification.

From an earlier study on acoustic consonant reduction in spontaneous versus read speech ([8],[10]) we selected four global acoustic measures of consonant reduction: Segment duration, the spectral Center of Gravity (i.e., the "mean" frequency, weighted by spectral power), the Intervocalic Sound-Energy difference (i.e., VCV Energy Difference, the difference in total sound power between consonants and their neighboring vowels), and the difference between the F_2 slopes at the CV and VC borders of the consonant. All of these measures are correlated to speaking style differences and vowel reduction and might be perceptually relevant ([8],[10]). It has been shown that neighboring vowels too play a role in the identification of consonants ([9]). Therefore, the degree of vowel reduction might also influence the identification of neighboring consonants. To assess this influence, we added the distance in the F_1/F_2 plane (in semitones) between the kernel of the tautosyllabic vowel (i.e., the point with the most extreme F_1 or F_2 value) and the center of vowel reduction, i.e. (250, 1300) Hz for this speaker. This distance quantifies the contrast in the vowel system [3].

The first three of these acoustic parameters (segmental duration, spectral Center of Gravity, and Intervocalic Sound-Energy difference are linked to the prosodic structure of the utterance [10]. The formant related acoustic measures are linked to the articulatory structure of the syllables.

2. MATERIALS AND METHODS

For this study we selected recordings of a single speaker who read aloud a transliteration of spontaneous speech recorded earlier (20 minutes of speech each). The orthographic script was transcribed to phonetic symbols and each recording was checked against this transcription and marked for sentence accent by one of us ([8],[10]). From the phonetic transcription, all Vowel-Consonant-Vowel (VCV) segments were located in the speech recordings (read and spontaneous). 1847 VCV pairs had both realizations originating from corresponding positions in the utterances with identical syllable structure, syllable boundary type, and sentence accent and lexical syllable stress. A subset of 791 pairs is used in this study (see table 1, these are the same realizations as used by [10]). The stimulus pairs were selected to cover all consonants and stress conditions present (except for /h/). The pairs were selected randomly for each individual consonant and stress condition (*syllable* stress only).

22 Dutch subjects, all native speakers of Dutch, were asked to identify these 1582 intervocalic consonant realizations in their original VCV context (791 pairs, 308 stressed, 483 unstressed). The outer 10 ms of the VCV tokens were removed and smoothened with 2 ms Hanning windows to prevent interference from the adjacent consonants and transient clicks. The order of

Table 1. Dutch consonants used in this paper and the number of matched Read/Spontaneous VCV pairs (ignoring voicing differences).

	Velar	Pal	Alv	Lab	Total
Plos	kg 63	-	td 65	pb 61	189
Fric	χ 77	ʃʒ 3	sz 63	fv 75	218
Nasal	ŋ 14	-	n 72	m 63	149
V-like	r 60	j 21	l 94	w 60	235
Total	214	24	294	259	791

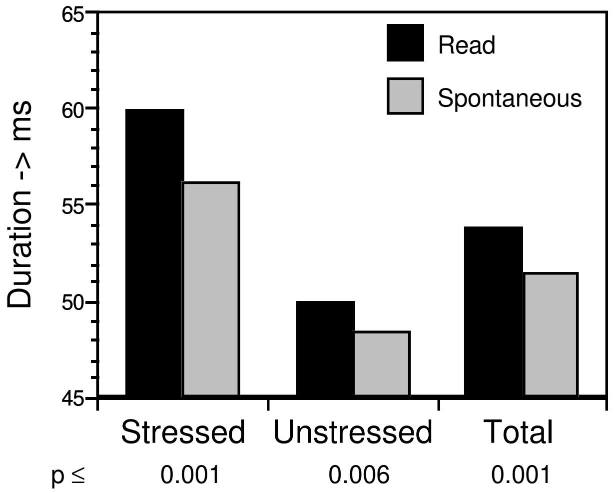


Figure 1. Mean durations of the consonant tokens (in ms), split on speaking style (read and spontaneous) and syllable stress. The significance levels of the differences between read and spontaneous realizations are calculated using the Wilcoxon Matched-Pairs Signed-Ranks test.

presentation was (pseudo-)random and different for each subject. The subjects had to select the Dutch orthographic symbol that corresponded to the sound heard on a computer CRT screen (this causes no ambiguity in Dutch).

In a first approximation, identification rates in a listening experiment can be modeled by a binomial probability distribution. For 22 responses to each stimulus it can be deduced that the standard deviation of the *difference* in the number of correct responses between members of read/spontaneous consonant pairs will be ~3 responses. This “error” has the same order of magnitude as the difference itself. The acoustical measurements add their own errors which too can be expected to be large with respect to the differences between speaking styles. For instance, segmentation errors, and therefore, durations, are comparable to a single pitch period, i.e., ~ 5 ms. However, the mean difference in duration between read and spontaneous consonants is less than 5 ms (see figure 1).

Considering these very large “noise” levels on our data, we decided to use only the *signs* of the differences between speaking styles to investigate the relation between acoustical measurements and identification errors (see table 2). Standard 2x2 Chi-square tests are used to decide on statistical significance of correlations. The strength of the correlation is expressed in terms of the *odds-ratio* between the frequencies found and the frequencies expected from the marginal distributions (see the example in table 2). The odds ratio, as calculated here, indicates the improvement in predicting the differences in identification that can be gained from using the acoustic parameter (and vice versa). Actually, only the signs of the differences between speaking styles are used. These odds ratios use the same tables as the Chi-square test and are especially useful in situations where categorical statistics are preferred.

Table 2. Example of frequency tables used to investigate the correlation between acoustic measurements and correct identification. Table entries are numbers of pairs with the sign of the difference between Read and Spontaneous realizations as indicated. Frequencies expected from the marginal distributions are given in brackets (Row Total · Column Total / Total, e.g., 263·177/609=76.44). The Odds are the sum of the diagonal terms divided by the sum of the off-diagonal terms, i.e., (110+279)/(67+153)=1.768 and (76.44+245.44)/(100.56+ 186.56) =1.121 (found and expected, respectively). The odds ratio is the odds found divided by the odds expected (i.e., 1.768/1.121=1.577). Also: $p \leq 0.001$, $\chi^2=35$, $\nu=1$, Contingency = 0.23. R: Read speech, S: Spontaneous speech, Rows: Duration, columns: identification rate.

	R < S	R > S	Total
R < S	110 (76.44)	153 (186.56)	263
R > S	67 (100.56)	279 (245.44)	346
Total	177	432	609

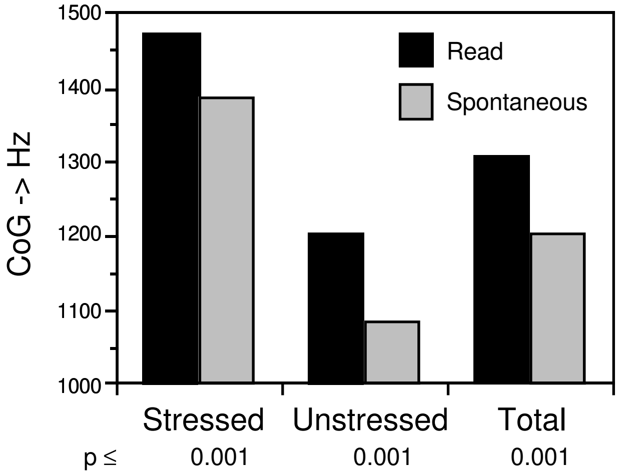
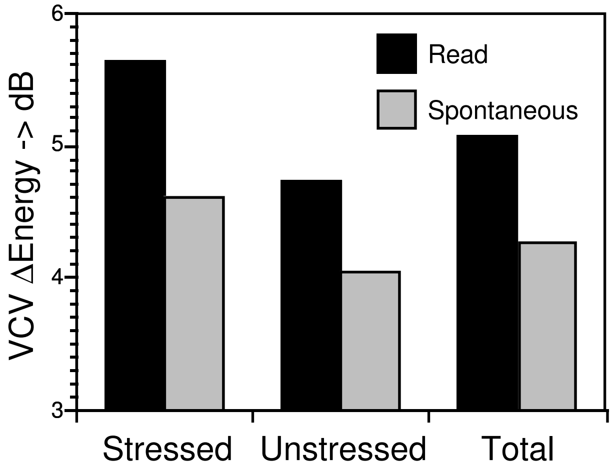


Figure 2. Mean Spectral Center of Gravity (in Hz) of the consonant tokens, split on speaking style (read and spontaneous) and syllable stress. Statistics as in figure 1.

3 RESULTS

Because acoustic consonant reduction is not well documented in the literature [10], we present the effects in detail. The acoustic characteristics of the tokens are summarized in the figures 1-5. It is clear that each of the acoustic measures shows a considerable difference between read and spontaneous speech and between stressed and unstressed realizations (only the former difference is displayed in this paper). The differences indeed point towards consonant reduction in spontaneous speech and unstressed segments, e.g., shorter durations, lower Center of Gravity, smaller intervocalic sound energy differences, shorter formant distances of the adjacent vowels, and larger differences of F_2 slopes. The last measure, a change in F_2 slopes, measures how well articulation speed can keep up with changes in duration.

The global error rates for Spontaneous and Read realizations are displayed in figure 6. The error rates are considerably larger for Spontaneous than for Read speech for both Stressed and Unstressed realization. For both speaking styles, the error rate was larger for unstressed than for stressed realizations ($p \leq 0.001$, $\chi^2>38$, $\nu=1$). We want to know to what extent a change in the value of each of the acoustic markers for consonant reduction is predictive for a change in identification errors. That is, when a marker indicates reduction, can we expect higher error rates and vice versa? This is analyzed as a correlation between the *signs* of the changes in the



p ≤ 0.001 0.001 0.001

Figure 3. Mean intervocalic sound energy difference (in dB) of the tokens, split on speaking style (read and spontaneous) and syllable stress. Statistics as in figure 1.

marker values and in the error rates due to speaking styles.

In figure 7 the result of this analysis is displayed as the ratio of the odds for predicting the correct direction of change using the observed and a random pairing of the identification scores and marker values. This odds ratio is by definition normalized for the mutual correlation with speaking style of both markers and error rates (i.e., the marginal distributions).

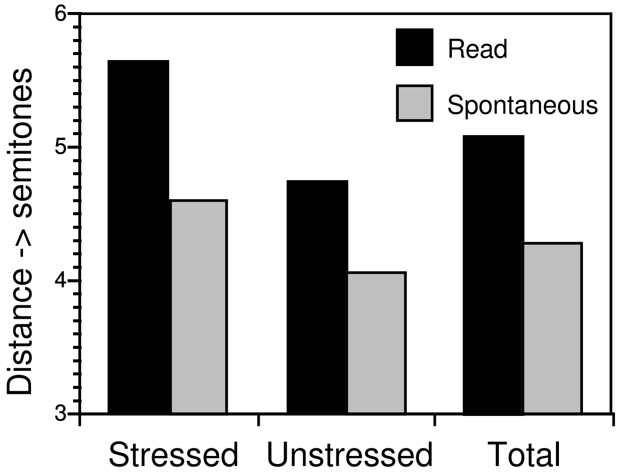
It shows that the correlation between the sign of the acoustic measurement and the identification rate is only statistically significant for the segmental duration and spectral Center of Gravity. It is clear that knowing either the direction of change in segmental duration or the Center of Gravity increases the odds ratio for guessing the correct change in identification rates to a maximum of 1.6. The perceptual relevance of the other markers for consonant identification seems to be marginal, at most.

4. DISCUSSION

In our earlier studies we have shown that consonant Duration, Center of Gravity, Intervocalic Sound Energy Differences, and F_2 slope difference are indicators of vowel and consonant reduction and are all correlated to changes in speaking style ([8],[10]). These effects of reduction are also apparent when differences in syllable stress are involved. It was to be expected that these differences in the acoustics of consonant realizations would affect consonant identification to some extent.

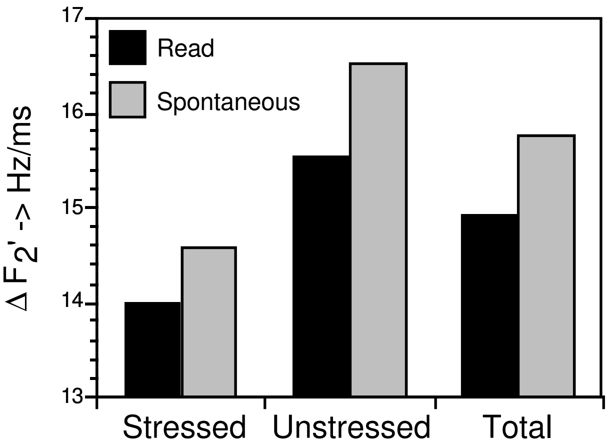
For the first two acoustic markers, segmental duration and spectral Center of Gravity, it could be shown that a "reduction" in their values is correlated to "reduced" identification. For both the Intervocalic Sound Energy Difference and the F_2 Slope difference no such a relation was found. Nor was any influence on consonant identification found of the amount of spectral reduction of the neighboring vowel.

It is evident that the predictive powers of segmental duration and spectral Center of Gravity are limited. The odds ratios are well below 2. That is, knowing the direction of change (i.e., reduction or not) in either of these acoustic factors not even doubles the odds for correctly predicting the direction of change in identification (i.e., better or worse).



p ≤ 0.001 0.001 0.001

Figure 4. Mean formant distance of the tautosyllabic vowel of the consonant stimuli to the center of reduction (250, 1300 Hz) in semitones, split on speaking style (read and spontaneous) and syllable stress. Statistics as in figure 1.



p ≤ 0.029 0.021 0.002

Figure 5. Mean F_2 slope difference between CV and VC boundaries of the consonant stimuli (in Hz/ms), split on speaking style (read and spontaneous) and syllable stress. Statistics as in figure 1.

One cause of this small predictive power is the large error in determining the identification rate and the acoustic measurements themselves. Large errors in the sizes of the differences give very "noisy signs". Even if the correlation between an acoustic parameter and the identification rate would be perfect, the large errors in determining them could give rise to quite "weak" apparent correlation strengths. This problem might have been the cause of the lack of any effect found for the Intervocalic Sound Energy difference and the formant measures. These three acoustic measures depend on the consonant *and* the vowel realizations. Such a dependence on several segments is bound to lead to increased measurement errors.

Another cause for a weak predictive power of individual acoustical factors is that consonant identification does

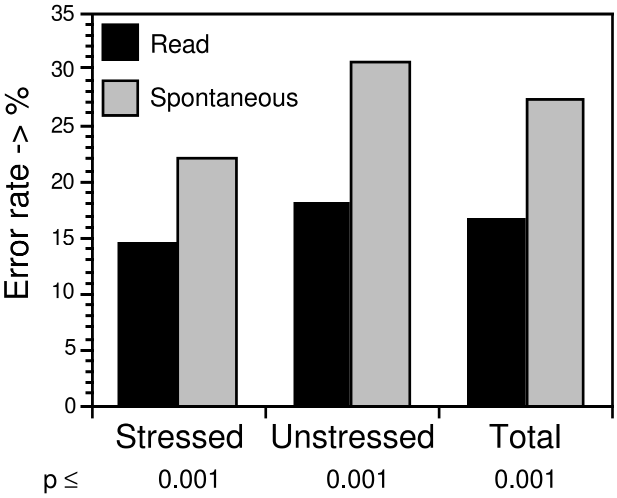


Figure 6. Mean error rates of the consonant stimuli, split on speaking style (read and spontaneous) and syllable stress. The significance levels of the differences between read and spontaneous realizations are calculated using McNemar's test.

not depend critically on a single acoustical feature. Many factors combine to determine consonant identity. Most of these factors too will be influenced by speaking style, and consonant reduction. The effect is again, to weaken the correlation between any single acoustic parameter and consonant identification scores. This is especially so for indirect parameters like the formant contrast that measures the amount of *vowel* reduction. The effects of the amount of vowel reduction on the identification of neighboring consonants might just have been too weak to measure.

There remains the question of how a longer duration and a higher spectral Center of Gravity can help consonant identification. In both cases, more information is available to the listener. It is straightforward that longer segments can carry more articulatory information, and they apparently do so. A higher frequency of the spectral Center of Gravity generally indicates a more level spectral tilt of the sound source at medium and higher frequencies [10]. That is, there is more energy at the higher frequencies and, therefore, more articulatory information in the signal. The fact that "reduction" of both factors actually reduces consonant identification in our experiment indicates that both duration and spectral tilt are "information limiting" in normal speech.

5. CONCLUSIONS

Of the five acoustic parameters tested, only segmental duration and spectral Center of Gravity could be shown to be perceptually relevant. For these two acoustic parameters, it was shown that the odds of predicting the correct direction of change in identification scores were increased by 30-60%. Together this suggests that the articulatory information in normal speech is limited by segment duration and the spectral tilt of the speech sound.

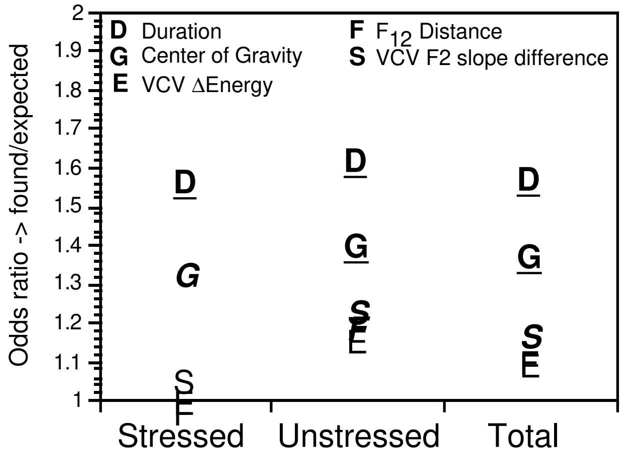


Figure 7. Odds ratios between acoustical measurements and identification rates, split on syllable stress. Underlined: p≤0.001, $\chi^2 > 13$, $\nu = 1$, *Italic*: p≤0.05, $\chi^2 > 3.89$, $\nu = 1$

If the other acoustic parameters, Intervocalic Sound Energy difference, F_2 slope difference and the formant contrast of the tautosyllabic vowel, did affect consonant identification, their values were either too erratic to give rise to discernible effects, or their effects were too small to be resolved by our experiment.

6. REFERENCES

[1] D. Duez. "On spontaneous French speech: aspects of the reduction and contextual assimilation of voiced stops", *Journal of Phonetics* **23**, 1995, 407-427.

[2] E. Farnetani. "The spatial and the temporal dimensions of consonant reduction in conversational Italian", *Proc. Eurospeech'95*, Madrid, 1995, 2255-2258.

[3] F.J. Koopmans-Van Beinum. *Vowel contrast reduction, an acoustic and perceptual study of Dutch vowels in various speech conditions*, Ph.D. Thesis, University of Amsterdam 1980.

[4] S.J. Moon & B. Lindblom. "Interaction between duration, context, and speaking style in English, stressed vowels", *J.Acoust.Soc.Am.* **96**, 1994, 40-55.

[5] L.C.W. Pols & R.J.J.H. Van Son. "Acoustics and perception of dynamic vowel segments", *Speech Communication* **13**, 1993, 135-147.

[6] D. Van Bergem. *Acoustic and lexical vowel reduction*, in *Studies in Language and Language Use* **16**. Ph.D. Thesis, University of Amsterdam, 1995.

[7] R.J.J.H. Van Son. *Spectro-temporal features of vowel segments*, in *Studies in Language and Language Use* **3**. Ph.D. Thesis, University of Amsterdam, 1993.

[8] R.J.J.H. Van Son & L.C.W. Pols. "What does consonant reduction look like, if it exists?", *Proc. Eurospeech'95*, Madrid, 1995, 1909-1912.

[9] R.J.J.H. Van Son & L.C.W. Pols. "The influence of local context on the identification of vowels and consonants", *Proc. Eurospeech'95*, Madrid, 1995, 967-970.

[10] R.J.J.H. Van Son & L.C.W. Pols. "An acoustic profile of consonant reduction", *Proc. ICSLP'96*, Philadelphia, 1996, 1529-1532.

This research was made possible by grant 300-173-029 of the Netherlands Research Organization (NWO)