

Pronunciation reduction: how it relates to speech style, gender, and age

Helmer Strik, Joost van Doremalen, and Catia Cucchiarini

Department of Linguistics, Radboud University Nijmegen, The Netherlands

[H.Strik | J.vanDoremalen | C.Cucchiarini]@let.ru.nl

Abstract

Many studies have shown that speech style affects pronunciation reduction, mixed results have been obtained for gender, and few results have been published regarding the relationship between age and reduction. In the present paper we investigate how pronunciation reduction is related to speech style, gender and age. Significant effects were found for speech style and age, while the effect of gender was not significant.

Index Terms: pronunciation reduction, speech style, age, gender

1. Introduction

Numerous studies have provided a considerable body of evidence indicating that pronunciation reduction is pervasive in everyday speech [1, 2, 3]. In addition, there are indications that pronunciation reduction might vary as a function of sociolinguistic variables such as speech style, gender and age.

In general, stylistic variations are interpreted as variations in the degree of formality of speech [4]. As speech becomes less formal, the syllabic structure of words may be reorganized, speech rate may increase, and there may be changes in pitch and loudness [5, pp. 66-69].

In general, a higher degree of pronunciation reduction is observed in spontaneous and extemporaneous unscripted speech than in more formal, scripted speech [6, 7, 8]. For the variable gender the results are less straightforward. In general, it seems that male speakers exhibit a higher degree of pronunciation reduction than female speakers [9, 10, 11], but several studies failed to find a significant difference between male and female speech [12]. With respect to the variable age less evidence seems to be available.

Several studies have addressed the relationship between speech tempo and age and gender, respectively. Since it has been found that a higher speech tempo induces more pronunciation reduction [13], it might be interesting to look at those findings. In general, it has been observed that younger speakers speak faster than older speakers [14, 15, 16]. Furthermore, on average, men appear to speak faster than women [12, 15]. However, comparisons among studies on speech tempo are made difficult by the fact that many different measures can be used to express speech tempo [17, 18] and, therefore, the results may vary depending on the measures used. An important factor seems to be whether pause time is included in the calculation of speech tempo or not. This becomes particularly apparent when studying the relationship between speech style and speech tempo. Measures that include pause time will indicate lower speech tempo in spontaneous speech whereas measures that do not include pause time are less likely to vary between speech styles [18].

To summarize, many studies have shown that speech style affects reduction, mixed results have been obtained for gender, and few results have been published regarding the

relationship between age and reduction. In the present study we decided to investigate whether pronunciation reduction is related to age and, if such a relationship exists, how this (possible) relationship compares to the way in which pronunciation reduction is related to speech style and gender, respectively.

2. Material and method

The corpus used for the current study is CGN, a database containing about 9 million words of contemporary Dutch as spoken in the Netherlands and Flanders [19, 20]. The broad phonetic transcriptions are manually checked for about 10% of the corpus, the so-called core corpus. In the present study we used the Dutch part of this core corpus.

The transcriptions of the realizations were aligned with the canonical transcriptions by means of the program ADAPT [21]. The output of this program contains some information regarding the alignment, such as the number of deletions (#Del), insertions (#Ins), and substitutions (#Subs). This information was used to calculate some additional measures:

- number of differences : $\#Diff = \#Del + \#Ins + \#Sub$
- percentage of deletions: $\%Del = 100\% * \#Del/LCa$
- percentage of differences: $\%Diff = 100\% * \#Diff/LCa$

where LCa is the length of the canonical transcription (i.e. the number of phonemes in the canonical transcription). All measures were calculated at the word level for all 563.380 words.

The CGN contains information regarding gender, age, and speech style. All speakers were divided in three age groups: 1. 20-39, 2. 40-59, and 3. 60-79. The CGN contains several components [19, 20]. We carried out analysis for all components together, and also for two groups containing components of similar speech styles (which will be referred to as speech style groups, or – more shortly – as speech styles hereafter): (1) *spontaneous*: components ‘spontaneous conversations’ and ‘spontaneous telephone dialogues’; (2) *read*: components ‘read speech’ and ‘broadcast news’.

In the present study, we thus have a 3x2x2 design: 3 age classes, 2 speech styles, and 2 gender values.

3. Results

We first analyzed the percentage of words for which the realization differs from the canonical transcription, either at least in one segment ($\#Diff > 0$), or at least in one deleted segment ($\#Del > 0$). The average values are shown in figure 1. It can be observed that the differences between male and female speech are small, while those between the age groups are more substantial for both variables. On average, in about 15% of the words there is a deletion, and in about 35% of the words there is a change (Del, Ins, or Subs).

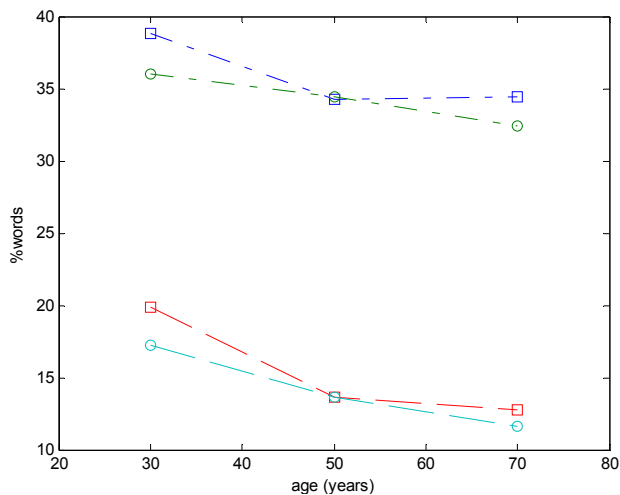


Figure 1. The percentage of words with a change. Symbols: square = male, o = female. Line types: dashed-dotted lines (at the top) - '#Diff>0', i.e. with at least 1 change (Del, Ins, or Sub); dashed lines (at the bottom) - '#Del>0', i.e. with at least 1 deletion.

Average values for different subsets of the data were then calculated. They are presented in table 1. Column 3 contains average values for all words, column 4 for only those words for which there is a difference between the realization and the canonical transcription ('#Diff>0', i.e. words with at least one difference: deletion, insertion, or substitution), and column 5 for words with at least one deletion ('#Del>0'). As is to be expected, in going from column 3 to 5 the number of words - on which the average is based - decreases. In rows 2 to 13, average values of four variables are presented: 1. #Del, 2. %Del, 3. %Diff, and 4. N (the number of words in each subset). For variables 1-3, average values were calculated for three different subsets of the data: 1. all data, 2. spontaneous speech, and 3. read speech.

In Table 1 it can be observed that the average values of the three reduction measures (#Del, %Del, %Diff) are always larger for spontaneous speech than for read speech, and that the average values for all components together are in between, but closer to those of spontaneous speech. The latter can be explained to a large extent by the fact that in our material more spontaneous speech is present.

The average number of deletions per word (#Del) is about 0.18 for all words, 0.52 for the words with a change ('#Diff>0'), and 1.19 for the words with at least one deletion ('#Del>0'). For %Del these numbers are approximately 5%, 14%, and 31%, respectively. And for %Diff they are 13%, 36%, and 38%, respectively.

If we look again at the average values of %Del for the different subsets, we see that the average percentage of deleted phones for all words is about 5%, for the 35% words

Table 1. In rows 2-10 average values of the three factors (#Del, %Del, and %Diff) studied for different sub-sets are presented, in rows 11-13 the total number of words (N) in each of these sub-sets is shown. Column 3 contains the results for all words, column 4 for those words with at least one change (deletion, substitution, or insertion), and column 5 for the words with at least one deletion.

		All words	Words with %Diff > 0	Words with %Del > 0
#Del	All data	0.18	0.52	1.19
	Spontaneous	0.24	0.61	1.21
	Read	0.09	0.27	1.08
%Del	All data	4.86	13.68	31.29
	Spontaneous	6.55	16.99	33.57
	Read	1.99	6.23	24.60
%Diff	All data	12.75	35.88	37.98
	Spontaneous	14.86	38.54	40.32
	Read	9.86	30.87	31.53
N	All data	563,380	200,133	87,491
	Spontaneous	263,241	101,517	51,367
	Read	90,519	28,918	7,328

with a change it is 14%, and for the 15% words with at least one deletion it is 31%. Consequently, the magnitude of the average values differs considerably depending on the subset chosen. However, in general trends appeared to be similar between subsets. The results presented below are based on (analysis for) all words. Furthermore, these numbers make clear that for the subset of words in which deletions take place, the average number of changes is very large: on average 31.29% of the phones is deleted, and 37.98% of the phones are not pronounced in the canonical way.

The relation between the average values of #Del, %Del, and %Diff and the factors age, speech style, and gender are visualized in figures 2a to 2c. For the sake of clarity, values (symbols) of the same speech style - gender combination are connected with lines. These lines thus make clear what the trends are as a function of age. Similar behavior for the three variables can be observed in figures 2a to 2c: the average values are always larger for spontaneous speech, the differences between age groups are smaller than those between speech styles, but larger than those between male and female speech. For spontaneous speech reduction gradually decreases with age. For read speech such a clear decrease with age is not observed.

Statistical tests were carried out to study the effects of the three factors age, gender, and speech style. We carried out an ANOVA for the 3x2x2 design. The results are shown in Table 2. The results in Table 2 show that, for all three variables studied, speech style and age have a significant effect ($p < 0.01$), while gender does not. Furthermore, the effect of the

Table 2. Results of the ANOVA test. For each of the three factors (#Del, %Del, and %Diff) the significance levels and eta-square values are shown in the columns. The rows contain the results of the main effects and the interactions.

Factor	#Del		%Del		%Diff	
	p-value	η_p^2	p-value	η_p^2	p-value	η_p^2
Gender	0.32	$2.33 \cdot 10^{-6}$	0.17	$5.34 \cdot 10^{-6}$	0.48	$1.40 \cdot 10^{-6}$
Speech style	0	$1.02 \cdot 10^{-2}$	0	$1.39 \cdot 10^{-2}$	0	$7.30 \cdot 10^{-3}$
Age	0	$6.96 \cdot 10^{-4}$	0	$4.40 \cdot 10^{-4}$	0	$2.70 \cdot 10^{-4}$

variables decreases in the following order: speech style – age – gender. In order to have an indication of the magnitude of the differences caused by these factors, average values of the differences were calculated (see table 3).

Table 3. Average differences between classes. In columns 2-4 the results for the three factors studied are presented. Shown in rows 2-4 are the differences for speak style, age, and gender, respectively.

	#Del	%Del	%Diff
$\Delta(\text{Spontaneous-Read})$	0.15	4.56	5.00
$\Delta(\text{Young-Old})$	0.09	2.47	2.59
$\Delta(\text{Male-Female})$	0.02	0.99	1.30

It can be observed that the largest differences are found for speech style, followed by age, and that the differences for gender are much smaller. For instance, the differences between genders are a factor 4-6 smaller than those between age groups. A decrease in significance, explained variance, and magnitude of the differences is observed in the order speech style, age and gender. Speech style and age have a significant effect, while gender does not.

4. Discussion and conclusions

The relationship between pronunciation reduction and three sociolinguistic variables was investigated in the study reported on in the present paper. Two of the three factors appear to have a highly significant effect on different indicators of pronunciation reduction. In this study the effect of the three factors was studied at word level, for a large amount of different words in a very global way. The drawback of using such an approach is that there will probably be much noise in the data, while the advantage is that the analyses are based on a substantial amount of data. Additional studies should then be carried out to gain more insight and to study the effect of the different variables in more detail, as will be explained below. Nevertheless, clear and consistent trends emerged from these large amounts of data.

The finding that pronunciation reduction is related to speech style is in line with previous findings [refs] and does not come as a surprise. The effect of gender, on the other hand, was less clear-cut in previous studies. Our results are in line with those reported by [5] who used speech from the same corpus.

What is somewhat surprising is that age appears to be more strongly related to pronunciation reduction than gender, a finding that, as far as we know, had not been reported previously. Age thus seems to be a factor that deserves more attention. The factors speech style and gender are often taken into account in studies in different fields, but for age this is less often the case. For instance, in studies on pronunciation (variation, reduction, tempo, etc.) in speech sciences, speech style and gender play a role much more often than age. In speech technology there are different ‘pronunciation’ models (e.g. in speech recognition) for males and females, and also for different speech styles. As to age, it plays a role in the extreme cases of children and elderly people, for which special measures have usually to be taken, but less so for youngsters and adults. Since the results presented here show that age has a larger effect than gender on pronunciation (reduction), especially in spontaneous speech, including age in pronunciation(-related) models should be considered.

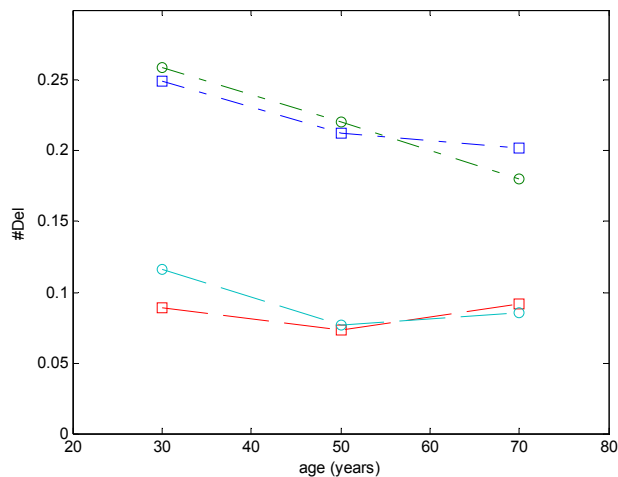


Figure 2a. #Del(age) for speech style & gender.

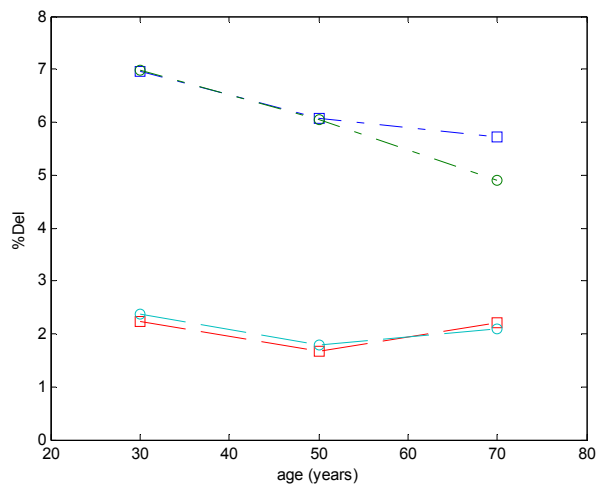


Figure 2b. %Del(age) for speech style & gender.

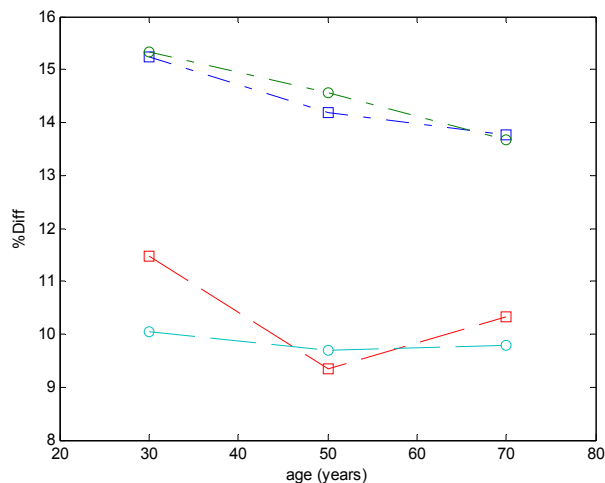


Figure 2c. %Diff(age) for speech style & gender.

Figure 2. #Del, %Del, and %Diff as a function of age. Symbols: square = male, o = female. Line types: dash-dotted (at the top) - spontaneous, and dashed (at the bottom) – read.

While pronunciation reduction clearly decreases with age in spontaneous speech, such a clear trend is not visible for read speech. This is very plausible. Reading out loud is a rather more formal task in which speakers tend to adhere to canonical forms and are more influenced by the orthography. This of course leaves less room for variation than spontaneous speech where orthography is not provided.

A possible limitation of these results is that they only indicate a significant relationship between measures of pronunciation reduction and sociolinguistic variables such as speech style and age and they do not provide information as to how varying speech style and age would affect pronunciation reduction. In other words, we are not in a position to make strong claims about possible causal relationships obtaining between speech style and age on the one hand and pronunciation reduction on the other.

In particular, with respect to age the question arises as to how the observed pattern should be interpreted. What we observe is a monotonic slope with age, which, as explained by Labov [17; see also 22] might be due to “age grading” or to “apparent time”. According to the “age grading” interpretation individuals change their pronunciation as they grow older. The alternative explanation, referred to as “apparent time”, holds that age groups that successively enter the speech community exhibit different pronunciation patterns, in our case characterized by increased reduction. According to this latter interpretation, a monotonic slope with age, measured at one point in time, would be indicative of change in progress [22]. Evidence for the “age grading” interpretation would come from studies on speech tempo that indicate that older speakers speak slower than younger speakers [refs]. As faster tempo is generally associated with more reduction, the lower amount of reduction observed in older speakers would result from slower speech tempo, which, in turn, is partly an aging phenomenon. Alternatively, it is possible that the older speakers in the corpus investigated always spoke in the same way and that the larger amount of reduction observed in younger speakers is due to language change. Obviously we are not in a position to resolve such a dilemma, as was the case with many of the sociolinguistic studies that were carried out in the 60s and 70s and which, later on, were followed by longitudinal studies aimed at disambiguating the two interpretations [22].

A final caveat about the results reported in this paper concerns the magnitude of the effects observed. Although for two of the independent variables significant effects were observed, the magnitude of the effect is relatively small, as appears from the partial η_p^2 shown in Table 2. The small magnitude of the effect is probably a consequence of the substantial between-subjects variation. As ever larger corpora are becoming available, the number of data (N) in statistical tests is often quite large. It is well-known, that for large values of N, factors can be highly significant even though their explained variance is small. Unfortunately, it is difficult to make comparisons with other studies, because many of the studies that address the relationship between speech characteristics such as speech tempo or reduction and sociolinguistic variables such as speech style, gender and age do not even report results on effect size.

To conclude, in this paper we have presented some global results on the relationship between pronunciation reduction and sociolinguistic variables like speech style, gender and age. For speech style and age we have found significant, albeit modest effects, which indicate that spontaneous speech is characterized by more reduction than read speech, and that

younger speakers reduce more in spontaneous speech than older speakers.

5. References

- [1] M. Ernestus, “Voice Assimilation and Segment Reduction in Casual Dutch,” Ph.D. dissertation, Free University of Amsterdam, 2000.
- [2] K. Johnson, “Massive Reduction in Conversational American English”, Yoneyama, K., Maekawa, K. (Eds.) *Spontaneous Speech: Data and Analysis*. Tokyo: The National Institute for Japanese Language, pp. 29-45, 2004.
- [3] C. Van Bael, H. Baayen, H. Strik, “Segment Deletion in Spontaneous Speech: A Corpus Study using Mixed Effects Models with Crossed Random Effects,” in *Proceedings of Interspeech*, 2007, pp. 2741–2744.
- [4] W. Labov, *Sociolinguistic Patterns*. University of Pennsylvania Press, Philadelphia, 1972.
- [5] J. Laver, *Principles of Phonetics*. Cambridge University Press, Cambridge, 1994.
- [6] D. Van Bergem, “Acoustic and lexical vowel reduction,” Ph.D. dissertation, University of Amsterdam, 1995.
- [7] R.J.J.H. Son and C.W. Pols, “An acoustic description of consonant reduction,” *Speech Communication*, vol. 28, pp. 125–140, 1999.
- [8] C.P.J. Van Bael, H. van den Heuvel & H. Strik, “Investigating Speech Style Specific Pronunciation Variation in Large Spoken Language Corpora,” in *Proceedings of ICSLP*, 2004.
- [9] D. Byrd, “Relations of sex and dialect to reduction,” *Speech Communication*, vol. 15, pp. 39-54, 1994.
- [10] A. Bell et al., “Forms of English function words - Effects of disfluencies, turn position, age and sex, and predictability,” in *Proceedings of the International Congress of Phonetic Sciences*, 1999, pp. 395-398.
- [11] M. Keune, M. Ernestus, R. van Hout, & R.H. Baayen “Social, geographical, and register variation in Dutch: From written mogelijk to spoken mok,” *Corpus Linguistics and Linguistic Theory*, vol. 1, pp. 183-223, 2005.
- [12] D. Binnenpoorte et al., “Gender in Everyday Speech and Language: A Corpus-based Study,” in *Proceedings of Interspeech*, 2005, pp. 2213-2216.
- [13] Fosler-Lussier and N. Morgan, “Effects of speaking rate and word frequency on pronunciations in conversational speech,” *Speech Communication*, vol. 29, Issues 2-4, pp. 137-158, 1999.
- [14] L. Ramig, “Effects of physiological aging on speaking and reading rates,” *Journal of Communication Disorders*, vol. 16, pp. 217-226, 1983.
- [15] J. Verhoeven, G. De Pauw and H. Kloots, “Speech rate in a pluricentric language: A comparison between Dutch in Belgium and the Netherlands,” *Language and Speech*, vol. 47, pp. 299-310, 2004.
- [16] H. Quené, “Multilevel modeling of between-speaker and within-speaker variation in spontaneous speech tempo,” *Journal of the Acoustical Society of America*, vol. 123 (2), pp. 1104-1113, 2008.
- [17] C. Cucchiari, H. Strik & L. Boves, “Quantitative assessment of second language learners' fluency by means of automatic speech recognition technology,” *Journal of the Acoustical Society of America*, vol. 107 (2), pp. 989-999, 2000.
- [18] C. Cucchiari, H. Strik & L. Boves, “Quantitative assessment of second language learners' fluency: Comparisons between read and spontaneous speech,” *Journal of the Acoustical Society of America*, vol. 111 (6), pp. 2862-2873, 2002.
- [19] N. Oostdijk, “The Spoken Dutch Corpus. Overview and first evaluation,” in *Proceedings of LREC*, 2000, pp. 887-894.
- [20] CGN website, <http://lands.let.ru.nl/cgn/ehome.htm>, CGN.
- [21] A. Elffers, C. Van Bael and H. Strik, “Adapt: Algorithm for Dynamic Alignment of Phonetic Transcriptions,” CLST internal report, 2005.
- [22] Sankoff, G. Age: Apparent time and real time. Elsevier Encyclopedia of Language and Linguistics, Second Edition, 2006. Article Number: LALI: 01479.