POSTVOCALIC /r/-DELETION IN DUTCH: MORE EXPERIMENTAL EVIDENCE

Catia Cucchiarini and Henk van den Heuvel
A²RT, Dept. of Language and Speech, University of Nijmegen, the Netherlands
{C.Cucchiarini,H.v.d.Heuvel}@let.kun.nl, http://lands.let.kun.nl/

ABSTRACT

In this study we investigate three properties of the left vowel context which we hypothesize to underlie deletion of /r/ in postvocalic, preconsonantal position in Dutch spontaneous speech: vowel type (schwa, full vowel), vowel length (long, short) and lexical stress (+,-). For each of five categories 90 instances with possible /r/-realizations were extracted from a large speech database containing man-machine dialogues in an automatic train-table inquiry system. The frequency of /r/-deletions in these 450 cases was investigated on the basis of variant selection by a CSR and human transcriptions of the same material. Loglinear analyses revealed that /r/-deletion was significantly more frequent after schwa than after full vowels, and that vowel length and lexical stress were not significant. This appeared from both CSR data and human transcriptions. The only discrepancy between the two sets of results concerns the absolute frequencies of /r/-deletion. Possible explanations are discussed.

1. INTRODUCTION

A well-known peculiarity of rhotic sounds is that they exhibit many similarities from a phonological point of view, while phonetically they appear to be rather diverse [7]. The greatest amount of variation is observed when /r/ appears in coda position [7, 11, 12]. Even within one and the same language a considerable amount of variation is observed [11, 12]. For instance, for Dutch up to ten different realizations have been distinguished for postvocalic /r/ alone [10: 130-131]. Another characteristic of postvocalic /r/ realizations is that they "tend to become vowels or disappear altogether" [7]. This is exactly what is going on with Dutch postvocalic /r/ in connected speech [10: 139].

This paper deals with /r/ in postvocalic position and before another consonant, since here /r/-deletion is most obvious in Dutch. In a previous paper [4], we accounted for the connected speech processes (CSPs) of postvocalic /r/-weakening and /r/-deletion in Dutch by referring to the linguistic context in which these phenomena are more likely to occur. On the basis of the data presented, it was concluded that a process of /r/-weakening is operative in Dutch connected speech, which, in certain contexts, may lead to complete deletion of postvocalic /r/. /r/-weakening can be observed in all positions where /r/ is preceded by a vowel and followed by a consonant, whereas /r/-deletion does not occur everywhere. An important factor to be acknowledged is the left vowel context [4].

The best way to test any hypothesis about these (and other) synchronic phenomena is to study them in real-life speech. For this reason in [4] we suggested the use of the large speech databases and the techniques that have been made available for the purpose of ASR in order to study these processes in real-life extemporaneous speech. Consequently, in a following paper [5] we set out to obtain experimental evidence for the occurrence of postvocalic /r/-deletion in Dutch by checking how often /r/-deletion occurred in a large spontaneous speech corpus (214,102 words). To this end a continuous speech recognizer (CSR) in forced recognition mode was used. The CSR had the task to decide whether /r/-deletion was applied or not in a word.

The accuracy of forced recognition in selecting the correct variant was checked in an experiment [6] in which the CSR's responses were compared with those of nine expert listeners who carried out the same task. For /r/-deletion the CSR turned out to select the correct pronunciation variant in 75.6% of the cases, while for the listeners this percentage varied between 74% and 93%. This indicated that there was a good correspondence between recognizer response and human ratings for /r/-deletion.

The results of forced recognition showed that in a corpus containing 214,102 words in which /r/-deletion could be applied 16,865 times, it was actually applied in 47.6% of the cases. Moreover, the results indicated that the frequency of occurrence of /r/-deletion was dependent on the left vowel context. First, we expected to find more instances of /r/-deletion when the left vowel is a schwa as opposed to any other kind of vowel. This was also confirmed by the analyses [5]. Second, we thought the frequency of /r/-deletion may vary depending on whether the preceding vowel is short or long, for the following reason. If an /r/ is deleted after a short vowel in polysyllabic words, then some restructuring has to take place because a short vowel is not allowed in syllable-final position in Dutch [2: 5]. This requirement might have an inhibiting effect on the application of /r/-deletion, with the consequence that /r/-deletion is more frequent after long vowels than after short vowels, at least in polysyllabic words However, the analyses showed that vowel length had no significant effect on /r/-deletion frequency. A possible explanation for this finding could be that in the experiment reported on in [5] the effect of vowel length was confounded with that of stress. If we consider that schwa is "unstressable" in Dutch [2: 20], then the findings presented in [6] suggest that the phenomenon of /r/-deletion might be related to stress: it is more frequent after a vowel that is never stressed like schwa than after any other vowel that can potentially bear stress. Since in the experiment in [5] no distinction was made betweeen stressed and unstressed short and long vowels, the relation between stress and /r/-deletion could not be investigated.

In order to study these factors separately, we decided to carry out another experiment that was directed at studying the phenomenon of /r/-deletion under various conditions of left vowel context and stress. Furthermore, although the performance of the CSR had been shown to be comparable with that of expert listeners, we thought evidence from human transcribers was needed for a good understanding of the relationship between /r/-deletion and left vowel context. Therefore, the speech material used as input to the

CSR was also transcribed by thirteen transcribers.

To summarize, the aim of the experiment reported in this paper is to determine whether /r/-deletion is dependent on the length and the degree of stress of the preceding vowel, or whether the distinction between schwa and full vowel is the only determinant of this phenomenon. To establish this two different types of evidence were gathered: evidence from the performance of a CSR and evidence from human transcribers. Both types of data will be presented and analyzed in the rest of this paper.

2. METHOD

2.1. Speech Material and Design

Since we wanted to investigate the phenomenon of /r/-deletion in real-life extemporaneous speech, a database was used that contains spontaneous speech recorded over telephone lines, which stems from man-machine interactions in an automatic train time-table inquiry system [9]. The waveform format of the speech files is A-law, sampled at 8 kHZ. The VIOS1 training database was used to train the CSR (see further section 2.3). The test utterances were selected so as to obtain a full factorial design for the relevant effects (vowel quality, length and stress). Accordingly, five classes for the left vowel of /r/ were distinguished: 1. schwa (which is always unstressed); 2. short vowel, stressed; 3. short vowel, unstressed; 4. long vowel, stressed; 5. long vowel, unstressed. Note that before /r/ only five vowels in Dutch are short: ═☆, ✷, ✣, ✡, }/. For each category 90 samples (target words) were chosen, giving a total of 450 samples. This rendered the task feasible for the human transcribers, while still permitting sufficient samples for statistical analysis of the data.

The 450 target words were taken from utterances that were not in the training set. Furthermore, care was taken that the target words were not monosyllabic. In monosyllabic words with a short full vowel, syllable reorganization is not required because the short vowel would not be syllable-final anyway. The 450 target words selected stemmed from 425 unique utterances from 385 different dialogues.

2.2. Automatic Variant Selection

On the basis of the /r/-deletion rule as specified in [4, 5], variants with /r/ and variants without /r/ were generated for all relevant words in the test set in which the relevant contexts were met, by means of a Perl implementation of the rule in question. For each of these words the pronunciation variants with and without /r/ were included in the lexicon that was used for the forced recognition. In forced recognition mode the CSR is not used to select the spoken words, but it is forced to choose from among alternative variants of the same word. This procedure is usually applied when one is not interested in determining which word was spoken, but which pronunciation form of that word, as in this study.

In the present experiment a standard CSR was used with context-independent HMMs for 35 monophones. The models were trained on the canonical transcriptions of the words. This means that during training /r/ was not deleted anywhere from the phoneme transcriptions of the words. /l/ and /r/ had separate models for prevocalic and postvocalic position in the syllable. The postvocalic model of /r/ is tested for the /r/-deletion experiments reported on here. Each monophone consists of three segments of two equal HMM states. Speech is coded as 14 mel-based cepstra (c[0]-c[13]) and 14 corresponding delta cepstral coefficients. The frame width is 16 ms and frame shift is 10 ms. For details about the CSR, the reader is referred to [9]. The phoneme models were trained on the training material of the VIOS1 database containing 25,104 utterances stemming from 3,530 different dialogues.

2.3 Human Transcriptions

Thirteen transcribers were asked to transcribe the above-mentioned 450 realizations of postvocalic /r/. The transcribers were Language and Speech Pathology students at the University of Nijmegen. They were all female. All had attended the same transcription course including 32 hours contact time. The transcription system used in this course is that of the International Phonetic Association (IPA).

The transcribers worked in small groups of two or three people (five duos and one trio) and based their transcription on auditory analysis of the full utterances without any kind of visual support. Each group produced a consensus transcription of one sixth of the material (75 realizations). The task was to transcribe only the target /r/. The transcribers were not just instructed to determine whether they heard an /r/ or not, but they had to use the full range of IPA symbols and diacritics at their disposal to describe their auditory impression.

3. RESULTS

3.1 General Findings

The transcriptions of five /r/-realizations turned out to be missing, so that we had to remove these cases from our analyses. Table 1 presents the frequencies and percentages of /r/-deletions in all categories both for the human listeners and for the forced recognition by the CSR.

		CSR		Human	
	possible	applied	perc.	applied	perc.
after schwa	89	50	56%	30	34%
after short vowel	179	40	22%	24	13%
+ stress	90	19	21%	12	13%
- stress	89	21	24%	12	13%
after long vowel	177	47	27%	13	7%
+ stress	90	21	23%	5	6%
- stress	87	26	30%	8	9%

Table 1. Number of possible applications, number of applications and percentage of application of /r/-deletion in the various contexts, for both the machine and the human responses.

For the CSR responses Table I shows that the type of vowel preceding /r/ has an effect on the amount of deletion: /r/-deletion is clearly more likely to occur after schwa than after any other type of vowel, as already observed in [5]. However, the effect of the factor stress is not very clear: for short vowels there is hardly any difference, while for long vowels there seems to be more deletion after unstressed vowels.

Table 1 further shows that the human listeners scored fewer cases of /r/-deletion, overall, than the CSR did. It can also be seen that the scores of the human transcribers show the same behavior as

those of the CSR. After schwa /r/-deletion occurs much more frequently than after full vowels, whereas the factors length and stress within the full vowels do not seem to influence /r/-deletion markedly. We find a tendency for long vowels to have somewhat more /r/-deletions after unstressed vowels than after stressed vowels, which was also observed for the CSR.

The above mentioned impressions for the /r/-deletions in the human and the CSR data sets have to be substantiated by a series of more thorough statistical analyses. The next subsections report on these analyses.

3.2 Automatic Variant Selection

Loglinear analyses were performed since they are typically suited to deal with frequency data in more complex factorial designs [8]. The first impression to be tested for our data is the hypothesis that factors 'vowel length' and 'stress' do not have a significant impact on /r/ deletion after full vowels. To test the effects of the two factors a hierarchical loglinear analysis was carried out on a subset of the data. The contingency table used as input to loglinear analysis is given in Table 2.

	short vowel		long vowel		Total
	+ stress	- stress	+ stress	-stress	
- r	19	21	21	26	87
+ r	71	68	69	61	269
Total	90	89	90	87	356

Table 2. Contingency table for full vowels with frequencies of /r/-deletion (-r) and /r/-retention (+r) as affected by vowel length and stress. Scores for CSR in forced recognition mode.

In carrying out loglinear analysis we started with a saturated model, that is one containing all possible effects: '/r/-deletion', 'vowel length', 'stress' and the four possible interactions. The results of this analysis show that none of the interaction effects is significant and that the factors 'vowel length' and 'stress' have no significant effect either; a significant effect was found only for the '/r/-deletion' factor ($z = -9{,}11$; $p < 0.001$). In other words, the frequency of /r/-deletion does not seem to be dependent on the length of and/or the amount of stress on the preceding vowel.

In the light of these findings we decided to pool the data pertaining to the groups of stressed and unstressed short and long vowels so as to form a new category 'full vowel' which could then be compared with schwa. Again a hierarchical loglinear analysis with a saturated model was carried out, this time with the factors '/r/-deletion', 'vowel quality' (with the two levels 'full vowel' and 'schwa'), and the interaction between the two factors.

Table 3 shows the data submitted to loglinear analysis. The results of this analysis show that only a saturated model is appropriate, since both effects and their interaction are significant: $z = 5.58$ ($p < 0.001$) for the interaction effect, $z = -3.58$ ($p < 0.001$) for /r/-deletion and $z = -10.06$ ($p < 0.001$) for vowel quality. In other words, the difference in degree of /r/-deletion observed between schwa and the other vowels appears to be statistically significant.

	full vowel	schwa	Total
- r	87	50	67
+ r	269	39	378
Total	356	89	445

Table 3. Contingency table for full vowels and schwa with frequencies of /r/-deletion (-r) and /r/-retention (+r) as affected by vowel quality. Scores for CSR in forced recognition mode.

3.3. Human Transcriptions

To substantiate our impressions of the human transcriptions we performed two hierarchical loglinear analyses with the same design as those outlined in section 3.2. First, the effect of factors 'vowel length' and 'stress' for the full vowels was tested with the data in Table 4.

	short vowel		long vowel		Total
	+ stress	- stress	+ stress	-stress	
- r	12	12	5	8	37
+ r	78	77	85	79	319
Total	90	89	90	87	356

Table 4 Contingency table for full vowels with frequencies of /r/-deletion (-r) and /r/-retention (+r) as affected by vowel length and stress. Scores found for human transcribers.

As for the CSR scores, only the factor '/r/-deletion' was significant ($z = -12.11$; $p < 0.001$), whereas the factors 'vowel length' and 'stress' and all interactions were not.

For this reason, we pooled the data for the full vowels and compared them with the /r/-deletions after schwa, as displayed in Table 5.

	full vowel	schwa	Total
- r	37	30	67
+ r	319	59	378
Total	356	89	445

Table 5. Contingency table for full vowels and schwa with frequencies of /r/-deletion (-r) and /r/-retention (+r) as affected by vowel quality. Scores found for human transcribers.

The corresponding hierarchical loglinear analysis revealed that only the saturated model fitted the data appropriately. Both '/r/-deletion' and 'vowel quality' proved significant ($z = -9.98$, and $z = -6.70$, resp.) as well as their interaction ($z = 5.23$), all at $p = 0.001$

level. This means that the difference in degree of /r/-deletion observed between schwa and the other vowels is statistically significant, which replicates the outcome for the CSR transcriptions.

4. DISCUSSION AND CONCLUSIONS

In this paper we have examined the process of postvocalic /r/-deletion in Dutch from two different perspectives: that of pronunciation variants selected by a CSR and that of phonetic transcriptions made by human transcribers. With respect to the factors influencing the /r/-deletion process both sets of data produce the same results. In particular, the quality and the characteristics of the left vowel context appear to be relevant only to a certain extent for /r/-deletion: if the preceding vowel is a schwa, /r/-deletion is favored, whereas if this vowel is a full vowel /r/-deletion is less likely. For the rest no other distinction within the 'full vowel' category seems to be relevant, neither in length nor in degree of stress. This pattern of results also seems to be rather stable, as it emerged from both types of data.

These findings would seem to suggest that the stress distinction that is relevant to the process of /r/-deletion is not a gradual one, but a dichotomous one in terms of stressable *vs* unstressable, with "unstressability" favoring /r/-deletion. Since distinctions are fainter in an unstressed environment [1:210], such ultimately unstressed syllables, like those containing schwa, can more easily tolerate further reduction, like that of postvocalic /r/, than any other type of syllable.

The data presented in the previous section also confirm that /r/-deletion does indeed occur in Dutch spontaneous speech. Instances of /r/-deletion are observed in all contexts investigated, albeit with different frequency.

Although the two sets of data are in concordance as to the impact of the factors under investigation, it is clear that they differ with respect to the percentage of occurrence of /r/-deletion. More precisely, the CSR data reveal higher percentages of /r/-deletion than the human transcriptions. There are good reasons for this.

First, the task carried out by the transcribers differed essentially from the CSR's task. The transcribers could choose from among many possibilities, while the CSR's choice was dichotomous. In other words, the CSR was forced to choose between a full /r/ and a deleted /r/, whereas the transcribers could indicate various degrees of weakening by using diacritics or other phonetic symbols like schwa. As a matter of fact, in many cases they indeed transcribed a schwa instead of an /r/. One might argue that these occurrences should be counted as instances of /r/-deletion, especially if schwa is the left vowel context. In fact, by considering schwa transcriptions as /r/-deletions the responses from the CSR and the human transcribers look quite alike in terms of absolute frequencies. However, for the present analyses we decided to adopt a more conservative approach and limited ourselves to factual /r/-deletion cases. In the near future we intend to analyze the transcription data in more detail in order to see how they differ from the CSR's data.

A second possible explanation for the lower percentages of deletion in the human data is that human transcribers are likely to be influenced by their knowledge of the orthographic representation of words [3]. Since in this experiment the transcribers listened to whole utterances, they knew which words the speaker was uttering and this might have induced them to actually 'hear' an /r/ when in fact it was not there. However, we should not forget that transcribing connected speech processes in categorical terms may be very difficult in certain cases, as has been pointed out by [2: 126]. Typically, certain phenomena, like the one discussed in this paper, may exhibit a gradual nature, so that in certain cases it can be very difficult to determine whether /r/ is extremely weakened or completely deleted.

Finally, another reason why there was such a discrepancy between the CSR data and the human transcriptions in 'perceiving' /r/-deletion is related to very nature of the HMMs employed in the CSR. These models are essentially static, in the sense that they do not take much account of neighboring sounds. With respect to human perception, on the other hand, we know that the way one sound is perceived very much depends on the identity of the adjacent sounds.

To summarize, the data presented in this paper reveal some differences in percentages of /r/-deletion between the CSR data and the human transcriptions. These differences, however, do not interfere with the main aim of this investigation. Our main objective was to determine which properties of the left vowel context favor postvocalic /r/-deletion. In this respect both types of data examined yield the same results: /r/-deletion turns out to be much more common after schwa than after any full vowel, irrespective of the length and stress characteristics of that vowel.

ACKNOWLEDGEMENTS

We would like to thank Mirjam Wester and Judith Kessens for preparing the tapes for the human transcribers, and for carrying out the forced recognition and making the results available to us. We thank Felix Vieregge for kindly integrating our transcription material in his course curriculum.

REFERENCES

[1] Boersma, P. 1998. *Functional phonology. Formalizing the interactions between articulatory and perceptual drives.* PhD Thesis. University of Amsterdam.

[2] Booij, G. 1995. *The phonology of Dutch*, Oxford, Clarendon Press.

[3] Cucchiarini, C. 1993. *Phonetic transcription: a methodological and empirical study* PhD Thesis, University of Nijmegen.

[4] Cucchiarini, C. and Van den Heuvel, H. 1995. /r/-deletion in Standard Dutch. *Proc. Department of language and Speech* 19, 59-65.

[5] Cucchiarini, C. and Van den Heuvel, H. 1998. /r/-deletion in Standard Dutch. How phonological research can profit from ASR technology. *Proc. ICSLP'98,* 5, 1759-1762.

[6] Kessens, J., Wester, M., Cucchiarini, C., and Strik, H. 1998. The selection of pronunciation variants in spontaneous speech: Comparing the performance of man and machine. *Proc. ICSLP'98, 6, 2715-2718.*

[7] Lindau, M. 1985. The story of /r/. In Fromkin, V. (ed.) *Phonetic Linguistics: Essays in honor of Peter Ladefoged*, Orlando, Academic Press, 157-168.

[8] Rietveld, T. and Van Hout, R. 1993. *Statistical techniques for the study of language and language behaviour.* Berlin, New York: Mouton de Gruyter.

[9] Strik, H., Russel, A., Van den Heuvel, H., Cucchiarini, C., Boves, L. 1997. A spoken dialogue system for the Dutch public transport information service. *International Journal of Speech Technology*, 121-13.

[10] Van de Velde, H. *1996. Variatie en verandering in het gesproken Nederlands (1935-1993)*, PhD Thesis, University of Nijmegen.

[11] Van Reenen, P. 1994. Driemaal -r- in de Nederlandse dialecten. *Taal en Tongval* 7, 54-72.

[12] Vieregge, W.H. and Broeders, T. 1993. Intra- en interspeaker variation of /r/ in Dutch, *Proc. Eurospeech'93*, Berlin, 267-270.