MODELING PRONUNCIATION VARIATION FOR ASR: COMPARING CRITERIA FOR RULE SELECTION

J.M. Kessens, H. Strik & C. Cucchiarini

A^{2}RT, Dept. of Language and Speech, Univ. of Nijmegen, the Netherlands
{J.Kessens, Strik, C.Cucchiarini}@let.kun.nl
http://lands.let.kun.nl/

ABSTRACT

In this paper we use a data-driven (DD) rule-based method for modeling pronunciation variation. Error analysis is performed in order to gain insight into the effect of pronunciation variation modeling. This analysis shows that although modeling pronunciation variation brings about improvements, deteriorations are also introduced. A strong correlation is found between the number of improvements and deteriorations per rule. This result indicates that it is not straightforward to improve the performance of automatic speech recognition (ASR) by excluding the rules that cause deteriorations, because these rules also produce a considerable number of improvements. Finally, we compare three different criteria for rule selection. This comparison indicates that the *absolute frequency of rule application* (F_{abs}) is the most suitable criterion for rule selection. For the best testing condition, a statistically significant reduction in Word Error Rate (WER) of 1.4% absolute, or 8.2% relative, is found.

1. INTRODUCTION

Most of the studies on data-driven (DD) pronunciation variation modeling show that such an approach can lead to improvements in ASR performance (for an overview of DD methods, see [12]). Showing that WERs can be reduced is often not enough, though. One would like to know exactly how such reductions come about and, possibly, obtain information that can be generalized to different contexts or situations. Unfortunately, such general knowledge is not often presented in the literature on pronunciation variation modeling for ASR [12].

In an attempt to increase our understanding of pronunciation variation modeling and the way in which this can improve ASR performance, we decided to carry out an error analysis aimed at determining how the improvements and deteriorations in ASR performance came about. This error analysis was carried out automatically at word level. So far, error analysis has been used in very few studies (e.g. [8], [9] and [16]). Furthermore, the kind of error analysis employed in these studies has some drawbacks compared to the error analysis used in the current study. For instance, in [9] the error analysis was performed manually, with the consequence that the amount of material that could be analysed was limited. A disadvantage of our previous analyses studies (see [8], [16]) is that they were performed at sentence level.

The main advantage of performing error analysis at word level is that the error analysis results are directly related to the WERs, and that the total number of words is usually much larger than the number of sentences. However, a disadvantage of our word level analysis is that the errors are considered to be independent, whereas it is well known that errors are interdependent.

In the research on modeling of pronunciation variation, rule or variant selection forms a vital part of the research methodology. In this paper, we examine the results of error analysis in order to find criteria that could be used to select rules. In the literature, a number of papers can be found that use different criteria for rule or variant selection, e.g., a maximum likelihood criterion [4], confusability measures [14], confidence measures [17], and entropy [18]. However, in none of these studies are criteria for selecting rules systematically compared. In the present study, three criteria to select rules are compared to each other: a rule selection criterion that emerged from our error analysis and two frequency measures. All three measures were tested on their suitability as criteria for rule selection.

This paper is organized as follows: In Section 2, more details are given on the speech material and CSR that we used. Furthermore, the automatic rule extraction procedure is explained. Subsequently, in Section 3, an initial rule selection is carried out in order to measure recognition performance and to analyze the results. The goal of this error analysis procedure is to find out how exactly recognition performance is affected by modeling pronunciation variation. In Section 4, we compare three criteria on their suitability for rule selection. Finally, in Section 5, we discuss the results and draw some conclusions.

2. METHOD

2.1. Speech material

The speech material used in these experiments is part of the VIOS database, which consists of recordings of an on-line version of OVIS. OVIS is a spoken dialogue system that gives information about train timetables (see [6] and [13]). We selected 99,400 utterances, which were divided into three corpora (see Table 1).

corpus	%	# utterances	# words
training	60	59,640	180,298
test	20	19,880	60,059
error analysis	20	19,880	60,087

Table 1: Statistics of the speech material

2.2. CSR

The continuous speech recognizer (CSR) uses 39 continuous density hidden Markov models (HMMs). For each of the phonemes /l/ and /r/, separate models were trained for post- and prevocalic position. For each of the other 33 phonemes, context-independent models were trained. In addition, one HMM was trained for non-speech sounds and a one-state HMM was employed to model silence. The baseline test and training lexica contain 1288 words and 1465 words, respectively, plus three extra entries: one for noise and two for filled pauses. In the baseline system, for each word, one transcription is present in the lexicon. This so-called 'canonical transcription' was obtained using a Text-to-Speech system (TTS) for Dutch [5] followed by a manual correction. The acoustic models and language models (unigram and bigram) are estimated on the training material. For more details on the CSR, see [10] and [13].

2.3. Automatic rule extraction

The DD rules were extracted from automatic transcriptions of all the utterances in the training corpus. Since our previous research [15] showed that many deletions occur in the VIOS material, and since deletions are more frequent than insertions and substitutions, we have restricted ourselves to studying deletions. The following five steps describe the whole procedure of automatic extraction of the deletion rules:

1. For each word in an utterance, the canonical transcription (T_{can}) is looked up in the baseline lexicon (see Section 2.2).
2. Pronunciation variants are generated by making each phone in T_{can} optional, with the constraint that one phone per syllable must remain present. For example: Suppose T_{can} is '/wIL/' (want), then the following pronunciation variants are generated for this word: /wIL/, /wI/, /wL/, /IL/, /w/, /I/ and /L/.
3. Forced recognition is performed using the baseline phone models and all variants generated in step 2 (including the canonical variant). During forced recognition, the CSR does not choose between all the words in the lexicon, instead, for each word in the utterance, it has to determine which pronunciation variant best matches the acoustic signal. In this way, data-driven transcriptions (T_{dd}) of all the utterances of the training corpus are obtained.
4. A dynamic programming algorithm is used to align T_{can} with T_{dd}. An example of the alignment of T_{can} with T_{dd} is the following:

   T_{can} | I k | w I L | (| = word boundary)
   T_{dd} | - k | w I L | (- = deletion)

5. Using the alignments obtained in step 4, we formulate candidate deletion rules. These rules are defined in the following manner:

   /L F R/$_{can}$ → /L - R/$_{dd}$

   This means that the focus phone F in T_{can} following the phone L (left context) and preceding the phone R (right context) is deleted in T_{dd}. The left and right context can be a phone or a word boundary. It should be noted that this rule formalism is different from the one that is normally adopted in knowledge-based studies. Knowledge-based rules usually have a more general scope, i.e., L and R can be classes of phones, instead of one single phone.

3. INITIAL RULE SELECTION

In order to perform recognition experiments, we performed an initial rule selection. The initial rule selection is necessary as too many rules are automatically generated. For the best test condition, the recognition experiments were repeated on an error analysis corpus and the errors were analyzed. The goal of this error analysis procedure is to find out how exactly recognition performance is affected by modeling pronunciation variation.

3.1 Frequency measures for initial rule selection

After applying the automatic rule extraction procedure to the training corpus, in total 1,392 candidate rules were obtained. These rules together describe the deletions of 6.6% of the total number of 686,909 phones in the training corpus. Examples of rules are given in Appendix 1. Since this number of rules is too large to take all of them into account, we used frequency measures to select rules. Rule frequency can be interpreted in three different ways:

- F_{cond} = the number of times the condition for rule application is met
- F_{abs} = the number of times a rule is applied
- F_{rel} = 100%*F_{abs} / F_{cond} (0 • •F_{rel} • •100%)

The relative frequency F_{rel} is also referred to as rule application likelihood, or rule probability.

Whereas all three frequency criteria have been used for selection in previous research, F_{rel} is probably used most often (see e.g. [1] and [2]). For this reason, we started off by selecting different sets of rules by varying the threshold for F_{rel}. Furthermore, only rules were selected for which F_{abs} is higher than 100. This was done because the automatically obtained DD transcriptions may contain errors due to artefacts of the CSR (e.g. contaminated HMMs). Since it can be expected that transcription errors do not occur systematically, rules that are based on transcription errors are probably not as frequent as rules that are based on genuine deletion processes. For this reason, we expect the errors to be filtered out if the threshold for F_{abs} is set to 100. In addition, we expect that a minimum number of occurrences of 100 is enough to ensure substantial changes in WER and to reliably estimate the probabilities of the pronunciation variants. By excluding the rules for which F_{abs} is smaller than 100, the rule set is reduced to 91 rules, which is 3% of the original size. These 91 rules describe 66% of the deletions that occur in the training material.

3.2 Recognition experiments with the initial sets of rules

By varying the threshold for F_{rel}, seven rule sets are obtained. These threshold values are shown in the first column of Table 2. In order to generate pronunciation variants, we applied the selected rules to the transcriptions in the baseline test lexicon.

By adding the generated variants to the baseline test lexicon, different multiple pronunciation lexica were obtained. Table 2 shows the statistics of the multiple pronunciation lexica. The second column displays the number of rules that were selected (# rules). The third column shows the total number of added variants (# added vars), and column four displays the average number of pronunciation variants per word present in the recognition lexicon (av. # vars/word). Finally, in the last column, the maximum number of pronunciation variants per word is given (max. # vars/word).

F_{rel} >	# rules	# added vars	av. # vars/word	max. # vars/word
50%	7	81	1.1	4
40%	10	322	1.3	8
30%	16	466	1.4	12
20%	25	702	1.5	12
15%	38	993	1.8	12
10%	53	1896	2.5	64
0	91	3528	3.7	128

Table 2: Statistics of the multiple pronunciation lexica

We modeled pronunciation variation at all three levels of the CSR (lexicon, HMMs, and language model), thus obtaining the following three test conditions:

- **T1:** The *lexicon* is expanded by adding pronunciation variants to it, thus creating a multiple pronunciation lexicon. The only difference with the baseline testing condition is that in testing condition **T1** the baseline lexicon is replaced by a multiple pronunciation lexicon.

For the other two testing conditions, an extra step is needed. In this step, automatic transcriptions of the words in the training corpus are obtained. This is accomplished by performing forced recognition with the baseline phone models and the set of variants which have been automatically generated with the selected set of rules.

- **T2:** The *HMMs* are retrained on the basis of the new transcription of the training corpus. The only difference with testing condition **T1** is that in testing condition **T2** the baseline phone models are replaced by the retrained phone models.
- **T3:** A new *language model* is calculated on the basis of the new transcriptions of the training corpus. In the baseline language model all pronunciation variants of the same word are assigned equal prior probabilities. In the new language model, different variants of the same word are assigned their own specific prior probabilities. These prior probabilities are calculated on the basis of the automatic transcriptions of the pronunciation variants in the training corpus. The only difference with testing condition **T2** is that in testing condition **T3** the baseline language model is replaced by the new language model.

The performance of the CSR for these three testing conditions is measured on the test corpus, using the seven different multiple pronunciation lexica which were presented in Table 2. The WER is defined as follows:

$$\mathrm{WER} = \frac{S + D + I}{N} \qquad (1)$$

where S is the number of substitutions, D the number of deletions, I the number of insertions, and N the total number of words. For the baseline CSR we measured a WER of 16.94%.

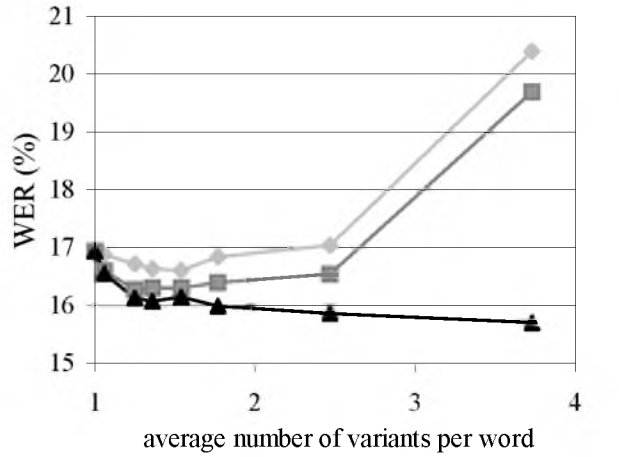

Figure 1: WERs for the test corpus for the three test conditions: **T1** (▫), **T2** (▪), and **T3** (▲)

Fig. 1 shows that if the average number of variants per word is larger than 2.5, expanding the lexicon without using variant-specific prior probabilities (**T1**) leads to a significant (t-test, α=0,05) deterioration in recognition performance. Retraining the HMMs (**T2**) leads to improvements in recognition performance compared to **T1**, but the improvements are small. When prior probabilities are used (**T3**), the WERs are always lower than the WER for the baseline testing condition. In the best testing condition (91 rules, **T3**), a significant reduction in WER is obtained of 1.2% absolute or 7.3% relative.

3.3 Error analysis

In Section 3.2, we showed that the WER can be reduced, provided that variant-specific prior probabilities are used. Since our intention was to go beyond merely reducing the WER and to gain insight into the processes that lead to such reductions, we carried out an error analysis at the word level, which is presented in this section.

For error analysis we used an independent error analysis corpus, which is about the same size as the test corpus (see Table 1). The WER for the baseline system on the error analysis corpus was 16.47%. The WER on the error analysis corpus measured for the best condition (91 rules, **T3**) was 15.44%. Subsequently, the recognition results were compared to the recognition results of the baseline system. The two resulting word strings were aligned and the differences (at word level) were categorized.

If a recognized word is a pronunciation variant, then the change is attributed to the rule(s) that generated this variant. In

other cases there is a change, but the word is classified as 'no-variant'. These latter changes cannot be attributed directly to a change in the lexicon; they are the result of other changes, e.g. changes in the HMMs and the language models, or changes in other words in the recognized utterance.

	Total	Variant	No-variant
Improvement	2219	489 (22%)	1730 (78%)
Deterioration	1613	301 (19%)	1312 (81%)
Net result	606	188 (31%)	418 (69%)

Table 3: Number of changes per category

Error analysis (for the 60,087 words in the error analysis corpus) showed that 3832 changes had occurred: 2219 improvements and 1613 deteriorations. The numbers per category are presented in Table 3. It can be observed that the majority of the changes fall in the category 'no-variant', and that for both categories there are more improvements than deteriorations. Consequently, the net result is positive: 606 improvements which cause a lowering of the WER from 16.47% (for the baseline) to 15.44%. The changes for both categories were studied in more detail. We will only present some results for the 'variant' category here, i.e. the 489 improvements and 301 deteriorations in column 3 of Table 3 (for more results, see [6]).

A change in the 'variant' category is by definition the result of a recognized variant. If a variant is generated by N rules, then the counters of those N rules are raised by 1/N. In this way, we determined for each of the 91 rules how many improvements and deteriorations the rule caused. The correlation between the number of improvements and the number of deteriorations turned out to be very high (Pearson's r [3] is 0.97), indicating that a rule that causes many improvements also causes many deteriorations.

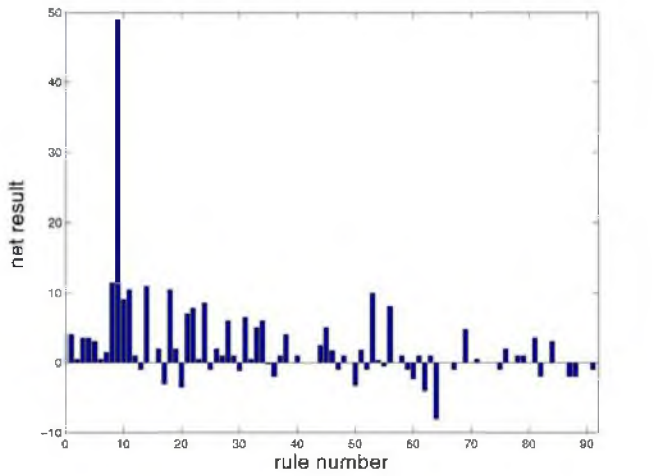

Figure 2: 'Net result' for each of the 91 rules

For each rule we also determine the *'net result'*, which is defined as the number of improvements minus the number of deteriorations. Fig. 2 shows that for most rules the *'net result'* is positive. Furthermore, it can be observed in Fig. 2 that not all rules contribute equally, e.g. rule 9 (condition /@n|/) has a large positive net effect, while rule 64 (/na:R/) has a large negative net effect.

4. COMPARING RULE SELECTION CRITERIA

The results of our error analysis indicate that the number of improvements and deteriorations per rule are highly correlated. Consequently, it is not straightforward to improve ASR performance by excluding the rules that cause many deteriorations, because these rules also produce a considerable number of improvements. The question that remains is which criteria are most suitable for rule selection. For this reason, we investigated the adequacy of three rule selection criteria.

There are various motivations for performing rule/variant selection. First of all, the addition of pronunciation variants to the lexicon increases confusability, especially if the lexicon is large. This means that the more variants are included in the lexicon, the more lexical confusability increases due the addition of variants. The large increase in confusability is probably the reason why adding many variants to the lexicon usually leads to small improvements or even to deteriorations. By making an appropriate selection of the pronunciation variants, the balance between solving and introducing errors could become more positive. A second reason for constraining the number of variants is to limit decoding time, since decoding time is directly related to the size of the lexicon. Third, in data-driven approaches, the data-derived variants are usually selected or filtered, as the variants might be based on transcription errors (caused by artefacts of the CSR) instead of being based on genuine pronunciation variation.

4.1 Three rule selection criteria

We investigated three measures that could be used to select rules; the first measure emerges from the error analysis in Section 3, whereas the other two measures concern the application frequency of the rules:

1. The *'net result'*
   In Section 3.3, the *'net result'* is defined as the number of improvements minus the number of deteriorations for each rule (see also Fig. 2). Rules are selected on the basis of their *'net result'*, which means that the rules with the highest *'net result'* are selected first. The following values of *'net result'* were used as thresholds: 45, 10, 5, 1, 0, -1.

2. F_{abs}
   Rules are selected based on F_{abs}, which means that rules with the highest F_{abs} are selected first. The following threshold values were used for F_{abs}: 5000, 500, 400, 300, 200, 140, 100.

3. F_{rel}
   Rules are selected based on F_{rel}. Since we already used F_{rel} as a selection criterion, we did not repeat the recognition experiments, and simply used the results reported in section 3.1.

4.2 Correlations with WER reduction

Subsequently, for each set of selected rules, the WER reduction is measured (on the test corpus). Table 4 shows the correlations (Pearson's r) between the WER reduction and the value of each selection criterion for each rule set.

'net result'	F_{abs}	F_{rel}
0.86	0.93	-0.83

Table 4: Correlations (Pearson's r) between WER reduction and the various rule selection criteria

These data suggest that F_{abs} and *'net result'* are better criteria for selecting rules than F_{rel}. Furthermore, it can be seen that a reduction in WER is associated with a lower F_{rel} (Pearson's r = -0.83). This is contrary to expectation, as one would expect the reduction in WER to be larger if F_{rel} of the rules in the set is higher. A possible explanation for this result is that of the two factors that play a role - namely F_{rel} and F_{abs}, - F_{abs} is more important. This point can be illustrated by the following example. A specific value of F_{rel} could be the result of two completely different situations. For instance, an F_{rel} value of 50% could be obtained in the following two situations:

1. F_{abs}= 1 and F_{cond}= 2,
2. F_{abs}= 10,000 and F_{cond}= 20,000.

It is easy to imagine that in relation to the total amount of material, situation 2 is bound to have a much greater effect on recognition performance than situation 1. While this difference clearly emerges from F_{abs}, it is completely blotted out in F_{rel}, which in turn explains why F_{rel} does not appear to be a good predictor of the reduction in WER.

4.3 Comparative recognition experiments

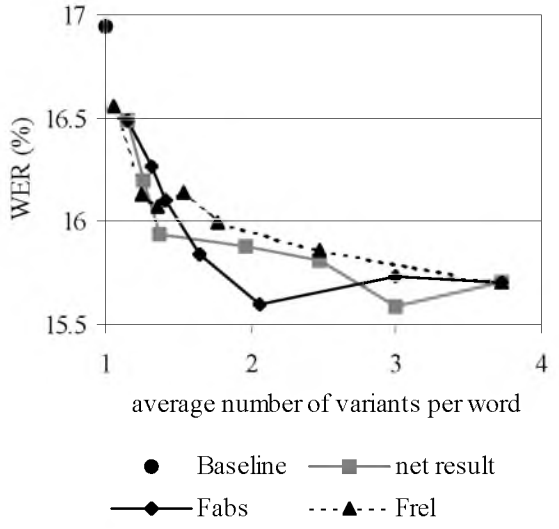

Figure 3: WERs for rule sets selected with the three different selection criteria

Fig. 3 shows the WERs on the test corpus (**T3**) for the rule sets that are selected on the basis of the three selection criteria. In Fig. 3, differences of more than 0.35% WER are significant. The lowest WER is found for $F_{abs} > 200$ (an average of about 2 variants/word); a statistically significant reduction in WER of 1.4% absolute, or 8.2% relative, is found. Consequently, the results of these experiments and the correlation results (Section 4.2) show that F_{abs} seems to be a better criterion for selecting rules than F_{rel}.

5. GENERAL DISCUSSION AND CONCLUSIONS

The recognition experiments demonstrated that the DD rules can be used effectively to improve recognition performance. Furthermore, our results shows that it is crucial to use variant-specific probabilities in the language model (**T3**) in order to ensure improvements in recognition performance.

The error analysis presented in this paper shows that many changes occur due to modeling pronunciation variation, but the net result is a small improvement in WER. Error analysis also revealed that the number of improvements and the number of deteriorations that the various rules cause are strongly related. This result indicates that it is not straightforward to improve ASR performance by excluding the rules that cause many deteriorations, because these rules also produce a considerable number of improvements.

As to the choice of which of the three selection criterion is most optimal for rule selection, our results showed that F_{abs} and *'net result'* are better criteria for selecting rules than F_{rel}. The question that remains is which of the two measures F_{abs} and *'net result'* is the better criterion. Let us compare the results of the two criteria. First of all, the correlation with the reduction in WER is higher for F_{abs} (0.93) than for *'net result'* (0.86). Second, *'net result'* clearly has the disadvantage that it can only be used after performing a recognition experiment and carrying out an error analysis. F_{abs}, on the other hand, can be determined directly from the transcriptions used for automatic rule extraction. Third, for F_{abs} the optimal WER is obtained using an average of two variants/word in the lexicon, whereas three variants/word are needed when *'net result'* is used as a selection criterion (see Fig. 3). Since decoding time is correlated with the number of entries in the lexicon, this means that decoding time is shorter when the optimal rule set is obtained by selecting the rules based on F_{abs} than on the basis of *'net result'*. For all of these reasons, of the three criteria compared in the present study, F_{abs} seems to be the most suitable one for rule selection.

ACKNOWLEDGEMENTS

We would like to thank Loe Boves, Mirjam Wester, Febe de Wet and two anonymous reviewers for their helpful comments on earlier versions of this paper.

REFERENCES

[1] Amdal, I., Korkmazskiy, F., Surendran, A., "Data-driven pronunciation modelling for non-native speakers using association strength between phones", *Proc. of the ISCA workshop ASR2000, Paris, France,* pp. 85-90, 2000.

[2] Cremelie, N., Martens, J.-P., "In search of better pronunciation models for speech recognition", *Speech Communication 29*, 115-136, 1999.
[3] Ferguson, G.A., *Statistical Analysis in Psychology and Education*, McGraw-Hill, Singapore, 1981.
[4] Holter, T., Svendsen, T., "Maximum likelihood modelling of pronunciation variation", *Speech Communication, Vol. 29,* pp. 177-191, 1999.
[5] Kerkhoff, J., Rietveld, T., "Prosody in Niros with Fonpars and Alfeios", *Proc. of the Dept. of Language & Speech*, University of Nijmegen, Vol.18, pp. 107-119, 1994.
[6] Kessens, J.M., *Making a difference: On automatic transcription and modeling of Dutch pronunciation variation for automatic speech recognitio*n, Ph.D. thesis, University of Nijmegen, The Netherlands, 2002. http://webdoc.ubn.kun.nl/mono/k/kessens_j/makia_di.pdf
[7] Kessens, J.M., Strik, H., Cucchiarini, C. "A bottom-up method for obtaining information about pronunciation variation", *Proc. ICSLP, Beijing*, pp. 274-277, 2000.
[8] Kessens, J.M., Wester, M., and Strik, H., "Improving the Performance of a Dutch CSR by Modeling Within-word and Cross-word Pronunciation", *Speech Communication., Vol. 29, ,* pp. 193-207, 1999.
[9] Ravishankar, M., Eskenazi, M., "Automatic generation of context-dependent pronunciations", *Proc. Eurospeech*, Rhodes, Greece, Vol. 5, pp. 467 – 2470, 1997.
[10] Steinbiss, V., Ney, H., Haeb-Umbach, R., Tran, B.-H., Essen, U., Kneser, R., Oerder, M., Meier, H.-G., Aubert, X., Dugast, C., Geller, D., "The Philips Research System for Large-Vocabulary Continuous-Speech Recognition.", *Proc. of EUROSPEECH '93*, Berlin, pp. 2125-2128, 1993.
[11] Strik, H., "Pronunciation adaptation at the lexical level", *Proc. of the ITRW Adaptation Methods for Speech Recognition*, Sophia-Antopolis, France, pp. 123-130, 2001.
[12] Strik, H., Cucchiarini, C. "Modeling pronunciation variation for ASR: a survey of the literature", *Speech Communication, Vol. 29,* pp. 225-246, 1999.
[13] Strik, H., Russel, A.J.M., van den Heuvel, H. Cucchiarini, C. and Boves, L. "A spoken dialog system for the Dutch public transport information service", *Int. Journal of Speech Technology, Vol. 2, No. 2*, pp. 119-129, 1997.
[14] Wester, M., Fosler-Lussier, E. A comparison of data-derived and knowledge-based modeling of pronunciation variation. *Proc. of ICSLP*, Beijing, China, Vol. 4, pp. 270-273, 2000.
[15] Wester, M., Kessens, J. M. and Strik, H., "Two automatic approaches for analyzing the frequency of connected speech processes in Dutch", *Proc. of ICSLP*, Sydney, Australia, 3351-3356, 1998.
[16] Wester, M., Kessens, J.M. and Strik, H., Pronunciation variation in ASR: Which variation to model?, *Proc. of ICSLP*, Beijing, China, Vol. 4, pp. 488-49, 2000.
[17] Wiliams, G. *Knowing What You Don't Know: Roles for Confidence Measures in Automatic Speech Recognition*, Ph.D. thesis, Department of Computer Sciences, University of Sheffield, Sheffield, United Kingdom, 1999.
[18] Yang, Q., Martens, J.-P., "Data-driven lexical modeling of pronunciation variations for ASR", *Proc. ICSLP*, Beijing, China, Vol. 1, pp. 417-420, 2000.

Appendix 1

Examples of DD rules, ordered according to descending F_{abs}. Rules are given for which $F_{abs} > 300$ and $F_{rel} \geq 10\%$. The rules are formulated as described in section 2.3, step 5. For each rule, F_{abs}, F_{rel} and *'net result'* are given in the corresponding columns.

deletion rule	F_{abs}	F_{rel}	*'net result'*
/@n|/ → /@-|/	5339	43%	49
/@Rd/ → /@-d/	2031	48%	11.5
/a:R|/→ /a:-|/	1089	10%	10
/st@/ → /s-@/	777	29%	-3
/@Rt/ → /@-t/	638	57%	3.5
/v@r/ → /v-r/	555	28%	10.5
/@nt/ → /@-t/	528	25%	7.8
/xt|/ → /x-|/	498	13%	0
/|ni/ → /|-i/	442	18%	0.5
/nd@/ → /n-@/	417	34%	10.5
/it|/ → /i-|/	416	18%	6.5
/d@r/ → /d-r/	333	30%	2
/d@|/ → /d-|/	317	19%	1
/At|/ → /A-|/	310	15%	1
/En|/ → /E-|/	310	13%	5

Table 5: Examples of deletion rules