# Regional pronunciation variants for automatic segmentation


**Nicole Beringer***, **Marcia Neff***

*Institut fuer Phonetik und Sprachliche Kommunikation
Schellingstr. 3, D-80799 Muenchen, Germany
{beringer,maneff}@phonetik.uni-muenchen.de



**Abstract**

The goal of this paper is to create an extended rule corpus with approximately 2300 phonetic rules which model segmental variation of regional variants of German. The phonetic rules express at a broad-phonetic level phenomena of phonetic reduction in German that occurs within words and across word boundaries. In order to get an improvement in automatic segmentation of regional speech variants, these rules are clustered and implemented depending on regional specification in the Munich Automatic Segmentation System.


## 1. Introduction

A fundamental property of speech is high variability, which means that no two utterances of the same word are ever produced exactly the same. This phenomenon can be observed among different speakers, as well as in repetitions of words by a single speaker. There are similarities which can be detected and summarized in pronunciation rules. These pronunciation rules can then be related to different dialects, sociolects and styles of speech, such as read or spontaneous speech.

Human beings can assign pronunciation variants to relevant words. To imitate this function in ASR, pronunciation rules must be included in the recognition process. For automatic segmentation, we developed the Munich Automatic Segmentation System (MAUS)(Kipp, 1999)), which is an HMM-based system for the automatic segmentation of spontaneous speech dealing with pronunciation variants.

The rule corpus presented here has been designed to improve MAUS. MAUS (Kipp, 1999; Beringer-Schiel, 1999) is based on HTK (htk, 1996) and statistically weighted rewrite pronunciation rules to automatically segment large speech corpora. This system is trained on German manually segmented data (approx. 5 hours of speech).

The aim of the work of this paper is to automatically segment various corpora according to phoneme. The corpora are available from the BAS (Schiel, 1999) e.g. RVG I (Regional variants of German (Burger-Schiel, 1998)) The segmentation thus creates a reliable segmentation and labeling for applications in speech processing (such as ASR and speech synthesis (e.g. PSOLA)) and phonetic research.

In previous research it was shown that MAUS can substitute manually segmented data for the VERBMOBIL corpus (Kipp et al., 1997) because it is less time-consuming and reaches a correspondence with human labelers of approximately 93.6 percent via a transformed forced alignment.

Segmenting other databases using MAUS can be misleading because the pronunciation rules used in the transcription are trained basically on the VERBMOBIL corpus.

This leads to unknown regionally specific variants which are in turn not considered in the segmentation. Thus, our aim is to transform this system to automatically segment other databases as well.

For this reason we developed an extended rule corpus of phonetic rules clustered by regions (Burger-Schiel, 1998).

This corpus consists of approximately 1500 basic phonetic rules presented in (Wesenick, 1996) and approximately 200 regionally specific rules for each of the 9 regions additionally (defined in (Burger-Schiel, 1998)).

These 1800 rules describe phonetic variation as seen in the RVG I corpus.

The following section describes typical phonetic processes in German regional variants. It includes an inventory of used symbols as well as the syntax of regional variation rules.

Section 3 gives an outline of typical phonetic processes in Regional Variants of German.

## 2. The representation of segmental variation in Regional Variants of German in phonetic rules

In dealing with spoken utterances in automatic speech recognition, the following problem always arises: The variability of the production of the same utterance by different speakers, as well as the repetition of an utterance by a single speaker (inter- vs. intra-speaker variability). Among other factors, the high variability of speech depends on the speech rate, speaking style and complexity of the semantic contents of utterances. Depending on the situation, utterances can be clear and precisely articulated e.g. in formal situations, or they can tend towards lowest possible articulatory effort (hyper- and hypospeech). In order to surmount these problematic phenomena in ASR, we have already developed a system which takes the canonic form found in a pronunciation dictionary and modifies it by statistical pronunciation rules. Therefore, we require sufficient manually segmented data which we do not always have available. To substitute for these kind of rules, phonetic processes that lead to the variability using phonemic or broad-phonetic information is taken into account. This procedure led us to implement the automatic generation of German pronunciation Variants by (Wesenick, 1996), which was extended recently for regional variants of the RVG I corpus and is described in this paper.

### 2.1. Inventory of used symbols

The phonetic symbols that are used in the regional pronunciation rules are taken from the SAM-Phonetic Alphabet for German (as reprinted in Pompino-Marschall (ed.)Pompino-Marschall, 1992)). The symbols stand for the phonemes of German. We clustered these symbols in the following way:

| phone category | phonetic symbols |
|---|---|
| unreduced vowels | a, a:, e:, I, i:, O, o:, U, u:, E, E:, 9, 2:, Y, y: |
| reduced vowels | @, 6 |
| diphthongs | aI, aU, OY |
| plosives | p, b, t, d, k, g, Q |
| fricatives | f, v, s, z, S, Z, C, j, x, h |
| nasals | m, n, N |
| liquids | l, r |

Table 1: complete list of phonetic symbols used in the rules (and the corresponding HMMs)

Aspiration and the burst of plosives are not marked. The phonological voiced-voiceless distinction refers to the difference in energy, namely the distinction fortis-lenis. For the evaluation and clustering of rules we define the following special symbols which we refer to in section 4:

| phone category | special phonetic symbols |
|---|---|
| vowels | !v (unreduced and reduced vowels and diphthongs) |
| consonants | !K |
| voiceless plosives | !P |
| voiced plosives | !B |
| voiceless fricatives | !F |
| voiced fricatives | !V |
| nasals | !N |
| liquids | !L |

Table 2: complete list of special phonetic symbols used in the rules

### 2.2. Syntax of the rules

A rule r(i), i= 0 ... N-1 of the rule corpus consists of a right and a left part which are separated by ">". Both parts consist of a string of symbols in SAM-PA. The canonic form is to the left of the ">" which is modified to the pronunciation variant appearing at the right of the ">". Note that each part has a left and a right context which is constant.

For example:

h-U-n> h- -n

U can be deleted between h and n (Region A).

d-a-s> d-E-s

a can be replaced by E between d and s (Region B).

s-t-f> s- -f

t can be deleted between s and f (Region C).

e-p-a> e-b-a

p can become voiced between vowels (Region D).

n-O-m> n-U-m

O can be replaced by U between Nasals (Region E).

t-s-t> t-S-t

s can be replaced by S between Plosives (Region F).

h-aI-m> h-O-m

Replacement of the Diphthong aI with O between h and m (Region G).

m-y:-n> m-Y-n

Replacement of y: by Y between two Nasals (Region H).

b-a:-n> b-o:-n

Replacement of a: by o: between b and n (Region I).

These rules correspond to the regions A - I described in the section below.

## 3. Description of typical phonetic processes in Regional Variants of German

Before describing typical phonetic processes in regional variants of German, we will first give a short introduction in the table below to the clustering method used to obtain the regions. The clustering algorithm used in RVG I is based on the dialectal subdivision by König (1978)(König, 1978), deleting all unspoken dialects. All regional variants with a small number of inhabitants or very small dialect regions are assigned to adjacent regions, while larger regions were divided into sub-classes. This division results in quite a stable number of regional speakers. Dialectal borders are aligned to county and state borders. For more details refer to (Burger-Schiel, 1998).

| RVG-cluster | name of dialect region |
|---|---|
| A | Lower Franconian (Niederfränkisch) |
| B | West Low German (Westniederdeutsch) |
| C | East Low German (Ostniederdeutsch) |
| D | West Middle German (Westmitteldeutsch) |
| E | East Middle German (Ostmitteldeutsch) |
| F | Alemannic (Alemannisch) |
| G | East Franconian (Ostfränkisch) |
| H | South Franconian (Sdfränkisch) |
| I | Bavarian-Austrian (Bairisch-Österreichisch) |

Table 3: regional classification of variants

For our purpose we will not consider the sub-cluster of the dialect regions described in (Burger-Schiel, 1998). To get an idea of the geographic location of the single regions see the map below.

The following subsections describe some typical occurrences of phonetic processes in German according to the corresponding region cluster.
Please note that we have just given a few typical examples of the phenomena found in the several regions to give an idea of typical phenomena for the corresponding region.

### 3.1. Assimilation

A frequently observed phenomenon in spoken language is the assimilation of one segment to an adjacent one within at least one parameter. We distinguish between assimilation of place of articulation, assimilation of manner of articulation and assimilation of the voicing parameter (which refers here also to centralization of vowels). Assimilations

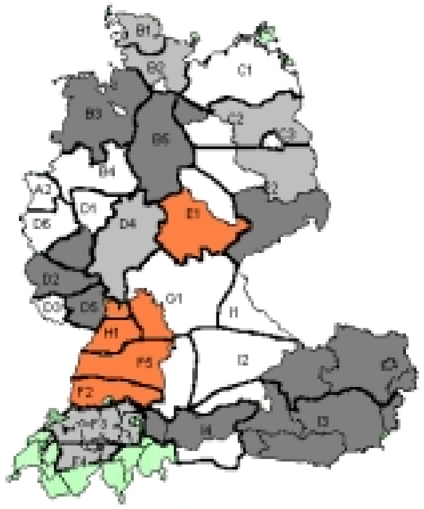

Figure 1: Map of German speaking regions belonging to the classification in Table 3

can occur within morphemes or syllables as well as across them. They also can be total or partial.

For a short outline of assimilations found in the several dialects see the section below.

#### 3.1.1. Assimilations in Lower Franconian (region A)

**Assimilation of Place of Articulation**

!P-!P,!F-!V> !P-!V-!V
< Inspektion> (inspection)
InspEk**ts**jo:n -> InspEk**z**jo:n

**Assimilation of Manner of Articulation**

!N-!L,I-!F> !N-!F-!F
< nämlich> (namely)
nE:m**lI**C -> nE:m**C**C

**Assimilation of the Voicing Parameter**

!N-!B-@> !N-!P-@
< angeführt> (leaded)
aN**g**@fy:6t -> aN**k**@fy:6t

#### 3.1.2. Assimilations in West Low German (region B)

**Assimilation of Manner of Articulation**

!N-!P-!L¿!N-!B-!L
<impliziert> (implies)
Im**p**lItsi:6t - > Im**b**lItsi:6t

**Assimilation of the Voicing Parameter**

!L-e:-!P¿!L-@-!P
<lebt> (lives)
l**e:**pt -> l**@**pt

#### 3.1.3. Assimilations in East Low German (region C)

**Assimilation of Manner of Articulation**

!P-!F-i:¿!P-!B-i:
<Ziel> (target)
tsi:l ->t**d**i:l

**Assimilation of the Voicing Parameter**

i:-!B,@-!N¿i:-!P-!N
<biegen> (bend)
bi:**g**@n -> bi:**k**@n

#### 3.1.4. Assimilations in West Middle German (region D)

**Assimilation of the Voicing Parameter**

!N-!B,@-!N¿!N-!P-!N
<sangen (sang)
zaN**g**@n -> zaN**k**@n
I-!B,@-!N¿I-!P-!N
<Widder> (ram)
vI**d**@r -> vI**t**@r

#### 3.1.5. Assimilations in East Middle German (region E)

**Assimilation of Manner of Articulation**

!F-!B-@¿!F-!V-@
<aufbessern> (to improve)
aUf**b**@s6n -> aUf**v**@s6n

**Assimilation of the Voicing Parameter**

!N-!B-@¿!N-!P-@
<anbelangt> (arrived)
an**b**@laNt -> an**p**@laNt

#### 3.1.6. Assimilations in Alemannic (region F)

I-!P-!F¿I-!F-!F
<Gipfel> (summit)
gI**p**f@l -> gI**f**f@l

**Assimilation of the Voicing Parameter**

e-a:-!P¿e-a-!P
<Theater> (theatre)
te**a:**t6 -> te**a**t6

#### 3.1.7. Assimilations in East Franconian (region G)

**Assimilation of Place of Articulation**

OY-!P-!P¿OY-!F-!P
<Häupter>(heads)
hOY**p**t6 -> hOY**f**t6

**Assimilation of Manner of Articulation**

aI-!B,@-!N¿aI-!P-!N
<bleiben> (stay)
blaI**b**@n -> blaI**p**@n

#### 3.1.8. Assimilations in South Franconian (region H)

**Assimilation of Place of Articulation**

!V-aI-!L> !V-i:-!L
< weil> (because)
v**aI**l -> v**i:**l

**Assimilation of Manner of Articulation**

i:-!B-@> i:-!V-@
< sieben> (seven)
zi:**b**@n -> zi:**v**@n

**Assimilation of the Voicing Parameter**

u:-!B-@> u:-!P-@
< trugen> (carried)
tru:**g**@n -> tru:**k**@n

### 3.1.9. Assimilations in Bavarian-Austrian (region I)

**Assimilation of Place of Articulation**
I-!N-!F¿I-!P-!F
<ins> (in the)
I**n**s -> I**t**s
**Assimilation of Manner of Articulation**
I-!F-!F¿I-!P-!F
<möglichst> (as well)
m2:glI**C**st -> m2:glI**k**st
**Assimilation of the Voicing Parameter**
-!B,a-!F¿-!P-!F
<das> (the)
**da**s -> **t**s

## 3.2. Lenition

Lenition is defined as the reduction of articulatory effort due to a higher speech rate and inprecise articulation. To get an idea of how lenition sounds, we suggest the example of voiceless consonants becoming voiced.

### 3.2.1. Lenition in Lower Franconian (region A)

!L-!P-aU> !L-!B-aU
< Turteltaube> (turtle dove)
tUrt@l**t**aUb@ -> dUrd@l**d**aUb

### 3.2.2. Lenition in West Low German (region B)

OY-!P-!F¿OY-!F-!F
<deutsch> (German)
dOY**t**S -> dOY**S**S

### 3.2.3. Lenition in East Low German (region C)

a-!P-I¿a-!B-I
<Gattin> (spouse, wife)
ga**t**In -> ga**d**In

### 3.2.4. Lenition in West Middle German (region D)

!N-!P-!F¿!N-!Ba-!F
<impfen> (vaccinate)
Im**p**f@n -> Im**ba**f@n

### 3.2.5. Lenition in East Middle German (region E)

U-!P-!P¿U-!B-!P
<Subklassifizierung > (sub-classification)
zU**p**klasIfItsi:6UN -> zU**b**klasIfItsi:6UN

### 3.2.6. Lenition in Alemannic (region F)

!N-!P-U¿!N-!B-U
<unkundig> (ignorant)
Un**k**UndIC -> Un**g**UndIC

### 3.2.7. Lenition in East Franconian (region G)

a-!F-O¿a-!V-O
<davon> (there of)
da**f**On -> da**v**On

### 3.2.8. Lenition in South Franconian (region H)

E-!P-@> E-!B-@
< Ecke> (corner)
E**k**@ -> E**g**@

@-!P-U> @-!B-U
< Getummel> (turmoil)
g@**t**Um@l -> g@**d**Um@l

## 3.3. Epenthesis

Epenthesis is when segments are inserted into a sequence of phonemes. This either occurs with consonants or with vowels.

### 3.3.1. Epenthesis in Lower Franconian (region A)

**vowel epenthesis:**
y:-!B- 6> y:-!V,e:- 6
< über> (over)
y:**b** 6 -> y:**ve:** 6

### 3.3.2. Epenthesis in West Low German (region B)

**vowel insertion:**
!B-e:-!N¿!B-e:-@-!N
<dem> (the)
d**e:**m -> d**e:@**m
**consonant epenthesis:**
a-!N-!F¿a-!N!B-!F
<anfangen> (begin)
a**n**faN@n -> a**mp**faN@n

### 3.3.3. Epenthesis in Alemannic (region F)

**vowel insertion:**
!B-aI-!B¿!B-Y,U-!B
<beiden> (both)
b**aI**d@n -> b**YU**d@n

### 3.3.4. Epenthesis in East Franconian (region G)

**vowel insertion:**
!F-y:-!F¿!F-Y,a-!F
<Schüchen> (small shoes)
S**y:**C@n -> S**Ya**C@n

**consonant epenthesis:**
I-!P-i:¿I-!L,!B-i:
<Philippinen> (Philippines)
fIlI**p**i:n@n -> fIlI**p**i:n@n

### 3.3.5. Epenthesis in South Franconian (region H)

**vowel insertion:**

!B-e:-!B> !B-E:,a-!B
< beben> (shake)
b**e:**b@n -> b**E:a**b@n

**consonant epenthesis:**
a-!N-@> a-!N,!B-@
< Wange> (cheek)
va**N**@ -> va**Nk**@

#### 3.3.6. Epenthesis in Bavarian-Austrian (region I)

**vowel insertion:**
!F-6-I¿!F-a-I
<Glycerit> (Glyzerit)
glYts**6**It -> glYts**a**It

### 3.4. Elision

In spoken language we often observe that parts of a word or segment are not realized although it is produced in the citation form. We refer to such deletions as elision. In German e.g. the elision of /@/ (schwa) or of the apical plosive /t/ in a word final syllable can often be observed.

#### 3.4.1. Elision in Lower Franconian (region A)

!V-E,!P-!F> !V-e:-!F
< jetzt> (now)
j**Et**st -> j**e:**st

#### 3.4.2. Elision in West Low German (region B)

-!P-aI¿–aI
<kein> (nothing)
**k**aIn -> **aI**n

#### 3.4.3. Elision in East Low German (region C)

!N-!P-E¿!N–E
<entdecken> (discover)
En**t**Ek@n -> En**E**k@n

#### 3.4.4. Elision in West Middle German (region D)

!F-@-!N¿!F–!N
<helfen> (help)
hElf**@**n -> hElf**n**

#### 3.4.5. Elision in East Middle German (region E)

!P-@-!N¿!P–!N
<denken> (think)
dEnk**@**n -> dEnk**n**

#### 3.4.6. Elision in Alemannic (region F)

@-!N,!B,@-!P¿@–!P
<sendet> (sent)
s@nd@t -$Y$ >s@t

#### 3.4.7. Elision in East Franconian (region G)

!B-@,!N-!P¿!B–!P
<Abend> (evening)
ab**@n**t -> ab**t**

#### 3.4.8. Elision in South Franconian (region H)

**/@/-Elision:**
I-!P,@-!N> I-!P-!N
< bitten> (request)
bI**t@**n -> bI**t**n

aI-!B,@-!N> aI-!P-!N
< bleiben> (stay)
blaI**b@**n -> blaI**p**n

!N-!F,@-!N> !N-!F-!N
< impfen> (vaccinate)
Im**f@**n -> Im**f**n

!N-!B,@-!F> !N-!P-!F
< angefangen> (started)
aN**g@**faN@n -> aN**k**faN@n

!N-!B,e:-!N> !N-!P-!N
< indem> (in that)
In**de:**m -> In**t**m

#### 3.4.9. Elision in Bavarian-Austrian (region I)

aI-@-!N¿aI–!N
<weihen> (dedicate)
vaI**@**n -> vaI**n**
aU-!F-!F¿aU–!F
<auffangen> (catch)
aU**f**faN@n -> aU**f**aN@n

## 4. Regional Variants - an Improvement in Automatic Segmentation

Having shown some typical phonetic processes in the diverse clusters of regional variants, we will now briefly outline how we have successfully included the rules in the MAUS-System. Further, we will give some preliminary results of the improvement in automatic segmentation with the inclusion of regional speech variants.

### 4.1. Integration of pronunciation variants in the Munich Automatic Segmentation System (MAUS)

The existing System works as follows:

The Input consists of the speech signal and its related orthographic representation. The Output is a fully automatically produced segmentation on the phonemic level.

The technical implementation includes the following components:

- hybrid approach consisting of statistic (HMM) and rule based components (HMM)
- possible pronunciation variants of German are taken into

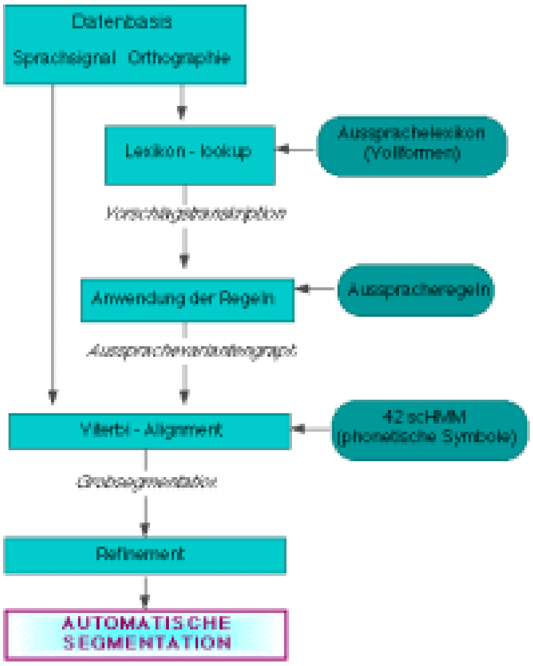


Figure 2: Illustration of the MAUS-System

account.
- applicable to read speech as well as spontaneous speech (in cooperation with the Technical University Munich).

As illustrated in Figure 2, the process includes the following steps:

LEXICON-LOOKUP:
a phonetic standard transcription is related to the given orthography using a phonetic lexicon that contains all inflected forms.
GENERATION OF PRONUNCIATION VARIANTS (refer to “Anwendung der Regeln”):
the set of approx. 1500 phonetic pronunciation rules of German (PHONRUL) are applied to the standard transciption of the concerned utterance, yielding a directed acycylic graph. This graph represents a multitude of hypothetical pronunciation variants of the utterance that is to be analyzed.
VITERBI-ALIGNMENT:
The incoming speech signal is segmented and transcribed using 42 semicontinuous HMMs that correspond to the 42 SAM-PA symbols of the phonemes of German. The result is a rough segmentation.
REFINEMENT:
The rough segmentation is refined and improved with time-domain methods. The result is the automatic segmentation of the speech signal.

We included the regional pronunciation rules in step 2 (Generation of Pronunciation Variants) by attaching them to the existing general phonetic pronunciation rules of German (Wesenick, 1996). It resulted in nine subsets of pronunciation rules.

### 4.2. Preliminary Results

We have successfully included the rules in the MAUS-System (as described above), and have obtained a resulting improvement in automatic segmentation of regional speech variants of approximately 19 percent as compared to the standard pronunciation variants.

## 5. Conclusion

This work has tested the hypothesis developed in (Beringer-Schiel, 1999) that we may need a more detailed set of general pronunciation rules to obtain better results. To obtain better results we are now testing the pronunciation rules after applying statistical weighting followed by a pruning of the rules to the most relevant ones. As with all statistically based methods, it is to be expected that the results will improve proportionally to the amount of available data.

## 6. References

Beringer, N.; Schiel, F. ,1999 Independent Automatic Segmentation of Speech by Pronunciation Modeling. *Proceedings of the ICPhS 1999, San Francisco* : pp.1653-1656

Burger, S.; Schiel, F. ,1998 RVG 1 - A Database for Regional Variants of Conte mporary German. *Proceedings of the FIRST INTERNATIONAL CONFERENCE ON LANGUAGE RESOURCES 1998, Granada, Spain*

The HTK Book Cambridge University Press, 2nd edition. 1996

Kipp, A. , 1999. Automatische Segmentierung und Etikettierung von Spontansprache. Shaker Verlag Aachen 1999.

Kipp, A., Wesenick, B., Schiel, F. , 1997. Pronunciation Modeling Applied to Automatic Segmentation of Spontaneous Speech. *Proceedings of the EUROSPEECH 1997, Rhodos, Greece,* pp.1023-1026.

König, W. , 1978. *Dtv-Atlas zur deutschen Sprache* München: Dtv-Verlag.

Pompino-Marschall (ed.) , 1992. *PHONDAT: Verbundvorhaben zum Aufbau einer Sprachsignaldatenbank für gesprochenes Deutsch* Forschungsbericht des IPSK (FIPKM) 30.99-128, 1992.

Schiel, Florian, 1999. Speech and Speech-Related Resources at BAS. *Proceedings of the FIRST INTERNATIONAL CONFERENCE ON LANGUAGE RESOURCES 1998, Granada, Spain* pp. 343-349. Cambridge University Press, 2nd edition. pp. 343-349.

Wesenick, M.-B., 1996. Automatic Generation of German Pronunciation Variants. *Proceedings of the ICSLP 1996. Philadelphia* pp.125-128