Generation of pronunciation rule sets for automatic segmentation of American English and Japanese

Nicole Beringer[1], *Tsuyoshi Ito*[1], *Marcia Neff*[1]

[1] Institut für Phonetik und Sprachliche Kommunikation, Schellingstr. 3, 80799 München, Germany

ABSTRACT

The goal of this paper is to create an extended rule corpus with approximately 2300 phonetic rules which model segmental variation on a three language task. The phonetic rules express at a broad phonetic level phenomena of phonetic reduction in German, English and Japanese that occur 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 language specification in the Munich Automatic Segmentation System.

1. INTRODUCTION

A fundamental property of speech is its 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 etc.

Human beings can assign pronunciation variants to the relevant word. To imitate this function in ASR and automatic segmentation, knowledge about pronunciation must be included in the process.

We already showed in ([5]),([7]) using the VERBMOBIL corpus and ([8]) using the RVG I corpus (Regional variants of German ([4])) that it is possible to get a (slight) improvement in ASR using pronunciation variants in regionally unspecific domains and to get even comparable results to manually segmented data using pronunciation variants in automatic segmentation. In the latter we reached a correspondence with human labelers of approximately 93.6% by using the MAUS technique (([1])).

In this work we want to find out whether phoneme based variants influence the segmentation and recognition results in other languages as well. Therefore we also concentrated on a tonal accent language: Japanese, where pronunciation variants are also based on accent.

First we created a rule corpus of approximately 2300 multilingual phonetic pronunciation rules which model segmental variation of pronunciation variants of English, Japanese and German.

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

Section 3 gives an outline of the performance using the language-specific rule sets in automatic segmentation.

Finally, results and future work are discussed in the last section.

2. EXPERT RULE SETS

The phonetic rules express on a broad phonetic level phenomena of phonetic reduction that occur within words and across word boundaries. Each language-specific rule set describes typical phonetic processes according to ([6]) such as assimilation of place and manner of articulation, assimilation of voicing, elisions, substitutions, vocalisation, consonant epenthesis or lenition.

These rules are used to automatically segment according to phoneme various corpora available from the BAS ([3]). Two native speakers of American English and Japanese produced the pronunciation rule sets (expert rules) for English and Japanese by combining corpus observations with standard phonetic literature.

The German rule set has already been described in ([6]). The extended version including dialectal variants is discussed in detail in ([8]).

2.1 Inventory of used symbols

The phonetic symbols that are used in the regional pronunciation rules are taken from the SAM-Phonetic Alphabet ([9]). The symbols stand for the phonemes of German. We clustered these symbols according to table 1.

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.

phone category	phonetic symbols
unreduced vowels	a,a:,V,A,A:,e,e:,I,1,i:,O,o:,Q,O:,M,7, U,u:,E,E:,9,2:,3:,Y,y:
reduced vowels	@, 6, {
diphthongs	aI, aU, OY, eI, OI, @U, I@, e@, U@
plosives	p, b, t, d, k, g, Q
fricatives	f,v,s,z,J\z\,S,Z,C,x,h,T,D,G,B
glides	j, w
nasals	m, n, N
liquids	l, r, 4

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

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).

The next subsections give a general description of the rule sets of American English and Japanese where they differ from former publications on this subject.

2.3 English Rule Set

All rules were compiled by using the rule descriptions in ([10], [11],[12],[13],[14],[15],[16]).

Rules for the unvoiced plosives: Unvoiced plosives are substituted according to their environment by their voiced counterpart (in voiced position), glottal stop (in wordend position), or their nasal counterpart (in velar position). It can be elided if wordend or can follow another plosive.

Rules for the voiced plosives: Voiced plosives can be deleted if wordfinal. They can be substituted by voiced fricatives of the same articulation place.

Rules for the voiced affricates: Voiced affricates may become voiceless, or even voiceless fricatives.

Rules for the unvoiced fricatives: Unvoiced fricatives are substituted according to their environment by their voiced counterpart (in voiced position). Fricatives can be substituted by plosives of the same articulation place.

Rules for the phoneme z: The voiceless counterpart in wordend position is substituted by z.

Rules for the elision of h: The phoneme h is elided in syllable initial position if followed by a vowel or a semivowel.

Rules for the phoneme l: l can be elided between vowels or when it follows liquids.

Rules for the phoneme r: r can be elided if following a rounded vowel or diphthong in wordfinal position and before dental consonants or semivowels. It can be substituted by a semivowel in combination with the vocalized form of R.

Rules for the phoneme j: j can be inserted after voiced plosives.

Rules for the phoneme w: w can be inserted before bilabial plosives.

Rules for diphthongs: Diphthongs may be centralized or even monophthongized if followed by r.

Rules for the elision of @: @ is deleted in unstressed syllables.

Rules for unreduced long vowels: Unreduced long vowels are shortened in weak positions.

Rules for rounded vowels: Rounded vowels are often unrounded.

Rules for checked central vowels: Vowels may often become more centralized in word final position.

2.4 Japanese Rule Set

All rules for Japanese pronunciation change are based on ([17], [18], [19], [20], [21], [22], [23], [24]).

Rules for the archiphoneme N/: The archiphoneme N/ is realized in the following way, according to its environment: as /m/ before labial, as /n/ before alveolar, as /N/ before velar, as uvular /N\/ before vowel, approximant, fricative or in word final position.

Rules for the phoneme /M/ to [1]: The phone [1] is a phonologically conditioned allophone of the phoneme /M/. The phoneme /M/ is realized as high central unrounded [1] after the alveolar fricative and affricate. The phoneme /M/ may be realized as [1] after the /n/ or /j/. The phoneme /M/ may be realized as [1] after the tap /4/ (spoken by young girls).

Rule for /M/ to /1/: The sequence of the phonemes /Mi/ may be realized as [1i].

Rules for /M/-deletion: The vowel phoneme /M/ may be deleted in some voiceless environments.

Rules for the unaccented vowel deletion: The high unaccented vowel phonemes may be deleted between voiceless consonants.

Rules for the phoneme /M/ to nasal: The phoneme

/M/ after the voiceless consonant may be realized as [n] before /n/, as [m] before /m/, as [N] before /N/.

Rules for the long vowels (these rules do not apply to the morpheme boundary): The sequence of the phonemes /ei/ and /7M/ are realized as long vowel /ee/ and /77/ respectively.

Rules for vowel devoicing (applied to the unaccented vowels): The high unaccented vowel phonemes /i, M/ and the allophone [1] of /M/ are devoiced between voiceless consonants.

Rules for the vowel devoicing of /7, A, e/: The non-high unaccented vowel phonemes /7, A, e/ may be devoiced in some voiceless environments.

Rules for the glottal stop: The glottal stop [?] may occur after the short vowel before the pause.

Rules for the /b/ to voiced fricative [B]: The phoneme /b/ may be realized as bilabial voiced fricative [B] between vowels.

Rules for the /g/ to voiced fricative [G]: The phoneme /g/ may be realized as velar voiced fricative [G] between vowels.

Rules for the phoneme /t/ to [cs\]: The phoneme /t/ is realized as voiceless affricate of palatal plosive and alveolo-palatal fricative [cs\] before /i/.

Rules for the /d/ to [J\z\], [z\], [dz], [z]: The phoneme sequence /di/ is realized as voiced affricate of palatal plosive and alveolo-palatal fricative [J\z\] at the word onset and after /n/, as alveolo-palatal voiced fricative [z\] after the vowel. The phoneme sequence of /dM/ is realized as [dz1] at the word onset and after /n/, as [z1] after the vowel. The phoneme sequence of /dj/ is realized as voiced affricate [J\z\] at the word onset and after /n/, as alveolo-palatal fricative [z\] after the vowel.

Rules for the conditioned allophone of /h/: The phoneme /h/ is realized as [p\] before /M/, as [C] before /i/. [p\] and [C] are phonologically conditioned allophones of the phoneme /h/.

Rule for the conditioned allophone of /s/: The phoneme /s/ is realized as [s\] before /i/. [s\] is the phonologically conditioned allophone of /s/. The phoneme sequence /sj/ is realized as [s\].

Rules for the conditioned allophone of /dz/ to [J\z\], [z\], [dz], [z]: The affricate phoneme /dz/ before /i/ is realized as [J\z\] after /n/ or at the word onset, as [z\] after vowel. The affricate phoneme /dz/ before /M/ is realized as [dz] after /n/ or the word onset, as [z] after the vowel, /M/ is realized as [1] after [z] and [dz].

Rules for the /M/-deletion: The phoneme /M/ after [p\] may be deleted before vowel /i/, /A/, /e/ or /7/.

Rules for the /z/ to [dz], [J\z\]: The phoneme /z/ before /i/ is realized as [J\z\] at the word onset. The phoneme /z/ is realized as [dz] at the word onset. The phoneme /z/ may be realized as [J\z\].

Rule for the /t/ to [ts]: The phoneme /t/ is realized as affricate [ts] before /M/, and /M/ is realized as [1] after [ts].

Rules for the palatalization: The phoneme /p/ is palatalized to [p\] before /i/. The sequence of phonemes /pj/ is realized as [p]. The phoneme /b/ is palatalized to [b\] before /i/. The sequence of phonemes /bj/ is realized as [b\].

Rule for the palatalization of /d/: The phoneme /d/ may be palatalized as [d\] before /i/ or /M/.

Rules for the palatalization of /k/: The phoneme /k/ is realized as [k\] before /i/. The phoneme sequence of /kj/ is realized as [k\].

Rules for the nasalization with palatalization of /g/: The phoneme /g/ at word onset is palatalized to [g\] before /i/. The phoneme /g/ before /i/ is nasalized and palatalized to [n\] after vowel. The phoneme sequence of /gj/ is realized as [N\].

Rule for the palatalization of /t/: The phoneme sequence /tj/ is realized as [cs\].

Rules for the palatalization of /z/: The phoneme sequence of /zj/ is realized as [J\z\] at the word onset, as [z\] after the vowel.

Rule for the palatalization of the tap /4/: The phoneme /4/ is palatalized to [4] before /i/. The phoneme sequence of /4j/ is realized as [4\].

Rules for the palatalization of /m/: The phoneme /m/ is palatalized to [m\] before /i/. The phoneme sequence of /mj/ is realized as [m\].

Rule for the tap /4/ to [l]: The phoneme /4/ is realized as [l] after /n/. The phoneme sequence of /n4/ before /i/ is palatalized to [n\l\]. The phoneme sequence of /n4j/ is realized as [n\l\].

Rule for the /g/ to [N]: The phoneme /g/ in intervocalic environment is realized as [N].

Rule of devoicing for geminated /g/: The geminated phoneme sequence of /gg/ may be devoiced to [kk].

3. AUTOMATIC SEGMENTATION RESULTS

Simultaneously the labellers segmented and labeled a portion of the Verbmobil II speech corpus to provide material for the automatic training of re-write rules.

For our purpose, we trained MAUS (([1]), ([2])) on English and Japanese acoustic models and substituted the German rewrite pronunciation rules by English and Japanese expert rules. Although we did not include tonal accent variation in our rule set for Japanese, the automatic segmentations showed a similar improvement as for German and English data.

Language	correspondence in percent
English	89.43%
Japanese	83.46%

Table 2: Comparison of automatic segmentation and manual segmentation of the same utterances compared to the mean correspondence of human labellers

4. CONCLUSION

This paper gave an outline of pronunciation rules for English, German and Japanese.

Using these rules we have shown how multilingual pronunciation rules can influence automatic segmentation of speech. 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.

References

[1] Kipp, A. : Automatische Segmentierung und Etikettierung von Spontansprache; Shaker Verlag Aachen 1999.

[2] Beringer, N.; Schiel, F. (1999) Independent Automatic Segmentation of Speech by Pronunciation Modeling. Proc. of the ICPhS 1999. San Francisco. August 1999. pp. 1653-1656

[3] F. Schiel (1998): Speech and Speech-Related Resources at BAS; Proceedings of the FIRST INTERNATIONAL CONFERENCE ON LANGUAGE RESOURCES AND EVALUATION 1998, Granada, Spain, pp. 343-349.

[4] Burger, F. Schiel (1998): RVG 1 - A Database for Regional Variants of Contemporary German; Proceedings of the FIRST INTERNATIONAL CONFERENCE ON LANGUAGE RESOURCES 1998, Granada, Spain

[5] Kipp, A., Wesenick, B., Schiel, F. (1997): Pronunciation Modeling Applied to Automatic Segmentation of Spontaneous Speech; in: Proceedings of the EUROSPEECH 1997, Rhodos, Greece, pp.1023-1026.

[6] M.-B. Wesenick (1996): Automatic Generation of German Pronunciation Variants; in: Proceedings of the ICSLP 1996. Philadelphia, pp. 125-128, Oct 1996.

[7] Beringer, N.; Schiel, F. ; Regel-Brietzmann P.: German Regional Variants - A Problem for Automatic Speech Recognition? in: Proceedings of the ICSLP 1998. Sydney, Dec 1998.

[8] Beringer, N.; Neff M.: Regional pronunciation variants for automatic segmentation; in Proceedings of the SECOND INTERNATIONAL CONFERENCE ON LANGUAGE RESOURCES AND EVALUATION 2000, Athens, Greece.

[9] http://www.phon.ucl.ac.uk/home/sampa/home.htm

[10] http://www.slanguage.com/

[11] http://www.ling.upenn.edu/phono_atlas/

[12] http://www.uta.fi/FAST/

[13] Tajchman, Gary, Jurafskz, Fosler, Eric. Learning Phonological Rule Probabilities from Speech Corpora with Exploratory Computational Phonology. Submitted to ACL-95.

[14] Stewart, Darrzl, Hanna, Philip, Ming, Ji, Smith F. Jack. An improved Method of Automatically Generating Alternative Lexicon Pronunciations. ICPhS99, San Francisco.

[15] Teaching American English Pronunciation, Avery and Ehrlich Oxford University Press

[16] The Carnegie Mellon Pronuncing Dictionary V0.1. Carnegie Mellon University. 1993. Akamatsu, Tsutomu. 1997. Japanese Phonetics: Theory and Practice. Muenchen / Newcastle (Lincom).

[17] Amanuma et al. [1978]1991. Nihongo Onseegaku. Tokyo (Kuroshio).

[18] Hinds, John. 1986. Japanese. London / N.Y. (Routledge).

[19] Homma, Yayoi. 1973. An acoustic study of Japanese vowels. Onsee no Kenkyuu [Study of Sounds]. Tokyo (Phonetic Society of Japan) 16:347-368.

[20] Kawakami, Shin. 1977. Nihongo Onsee Gaisetsu [An Outline of Japanese Phonetics]. Tokyo (Oofuusha).

[21] Shibamoto, Janet S. 1985. Japanese Women's Language. Orlando etc. (Academic Press).

[22] Tsujimura, Natsuko. 1996. An Introduction to Japanese Linguistics. Cambridge / Oxford (Blackwell).

[23] Vance, Timothy. 1987. An Introduction to Japanese Phonology. Albany (State Univ. of N.Y. Press).

[24] Wenck, Guenther. 1966. The Phonemics of Japanese: Questions and Attempts. Wiesbaden (Harrassowitz).