Rule-based lexical modelling of foreign-accented pronunciation variants

Stefan Schaden
Institute of Communication Acoustics
Ruhr-Universität Bochum
D-44780 Bochum, Germany
schaden@ika.rub.de

Abstract

This paper describes a novel approach to generate potential foreign-accented phonetic transcriptions using phonological rewrite rules. For each pair of a native language (L1) and a target language (L2), a set of postlexical rules is designed to transform canonical phonetic dictionaries of L2 into adapted dictionaries for native L1 speakers. Some general considerations on the design of such a rule-based system are presented.

1 Introduction

Pronunciation dictionaries are a crucial component of speech recognition and speech synthesis systems, as they form the link between the acoustic and symbolic level of automatic speech and language processing. Typically, each entry in a lexicon is assigned a phonetic transcription that represents its *canonical form*, i.e. its standard pronunciation in the language the system is designed for.

Canonical lexicons, however, have the general drawback that every marked deviation from the standard form will lead to a mismatch between lexicon transcription and actual pronunciation. In Automatic Speech Recognition (ASR), this may cause a significant decline of the recognition performance.

In recent years, a number of approaches to compensate for this mismatch by various *lexical adaptation* techniques have been proposed (for an overview see Strik, 2001), e.g. by adding alternative pronunciation variants to the lexicon, by generating these variants using phonological rules, or by building pronunciation networks. Usually these techniques are applied to model frequently occurring stylistic variations such as within-word or cross-word assimilations or elisions in informal speech.

It is the aim of our current research to extend the lexicon adaptation approach from intra-lingual variation to the domain of foreign-accented pronunciation. Non-native speakers frequently produce variants that deviate markedly from the canonical form. They are characterized by phenomena such as changes in allophonic realizations, phoneme shifts, word stress shifts, or alternations in syllable structure caused by epenthesis or deletion of speech sounds. A primary (though not the only) source of these mispronunciations is a transfer of phonetic elements and rules from the speaker's native language onto the target language.

The idea to model these errors by lexicon adaptation is based on the assumption that for each language direction – i.e. a pair of a native language (L1) and a target language (L2) – a number of characteristic pronunciation errors can be identified. Although there is a considerable range of inter-individual variation even for speakers with the same native language background (due to variables

This study was carried out at the Institute of Communication Acoustics, Ruhr-University Bochum (Prof. J. Blauert, PD U. Jekosch). It is funded by the *Deutsche Forschungsgemeinschaft* (DFG).

such as L2 proficiency, age, education, dialectal origin, etc.), it is assumed that common mispronunciations can be formulated as rewrite rules to generate prototypical interlanguage transcriptions.

Currently, the languages investigated are German (GER), English (ENG), and French (FR) in different L1/L2 combinations; an extension to additional languages is envisaged.

A prototype of a task-specific rule interpreter was implemented, and phonological rule sets for the language directions ENG → GER, GER → FR, GER → ENG, and FR → GER were developed and are constantly being updated and modified. These rules are based on actual pronunciation variants observed in a non-native speech database (see below). They are currently limited to the domain of foreign city names; yet it is expected that the findings can be generalized to other lexical domains.

2 Speech data

For the purposes of this research project, a speech database of non-native speech was built up. The data collection and the experimental setting for the recordings are described in full detail in Schaden (2002). It includes non-native pronunciation variants of city names/town names from five European languages (English, German, French, Italian and Dutch) spoken by native speakers of English, German, French, Italian, and Spanish. In order to account for potential inter-speaker variability, at least 20 speakers per native language were recorded. The recordings included both a reading task and a repetition task, using the same words for both tasks. This allows to spot the particular influence of spelling pronunciation on the production of the speakers.

3 Inter-speaker variability

As a general prerequisite for modelling pronunciation variation of any kind – be it speaker-specific, dialectal, or foreign-accented –, knowledge about the target forms to be modelled is required: For obvious reasons, pronunciation rules can only be established after having specified the target rule output. The required knowledge can either be inferred from speech data or extracted from the literature.

However, contrary to intra-lingual (e.g. dialectal or stylistic) variants, which are relatively well documented, the definition of appropriate target forms is not as straightforward in the case of non-native speech. A primary reason for this is the heterogeneity of the speaker group: While e.g. in dialectal speech, phoneme shifts and other deviations from the standard are relatively consistent over large speaker groups, foreign-accented pronunciations will vary considerably according to individual speaker characteristics (some of which were mentioned above). Although it is certainly possible to detect prevalent pronunciation errors for speakers of the same L1, a common native language background does not constitute a homogenuous non-native speaker group. It is therefore not adequate to model variants for a particular L1/L2 combination by adding just *one* single prototypical L1-specific variant for each L2 lexicon item. Rather, there is a continuum of potential mispronunciations ranging from slightly accented forms with only minor allophonic shifts up to heavily accented pronunciations with extreme deviations from the L2 standard.

4 Prototypical accent levels

In order to model inter-speaker variability, it is not a practical aim to take *all* potential variants into account. Instead, a different approach is pursued: As a working hypothesis, it is suggested to break up the continuum into discrete categories by defining a number of prototypical foreign-accented pronunciations per word, where each of these prototypes represents a particular *accent level*. Accent levels range from near-native pronunciation to gross mispronunciations. Currently, the model is based on four accent levels, where higher integers indicate increasing deviations from the canonical L2 pronunciation:

Accent level	Description
AL 0	Canonical L2 pronunciation (no accent)
AL 1	AL 0 + Minor allophonic deviations
AL 2	AL 1 + Allophone/phoneme substitutions
AL 3	AL 2 + Partial transfer of L1 spelling pronunciation (GTP correspondences) to L2
AL 4	Almost full transfer of L1 spelling pronunciation to L2

Table 1: Accent levels

Accordingly, the rule system is built up in such a way that for each input word, multiple variants representing the accent level prototypes can be generated. By this, the probability that one of the automatically generated variants approximates the actually observed pronunciation is increased. It is expected that for speech synthesis and recognition purposes, a sufficient approximation to actually occurring variants can be achieved in this way.

Furthermore, it is attempted to design a modular rule system that operates *incrementally*, as indicated above in Table 1: Each rule module models a specific accent level, and a sequential application of the modules should ideally generate phonetic forms of increasing accent degrees.

5 Modelling phoneme substitutions

It is one of the most salient characteristics of foreign-accented pronunciation that non-native speakers tend to substitute L2 speech sounds by similar, yet not identical L1 equivalents. The first idea that suggests itself in order to model these substitutions are phoneme/allophone *mapping tables* that replace particular L2 sounds by similar speech sounds from the L1 inventory. However, simple context-free phoneme mapping is problematic in at least two respects:

First, for many L2 sounds it is not clear what the 'best' L1 equivalent is. Acoustic or articulatory proximity of an L1/L2 allophone pair is not always a reliable predictor of the sound shifts that speakers actually produce. Secondly, our data clearly indicates that in many cases, the choice of the substitution phoneme/allophone is related to the phonetic or graphemic surroundings of the substituted phoneme. Therefore, in order to restrict their application to appropriate contexts, most rewrite rules require context conditions on the phoneme level and/or on the orthographic level (see below).

5.1 Phonemic context conditions

Rules that do not require information from linguistic levels other than the phoneme/allophone level can be formulated using the established rule notation adopted from generative phonology:

$$X_{L2} \rightarrow Y_{L1} \quad / \quad LC \_\_ RC$$

Here, a phoneme/allophone X_{L2} (element of the L2 inventory) is substituted by Y_{L1} (element of the L1 inventory) if the immediate left and right contexts LC and RC are valid. In the rule system presented here, X and Y are usually phoneme or allophone segments. In cases where a rule applies to entire phoneme classes, X and Y (likewise LC and RC) can also be written as phonetic feature arrays:

$$\begin{bmatrix} +\text{obstruent} \\ +\text{voiced} \end{bmatrix} \rightarrow \begin{bmatrix} +\text{obstruent} \\ -\text{voiced} \end{bmatrix} \quad / \quad \_\_ \quad \#$$

This is a useful abbreviatory device if a generalizable phonological rule of L1 is transferred to L2 (e.g. the German rule of final obstruent devoicing applied to English).

5.2 Graphemic constraints

In the particular case of *read* speech, mispronunciations by non-natives are often triggered by a projection of L1 grapheme-phoneme correspondences to L2. Here, speakers apply letter-to-sound rules of their native language to L2, provided that L2 target words contain orthographic sequences that allow such a transfer.

One technique to model this particular error type is the application of L1 grapheme-to-phoneme (GTP) converters to L2 orthographic input. This approach was explored e.g. by Cremelie & ten Bosch (2001) in a speech recognition experiment in the proper names domain. But although GTP conversion by L1 rules proved to be beneficial in this recognition scenario, it does not model speaker behavior adequately, since non-native pronuncia-

tion variants are rarely based on unmodified L1 GTP rules applied to L2. There are various reasons for this: Many speakers have an awareness of at least some pronunciation rules of L2 (e.g. the pronunciation of German <sch> as [S] is familiar to many European speakers). Secondly, for some L2 orthographic sequences, a straight transfer of L1 GTP rules would yield 'unpronouncable' clusters; hence the L1 rules can only be applied to parts of the L2 grapheme string.

As an alternative to letter-to-sound conversion by L1 rules, where the entire string is *globally* transcribed according to L1 letter-to-sound rules, it is therefore suggested to apply *graphemically constrained phoneme substitutions* in order to model spelling pronunciation errors *locally*. In this rule type, phoneme substitutions are tied to particular graphemic representations. For example, native English speakers frequently mispronounce German [v] as [w]. This substitution, however, only occurs if [v] is orthographically represented by <w>, while [v] represented by orthographic <v> fails to undergo this rule. Such a restriction can be formalized as follows:

```
PHONEME LAYER:    [v]   →   [w]
                   |
GRAPHEME LAYER:   <w>
```

For this rule type, it is required that the phoneme string is aligned with the grapheme string in order to map each phoneme correctly to the grapheme segment or cluster representing it. A rule-based grapheme-phoneme alignment module for English, German, and French is therefore included in the presented rule system.

According to the experience gained up to now, graphemically constrained substitution rules are capable of modelling a wide range of typical spelling pronunciation errors adequately – from insignificant misreadings up to strongly accented variants that follow almost completely the L1 letter-to-sound-rules. Furthermore, this approach has the advantage over GTP conversion by L1 rules that all errors (reading errors included) can be modelled *postlexically* without interfering with the canonical input lexicon.

6 Summary, future extensions

In its present status, the rule system outlined in the previous sections includes sets of postlexical accent rules for English, French, and German in all L1/L2 combinations. Currently, the number of rules per language direction is 80-100. The rules generate several prototypical foreign-accented variants per input word, using phoneme substitution rules of the type described above.

Future extensions of the rule system will focus on two issues: (i) Modelling shifts in word stress patterns that can frequently be observed in non-native pronunciation variants (L1 stress patterns transferred to L2); (ii) the role of morphemes and lexemes which are part of the learned vocabulary (of speakers with some L2 proficiency). The data indicates that these elements (e.g. *-stein* or *-bach* in German city names) are less susceptible to accented pronunciation and may thus escape the effects of the phoneme substitution rules. Furthermore, an extension to additional (native and target) languages is scheduled. Rule sets for Italian (as L1 and L2) and Dutch (as L2 only) will be set up.

For an evaluation of the automatically generated pronunciation variants, a comparison to the pronunciations of new (i.e. non-database) speakers as well as speech recognizer performance tests using the adapted dictionaries will be essential.

References

Cremelie, N. and L. ten Bosch. 2001. Improving the Recognition of Foreign Names and Non-Native Speech by Combining Multiple Grapheme-to-Phoneme Converters. *Proceedings ISCA ITRW Workshop 'Adaptation Methods for Speech Recognition'*, Sophia Antipolis, France [on CD-ROM].

Schaden, S. 2002. A Database for the Analysis of Cross-Lingual Pronunciation Variants of European City Names. *Proceedings Third International Conference on Language Resources and Evaluation (LREC 2002)*, Las Palmas de Gran Canaria, Spain, Vol. 4, 1277–1283.

Strik, H.. 2001. Pronunciation Adaptation at the Lexical Level. *Proceedings ISCA ITRW Workshop 'Adaptation Methods for Speech Recognition'*, Sophia Antipolis, France [on CD-ROM].